The Green Book

A Guide to Effective Graduate Research in African Agriculture, Environment and Rural Development

The Green Book:

A Guide to Effective Graduate Research in African Agriculture, Environment, and Rural Development

Edited by

Bharati K. Patel, Kay Muir-Leresche, Richard Coe and Susan D. Hainsworth

The African Crop Science Society
P.O. Box 7062, Kampala, Uganda

2004

This publication is an output from the FORUM for African Agricultural Resource Husbandry Programme (1991–2002), Food Security Division, The Rockefeller Foundation.

The Green Book is published with the financial assistance of the Rockefeller Foundation, which is knowledge-based, global foundation with a commitment to enrich and sustain the lives and livelihoods of poor and excluded people throughout the world. This guide is a product of the Foundation's efforts to support activities aimed at informing policy development, programme design and resource allocation by providing local level, multifaceted information on food, work, health and other human conditions that are needed to understand and address the root causes of poverty.

Patel, Bharati K., Muir-Leresche, Kay, Coe, Richard and Hainsworth, Susan D. (Eds.). 2004. *The Green Book: A Guide to Effective Graduate Research in African Agriculture, Environment, and Rural Development*. The African Crop Science Society, Kampala, Uganda. 248 pp.

ISBN 9970-866-00-1

This publication houses a CD-ROM that contains the printed edition, together with extensive appendices.

Editorial Secretariat, *African Crop Science Society*, Faculty of Agriculture and Forestry, Makerere University, P.O. Box 7062, Kampala, Uganda.

Contents

Acknowledgements

Along the route to publication the editors and authors were assisted by many people without whose help the Green Book could not have been published. We thank all of them and particularly:

Roger Stern for his wise council and assistance to the authors of Part 4.

WARDA – The Africa Rice Center and the Technical Centre for Agricultural and Rural Cooperation **(CTA)** for their permission to use the full text of Stapleton *et al.* (1995) as Appendix 5, and the World Agroforestry Centre **(ICRAF)** for making this happen.

The **World Agroforestry Centre** for the use of Appendices 9, 10 and 11 on the CD.

Paul Woomer and **Adipala Ekwamu** for their PowerPoint presentations Appendices 1–4, 6 and 7 on the CD.

Eric McGaw for Appendix 8.

John Tenywa at the African Crop Science Society **(ACSS)** for checking the references.

TNG Sharma for patiently co-ordinating all the e-mails and for his editorial assistance.

Deanna Hash for helping us put the book on bed.

Pragati Offset Private Limited for their kind cooperation at all stages of design, typsetting and printing.

Heritage Media for holding our hands as together we developed and produced the CD.

We would like to keep The Green Book relevant. Please help us to update the book and CD with examples, references and any comments, suggestions and criticisms you may have. Please e-mail the African Crop Science Society acss@starcom.co.ug for the attention of Dr John Tenywa, who will forward your comments. Credit will be given to inputs received in subsequent editions.

Foreword

Achieving food security and poverty alleviation in sub-Saharan Africa will require a large cadre of well-trained agriculturists, and development partnerships that derive their agenda from the needs and voices of the continent's populace. It will require team work, combined with a new set of leaders who use participatory approaches and tools relevant to Africa's socio-economic situation. Institutions in Africa must develop new ways of teaching with tools that can be adapted to the training needs of African scholars so that they can re-orient their students to: think creatively, value team work and partnerships, recognise that they belong to a global world, and that whatever they do, should contribute to the improvement of the welfare of African's people. Unfortunately African scholars and fieldworkers often lack reference materials that are based on local experiences.

The Green Book is intended to equip young African scholars with guidance for their thesis study, and to prepare them for the leadership roles that will be expected of them on completion of their studies. The book is written as a guide and examples from the continent are used to highlight the issues. It implores students to always be aware of the bigger picture – the development goal – and to realise that he/she is not operating in a box. We are all part of a community, and indeed, of a global world that is changing – and we must adjust in order to cope with the changing needs of society. This first edition is relevant to all students, but is biased towards those in eastern and southern African universities. It is hoped that the second edition, that will be published in both French and English, will be broader in scope and will include West African reality. The editors welcome all comments and suggestions for improvement in future editions.

The Green Book is linked to many other books, organisations, institutions, and dedicated persons, we acknowledge their contributions and those of all contributors. This book is published under the aegis of the African Crop Science Society (ACSS) which is pan-African in its coverage. The ACSS was the brainchild of staff at Makerere University where it is based, thus, it was a natural choice as publisher. The ACSS has been closely associated with the Rockefeller Foundation since its inception and is grateful for the provision of funds to develop and publish the Green Book.

Professor Adipala Ekwamu
Founder President of the African Crop Science Society


MEET GEORGE AND CHARITY..
GRADUATES, JUST LIKE YOU!
ALEYA

The Green Book and how to use it

There are many books and materials available to graduate students but most are targeted to students in the developed world and very few are specifically directed at rural development and African students. Graduate studies play an important role in contributing to research and development in African agriculture. Combining research and development, working for and with rural people, and making a real difference to the future requires special skills and approaches that are not often required in the industrialised world. Students need to be able to balance their academic requirements with the needs of projects and rural participants. They need the skills and adaptability to work in interdisciplinary teams and to understand what is required both for their specific discipline and the broader requirements.

This book arose from the experience of the editors who have worked closely with agriculture, environment and rural development postgraduate students in African universities. All the authors who were invited to contribute have wide experience and close connections with universities mainly in eastern and southern Africa.

The information is presented in four main parts, each containing several chapters. Each chapter's title page has a set of bullet points in the left-hand margin that guide you to the chapter's content. These four main parts are followed by a short chapter that is intended to inspire further efforts. Within the text are hints and highlights printed in italics and bold type. These are not an excuse to skip the main text, but rather as aide-memoires to points of emphasis in the chapter.

The Green Book CD includes the full text linked to additional information and resources in publications, software and PowerPoint presentations. Cross references in the text and to the CD are printed in bold green type.

Enjoy the Green Book, we hope you will find it useful, and that it will fulfil part of the Declaration on the African Universities made in Nairobi on 9 February 2001.

'Through the promotion of research and free enquiry; the open contestation of ideas, and the appreciation and tolerance of difference, African universities must generate and disseminate knowledge and understanding, foster the values of openness and respect for merit, and enrich the general quality of the social life of their communities'.

The Editors


RESEARCH
EXTENSION
FARMING
Service Industries
Marketing
MANUFACTURING
Tourism
FORESTRY
?
DECISION TIME!
ALEYA

1.1 Research? What, why and how?

Tony Greenfield

- **Why do you want to do research? Have you chosen a research project? Do you feel strongly about an outstanding problem or will your study be guided by your professor or research organisation?**
- **Research is about solution of problems but they must be problems that attract funds, a supervisor and collaborators**
- **Will your project fit into a wider strategy? Will it make a valid contribution to the big picture?**
- **Your abilities will yield surprises, results and benefits that nobody predicted. You can be creative**
- **Write clearly and honestly. Avoid deceit and obfuscation**

> *'We have three principal means: observation of nature, reflection, and experiment. Observation gathers the facts, reflection combines them, and experiment verifies the results of the combination. It is essential that the observation of nature be assiduous, the reflection be profound, and that experimentation be exact. Rarely does one see these abilities in combination. And so, creative geniuses are not common.'*
>
> **Denis Diderot (1753)**
> On the Interpretation of Nature

Introduction

You are stuck. You are a graduate on the ground floor and you have been there for several years. There's a hard ceiling to your career and you can't climb the staircase to the floor above without showing a ticket to pass. What is that ticket? It is, at least, a portfolio of published papers. Better, it is a research degree, a master's or a doctorate. Your work experience has sparked many worthy ideas that you know you could develop with well-supported research. So, how do you start?

Even if research doesn't immediately attract you, you realise that a higher degree is essential for any promotion. You may be working in a government department or in some non-governmental organisation or in a company that does not have a research mandate, yet you know that an MSc or a PhD will enable you to climb the staircase to higher management, greater responsibility, and wider opportunities. Research must be part of this process, for it is through doing research that you will learn to learn, to observe, and to marshal new knowledge to the greatest effect.

Or: You are about to graduate. Your head is bursting with knowledge and ideas. The world, you hope, is waiting for you to make it a better place. How will you do this? There must be a job just right for you. Farm management? Well, you'd have to start as a junior and work your way up. How long would that take? Supplier of materials such as fertilizers, pesticides, seeds, machinery? That might be quite interesting, even exciting: travelling about in a company car as a representative and discussing the needs of your customers, the farmers. Or even politics? Could you persuade governments, local, national or international, to direct their resources to where, in your opinion, they are most needed?

But: "You've done well," your professor says. "You might consider staying here and doing research."

Research? Now here's a chance to use your knowledge and ideas to make the world a better place, to improve the quality of the lives of fellow humans, to give them the security of good food supplies, to protect and improve the environment. At the same time you will gain a qualification that may assure you of an interesting and well-paid(?) career; you will publish reports of your work, present them at international conferences; you'll feel good. You might pursue your theoretical studies in Europe and do your field studies in Africa where the farmers' problems exist. If your work is in a university, you will mix with researchers in other sciences, in engineering, in economics and in business. You could teach, become a professor. You might even create a new product or process that will lead you to commercial as well as academic success.

"Show me the lab," you say. "I'll put on my white coat and start to work."

But wait. What will you do? Who will pay you? Who will pay for your research? Will you be paid at all? Perhaps, but not a lot. Are you sure you have the skills for research and understand the procedures?

You will need a grant: money to live on, funds for materials and equipment, travel, conference fees. So you must persuade somebody that what you propose to do is worthy of their investment. So here's a skill that you might not have contemplated: how to write a research proposal.

To me, research is an art aided by skills of enquiry, experimental design, data collection, measurement and analysis, by interpretation, and by presentation. A further skill, that you can learn and develop, is creativity or invention. These are a few of the many skills and methods of research that you should know and understand.

So this book is for you, for graduate students, whether MSc or PhD, at universities throughout Africa if you want to work in agriculture, in natural resources, in rural development or the environment, with people whose livelihoods depend on the land. It is for your supervisors too: those who are appointed to guide you through your first major research project.

Research is a big subject and it would not be possible to write a single volume about it in any depth. This book is intended to be a general reference on all aspects of research methods and should be used as notes for guidance. Its content is intended to be fairly simple and easily intelligible by most readers. There are references to more substantive texts and to websites, and plenty of additional information on the CD.

The many viewpoints and components of research methods demanded that several contributors would be needed. Fortunately, there are enough qualified people in universities, research organisations and consultancy, who volunteered eagerly to write one or more chapters each. The editors asked them to write in a light style that you could read easily so that you can pick up the general themes. We have tried to make it a book not only of general guidance but also of relevance to the reality of current research and agriculture in Africa and to hopes for the future of this continent.

If there are parts that you don't understand, or that could be expressed more clearly, or if there are important omissions, please write to the editors or the publishers. Everything can be improved, especially the first edition of a book, and your opinions will help us.

These notes for your guidance have been divided into four main parts, with several chapters in each. Look through the contents list and see how the topics have been grouped. You may feel that some of the chapters are not for you. For example, do you know how to design a survey? Of course you do! You have often been asked to answer survey questions so you know what questions should be asked. But there is much more to a survey than that. How do you choose your target population and then a sample size? How do you ensure that you get a good response? How do you code the questionnaire for analysis? How do you analyse and interpret results. Read **Chapter 4.4** and learn. How do you attract financial, material and intellectual support? There are chapters to help you with these. You will run into difficulties. You will find problems of management, of resources, of people. There's a chapter telling you who can help. I suggest you read it before you meet those problems. There are also chapters on planning your work, about keeping documents, about examining your research process and keeping it on course.

Glance quickly at the chapter on data analysis and you may think 'I can leave that until much later, when I have some data to analyse.' Scientific method is about observing the world and collecting information so that you can understand the world better. The way in which you do this must surely depend on how you will process the information when you have collected it. The data you collect will depend on how you will analyse the data. Analysis is an essential feature of research and you will make easier progress with your research the more you understand analysis. To some people it is hard and daunting. They would prefer to ignore it. To other people it is a challenge. Whichever is your viewpoint, make it a challenge and face it now. Honestly, the more you understand about how you will analyse and interpret data, the better will be your planning and management of the way you collect it. The design of a good experiment depends on how the data from the experiment will be analysed.

How do you communicate your results so that other people will notice and act on your advice? Whatever research you do, you must present your results: in a thesis or dissertation, in reports and published papers, and in stand-up talks to live audiences. There are many books about presentation and some are recommended. The CD also contains helpful guidance.

Your problem within a strategy

Your first problem is to find a problem; the solution of problems is what research is about. You must also find a problem that is of interest to other people: those who might provide your funds and those who might support you, supervise you and collaborate with you. Does that restrict you too much? You have already chosen to work in agriculture, natural resources, the environment and rural development so you have defined the domain in which to find a research problem. Thousands of research problems have already been identified. These have led to research projects that have yielded results: published papers, dissertations and theses, books and, most important, changes in the way things are done out there on the land and in the lives of real people.

So, are there any problems left for you?

The more we know, the more we know we don't know

The more problems we solve the more problems we find that need to be solved. There is no shortage of problems, there never will be, even if you narrow your search down to problems of farmers.

The finding and expression of problems and their translation into research proposals are discussed at length in **Chapter 3.2.** But that chapter starts with 'You have decided on the area you want to research'. Have you?

There are strategies at work and you would do well to consider them because it is within these strategies that research projects are created, are inter-related and are funded. If you can identify a problem for your research that fits into an existing strategy, then you will have a good chance of approval by colleagues, by assessors of your ideas, by grant-funding panels.

The strategy is the big picture. Your project is just a part of it. But if you can understand how a strategy is constructed you will be well equipped to challenge it with such questions as: 'Has it been motivated by political and economic considerations rather than the needs of the farmer?' and to ensure that your research project makes a valid contribution.

Strategy is about answering two questions:

1. How will things be done in the future?
2. What changes and investments must be made to achieve that future?

Answers will be found by following nine concurrent steps (while you read this, think about the farmer and his problems and think about the research that you might do to solve his problems)

1. Understand how well the current products and practices meet current needs so that you can measure the gap in performance and identify the constraints against closing the gap.
   - Specify the needs:
     - eradicate hunger, reduce poverty and safeguard the environment
     - larger harvests through higher yields
     - better nutrition through improved food quality
     - empower people to build their own capacities, self-confidence and self-reliance
     - develop agriculture where water is very scarce, with poor and degraded soils, and where social infrastructure is weak or non-existent
     - provide subsistence farmers with opportunities to increase their income by pursuing opportunities for commercialisation.
   - Identify the constraints:
     - low rainfall
     - inadequate water storage and distribution
     - poor and degraded soils
     - weak or non-existent social infrastructure
2. Determine local trends along four main themes.
   - Changing customer needs (often driven by geopolitical changes such as climate changes, urbanisation, wealth, labour and other resources, irregular markets, and perceptions of what might be)
   - Legislation (genetic modifications, crop type restrictions, land access)
   - Science and technology
     - improved water use through water conservation
     - improved water use and drought tolerance in crop genotypes with plant breeding and biotechnology
     - improved farming machinery, processing, storage, transportation
     - improving social structure, education and health.
   - Competition (what products are wholesalers and distributors asking for and at what prices?).
3. Who are the farmer's current and future customers? (Self-sufficiency defines small-scale farmers' prime customer: themselves followed by local markets and cooperative marketing companies. Large-scale farmers must think about wholesalers, packaging companies, super-markets and exporters).
4. Extrapolate the science, technology, political and social structure, economy, environmental management and consumer trends to create the vision of rural transformation and of how African agriculture will look 3–10 years into the future.
5. Determine what farmers must do to continue to be successful in their future markets (change soil management and irrigation, crop varieties, harvesting, pest control, storage, processing, join co-operative partnerships).

6. Determine what investments may be needed to deliver these changes (investments in the research need to find out how to deliver the changes as well as investments in equipment and materials).
7. Convince farmers themselves that the changes are worth the investments and convince other providers of funds (governments, donors, marketing companies) that their support is needed and will, somehow, pay dividends.
8. Promote partnerships and sponsor projects that bring results of research to rural communities, farmers and their families.
9. Translate the investment plans into implementation plans.

A strategy that has room for you may already exist within the university department or in some research institution. With good fortune, enthusiasm and guidance, you will find or be directed to a problem that will occupy you for 2–3 years: the time you need to complete your post-graduate research.

But .... Your head is bursting with knowledge and ideas. The world, you hope, is waiting for you to make it a better place. "If the strategy is already defined, if my problem is but a small part of that strategy, if my proposal for research must be written so as to earn the approval of my supervisors and my fund granters and has to be so tightly defined as to guarantee a satisfactory conclusion within 3 years, where", you might ask, "is there any opportunity for my originality?"

Don't worry. You have been encouraged to do research because you have already shown a potential, if not already proved ability, for problem-solving. Your supervisors know that you can:

- Assimilate, analyse and evaluate complex information
- Identify key issues and principles
- Think critically
- Learn from mistakes
- Challenge established assumptions
- Avoid prejudices
- Take a broad view
- Think conceptually and creatively

These are abilities that you already possess – abilities that will ensure that, no matter how tightly defined, your research project will yield surprises, results and benefits that nobody predicted.

Creativity

'Creativity is what cannot wait, cannot stop, cannot back step: faster or slower, it always goes ahead: through, alongside, above, regardless of crises or systems.'

Jose Rodrigues Migues

There are many ways to solve a research problem. A formal procedure will often yield a solution, provided you keep an open mind and look for the unexpected.

If you look at the most outstanding of creative leaps in the history of science you will see that they were all founded on an irrationality of thought. Well known examples are: Watt's invention of the separate condenser for the steam engine as he strolled in the country; Poincare's theory of Fuchsian functions as he boarded a bus; Kekule's discovery of the benzene ring as he dozed by the fireside. So, be prepared to note any odd thought you might have at an unexpected time in an unexpected place. And don't discard unexpected results.

'If you do not expect the unexpected, you will not find it; for it is hard to be sought out, and difficult.'

Heraclitus of Ephesus
in Khan (1979)

'Just because something doesn't do what you planned it to do doesn't mean it's useless.'

Thomas Alva Edison

You don't have to wait for that magical moment or a bang on the head to spark an original idea. There are some well-established methods of intellectual discovery that you can apply to your problem:

- **Analogy.** Look for similarity between your problem and one for which you know the solution. Electrical circuits are perceived as water flowing through tanks, pipes, pumps and valves; brain function is studied by comparison with computers. The more remote your analogy is from your problem, the more creative will be your solution
- **By parts.** Break the problem into a series of sub-problems which you hope will be more amenable to solution
- **By random guesses.** Edison used it extensively and brain-storming is a modern version
- **Generalise.** If a specific problem is baffling, write a general version of it; an algebraic model leads to simplified solutions compared with tackling complicated arithmetic head-on
- **Add.** A difficult problem may be resolved by adding an auxiliary sub-problem
- **Subtract.** Drop some of the complicating features of the original problem; this is a trick used in simulation to make it more tractable
- **Particularise.** Look for a special case with a narrower set of conditions, such as tackling a two-dimensional example of a three-dimensional problem
- **Stretch or contract.** Some problems are more tractable if their scale or the range of variables is altered
- **Invert.** Look at the problem from the opposite viewpoint; instead of 'When will this train arrive at Nairobi?' ask 'When will Nairobi arrive at this train?'
- **Restructure.** In clinical studies we do not ask if a treatment will cure a disease, but if an inert treatment will fail to cure the disease
- **The method of Pappus.** Assume the problem is solved and calculate backwards
- **The method of Tertullus.** Assume a solution is impossible and try to prove why.

Check each of these approaches, asking yourself how you might bring it to bear on your problem. Then, if you need any more stimulation, read:

The Art of Scientific Investigation a book by W I B Beveridge published in 1950 but still, half a century later, stimulating to read; G Polya's *How to Solve It* offers practical recipes; and Arthur Koestler's *The Act of Creation* for a discussion of the working of the mind.

My personal approach to solving a problem that has defeated me at my desk:

- Go for a long and lonely walk in the country

  or
- Drowse in a hot bath................Inspiration comes.

Ethics of research

'Creativity is great but plagiarism is faster.'

Anon

If you are researching into some aspect of the environment, or into the development of genetically altered viruses for the control of crop pests or in some area of medicine, either human or animal, you will almost certainly have thought about ethical aspects of your intentions. But if your research is in some other area such as sociology, education, water storage and distribution or poor and degraded soils, you may think that there are no ethical questions for you to consider.

You would be wrong to think that.

Fraud is an obvious ethical matter but surprisingly so are experimental design, planning, management and execution; and so is publication.

If you know yourself to be thoroughly honest, you must be confident that you will never be deliberately unethical. Unfortunately, no matter how good a person you are and how well intentioned, there is the possibility, indeed it is very likely, that you will be inadvertently unethical, insomuch as you infringe the accepted code of research behaviour. Anybody who embarks on research is at risk of such inadvertent unethical behaviour. Avoidance demands good advice at all stages. Where will you find that advice?

Ethics, in its widest sense, is the set of principles of good human behaviour.

Most professional organisations have their own codes of conduct that are largely about the ethical standards that are expected of members. I have distilled the following points from several professional codes:

- It is unethical to conduct research which is badly planned or poorly executed
- Research must conform to generally accepted scientific principles based on adequately performed experimentation and on a thorough knowledge of the scientific literature
- Every research project should be preceded by careful assessment of predictable risks in comparison with foreseeable benefits
- In publication of the results of research preserve the accuracy of the results. Inaccurate and incomplete reports of experiments should not be accepted for publication
- The research proposal should always contain a statement of the ethical considerations involved
- Special caution must be exercised in the conduct of research that may affect the environment
- Within your chosen field, you must have an appropriate knowledge and understanding of relevant legislation, regulations and standards and comply with such requirements
- Have regard to basic human rights and avoid any actions that adversely affect such rights
- Accept responsibility for the social consequences of your work
- Seek to avoid being put in a position where you might become privy to or party to activities or information concerning activities that would conflict with your responsibilities
- Never cast doubt on the professional competence of another without good cause
- Do not lay claim to any level of competence that you do not possess
- Any professional opinion shall be objective and reliable
- You must not allow any misleading summary of data to be issued
- Views or opinions based on general knowledge or belief should be clearly distinguished from views or opinions derived from the statistical analysis being reported.

It is worth remembering that:

'Precise conclusions cannot be drawn from inadequate data.'
Pearson and Hartley (1966)
Biometrika Tables for Statisticians

Fraud

While much unethical science is inadvertent, caused mainly by poor management, there is a long history of scientific fraud reaching back several centuries. Charles Babbage, who was Lucasian Professor of Mathematics at Cambridge University (a chair held by many great scientists including Isaac Newton and Stephen Hawking), published a book in 1830 entitled *The Decline of Science in England*.

One chapter in his book was about scientific fraud under which he described four methods of fraud: Hoaxing, Forging, Trimming, and Cooking. To these I would add Obfuscation. For the first four, I cannot do better than quote him directly.

Hoaxing

'In the year 1788, M Gioeni, a knight of Malta, published an account of a new family of Testacea of which he described, with great minuteness, one species. It consisted of two rounded triangular valves, united by the body of the animal to a smaller valve in front. He gave figures of the animal, and of its parts; described its structure, its mode of advancing along the sand, the figure of the track it left, and estimated the velocity of its course at about two-thirds of an inch per minute. ... no such animal exists.'

There have been many more hoaxes since Babbage's day, including the saga of the Piltdown man.

Forging

'Forging differs from hoaxing, inasmuch as in the latter the deceit is intended to last for a time, and then be discovered, to the ridicule of those who have credited it; whereas the forger is one who, wishing to acquire a reputation for science, records observations which he has never made. ... The observations of the second comet of 1784, which was only seen by the Chevalier d'Angos, were long suspected to be a forgery and were at length proved to be so by the calculations and reasoning of Encke. The pretended observations did not accord amongst each other in giving any possible orbit.'

Statistical methods now exist to discover forged data. Examples may be found in industrial research and in clinical trials. If you are tempted to forge your data, be warned. A good examiner will detect your forgery and you will be humiliated.

There can be great pressure on students to complete a research project within the time specified by the university rules or before their grant expires. Under such pressure the student may be tempted to forge data which they have never observed. Or, if they have made some measurements that don't properly meet their expectations – they may be tempted to cook the results. Cooking is described below by Babbage.

Trimming

'Trimming consists in clipping off little bits here and there from those observations which differ most in excess from the mean, and in sticking them on to those which are too small ... the average given by the observations of the trimmer is the same, whether they are trimmed or untrimmed. His object is to gain a reputation for extreme accuracy in making observations .... He has more sense or less adventure than the cook.'

Cooking

'This is an art of various forms, the object of which is to give to ordinary observations the appearance and character of those of the highest degree of accuracy.

'One of its numerous processes is to make multitudes of observations, and out of these to select those only which agree, or very nearly agree. If a hundred observations are made, the cook must be very unlucky if he cannot pick out fifteen or twenty which will do for serving up.

'Another approved receipt, when the observations to be used will not come within the limit of accuracy, is to calculate them by two different formulae. The difference in the constants, employed in those formulae has sometimes a most happy effect in promoting unanimity amongst discordant measures. If still greater accuracy is required, three or more formulae can be used.

'It sometimes happens that the constant quantities in formulae given by the highest authorities, although they differ amongst themselves, yet they will not suit the materials. This is precisely the point in which the skill of the artist is shown; and an accomplished cook will carry himself triumphantly through it, provided happily some mean value of such constants will fit his observations. He will discuss the relative merits of formulae ... and with admirable candour assigning their proper share of applause to Bessel, to Gauss, and to Laplace, he will take that mean value of the constant used by three such philosophers which will make his own observations accord to a miracle.'

Obfuscation

Obfuscation means 'to make something obscure'. It is a deliberate act which is intended to convey the impression of erudition, of being learned, of great scholarship. Hence it is fraudulent. There is a style of academic writing, increasingly common in recent years, i.e., long-winded with long paragraphs, long sentences, long words, passive statements and tortuous structures (see **Part 3**). It is intended to deceive and it does so easily because the reader, even an examiner, is tempted to skim such verbosity and subsequently fears to confess he or she has not understood every word.

It is a trick that is apparent today in many academic papers and theses but it was not uncommon a hundred years ago:

'The researches of many commentators have already thrown much darkness on this subject, and it is probable that, if they continue, we shall soon know nothing at all about it.'

Mark Twain (1899)

... or even 400 years ago:

'The ill and unfit choice of words wonderfully obstructs the understanding.'

Francis Bacon (1620)
(in Novum Organum)

Perhaps some people can't help writing obscurely but if a post-graduate research student does so we should be suspicious.

'People who write obscurely are either unskilled in writing or up to mischief.'

Peter Medawar (1984)

Resource material and references

Babbage, C. 1830. *Reflections on the Decline of Science in England.* (Republished 1971.) Irish University Press, Dublin, Ireland.

Khan, C.H. 1979. *The Art and Thought of Heraclitus.* Cambridge University Press, Cambridge, UK.

Koestler, A. 1973. *The Art of Creation.* The Dell Publishing Co., New York, USA.

Medawar, P. 1984. *Science and Literature in Pluto's Republic.* Oxford University Press, Oxford, UK.

Pearson, E.S. and Hartley, H.O. 1966. *Biometrika Tqbles for Statisticians*, Vol. 1. Third edition. Cambridge University Press, Cambridge, UK.

Twain. Mark. 1899. *The Sciences* (September–October).

The Green Book HELPS YOU THROUGH...
MSc
PhD
WELCOME
ALEYA

1.2

Research: a path through the maze[1]

Tony Greenfield and **Tom Bourner**

- **There are many paths through the research maze. Some paths will pose more difficulties than others, present more limitations, and take you down fruitless side tracks; others will present greater insight and surprises**
- **Every path should have four parts to it: review; theory building; theory testing; and reflection**
- **A review will identify a gap in the literature and a problem that is worth solving**
- **Theory building is the most personal and creative part of research**
- **Theory testing is a challenge to your theory, using all available information including your experimental results and their analysis**
- **Reflection: You must reflect on how your research findings relate to current thinking in the field of your research topic and identify further questions and new avenues to explore**

'Life is a maze in which we take the wrong turning before we have learned to walk.'

Cyril Connolly (1944)
The Unquiet Grave

Introduction

Research may seem, at first sight, like a mysterious world full of very clever people, remote from the rest of us, talking in strange languages about weird subjects, chalking obscure symbols and diagrams on blackboards, writing unintelligible papers for academic journals: an unreal world like a nightmarish fairground full of wonders and coloured lights but also strewn with hazards and pitfalls. Certainly it will seem like a box of tricks, even a disorganised mix of activities, a maze in which you could easily lose your way.

In this chapter, we offer a different view of research that we hope will encourage and guide you along a fascinating and adventurous path of exploration. Yes, research is a maze but there are many paths through it and, by example, we shall take you along those paths.

Research is about solving problems: overcoming the obstacles between the first statement of a problem and its solution.

We can start with a problem that might face you during your post-graduate studies. When you have decided on the topic for your research you may be encouraged to visit a research establishment, an agricultural college or a university in South Africa or Tanzania for the theoretical aspects of your research. The plan might be to return home to run some field trials and then to visit southern Africa, again, for support in the analysis of your results. Your immediate problem is: how should you travel to South Africa from, say, Nairobi? What is the most economical and enjoyable way, following a route that is within your means and takes you through some interesting places? How would you go about it? Well, you would probably start by opening an atlas. You would mark a possible route, putting rings round the names of places that you would like to visit: Lake Tanganyika, Lake Rukwa, Zambia, the Victoria Falls, Okavango Delta, Botswana and on to Johannesburg. Or you might consider travelling through eastern Tanzania to Dar es Salaam, Lake Malawi, Mozambique and Zimbabwe. There are a number of possible routes from Nairobi to Johannesburg; there may be several possible ways to solve any problem. You could find

1. This chapter is adapted from a chapter by Tom Bourner, with his permission, in the book *Research Methods for Postgraduates*, Second Edition, Arnold (2002).

out from friends and relatives what they know about travel options. You could go to the bus stop and make enquiries of people returning along those routes. You could go to the library and read some travel books to learn about modes of travel in each of the countries through which you might pass. The Lonely Planet series can offer plenty of advice. Is hitchhiking safe? Would it take too long? What are local buses like, where is it possible to use trains? How frequent are the ferries? How much would they cost? What sort of accommodation would you expect? Do people in the villages offer accommodation and food to travellers? Are there hostels or cheap hotels? Perhaps you would carry a small tent and sleep in the open. Local weather patterns, local laws and safety might influence you. You might plan three or four possible routes, write down the details of each and a list of pros and cons. You might even devise a scoring system to help you to decide which route to follow. Later, when you have reached your destination and are still marvelling at the differences between Nairobi and Johannesburg, and some similarities, as well as all the wonderful adventures you have had on the way, you will probably reflect on the process: the extent to which your journey met your original aspirations, what your first-hand experience has told you about travelling cheaply across Africa, and what you have learned from your experience.

Now if you take off your traveller's spectacles and put on instead a pair of researcher's spectacles you will observe some similarities between that process of planning a long journey and the process of research. First, you did a literature review (atlas, travel books, bus and rail time tables, tourist leaflets) to get an overview of the field. Second, you developed a theory of which of the available routes would be to your requirements (your short list). Third, you tested the theory by inspecting and scoring those on your short list. The testing continued by making your journey, by doing your fieldwork. Finally, you reflected on the experience and your results. Stated formally, the process contains four parts:

Part 1 Reviewing the field
Part 2 Theory building
Part 3 Theory testing
Part 4 Reflecting and integrating

Perhaps this sequence seems familiar. Perhaps you recognise it from other significant decisions you have made in your life: choosing an undergraduate course, buying a new suit, finding a vacation job.

With some decisions, it's not possible to go through all the stages. For example, when you choose a job the final test of your theory that you have chosen the right job is by doing the job. Unfortunately, this is possible only after you've committed yourself to the job. Perhaps that's why so many unsatisfactory job decisions are made. The literature on labour turnover often refers to the period immediately following recruitment as the 'induction crisis' when job expectations are tested by the job realities.

The four stages are four parts of the path through the research maze. But this is no ordinary maze in which there is a unique path to follow. Side turnings in an ordinary maze will take you to dead ends. Some side turnings in the research maze will also have dead ends but others will reach the goal, a solution to the problem. But will it be the best solution? Another path through the maze might have been better. There are usually several ways to solve a problem, to do research. Some ways will pose more difficulties than others, present more limitations, take you down fruitless side tracks.

Once you recognise that you are already familiar with each of the major parts of the research process through your experience of making the larger decisions of your life you will have a valuable resource to draw on. Reflection on those experiences will also give you an indication of the possible pitfalls.

That four-part process can help you to put what you are doing into a broader picture when you start to get bogged down in the detail of research. It can also be useful in designing your research project.

Let us examine the parts of the process in more detail.

Part 1 Reviewing the field

Many research projects arise from a study of current thinking in a field. The research project follows from identifying a gap in the literature. Most other research projects arise from awareness of a problem that is worth solving. In either case, a good start is an overview of current thinking in the field.

In case you are impatient with this part of the process and want to start immediately with fieldwork, here are some reasons for spending time and effort on a review of the field. It would help you to:

- Identify gaps in current knowledge
- Avoid reinventing the wheel (at the very least this will save time and it can stop you from making the same mistakes as others)
- Carry on from where others have already reached (reviewing the field allows you to build on the platform of existing knowledge and ideas)
- Identify other people working in the same and related fields (they provide you with a researcher network, which is a valuable resource indeed)
- Increase your breadth of knowledge of the area in which your subject is located
- Identify the seminal works in your area
- Provide the intellectual context for your own work, (this will enable you to position your project in terms of related work)
- Identify opposing view
- Put your own work in perspective
- Provide evidence that you can access the previous significant work in an area
- Discover transferable information and ideas (information and insights that may be relevant to your own project)
- Discover transferable research methods (research methods that could be relevant to your own project).

Part 2 Theory building

Theory building can be the most personal and creative part of the research process. Some people find it the most exciting and challenging part of the whole business.

In some cases, data collection precedes theory building and, in other cases, it follows it. Have you ever bought a second hand bicycle? If so, you may have identified some possibles before narrowing down to a few probables. You collected data and then formed a theory about which of the bicycles would best meet your needs. In that situation, theory building followed data collection. The process of developing a theory by inspecting individual cases has a special name: **induction.**

Our journey from Nairobi to Johannesburg is another illustration of induction. If each time you are sent the times and prices of a train journey from one city to the next in a faraway country you notice that it is more expensive than you can afford, you may develop the theory that all the train journeys in that country are too expensive for you. Acting on that theory, you may ask the travel agent to stop sending details of train journeys in that country and you ask for details about buses instead. That is the process of induction at work again: forming a theory from information about specific instances. Induction is a type of generalisation.

The other side of the coin from induction is **deduction** that involves reaching conclusions about specific instances from general principles. Here is an example of deduction: 'I can't afford to stay in a Lagos hotel so don't bother to send me the details of hotels in Lagos'. In this example 'I can't afford to stay in a Lagos hotel'– is the generalisation and deduction leads you to the conclusion about any specific Lagos hotel, i.e., that you can't afford it.

Induction is a thought process that takes you from the specific to the general. Deduction is a thought process that takes you from the general to the specific.

We have seen how a theory can emerge from the data. However, theory can also emerge from armchair theorising, introspection, deduction following a review of the literature, personal experience, a fortuitous remark, a brainstorm, an apt metaphor, or pure inspiration. Creativity has a role to play in all aspects of the research process, but especially in the theory-building part.

We said earlier that data collection can precede theory building and that it can follow it. In the case of induction, data collection comes first. When data collection follows theory building then it is usually for the purpose of testing the theory. That is the part of the research process to which we turn next.

Part 3 Theory testing

'Experience has shown each one of us it is very easy to deceive ourselves, to believe something which later experience shows us is not so.'

C. Rogers (1955)

When we were planning our long journey, we wanted to check whether those attractive routes that we marked in our atlas would really meet our needs. Likewise, when we are doing research we will want to check if the theory (or theories) that we have formulated fulfil our hopes and expectations.

The sort of theory testing we do will depend on our ambitions and claims for our theory. If we want to claim that our theory applies generally, e.g., 'All hotel rooms in Lagos are more expensive than all hotel rooms in Dakar' then we may want to use statistical methods (known as inferential statistics) which have been developed to enable us to make claims about whole populations from information about a sample from a population.

But if your claims are only about the accuracy of your theory in the context of a particular situation, e.g., 'The route that suits me best is via Zambia and Botswana' then theory testing may involve checking your conclusions (theory) from other perspectives. You may have gathered a lot of information about a particular route (bus and train fares, hostel prices, sights to see) but you might find people who have already visited some of the places on the route. They could tell you from their own experiences what to expect. In research in the social sciences, the term **triangulation** is used to describe the process of checking if different data sources and different methods allow you to reach the same conclusions.

Testing theory can take many forms. At one extreme, you may simply invite the readers of a research report to test the conclusions against *their* own experiences. The test is: does the reader say 'Aha! I can now make sense of my own experience in a new and convincing way'? But if the reader is unlikely to have first-hand experience for testing the researcher's theory, or if the claims being made involve a high level of generality, then the theory testing stage will be more formal and elaborate. At some level, however, theory testing is likely to be part of any research process.

Part 4 Reflection and integration

'Knowledge doesn't exist in a vacuum, and your knowledge only has value in relation to other people's.'

A.D. Jankowitz (1991)

Reflection and integration is the last stage of the research journey. There may be many things on which you want to reflect: what you have learned about the process of research; what you could have done differently; what you have learned about yourself. But there is one matter for reflection that is a crucial part of the research process itself. It will affect how your research is judged and the impact of your research. You must reflect on how your research findings relate to current thinking in the field of your research topic.

Your reflection on how your research results relate to current thinking will include your assessment of where your research fits into the field of knowledge. It will contain your assessment of your contribution to the field. In this part of the research process you are likely to return to your review of current thinking that you made at the outset and reassess it in the light of your results. It's as if the current thinking in your field of study is a partially complete jig-saw puzzle and you are detecting where your own new piece of the jig-saw fits in.

Relating the outcomes of your research to current thinking in the field may simply involve showing how it adds to what is already known in the field. This would be the case when you have filled a gap in the literature or found a solution to a particular problem in the field. It may involve seeking connections with current thinking. It may involve challenging some parts of the map of the current thinking in the field, so that you will be proposing some reconstruction of that map. It may involve testing the consistency of your research findings with current thinking. It may involve asking 'What if?' questions of your research findings.

Any of these ways of relating your research findings to current thinking in the field may present further questions and new avenues to explore. Successful research usually answers some questions but also raises new ones. It enables researchers to ask questions that would not have been asked before the research. New questions can be an important outcome of research. It is small wonder therefore that the final chapter of most research reports has a section containing suggestions for further research.

A good practical question to ask yourself is: 'What are the implications of my research results for our understanding in this area?' The implications can take many forms. Here are a few:

- You may have filled a gap in the literature
- You may have produced a possible solution to an identified problem in the field
- Your results may challenge accepted ideas in the field (some earlier statements in the literature may seem less plausible in the light of your findings)
- Some earlier statements in the literature may seem more plausible in the light of your findings
- Your work may help to clarify and specify the precise areas in which existing ideas apply and where they do not apply (it may help you to identify domains of application of those ideas)
- Your results may suggest a synthesis of existing ideas
- You may have provided a new perspective on existing ideas in the field

- Your work may suggest new methods for researching your topic
- Your results may suggest new ideas, perhaps some new lines of investigation in the field
- You may have generated new questions in the field
- There may be implications for further research.

Most of all, this last stage in the research process is about seeking to integrate the fruits of your own research with current thinking in the field.

Summary and conclusions

It is sometimes difficult to keep in mind the whole research journey when all of your attention is focused on crossing some particularly difficult ground. Our purpose in this chapter is to help you to keep the whole research process in perspective when you are engaged in a particular research activity. We have done this by giving you an overview map on which the whole journey is plotted in outline. We hope this will help you to plan your research journey.

We have related the process of research to the way that you find information needed for the larger decisions in your life. You already have much experience to draw upon in planning and doing your research.

We have suggested a four-part research process: 1. Reviewing the field, 2. Building theory, 3. Testing theory, 4. Reflecting and integrating.

There is considerable diversity of approaches to research in different fields but this four-part framework is sufficiently broad to encompass most research in the sciences, the agricultural and environmental sciences and the humanities. Much of the literature on research focuses on different parts of the process. For example, in the social sciences it usually focuses on theory-building, whereas in agricultural sciences it may focus on theory-testing.

Your four parts may not follow this sequence strictly. For example, after you have reviewed the literature you may want to monitor developments in current thinking while you are collecting and analysing data. You may engage in some parts of the research process more than once. For example, you may find that data you collect for theory building enables you to test statements found in the literature. Or data collected to test a theory may suggest a new theory so that it becomes an element of theory building.

You may not want to spend the same amount of time and energy on each of the four parts of the process. For example, theory building may be only a token part of your research project if your main contribution lies in testing a theory that you found in the literature. On the other hand, you may direct most of your effort towards theory building, so that theory testing may be little more than establishing the plausibility of your theory in the light of the data you've collected.

The four parts will be present in almost all research projects, at least conceptually. If one of the four parts seems to be missing from your own research project, you should discuss it with other researchers and, if you are registered for a research degree, with your supervisor. If you intend to omit one of the parts from your own research project, you must be able to state clearly why this part has no role.

Resource material and references

Greenfield, T. 2002. *Research Methods for Postgraduates*. Second edition. Arnold, London, UK. 384 pp.

Jankowitz, A.D. 1991. *Business Research for Students*. Chapman and Hall, London, UK.

Rogers, C. 1955. Persons or science: a philosophical question. *American Psychologist*, vol. 10. Reprinted in *On Becoming a Person*. pp. 267–279. Constable, London, UK.

Part 2
Your research in perspective

Agriculture is an important component of rural life. Rural transformation is part of the development process and it implies change. To alleviate poverty and become food- secure we need to improve the productivity of agriculture. The smallholders who are the backbone of agriculture deserve our attention and support. National governments, professionals and outside agencies must all work together to achieve this common goal. Agricultural professionals need to become active change agents. As a scientist or an agriculturalist you need to focus on 'problem solving' and remember that both social and scientific skills are required to navigate and manage rural change.

Everyone's life or research journey will be different, but our target and beneficiaries will be the same. This part of the book focuses on many issues with which you will have to deal or confront throughout your career. You will have to work as a team member in a department, programme, or project. This involves cooperation, sharing ideas, openness, transparency, and honesty. You will be a link in the chain that binds the team together. Be sensitive to the needs of others, especially the small-scale farmers – they deserve feedback and all our professional assistance.

Three chapters of this part are rather personal appeals to adopt certain values and approaches in your research. In the first it is suggested that researchers should see themselves as elements in a change process rather than the bringers of answers. The importance of recognising and using farmer's local knowledge is emphasised. In the second you are encouraged to pursue research which will actually lead to solving substantive farmer problems and improving the welfare of the poorest farmers. The third outlines some approaches to working with farmers that might help this to happen. The authors' views are generally mutually consistent, and are based on years of experience of what has worked and what failed. However, the ideas they present are all the subject of debate, and others with similar experience may choose different points to emphasise. As a student you should become familiar with these discussions, work out their relevance to your situation and think about the extent to which your experience confirms or modifies them.

As a student preparing an MSc or PhD thesis, your work will probably be embedded in a larger project. You may therefore have limited opportunities for influencing the overall strategy and approaches of that project. However you do have to be fully aware of how your work links to the rest of the project. The chapter on developing and managing projects should help with that. If your project meets all the aspirations described here then you are lucky! If not, you have a responsibility to challenge the strategy adopted by the project, and to make your own component as effective as possible.

Bharati K. Patel

?
ROLE OF THE PROFESSIONAL in Rural Transformation
WHO ARE YOU RESEARCHING FOR?
WHAT LEVEL OF TECHNOLOGY CAN THEY ADOPT?
The AGRICULTURAL KNOWLEDGE TRIANGLE
EXTENSION
UNIVERSITY
RESEARCH
TEAM Building
PUTTING A PROJECT TOGETHER
Project Writing
BUDGETING
• TIME • MONEY
• EQUIPMENT
APPROACH
Respect
Empathy
Creativity
ALEYA

2.1 The role of professionals in rural transformation

Joseph Opio-Odongo

- **Science and technology can make a difference in rural transformation**
- **Agricultural scientists must play a positive role in rural development**
- **Understanding farmer reality and good communication skills are essential for servicing rural communities**
- **Rural communities have the capacity and capability to change their reality**
- **Technologies and innovations must match the needs of the target communities**

'Agricultural research scientists have been reluctant revolutionaries, because while they have wanted to revolutionise technology, they have preferred to neglect the revolutionary impact of technology on society.'

Vernon Ruttan (1982)

Introduction

The development and application of science and technology by agricultural research scientists in sub-Saharan Africa (SSA) are part of the global agenda to liberate people from the limitations of the natural world. However, although both positive and negative socio-economic effects of the liberation should be expected, these have not preoccupied the minds of the agricultural research scientists. Yet how society deals with such change affects both the nature and pace of rural transformation.

Agricultural research scientists, as agents for this transformation, need new ways to deliver the results of applied science if they are to empower farmers to operate as subjects rather than the objects of rural transformation.

Rural transformation

When you read the literature on agricultural and rural development, you will find that the concept of rural transformation is often talked of in terms of the:

- Changing composition and roles of the agricultural and other sectors in national development
- Dwindling size of the rural population engaged in agriculture
- Increasing prominence of agro-processing in the rural economy
- Declining proportion of the population deriving its livelihood from agriculture
- Substantial reduction in the contributions of agriculture to the gross domestic product (GDP).

All these are often used as the indicators of rural transformation. But, an equally important aspect of that transformation is the changing capacity of the rural population to drive the process of change. This capacity is expressed in the people operating as the subject rather than the object of change as they shape their destiny by tactfully dealing with the present situation and strategically working towards capturing the future. It is this aspect of the transformation that is the focus of this chapter. Of particular interest is how agricultural research scientists could work with the farming communities to help them manage self-propelling and people-driven processes of change that contribute to sustainable livelihoods and the transformation of the rural economy.

For this to happen, four main conditions must be fulfilled.

1. Recognition that rural transformation, as a self-propelling and people-driven process, is possible in any society and that the knowledge system that drives it must be internalised and used appropriately.
2. The internal dynamo for rural transformation has to be recharged after years of debilitation by colonial and post-colonial policies and institutions.
3. Agricultural scientists need to be animators and facilitators rather than interventionists to increase local initiative and reduce dependency.
4. Agricultural scientists must invest in acquiring the skills and art of animation, appreciative inquiry and the others needed to facilitate rural transformation.

Internal dynamo for rural transformation short-chained

Colonial rule influenced agriculture in SSA in three important ways. Firstly, it attempted to divest it of its foundation that was built on a theory of knowledge learned from tried and tested traditional methods. The professional and administrative assault on traditional farming practices was meant to transform agriculture from an African way of life into business-based European models of farming. Traditional agricultural practices such as intercropping, although poorly understood by the Europeans, were deemed inferior and were to be replaced by modern ones.

Secondly, it attempted to develop a cadre of 'master' or 'model' farmers who were to be entrusted with the responsibility of recreating the agricultural system in line with the European model.

And thirdly, it introduced a research support system that predominantly served the interests of science and private capital rather than the developmental needs of local farmers. Champions of the traditional knowledge systems and the innovations developed from them fizzled from the limelight, thereby creating space for the new cadre of extension workers to disseminate information from the research stations.

It was in the process of implementing these changes that the **internal dynamo** for rural transformation was **short-chained**. In many SSA countries this situation changed little during the first half of the 20th century. It was not until the mid-1960s that agricultural research scientists began to validate the theories that underpinned many of the traditional agricultural practices such as intercropping. Until very recently, such validation was performed largely within the framework of a delivery model that treated local farmers more as objects rather than subjects of rural transformation. The internal dynamo for rural transformation still remains stalled in many rural communities in SSA.

Trained incapacity to reckon with local situations

Formal education is a powerful socialisation tool that professional disciplines apply in training those of their calling to enable them to view and interpret the world appropriately. Under the controlled condition of experiments in laboratories, glasshouses or experimental plots, that tool has served agricultural scientists very well. However, for situations outside these controlled conditions, especially in the world of farming where a number of conditions influence farmers' behaviour towards a technology that is being promoted by agricultural scientists, that tool in its conventional form has been less powerful.

The trained incapacity which comes from conventional higher education, results in a tendency to blame farmers for being conservative or backward when they don't implement new technologies, rather than assessing where the recommendations are failing farmers' needs.

> *'Agricultural scientists and researchers do not usually understand the positive and negative forces and changes facing farmers. Current training does not encourage an understanding of traditional husbandry practices. Even students coming from rural communities consider their new knowledge to be superior and have little real appreciation of the inherent logic of traditional systems.'*
>
> **Joseph Opio-Odongo (1992)**

Students are not trained to develop an appreciation of the basic foundation of traditional agriculture. There is growing evidence that farmers' planting strategies, seed selection and preservation, and crop rotation, are based on knowledge systems developed over generations. The same holds true for the local system of soil classification. Farmers continue to develop a variety of agricultural innovations that help them to deal with the exigencies of rural life but which are unknown or poorly understood by agricultural scientists. Researchers and students rarely appreciate why farmers may adopt, but then adapt the new technologies. An example from Uganda in the 1960s helps to illustrate this.

Despite strong recomendation by the research and extension establishments in Uganda that cotton farmers should plant early and use specific fertilizer applications, the majority of farmers developed their own variations of what was recommended. They planted later and applied lower dosages of fertilizers. When the results of field trials based on farmer practices were published – the farmers' deviant behaviour was proved to be rational and valid.

It is also imperative that students and agricultural scientists be trained to communicate effectively with farmers, and to be exposed to farmer situations so that they can develop vital empathy.

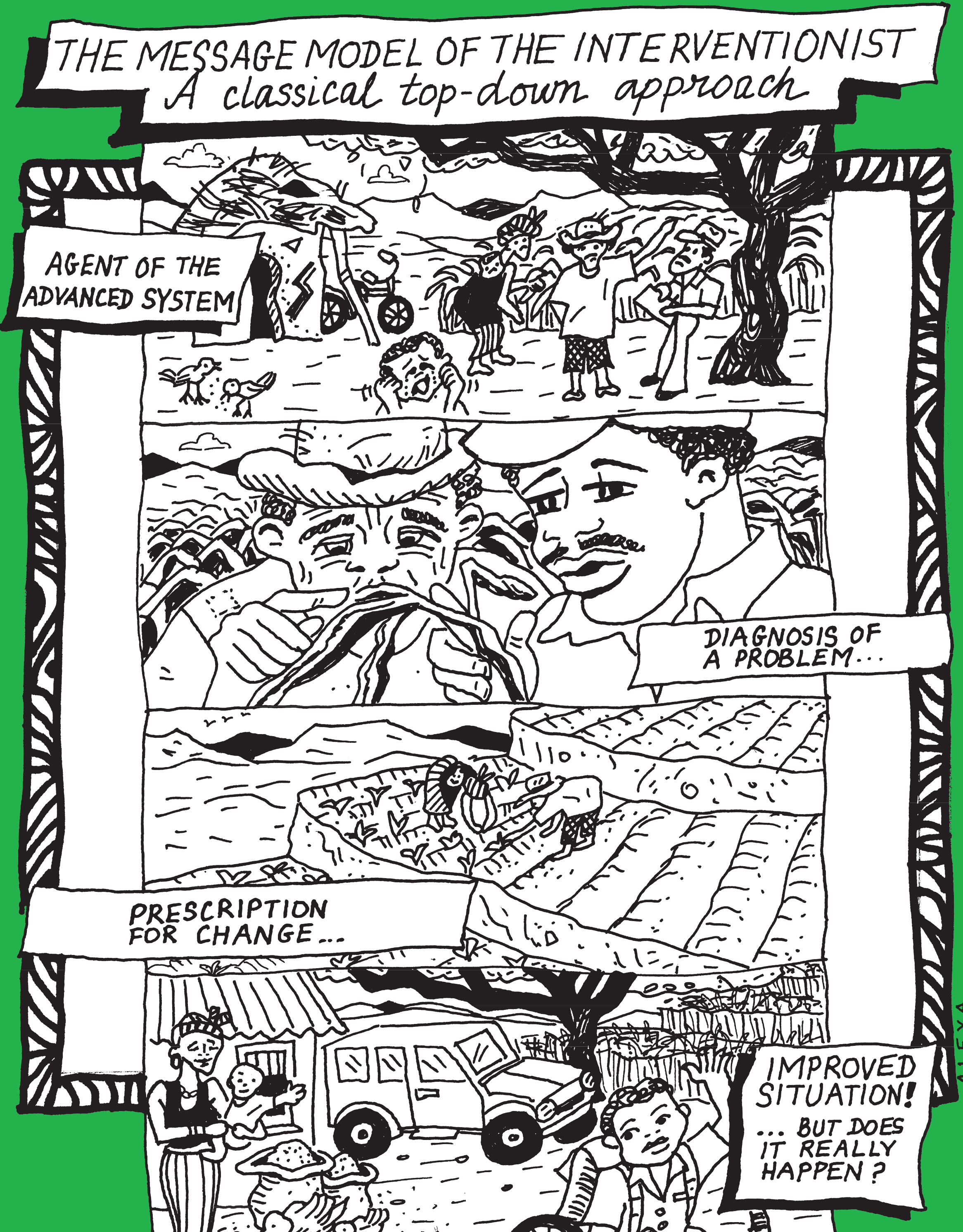
The message model of the interventionist

The usual way services are delivered by agricultural scientists in SSA is to use external expertise to define a problem and then to institute measures to resolve it. The role of farmers in dealing with the situation tends to be largely that of passive implementers. It is an example of the classical benefactor–beneficiary relationship. How it was thought to work is shown on page 24, but in reality it often fails.

An example of this 'message model of the interventionist' is one where a representative of a public health system notices a problem and diagnoses it as by poor nutrition. Arrangements are therefore made to enrich the flour consumed in the area with the aim of improving the nutritional status of the community.

Replace the nutritionist with an entomologist or agronomist and the approach remains valid. The entomologist, for instance, notices a substantial drop in the average household income in a community dependent on maize for its income and food security. His/her diagnosis is that poor pest control is resulting is considerable post-harvest losses. The 'prescription' is that farmers should adopt modern ways of controlling post-harvest losses in order to ensure food and income security in the community.

The gravity of the problem and the inherent superiority of the modern methods being introduced are expected to compel farmers to adopt the modern system of reducing post-harvest losses. To encourage adoption, on-farm trials are conducted to enable farmers to appreciate the efficacy of the modern methods. Little attention is usually given to such


THE MESSAGE MODEL OF THE INTERVENTIONIST
A classical top-down approach
AGENT OF THE ADVANCED SYSTEM
DIAGNOSIS OF A PROBLEM...
PRESCRIPTION FOR CHANGE...
IMPROVED SITUATION!
... BUT DOES IT REALLY HAPPEN?
ALEYA

issues as the affordability, accessibility and sustainability of the recommended methods given local constraints.

The interventionist model is basically one that aims to change the direction of human activity in what the intervener deems to the most appropriate way. The model operates as if its application is based on full information. Yet in most instances agricultural scientists rarely have the full picture of the situation within which they are intervening, especially given their indication to take a mono-disciplinary approach to problem solving. It is rare to find them working in concert with social scientists in multidisciplinary teams. Neither would they normally try to seek information on what solutions have been applied by farmers in response to the problems at hand and with what results.

Becoming an animator or facilitator of rural transformation

Notwithstanding the general intellectual arrogance that delayed the understanding of the traditional knowledge based African farming practices, since the late 1960s there has been an upsurge of interest in such practices, notably in farming systems research. The researcher–extension worker–farmer linkages that farming systems research promoted was consistent with the late President Julius Nyerere's plea that: *"In the interest of becoming more effective in what they do, African intellectuals have to be part of the society, which they are changing; they have to work from within it, and not try to descend like ancient gods, do something and disappear again."*

This was a wake-up call for agricultural research scientists to operate more like animators if they expected to have any impact on agricultural and rural development in Africa. It challenged them to shift from the message model of the interventionist to one that is based on the knowledge and value systems of communities and through which the agricultural research scientist facilitates the process of agricultural and rural transformation. Such a model was described as a 'system model of the explicator' that could be applied by agricultural scientists in facilitating rural transformation.

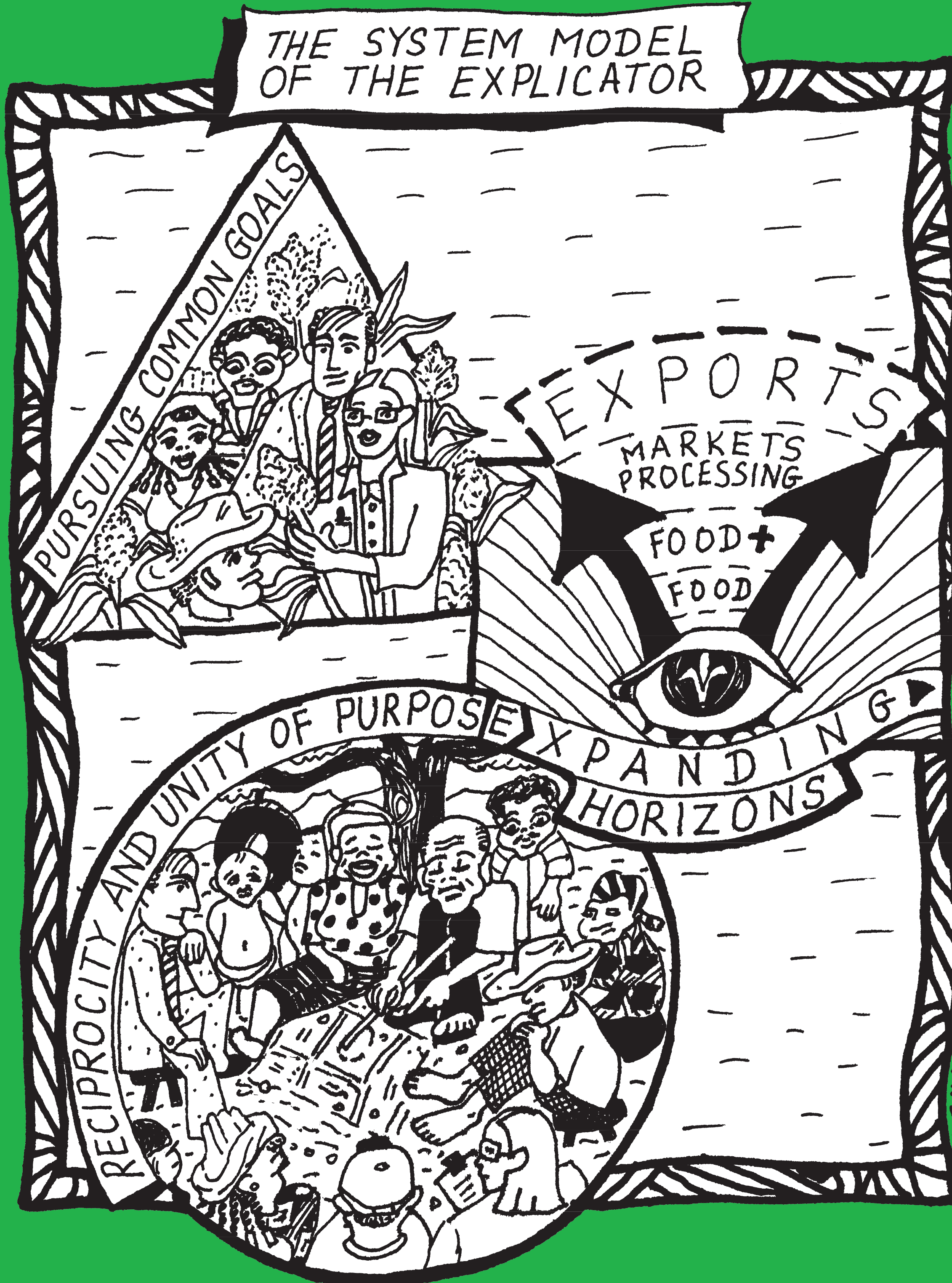
In the system model of explicator, agricultural scientists accept the inherent capacity of people to take charge of their own destiny, thereby drawing upon the internal capacity of the stakeholder community to activate a self-driven transformation process. The model is underpinned by the principles of communication, reciprocity and partnership that are essential in enabling rural people to apply their knowledge and capacities to activating the process of rural transformation. How it could work is depicted on page 26.

The system model of the explicator

A fundamental assumption of this model is that a farming community is an information-processing structure that draws upon internal and external information to act on situations. The model recognises that farmers have a stock of knowledge and wisdom that agricultural scientists have tended to ignore to their disadvantage.

It is through dialogue with members of any community that agricultural research scientists can tap the farmers' rich wealth of experience and wisdom in agriculture and use it to stimulate rural transformation.

Studies in Africa have validated the merit of using such a model, e.g., the indigenous knowledge on soils that was tapped from the Bété and Senufu communities in Côte d'Ivoire, read Birmingham, 1998. A number of techniques have already been developed to enable researchers to access such knowledge including:


THE SYSTEM MODEL OF THE EXPLICATOR
PURSUING COMMON GOALS
EXPORTS
MARKETS
PROCESSING
FOOD+
FOOD
RECIPROCITY AND UNITY OF PURPOSE
EXPANDING HORIZONS
ALEYA

- Participatory Rural Appraisal (PRA)
- Methods of Active Participation (MAP)
- Participatory Action Research (PAR)
- Appreciative Inquiry (AI).

Fundamentally, the system model of the explicator does not pretend that every popular practice within an agricultural system is welfare-promoting, or that there is never room for improvement or innovation. Rather, its most powerful message is that those improvements or innovations are more likely to succeed and become sustainable in the long run if they are introduced with due cognisance of the existing knowledge systems, needs and capacities of the target communities. Using this model, an agricultural scientist interprets the world of rural people not as a set of unsolved problems, a series of gaps and deficiencies and failures, but rather as a set of brave attempts, a series of partial achievements, and a sequence of possibilities that could yield rich rewards if the community and the agricultural research scientists worked together.

Implications for your graduate research training and research in agriculture

If you are to follow this model then you need to:

- Develop skills that allow you to interact effectively with the communities in which you will carry out your research – take some social science courses and try to work with interdisciplinary teams
- Improve your communication and listening skills to capture local knowledge and use it in designing your project
- Be sure that your project really addresses the needs of the people you are trying to help.

The impressions that members of the rural community have of who they think you are and what interests you, substantially influences what information can and cannot be shared with you. If you do not quickly detect and help to dispel a negative perception, it will affect the quality of information you are receiving from the community. If that happens, your interaction with the community could do you more harm than good.

Learn the art of using observations and knowledge of the agricultural systems where you work so that you appropriately factor them into your research design and respond to the priority needs of both the scientific and the farming communities. You also need to understand the interface between science and the economy, or science and commerce. Indeed, many people believe that the increasingly closer interface between science and commerce has been a main factor in the development of biotechnologies. The biotechnology movement did not begin from the usual pursuit of knowledge for knowledge's sake but rather from the decision to strategically apply scientific knowledge to enhance the competitive edge of those investing in the generation of new technologies in a very highly competitive global economy.

You need to be conscious of the primary need that you are serving as an agricultural scientist in the public domain. It may be difficult to hide behind the cloak of neutrality.

The system model of the explicator is not a panacea for all research or rural development problems. Each development situation as it unfolds presents new challenges that pose difficulties in the application of the model. Indeed, given the predominance of the message model of the interventionist in much of rural Africa, farming communities may not initially think that you are serious when you do not provide ready answers to their problems. Helping them to begin taking charge of their situation and destiny is the art that you must possess. It is therefore important that if you try to apply the system model of the explicator, you also make a deliberate attempt to determine what works and what doesn't. Out of that experience, a local research team can then evolve a modified version of the model that is more appropriate to the specific situation in which they are working.

Resource material and references

Birmingham, D.M. 1998. Learning local knowledge of soils: a focus on methodology. *Indigenous Knowledge and Development Monitor* 6(2): 7–10.

Hagmann, J. 1999. *Learning Together for Change. Facilitating Innovation in Natural Resource Management Through Learning Process Approaches in Rural Livelihoods in Zimbabwe.* Margraf Verlag, Weikersheim, Germany. 337 pp.

Hagmann, H., Chuma, E., Murwira, K., Connolly, M. and Ficarelli, P. 2002. Success factors in integrated natural resource managemnt R&D: lessons from practice. *Conservation Ecology* 5(2): 29.

Opio-Odongo, J.M.A. 1992. *Designs on the Land: Agricultural Research in Uganda*, 1890–1990. African Centre for Technology Studies (ACTS), Nairobi, Kenya.

Ruttan, V.W. 1982. Agricultural scientists as reluctant revolutionaries. *Interdisciplinary Science Reviews* 7(2): 170–177.

Young, F.W. 1968. Intervention versus explication in applied science. *Journal of Home Economics* 60(4): 262–267.

2.2 Research for whom?

Bharati K. Patel

- Two-thirds of the African population reside in the rural areas and are involved in agriculture
- Farmers voices must be heard and their participation is essential to bring about change
- Africans can and must solve African problems
- You can make a difference
- Universities must work closely with research and extension services
- Learn by doing – get your hands dirty
- Be aware of the limitations of your research methods
- You should honestly assess the value of your research
- There is great satisfaction in being of service to others

'Whenever in doubt apply the following test: Recall the face of the poorest and the weakest man whom you may have seen and ask yourself if the step you contemplate is going to be of any use to him. Will he gain anything by it? Will it restore him to a control over his own life and destiny? In other words, will it lead to self-reliance for the hungry?'

Mahatma Gandhi
Advice to policy-makers

Research for whom?

Research for the millions of small-scale, resource-poor farmers who have to eke out their livelihoods from agriculture. Nearly two-thirds of the population on our African continent is dependant on, or involved in agriculture. If we want to reduce poverty, feed all our people and develop then agricultural productivity must increase. We cannot ignore or bypass this large sector of our population.

There are many problems to solve in agriculture. Much research has been done and many technologies have been developed, but why are so many of them still on 'the shelf'? There are a number of reasons why this has happened but for the most part it is because the technologies are not relevant to farmer's circumstances. To be effective new technologies must fulfil an identified need.

Farmers' voices must be heard through their active participation in helping to set research priorities. Only then will research and progress be in step.

Agriculture and rural development have progressed significantly in Asia and Latin America, why not in Africa? The development journey began at about the same time for all three regions, but Africa has remained stagnant at best and in some parts has even declined. We can only change this situation if each one of us makes a commitment to serve and better the lot of the majority of Africans.

Some of our governments have slowly come to realise that the small-scale farming sector really needs better support because agriculture is the engine for growth and development, and because poverty-reduction goals cannot be achieved without such support. The Asian governments supported agriculture and benefited from the Green Revolution in feeding their large populace, but African governments wanted to bypass agriculture and leapfrog into industry. Even in those African countries where there was investment in agriculture, it was only in the large-scale sector. Kwame Nkrumah did not heed the report of the economist W Arthur Lewis who advised that Ghana should give priority to food production before embarking on the road to industrialisation. Nkrumah even abolished the national extension system! Korea and Ghana in 1957 had the same per capita income and Ghana's gross domestic product (GDP) was higher than Korea's. Now Korea is a developed country and Ghana is at long last heeding Lewis's advice, having learnt a bitter lesson. Nigeria's experiment with industrialisation is another glaring failure, yet even today many African countries do not give agriculture the priority it deserves. Small-scale agriculture can be productive; it helps retain earnings in the rural areas and can have a significant impact on unemployment, as experience in Asia has proved.

The World Bank's lending for agriculture is currently at an all-time low. The Millennium Goals give very little emphasis to agriculture. The environment, infrastructure and population density make things different in Africa. The situation calls out for research into what is limiting progress. The contribution you, as a postgraduate, can make is significant.

We need young Africans to take up the challenge and invest their time and brain-power in agriculture. You must lead the way.

My personal research journey

I joined the research service in Zambia in 1971 immediately after getting an MSc degree in plant pathology that required a dissertation and a research study. I did my MSc when chemical control was the in thing and you could always find a chemical control measure to get rid of a disease or pest problem.

The dissertation was on brown spot and blue mould (Wild Fire) of tobacco, and my practical research was on potato dry rot (R*hizoctonia solani*). During this practical research I learned a lot about varietal reaction to the same pathogen, how size and time of infection affects the quality of potatoes, how long the disease survives in the soil, and how farmers were dealing with the problem.

My first sole assignment when I started work was to solve a disease problem on the greens of the largest golf club in town! I knew nothing about fairy rings, but did manage to identify the cause and provided a chemical solution to the problem.

In those days pathologists and entomologists had to do research and provide a diagnostic service by identifying pests and diseases on samples sent in by farmers – mainly commercial large-scale growers because very few small-scale farmers took advantage of this service. This taught all the staff a great deal about holistic diagnostics and stood them in good stead when they went to the field.

One redeeming feature of the job was an annual tour of one whole province of Zambia and that included visiting farmers of all types – in some areas there were only small-scale farmers. These tours allowed me to observe and learn much about development and problems in the rural areas. We travelled by landrovers over every bump and got stuck many times because the tours were made during the rainy season. It was a wonderful way to see, and learn while building relationships and networking with colleagues from other agricultural divisions – extension, veterinary services, land use planners and those working in the University and the National Council of Scientific Research. In those days I remember we used to have unspent budgets! so adequate funds were available for our research, for travel within the country, to buy books, and to print advisory booklets or pamphlets. Farmer training institutes in the rural districts were well funded. Research centres based in the provinces did do research for the smallholders. They bred beans for local conditions that were based on taste preferences and would fit into local intercropping patterns, and developed sunflower varieties for smallholder production.

Farming systems came to the forefront in the early 1980s

Integrated research teams were formed involving economists, agronomists, and sociologists. Adequate funding was provided by donors for these teams, and even though they were looked upon as the 'prima donnas' of the research systems, they did do research in the rural areas on smallholder farms. Funds from government coffers were stable but started to shrink in the mid-1980s and fell to critical levels in 1990s. I did my PhD in Australia during this period. Major training efforts were launched in the national agricultural research systems (NARS), extension and other divisions of the ministries of agriculture with support from donors at this time. An expansion in capacity development that was much needed. Interaction with the international agricultural research centres (IARCs) became stronger in the 1980s and in the 1990s non-governmental organisations (NGOs) entered the agricultural scene with a bang! Today there are many players/actors in the agricultural research and extension arena and agriculture involves many more institutions and organisations than the traditional NARS, IARCs and universities.

The two decades after independence can be identified as the period of growth for agricultural research in Africa

Unfortunately small-scale farmers were not clearly identified nor targeted by research services during this period – they just loomed in the background – forming a backdrop – hence, the road to enlightenment, awareness and duty to identify and assist those who most require our assistance was a long one. The smallholders slowly came into the picture and focus and now hold the centre stage for all agricultural developmental activities. I witnessed the transition and participated in the slow change from a classical model of conducting research on-station to advocating and supporting on-farm research by all scientists in the research service.

My personal views on research and development became more focused with time, experience and exposure. What stood out was that **good science** – the currency of researchers, whether in the laboratory or field or participatory-based **is a must**. And that you need well-trained scientists to do the research. You reflect on your life and achievements from time to time – and you do ask yourself the following questions 'What have I achieved/whom have I helped?'; 'Have I made the best of my opportunities?' Honest reflection does often give you the impetus to go the extra mile. Throughout my journey I had ample learning opportunities, met a lot of interesting and dedicated persons, and was lucky to have good bosses, colleagues, and seniors. You have to be curious, open-minded and keen to learn to take advantage of all the opportunities that comes your way. I was very lucky to have worked for my country's research service, for an IARC and finally for a foundation that firmly believes in science-based development. My last job gave me the opportunity to develop and operate a programme which allowed me to put into practice all that I had learned about research and development over time. The programme focused on two fundamental areas of importance in Africa – capacity building through problem-solving research. Mine has truly been an experiential journey.

Your research journey

Researchers of today have a clear-cut focus/goal because smallholders are now specifically identified as the target and beneficiaries of research.

You must be aware that even if your research does not benefit the farmer immediately, ultimately it should. You must also to be aware that you cannot solve all their problems – you have to empower them to be able to carry out some research on their own in order to find solutions to suit their reality. It still stands that farmers need to be provided with a basket of options, such that they can choose or select options or components of a technology that suit their particular conditions.

Many advances have been made in conducting research in the field and in ways of involving the smallholders – their knowledge and participation is now fully recognised as being of core importance. Different ways of engaging stakeholders and integrating their participation are being tried out in many developing countries of the world. Farmer field schools were first tried out in Indonesia in the rice-based systems and have now spread to our continent. **The cost of on-farm research is high.** This high cost of field research makes it imperative that the farmers become part and parcel of the research and development process. The pros and cons of each approach or methodology should be well scrutinised, revised and refined before being used under African conditions. The costs of field research need to come down through the development of new approaches so that we can reach many more farmers. This is a challenge for all of you.

Africa has all the natural and human resources to produce adequate food for its people. Good supportive proactive policies for smallholders, that hold steady over a period of time, would allow the farmers to adapt and grow. **There is much to be done and it can be done.** We need more innovative indigenous approaches that are based on good science if we are to reach enough farmers.

Mapping the landscape/situation analysis

Mapping the landscape and analysing the situation are essential to any research. You need to collect all the available information or data on your topic/pest/disease/soil/system area so that you can sense where your research fits into the agricultural research agenda. This will help you understand what type of research you are doing and where it fits in the research continuum. It took over 20 years of research and interaction with NARS by an IARC for one chickpea variety to be released to farmers. Not much work had been done with chickpeas when the research project started. It began with germplasm collection, evaluation, and selection, and progressed through the initiation of a breeding programme, to field trials, seed multiplication and the eventual release of the new variety. Remember basic or strategic research will take longer to reach the main beneficiary (the small-scale farmer) than applied or adaptive research. The information that you collect while preparing for your research proposal will give you an indication of the importance and relevance of your research. Research results from NARS are usually found in their annual reports. Scientists of the NARS systems do not often publish their results in peer-reviewed journals as do the academics and scientists from IARCs. Most of our National Research and Extension Services (NARES) have a backlog of annual reports. Hence you will need to make personal contact with the staff of these vital services to get the information you need.

The role of universities

The main function of the university is to create and disseminate knowledge. The primary function is to provide degree training. But, since universities have highly trained staff, their research capacity and involvement in carrying out relevant problem-solving research for the country is vital.

In his lecture on 'The Institutions and the African Farmer', Eicher (1999) makes a logical case for the university to be part of the triangle which includes the national research institutes and extension services. The university/faculty needs to be part of this triangle if it is to provide you with the best possible training in agriculture. It is the university that produces the future staff of the national institutions. The three institutions should be closely linked. The need for them to work together collectively is obvious but not so easy to put into practice. In some countries integration has been attempted at some level as in Kenya where universities can apply for grants from the Kenya Agricultural Research Institute (KARI) to carry out specific pieces of research. Uganda has also gone some way along the road to integration. The World Bank's Agricultural Knowledge and Information System (AKIS) also supports closer integration of the three main national agricultural institutions: research, extension and the universities.

At the individual level interaction is possible but institutionalising integration requires much more effort and appreciation of each other's strengths and weaknesses. But working at the field level demands not only the input of research scientists based in provincial research stations but also the extension staff based at district/county or sub-county level.

Graduate research: learning by doing

You may have done a small research study as part of your Batchelor's degree. This might have involved some simple tests or measurements such as monitoring the spread of a disease in a plot or field or examining the effect of the some chemical on disease or pest control. This will have given you an insight as to what goes into a research study and how to measure the effect of something on a plant, or soil. Such experience helps you to become proficient in spotting a trend or an effect and in turn allows you to make a judgement or perhaps reach a conclusion.

The best way to learn is by getting your hands dirty. Share with the farmers, live with them, and learn their problems first-hand.

You learn from practical experience what effort farmers expend on agriculture; the time, energy, opportunity costs and risks they take to eke out a living. The understanding gained will be most useful to you for the rest of your life, and your post-graduate years are the best opportunity many of you will have to gain that first-hand experience.

At the MSc level the research period is usually about 2 years. At the field level this time would allow you to gather data over one or two rainy seasons depending on the crop you are studying and where you are situated in the unimodal or bimodal regions of Africa. At the PhD level your work will cover a period of 3–4 years. In this time you may or may not get a conclusive result but will gain an understanding of the problem or situation and realise that research can be a long process.

More ways than one

You should be aware that there are many ways to get an answer and that your chosen approach may or may not have added advantage over other approaches. You must be realistic in selecting an approach and be objective when drawing conclusions. You need to be aware of the pros and cons of each methodology. Discuss this with the biometricians in your faculty before embarking on your fieldwork. You should also be aware that in a biological living system everything is affected by everything else. The system is complex and the effect can be compounded. We tend to study effects of major variables – the variety being used, the status of the soil, rainfall pattern and management. You can look at the effect of one insect or pathogen or you may choose to look at the effect of a combination of factors such as the effect of diseases or pests on growth, or to study the interactions between the factors on the crop.

Your research could fill a gap in the knowledge, or it might be of an exploratory nature, that focuses on listing of all problems and issues linked to a certain crop and then prioritising the problems from the farmers' perspective and experience. Your research might just validate a result already obtained elsewhere in the country or region, or it could adapt technologies from elsewhere to local conditions.

You should assess the value of your research

Where does it fit in the farmers' priorities list? It may also be useful to see how it fits into district, provincial or national level priorities. What type of research? Is it basic or strategic, applied, or adaptive? Discuss all these aspects with your supervisor, NARS scientists, extensionists and others working in the field.

If possible you must try to work on a relevant and real problem that farmers face. Seek out academics who prefer to work in the field as your supervisors. 'Outreach' is now accepted as function of the universities and hence it is possible to do research in farmers' fields.

'Service orientation' (service to others, to your nation and to the farmers) should be part of the curriculum**.** Whatever type of research you undertake from laboratory-based molecular biology to economic/policy research – what you have to keep in mind is that the ultimate beneficiary is the small-scale farmer. We must all acquire a developmental mind set. As my boss used to say frequently *"Africans must solve Africa's problems"* – outsiders can assist or facilitate but Africans must take the lead in charting their destiny. **Your role is to take up that challenge and serve your country and the rural poor by applying your talents positively.**

At the MSc and PhD level you are just beginning the journey – which path you take will depend on your ambition and/or the jobs or opportunities available. Your first research project is a learning exercise – your first hands-on experience.

The journey is long but the satisfaction of being of service – bettering the lives of others, creating knowledge, training and seeing your name in print – makes it all worthwhile.

Important things to remember

1. Respect for all – everyone is important – the farmers, the extension worker, everyone involved in rural transformation.
2. Learning together – make this your motto – observe, question and keep on learning – you never stop learning because you do not know everything!
3. Be transparent in all your dealings – when you do not know something, say so – you can always find out!
4. Empower others – give them space to learn and lead.
5. Build solid relationships with the farmers – make their reality matter.
6. Development does not occur overnight – it takes time; be patient.
7. You will have to play many roles, researcher, advisor, listener, convenor, negotiator, facilitator, social worker and more......Be prepared........
8. Be a good team player.
9. Make room for reflection and enhance your analytical abilities.
10. Have a vision...an individual can make a difference ....one little mosquito can and does!

Resource material and references

Blackie, M. and Gibbon, D. 2003. *Enhancing Impact: Strategies for the Promotion of Research Technologies to Smallholders in Eastern and Southern Africa*. Natural Resources International Limited (NRIL), Aylesford, UK. 60 pp.

Eicher, C.K. 1999. *Institutions and the African Farmer*. Economics Program, CIMMYT, Third Distinguished Economist Lecture (15 January, 1999). Centro Internacional de Mejoramiento de Maíz y Trigo (CIMMYT), Mexico DF, Mexico. 64 pp.

Eicher, C.K. 2001. *Africa's Unfinished Business: Building Sustainable Agricultural Research Systems*. Staff Paper No. 2001-10. (May 8). Department of Agricultural Economics, Michigan State University, East Lansing. Michigan, USA.

Hagmann, J. 1999. *Learning Together for Change. Facilitating Innovation in Natural Resource Management through Learning Process Approaches in Rural Livelihoods in Zimbabwe*. Margraf Verlag, Weikersheim, Germany. 337 pp.

Hagmann, H., Chuma, E., Murwira, K., Connolly, M. and Ficarelli, P. 2002. Success factors in integrated natural resource managemnt R&D: Lessons from practice. *Conservation Ecology* 5(2): 29.

Kolb, D.A. 1984. *Experiential Learning. Experience as the Source of Learning and Development*. Prentice-Hall Inc., New Jersey, USA. 256 pp.

Sayer, J.A. and Campbell, B. 2001. Research to integrate productivity enhancement, environmental protection and human development. *Conservation Ecology* 5(2): 32.
url: http://www.consecol.org/vol5/1552/art32.

World Bank. 1999. *Integrating Universities into National Agriculture Research and Extension Systems; Good Practice for Investment in Agricultural University Programs*. Agriculture Knowledge and Information Systems (AKIS), Good Practice Note, World Bank, Washington DC, USA.

2.3 Approaches to impact-oriented agricultural research

Paul L. Woomer

- **You need patience, vision and team-work to bring about change**
- **Rural farmers need your help**
- **If possible work through local farmer organisa-tions**
- **You must know what type of research you are doing and where your research fits on the research continuum**
- **There are many successful participatory scaling up projects**
- **Be a life long learner- learn from others**
- **Do not let your research vision be narrowed by disciplinary bias**

'Providing insightful answers to difficult questions in ways that are useful to society is the principle responsibility of the scientific community.'

W.C. Booth *et al.* (1995)

'Documenting good ideas and potentially useful solutions through carefully conducted experimentation and case studies, in a manner that disentangles them from research fads and personal opinions, is the means to achieve these answers.'

Carl Sagan (1996)

Introduction

There is a renaissance in African agriculture. Many farmers in sub-Saharan Africa are moving from subsistence, cereal-based farming to market-oriented and mixed-enterprise agriculture. Threatened households find new uses for available resources that lead first to greater self-sufficiency and then to local and more distant markets. Farmers are diversifying into new enterprises such as confined livestock and poultry keeping and market gardening. They are forming community self-help and conservation groups and independent marketing associations and they are joining out-grower schemes. Several international, national and non-governmental research and development organisations are helping. A decades-long commitment to increasing human resources in agriculture adds force to the changes.

Stephen Carr (1997) identified five components for attaining rural transformation that provide guidelines on where your research could make an impact:

- Access to suitable improved germplasm
- Access to inputs
- Closely linked crop and livestock enterprises
- Road and communication infrastructure
- Information and market access.

What skills are required?

Professionally trained young Africans are keen to contribute to this rural transformation and to apply their knowledge. This is not easy. You will need patience, vision and teamwork: skills that may not be taught to undergraduates. You need patience because progress is always too slow for the eager. Agendas are inflexible, administrators are bureaucratic, donors are cautious, colleagues are preoccupied, sites are remote and farmers are conservative. A project rarely has impact within 2 or 3 years.

Teambuilding and teamwork are essential because successful research and development projects demand a spectrum of skills. The need for particular skills changes over time as the focus of efforts shifts from applied research to its adaptation and dissemination. Each team must be flexible and its members must be willing to move both forward and aside, depending on project needs and achievements. Honest criticism is essential to appraise team performance, and team members must remain open to the suggestions of others, even when they point to shortcomings and mistakes of individuals. As a team member you must be proud of the team's achievement as well as you own contributions; you must also learn to trust the instincts and abilities of others.

Two types of small households that need help

The poorest farmers suffer from food insecurity, and research efforts should focus on improving their food production and storage systems (Figure 1). When food supplies are exhausted from season to season, any improvement must be achieved within one season. The poorest households lack land, labour and the capacity to invest in farm inputs, so you should be realistic about the candidate technologies you explore.

You can help the poorest farmers by examining their management practices, discovering changes that will bring them the greatest improvements, and advising them on how to make more efficient use of their resources. You will find many practical limitations such as nutrient or moisture supply, pests and diseases, weed competition or crop genetic constraints, so a wide range of agricultural disciplines and perspectives are needed within a research team. Food insecurity also affects household nutrition, so you may be able to improve diets with advice on the productivity and diversity of the home garden. Households on very small holdings in peri-urban settings cannot be expected to achieve food self-sufficiency; they are more likely to benefit from research that promotes market-oriented gardening than increasing food production.

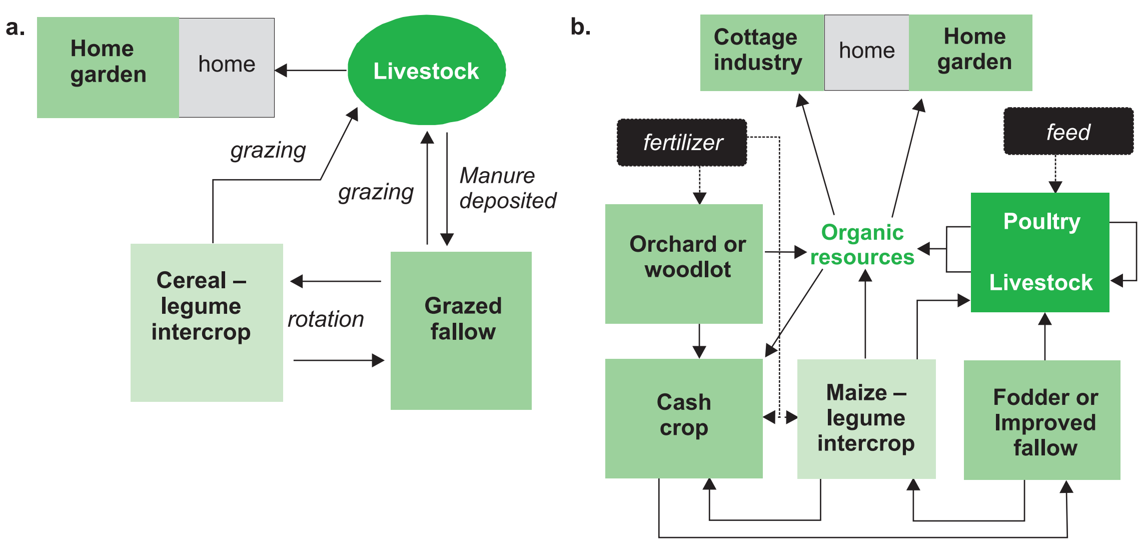


Figure 1. Major organic resource flows within: a. subsistence, cereal-based farming, and b. mixed enterprise, market-oriented agriculture. Note that resource flows within subsistence farming are more passive in nature and that more uses of organic resources emerge as farm operations diversify

Food-secure households offer a more diverse pallet of research and development opportunities. Management systems which economise on inputs and labour, or that substitute one for the other, offer special opportunities for research. Food-secure farmers are market-oriented, more open to changes and place more value on their time (Figure 1).

Working with farmers

Establishing a healthy, productive working relationship with farmers is an essential component of applied and adaptive research. Farmers' involvement in the research process may vary with different types of investigations. In researcher-designed and -managed off-station studies, farmers simply provide the research field and are consulted about local growing conditions. In researcher-designed, farmer-managed studies, the farmers' actions determine the experimental outcomes. In farmer designed-and-managed studies, as a researcher you can play a facilitating role in assisting farmers to better interpret results, and by treating farmers' impressions as sources of data. Regardless of the division of duties between you and the farmers, it is extremely important that both parties understand their rights and responsibilities (Table 1).

Table 1. Guidelines for successful collaboration between farmers belonging to self-help groups and researchers conducting on-farm studies

While doing research you should
- Involve cooperating groups and farmers in an earlier stage of research planning
- Rely upon simplified experimental designs and relatively few treatments and explain which treatments are intended as candidate improved technologies
- Establish a clear timetable and division of responsibility for field operations, data collection and record keeping
- Interpret their research findings into terms understandable by client farmers, particularly their costs and returns
- Be prepared to modestly compensate collaborators for their efforts and harvest removal
- Encourage farmers to conduct their own satellite experiments adjacent to the field trials.

You should not
- Perform unplanned on-farm field operations without the knowledge and consent of collaborators
- Fail to keep appointments or rearrange schedules without consulting collaborators
- Ignore collaborators' impressions of different management practices, particularly unrealistic reliance upon additional labour, land or expenses
- Exclude acknowledgement of community groups and key individuals in your publications.

Collaborating farmers should
- Make their own observations concerning field trials and express them at group meetings and to research partners
- Organise local field days that demonstrate the tested technologies to their communities
- Make a genuine effort to understand the scientific basis for treatment selection and sampling procedures so that promising results can become adapted into farm practice.

Collaborating farmers should not
- Falsify data collection records, disguise experimental failures, or exaggerate claims for compensation
- Remove crop harvests without the knowledge and agreement of research partners
- Expect researchers to engage in lengthy social interactions during intensive field campaigns.

You must not dominate the research process, and should take time to explain your intentions and approaches to farmers. Establishing regular consultation and firm timetables, and relying upon simple experimental designs with understandable treatment combinations are important means to this end (Box 1). Hiring family members to assist in routine measurements is one possible reward mechanism and you could offer a small compensation for the use of the land. Don't intrude too much and keep your social interactions with farmers to a respectful minimum.

Box 1. On-farm miscommunication

- Patty's MSc research involved large field experiments on three farms. She would arrive at irregular intervals, deride farmers for lateness of field operations, refuse to consider claims for compensation and argue with fellow students and technicians during data collection. During her defense, she acknowledged the assistance by 'the farmers' without referring to them by name. Three years later she received a PhD scholarship and sought to work in the same general locations but experienced difficulty in securing collaborating farmers!
- Bill and his team of field technicians were experts in installing on-farm trials, completing as many as 10 per day. They would arrive, mark the plot boundaries, apply pre-packaged inputs, plant the seeds, label the plots, thank the farmer, Cornelius, and be gone in a matter of minutes. Cornelius established an independent satellite experiment by carefully marking the plots in string and planting along a marked 10-m line, just as the researchers did. But there was no improvement in crop performance because nobody had bothered to explain which inputs were being examined and why!
- A team of researchers scouting for new field sites chanced across a farmer burning brush next to a slightly chlorotic stand of groundnuts. They observed satisfactory root nodulation and concluded that the groundnut suffered calcium deficiency. They carefully demonstrated how to apply wood ash to the groundnut pegs and then recommended that the farmer do the same. He replied "Why should I waste this wood ash on my neighbour's groundnuts?" Because no obvious boundary existed, the researchers had assumed a contiguous landholding!

Advantages of working through local farmers' organisations

Projects designed to achieve agricultural impacts almost always involve a stage of on-farm research. It is common to select collaborators from a list of farmers obtained from local authorities. Where possible, work with self-help groups. Self-help groups are formed as a means to access information and to assess new technologies. Groups of neighbouring farmers share common concerns and it is reasonable that they organise themselves for collective action. They share information, learn new technologies and pool resources to acquire inputs or to market surpluses. Farmers who belong to self-help groups are often enthusiastic and capable collaborators.

Among the advantages of working with local groups is that it is easier to assemble farmers to explain the goals of a research project and to identify collaborators. Training in experiment installation, measurement and record keeping can be organised more easily through the officers of a local self-help group, whose keenest members can be recruited as trainers. Identifying one farmer as a local coordinator tends to invoke less rivalry when all of the collaborators belong to the same self-help group, especially when the appointment is made in a participatory manner. It is easier to establish experimental schedules and to adjust them when working with groups of farmers. During experiments, peer pressure among group members will ensure that tasks are performed correctly and on time, such pressure does not exist among independently recruited farmers. Similarly, the rights and responsibilities of collaborating farmers are more easily established and enforced. Farmers within groups will tend to voice their opinions more openly and to be bolder in challenging your actions and highlighting your misconceptions. Local self-help groups are in a better position than you are to organise farmer field days to promote research findings. They often have important contacts among neighbouring groups and local authorities that will ensure these events are well attended. An example of a very successful self-help group in rural Kenya, the St. Mark's Women's Group, is presented in Box 2.

Box 2. St. Mark's Women's Group, Amagoro, Teso, Kenya

Teso lies to the south of Mount Elgon in Western Kenya. It has infertile sandy soil and until recently was primarily used for grazing. The conversion to sedentary agriculture resulted from increased population and establishment of land titles but was accelerated by an epidemic of East Coast fever, a viral disorder that decimated the local cattle population. The St. Mark's Women's Group was started by 30 church members in 1998 as an outgrowth of a prayer group. Its original goals were poverty alleviation and improved child nutrition. The group has five elected officials; a Chairperson, Vice-chairperson, Secretary, Vice-secretary and Treasurer who are elected for 3-year terms. The membership in 2003 was 52 and the group is locally recognised as an effective and equitable community-based organisation, partly because of its widely attended field days and its successful efforts in processing and marketing traditional crops.

The group's primary collaborator is the Sustainable Agriculture Centre for Research Extension and Development in Africa (SACRED-Africa) that initiated a local outreach project in partnership with St. Mark's and other local organisations in Teso in 1999. Relying upon participatory methods for problem identification and a simple adaptive research process, progress was made in the areas of composting, soil fertility management, tree seedling establishment, integrated pest management, crop diversity, marketing farm surpluses and gender roles in agriculture. The St. Mark's group also serves as one of seven co-operators in the Best-Bet Network, a group that evaluates alternative land management recommendations side-by-side on 140 farms in Western Kenya. Exposure to different maize–legume intercropping technologies has demonstrated to members how to raise yields from less than 2700 kg/ha to over 4200 kg/ha, increasing net returns by US$200 per crop. In the process, members developed new skills in recordkeeping and fertilizer use, increasing their experimental successes from 70–87% within a single year.

After 5 years of operations, several impacts from the group are evident. Their rapid bulking and broad distribution of cassava resistant to the mosaic virus promoted food security within the group and among neighbouring farmers. When most other cassava in their district was failing, this group had established over 240 ha of cassava throughout the area. The adoption of a maize–lablab relay fallow has demonstrated that sustainable field cropping can be achieved on the worst of N-deficient sands. Traditional green vegetables and small grains that were previously considered a home gardening activity now have established markets. But the benefits from the group's activities extend beyond technical adoption because its members now view agriculture in a more holistic and positive manner.

Members are able to diagnose new problems as they arise and to better apply past lessons to emerging situations. The underlying mechanisms for the degradation of agricultural resources are now better understood, as are the relationships between various conservation measures. As Jenipher Etiang', the group's Chairperson, stated "We discovered that we had many resources at our disposal that we were not using well and the relationship between the problems that we were having and our present and past actions. It was a turning point in our lives." The group is frequently visited by members of other organisations from Kenya and neighbouring countries, by officers from the local Ministry of Agriculture and by local politicians who attend field days to make modest donations. Members assist one another with medical and funeral expenses and through small loans because they know their neighbours can now generate income from farming. Even domestic lives have improved, as is evident from Jenifer's comment "Women no longer bother husbands for money to buy salt, sugar or tea leaves and this has improved our family relationships."

How to navigate the research and development continuum

Along the research and development continuum (Figure 2), 'upstream' theory and basic research is translated into potentially useful technologies that are then applied under field conditions. Next, these candidate technologies undergo 'downstream' adjustments based upon the needs of potential users and their site-specific conditions become captured into

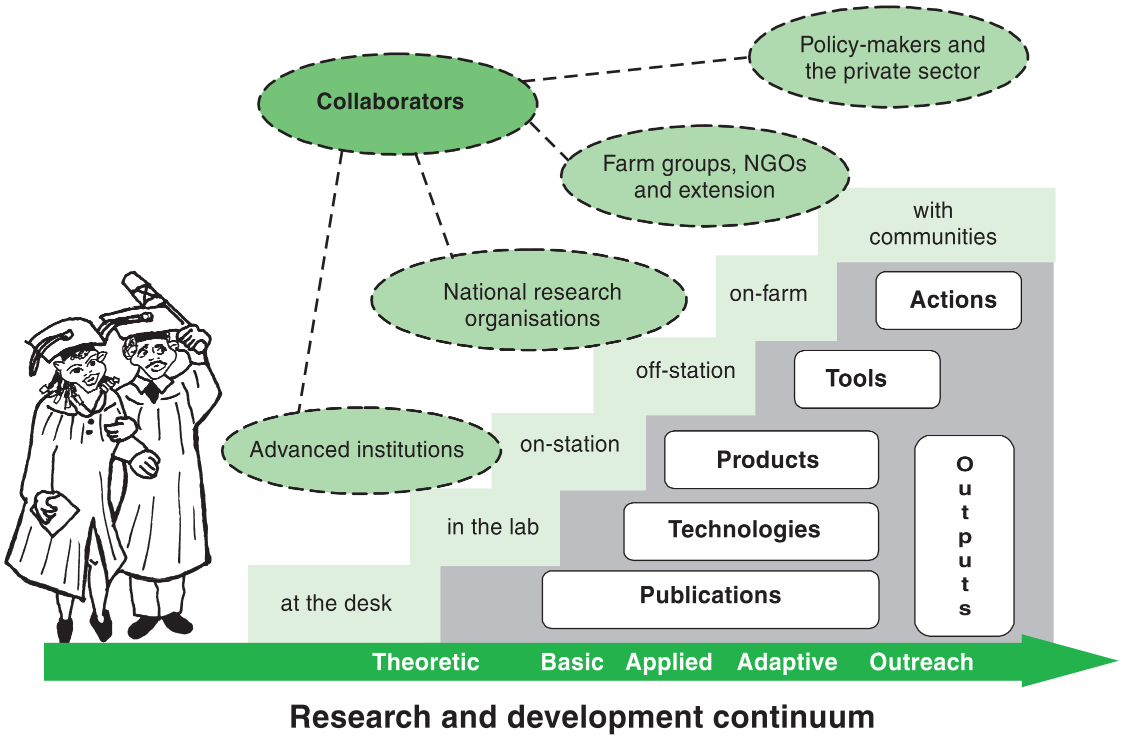


Figure 2. Research intended to achieve impacts through collective action is conducted in a series of settings and with changing sets of cooperators, leading to a sequence of more widely distributed research products

products that are tested under a wider range of conditions, and are then offered to communities as a possible solution to common problems. This process requires several years at best, and while relatively few candidate technologies are likely to become widely adopted by society, actions and choices taken by scientists as they escort research products along the continuum can greatly affect the fate of their innovations.

While the process appears straightforward, navigating the research and development continuum requires teamwork, vision and persistence. Innovators must select the appropriate setting and collaborators, both of which change as their innovations undergo testing and become refined. Theoretical and basic research leading to new ideas, is performed at the scientist's desk and in the laboratory, and is best facilitated through collaboration with leading scientists from advanced institutions. On-station studies translate these ideas into candidate technologies, with emphasis on better understanding of the underlying mechanisms that govern their effects.

When studies are moved away from the research station, different collaborators are needed with expertise in packaging technologies and examining their cost-effectiveness. If these technical packages offer promise, they can be scaled up. You must remember that farmers will innovate and adapt your technology while trying it out. Accept this – they know the reality of their situation better than you. If you choose your research setting and collaborators wisely as you escort your innovations or your contribution to an innovation, 'downstream' you are much more likely to achieve impacts than someone who does not. Make sure that you have solid well tested technologies to pass on.

The case of staggered intercropping (MBILI) in Western Kenya presents an interesting example of movement along the research and development continuum. The farmers' main enterprise, maize–bean intercropping, was performing poorly, primarily because of inadequate nutrient supply to the maize and a severe pest-and-disease syndrome on the beans. Researchers had examined a wide range of solutions over many years, including mineral nutrient replenishment, short-term improved fallows, green manure and cover cropping, breeding for stress, pest and disease tolerance and detailed integrated pest management (IPM) strategies but none of these interventions were reaching farmers, partly because they overestimated farmers' land, labour and ability to invest. A simpler solution was required, one that involved reconfiguration of farmer-available resources.

By simply staggering the 75-cm maize rows into alternating 50-cm and 100-cm rows (Figure 3), and orienting the rows in an East–West direction when possible, substantial improvement was achieved. This adjustment allowed for the inclusion of additional, higher-value legumes such as groundnut and green gram into the intercropping system, crops the farmers already knew but only grew as occasional monocrops. These legumes fixed more symbiotic nitrogen, resulting in strong residual benefits to maize, and, when grown as a

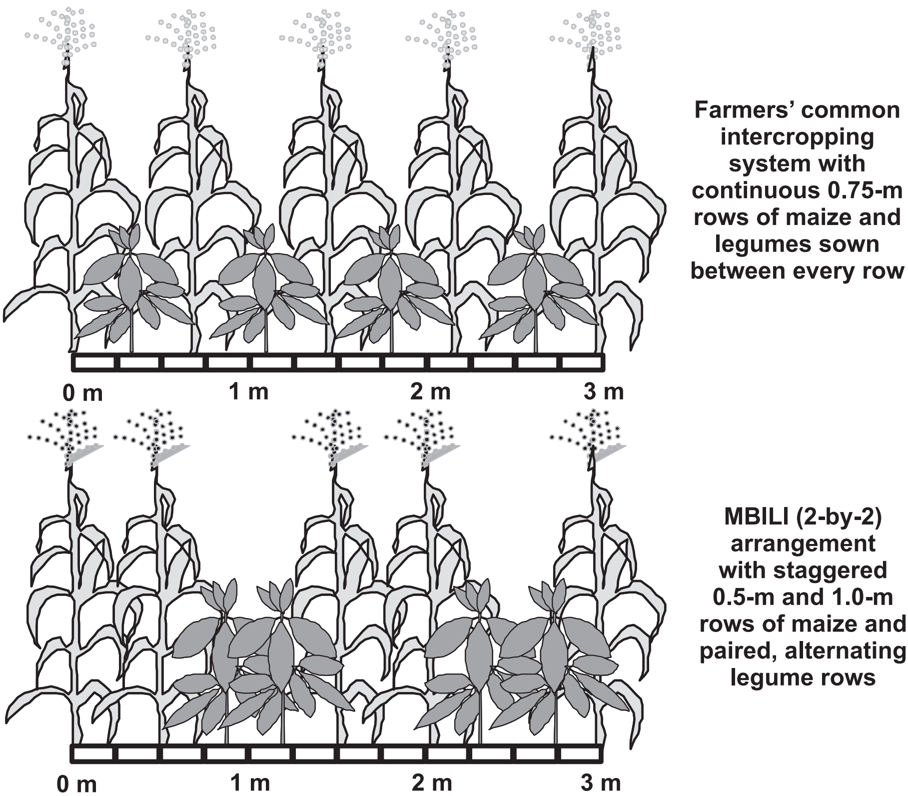


Figure 3. The MBILI staggered maize–legume intercropping system that improves maize and legume production and provides better market opportunites in Western Kenya

maize–legume intercrop rotation (maize–groundnut, maize–green gram, maize–bean), greatly reduced the incidence of legume pests and diseases.

This MBILI intercropping adjustment was first proposed for Western Kenya by a university scientist in 1999 based upon an understanding of current farmers' problems and the volumes of literature on intercropping, relay and strip-cropping. MBILI was field tested at a farmer field school later that year, and then expanded to eight farms in 2000. In 2001, it was tested on 32 additional farms by a local NGO and featured at several agricultural shows and farmers field days. In 2002, it was included as one of eight candidate intercropping technologies for testing on 140 farms over 2 years by a research network, where it emerged as the most cost-effective management system. In 2003, MBILI was featured in a regional farming magazine and formal collaboration was established with officers from the Ministry of Agriculture, allowing the practice to be further promoted by extension agents. Only 5 years after its inception, many of the poorest farmers in Western Kenya credit MBILI for their escape from hunger, while better resource-endowed farmers rely upon it to produce greater crop surpluses for market. The progress of MBILI along the research and development continuum serves as an example of how rapidly a new technology can reach its intended clients when different partners within the agricultural community all play their respective roles, and when that technology matches farmers' needs and resources (see Box 3).

Box 3. Two perspectives on MBILI's benefits

Agronomic A baseline study of 107 farms in Western Kenya revealed intercrop yields of 1197 kg maize and 192 kg beans/ha, offering a net return of US$72/ha. Introduction of the MBILI intercropping package, involving row adjustment, substitution of groundnuts for beans and modest fertilizer application (31 kg N and 20 kg P/ha) resulted in 2431 kg maize and 360 kg groundnut/ha. The MBILI package required additional investment of US$63/ha, but increased returns by US$183.

Personal "I am a housewife with 6 children. I plant MBILI on my farm and I have found the system very paying. When I used to plant maize and groundnuts using conventional methods I could get 5 bags (250 kg) of maize and 1 bag (33 kg) of groundnuts from 0.2 hectares of my land. Under MBILI planting, I got 12 bags (600 kg) of maize and three bags (100 kg) of groundnuts which was a record yield for me. The rows in MBILI receive enough sunlight and it reduces wastage of land..."

Purity Nalianya, Chililila Woman's Group, Bungoma, Western Kenya

From project outputs to farm impacts

Dissemination and outreach strategies are at the other end of the research and development continuum. From the social perspective, science exists to provide useful solutions to human difficulties, a goal that requires that the research community be linked to the public. In developing nations science is largely confined to academia, and society has yet to recognise a substantial improvement through investments in research. Donor investment in applied research is usually linked to such pressing social concerns as food security or public health, and important scientific findings are expected to become translated into flexible tools for development. Thus, dissemination and community outreach (Figure 4) are natural conclusions of a successful project.

Within the context of agricultural research, the progression from on-station to off-station and to on-farm research proceeds in a stepwise manner. You must understand the mechanisms, ranges and suitability of a new technology, so that you can then identify opportunities, options and potential impacts (Figure 4). Many farmers consider all technologies tested on-

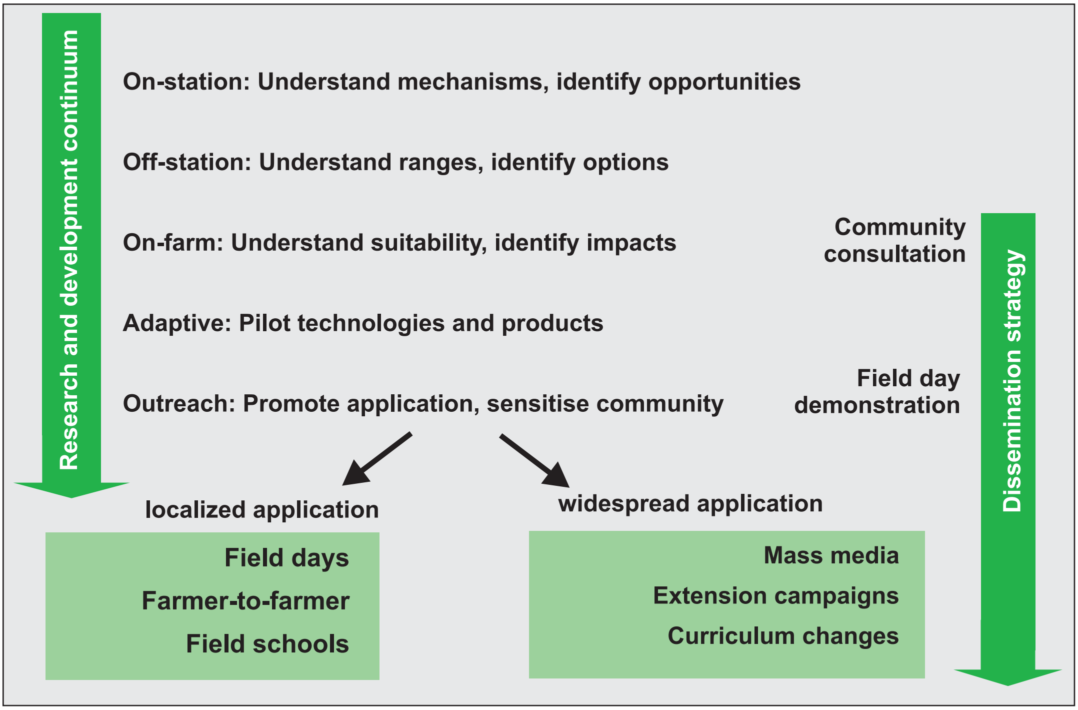


Figure 4. As the research and development continuum proceeds downstream, it interacts with different possible dissemination and community outreach strategies. Innovators are encouraged to escort technologies beyond their experimental stages in order to achieve greater impacts from their research

farm to be 'recommended', so you must make it clear to farmers that you are refining or testing a technology and that it is not a recommendation. As research becomes more participatory and demand-driven, dissemination moves further 'upstream', a phenomenon that requires you as a researcher to be more responsible and conservative about the technologies you might propose as candidate solutions to farmers' problems. The risk of miscommunication is reduced as farmers increasingly recognise their roles as full partners in adaptive research.

One useful approach to on-farm adaptive research is to test current recommendations and various 'best bet' technologies side-by-side across a range of cropping conditions, because many competing 'recommendations' are developed in relative isolation from one another. For example, the effects of applying composts are seldom tested against mineral fertilizers, or fallow systems are compared to one another but not to direct soil amelioration. You should welcome the opportunity to test your candidate solution alongside others. The fusion of alternative management practices represents an exciting avenue of research, particularly in partnership with farmers empowered to combine and manage different recommendations as they see fit. If you treat farmers' planning and reactions as useful information if will be an important step in realising fuller partnership with them during your adaptive research.

During project planning you should devote funds to the promotion of your findings. This can be done through farmers' field days, or the facilitation of adjacent farmer-to-farmer training, particularly through local extension agents and nearby self-help groups. Widespread application is effected through mass media campaigns, exhibits in regional or national agricultural shows, designing extension campaigns, and inducing curriculum changes within public school systems. It is generally difficult to plan and budget beyond localised outreach at the onset of a research project but you should try to liaise with senior extensionists, educators and journalists in order to popularise your findings to wider audiences. Table 2 identifies different outreach options, indicative costs and likely beneficiaries.

Table 2. Returns to US$1000 investment in various dissemination options

Option	Audience		Unit cost (US$)
Field day attended by	100	participants	10.00
	500	participants	2.00
	1000	participants	1.00
Printed media prepared and distributed	2000	copies	0.50
later each copy read by nine more farmers	20000	readers	0.05
Video documentary prepared and broadcast	50000	viewers	0.02
Video documentary taped and distributed	100	recipients	10.00
later seen by 99 additional viewers	10000	viewers	0.10
Field school initiated with	25	members	40.00
Initiated with	50	members	20.00
and each member trains 9 farmers	500	trainees	2.00

Pitfalls you should avoid

Intense disciplinarity

Agricultural science is not a single discipline, but rather the sum of numerous sub-disciplines. These pursuits include agronomy, animal husbandry and nutrition, crop improvement, economics, food science, horticulture, pest and disease management, post-harvest handling, rural sociology, soil science and many other fields. Modern society is complex, and this is especially true within science. Indeed, most of the greatest scientific discoveries result from several years, or even decades, of intense, highly specialised, often obsessive study. But this model for scientific achievement is less applicable within the more-applied disciplines such as agriculture. Agricultural science still involves discovery, but within an iterative, problem-solving context. **Support for agricultural science is society's insurance against malnutrition and famine.** Most of agricultural science starts not with theory, but in the field, where plant production is failing. This situation is especially true in Africa, where the continent's capacity to feed its people is failing.

Even at the undergraduate level, young agriculturalists are expected to declare an area of specialisation. Graduate candidates specialise further, often leading to detailed knowledge within a sub-sub-discipline, and little else. This situation is most severe in universities where graduate studies do not include coursework. **Do not let your research vision be narrowed by disciplinary blinkers.**

I recall a farm visit in the Central Kenyan Highlands accompanied by a noted virologist. This farm was extremely diversified, and included a number of different vegetables intended for Nairobi markets. After discussing crops and markets with the host farmer, and touring the fields, the virologist confided that she did not enjoy the visit because the crops were all healthy and none displayed symptoms of viral disorders. At the time I sympathised, and promised to better direct future field visits, but afterwards I recognised a perverse logic. The absence of pest and disease is not random, or the result of good fortune, but rather nested into farm practices and the farmer's skills and experience. Even a famous virologist should be prepared to learn something from a farm where viral disorders are controlled. This analogy may be extended to soil fertility specialists and the lack of nutrient deficiency symptoms, economists and strong markets, entomologists and harmful insects, agroforesters and trees and all other agricultural specialists. It is naïve to assume that farm productivity is necessarily restricted by factors related to your chosen sub-discipline, and this becomes more so as you specialise further. When agricultural scientists behave like the blind men attempting to describe an elephant, then they are practicing precise, and most likely inaccurate, disciplinarity.

Survey-mania

Too many talented scientists focus upon describing problems that are already known, rather than exploring possible solutions. As a field researcher you should conduct your own 'participatory, diagnostic survey' at an early phase of every investigation. Concise survey instruments consisting of less than 20 short questions that are based upon important research questions are a useful first step. Many farmers grow weary and wary of answering questions about family members, land tenure, income and education, so your survey should be short and conform to the task at hand.

Over-delegation

Delegation is a responsibility in a research setting, it entails the optimal use of available time, funds and staff to ensure scientific quality and meet project goals and it requires that delegated tasks be closely supervised. Responsible delegation breeds scientific teamwork and mutual satisfaction, irresponsible delegation leads to division and inferior science. For example, I once visited an agricultural research station and noted a large, vigourous stand of an indigenous legume that was being examined elsewhere as a candidate for land restoration efforts. Some controversy surrounded the nodulation of this legume, so I asked a station scientist if I might borrow a shovel and conduct an impromptu investigation. He agreed to help, and consulted the Chief Technician. After about one hour, a stranger arrived to deliver my 'bag of soil'. When I asked where the soil came from, he could not say. Evidently, the Chief Technician consulted the Laboratory Technician who informed the Field Technician who instructed the Field Labourer who put some soil in a bag for reasons not understood. Obviously, my intentions were mistaken in the process. This 'tradition' of over-delegation, where scientists and senior technicians avoid 'unnecessary' field labour and where important tasks are performed by unqualified casual workers, poses a serious hazard to quality research.

Compulsive home-area focus

Many young agricultural scientists in Africa feel compelled to conduct studies in their home areas. Clearly, advantage exists for these studies. They have an intuitive understanding of production constraints, are able to communicate with the poorest farmers in the most local dialects and have ready access to field sites. During field visits, workers are often able to economise on accommodation. Moreover, working within your home area fulfils a social obligation to return the benefits of education to your family and immediate community. All of these factors offer clear advantages in terms of research position and motivation.

But a suite of disadvantages also exists. Your home area may be remote, and may suffer from similar constraints to a closer more easily accessible location. You may not have the expertise to address the constraint that exists in your home area. Furthermore, social obligations can become confounded with research strategies in a compromising manner. Devote yourself to good science in the place your work can have most impact (Box 4).

Box 4. Where will your work have impact?

Mary receives an advanced degree in plant pathology overseas, finds employment in an African public university and returns home intent upon 'proving' herself. She has specialised in bacterial diseases of a crop that is not widely gown in her home area, but nonetheless she bases her research project upon it. After several years, she has produced a research publication but failed to impact upon farmers lives. Next she employs her skills within an interdisciplinary team active in another agroecological zone where the crop is very important and within one year is assisting farmers to better understand their production problems and improve their yields. Where, and with whom should she have initiated her research?

Conclusion

Well-trained scientists practice credible research within their immediate sub-discipline almost by instinct.

As practicing agriculturalists in Africa, given the continent's persistent problems of under- and malnutrition, resource degradation and rural stagnation, identifying important questions requires that we look no farther than into the lands, crops and practices of the poorest farmers, particularly into their main food production enterprises. Escorting communities from failing subsistence farming to mixed, market-oriented agriculture is a major responsibility of the agricultural research community, and if you cannot identify your particular role within this key transformation perhaps you should rethink your longer-term research interests. Let us not merely identify the farmers' problems, but work together to solve them. Let us not be dazzled by proliferating agendas and research fads, but rather focus upon the basic needs of the poor and their abilities to secure better lives. Let us not work as individuals on those problems that comfortably address our disciplinary bias, but work together with other specialists and generalists to ensure that food security and better rural livelihoods result from our combined efforts. This process requires patience, vision and teamwork, and its completion is not likely to be repaid in riches, but rather rewarded through personal satisfaction and social wellbeing.

Resource material and references

Appendix 2. Innovation, Problem Solving and Operational Research Strategies. Paul L. Woomer. PowerPoint on CD.

Appendix 3. Designing Research Around Client Needs. Paul L. Woomer. PowerPoint on CD.

Abate, T. and Ampofo, J.K.O. 1996. Bean pests in Africa. *Annual Review of Entomology* 14: 45–73.

Amadalo, B., Jama, B., Niag, A., Noordin, Q., Nyasimi, M., Place, F., Franzel, S. and Beniest, J. 2003. *Improved Fallows for Western Kenya: An Extension Guideline*. World Agroforestry Centre (ICRAF). Nairobi, Kenya. 48 pp.

Booth, W.C., Colomb, G.C. and Williams, J.M. 1995. *The Craft of Research*. The University of Chicago Press, Chicago, USA. 294 pp.

Carr, S.J. 1997. A Green Revolution frustrated: Lessons from the Malawi experience. *African Crop Science Journal* 5: 93–98.

Carr, S.J. 2003. More people more soil degradation; the Malawi experience. *African Crop Science Conference proceedings*, Vol. 6. (In press).

Chambers, R., Pacey, A. and Thrupp, L.A. 1989. *Farmer First: Farmer Innovation and Agricultural Research*. Intermediate Technology Publications, London, UK. 219 pp.

Den Biggilaar, C. 1991. Farming systems development: Synthesizing indigenous and scientific knowledge systems. *Agriculture and Human Values* 8: 25–36.

Eicher, C.K. 1999. *Institutions and the African Farmer*. Economics Program, CIMMYT, Third Distinguished Economist Lecture (15 January, 1999). Centro Internacional de Mejoramiento de Maíz y Trigo (CIMMYT), Mexico DF, Mexico. 64 pp.

Food and Agriculture Organization of the United Nations (FAO). 2002. *FAO Yearbook 2001: Production*. Vol. 55. FAO, Rome, Italy.

Harrison, P. 1996. *The Greening of Africa: Breaking through the Battle for Land and Food*. Academy Science Publishers, Nairobi, Kenya. 380 pp.

Lacy, W.B. 1996. Research, extension, and user partnerships: Models for collaboration and strategies for change. *Agriculture and Human Values* 13: 33–41.

Lynam, J.K. and Blackie, M.J. 1994. Building effective agricultural research capacity: The African challenge. In: *Agricultural Technology: Policy Issues for the International Community*, Anderson, J.R. (Ed.), pp. 106–134. CAB International, Wallingford, UK.

Misbahulmunir, M.Y., Sammons, D.J. and Weil, R.R. 1989. Corn–peanut performance in relation to component crop relative planting dates. *Agronomy Journal* 81: 184–189.

Mureithi, J.G., Gachene, C.K., Muyekho, F.N., Onyango, M., Mose, L. and Magenya, O. 2002. *Participatory Technology Development for Soil Management by Smallholders in Kenya*. KARI Legume Network Project, Kenya Agricultural Research Institute, Nairobi, Kenya. 551 pp.

Nicholson, S.E. 2001. Climatic and environmental change in Africa during the past two centuries. *Climate Research* 17: 123–144.

Norman, D.W., Francis, C.A. and Henrich, G.M. 1994. Providing relevant education and training for sub-Saharan African agricultural scientists: foundation for a sustainable future. *Journal of Sustainable Agriculture* 4(3): 79–90.

Omare, M.N. and Woomer, P.L. 2002. Innovations in organic resource management: Experiences from Kenya's Forum for Organic Resource Management and Agricultural Technologies. *Outlook on Agriculture* 31: 235–242.

Patel, B.K. and Woomer, P.L. 2000. Strengthening agricultural education in Africa: The approach of the Forum for Agricultural Resource Husbandry. *Journal of Sustainable Agriculture* 16(3): 53–74.

Sagan, C. 1996. *The Demon-Haunted World: Science as a Candle in the Dark*. Ballantine Books, New York, USA. 457 pp.

Saint, W.S. 1992. *Universities in Africa: Strategies for Stabilization and Revitalization*. World Bank Technical Paper No. 194. Africa Technical Department Series. World Bank, Washington DC, USA. 140 pp.

Smaling, E.M.A., Nandwa, S.M. and Janssen, B.H. 1997. In: *Replenishing Soil Fertility in Africa*. Bursch, R.J. Sanchez, P.A. and Calhoon, F. (Eds.), pp. 47–61. Soil Science Society of America (SSSA) Special Publication No. 51. Madison, Wisconsin, USA.

Squires, V.R., Glenn, E.P. and Ayoub, A.T. 1998. *Combating Global Climate Change by Combating Land Degradation*. United Nations Environmental Programme (UNEP), Nairobi, Kenya.

Terrent, J. and Poerbo, H. 1986. Strengthening community-based technology management systems. In: *Community Management, Asian Experience and Perspectives*. Korten, D.C. (Ed.), pp. 172–182. Kumarian Press, West Hartford, Connecticut, USA.

Tungani, J.O., Mukhwana, E.J. and Woomer, P.L. 2002. MBILI is Number 1: A *Handbook for Innovative Maize-Legume Intercropping*. Sustainable Agriculture Centre for Research Extension and Development in Africa (SACRED)-Africa, Bungoma, Kenya. 20 pp.

Vedeld, T. 2000. Village politics: Heterogeneity, leadership and collective action. *The Journal of Development Studies* 36: 105–134.

Wahua, T.A.T., Babalola, O. and Aken'ova, M.E. 1981. Intercropping morphologically different types of maize and cowpeas: LER and growth attributes of associated cowpeas. *Experimental Agriculture* 17: 407–413.

Woomer, P.L, Bekunda, M.A., Karanja, N.K., Moorehouse, T. and Okalebo, J.R. 1998. Agricultural resource management by smallhold farmers in East Africa. *Nature and Resources* 34(4): 22–33.

Woomer, P.L., Mukhwana, E.J. and Lynam, J.K. 2002. On-farm research and operational strategies in soil fertility management. In: *Integrated Plant Nutrient Management in Sub-Saharan Africa: From Concept to Practice*. Vanlauwe, B., Diels, J., Sanginga, N. and Merckx, R.. (Eds.), pp. 313–331. CABI Publishing, Wallingford, UK.

Woomer, P.L., Musyoka, M.W. and Mukhwana, E.J. 2003. *Best Bet: Maize-Legume Intercropping Technologies and Summary of 2002 Network Findings*. Sustainable Agriculture Centre for Research Extension and Development in Africa (SACRED)-Africa, Nairobi, Kenya. 18 pp.

Woomer, P.L., Omare, M.N. and Mukhwana, E.J. 2003. The Operations of Rural Self-Help Groups. In: *Organic Resource Management in Kenya: Perspectives and Guidelines*. Savala, C.E.N., Omare, M.N. and Woomer P.L. (Eds.), pp. 131–145. FORMAT, Nairobi, Kenya.

2.4 Development and management of projects

Bharati K. Patel

- **Projects provide a framework for a piece of research**
- **The logical framework matrix is commonly used in project design**
- **You will always be a part of a research project and a research team**
- **Be a good team member**
- **A large research undertaking can be divided into smaller projects**

'The success or failure of the many community projects and farmer associations that are being formed in Africa will depend on the availability of profitable technology for small-scale farmers. This situation harkens to some of the hard choices that Asian political leaders made in the 1960s to build agricultural knowledge triangles. Will African political leaders make the same hard choices in building strong national agricultural science bases or will they continue to rely on bottom-up, quick-fix projects and a gaggle of food aid subscriptions? Only time will tell.'

Carl K. Eicher (2003)
Agriculture in the Global Economy

As a graduate student you will first need to write a research proposal that will be approved by a research committee before you can do your research project. Your main aim is to get a degree and thus your project should be 'do-able' in the timeframe available. Projects and proposals are closely linked and have common features. Smaller discrete projects can be developed from a large proposal or small proposals for funding can be developed from a large project.

This chapter deals with how a project is developed and managed. A project has many facets and provides a framework for a piece of research or work. It usually involves other persons, which implies partnerships and interactions with colleagues, peers, students or clients. A project is only a part of a larger agenda. Everyone involved in a project is important and has a role to play and it is essential that all those involved in a particular project are committed to it and share a common vision. Throughout your research career you will be involved with projects and will be a part of a team.

Most agencies and organisations today utilise a project design approach based on some version of a logical framework matrix (LFM) and a work breakdown structure (WBS) of project activities. The LFM and WBS facilitate not only project design but also progress reporting and evaluation. The LFM is a one or two-page overview of a project which summarises its design information. The WBS is a one-page graphical presentation of the project which links the goal, purpose and outputs to specific project activities.

Logical framework matrix (LFM)

Begin by preparing the LFM as shown in Table 1. On the vertical axis (first column) of the LFM Goal refers to the broad programme or sector goal to which the project is expected to contribute. It represents a development objective.

Table 1. Logical framework matrix (LFM) for project design

Narrative summary	Objectively verifiable indicators	Means of verification	Critical assumptions (beyond the control of the project team)
Goal			
The reason for the project, the desired end toward which the efforts are directed (*e.g., to raise farm yields*)	Conditions which will indicate that the goal has been achieved (*e.g., higher yields in farmers fields*)	The way that goal achievement can be objectively verified (*e.g., seed sale records, government yield statistics*)	Assumptions which must be met if purpose is to result in achievement of the goals (*e.g., effective seed systems exist*)
Purpose			
The anticipated benefit if the project is completed successful and the outputs are actually utilised (*e.g., improved varieties released by* NARS)	The signs which will indicate that the purpose is achieved (*e.g.,* NARS *release at least three varieties*)	The way that purpose achievement can be objectively verified (*e.g., official varietal release documents*)	Assumptions which must be met if outputs are to result in achievement of the purpose (*e.g.,* NARS *skilled and funded to effectively test and release varieties*)
Outputs			
The specific kind of results that are expected from good management of the project inputs (*e.g., improved elite lines*)	The tangible form in which the outputs can be observed (*e.g., breeder's advanced germplasm collection*)	The way these outputs can be objectively verified (*e.g., yield trial results*)	Assumptions which must be met if Inputs are to result in achievement of the outputs (*e.g., useful genetic variation exists*)
Inputs			
Resources for activities necessary to produce the outputs (*e.g., basic germplasm*)	The tangible form in which the Inputs can be observed (*e.g.,* 1000 *new accessions used*)	The way these inputs can be objectively verified (*e.g., field books and plantings*)	Assumptions which must be met if Inputs are to be attained (*e.g., importation of exotic germplasm is allowed by national authorities*)

Source: ICRISAT, Guidelines for proposal/project development (personal communication)

Purpose is the immediate objective that describes the intended impact of the project on the direct beneficiaries, but is beyond the direct control of the project team since it relies on how the beneficiaries will make use of the project outputs. A project should only have one immediate objective or purpose.

Outputs are the specific kinds of results that can be expected from good management of the project inputs and activities. The project team should be held accountable for the production of the outputs.

Inputs are the resources (funds, personnel and goods) that are needed for the production of the outputs.

Activities are not listed in the LFM since they are provided in the work breakdown structure (WBS).

On the horizontal axis of the LFM, **indicators** are parameters, preferably those that can be quantified, which verify the achievement of the goal, purpose and outputs. Indicators provide a basis for monitoring and evaluation, they should define attainment in terms of target group (for whom), type of output (what), quantity (how much), quality (how good), time (by when) and location (where).

Means of verification should specify both the instrument for measuring the indicator and the sources of information necessary to use the indicators (e.g., questionnaires and structured interviews, the results of which are found in ministry statistical reports, project technical and financial progress reports).

Critical assumptions are the events or conditions over which the project team have little control but which must be assumed to exist if the inputs are to be applied, the outputs delivered, and the objectives achieved. These external factors determine the risks of the project. They clarify and set the limits of responsibility for the project management. Projects may be unsuccessful because:

- Unreasonable assumptions were made during the design phase
- Reasonable assumptions do not hold up
- Inputs are poorly managed.

The identification and clear expression of the assumptions are therefore extremely important in the evaluation process.

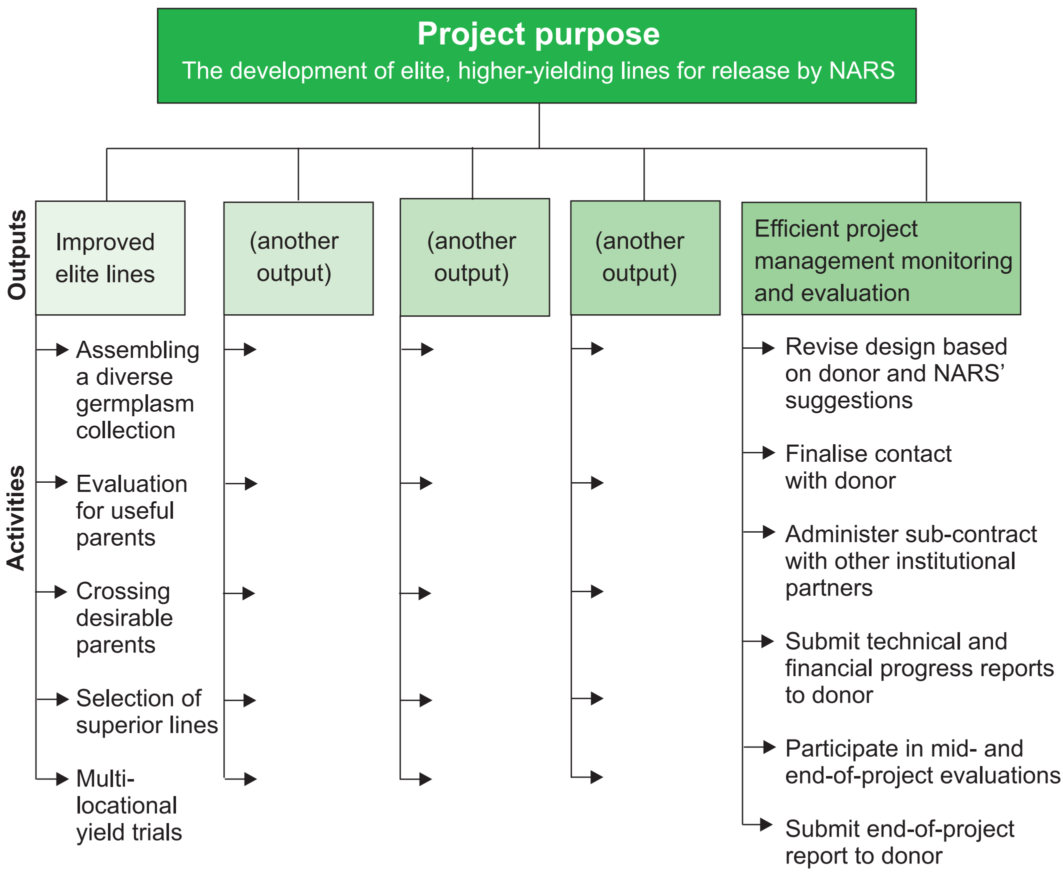


Figure 1. Example of a work breakdown structure (WBS) linking project activities to outputs

Work breakdown structure (WBS)

After completing the LFM, the next task is to prepare the work breakdown structure (WBS).

The WBS is a graphical diagram which groups project activities around specific outputs. The outputs of the WBS (Figure 1) provide the link to the LFM.

Activities identified in the WBS serve as the basis for:

- Project monitoring
- Preparing both technical and financial progress reports to the donor
- Providing the framework for mid-term and/or end-of-project evaluations.

The following brief descriptions of two projects provide insights into how projects can be developed and managed. One project is 10 years old and the other has just completed its second year. They differ in scale, scope, approach, conceptualisation and implementation.

Cowpea Improvement Project, Makerere University, 1993–2003

Starting point (Phase 0)

In 1992, during a needs assessment by the National Agricultural Research Organisation (NARO-Uganda) and Makerere University researchers identified cowpea (*Vigna unguiculata*) as a key legume crop supporting livelihood of about 30% of the Ugandan population (Professor Adipala Ekwamu, personal communication, 20 October 2003). Strikingly, no research was being done on cowpea. A multidisciplinary team was subsequently formed to develop a research agenda for this important crop of the semi-arid regions of Uganda. The starting point was clearly to conduct a baseline study (survey) to:

- Establish the status of the crop (its importance in relation to other crops)
- Gather production statistics
- Identify and categorise the available germplasm
- Identify production constraints (social and biological)
- Establish the competitiveness of this crop vis-a-vis other crops grown by the farmers.

A proposal was submitted to the Rockefeller Foundation's Forum on Agricultural Resource Husbandry Programme, which agreed to support it. This initial phase (Phase 0) established that a research agenda would be formulated with defined deliverables (outputs and indicators) and the steps (based on available information) to be followed in subsequent years.

Phase I Diagnostic survey 1993–95

A national baseline survey helped to set the priorities and define the future research agenda. A multidisciplinary team that included staff from the University, NARO, and the Extension Service further agreed to oversee the research agenda even though the project was housed at the University. Seven MSc students were recruited to implement the baseline survey. A structural questionnaire was developed, pre-tested and used to obtain information from various stakeholder groups. Secondary data was also collected. After data collection and analysis, additional field visits were made. The biological science students' field visits were made over a period of two seasons to verify and quantify biological information provided by surveyed farmers. The outcome of this phase helped to identify the most pressing problems faced by the farmers, and the researchable issues, and to establish benchmarks for future research for development (R&D) for cowpea.

Phase II On-station trials and socio-economic studies 1995–97

The key issue that emerged in Phase I was poor yields (<200 kg/ha). These were mainly due to pests, and to a lesser degree to diseases, weeds, a narrow germplasm base (7 cultivars/

varieties), and poor agronomic practices. The team also considered that more-detailed analysis of market competitiveness and the underlying marketing issues was needed. Five MSc students were recruited for this phase. On-station trials were conducted within the University and at NARO field stations in the cowpea-growing agroecological zone. This phase marked the beginning of efforts to develop technologies to address some of the key constraints identified in Phase I.

Phase III On-farm verification trials 1998–2000

The promising technologies identified in Phase II were tested in farmers' fields but were managed by the research students. Efforts were made to integrate some of the technologies, principally, to develop integrated pest management (IPM) technologies, since pests had been identified as key constraints. Various options such as time of planting, host resistance, chemical control, spacing, weeding frequency and market preferences were studied. These trials resulted in a basket of options that farmers could potentially use to improve the productivity of their cowpeas. They targeted different farmer groups and market demands, and involved four MSc students.

Phase IV Dissemination of IPM options, 2000–02

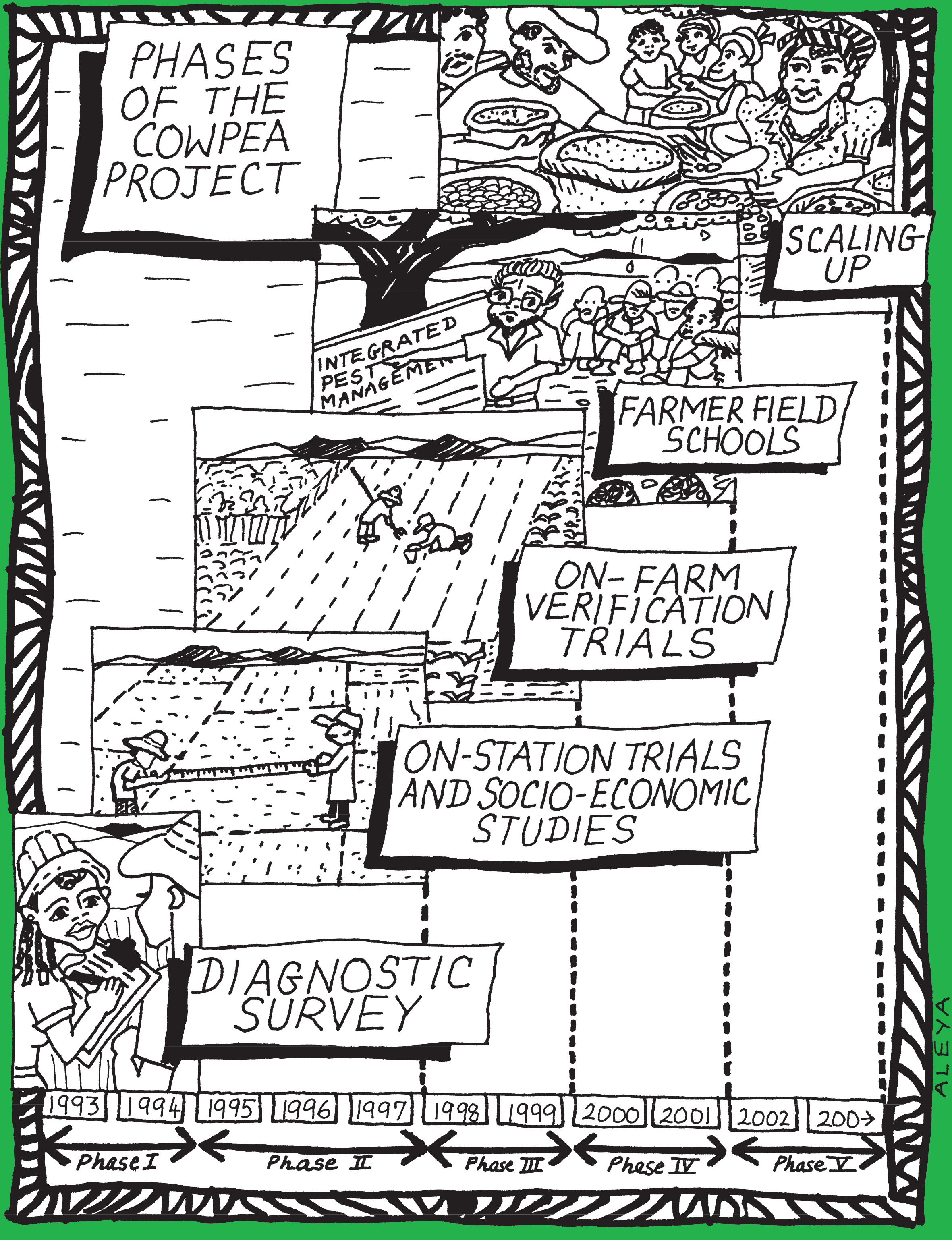
This phase involved refinement and dissemination of the IPM options developed in Phase III. A farmer field school approach was used involving 6 farmer groups (each group consisting of 20–30 farmers). The farmers conducted season-long evaluations of the different options. Pre-tests were conducted to establish baseline knowledge and background of the participating farmers and at the end of the second year, post-tests were done to establish the acquisition of knowledge and changes in practices. Based on this phase, strategies for scaling up were developed in partnership with farmer groups and district extension officials. Three students were involved in this stage, during which other partners, such as the Integrated Pest Management / Collaborative Research Support Program of the United States Agency for International Development (USAID) (IPM/CRSP) were incorporated into the research team and helped to co-fund the study. After 10 years of study three improved varieties were finally identified and recommended to the Ugandan National Variety Release Committee.

Phase V Scaling up dissemination of IPM options, 2003

In order to reach more farmers in all regions the Makerere University Cowpea Improvement Project has now joined up with a number of partners to support dissemination and adoption. Different district extension departments are now engaged in dissemination of such technologies and programmes as I@mak.com (Innovations at Makerere Committee), IPM/CRSP and the National Agricultural Research Organisation–Department for International Development) (UK) (NARO–DFID) are now supporting the scaling up activities. The farmer field schools have been expanded to cover six districts based on earlier experience. Three graduate students are currently involved in scaling-up. A proposal for further scaling up has been prepared for funding.

Summary

10 years of phased studies led to the development of technologies that are now being adopted nationally. It has also strengthened linkages and partnerships and produced graduate students that are highly practical and more responsive to society's needs. Throughout the phases a **total of 22 graduate students** have been trained in various disciplines (see Table 2)



PHASES OF THE COWPEA PROJECT
SCALING UP
INTEGRATED PEST MANAGEMENT
FARMER FIELD SCHOOLS
ON-FARM VERIFICATION TRIALS
ON-STATION TRIALS AND SOCIO-ECONOMIC STUDIES
DIAGNOSTIC SURVEY
1993
1994
1995
1996
1997
1998
1999
2000
2001
2002
200→
Phase I
Phase II
Phase III
Phase IV
Phase V
ALEYA

Table 2. Graduate students trained during the cowpea improvement project

Discipline	Number of students per phase I	II	III	IV	V
Agricultural economics	2	1			
Entomology	1	1	2	1	1
Plant pathology	2	1			
Agronomy	2	1	1	1	1
Weed science		1			
Social sciences			1		
Extension				1	1
Total students	7	5	4	3	3

and the project has to date produced **46 scientific publications** (30 journal articles and 16 proceedings publications). Extension brochures have also been produced.

Consortium for scaling up options for increased farm productivity in Western Kenya (COSOFAP)

Why was a consortium needed?

COSOFAP (2003) reported that: 'Because there was an identified need – a glaring one – to reach more people, more quickly and more lastingly'. COSOFAP also highlighted the following farmer from Soso village who said: "If you work with 3 or 4 farmers in our village, we do not see anything tangible, the whole village needs to be involved if we are to prosper".

Many development workers in Western Kenya felt that a joint effort would be the best way forward, and thus COSOFAP was born out of a stakeholders meeting held in 2000. It currently has over 70 members and more are expected to join. The Consortium is made up of the Ministry of Agriculture and Rural Development Extension Branch, research institutions both local [Kenya Agricultural Research Institute (KARI), Kenya Forestry Research Institute (KEFRI)] and international [World Agroforestry Centre (ICRAF), Tropical Soil Biology and Fertility (TSBF)], national and local non-governmental organisations (NGOs), community-based organisations (CBOs), the Swedish-supported Regional Land Management Unit (RELMA), private companies, and most important of all – farmers' representatives.

COSOFAP's vision is to alleviate poverty among the resource-poor farmers in Western Kenya; its mission is to increase productivity through increased access to technologies through better delivery mechanisms and its purpose is to do this via better networking among the development partners. Figure 3a shows the structure of COSOFAP and 3b its interactive learning sites.

Phase I

The first thing that the Consortium did was to set up a temporary functional secretariat. Next it established management structures and developed governance rules that could function and apply at the regional and sub-regional levels. The region was divided into three sub-regions and three sub-regional coordinating committees were formed to efficiently manage the whole region. The secretariat also prepared a directory of partners, identified learning sites and prepared training and extension notes. A website was developed. Other secretariat duties included proposal preparation for resource mobilisation.

Together with the farmers many activities were identified for support, these included:

a.

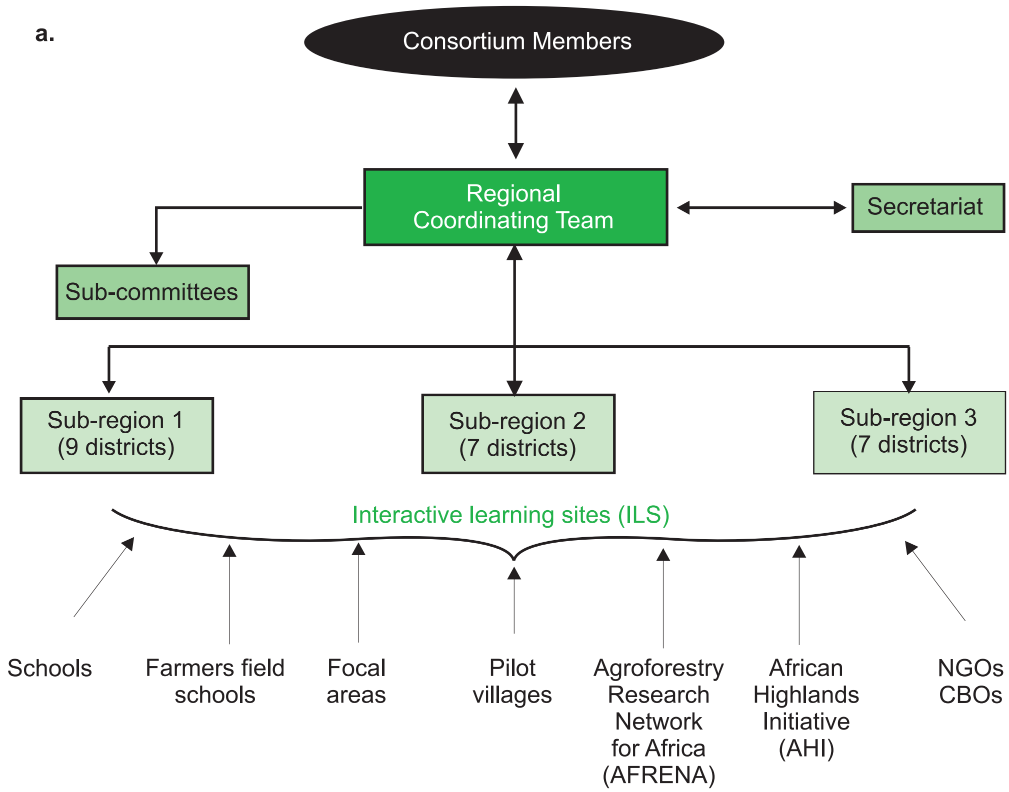


Figure 3a. Structure of the Western Kenya Consortium of Partners (COSOFAP)

b.

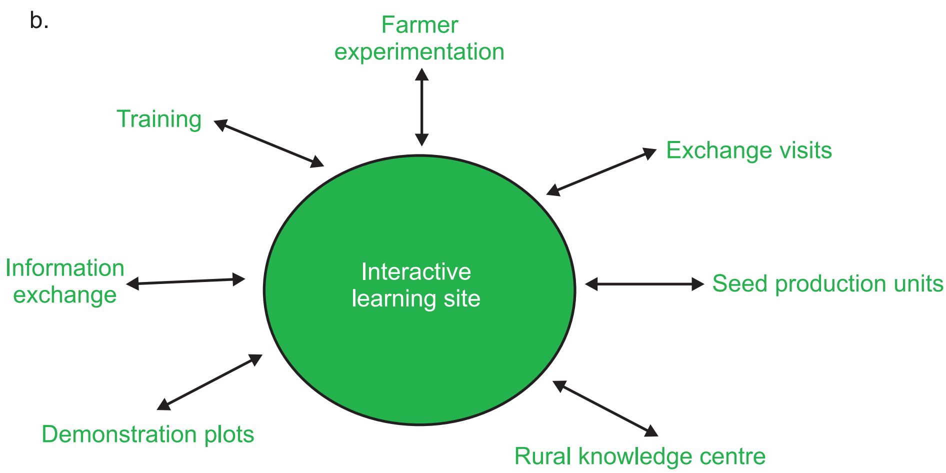


Figure 3b. Interactive learning sites of COSOFAP partners

- Training in soil fertility options
- Grafting of trees
- Poultry and beekeeping
- Exchange visits, field days and training at interactive learning sites (ILS), germplasm distribution.

At the end of the first year what lessons had been learned?

The Consortium learned that much can be achieved in a fairly short time if there is buy-in from all the players at an early stage. That input and commitment from decision makers is essential. That partnership appraisal is also needed to identify the critical and willing partners. That it is better to start small and then grow because there is a need to scale up both options and processes. That it pays to involve all stakeholders from the conceptual stage to the action phase and that it is necessary to keep everyone well informed. In order to do this more meetings are required during the start-up phase of a project. That it is best to develop monitoring and evaluation procedures together with the farmers as soon as possible, and that these should be immediately instituted. All inputs should be in place in time and in adequate quantities.

The Consortium also learned that the expectations were high and varied – some stakeholders viewed COSOFAP as a donor! Management and coordination of partnerships require special attention such that no one can exert undue influence. COSOFAP found that the administrative load was very heavy for a part-time officer and that fundraising is not easy! Grassroot organisations tended to make direct demands rather than following the designated channels. The need to identify recommendation domains became evident as did the need for decision-support tools.

Insights and learning from Cowpea and COSOFAP Projects

The two projects were selected as models because of their differences. But similarities abound as well, see Table 3.

The University-based cowpea project provided many additional learning opportunities for students:

- The area of research was defined prior to selection of students to secure funding, so they had to fit into an identified project and developed their sub-research projects around a given topic
- Students were involved in real problem-solving research
- Different disciplines worked together sharing experiences and knowledge during the whole crop chain from production constraints to markets
- Students were exposed to several participatory approaches
- Students were supervised by specialists in their field of research, and were required to seek the help of the Biometrician in the design of their experiments to ensure continuity
- Successive batches of students followed a progression along the Research Continuum
- Students interacted with scientists from NARS, the extension department, and farmers
- After 2 years of research students published papers in reputable journals, attended meetings, and presented their research data via posters or presentations
- Students also learned how to arrange and manage meetings.

Table 3. Comparison of two projects to highlight similarities and differences

Cowpea	COSOFAP
Goal	**Goal**
Improve the livelihood of the cowpea-growing farmers	Reduce the poverty level of the resource-poor farmers in Western Kenya
Purpose	**Purpose**
Through research develop technologies for increased productivity of the crop approach	Reach more farmers through a co-ordinated effort via the Consortium
Outputs	**Outputs**
Initially graduate training and then farmers in the scaling up phase through a farmer field school	Farmer training from the start of the project using the interactive learning sites
Development of a database	Development of management and operational structure
On-station and on-farm research	Setting up of interactive learning sites
Publications and posters	Training pamphlets
Multidisciplinary team work – 3 public institutions – different faculty members and students working together setting research agenda	Multidisciplinary teams working together on jointly identified priorities and setting work agenda
Joint production of reports	Joint report prepared by regional teams and secretariat
Farmer involvement	Farmer empowerment
Papers presented at scientific meetings	Pamphlets and other training materials prepared for farmers
Continuity and networking	Improved networking in the whole Province
Transparency in process management and resource use	Transparency in management, resource acquisition and use
Strengthened scientific collaboration and interaction	Strengthened the community
Participatory management	Participatory management
Inputs	**Inputs**
Funding resources available from donor in blocks of 2 years because tied to MSc training	Some resources available; funding to be sought through proposals
Students available for training – supervisory expertise available	Expertise available for most identified activities
Financial management – applications for funds by one faculty member; budget development involved everyone including students. A good example of transparency and learning by doing.	Finances managed by sub-regional teams and secretariat
Clear division of duties – who does what and when	Clear definition of roles and duties

How to survive your graduate years and beyond.........

It is always useful to find out as much as possible about the institution, faculty, department and the supervisor prior to embarking on your graduate work. If you are not careful you could have a miserable time and not achieve your academic or intellectual goals. Students within the faculty/ department and ex-graduate students can be a good source of information that could help you to avoid or navigate the hurdles that might crop up along the way. Remember:

- Everyone has a reputation and their own way of working in an institution
- Each institution has its own culture and networks
- Work ethics vary.

Things do not always go as smoothly as you wish. So what do you do when problems crop up? You may have a falling out with your main supervisor – he or she may not provide the guidance and supervision you need; funding problems frequently crop up; you may not feel so motivated as when you started...........

Honesty is the best policy – sometimes it is not possible to apply this rule fully when you are a student, but it is possible to find a solution through other means. Try to discuss the problem or issue with the concerned person. If you do not get any positive results then you have to look for an alternative approach. You could approach the Dean of the faculty or Department Chair if the problem persists, or most faculties usually have a student representative who can take up an issue with management, and who could try to resolve things within the department.

The point is that you are there to get a degree and training, and it is the duty of the institution to provide the best possible education but you also have responsibilities and duties. You do not have to put up with an inadequate supervisor or a poor research project, but you must weigh up the costs before taking a decision or stand.

Always try to keep the communication channels open with your main supervisor and your research committee whose members you can also approach for help. Keep them well informed, meet them as a team as often as possible so that when the project appears to be off track, you can get a consensus opinion and devise a joint way forward. If you see people individually you might get diverse opinions and advise, and end up more confused than ever.

Be open, ask questions and seek clarification when you do not understand a view point or instruction... and Good Luck!

Resource material and references

Appendix 2. Innovation, Problem Solving and Operational Research Strategies. Paul L. Woomer. PowerPoint on CD.

Appendix 3. Designing Research Around Client Needs. Paul L. Woomer. PowerPoint on CD.

Conway, G. 1997. *The Doubly Green Revolution.* Penguin Classics, UK. 335 pp.

Eicher, C.K. 2003. Agriculture in the global economy. In: *Hunger: 13th Annual Report of the State of World Hunger.* Prepared for the World Resources Institute (WRI), Washington DC, USA. 164 pp.

Kaner, S., Lind, L., Toldi, C., Fisk, S. and Berger, D. 1998. *Facilitators Guide to Participatory Decision Making.* New Society Publishers, Canada.

Lacy, W.B. 1996. Research, extension and user partnerships: models for collaboration and strategies for change. *Agriculture and Human Values* 13: 33–41.

Milton, S. 2003. Systems Thinking–Putting Agricultural Programmes in an Ecological Context. Curriculum Development and Transformation in Rural Development and Natural Resource Management. FORUM *Working Document Number* 7, pp. 58–65.

Quereish, N. and Njui, A. 2003. *Consortium for Scaling-up Options for Increased Farm Productivity in Western Kenya* (COSOFAP), First Year Project Report. Prepared for The Rockefeller Foundation, New York, USA.

Senge, P.M. 1994. *The Fifth Discipline: The Art and Practice of the Learning Organisation.* Doubleday Dell Publishing Group, New York, USA. 423 pp.

Part 3
How to approach academic research

This part of the book is designed to help you to get started, select your topic and develop and write up your thesis in a way that meets academic requirements. All academic research is primarily directed toward a better understanding of our world. As a graduate student your research is also required to give you practical experience in doing research and to test your ability as a future academician. Graduate research work, particularly in Africa, can make a significant contribution to producers, entrepreneurs, communities and/or policy-makers and development agencies.

Your objective is to obtain your qualification. If your research is funded outside the university, then the sponsors will require that your research contributes to their project objectives. If your work is applied, then those with whom you will be researching or working will expect your research to contribute to their own productivity or development agenda. This can place you, the student, in a difficult situation. The requirements of the project are likely to be made explicit to you. You need to be aware of the requirements for your academic thesis. You can then take steps to balance this with what you are asked to do for your project. The most important thing is to be very clear with the funding agency what you are required to deliver for them. This way they cannot keep changing the goal posts. It also means that you can be sure to adapt your research from the start so that it is able both to achieve the required funding objectives and your personal objective of obtaining a higher degree. The first three chapters of this part provide guidelines on your objective of obtaining a qualification and guidance on initiating, developing and writing an academic thesis. Chapter 3.4 complements the research proposal chapter with specific recommendations for a research grant proposal.

Kay Muir-Leresche

Making it HAPPEN
SILENCE
SEEK & THOU SHAL FIND
CLICK!
ALEYA

3.1 The first steps – literature reviews and references

Kay Muir-Leresche

- **What research would fascinate you and what are you good at?**
- **Consider the problems that need to be solved – use your everyday experience to help you**
- **Look for projects working in this area which you could join**
- **Carry out a preliminary literature scan and Internet search**
- **Narrow down the broad topic to the specific area you would like to research**

'Every man is a borrower and a mimic … and literature a quotation.'

Ralph Waldo Emerson (1860)
Society and Solitude

The first steps

1. Carefully assess your strengths and weaknesses. Consider first in which subjects you are strongest and which most fascinate you. Remember that you will have to maintain your interest and enthusiasm for the topic for a long time – often many years. Enjoying and having a good grasp of the subject area you work in is important. Then consider whether you would be better working with abstract concepts emphasising theory and mathematical models (basic research), or if you would be better working more closely with people in a practical and development, or policy-oriented way (applied research). You do not need to decide anything at this stage. You just need to have a realistic view of what you can do and where your interests lie. You will be most effective if you are able to be passionate about your research so it is very important to chose interesting topics and an approach which suits your personality and abilities.
2. When you have thought of your strengths and interests, you need to think about the world in which you live. In your own opinion, and based on your experiences, what do you think needs to be researched? Ask yourself such questions as:
   - Is there a particular soil type that appears to be uneconomic for a particular crop or in a certain area?
   - Are farmers' decisions relating to resource use leading to degradation?
   - When you were studying biochemistry, the only examples used were analysing samples from another continent – would it be interesting to run similar tests on local soil?
   - Does goat manure appear to be more economic than cow manure in improving the productivity of soils for growing maize?
   - Are the currently used physical measurements of soil structure relevant in a particular situation?

   These sample questions are from different disciplines but are all directed at considering the productivity of soil in one particular area. It is you who decides what the most interesting research needs might be in the particular topic and area you are considering. A few students will focus on basic research, but most African graduate students will be primarily focused on addressing their research to real-world problems. Your own life experience can be an important contributor to your understanding of issues and problems:

- What problems do your parents face on the farm?
- What products or services are not meeting your needs as a consumer?
- Which government policies are making it difficult for your uncle to establish a business or for your cousin to market the commodities she grows?

These and other questions could all be inspiration for the starting point of your research. Your observations of the world – even simply that some farmers get better results than others – can all lead you to problems to research as you consider why. You will then use theory as a central component to solve the problems, even in applied research.

Your starting point could even come from a news report or from a crisis the country is going through, or from a change in government policy. Reading literature, journals, Internet articles, books and even your studies will provide most of you with ideas for applied research. A literature search and your own laboratory experience are often the main source of inspiration if you are more interested in basic or theoretical research.

3. When you have some idea of your field or fields of interest, approach your university lecturers and find out if there is an on-going project in this area which you may be able to join when you carry out your research. It is useful to indicate your interest early.

Do not make an approach to them, even informally until you have briefly looked at some literature so that you are able to talk with some confidence on the topic when they question your interest in the research.

Look outside the university as well. Research goes on in many different organisations – national, regional, international and even some commercial. Many have schemes for attaching students from universities, and often have better-funded projects that can help support students.

If you are prepared to finance your own research then you will have more control over what you research and the approach you use and you can go directly to a more-detailed literature search. If you have already been recruited to a project, then your choice of field and approach to be used may be prescribed for you, but the emphasis you place on different aspects will still be up to you.

4. Once you have some idea of the topics you will investigate, carry out a preliminary search of the literature. If you have access to the Internet that is a good place to start. If your university does not provide good Internet access then you might have to invest a bit of your own money in using a public Internet access point ('cyber café'). If you do not have Internet access, then go to your libraries and search by subject matter. There are some journals which provide abstracts of recent publications in particular fields. Also go to relevant journals and glance through the article titles, occasionally stopping to look at abstracts. Many African agriculture faculties and libraries have access to CD-ROM literature databases, like the Essential Electronic Agriculture Library (TEEAL) which is a full text collection of core journals in agriculture and related fields. TEEAL is compiled by Cornell University. It is most useful, particularly as you can scan and print the abstracts, and only print a few of the most relevant articles. Given the high cost of printing, it is not advisable to print much when you carry out your initial exploration of the literature.
5. After you have looked at some of the literature relating to the topic of your interest, you should be in a position to begin considering exactly on which area you are going to focus

and to start to narrow this area to a researchable problem. Now go and visit lecturers, research funders, stakeholders and other persons involved in the work you are interested in pursuing. Discuss with them your ideas and what you hope to achieve. You can then begin to put together your research proposal. This will require you to consult the literature in more depth and to start looking at the theoretical literature and the data collection and analytical methods that may be appropriate for what you want to achieve.

Research is an on-going process and your thesis will represent one cycle. It will not be the end of research into the topic. One of the purposes of academic enquiry is to develop further questions for research. (Figure 1).

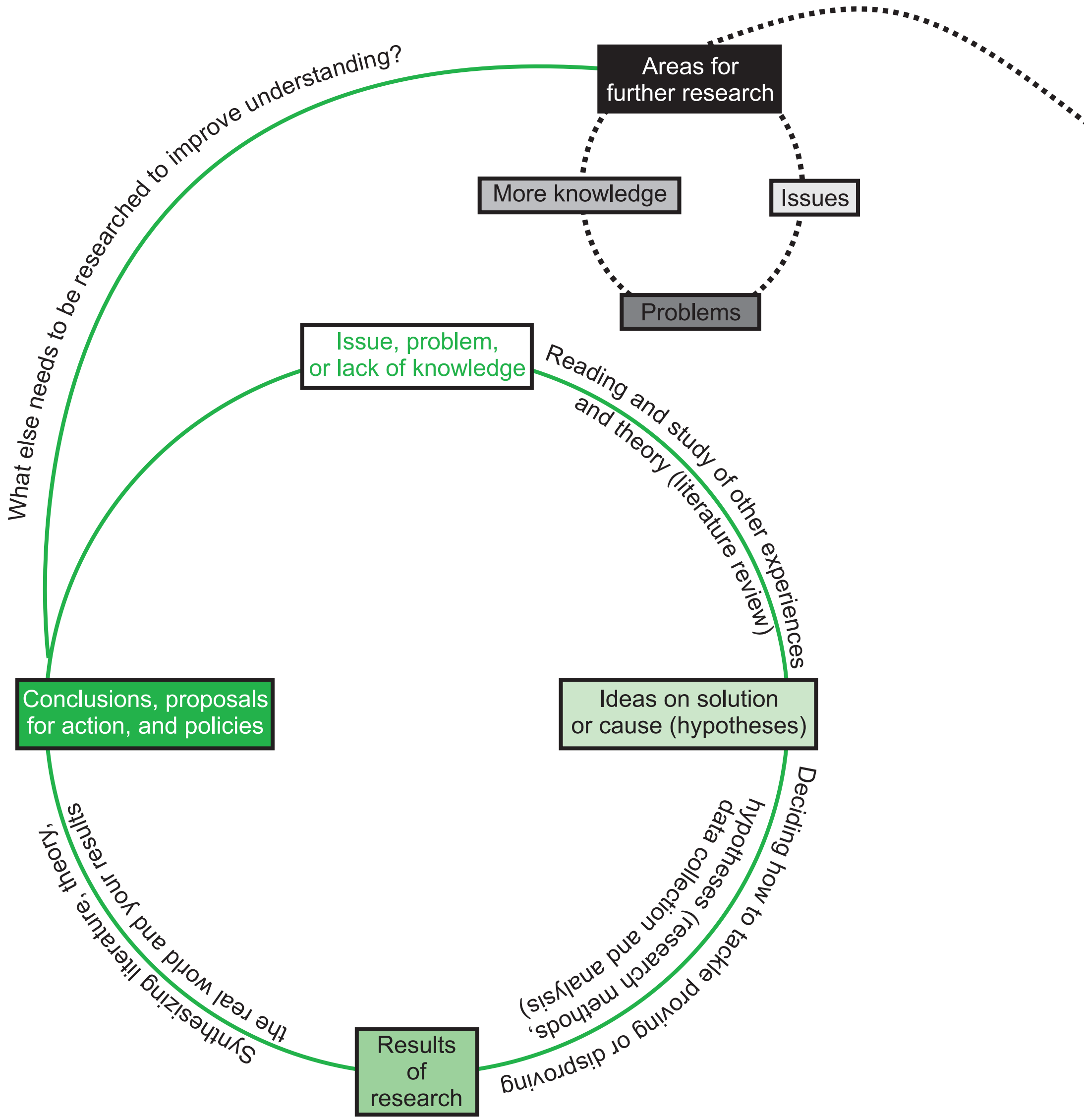


Figure 1. The cycle of research

The purpose of research is to increase knowledge. The scientific process starts with a problem that needs to be addressed. After reading and consultations, you should be able to come up with some ideas of cause and effect and to present the issues to be researched as one, or a series of hypotheses. You will collect data in the form of experiments, or surveys, or literature and secondary data sources. This data will then be analysed and interpreted and, where appropriate, presented as recommendations and areas for further research. Your conclusions will be based both on analysis of the data on theoretical principles and the literature review. With the new questions posed by your research, the cycle will begin again.

Your literature review

- Identify earlier work in your field – avoid reinventing the wheel
- Use the literature to show how your research will contribute to greater understanding of the subject
- Synthesise and summarise the earlier studies in a way that links together and highlights and resolves conflicts between different approaches
- Be sure that your review is critical and keeps showing the reader its relevance to your study.

It is a tragic waste of our precious human resources to continually re-invent the wheel.

Literature review

Purpose

Reviewing literature takes up considerable time for all academics – even those involved in the most theoretical, mathematical or practical research projects. It is particularly important to those addressing descriptive or policy issues. One of the main reasons that reviewing literature is so important is to avoid your spending long hours investigating something that has already been investigated! Re-investigation is only appropriate if you consider the previous research process inadequate, or if a different research method or new technology could produce different results.

For both masters and doctoral theses you will need to demonstrate your understanding of and competence in the research process, research techniques, interpretation and presentation. Your work must contribute to the body of knowledge but only doctoral theses have to demonstrate that the research constitutes an original investigation or testing of ideas or, if it is predominantly descriptive research, that it is in an unexplored area.

In order to demonstrate this you will need to provide a review of the literature to show you are not re-inventing the wheel. You will also require the literature to describe the problem you are researching and to explain why you consider it worthy of research. You will need a review of the theoretical literature in order to provide for the logic behind your hypotheses and for the interpretation of your results. You will need to use literature as evidence to support many of your statements, both theoretical and empirical, during the course of writing your entire thesis.

A thorough understanding of the literature is an essential component of graduate studies.

All dissertations, including honours theses, must demonstrate your ability to make critical use of literature and to provide a clear link between the literature reviewed and the research being undertaken. You must not review conflicting texts without highlighting the conflicts and interpreting them in relation to your own thesis.

Your literature review must avoid simply summarising available texts.

All graduate students need to ground their research in an explicit theoretical or conceptual framework, even when it is based on problem-solving or policy-oriented applied research. You will need to include a specific section related to those theoretical principles which underlie your assumptions and which inform your analysis and conclusions.

For your review of theoretical literature, it will be useful to scan old notes from relevant courses and then do the additional readings you may have skipped at the time.

Doing the review

In the first place you will need to make a preliminary review of the literature to provide you with the inspiration you need when developing your research topic and your research proposal. Then you will need a thorough review of the literature on the problem you are addressing. This will normally, but not always, be presented in a literature review chapter of your thesis.

The literature review chapter presents the views of other authors on the problem and issues your research addresses. Your opinions should be kept out of this chapter except where you critique what a particular author is saying. You are also required to show the reader how the literature you are reviewing relates to your research. This must be a well thought-out review of the literature. You use it like building blocks to show what other research has been carried out in the field and to review the theoretical principles which form the basis for the hypotheses you are postulating and/or any principles you may actually be testing in your research. You need to show:

- What other studies have been done, what their findings are and how they relate to what you propose to research. You need to report the results of others' findings and provide a critical review of these and of the methods which were used
- What theory has to say about the relationships you are investigating in the problem you are researching
- That you are familiar with the literature (not that you have seen it all!) and that you are able to understand and interpret the work you are reviewing. You need to make a particular effort to access current literature.

Links between each reviewed author and the research problem, and those between the reviewed authors must be made explicit.

It may be appropriate to include an analytical model which summarises the theoretical framework.

Many students are afraid of models. Don't be! A model is just an abstraction of reality. We build models to make it easier for us to deal with a very complicated world. A model can be descriptive – when you describe the real world, but do not include everything; schematic – when you produce an illustration which simplifies the relationships; or you can develop mathematical models. Mathematical models are the most useful when you need to make predictions, whereas descriptive and schematic models are useful when trying to understand the world better. Mathematical models are also very useful in establishing which variables are important in a system and how they are linked. See **Chapter 4.8** for a detailed discussion of models.

The literature review provides the validation of your research. It also provides legitimacy to your approach to data collection, analytical techniques and to your interpretation. It should be an interesting synthesis of available literature which will make it clear to your examiners why you are undertaking this research and why you have chosen a particular approach. It will also lead them towards an understanding of your interpretation of the results and of your recommendations to producers, or policy-makers. The literature review will also inform them about areas for further research which are such an important component in graduate research. Most importantly it shows readers that you are not repeating earlier work.

Two common problems to avoid

1. Taking a literature review from another paper and giving the impression that you have yourself looked at the articles referenced
2. Plagiarism through copying large sections from other authors or published abstracts.

What is literature?

Literature is really just information that has been presented by other individuals or institutions. It includes text books, journal articles, graduate theses, working papers, occasional papers and briefings, Internet articles and data sets. It also includes unpublished reports, documents in files, pamphlets, draft research and notes. Published literature is that which it is possible for other readers to access. 'Grey' literature is that which is only available directly from the author or institution who produced it and is not available for public orders. Text books from reputable publishing houses and journal articles will normally have more credibility than unpublished reports, or pamphlets. This is because the published work will have been reviewed by peers and the data collection and analytical techniques will have been open to scrutiny. Unpublished work may be based on poor data or analysis or may even be deliberately biased to a particular outcome. When you are using literature to provide evidence for some of your assertions you need to keep this in mind.

Don't be fooled – not all published journal articles are based on sound data or evidence. If the results or conclusions are not logical, carefully question the author's assumptions, data and/or analytical methods.

Be warned!

In my own experience, I published data from very dubious records by District Administrators on estimated crop areas and yields in communal farming areas. I placed caveats everywhere warning people not to use the data which was only included as a matter of historical record since there was no other data available between 1890 and 1960. The supply response studies I did using the data were not included in my thesis because of the unreliability of the data. Some 10 years later I saw an article in the American Journal of Agricultural Economics, estimating supply response using my data with no acknowledgement of the problems that existed with that data!

You need to be aware that even the most respected publishers and journals may publish material that is false – the review process is not infallible. You must always look at what is presented critically and assess the source of secondary data and the method of primary data collection carefully. For the most part you can assume that information from published sources is reliable – but keep a look out for errors. You need to be critical.

You need to be particularly critical of information coming from vested interests – farmers associations, government departments, or private companies. It is often very good information but you need to assess it for bias. For the social sciences grey literature is widely used, but you have to be constantly aware of the need to assess its validity.

Literature is central to providing evidence for your choice of topic, your hypotheses, and the interpretation of your results. It also informs the methods you use to collect and analyse data. An effective review of literature provides your examiners with evidence that you understand your topic. Do not overlook its importance.

How to go about a literature search

Where to look

1. Look at the materials that have been recommended to you in reading lists for your courses. These are usually the easiest to obtain. These articles and books will in turn include references to which you can go for more depth. If your library does not have the text you need to access, you can ask your librarian to access it from a larger library. This was common practice until the 1980s when, in most of Africa, paper and postage became too expensive and so for several decades the inter-library loan facilities were not used very much. However, now with Internet and scanning facilities it has become easier for your librarian to arrange for you to access the material you need by using electronic transfers.
2. If you have Internet access this is probably the most useful way to get access to literature and some suggestions on how to access the information you need is included at the end of this chapter.
3. Use the subject index in your library to go to the shelves and see what is easily available in your area. Increasingly our university libraries have very little modern information, but quite often the old information they do have can be most informative. Many of the studies that were carried out in Europe, Japan, India and North America in the 1950s–1970s can be very relevant to what you want to do. The experiments and analyses reported were relevant to their particular natural and social conditions, but there are others which can usefully provide guidelines for undertaking similar investigations in Africa.
4. Go to the journal section in your library and scan the article titles to see which ones may be relevant to your work.
5. Use some of the journals and CD systems available which provide abstracts (and sometimes the entire article) of recent research. A list of some of these sources is provided at the end of **Chapter 3.1**.
6. Go to government offices, farmers' organisations, NGOs, research organisations and the information libraries of various international agencies and embassy services. Some government departments have a surprisingly good collection of material that is not available elsewhere. Note that in some African countries national and international agricultural research organizations have been in existence for a very long time [the Kenya Agricultural Research Institute (KARI) in Kenya are celebrating their centenary]. Of course some information they hold becomes out-of-date but much is still relevant.

How to look

1. You should learn how to scan very quickly through books, articles and papers in such a way that you are able to see if they are likely to be useful. You need to look at the abstract first and if it does not appear to be relevant put it aside. This should take you no more

than a few minutes for each book or article. If you are looking through a publication of abstracts then use the Table of Contents to find the relevant sections and then glance down the page and only stop to read the abstract if the title seems relevant. You may miss some relevant material in this way but you have to learn to make a trade-off between time available and the depth of your hunt for literature.

2. If it appears that the publication may be useful then turn to the Table of Contents and if that also looks useful then make a note of the author and title.
3. Next, turn to the conclusion and scan that. If it still seems relevant make a few very brief notes from the abstract and conclusion and be sure to fully complete the reference note with publisher and date and place of publication. Then put it aside to check out (if it is in a library) so that you can read and make notes at leisure. If it cannot be taken home, then make a quick review of the rest of the document, paying particular attention to the results and make a note of the data collection method and the analytical techniques. If there are any really interesting assumptions or conclusions, make a note of these and perhaps even copy out a quote. This should not take you more than 15 minutes per book or article. If it seems that the publication is essential to your research then see if you can persuade the holder to let you borrow it, or at least to photocopy some pages or arrange to come back when you have more time.

Get a hardback notebook and when you go somewhere head the page with the name of library/office and the date. Use the book for your references and your notes. One of the most difficult things is retracing where you found the information you are using. Be sure to enter these regularly in your computer in case you lose your book! Then back up the computer as well.

Constructing your review chapter

The type of research you are undertaking, and your discipline, will affect how your literature review is constructed. In some cases the literature review may form an introductory section to several chapters and not stand alone. In others you may actually have two full literature review chapters, one related to description and history and another related to theory and research methods. Whatever the form your literature review takes it must at some point include the following aspects:

- An overview of the subject, issue or theory under consideration, along with the objectives of the literature review
- A categorisation of the literature into those items that are in support of a particular position, those against, and those offering alternative theses entirely
- Explanation of how each work is similar to and/or how it varies from the others
- Conclusions as to which pieces are best considered in their argument, are most convincing of their opinions, and make the greatest contribution to the understanding and development of their area of research.

In assessing each piece, consideration should be given to:

- Provenance – What are the author's credentials? Are the author's arguments supported by evidence (primary historical material, case studies, narratives, statistics, recent scientific findings)?
- Objectivity – Is the author's perspective even-handed or prejudicial? Is contrary data considered or is certain pertinent information ignored to prove the author's point?

- Persuasiveness – Which of the author's theses are most/least convincing?
- Value – Are the author's arguments and conclusions convincing? Does the work ultimately contribute in any significant way to an understanding of the subject?
(for more details see Lyons, 2003)

The most important factor in a literature review is that you continually link what you are reviewing to your research. You must show your reader why you are including this work and how it helps to develop your thesis. You need to make these links as explicit as possible.

Helpful hints in writing a review

You need to ensure that your literature review is not just a jumble of different summaries – you need it to tell a story. The best approach in most cases is to separate the literature into different themes:

- Historical and descriptive – or what research has already been done on this topic
- Methodological – how other studies collected data and analysed it. Here is where you look at various data collection and analytical methods including those you are **not** going to use and say why they are not right for your research. In some theses this section can also go before the chapter you write on data collection and/or in some cases before the analysis chapter
- Results, interpretations and controversy – here you discuss the major points these writers were making and how your research fits in. You can assess weaknesses of the earlier research and contribute to the discourse.

Remember

- Use literature as evidence to back up what you are saying
- Be selective
- Summarise the key points made in your own words – don't use too many quotations
- Be careful when paraphrasing – you must represent the opinions and information accurately.

Referencing, citations and bibliographies

It is essential that you always acknowledge the source of what you are saying. **Plagiarism and falsifying research data are totally unacceptable in academic circles.** Either of these crimes will result in your losing employment, funding, and credibility, and lead to expulsion from your degree course or to failure. **Plagiarism is defined as stealing the writings or ideas of another.** It occurs when you are using work that you do not acknowledge, even if you have put those ideas into your own words. You are expected to use other people's ideas and even on occasion their words, but you must always acknowledge that you are using them. There are many different styles of providing the sources of your information, but all systems involve citing the author in the text where you use the ideas (either in brackets or in a footnote). In addition you must provide the full reference to the citation – usually in a reference list at the end but sometimes by providing the information in footnotes throughout the text. The detailed references must provide all the information needed for your reader to access the original source.

Evidence and acknowledgement are the heart of academic research

All written work must include the sources of information you have used. These should be the main sources you have consulted in writing your report or dissertation, i.e., those that you cited in the body of your thesis. Your reference list does not include everything you might have looked at. In some academic writing a full **bibliography** is given at the end of the publication that includes most of the documents consulted, even if they are not cited in the text. More commonly a **reference list** is used and this includes only cited work.

Why do you need to provide citations and references?

- To provide evidence – the most important part of contributing to humanity's store of knowledge is to be sure that the contribution is valid (within the limits of what is known at the time). As an academic you should never make unsupported assertions. Evidence and support come from:
  - empirical facts which you obtain by carrying out experiments or collecting data which 'prove' that this is a reflection of the real world
  - theoretical principles (and in some cases you may need to show how these were derived or why they are valid)
  - literature – what other people have found out. You need to say where you got this information from – hence citations and references
  - allowing your readers to check what was said and to ensure that you have interpreted it correctly
  - allowing your readers to discover more information and read more deeply on the topic.

Providing citations and references is what separates academic research from political rhetoric, or religious dogma, or uninformed opinion. It is one of the most important parts of your research and your literature review and the references you provide may be the most useful contribution of your research to academic enquiry.

A **bibliography** or **reference list** gives all the work you have cited in alphabetical order by author's surname, or institutional (corporate) name, or by title where no name is given and is usually situated at the end of the main body of your work, but before the appendices. There are many different styles of referencing and you can chose to use any system, even a numbers system, provided you are consistent. Check with your university regulations to ensure that the system you use is appropriate. A slightly adapted version of the Harvard system is the one used in this book that follows the style of the African Crop Science Society Journal.

In the text you should include the author or editors surname in brackets after the presentation of her/his ideas or a direct quotation (e.g., Rukuni and Eicher). It is also common to include the date with the surname/s (e.g., Rukuni and Eicher, 1994) but this is only essential when you have more than one publication by the same author/editors. If you have more than one by the same author in the same year, then you also use alphabetical letters (e.g., Rukuni and Eicher, 1994b) to distinguish them.

It is only for quotations (using the other author's words) or where you want to provide your reader with direct access to the source, that you include the page numbers in your citation in the text. When you quote directly from another work you must indicate this with quotation marks and then beneath where it appears in your text, insert the author/editor name and the page number(s) of the original text. This will correspond to the name/s as they appear in the alphabetical sequence of citations in your bibliography or reference list. Generally textural quotations should correspond precisely to the original in wording, spelling, capitalisation and punctuation. With short quotations incorporate passages of not more than four lines directly into your text, identified by quotation marks, e.g., Ojwang (1994) 'The natural cycles of biodiversity are affected where there is habitat destruction,.....monoculture crops or forest plantations.'

With longer quotations (two or more sentences or more than four lines) set the quotation off from your text and indent from the left margin.

> 'We have three principal means: observation of nature, reflection, and experiment. Observation gathers the facts, reflection combines them, and experiment verifies the results of the combination. It is essential that the observation of nature be assiduous, the reflection be profound, and that experimentation be exact. Rarely does one see these abilities in combination. And so, creative geniuses are not common.'
>
> (Diderot, 1753)

The reference list

You should find out if you are required to follow a particular system of referencing from your university and search for a detailed description of that system. When you are publishing journal articles you will need to check and see what system is used by that journal. The elements are the same in all reference lists but they may use a different order and different punctuation. The Royal Agricultural College Library has compiled a comprehensive description of the Harvard System also known as the Author-Year or Name-Date system. It gives very detailed and useful information on how to reference and the website details are included in the reference list to this chapter. Modified extracts of their website information, using African examples are provided below.

Main elements of a reference list – how to prepare a citation

Author

- Surname first, followed by initial(s)
- Where more than one author, list them with 'and' before the last one. (When citing in the text if there are more than three, put first named followed by... *et al*. (meaning 'and the rest'. But all their names must appear in the Reference List at the end).
- If they are editors, put ed. or eds. in brackets after the name(s).
- Corporate authors: where initials are commonly used, e.g., FAO (Food and Agriculture Organization of the United Nations), make sure that somewhere in your paper, (perhaps as a separate appendix) you have a list of acronyms giving their full expansions.

Title

- Must be in *italics*
- Capital letter for the first word only, or second word if the first is an article.
- Take the wording from the title-page of the book (not the cover as it may differ).
- A colon precedes the sub-title. (Always include sub-titles because they often give extra useful information.)

Books: Personal author

Surname, initial(s). (Year) *Title*: *sub-title*. Edition (if more than one). Publisher. Place of Publication

Venkatesan, V. (1994) *Seed Systems in Sub-Saharan Africa: Issues and Options*. World Bank, Washington DC, USA.

Books: Corporate author

Name of Organisation. (Year) *Title*. Publisher. (add Series and number if appropriate)

ISSER (Institute of Statistical, Social and Economic Research). (1994) *The State of the Ghanaian Economy* Legon: University of Ghana, Accra, Ghana.

Chapter in book

Cite the author(s) of the chapter, use the word In: to indicate where that chapter is, then give the title of the book in italics and editor name(s) publishing details and lastly the page numbers in the book.

Ondiege, P. (1996) Land tenure and soil conservation. In: C. Juma and J.B. Ojwang (Eds) *In Land We Trust: Environment, Private Property and Constitutional Change* Initiatives Publishers, Nairobi, Kenya, and Zed Books, London, UK. pp. 117–142.

Conference proceedings

A paper in a collection of conference proceedings is cited in similar fashion:

Boardman, J. (1995) Soil erosion and its impacts. In: *Soil erosion and land use: towards a sustainable policy: proceedings of the seventh Professional Environmental Seminar, February 1995, Cambridge.* B. Evans (Ed). Cambridge Environmental Initiative: 37–53.

Journal articles

Author(s). (Year) Title of article. *Title of journal.* Volume (part number): page nos

Borlaug., N.E. and Dowswell C.R. (1995) Mobilising Science and Technology to get Agriculture moving in Africa. *Development Policy Review* 13(2): 115–129.

Unpublished dissertations or theses

Sakala, W. (1994) Crop Management Interventions in Traditional Maize–Pigeonpea Intercropping Systems in Malawi, unpublished M.Sc. thesis, Bunda College of Agriculture, University of Malawi, Lilongwe, Malawi.

Personal communication

Sukume, C. (2003 Sept 20) Conversation about the importance of tenure to investment in production. University of Zimbabwe, Harare, Zimbabwe. [Note, usually only cited in the text, not in the reference list.]

Internet

Author. (Year) *Title*. [online]. Edition. Publisher. Available from: URL
[Date accessed]

Environment Agency. (2002) *Fisheries action plans* [online]. Environment Agency. Available from: www.environment-agency.gov.uk/subjects/fish [Date accessed: 30.9.02]

It is worth keeping copies of key Internet pages on your computer as they can disappear.

The term **Publisher** is used to cover the organisation responsible for maintaining the site on the Internet (corporate author). If it is not immediately obvious, delete sections of the URL one by one, going higher up the chain of the Internet address to find out who created the website.

Electronic journal articles

Author (Year) Title of article. *Journal title*. [online or CD-ROM] volume (issue), pagination. Available from: URL. [Date accessed]

Horesh, R. (2002 Sept) Better than Kyoto: climate stability bonds. *Economic Affairs*. [online]. 22(3):48-52 Available from: www.iea.org.uk [Date accessed 30.9.02]

E-mail communications

Discussion lists generate e-mail messages which are sent directly to the subscriber. References to these messages should be treated in a similar fashion to journal references. Use the list name in place of the journal title and the subject line of the message in place of the article title. After 'Available from', use the e-mail address of the list administrator or URL of the archive.

Author (Year, month, day). Subject of message. *Discussion list* [online]. Available from: list e-mail address [Date accessed]

Ladley, L. (2001 Feb 27) Customer feedback: summary of responses. *Lis-link* [online]. Available from: www.jiscmail.ac.uk. [Date accessed: 30.9.02]

Searching the Internet

The purpose of this section, that was compiled by ***Liliosa Maveneka***, is to provide you with a simplified overview on how to use the Internet to obtain information and to provide a guide to some of the agriculture-related databases which you can access through the Web.

The Web (or WWW) is a collection of documents, resources, Internet sources or websites that can be accessed on the Internet via a Webbrowser. Microsoft Internet Explorer is a popular multimedia browser. **Search engines** are the primary tools that are generally used to search for information from the Internet.

Search engines

Search engines are of two main types. Some have robotic software that crawl all over the Internet looking for webpages and indexing them in their databases. They index any type of webpage.

Other search engines have human indexers who index web pages. These are generally better than robotic search engines because their websites are checked for their quality. Their search programmes give the entries found in order of relevance.

General search engines:

Google	http://www.google.com/
Alltheweb	http://www.alltheweb.com/
AltaVista	http://www.altavista.com/
HotBot	http://www.hotbot.com/
WiseNut	http://www.wisenut.com/

Meta-search engines:

Dogpile	http://www.dogpile.com/
Iboogie	http://www.iboogie.tv/
Metacrawler	http://www.metacrawler.com/
Profusion	http://www.profusion.com/
Infogrid	http://www.infogrid.com/

Meta-search engines search different engines at the same time and then present results to the user

Some of the more serious academic research engines are, AltaVista and ENCARTA.

Directories, also known as subject directories, group websites by subject and make it easier to select appropriate information. Directories are indexed by subject specialists and therefore have superior quality websites. However search engines are more up-to-date than subject directories.

Examples of directories are:

Yahoo	http://www.yahoo.com/
Pinakes	http://www.pinakes.com/

Searching techniques

Most people have problems because of the overwhelming amount of information available on the Internet. After you have determined a suitable search engine, directory or database, you will find the following strategies useful.

1. Define the key information you are looking for and come up with key terms. You may also use the 'help' function found on the engine.
2. If you have multiple key terms use Boolean operators to specify the search action and limit the list of hits to an acceptable and relevant number. The **Boolean operators AND, OR** and **NOT** specify relations between search terms:
   AND to indicate certain terms to appear on a page simultaneously,
   NOT to indicate that certain terms should certainly not occur on a page,
   OR expands the search to contain either of the words.
3. Use phrase searching to get highly relevant results. Phrases help to refine the search.
4. If you are unsuccessful in your first search, use alternative search terms.
5. Try to formulate as many search terms as possible.
6. Try a different search engine or database.

Searching requires time but is not stressful.

Example: Searching for literature on writing a research proposal

- Double click on the **Internet Explorer** icon
- Click **Search**. A search engine/ directory comes up
- Click **Find a Web page**. Write the word **Google** and click on **search**.
- One of the sites in the search results is **Google**. Clicking this makes the google engine active
- Enter the word **AltaVista** and click on **google search**
- Click on the search result **AltaVista** and this search engine becomes active
- Enter words **writing research proposals** and click **Find**
- AltaVista gives the option to focus the search term. Focus on **proposal writing**
- Looking through results one of the results (as at 26/08/2003)
  Reid A, Proposal Writing – A Practical Writing Guide. The website is given as
  http://members.dca.net/areid/proposal.htm

Useful websites/databases on agriculture and related sciences

- AGRIFOR – The UK's gateway to high-quality Internet resources in agriculture, food and forestry
  http://agrifor.ac.uk/

- African Journals Online AJOL – for African-published journals in the sciences
  http://www.inasp.info/ajol/
- EBSCO – provides access to academic search in social sciences and other areas as well as access to economic journals (Library to register – updated daily via EBSCO host)
- Institute of Development Studies, Sussex, UK
  http://www.ids.ac.uk/
- Consultative Group on International Agricultural Research (Provides information on all the 16 centres in the Group)
  http://www.cgiar.org
- International Network for the Availability of Scientific Publications, UK
  http://www.inasp.info/ajol/links.html e-mail inasp@inasp.info
- Food and Agriculture Organization of the United Nations
  http://www.fao.org/
- World Bank
  http://www.worldbank.org/

Resource material and references

Appendix 5. Stapleton, P., Youdeowei, A., Mukanyange, J. and van Houten, H. 1995. *Scientific Writing for Agricultural Research Scientists*. WARDA/CTA, Ede, The Netherlands. On CD.

Appendix 6. Publication as an Output of Science. Adipala Ekwamu. PowerPoint on CD.

Appendix 7. The Art and Ups and Downs of Scientific Publication. Adipala Ekwamu. PowerPoint on CD.

Appendix 8. Presentations and Style – Tips on Photography and Writing. Eric McGaw. On CD.

Boling, E. 1997. *Research Writing: Literature Review*, University of Indiana, USA.
Available from: http://education.indiana.edu/~istcore/r519/lit.html [30/07/2003]

Coghlan, L. 2000. *Referencing*. Portsmouth University, Portsmouth, UK.
Available from: www.pbs.port.ac.uk/~coglan/bus102/ref.pdf [30/07/2003]

Cooper, H. 1998. *Synthesizing Research: A Guide for Literature Reviews*. Sage Publications, Thousand Acres, California, USA.

Daniel, H. and Fisher, E. 1999. *Enjoy Writing Your Science Thesis or Dissertation*. Imperial College, London, UK.

Galvan, J.L. 1999. *Writing Literature Reviews*. Pyrczak Publishing, Los Angeles, USA.

International Network for the Availability of Scientific Publications (INASP). 2003.
http://www/inasp.info/ajol/links.html

Kung, H.T. 1987. *Useful Things to Know about PhD Thesis Research*, prepared for 'What is Research' Immigration Course. Computer Science Department, Carnegie University, USA.
Available from: http://www.eecs.harvard.edu/~htk/thesis.htm [30/07/03]

Lyons, K. 2003. *How to Write a Literature Review*. University of California, USA.
Available from: http://library.ucsc.edu/ref/howto/literaturereview.html [30/07/2003]

Macauley, P. 2001. *The Literature Review*, Deakin University. Geelong, Victoria, Australia.

Available from: http://www.deakin.edu.au/library/findout/research/litrev.html [15/10/2001]

Mouton, Johann. 2001. *How to Succeed in Your Master's and Doctoral Studies: A South African Guide and Resource Book*. Van Schaik Publishers, Pretoria, South Africa. 119 pp.

Ojwang, J.B. 1994. National domestication of the Convention on Biological Diversity. Chapter 16, In: *Biodiplomacy: Genetic Resources and International Relations*, Vicente Sanches and Calestous Juma (Eds.). African Centre for Technology Studies (ACTS), Nairobi, Kenya.

Royal Agricultural College Library Services. 2003. *Bibliographic References – The Harvard System*. Royal Agricultural College, Cirencester, UK.
Available from: http://www.royalagcol.ac.uk/library/referencing.htm [30/07/2003]

Rukuni, M. and Eicher, C. (Eds.) 1994. *Zimbabwe's Agricultural Revolution*. University of Zimbabwe Publications, Harare, Zimbabwe.

Saint Louis University. 1998. A *Student's Guide to Research with the* WWW. Saint Louis University, St. Louis, Missouri, USA.
Available from: http://www.slu.edu/departments/english/research [21/08/2003]

Sullivan, D. and Sherman, C. (Eds.) 2001. *Web Searching Tips*. Search Engine Watch.
Available from: http://searchenginewatch.com/facts/article.php/ [21/08/2003]

Taylor, D. and Proctor, M. 2001. *The Literature Review: A Few Tips on Conducting It*. University of Toronto, Toronto, Canada.
Available from: www.toronto.ca/writing/litrev.html (30/07/2003)

Tilburg University. 2003. *Searching the World Wide Web: A Basic Tutorial*. Tilburg University, New Zealand.
Available from: http://www.tilburguniversity.nl/services/library/instruction/www/onlinecourse/introduction.html [21/08/2003]

University of Zimbabwe Libraries. 2002. *Dissertation and Thesis Writing*: Online Resources, University of Zimbabwe, Harare, Zimbabwe.
Available from: http://www.uz.ac.zw/library/resources.html

3.2 Your research proposal

Kay Muir-Leresche

- **Planning research requires you to reduce and organise information**
- **What you can achieve is constrained by your time, skills, contacts, physical and financial resources**
- **Do not overstate what you will research - rather understate and achieve more**
- **You will be judged on how rigourously you approach the hypotheses and objectives you specify**
- **Continually involve your supervisor in the development of your research proposal**
- **Objectives should be simply and clearly stated and make it clear to the reader what you want to achieve by solving the problem**
- **A hypothesis is a statement of what you think is true based on available evidence. Your research will then prove or disprove this hypothesis**
- **You must provide the rationale for your hypotheses and you must be sure you have the resources to test them**

'Extraneous information and ideas are eliminated as foreign matter might be filtered in a funnel'

Andrew and Hildebrand (1982)

This chapter is directed towards the requirements for an academic degree. But it is important to keep in mind that if you are involved in a larger project that this will have demands of its own. **Chapter 3.4** covers the development of project proposals to obtain grant funding and it complements this chapter.

Where you are part of a project you may find yourself caught in the middle when trying to reconcile your thesis requirements with those for the project. Your research will be just one component of the larger project. This means that you will need to adapt your research to fit in with the requirements of the project.

Study the grant proposal and project documentation before you begin to refine your research proposal. Do not rely on what someone else says is needed. When you have done this, then carry out the steps outlined in **Chapter 3.1** before going on to develop your research proposal.

Getting started

You have decided on the area you want to research, you have gathered information on the topic and you have held discussions with prospective supervisors – you have established what you want to do. Now you need to work on your research proposal and you need to clearly define your research. In due course this research proposal will form the basis for the introduction to your thesis.

Your primary task is to narrow down and clearly specify what you hope to achieve. You need to specify the problem or issue you are addressing in your research. You need to show why this is an important topic. You can specify the problem in a broad context at first but then you will need to narrow this down to exactly what you will be researching and show how this will contribute. Planning research requires that you reduce a large volume of information to manageable proportions. Each part of your research proposal should be designed to make what you are researching clearer to your reader (and yourself!). To follow Andrew and Hildebrand's model, you start of with the wide section of the funnel and the general setting, this narrows to problems faced, then specifically to the problem you address and then to the hypotheses you will test, your objectives in testing them and the questions you will pose to gather the information. You need to keep filtering out surplus and less directly relevant information to make the orientation precise. The constraint is the narrow part of the neck of the bottle into which the filter must fit. This bottleneck is the resources available to you – time, skills, contacts, and the physical and financial resources for the research.

The research proposal is primarily an exercise in cutting back. For an academic thesis you can do more than you indicate, but do not state that you are going to do more than you are able to achieve.

What does this mean? It means that less is more. You should limit what you set out to achieve to the bare minimum required by your supervisor. If you achieve more in the course of the research that will be good and it will give you more information to back up your conclusions and to develop areas for further research. However, if you state that you are going to test many hypotheses, or one very complicated hypothesis, or achieve multiple objectives, and your thesis does not make a full and thorough attempt to address these, then an examiner could deem you to have failed. You will be judged on whether you make an acceptable attempt to carry out the research on those hypotheses and objectives you state you are addressing. This means that you should show a thorough understanding of the various approaches to the research as well as providing a thorough theoretical review, analysis and well structured testing of your hypotheses. It is extremely important that you remember that when determining your academic research **less is more.** Be precise about what you will achieve, limit it to fit your resources and then be sure that you are very rigourous in your academic enquiry.

Always limit what you say you are going to do. If you say you are going 'to discover why some small-scale farmers are successful and others are not' then you will be judged on whether you succeed in achieving this objective. Rather say your objective is 'to discover some of the reasons why some smallholders are more effective at x,y z than others'. That is far more realistic.

If you need to incorporate additional research in order to meet the requirements of other stakeholders (the communities/farmers with whom you are working, the funding agency and others) then provide them with a supplementary proposal. This should detail those aspects that you will include but which are not being included in the academic research proposal. The aspects covered in your academic proposal will receive a more academic and theoretical approach than those included in the supplementary proposal. **The requirements of a research report to stakeholders are different from the requirements of a thesis.** The former is more interested in your findings and in ensuring that the research process used is valid and that the findings are legitimate. In addition to these criteria, your thesis needs to show your examiners that you understand the theoretical concepts, are aware of the available research and analytical tools and of the literature. A thesis proposal must indicate to the academic committee how you will address these issues, whereas a research proposal to prospective funders or clients needs to be more explicit about how your project meets their needs. In rural development you also need to show how you will include affected communities in the research process and how they will benefit.

The research proposal for your thesis is a vital component of your research and can take up to a quarter of the time allocated to the thesis. Specifying the problem clearly is essential to avoid gathering the wrong information and/or using the wrong research tools and analytical methods and then having to start again, or to base conclusions on inadequate evidence.

If you include the context in which your research will operate you may find that this will help you to limit what you are planning to achieve. If you are explicit about the geographic, social, and economic bounds of the problem and solution you can avoid the trap of producing a grandiose but unachievable proposal.

You need to continually involve your supervisor and other lecturers in your research proposal phase. In many ways this is the most important time of your research. It is also the time when you are likely to interact most closely with your supervisor and is a good time to really get to know her/him.

The rush to the field

Getting out and being active is what most people want to do. If you really have to start before the proposal is ready, use your early fieldwork to refine the proposal. Use rapid rural appraisal to gain an overview, informally pilot some approaches, and pretest your questionnaires. Avoid starting to collect data before you have a research proposal accepted by both your academic committee and, if appropriate, those funding your research.

The research proposal

Setting the stage

You need to provide an introduction that gives the background to the topic, explains your rationale for choosing this problem and briefly includes a review of some of the other work carried out on the subject. You should avoid giving a detailed review of agriculture in that region – rather just highlight the elements most relevant to your research and provide references. However, you do need to provide the general background for the particular research problem you will address.

Example

Smallholder farmers in Africa (or in x country, or in y region) have very low yields and low incomes/poor health, etc. (provide references). The poor soil fertility and the impact of pesticide residues negatively affect productivity (references) and contribute to unsustainable production systems and a degrading environment (references).

This is a key part of your research proposal but you should avoid providing detailed information. It is more important for you to extract the essential elements and then to provide references indicating where information can be found. The length of this section depends on the type of problem you are addressing and on whether you will include a brief literature review after providing the statement of the research problem. You will need to get guidance from your own university on their limits and requirements. Some universities require you to have both a long and a short proposal. The long proposal would include a comprehensive literature review. However most academic committees when deciding on the merits of your research proposal will use a short proposal and will not want the introduction and literature review to be more than one or two pages.

Start off with as much of a background as will help you to set the stage. Then go ahead in specifying the problem and developing hypotheses and research questions. After this is done go back and cut the introduction to the required length, making sure what you leave in is directly relevant to your proposed research.

If you are applying for an independent research grant then you will need to have a long proposal with a detailed literature review to provide background to the short proposal. If, however, you are applying to be part of an existing or proposed project then your own proposal does not need to be so long. It does, however, need to be very clear on how it fits in with the project. What will your research contribute? How will it meet some of the specific objectives of the project? Will it address any already established hypotheses? If not, how do your hypotheses add value to the project?

The research problem

Your clear statement of the research problem can be in a section on its own or it can fit into the end of the Introduction. This is your opportunity to clearly indicate which component you will address from the general issue/problem outlined in your introduction. You need to use it to show how this aspect fits in to the general subject area and you can also use the section to show why addressing this particular aspect is important – to academic enquiry, to your client population, or to society in general.

At the end of this section you need a paragraph which briefly and clearly defines the research problem you are addressing.

Example

Pesticides and chemical fertilizers are expensive and frequently unobtainable. Some smallholders in region x have been successful in increasing incomes and maintaining soil fertility by using an integrated pest management scheme and crop rotation. However other farmers have been less successful. This study is particularly concerned with the problems of soil type (or distance to market; or access to draft to ensure good timing; or access to manure; or crops included in the rotation; etc) on successfully using IPM and crop rotation to increase income and sustainability.

It is important that your research problem statement includes the specific aspect that you will be considering. The first three sentences would be inadequate as the statement of your research problem. They need the final qualifying statement that will clearly indicate what you are going to address.

Literature review

In short proposals this may be included in the Introduction. In academic proposals it is important as it provides evidence to the academic committee approving your research that you understand the context of your research and what other work has been done in this area. The information provided in **Chapter 3.1** is relevant here, although, obviously the proposal will not provide a detailed analysis of the literature. In the proposal, the Literature Review is a brief summary of the most important information, highlighting its relevance to your choice of topic.

Objectives

What is the purpose of your research? What do you hope to achieve? In answer to these questions you will develop objectives for your research. Note that there is no particular order in which the objectives, hypotheses and research questions must be presented. Each school of thought will have its own approach and you must consult with your supervisor to see whether you can choose how you frame your proposal or if you must follow a fixed format. The objectives are also a way of identifying your clients, defining the limits of the research and describing the expected outputs in a clear and succinct way. It is useful to crystallise your objectives before you finalise your hypotheses, even if in the research proposal you put the objectives after the hypotheses.

The objectives need to be simply and clearly stated. They need to include both the general objectives related to stakeholder welfare and the specific objectives of the type of approach you will use and the aspects you will emphasise.

Objectives based on the above research problem statement:

1. To assist smallholders to improve livelihoods/ reduce health risks, etc.
2. To contribute to improved sustainability of the environment by isolating those factors that reduce the adoption/success of using IPM and crop rotations.
3. To use xyz baseline study to differentiate between farmers' adopting/successfully using the recommended technology and those who are less successful or non-adopters. If you are doing a doctorate, or a masters which is part of a larger study, then you could be involved in the baseline study. If you are doing the masters on your own it is too ambitious for you to carry out a statistically valid baseline study without available secondary data in the time limit. You could only do this if you worked in a team with other masters students on a collaborative research project.
4. To collect detailed information on the soils (availability of draft, access to market or whatever) of selected farms in the differentiated groups from the baseline study area.
5. To analyse the information and determine if the type of soil (whatever) affects adoption/ success.
6. If time is available (otherwise put in areas needing further research) – to conduct experiments varying the recommendations according to your preliminary findings. (This is usually applicable to doctoral rather than masters theses).
7. To provide recommendations to farmers/ technical agencies/policy-makers based on the outcome of this research.

Some parts of the above objectives are moving towards methods and are not strictly objectives in that they address more how to solve the problem than what the problem is. However, it can be useful to include an indication of method to help clarify it for yourself and your readers. But do not go into any details on the method part – the objective should be directed at what you will achieve by solving the problem – not how to solve it.

You will need to include some justification and background for these objectives, unless earlier sections have made the logic apparent.

Hypotheses

A **hypothesis** reflects what you think is true but which still needs to be proved. It is a supposition or a provisional explanation of something. Students are often overwhelmed when asked to present their hypotheses. In fact it is simple: A hypothesis is a statement of what you think is true based on available evidence. Your research will then set out to prove or disprove the validity of this hypothesis.

This is something you do every day. In countries with shortages a person might hypothesise that flour/sugar/fuel is more easily available in the low-density suburbs. They will then test this hypothesis by going there and trying to find the commodities. Another person may hypothesise that these commodities are more easily available in the poor suburbs and go there to find out. One researcher may hypothesise that these commodities are more easily available in the wealthy suburbs because people are able to afford the very high prices. Another researcher may hypothesise that they are more easily available in the poor suburbs because illegal exchanges are more easily facilitated in the open, informal markets.

It does not matter which researcher or individual is correct. That is the task of the research – to gather evidence and then to prove or disprove the hypothesis. What you do need to do when presenting your hypothesis is to make it clear what basis you have for presenting it. In the above example the two researchers will go on to test:

- Whether the commodities are more common in one area or another
- Whether the reason for this is the one provided in the hypothesis.

You need to show the logic of the hypothesis. When it is simple it can be incorporated into the hypothesis as above. In other cases the rationale needs to be included in the hypothesis section, unless the introduction or literature review make it obvious.

The 'why' part of the hypothesis is critical in most studies, and too often neglected. I see too many pointless hypotheses like 'Farmers vary in their willingness to adopt...' Of course they do! Such studies usually have finding out 'why or how' as the real objectives, but if their reasons are not hypothesised it is impossible to set up relevant methods, in particular to know what to measure.

In addition to providing the rationale for your hypotheses you must be sure that the hypothesis(es) you propose to test are manageable. Will you have the resources (time, money or skills) to test them, and are they in fact testable? Do not concern yourself with whether your hypothesis is likely to be proved correct. The purpose of research is to determine if the hypothesis is true and your research is just as valuable even if you prove that, against your expectations, it is actually not true.

It may help to specify a general hypothesis that drives your research. This may be one underlying hypothesis that you are not necessarily planning to test, for example,

'That farmers with the same physical resources will not always achieve the same results.' This is more of a tautology, but it could be useful to articulate it and provide the overview of factors which affect this before going on to specify the particular hypothesis you will assess.

'Farmers who are risk-averse will have lower gross margins over time'. In order to test this hypothesis you will need to be able to access time-series data. You would need to take into consideration that the period covered by the series is not exceptional, i.e., that there were not an abnormal number of drought years or greater or fewer policy upheavals etc. This is a testable hypothesis. However if you are not intending to test it, but to use it as an assumption underpinning much of the other research, then you need to discuss it in detail and provide the logic in your use of it in this way. You can then go on and present the specific hypotheses you will be testing:

'Farmers who do not have reliable access to remittances will be more risk-averse'. You can then go on and give your rationale 'These farmers cannot afford to risk the failure of basic food crops and are less able to experiment with production methods and commodities until these are proven to be as reliable as the traditional systems. Thus even if they have the same soils and inputs they will be less successful than farmers who are more food-secure because they are unable to risk growing unfamiliar commodities or using unproven production technologies.'

A **null hypothesis** presents your supposition in such a way that you would expect it to be disproved. The null hypothesis posits that there has been no change or difference. If there has been a change the null hypothesis is rejected which indicates that the alternative hypothesis is accepted. Null hypotheses are the usual form in statistical analysis where tests of significance are set up so that one first assumes no change or difference. In statistical analysis the onus of the proof rests with the hypothesis of change. It sometimes makes sense to present null hypotheses even when not using statistical analysis, but this is not usual.

Example of a null hypothesis

'The inclusion of legumes in the crop rotation will not affect maize yield'.

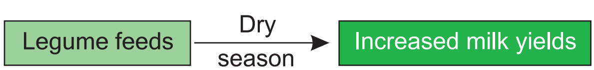
If the analysis shows that it does affect yield then the null hypothesis is rejected and the alternative hypothesis: 'That the inclusion of legumes does affect yield' is accepted. The problem is that it is always easy to 'fail to reject the null hypothesis' simply by doing a poor study. Thus the statistical concept of null hypothesis may be useful sometimes, but it does not help design an effective study. However the concept of 'null models' in ecology is important and under-used. The idea is that the consequences of 'no effect' may not be as simple as you think.

More often in applied research we know the null hypothesis of 'no effect of legume' must be false but are actually interested in things like 'Is it large enough to be of use to farmers?' So it would be better to change the hypothesis to reflect what is needed.

In applied research a hypothesis should also be useful, i.e., it should take you further in the problem-solving or remove a critical knowledge gap or blockage.

Formulating and testing hypotheses is the essence of scientific activity whether the field of study is in the natural sciences, social sciences or even the arts. The hypothesis is a theoretical proposition that can be right or wrong, true or false. It differs from a tautology which is self-evident and is always true.

A hypothesis will normally consider how an independent variable or concept is affected by a dependent variable or concept, it indicates that x is related to y, or x causes y. 'Access to supplementary legume feeds during the dry season will improve milk yield'. It can help if you put your hypothesis into a small illustrative model:



You do not necessarily do this in your proposal or thesis but it is a good way to see if you are specifying a testable hypothesis. And it also helps to identify methods, for example, by avoiding the effects of confounders. If you find three things pointing to 'milk yield', you have to have methods that will separate their effects.

The research process involves examining your subject area, finding a specific problem to address, formulating hypotheses, testing them against facts, and then either accepting or rejecting the hypothesis – or more frequently reformulating, amending and retesting in a continuous exploration. Formulating hypothesis is therefore an essential component of your academic research. **The key hypotheses form the core of your research and you should be sure that they are clearly specified.** During the course of your research you may formulate and test many sub-hypotheses in the body of your thesis that you are not including in either your research proposal or the introduction to the main body of your thesis. What you need for the proposal are the major hypotheses that are to be tested in order to assist in addressing your research problem.

Research questions

You have established:

- What you are going to research (the research problem)
- Why you are going to undertake the work (objectives)
- What you are going to test (hypotheses).

Now you articulate what you need to find out in order to be able to carry out this research – the research questions. The questions you pose will also affect the research methods you use to obtain the data for your analysis. They are discipline-specific.

If you are an animal scientist you may be most concerned with comparing milk yield for cows/goats with and without leguminous supplementary feed. If you are a crop scientist you will want to consider the impact on the yields of other crops grown in the rotation, whereas a soil scientist will be particularly concerned with the impact on soil fertility. If you are an economist you will want to assess all the benefits and costs involved for the farmer of including legumes in the crop rotation and you will need the results from all the other disciplines to determine the benefits. A sociologist or anthropologist will want to consider the effect of the changed rotation on social relationships – does it reduce or increase the income of women, or affect the nutritional status of children.

The research questions will help you to determine the required research methods and analytical approach.

Research methods and analytical framework

Although you may not be in a position to provide the exact research approach you intend to use, your research proposal must provide your approval committee with an indication of what approach you intend to pursue. You need to show them that you understand the requirements of different approaches and that you will be in a position to test the hypotheses specified, or to meet the objectives.

You should include an overview of the research and analytical techniques which may be applicable, and if you know, indicate which you will probably use.

Budget and timeline

In some universities the academic committee is not involved in how you will finance your studies. They judge the proposal only on its academic feasibility. At other universities they do need to know that your project is sustainable and a budget is required.

If you are applying for a research grant from the university or as part of a project, then the budget is critical. You need to be sure that you provide enough information to justify your requests. You also need to ensure that you cover all the expenses you will face. You are unlikely to be awarded further funds during the process of the research.

Take your budget to someone who has carried out a similar project and ask her/him to check that you have everything covered. Also be sure to discuss it with your supervisor.

It is important that you think carefully about the time you expect to take. With a time-limited dissertation and with funded research, you will have to stick closely to this timeline. If you are doing a thesis with a flexible timeframe, it is also important to stick to your timeline. Many theses are abandoned because they have dragged on for too long. If there had been more rigid time requirements, this may not have happened.

Well done – you now have a workable research proposal. You are much further down the road towards getting your MSc, MPhil or PhD! Your supervisor is happy with the proposal and the funders have accepted it. Now is the time for you to go into the field. Do not necessarily wait for formal academic committee approval. One of the biggest problems you might face are the long delays in getting proposals approved and theses examined. In some universities this can be delayed by cancelled meetings, strikes and committee members away on consultancies. If the committee return the proposal with requests for changes you can make them as required. The problem will come if they want you to change the data to be collected after you have already started.

See if you can get someone (other than your supervisor) who is on the academic committee to provide comments on your proposal before you go into the field. With two different perspectives accepting it, you should not have to make big changes.

Resource material and references

Appendix 1. The Craft of Research. Paul L. Woomer. PowerPoint on CD.

Appendix 5. Stapleton, P., Youdeowei, A., Mukanyange, J. and van Houten, H. 1995. *Scientific Writing for Agricultural Research Scientists*. WARDA/CTA, Ede, The Netherlands. On CD.

Appendix 6. Publication as an Output of Science. Adipala Ekwamu. PowerPoint on CD.

Appendix 7. The Art and Ups and Downs of Scientific Publication. Adipala Ekwamu. PowerPoint on CD.

Appendix 8. Presentations and Style – Tips on Photography and Writing. Eric McGaw. On CD.

Andrew, C.O. and Hildebrand, P.E. 1982. *Planning and Conducting Applied Agricultural Research*. pp. 14–33. Westview Press Inc., Colorado, USA.

Dixon, J., Bouma, M. and Atkinson, K. 1987. A *Handbook of Social Science Research*. Oxford University Press, Oxford, UK.

Greenfield, T. 2002. Writing the thesis. In: *Research Methods: Guidance for Postgraduates*, Greenfield, T. (Ed.), pp. 307–316. Second edition, Arnold Publishers, London, UK.

Ronald, M. and Bernauer, T. 1998. Empirical research on international environmental policy: Designing qualitative case studies. *Journal of Environment and Development* Vol 7: 4–31.

3.3 Writing your thesis[1]

Kay Muir-Leresche

- **You must make the essential components of your work readable by a broad audience**
- **Be logical and consistent in your arguments**
- **Be rigourous in your approach and do not make unsupported statements**
- **Be aware of your bias and avoid allowing it to influence your work**
- **Be aware of which approach you are using for each hypothesis. Use historical analysis, literature reviews and case studies for deductive reasoning. Use statistically valid samples and experiments for inductive generalisation**
- **Writing a thesis needs 'soul' time. Be sure to allow yourself the space to be creative**
- **Set yourself deadlines and be strict in enforcing them**
- **Remember your work is only one imperfect step towards the transformation of rural areas**

'A poorly written thesis can result in failure of otherwise excellent research. A well-written thesis cannot rescue poor research but it can help a marginal thesis pass.'

Derived from **Stan Taylor (1996)**

Introduction

Your thesis is where you bring together all your work and make it intelligible to your peers. It is very important that you make every effort to present your research professionally. Do not throw it all together and hope that the value of your research will enable you to pass.

You should produce a thesis which is clear, concise and both comprehensive and comprehensible. Any person who understands their topic well enough is able to provide a clear explanation. The analysis and the proofs contained in the thesis may only be clear to a specialist in that area but your interpretation of them and your rationale for using a particular approach need to be as simple as possible. This is particularly relevant for work in rural development where problems are not discipline-specific and people from other disciplines must understand your research.

Avoid using **jargon** where possible. Your examiners are more likely to be impressed if you can simplify what you are writing so that it is clearly understood. Jargon is often open to the dictates of fashion and your work will stand the test of time better if you can avoid using it. It should be distinguished from technical language. Jargon is language characterised by pretentious syntax, vocabulary or meaning. **Technical language** uses words or phrases that have special meanings in a specialised subject; such language is useful provided the special meanings are defined explicitly.

Your thesis will be available from your university library for others to read, even if you do not publish it as a book or in journal articles. It reflects on your university, your faculty and in particular on you and your department. You need to produce a high-quality thesis because it will affect your reputation.

1. Common practice uses thesis and dissertation interchangeably but they are different. A **dissertation** is a formal written discourse on a subject, whereas a **thesis** sets out to provide original research to prove or disprove a particular argument. A thesis is normally submitted for a doctorate and a dissertation for undergraduate and taught masters degrees. The advice given in this book for theses, applies equally to dissertations.

The structure

You need to set out the thesis so that your reader is able to follow a logical path. This path outlines:

- Why you undertook to do the research (the research problem) and how it fits into a wider research and development strategy
- What work had gone before (literature review)
- How you collected the information, and the information collected (research methods and data tables or historical information)
- How you analysed the information (analytical tools)
- What the results were (results of analysis)
- What these results mean in terms of your hypotheses and objectives (interpretation and conclusion)
- What they mean for policy
- What actions can be taken
- What further work would contribute to better systems.

Stan Taylor (1996) uses the analogy of an explorer producing a guidebook to clarify your role as a researcher writing a thesis, or dissertation. As the author of the guidebook you need to explain:

- Your starting point and why you decided to embark on the journey
- How you decided to undertake the journey
- The route you followed and the discoveries you made on the way
- How, in the light of the above, you redrew the route
- Where you arrived at the end of your journey
- How it differed from your starting point
- Where you go from here.

This logical progression is normally achieved by following a fairly well laid-out structure. The structure of the main body of the thesis will vary, and will be subject to the specific requirements and norms of your own university. Whatever the structure all the following elements need to be incorporated into your thesis.

Title

A title should be as brief as possible but should still indicate precisely what is covered. It may be useful to provide a sub-title which places the research into its precise geographic, social or other niche.

The title page of your thesis should include the Title (and sub-title if used), your full name, the name of the degree to be awarded, your Department, Faculty and University and the month and year submitted for examination.

Abstract

Think of your abstract as your explanation to an aunt of what your research is about, what you discovered, and what it means. It is very important that you do not use the abstract as an introduction. It must give a synopsis of your most important results as well as indicating why the research is important and what its implications are. The abstract needs to either explain or avoid technical terms. If, for example, your results include a domestic resource cost (DRC) ratio, then say: The DRC of .... indicates that there is (or is not) a comparative advantage in producing x.

The abstract should be no longer than one page of single-spaced text.

Acknowledgements and dedication

You do not need to include a dedication but you may like to. Alternatively you can include your dedication in the acknowledgements. The acknowledgements should highlight your supervisor(s) role and those of any other mentor who has assisted you. You can use the acknowledgements to indicate where part of your thesis may be particularly beholden to someone else, e.g., if you used someone to develop or help you develop a mathematical model. Keep this section as brief as you can and avoid a long list of names that includes everyone to whom you spoke.

Acronyms, abbreviations and glossary of terms

An acronym is a word formed by the initial leters of other words, such as ICRISAT (International Crops Research Institute for the Semi-Arid Tropics). An abbreviation may be formed by the initial letters of words or in some other way but it does not create a new word, such as UN. If you use very specific terms, or vernacular terminology they should be explained in a glossary.

Where you put this list differs, and you should check your faculty regulations. If you have a choice, it is more useful to your readers if it is near the front of the thesis.

Table of contents

In a Table of Contents you should use roman numerals for all pages leading up to the Introduction. It does not matter where you put page numbers, unless your University requires them to be placed in a particular place. In the table of contents first list all those pages with Roman numerals, then go on to provide a comprehensive list of each chapter including all the headings and sub-headings with the relevant page numbers.

Provide lists of all tables, illustrations and appendices and indicate the page numbers for each one.

The body of the thesis

Within the body of the thesis the Introduction will be based on your research proposal, taking into account the adaptations you have made along the way. This is where your reader learns what led you to undertake this research (not your personal path – but the literature and evidence of problems that led you to decide on the research), the objectives of the research and the hypotheses you will be testing. You can also use it to explain any circumstances that are peculiar to the situation you address or which made the research more or less difficult. Look at **Chapters 3.2** and **3.4** for input on hypotheses and objectives.

The literature review may be a separate chapter on its own, it may be only part of the introduction and part of the research methods chapter, or it could form more than one chapter and be part of other chapters. (See **Chapter 3.1** for more details on accessing and using literature).

Research methods and data presentation may be contained in one or more separate chapters. For some theses gathering the data is a major component of the contribution and presenting the data collected is an important component of your work. Refer to **Part 4** for information on how to present data, and on selecting the data presented. You need to be sure to provide an overview of which research methods are available and why you chose the particular method you did. You should also, show why you chose to collect those variables, the sites selected, time periods and any other important factors. Your examiners will be unimpressed if you just went out and collected everything that was available. An important part of the process is to know what to collect. You waste both your time and that of the communities involved in the data collection if you collect more than you need for your specific objectives and hypotheses.

The analytical techniques that are possible (with more detail on those used) can be put in front of the results in a chapter covering both analysis and results, or they can be in separate chapters. If your thesis is very involved in establishing effective models or in finding the best analytical techniques, then this may cover several chapters. If you are using a widely accepted methodology, then a simple introduction in front of the analysis and results chapter is fine.

In most cases it is best to present your results simply and clearly giving only a brief explanation of what they mean but without providing discussion and interpretation. However, for some research it may be appropriate to include some discussion with the results and they may be incorporated into several chapters leading to the overall conclusions.

The best way is to present the results tables is with a sentence or two below each one simply explaining what it means without going into discussion of the implications. These explanations are to show the examiner that you understand the results and to allow people from other disciplines some understanding of what the results mean.

The discussion of the results can be a separate section at the end of the presentation of the results or be in a chapter on their own. This discussion can be used particularly to provide input from your results to the different options relating to the problem and it is here that you would make recommendations arising from your research.

Your conclusion should not be too long. It should highlight your objectives and your hypotheses again and show how your thesis has addressed them. You should then summarise the discussion of the results, presenting them as implications and policy recommendations. For many theses, the most valuable contribution will come from your recommendations on areas for further research. This is a component that is often tagged on at the end with little thought. It should be given much more attention. You should not only indicate what needs to be done but, if you have the information, you should indicate how it should be done.

Throughout your work you should keep making a note of where you are frustrated because you don't have access to certain information. Whenever a villager or anyone asks you what your research shows should happen with respect to x, and in fact there is more work that needs to be done for you to answer the question, then note it down. When you come to consider the areas for further research you then select from these notes those you consider the most important and most interesting.

References

Every thesis must include a comprehensive list of all the references cited in the text. Examiners are very particular about references because it is a hallmark of academic work that the reader is able to follow up on what is being presented. Every reference cited must be included and the references must be presented in the style required by your university. Refer to **Chapter 3.1** for more information on reference lists and bibliographies.

Appendices and footnotes

Appendices and footnotes are very useful because they provide you with the opportunity to present information which may be of interest, but which, if included in the body of the text, would divert attention from the main topic. Appendices allow you to present information both as data and analyses which are relevant to the topic but outside the direct purpose of your research. However you should not simply include all the data you collected or all the analyses. You should only include information in an appendix if it is relevant to your thesis and if it would be useful to a reader. If you include too many appendices it will seem that you are unable to discern what is relevant. It may also make your thesis too heavy to carry!

Footnotes are very useful in providing some additional information but if you have too many of them then again it will appear that you are unable to select what is important, and they can also affect the readability of your work.

Presentation and style

Modern technology has made it easy for work to be well presented. You need to ensure that you are consistent in the way you present headings, tables and other information. This consistency is also important in the use of fonts, bold, spacing, bullets and other format issues.

You must follow the style required by your university for titles on tables, figures and illustrations; for headings and sub-headings and indentations.

For a discussion of style you should refer to *Scientific Writing for Agricultural Research Scientists* by Stapleton *et al.*, the full text of which is available as **Appendix 5** on the CD, as are several more helpful appendices. You can also consult Greenfield (Chapter 37, 2002) and Levy (Chapters 38 and 39, 2002). There is no one acceptable style but in general a formal, impersonal style is normal for theses and dissertations. It is interesting to note that Greenfield (page 312) argues strongly for using the pronoun, 'I' and a more personal approach. Levy on the other hand argues for a more formal and impersonal style (page 319). This contradiction reflects the different schools of thought and highlights the need for you to make your own choices. These choices are constrained by the requirements of your university.

Before you start to finally write up your thesis, you should look up the regulations at your institution. You should format your chapters to match these requirements when you first start drafting your chapters.

Your university will normally require you to present your thesis in a very specific way. It is useful to format your chapters correctly from the start. This will save you time. It can be very difficult to adjust your layout to new margins or to redo all the headings to meet university regulations.

Also check on your supervisor's preferences beforehand! You can ask your supervisor about format issues, but I suggest you read some of your supervisor's work (in fact the work of all the members of your advisory board if you have one) to be sure that you understand the style he or she prefers. You can choose to have a different style but then you should be prepared to defend it.

Writing

The approach you use

When you come to writing up your work it is important that you understand what underlies the logic of your arguments. You need to know where it is appropriate to rely on different approaches. Most theses will include an element of each of these approaches for different aspects, but one form of reasoning should provide the driving logic for the thesis and be the main focus for your 'guide book'.

You need to understand whether you are relying on **deductive reasoning** – this is where you derive your hypotheses from theory and is sometimes referred to as **conceptual reasoning.** It will be most applicable to theses where you are going from the general to the particular. This is the approach popularised by Karl Popper (1934, English version, 1977) for the social sciences and is normally followed by the testing of these hypotheses using historical analysis, literature reviews, or through the testing and collection of case study and sample data.

Inductive generalisations are used when you go out and collect specific information (from samples) and then apply the results to a broader population. This is the most common method of testing hypotheses in the natural sciences. In order for the specific to be valid for generalisation to a population, the sample will have to be statistically valid. **Case studies** are useful in providing a fuller understanding of a particular situation but can seldom be extrapolated to a valid generalisation.

Retroductive reasoning is a less-common research approach than deductive reasoning and is most relevant to developing and shaping theory. It is appropriate when the situation appears to be different from what we would expect theory to predict. It is where you gather empirical data and then develop hypotheses (make inferences) based on the information collected.

In the writing of your thesis you need to be sure that everything you claim is justified so that anyone challenging your results cannot fault you on the logic of your approach but only challenge the data you may be using.

Common fallacies of scientific reasoning (derived from Mouton, 2001)

Unsupported generalisations

Unsupported genrealisations present a very common trap and you need to be extra careful to avoid them – economists and other social scientists are particularly vulnerable. You must have evidence to support claims. Every time you state something is ... check to see if you have included your basis for saying it is so! Try and avoid 'all' or 'every' unless you are sure there are no exceptions.

Appeals based on authority

It is not always enough to claim that your evidence derives from the opinion of another 'expert'. If the source is internationally recognised as an expert you may get away with it. However it is better to get more than one source to support what you are saying or to use other evidence as well.

Impressing by large numbers (bandwagon)

Remember that all numbers are relative and that you need to put the data into context. Even where statistical analysis bears out the results you would like, it is useful to put it into the context of findings of similar research elsewhere.

Affirming the consequent (*post hoc* fallacy) otherwise known as illogical reasoning

You are probably aware of the requirements for statistical validity and may not fall into this trap but you need to be very careful when linking cause and effect. Just because when you harvested pearl millet, the clouds came over does not mean that millet-harvesting causes overcast conditions! If something appears illogical to you, then very carefully check how the researcher has linked cause and effect and what evidence they have for that link.

False analogy

False analogies are most frequently found when making comparisons. Because some things are similar it doesn't follow that all are, for example, given that Mozambique and Zimbabwe are both African and both developing it does not necessarily follow that they are the same country.

Circular reasoning

Circular reasoning happens when you try to prove a point by returning to the point itself.

A*d hominem* reasoning (attacking the person)

A*d hominem* reasoning occurs when you attack the credibility of the person instead of addressing the issues and attacking the argument.

Non-sequitur reasoning

This is where the reasoning is not logical and does not actually follow, e.g., it is a *non-sequitur* that: 'violence in films has a bad effect on children and so people should not see violent films.' The finding refers to children so you cannot make the conclusion for adults.

Red herring argument

A red hering argument in what happens when you bring in a side issue with no real relevance to your research. It only serves to distract the reader from looking for the proof that should be presented.

Some tips on writing up

Writing up your research is one of the most difficult tasks you will face. This is particularly true for those who are more used to practical work either in the field or the laboratory. It can also be the least-interesting aspect of the thesis because you now know what you found and it is tiresome to have to spend as much time writing it up for others as you spent in finding out the answers. You should avoid thinking in this way as it will affect how you write. Rather consider it an equal challenge to make what you know accessible; to prove to the world that you are capable. By making your work accessible you can make a difference to the world. If you do not take up this challenge then what you have found out will be as useful to the world as playing with a puzzle.

Another problem, even if you do see the written work as important, is that you may face writer's block. This is a problem almost all writer's face at one time or another. A suggestion is that you first give yourself a few days of total space to allow time for your subconscious to work. If that still doesn't help then just force yourself to write whatever it is as a letter to your grandparents or a cousin in the rural areas. In this way you make yourself simplify the problem and do not feel overwhelmed by the need to meet some outside standard. Once you have the basic concepts down, it is much easier to translate them into the more formal and detailed style required for your thesis. Try it – it really does work and it can help your thesis!

You also need to keep in mind that every person operates differently. For some it is important to stick to the routine of writing a certain number of pages each day regardless of how slowly. For others, the subconscious is always important and it may be necessary to give your mind time free from distractions that will then give you the chance to be much more productive when you do write. I am one of the latter and if I am producing original writing, I normally need time with no distractions pottering in my garden and reading light novels before I can really be effective. But such periods need some limits or they become simple indulgence.

Writing a thesis requires 'soul time'. You must recognise that you need to be able to dedicate all of yourself when you are writing your first draft. It is difficult to be creative and innovative and to really assess the implications of your results if you are not totally focussed on your research. You need to be aware of this and block out time completely away from other work and family duties in order to produce something really worthwhile. When you are feeling stressed or overworked, or when you have writer's block, then edit what you have already written. Design tables and figures and improve the style and presentation. These things allow your subconscious to continue to work and help you to overcome your block.

When you are in the process of writing, get up at frequent intervals – go for a short walk, stretch, or go and get a cup of tea. Avoid interaction with others so that when you come back and sit down you will find your mind has already arranged what you are going to say.

You need to be sure that you give yourself strict deadlines. With an MSc this is normally done for you and you are forced to complete. However for research degrees and doctorates, supervisors tend to leave you to determine your pace for yourself. There is a very real danger that if the project with which you are working finishes before you hand in your thesis, that you will never complete your degree. You really have to set yourself very strict deadlines and you cannot take on other consultancies or work, no matter how pressing your financial needs, if it means that you won't complete. You have to be very strict with yourself. If you leave your research for more than a year, rural development progresses so fast that you may find what you were working on has already been overtaken.

No matter how badly written it is imperative to produce a first draft promptly. It is much easier to rewrite and to see where you have gone wrong once you have it down on paper.

Another really important point is to remember that no matter how long you work on your research and on your thesis,

it is NOT possible to produce the perfect research, and it is not possible for your research to solve all the problems.

You have to draw the line and you have to accept that your work is not necessarily going to win a Nobel prize or transform rural areas or solve food security. Your work is only one imperfect step towards that goal. Provided it makes that step honestly and boldly and to the best of your ability, you can do no more.

After you have reviewed the literature in some depth and done your fieldwork and while you are in the labourious process of entering, or supervising the entry of your data, you should write your entire thesis. It will give you a needed break from entering numbers or working with your experiments. Before you have analysed the data and even before collecting all the data, write your thesis. In the results chapter you can put in what you think the results will be. It also helps to even outline the conclusion and fill in what you expect to find with respect to your hypotheses. This will assist when, almost inevitably, your results are different. You can then look for the explanations and see why the results differ from theory.

Although you will significantly change what you have written, even for the introduction, literature and methods chapters it will really help you to have the overall thesis structure in place. You will be writing it while you are living it, and this can also help you to chart a clearer course. You won't find when you get to the end that you have left out some vital component because by writing it first you will have identified what needs to be in place. Best of all, you can avoid the tension created when all the research is complete and you now have to write. This is advice that is seldom followed – but if you take it, you will find completing your postgraduate degree much more likely, and more fun.

Resource material and references

Appendix 1. The Craft of Research. Paul L. Woomer. PowerPoint on CD.

Appendix 5. Stapleton, P., Youdeowei, A., Mukanyange, J. and van Houten, H. 1995. *Scientific Writing for Agricultural Research Scientists*. WARDA/CTA, Ede, The Netherlands. On CD.

Appendix 6. Publication as an Output of Science. Adipala Ekwamu. PowerPoint on CD.

Appendix 7. The Art and Ups and Downs of Scientific Publication. Adipala Ekwamu. PowerPoint on CD.

Appendix 8. Presentations and Style – Tips on Photography and Writing. Eric McGaw. On CD.

Anson, C.M. and Schwegler, R.A. 2000. *The Longman Handbook for Writers and Readers*. Second edition. Longman, New York, USA.

Daniel, H. and Fisher, E. 1999. *Enjoy Writing Your Science Thesis or Dissertation*. Imperial College, London, UK.

Greenfield, T. 2002. Writing the thesis. In: *Research Methods: Guidance for Postgraduates*. Second edition. Greenfield, T. (Ed.), pp. 307–316. Arnold, London, UK.

Jones, R., Bizzaro, P. and Selfe, C. 1997. *The Harcourt Brace Guide to Writing in the Disciplines*. Harcourt Brace, New York, USA.

Kung, H.T. 1987. *Useful Things to Know About PhD Thesis Research*, prepared for 'What is Research' Immigration Course. Computer Science Department, Carnegie University, USA.
Available from: http://www.eecs.harvard.edu/~htk/thesis.htm [30/07/03]

Levy, P. 2002. Presenting your research: reports and talks. In: *Research Methods: Guidance for Postgraduates*. Second edition, Greenfield, T. (Ed.), pp. 317–333. Arnold, London, UK.

Mouton, J. 2001. *How to Succeed in Your Master's and Doctoral Studies: A South African Guide and Resource Book*. p. 119. Van Schaik Publishers, Pretoria, South Africa.

Popper, Karl, 1977. *The Philosophy of the Logic of Scientific Discovery.* (Originally published 1934 by Springer in German.) Routledge Publishers, New York, USA.

Taylor, S. 2002. Managing your PhD. In: *Research Methods: Guidance for Postgraduates*. Second edition. Greenfield, T. (Ed.), pp. 13–27. Arnold, London, UK.

Tham, M. 2001. *Writing Research Theses or Dissertations – Guidelines and Tips*. University of Newcastle-upon-Tyne, UK. Available from: http://lorien.ncl.ac.uk/mingDept/Tips [30/07/03]

University of Zimbabwe Libraries. 2002. *Dissertation and Thesis Writing: Online Resources*. University of Zimbabwe, Harare, Zimbabwe. Available from: http://www.uz.ac.zw/library/resources.html

3.4 Preparing a grant proposal

Paul L. Woomer

- **Some donors require specific formats but all include similar elements**
- **Proposed research methods must be cited and be suitable to test the stated hypotheses**
- **Highlight the expected impact of the research on beneficiaries**
- **Indicate how you will communicate results**
- **Keep what you say simple, clear and easy for a non-specialist to follow**
- **Get background information on your donors and be sure to direct grant proposals to the appropriate organisations**
- **You can apply to more than one donor – but only accept funds from one donor unless there is explicit approval for multiple donor funding**
- **Remember to always acknowledge donors when publishing**

'A proposal doesn't succeed. It's the project that succeeds. A good communication process is important, including calls, writing, and follow up.'
Joseph H. Cruikshank (1993)
The Clark Foundation

Introduction

Over the past several years, I have had the opportunity to familiarise myself with grantsmanship in a variety of ways including having several research grants awarded by different donors, by assisting other scientists to develop research proposals and by acting as a technical reviewer for the donor community. If you are an enthusiastic scientist with burning ideas and little experience in preparing research proposals for the international donor community, I offer the following advice on proposal format and considerations. The proposal format is generalised and stresses continuity between proposal components. Remember that some donors require very specific formats while others prefer to evaluate a proposal, in part, by its very composition.

Rember too that whatever the format, you proposal should be as clear and well writen as you can make it. Consult Stapleton *et al.* (1995) **Appendix 5** and other appendices on the CD for useful guidance.

Proposal format

A simple, short proposal is often best. One format that has been very successful in the past follows:

Title page. The title page includes the proposal title, principal investigator, co-operating investigators, complete contact details of principal investigator, proposal duration, funds requested, and a brief scientific summary (1 page). The summary should not be more than half a page, start with a one-sentence description of an agricultural or natural resource problem, how you plan to approach the problem, the time required to do so and the funds that are requested.

Introduction, justification and literature review. A clear statement of the problem and a state-of-the-art review of the research topic is necessary to convince donors that the proposed research is original and meaningful. This is the best place to introduce any conceptual or mathematical models upon which your proposed research is based. Any reference citations you include must be of the latest published work in the field. This is an area where access to computer-based bibliographic software is often extremely important because claims that a certain topic remains poorly understood must be substantiated through strong and current citation. In most cases, 2–3 pages of tightly worded introduction are sufficient, followed by a one-page justification and a comprehensive yet concise literature review (3–8 pages). It is important to include either original or properly cited conceptual diagrams and syntheses tables in the literature review.

Objectives. State a general and a few more specific objectives of the proposed research. These objectives must be well worded (and short). The objectives will lead to the working hypotheses of your proposal (0.5 page).

Hypotheses. The statement of a clear general, global hypothesis and a few specific (working) hypotheses is essential to a strong scientific proposal. What do you intend to prove or disprove? State these very carefully because this sets the standard by which the success of your research progress will eventually be judged. Many donors prefer research proposals that are hypothesis-based and most reviewers pay particular attention to the clarity of stated working hypotheses. You must avoid tautological statements and those that containing jargon in your hypotheses Social scientists often prefer to pose 'Research Questions' rather than hypotheses. You should be able to state your hypotheses in half a page.

Research approach. This section is similar to 'Materials and Methods' except that special effort must be placed upon how the approach you are proposing is suitable to test your previously stated hypotheses. You need not go into exact methodological detail but all proposed research methods must be cited in this section usually within 2–4 pages. This section can be sub-divided into:

General experimental approach and site characteristics. Does your research strategy fall into discrete phases? Examples of different phases are farmer survey, glasshouse, and field investigations. A description of the proposed research site, or sites if you are comparing two or more, should be included in this section.

Treatments and their rationale. Which treatments will be compared to test specific hypotheses?

Experimental design. This sub-section should specify an exact experimental design or key treatment contrasts, identify sample populations and, whenever possible, include a diagram that identifies treatments, plot sizes and the number of replicates.

Measurements. Which measurements need be taken to prove your hypotheses, how often will you take them?

Analysis of results. How will the experiments mentioned in the above section be analysed, what facilities are available to you for state-of-the-art analyses (co-operators, computer facilities, software)?

Research outputs and impacts. What do you anticipate the key accomplishments of this research topic to be and how will these be popularised? Include such activities as scholarly publications, popular publications, student degrees and agricultural field days (0.5–1 page). Research outputs are often well expressed as research products. What are these products, who are their likely beneficiaries, how might these be disseminated and what are their likely impacts on immediate clients and society as a whole? Remember that applied science exists to offer society solutions to its concerns! (1–2 pages, often presented as a numbered list of complete statements).

Timeframe and logistics. What will be done when? How will different components of the research project interact and complement one another? The time frame is best displayed as a table or figure. The logistics indicates to whom the funds will be dispersed and how equipment will be purchased and acquired. Many donors require specific details on how funds will be administered within a given organisation, including whether or not principal investigators serve as signatories and if funds intended for hard currency purchases are first converted into local currency upon release (1 page).

Budget and budget notes. This is a simple table with which most scientists are certainly familiar. What funds are required for general budget items by year? These should be subtotalled by year and item in US$ or the hard currency specified by donor, e.g., £ Sterling or Euros, and also, if you choose in local currency. Typical items include Equipment, Supplies, Technical Support, Travel (local and international), Student stipends, Communication, Office supplies and Miscellaneous. Details of these items should be included as footnotes (1 page).

Literature cited or references. Place these in the format of an international journal within your discipline but do not abbreviate journal titles (See Chapter 3....?) (1–3 pages is usually sufficient). Be aware that many technical reviewers will evaluate your literature by conducting a literature key word search using bibliographic software and will compare the recovered references to those listed in your Literature Cited.

As you can see, an acceptable proposal may be as short as 12 pages in length but will clearly specify why the research is important, what will be tested, how it will be tested, what measurements will be made, how data will be compiled and analysed, who is likely to benefit from the research, when it will be completed and what funds are required. The above is to some extent a general guideline that has proven effective in the past but is subject to modification depending on your preferences and grantsmanship style. As we are now well into the personal computer revolution, donors expect potential grantees to be computer literate and to prepare their proposals with word processing and graphics software. The overall appearance of a proposal is indicative to donors of an applicant's ability to later publish their research findings. Incorrectly spelled words, inconsistent heading and subheading structure, poorly constructed tables and improperly cited references are liabilities to an otherwise strong proposal. Additional information on proposal formats is available from Stapleton *et al.* (1995, **Appendix 5**). The process of identifying and refining a research topic, and the transition from preparing an outline, writing a first draft and completing a final draft are described by Baugh (1995).

Funding level. Well written research proposals in the area of agricultural resource management that seek between US$15,000 to US$30,000 per year for 3 years are frequently funded. Feel free to ask for less ($45,000) but be reluctant to ask for more ($90,000) especially if the grant is your first proposal to a particular donor. Given the expense of field experimentation and costs of technician and field labour, this level of funding is able to keep a research team very busy and to partially re-equip a laboratory as well.

Timeframe. Whenever possible the completion of a research proposal should be a fairly rapid process. Very few donors will consider funding a single research project for longer than 3 years and many prefer a 2-year duration. This time frame allows donors to assess research progress and then encourage successful grantees to submit an extension study. Also, remember donor's funding cycles. In general, donors begin to exhaust their funds by mid-year but it is never too early to submit something for the following year. Many donors have well established technical review procedures that require several weeks or months to complete. Whenever possible, ask the donor when is the best time of year to submit your proposal (and then do it!). Sometimes life is full of surprises and a donor may offer seed money on the spot to test the feasibility of a research proposal, especially one that is as near the cutting-edge of current thinking on a topic. Do not hesitate to accept these funds at this point even though the offered funds are less than you have requested. Offers of seed money may signal that the donor is very interested in your proposal and wants to see how successfully you can initiate it.

Continuity between proposal elements

The key to a successful grant proposal is continuity between proposal sections. The objectives must be few and clearly stated and in turn lead to well worded working hypotheses. These hypotheses must be stated in such a way that the experimental design and important measurements are obvious to the reader. If more than one experiment is being proposed within a single grant proposal, these different experiments must be easily distinguishable from one another (sometimes a donor may be interested in funding only part of the proposal). When writing your grant, you must be prepared to anticipate different experimental outcomes and identify the potential significance of these research outcomes upon outputs and impacts within the proposal itself, rather than later when experimentation is completed. **Keep-It-Simple-Scientist (KISS):** remember that often a donor may not be a specialist in your field of science but is generally interested in the impacts of many areas of science on agricultural development, food security, environmental awareness or curriculum development.

Avoid proposal drift

Sometimes people's thinking and ideas change or further develop during the proposal writing process. This must not be considered undesirable because it is a fundamental part of the learning process. What is undesirable, however, is when these changes are reflected in the finished proposal. Every next section of a proposal must be a sound reflection upon the parts preceding it. One all-too-common example of 'proposal drift' is to add an unexplained experiment or set of treatments when the proposal reaches the Experimental Design section without returning to the Introduction, Objectives and Hypotheses sections to reinforce the necessity of that experiment. Beware, sometimes research activities are included seemingly as an afterthought. One sure way to not be funded is to confuse the donor representative who reads your proposal!

Emphasise the role of students

One of the advantages of university-based research is the potential to involve students within research programmes. The enthusiasm, commitment and cost effectiveness of graduate students in conducting research is well known to donors, and some donor programmes require student involvement. In this regard, proposal authors are encouraged to identify their candidate students and specific research topics within a proposal. At the same time, investigators must not give the impression that they are imposing tightly prescribed research programmes but rather have opportunity for creative input. What is more, authors must be careful not to appear as delegators of responsibility to co-operators and students, but rather they must provide scientific leadership and maintain their own responsibilities to the project as a whole (Patel and Woomer, 2000). If you know your research could fit into a large project, be sure to get involved in writing your bit of the proposal.

Paragraph and section structures

One of the keys to successful scientific writing is your ability to adhere to sound paragraph structure. **Paragraphs** are the units through which concepts are presented and logical arguments constructed. Many paragraphs should consist of a single **claim** in the opening sentence, followed by **evidence** in support of that claim in the next few sentences and conclude

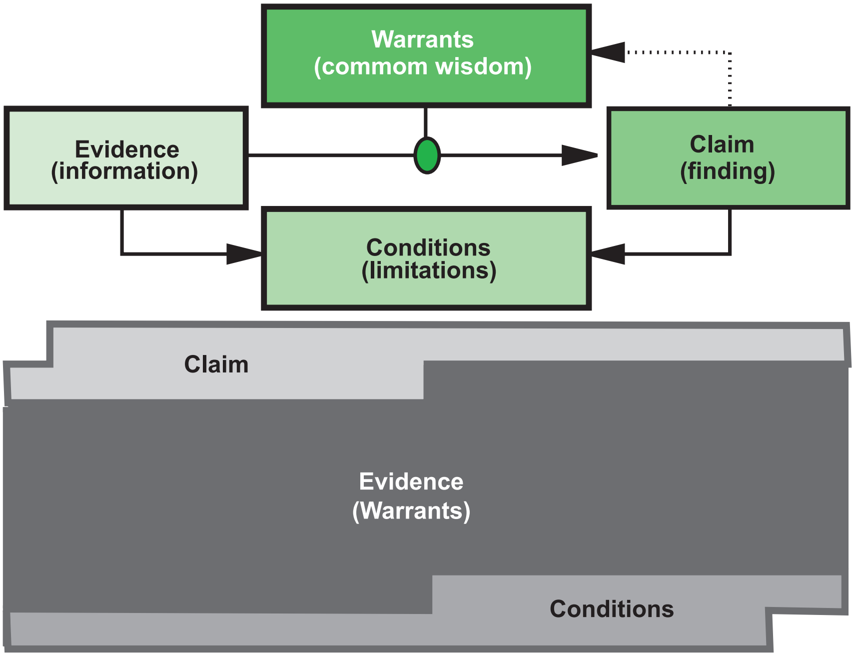


Figure 1. Constructing logical arguments as paragraphs requires claims, evidence and conditions

with a sentence that places **conditions or limitations** upon that claim (Figure 1). This structure is particularly important in the Introduction, Justification and Literature Review sections as well as in the Discussion section of scientific papers. Further information on establishing claims and supporting them with evidence is provided by Booth *et al.* (1995).

It is also important to maintain a consistent hierarchy of section, sub-section and sub-sub-section headings throughout the proposal. For example,

Sections centred, bold capitalised
Sub-sections bold, left justified
Sub-sub-sections bold italics, left justified is often used.

It is reasonable to structure the headings and sub-headings in the same way as those of a leading scientific journal within your area of interest.

Tables, figures and conceptual diagrams

Every proposal should include tables, figures and conceptual diagrams. These tools demonstrate an ability to compile and synthesise diverse sources of information and to prepare publication-quality material. **Conceptual diagrams** are best designed as graphic presentations of working hypotheses that identify likely mechanisms and how they might be elucidated. In many cases, quality graphics greatly reduce the need for lengthy text explanations when 'one picture is worth a thousand words'.

Additional documentation

The submitted proposal should be accompanied by a short covering letter, letters of institutional support, and a brief description of the investigators' qualifications. When possible, the covering letter should identify the title of the proposal and the specific donor programme to which the proposal is submitted. You must indicate your willingness to provide additional information to the donor in the future but the letter must not go into detailed explanation of the proposal's intention or methods.

Letters of institutional support

Such letters from a high-ranking member of the principal investigator's organisation are usually a prerequisite for processing an applicant's proposal. The letter of institutional support should mention the proposal by title and principal investigator by name and express commitment to accommodate the project. It is not usually necessary to include a full resume of all investigators with a proposal, but each investigator should prepare a 1 or 2-page profile of their qualification with emphasis on their educational and professional backgrounds, publications and previous experience in grant administration. You must not overwhelm a donor with enclosures or attachments accompanying a proposal because these could distract from the strengths of the proposal itself.

Know your donors

While it is not possible for most scientists to know every donor representative personally or to be assured that an individual proposal will appear attractive to a donor organisation, it is possible to target a proposal to a given donor. Many donor organisations maintain home pages on the Internet that describe their aims and programmes. Some post their Instructions to Authors and Application Forms over the Internet as well. You can gain an insight by examining the Acknowledgements sections of recent publications because the donor organisation and its specific programme that funded an investigation are usually identified there. Most donors 'specialise' in areas of food security, natural resource management, privatisation/liberalisation, forestry, environmental conservation and in specific commodities or agroecological zones and it is important to learn which topics are funded by which organisation. Such knowledge is gained through experience, as there is no single source and donor priorities change with time.

Emphasise substance, not superficial structure

Be aware that many donors rely on experienced technical reviewers to evaluate incoming proposals and that these reviewers are expected to comment on the feasibility, relevance and potential impacts of the proposed research. Some proposals highlight structure that is the administrative mechanisms through which a project is managed, rather than scientific substance and these proposals tend to be negatively reviewed. Avoid establishing 'management committees' for a project, the principal investigator should assume responsibility for completion of activities once the project is awarded. Also, be careful not to reflect top-down administrative and client attitudes in work plan diagrams, but instead emphasise interactions between research partners and stakeholders. Highlight the quality of your research experience rather than the size of your parent organisation. The proposal should reflect your stature as a developing scientist, not your ambitions to become a technocrat. Also, you must remember that donor representatives are by necessity generalists and that grant submitters will usually possess much greater knowledge of their individual subject area. Take care to explain the scientific approach and key measurements in understandable terms, and to rely upon the references you cite to represent extremely technical details.

Avoid disciplinary jargon and excessive abbreviation as this will be interpreted as an inability to communicate with the wider scientific community.

Grantee ethics

As a proposal writer you must be fully aware of the ethics surrounding proposal writing and research (see Booth *et al.*, 1995). All scientific ethics apply to grantsmanship, including the requirement to accurately cite and fully acknowledge the ideas and contribution of others, and not to misrepresent or obscure contradictory evidence. The following must be kept in mind throughout the proposal preparation and research process:

- It is ethical to submit the same or similar proposals to more than one donor at the same time, but unethical to accept funds from more than one donor for a single or similar research project. When one donor funds a research project, notify other donors who have been sent similar proposals that you no longer require funding for that particular research project, thank them for their consideration and mention that you look forward to sending additional proposals to them in the future.
- Grant contracts are legal documents and as the grantee you should feel legally and ethically bound to complete these contracts to the best of your abilities. Always comply with the contractual schedule for technical reports and financial statements in a timely manner. Remember that in these days of competitive research, entitlements and endowments are few and far between. Donor organisations have long institutional memories, and donor agencies willing to fund a particular topic in a particular area are not limitless. Do not feel trapped by your grant. Experience suggests that donor representatives are usually sympathetic to difficulties encountered and changing circumstances, so remember to keep channels of communication between yourself and donor representatives open.
- Avoid double reporting. Often different research projects are complementary to one another and may constitute components of an overall research question. Some donors encourage or readily accept 'piggy-backing', but others shy away from such arrangements. In general, donors do not approve of combining different funding sources into a general pool that does not distinguish the research activities that arise from each funding source. **Double reporting** results when investigators report all their research activities to all funding agencies, regardless of which agency actually funded each individual study. Remember that the members of the donor community are often in close communication with one another and will become extremely disappointed in you if you double-report your research accomplishments.
- Different donors have specific acknowledgement conditions with which you must familiarise yourself and comply. Some donors require that funding be acknowledged in all publications arising from your research. In general, always acknowledge donor assistance unless explicitly forbidden. Some acknowledgement conditions extend to specific wording. For example, one donor organisation discourages grantees from stating 'This research is supported by DONOR X' preferring 'this research was conducted through the financial assistance of DONOR X'.

If you fully acknowledge sources, cite contrary findings, recognise the limitations of your own findings, assert claims only as strongly as warranted and then strive for publication in the leading, peer-reviewed scientific journals, you will not only avoid moral dilemma but also accrue scientific credibility!

More on outputs and impacts

If you seek to improve a situation, its boundary conditions and underlying mechanisms must be defined using broadly applicable scientific procedures, but there is also a need to identify client groups, candidate solutions and the agents and pathways for change. Because of a paucity of demand for research products in the form of publication by those expert at translating findings into profitable technologies, you must reach further towards your potential clients and create research products that your clients can easily understand and test. Publication must remain a major currency of science, but not the only one and as a scientist in a developing country you must become more adept at shaping research products into other forms, particularly as pioneering technologies and pilot products. Good research can often be further synthesised into physical form, whether this be an invention, an extension package or a computer software model, and further testing of these products could in turn signal a new opportunity for research as candidate solutions. You must not merely position yourself within the research and development continuum where you find the most comfort, but you must be prepared to reach in either direction as circumstances dictate. The needs of the end clients are great but it is the intermediate agents for change, who are absent, unaware or indifferent, and you could help to bridge the gap between clients' difficulties and candidate solutions by striving to see your tested ideas formalised into wider practises.

The development of relevant research products begins with the earliest stages of research planning. Well designed research consists of a continuity of elements that start with clearly stated and substantiated objectives that in turn lead to key research questions and counterpart working hypotheses (Figure 2). Properly stated hypotheses identify treatment contrasts and necessary measurements which are in turn translated into experimental design and procedures. Too many proposals become truncated at this point by merely indicating statistical procedures and an intention to publish future findings, but failing to elaborate upon the immediate and end-clients of the research, which additional research products are anticipated, how these products will be delivered and what impacts may be expected from that delivery. Including the latter elements into a research proposal distinguishes that research as 'demand-driven' and 'impact-oriented' and indicates that you are aware that you must find new ways to formalise and popularise your research accomplishments.

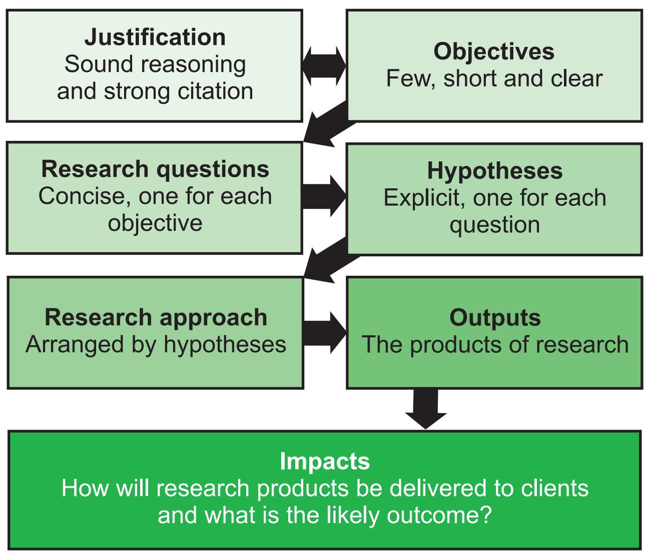


Figure 2. The confluence of proposal elements leads to research outputs and anticipated impacts

Diagnostic indicators of proposal and project viability

There are several indicators of the viability of a given proposal or project that are independent of scientific approach and technical details, but are related to the philosophy and openness of proposal preparation and project management. Healthy approaches to proposal preparation include true collaboration in its writing, internal review of a proposal prior to its submission, circulation and open discussion of reviewer's comments and careful selection and preparation of references. Similarly, there are several indicators of successful project management including scheduling of regular project meetings, assumption of research responsibilities by the principal investigator rather than operating through top-down delegation, and the clear assignment of tasks and funds to co-operators, including those from outside the principal investigator's organisation. The attitudes of the principal investigator and senior investigators are easily ascertained by reviewers and donor representatives from the proposal contents and short visits to their offices and laboratories.

As a student, your inclusion within a university-based proposal is usually an asset, but only if you are truly integrated into the research team. You should be provided with a copy of the complete proposal so that you may be in a position to better understand your role within a larger context. You should be aware of the funds available to you and be expected to account for these funds as part of your training. Project resources, particularly vehicles and computers must not be treated as personal property by the principal investigator and senior investigators, but should be made available to help all members of the team to complete their work. You must be assigned desk and laboratory work space so that you can work efficiently and be easily accessible to other team members. Your stipend must be sufficient to meet your living needs and must be paid in a timely fashion, otherwise it will be difficult for you to devote your full efforts to your thesis and project responsibilities. The opportunity to creatively contribute to a project and your treatment as a developing professional must both be reflected within any proposal with which you are involved, and again, these considerations are very easily detected by reviewers and donors.

The entire research team must strive for project accomplishments that reflect a healthy and productive research environment, particularly if the senior investigators seek further funds and a lasting relationship with their donors. The principal investigator must meet all project deadlines and goals and must not conveniently ignore, or attempt to redefine those goals at the end of a project. Late technical reporting and lax financial accounting reflect poorly upon a project and its investigators. Projects that lead to publication in leading journals will always be regarded in a favourable light that is not achieved by those that simply produce technical reports with limited circulation. Such other opportunities for research products as pilot products and pioneering technologies must not be overlooked. Senior investigators are better positioned to lead productive research teams when they assume direct research and writing responsibilities rather than operate exclusively through delegation and they must not demand co-authorship as a 'courtesy'. Project accomplishments and the manner in which they are obtained present an entry point for further donor relations and future project funding.

Indicators of healthy and weak approaches to proposal preparation and project management and outputs are shown in Table 1. If you find yourself working in a weak research environment you should to reconsider your position. If you are in agreement with the principles described in the healthy research environment you should take care that your intentions are clearly reflected in the proposals, project reports and research products of your team.

Table 1. Indicators of impacts on MSc programmes through grants

Indicator of positive impact	Indicator of no or negative impact
Proposal preparation	
Proposal is jointly prepared by principal investigator (PI), other investigators and, possibly, candidates for MSc scholarships	Proposal is prepared by PI in isolation, no inputs from other investigators, candidate students are not identified in the proposal
Proposal is internally reviewed by other FORUM grantees before submission	Proposal is sent to FORUM without the knowledge or comments of other local grantees
Reviewers' comments are circulated and revisions made by the project team	PI responds to reviewers' comments in isolation does not circulate reviews and revised proposal to others
Project activities and administration	
Regular project meetings held and attended by all investigators and students	Project meetings seldom held, PI assumes complete authority through delegation
PI pursues individual research interest as well as supervising students	PI delegates all research responsibilities to students or other investigators
Co-operating investigators assigned research and supervision responsibilities	Outside co-operation exists in name only to meet requirements in proposal guidelines
Co-operating investigators and students assigned research budgets	PI withholds budgetary information, preparing the financial statement in isolation
Student relations	
FORUM students group organised and meets regularly to discuss shared concerns	FORUM students discouraged or prevented from organising
Students provided with complete proposal and assigned some financial responsibilities	Proposal or sections of it withheld from students and other investigators
Students assigned designated desk and laboratory work space	Students lack designated work areas due to vaguely worded 'policies'
Students with ready access to project or student computer room established	Computers locked away from students for use by computers PI or other investigators and students require special permission for their access
Students regularly travel to field in project vehicle, vehicle available to other FORUM projects	Students must travel by public means, PIs use vehicle as personal transportation
Students work solely on research and thesis preparation, stipend sufficient for student's needs	Students distracted from thesis by outside employment
Project outputs	
Students submit thesis within 2 years then assist PI in preparing renewal	Students leave campus after conducting field work, completing thesis elsewhere
Project findings results in publication and completed MSc thesis	Research results only appear in submission of MSc thesis
PI serves as an active co-author in the preparation of scientific papers	PI's name appears as co-author as a courtesy claiming time not available for writing
Thesis and papers contain numerous, current citations from leading, peer-reviewed journals	Thesis dominated by grey, in-country literature or unpublished sources

Proposal refinement

If you submit a first version as a 'finished' proposal you are well advised to not consider this the end of your effort, but rather an important stage toward reaching a desired opportunity. The proposal award process may be viewed as a series of interactions between donors and applicants with many important steps required subsequent to submission (Figure 3). A donor's call for submission is met by applicants and then acknowledged by donors who will often request additional documentation. A request for additional documentation should be regarded as a sign of interest by the donor that you should satisfy in a timely manner.

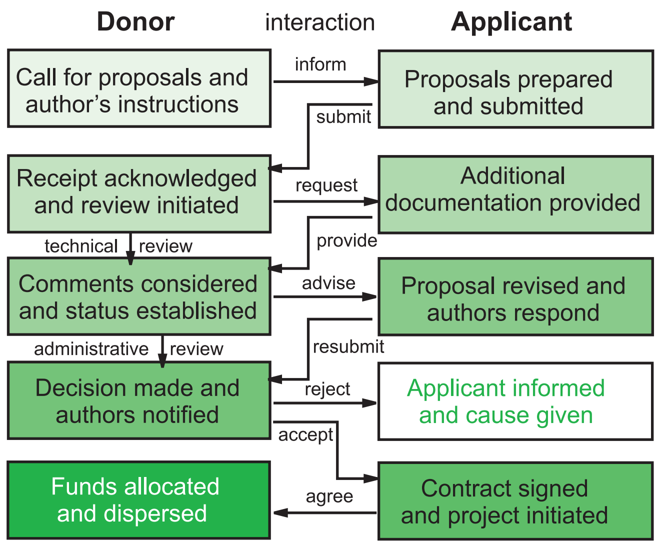


Figure 3. Interactions between the donor and applicants in the proposal submission, evaluation and award process

Proposals are often sent by donors for technical review. This may require several weeks or a few months. Reviewers' comments are then returned to donors and forwarded to applicants. These comments almost always require moderate to major proposal revision and you are well advised to accept and respond to reviews in a constructive manner, but not necessarily to comply with every one of the reviewer's suggestions. In some cases, proposal refinement occurs through several review/revision steps and you should not become discouraged unless it becomes clear that irreconcilable differences in scientific viewpoint exist between yourself and the reviewers. Even then, a proposal could be rescued by preparing a letter to the donor that documents and describes the merits and potential weaknesses of each perspective.

Ultimately, your proposal will be either accepted or rejected (Figure 3). If rejected, you should be told the reasons for this decision, and you should be aware that proposals rejected by one donor based upon their present criteria and priorities may be more acceptable to another donor. If accepted, you will be notified in a letter that is often accompanied by a contract that must be signed and returned to the donor before funds are released. If the donor's acceptance letter requires acknowledgement and requests banking details, you must reply as immediately as possible.

Conclusion

If you are preparing proposals that address agricultural development and natural resource management in Africa you must remember that you are in a crucial and privileged position. Some final advice on proposal preparation and interactions with donor representatives and reviewers is provided in Table 2.

Note particularly the caution against presenting minor and superficial changes as suggested major revision because this is particularly likely to anger the ire of reviewers and

Table 2. Hints on submissions of proposals to donors

Applicants should	Applicants should not
• Inform themselves of donor programme objectives and author's instructions	• Submit proposals without fellow cooperator's knowledge
• Respond to requests for additional information promptly	• Send frequent, unsolicited enquiries about proposal progress
• Respond to reviewer's comments in a constructive, interactive manner	• Respond to reviewer's comments in a dismissive or defensive manner
• Revise and resubmit proposal in a timely manner	• Present superficial changes as major revisions

alienate potential donors. Keep in mind that many donor organisations are extremely concerned over the continent's growing food shortage and depletion of its natural resources and are looking for creative ideas and innovative approaches that may serve to reduce or reverse these ominous trends. The ultimate clients of many agricultural scientists are the rural poor who are relying on society to help them solve their production problems and to obtain a better standard of living. If these societal responsibilities and concerns are reflected within your proposal in a scholarly, but relevant manner, most donors will consider it has promise.

Resource material and references

Appendix 4. Preparing and Refining a Research Proposal. Paul L. Woomer. PowerPoint on CD.

Appendix 5. Stapleton, P., Youdeowei, A., Mukanyange, J. and van Houten, H. 1995. *Scientific Writing for Agricultural Research Scientists*. WARDA/CTA, Ede, The Netherlands. On CD.

Appendix 6. Publication as an Output of Science. Adipala Ekwamu. PowerPoint on CD.

Appendix 7. The Art and Ups and Downs of Scientific Publication. Adipala Ekwamu. PowerPoint on CD.

Appendix 8. Presentations and Style – Tips on Photography and Writing. Eric McGaw. On CD.

Baugh, L.S. 1995. *How to Write Term Papers and Reports*. Future Vision Multimedia. NTC Publishing, Chicago, USA. 259 pp.

Booth, W.C., Colomb, G.G. and Williams, J.M. 1995. *The Craft of Research*. University of Chicago Press, Chicago, USA. 294 pp.

Geever, J.C. and McNeill, P. 1993. *The Foundation Center's Guide to Proposal Writing*. The Foundation Center, New York, USA. 202 pp.

Patel, B.K. and Woomer, P.L. 2000. Strengthening Agricultural Education in Africa: The Approach of the Forum for Agricultural Resource Husbandry. *Journal of Sustainable Agriculture* 16(3): 53–74.

Visit the Forum on Agricultural Resource Husbandry website at: http://www.rockforum.org

Part 4
Tools and data sources

You may be surprised that up to this point in the book there has not been much about data collection and statistical analysis. That changes now! But you will still find little in the way of mathematics or technical details. There are two reasons. First, many of the *ideas* you need to design and analyse good studies can be explained and understood without using mathematics. Secondly, there are many books around that describe the mathematics, and many of the courses in 'research methods', or 'statistics' that you will have followed will have used a mathematical, rather than an intuitive, approach. We want to provide an alternative.

These chapters can only be introductions to important ideas and methods. Maybe they will be all you need. It is more likely that they will raise all sorts of questions that are important in your research, and prompt you to seek out further understanding. They may even help you make sense of that statistics course you took and hated so!

The more technical aspects of a research project are important, and sadly many students have failed, or had to redo parts of their project, through failing to understand them early enough. If the material here raises any questions or uncertainties then you should get help. Biometricians and statisticians are experts in this stuff, so consult them! And any successful researcher must also have a sound grasp, and should be able to help you.

The topics covered in Part 4 are only a selection from those that could have been included. We have used two criteria to select them:

1. The topic is essential to most projects, yet commonly misunderstood. (Included here are the chapters on design of experiments, planning surveys, measurement, managing and analysing data).
2. The topic is important but often omitted from research methods courses. (We have included chapters on finding and using secondary and spatial data, and modelling).

The chapters are roughly arranged in the logical order in which they might occur in a project. But please don't wait until the end of your project to look at later chapters! The design of the study depends on how you will analyse it, so you must be aware of the later steps during early stages.

Richard Coe


TOOLS and DATA SOURCES
SURVEY
NIKON
ALEYA

4.1 Using secondary data sources

Jayne Stack

- **Primary data are observations you collect yourself to meet a specific research objective**
- **Secondary data are observations collected by others for other purposes**
- **Every research study should start with a review of relevant secondary data before planning collection of primary data**
- **Secondary data can help define the scope and extent of a problem**
- **The value of secondary data may be limited by availability, accuracy, timeliness and the definitions used**
- **There are very many sources of secondary data. International data are increasingly available free on the Internet**
- **Techniques are available to help you use secondary data effectively**

Overview

At the broadest level information sources that are available to you can be classified as primary or secondary. Primary data are those which you collect yourself for a specific research purpose. Secondary data is information that has been previously collected by individuals or agencies, usually for purposes other than your own particular research study. Secondary data may be qualitative or quantitative. Qualitative data is generally thought of as subjective, verbal and descriptive and includes information captured by a wide range of media. It includes photographs and maps, case studies, reported happenings, in-place observation, and tape or video recordings of conversations and/or activities. In contrast, quantitative data is generally numerical data, collected using some form of measurement and amenable to mathematical analysis. Quantitative data includes information captured by direct measurement through field observation (rainfall, temperature, crops yields, price-monitoring) and by direct measurement within structured questioning (household income and expenditure data collected by governments as part of national households surveys). Although the nature of secondary data influences the selection of tools that can be used to manage or interrogate a data set it is easy to exaggerate the differences. Both types of information are collated, sifted and organised into some sort of meaningful form by looking for connections or relationships in the data. The guidelines for reviewing secondary material outlined in this chapter can be adapted to a wide range of secondary sources compiled by other people.

The aims of this chapter are to:

- Show that no study should begin without first reviewing existing knowledge, regardless of the data-collection techniques to be followed later
- Identify the main sources of secondary information
- Point out and illustrate the need to critically examine the concepts and definitions used in secondary information
- Illustrate some of the conceptual and analytical tools that can be used to 'interrogate' data from secondary sources
- Secondary data analysis is not just a first-stage activity but can and should contribute to every stage of the research cycle.

Why is secondary data useful?

All too often inadequate attention is given to reviewing existing knowledge before embarking on primary data collection. No research study should be undertaken and without a prior search of secondary sources (also termed desk research). There are several grounds that give us confidence in making such a bold statement. (The following material was adapted from Crawford and Wycoff, 1990):

- Secondary data helps you to: define a research problem, formulate research questions and hypotheses, and select a research design. The assembly and analysis of secondary data almost invariably makes an important contribution to the research process. A review of existing knowledge will improve your understanding of the research problem, including the key issues, core concepts, and on-going debates. It will reveal approaches to data collection (e.g., useful conceptual models, variables for concepts of interest, appropriate analysis techniques) that may improve or complement your own initial research design. In sifting purposefully through secondary data, you may find something else that sends you exploring new regions or ideas you may not even have thought of before. And, you might find evidence that will actually change the shape of your ideas.
- Secondary data may be sufficient to answer the research question. Occasionally you may find the available data are so adequate that primary data collection is unnecessary. If useful secondary data are available, they can be used to substitute for primary data collection at any stage during your research. It is not always necessary for you to collect all the information required for the analysis yourself. For example, daily rainfall records for the last 10 years obtained from the Meteorological Office allow you to draw conclusions about the adequacy of the growing season and the problem of dry spells, or agricultural data from a national sample survey can provide good information on the major characteristics of a farming system.
- Data costs are substantially lower for secondary data than for primary data. A thorough review of secondary sources can be completed at a fraction of the cost and time it takes to complete even a modest primary data collection exercise. Finding a 'ready made' solution in existing sources is unlikely, but even partial solutions help primary data collection needs, and therefore save time and money. For example, the current livestock situation in a country in terms of stocking densities, grazing pressure, herd structure, and management practices could be studied using a combination of secondary livestock data from the Ministry of Agriculture, the veterinary services, and reports of past research studies.
- Secondary sources of information can yield more accurate data than that obtained through primary data. This is not always true, but when a government or international agency has undertaken a large-scale survey or even a census, their results are likely to be far more accurate than your own surveys when these are based on relatively small sample sizes. For example, a national income and expenditure sample survey is likely to yield more accurate results than an income and expenditure study of 200 sample households in a single area. However, it should be remembered that all secondary data was once someone else's primary data. Some people who work with official statistics wrongly conclude that their own analysis is more objective than analyses of primary data, which is 'soft' data.
- Secondary sources help define the population. They can be extremely useful both in defining the population and in structuring the sample you wish to take. For instance, government statistics on a country's agriculture will help to stratify a sample and, once you have calculated your sample statistics, the stratified sample can be used to project those estimates from the sample to the population.

- Secondary data can be used to make comparisons. Within and between nations and societies, comparisons can enlarge the scope for generalisations and insights. Global and regional data sets (e.g., those of the Food and Agriculture Organization of the United Nations (FAO), World Resources Institute (WRI), or the World Bank) are a valuable source of secondary data for between-country comparisons on a vast range of topics including poverty issues, food security, trade patterns, growth rates, and technical change. Within-country comparisons can be made using national data sets disaggregated by administrative or natural regions.
- The availability of secondary data over time enables the employment of a longitudinal research design. One can find baseline measurements in studies made in the past and locate similar data collected more recently. With an increasing emphasis on understanding patterns of change, the use of secondary sources can also be critical to single point surveys, which lack a time dimension.
- Secondary data can be used to increase the credibility of research findings obtained from primary data. The comparative use of other research together with a comparison of data collected during your study with official statistics on the same topic can be very valuable when you reach the analysis stage. **Research results are more credible when supported by other studies.**

Limitations of secondary data sources

The following material was adapted from Crawford and Wycoff (1990). Whilst the benefits of secondary information can be considerable, like any other data collection method, the validity of the data must be carefully assessed. The main problems include:

- **Access.** Once potential sources of useful secondary data have been located there may be difficulties in accessing variables of interest if the data are not in the public domain or are unpublished. If this is the case, you will need to approach the organisation or individual holding the data to seek permission to use the information it contains. Unpublished data sets residing with government or non-governmental institutions are usually made available once permission has been sought in writing, clearly explaining the purpose for which the data will be used and the user's willingness to adhere to specific conditions of use. A supporting letter from the institution sponsoring your research may be helpful. Sometimes the original investigator will not make data available, particularly if they are still using it to pursue their own research. This can be frustrating, especially if data analysis is taking a long time. Sometimes researchers may be willing to provide some data in aggregate form ahead of publication. This can be used providing its source is acknowledged.
- **Relevance.** There is an inevitable gap between primary data collected personally by an investigator with specific research questions and hypotheses in mind and data collected by others for different purposes. It often happens that there is an abundance of secondary data, but much of it is not of direct relevance to your specific research problem. During the early stages of examining secondary data, you explore and gather anything and everything that you think might be of interest or use. However, as you begin to organise the material you have collected to support one or another of your ideas some secondary data will not be relevant. Every bit of evidence that you include must justify its existence; it can only do so in support of an idea. Use the list above to ask yourself on what grounds the secondary data you have collected is useful.
- **Reliability.** The reliability of secondary sources may vary substantially and it is difficult to ascertain if insufficient information is available about how the data were collected and

potential sources of bias and errors. It helps considerably if you are able to speak to individuals involved in the collection of the data to gain some guidance on the level of its accuracy and limitations.

- **Definitions.** A common problem in using secondary data is how various terms were defined by those responsible for its preparation. Terms such as family size, income, credit, farm size, output sales, and price need very careful handling. For example, a family size may refer to only the nuclear family or include the extended family. In census data a household is often a group of people who stayed the census night in the dwelling unit, irrespective of whether they are part of the nuclear family or not. Income data often exclude the value of own-produced goods. Credit and sales statistics often ignore transactions that pass through the informal sector. Even apparently simple terms like the year for which the data apply may need care in interpretation. For instance, in Zimbabwe, the marketing year 2002/2003 refers to the period 1 April 2002–31 March 2003. Any crop sales data recorded against 2002/2003 refers to sales from the 2002 harvest. Sales from the 2003 harvest are recorded under the 2003/2004 marketing year! Special care in interpreting definitions and years is necessary in combining secondary data from several sources to produce a derived data set.
- **Timescale.** Most secondary data has been collected in the past so it may be out-of-date when you want to use it. If the data source includes estimates of growth rates this information may be used to extrapolate figures for subsequent years. For example, population censuses usually include an estimate of population growth that can be used to estimate inter-census population data.
- **Source bias.** You should be aware of vested interests when you consult secondary sources. The objectivity of officials may be affected when it comes to reporting situations for which they themselves are partly responsible. Similarly respondents may provide biased information depending on their perceptions of the purpose of data collection (e.g., planning drought relief, forced destocking,). Further, official economic data may be a very inaccurate source of statistics in situations where the informal economy and/or black market account for a significant share of economic transactions.

Sources of secondary data

Secondary data sources can be divided into two categories, internal and external.

Internal information sources

All organisations collect a range of information during their daily operations. For instance, a marketing board records deliveries of crops, payments made to farmers, stocks, and orders from buyers that are dispatched, and invoices sent out. Such information may be available in a more disaggregated form than is reported in the organisation's internal reports. Much of this internal information is of potential use to researchers, but surprisingly little of it is actually used.

You may be unaware of some of the data collected and the regular reports submitted by the organisation for which you work. Begin your secondary data search with an internal audit. Familiarise yourself with available internal information whether you are a researcher in a government body, non-governmental organisation (NGO), or a business organisation.

External information sources

The primary sources of official and semi-official statistics are:

- Government statistics. These may include population censuses, national income data, agricultural statistics, poverty surveys, trade data, cost of living surveys, nutritional surveys, the results of commissions of enquiry into particular issues (e.g., land tenure) and possibly data on market prices.

Secondary sources can include:

- Marketing boards, which are likely to have information on quantities purchased of different commodities, imports and exports, buying and selling prices, and stocks
- Extension organisations who will have crop area and production estimates for various crops and probably farm budget data for different enterprises
- Agricultural research institutes that are an important source of information on such agronomic issues as soil fertility studies, crop and livestock breeding programmes and technology
- Veterinary departments who may have data on livestock numbers and disease control measures, e.g., dip tank records
- Hospitals and clinics might have data on incidence of malnutrition, particular diseases and causes of death
- Local administration offices often have lists of households which could be useful in the construction of sampling frames. They might also provide information on project activities in the district, e.g., active NGOs, or registered cooperatives
- Archives are a useful source of information to help you understand patterns of change
- International organisations may have country studies available at their local information centres or offices
- Websites. With the rapid development of information technology and computerised databases, the scope for you to carry out a search of secondary sources and to use secondary data sets compiled by other organisations and posted on websites, has increased dramatically.

The following is a selection of key websites providing access to statistical data of particular interest to African agricultural, environmental and rural development researchers.

Main collections of wide-ranging development statistics

- World Bank website offers on-line access to country statistics and prepared tables for 207 countries and 18 country groups. 54 time series indicators on people, economy, environment, spanning 5 years are available and you can choose several ways of displaying the data: index, percentage change and graphs and can export the results to other documents. http://www.worldbank.org/data/
- UNDP Human Development Report and Indicators provides statistics on human development indicators including poverty, health, education, food security, employment, urban development, population, environmental degradation and national income accounts. http://www.hdr.undp.org/
- International Development Statistics (OECD/DAC) on-line database covering debt and aid. http://www.oecd.org/department/0,2668.en_2649_34447_1_1_1_1,00.html
- United Nations Statistical Organization (UNSO) has a wide range of statistical databases on-line on trade, national account, demography, population, gender, industry, energy, environment, human settlements and disability. http://www.un.org/Depts/unsd/stadiv.htm
- World Health Organization (WHO) website for health related statistical information. Also provides supporting materials including analysis software. http://www.who.int/whosis/

- Millennium Development Goal Indicators has 48 social and economic indicators for 1985–2000 used to monitor the implementation of the goals and targets of the United Nations Millennium Declaration. http://milleniumindicators.un.org/

Subject-focused development statistics websites

There are numerous subject-focused websites that provide data on specific topics (see http://www.eldis.org/statistics/index.htm). Those of particular interest include:

- United States Department of Agriculture (USDA) provides global and US agriculture data. http://www.ers.usda.gov
- Food and Nutrition and Crop Forecast. 'Food Outlook' is a report produced by FAO five times a year. It provides a global perspective on the production, stocks and trade of cereals and other basic commodities. Food Outlook can be downloaded from http://www.fao.org/giews/english/fo/fotoc.htm
- A helpful guide to other sources of food security statistics is the ELDIS Food Security Resource Guide at http://www.eldis.org/food/statistics.htm.
- HIV/AIDS. A resource guide for information and data is available at http://www.eldis.org/hivaids/aidsstats.htm.

Country-focused development statistics websites

A good starting point is the ELDIS country profile service. http://www:eldis.org/statistics/

Non-official sources

- Consultants reports (which may be gathering dust on the shelves of the body sponsoring the research!)
- Records of NGO activities including drought relief and supplementary feedings schemes
- Baseline surveys and project documents.

As you can see, secondary information can come from a bewilderingly large number of sources. Perhaps the most efficient and effective way to begin is to talk to people. Find the authorities in the field; search out the researchers working in your areas of interest. Conversations with them can get you further faster than almost any other search method. Researchers outside your own country can usually be contacted by e-mail and many are happy to forward copies of their own publications. Develop a network of contacts in key positions and cultivate them over time. Such contacts are particularly useful sources of semi-official and unpublished reports from research institutions and universities. In addition, experienced researchers have usually built up their own list of favourite websites that provide material on key research themes in development.

Recording details of secondary data material

Mountains of information can grow alarmingly quickly and it is imperative that you keep a record of the material that you have consulted in the course of your research so that you can acknowledge all the sources. The most important aspect of collating secondary data is to establish a 'trail' so that you or anyone who wishes to check your sources can easily find them again. Note the source of every piece of information you find useful. There is no single universally accepted format for referencing but a common order for the required information is:

- Author(s)
- Date of publication
- Title of the work cited

- Publisher
- Place of publication.

Evaluating secondary information

Information obtained from secondary sources is not equally reliable or equally useful. As mentioned earlier, just because data is published it does not mean that it is accurate. Just because data is available it does not mean it is useful for your particular study. If you are using secondary data, be it quantitative or qualitative, you should routinely ask the following questions, according to Dillon *et al.* (1990).

- What was the purpose of the study? Data are usually collected for some specific purpose, that ultimately determines the study variables of main importance, the reporting domains and the degree of precision
- Who collected the information? Because you are not collecting the data yourself, a natural question concerns the expertise and credibility of the source. Find out how the data were obtained and what sort of training and expertise is present in the organisation providing the data
- What information was collected? It is important to check this exactly. For example, in a study on household income were all income sources included, or only cash income?
- When was the information collected? The time the data were collected plays a role in its interpretation. For example, information on the nature of the season should be examined when interpreting household information on incomes or food security
- Which geographical area does the data represent? Not all data is collected for the same spatial area. Administrative boundaries often differ from geographical boundaries and may also vary depending on the organisation collecting the data. Boundaries also change over time as new administrative districts are formed by splitting or amalgamating existing units. The boundary issue is generally most problematic if data from different secondary sources are being combined and/or information from various points in time is being compared.
- How was the information obtained? The method used to collect data is an essential ingredient in evaluating the quality of secondary information: for example, the size and nature of the target sample, whether it is based on observation or recall, how it is collected (key informant interviews, household surveys, focus group, satellite imagery, etc.) and if surveys were from single or multiple visits. Some methods or combinations of methods are better than others at providing specific types information. Familiarise yourself with the alternative ways of obtaining information so that you can make an informed assessment of secondary data quality.
- Is the information consistent with other information? A valuable principal in data collection is that of triangulation where information is collected from multiple sources. If similar conclusions can be drawn from different sources of data this lends credibility to the findings. If differences exist, you should try to find out why, and which source is more reliable. The consistency of information is frequently a problem with agricultural production statistics from different sources.

Working with secondary data

Research studies use secondary data in several ways, the following are three broad types:

1. Research which uses aggregated secondary data to inform a study that will generate its own primary data as a major source of information.

For example:
- Study of consumption and marketing decisions of smallholders where the major source of information is a household sample survey, but where secondary data on grain production and marketed surplus by region are combined with official population data to examine past trends in agricultural production and marketed output. Here the analysis of secondary data provides a context for the analysis of the primary data.
- Investigation into the feasibility of edible insect farming using an experimental farm, where secondary information on artificial feeding is used to identify alternative feeding methods for field trials.

2. Research which uses aggregated secondary data as a major source of information, when interpreting this information. For example:
   - International comparison of various development indicators using a World Bank's global data set.
   - Regional human poverty comparisons made by the United Nations Development Programme (UNDP) for Zimbabwe using a poverty assessment study survey undertaken the Ministry of Public Service, Labour and Social Welfare and other secondary data sets (UNDP, 1998).
3. Research which uses disaggregated secondary data, perhaps in raw form, as a major source of information, with a new analysis of the same data. For example:
   - Modelling agricultural supply response using a data set derived from secondary data found in official statistics
   - Construction of a food balance sheet using official statistics
   - Lenin's' famous analysis of peasant differentiation using Zemstov house-to-house census data as his major source of data (Lenin, 1961).

The conceptual and analytical tools used to interrogate secondary data will vary depending on the role that secondary data play in the study. For instance, if secondary data are the major source of data for your research task, the analytical process, (specification and estimation) is likely to be a central component of your thesis. On the other hand, if you are assembling secondary data to improve your understanding of the socio-economic conditions in a field study area you are more likely to use simple descriptive statistics to highlight important trends and characteristics.

Regardless of the way you intend to use secondary data some general comments can be made about methods of interrogating it.

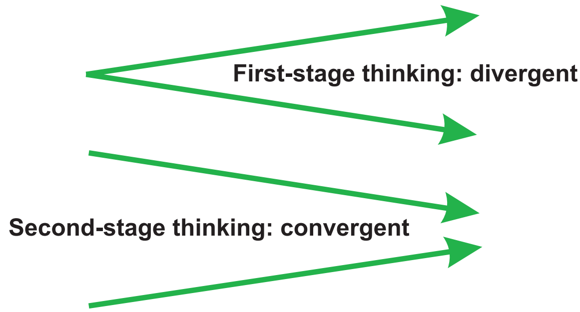
General

Research, like any other types of thinking, can be thought of as involving two stages. [The distinction between first stage and second stage thinking was first brought to my attention in a highly recommended course called 'Writing for effective change' distributed by an NGO called Fahamu, Learning for Change (Fahamu)]:
- First-stage thinking: exploration, discovery, generating ideas
- Second-stage thinking: collating, sifting, organising the ideas into a robust structure

First-stage thinking. Sometimes called 'divergent' or 'radiant' thinking; during this stage, you explore and gather anything and everything that you think might be of interest or use to your study.

Second-stage thinking. By contrast this is sometimes called 'convergent' or 'focused' thinking. It organises the material you have collected to support one or another of your ideas.

We tend to be much better at second-stage thinking than at first-stage thinking. So much so that we often fail to see first-stage thinking as thinking at all, but we ignore it at our peril. No amount of excellent second-stage thinking can compensate for poor or inadequate ideas. You must spend time generating ideas from secondary information **before** trying to assemble them into a structure.



Two techniques can help you:

- Mindmaps are powerful devices in first-stage thinking, they will help you gather and initially sort ideas
- Grouping and summarising is a second-stage thinking technique that helps you to organise your ideas.

Mindmapping

Mindmapping has been around for a long time, but the person who has done most to explain it and make it popular is Tony Buzan (1993). Mindmapping exploits our mind's extraordinary ability to create meaningful connections between ideas. Mindmapping helps us to see – or make – connections in our thinking, increasing our creativity and making thinking more efficient.

Brainstorming is the first step in mind mapping. Figures 1–3 show how a mindmap was developed to think about how to improve feeding systems using traditional practices (The mindmap example is adapted from 'Writing for effective change' distributed by an NGO called Fahamu, Learning for Change). Begin by writing the main research question or concept in a circle in the centre of a page. Then, jot down any ideas that come to mind when you think of this concept. (Figure 1). As you think of each new idea, new branches are created from the central balloon and the idea is written along the line (Figure 2). The next step is 'free associating' on each idea to build a verbal map of words or images that are connected to it. Sub-sets of

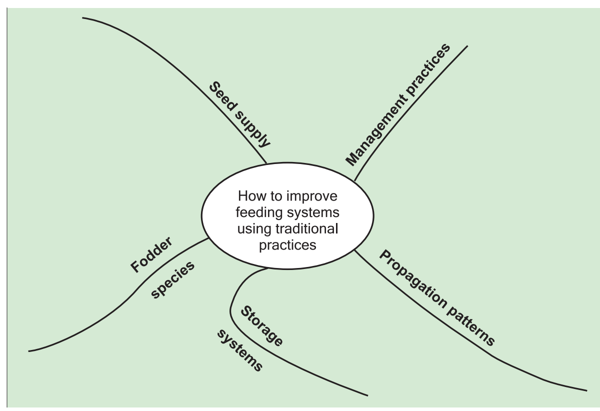


Figure 1. Example mindmap (4)

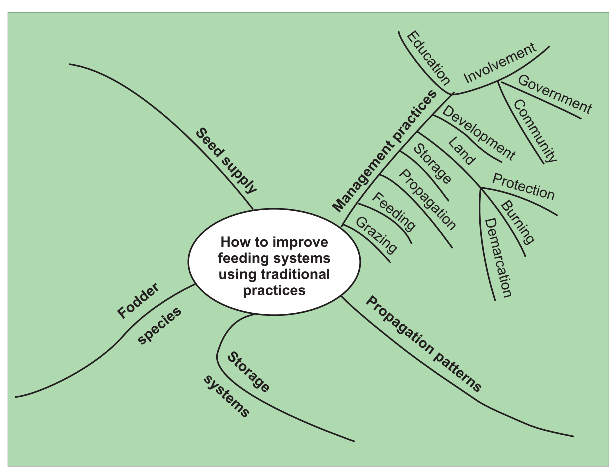


Figure 2. Example mindmap (5)

ideas are drawn as twigs from the appropriate branch line (Figure 3). Gradually a verbal mindmap tree of associations is built up. The final stage is to introduce hierarchies and categories to order or structure your mindmap. In the final mindmap you could use colour to emphasise hierarchies of ideas. The five main ideas for improving feeding systems radiating from the central image are shown in white whilst the sub-sets of ideas radiating from each of these ideas are black (Figure 3).

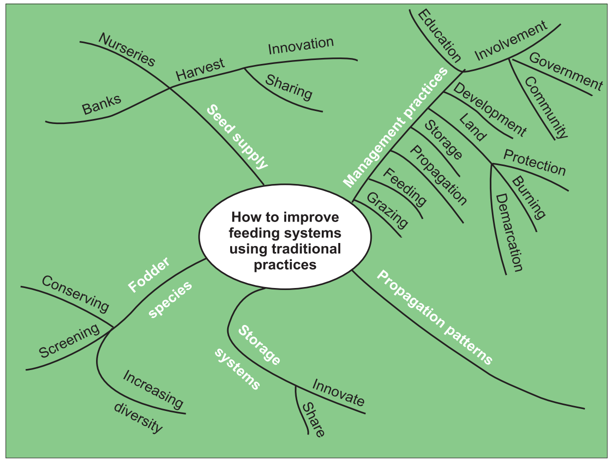


Figure 3. Example mindmap (final)

Grouping, summarising and organising ideas

It is unlikely that secondary data can be presented at the same level of detail in which it was previously collected. In other words you will not be able to report all the secondary material you have reviewed in its raw form and will need to summarise and present your data in a way that reveals patterns and trends in the data set. In order to summarise and present data it must be organised. There are various ways to do this. Some of the most common techniques that enable you to interrogate both qualitative and quantitative secondary data include:

Selecting categories in which the data can be summarised

A simple but effective way of revealing patterns in secondary information is to reorganise the information into new categories. The categories selected will depend on what is relevant to the topic being considered but if you are investigating food security you could use a data set of district-level information on estimates of per capita food availability to identify different categories of districts on the basis of their potential vulnerability to food insecurity. A study concerned with trade could rework information on the value of exports and imports between different countries by country groups to show the relative importance of different groups of trading partners. Putting available information into a tabular format is another way to organise either qualitative or quantitative information. Information from earlier studies was used in one study to compose a table showing different types of natural resource-access systems [individual, regulated common property, and unregulated common property (open access)] cross-tabulated with information about who controls access, who harvests, who benefits, who is included, and management implications. Organising the information in a tabular format highlighted the differences and similarities between each system more clearly than if the same information was just presented as paragraphs of written text (see Table 1).

Production of derived (secondary) data – new data sets

Very often the use of simple descriptive tools such as percentages, means, indices, or rates of growth can highlight patterns and trends in a data set that are not obvious in the original format of the data. For example, if crop production and sales data are available for two different farming sectors (smallholder and large-scale commercial) calculating the percentage share of each sector in total output and sales is a useful way of examining the relative

Table 1. Categorising and characterising different mopane woodland access systems in southern Africa

Land tenure institution	Who controls and how?	Who harvests?	Who benefits?	Who excluded?	Management implications
Individual	Individual	Individual	Individual	All others	Every owner has to exert effort to protect Cost of protection low if resource close to residence
Regulated common property	Rural Councils / Community resource management groups	Community members Licensed outsiders Community members Outsiders Licensed outsiders	Harvesting members Non-harvesting members (from licenses) Licensed harvesters	Unlicensed harvesters	Organised collective action to manage and protect resource. Potential economies of scale in protection activities. Transaction cost of formulating management rules and enforcing them may be too high for resource to be effectively regulated
Unregulated common property (open access)	Traditional control mechanisms		All harvesters	No one	If traditional regulations on extraction of resources break down, tragedy of commons results in overexploitation and deterioration over time
Centralised management of common property (e.g., State land)	Forestry Commission		Licensed harvesters State (from licenses)	Unlicensed harvesters	Organised centralised management system Potential economies of scale but cost of protection high for minor forest products Weaknesses in management may result in resources being ineffectively regulated

importance of each sector. If information on the volume, value and prices of exports of a particular commodity are available over a period of time, then calculating instability indices for each variable will demonstrate the level of export earnings instability and the extent this is due to either export price instability or instability in quantity exported.

Combining secondary data sources to form a derived secondary data set

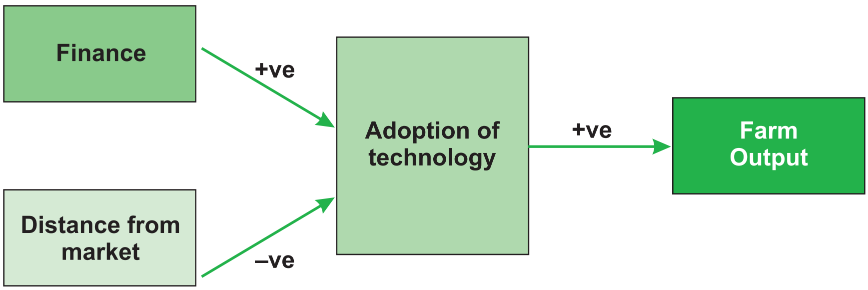
Combining secondary data sources may create a new data set that is more informative than each separate data set. A common calculation that illustrates this technique is the use of population data to express data sets of such variables as production, income, or cultivatable land in per capita terms.

Conceptual models and diagramming

Diagrams are very powerful tools for organising qualitative data. At its very simplest this could be diagramming a hypothesis about the main factors affecting a variable of interest using descriptive information from earlier studies (see **Chapter 4.3**).

Diagramming hypotheses (Dixon *et al.* 1995)

For example, in the following diagram, two concepts, finance and distance from market are hypothesised to be related as independent concepts to the dependent concept, adoption of technology. One of the independent concepts is seen to be positively related and the other negatively related to the dependent concept. Technology adoption is in turn hypothesised to positively affect farm output. Diagramming hypotheses promotes clear thinking and it is a useful way to summarise information from earlier studies. Use secondary data to diagram what you plan to study and even beyond the immediate research issue to show where your research fits in to the larger frame of reference.



However, diagrams can also be a useful way of conceptualising links and feedbacks within a system. For example, the livelihoods framework illustrated in Figure 4 is useful for thinking through livelihood circumstances of individuals, households, villages, and even communities and districts. The limitations of any such 2-dimensional representation of a process as complex as livelihood formation are recognised from the outset. The purpose of such a diagram is to organise ideas into manageable categories and identify the main components (assets, mediating processes, activities) and the critical links and dynamic processes between them (Ellis, 2000).

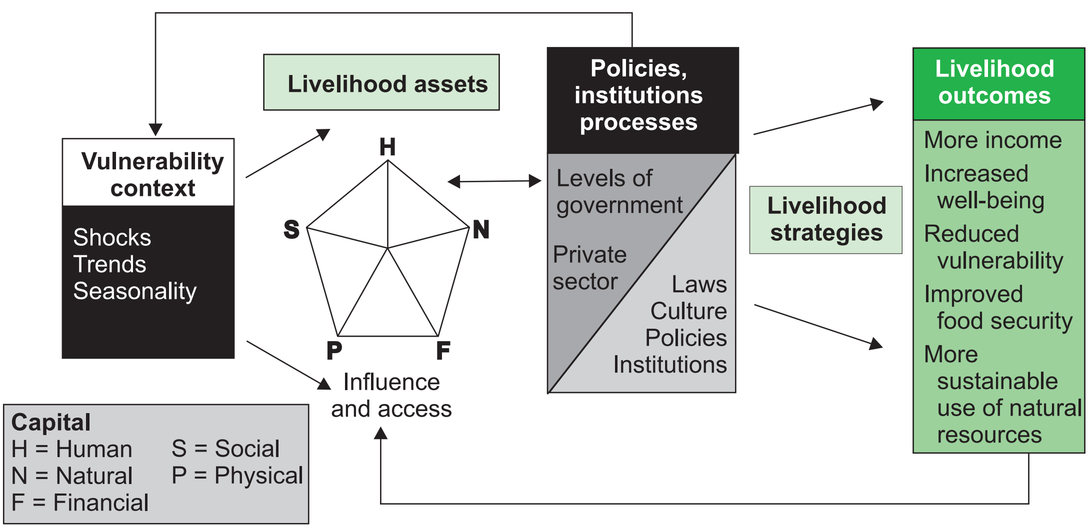


Figure 4. Sustainable livelihoods analytical framework

Analytical frameworks

Analytical frameworks are useful tools that enable you to concentrate on the broader picture. For example, food policy analysts may compute a food balance sheet using secondary information to examine food availability and identify key characteristics of domestic food consumption. Economic statisticians use accounting frameworks to prepare a country's national accounts and balance of payments based on secondary data.

Conclusions

A lot can be learned from secondary data and you should be prepared to explore various alternative ways of interrogating available information. Data sources should always be acknowledged and some guidance provided on the reliability and limitations of data used. In practice the collection and interrogation of secondary data is not just a first-stage activity but is something that can and should contribute to every stage of the research cycle. As noted in the opening section, secondary data can assist in designing a sampling frame, and in identifying a potentially useful method of analysis or appropriate conceptual framework. Secondary information provides a context for the analysis of primary data. The comparative use of secondary data can be especially valuable at the analysis stage and a good researcher will highlight areas of contrast and similarity between their own data and research findings of earlier studies on similar topics. Whilst findings gain more credibility if they are supported by a number of other studies, you should not be afraid to indicate where findings are different.

Resource material and references

Buzan, T. 1993. *The Mindmap Book*. BBC Books, London, UK.

Crawford, I.M. and Wycoff, J.J. 1990. *Market Research Teaching Notes*. FAO Report no. GCP/RAF/238/JPN (December). pp. 15–16; 16–17. Food and Agriculture Organization of the United Nations, Rome, Italy.

Dillon, W.R., Madden, T.J. and Firtle, N.H. 1990. *Marketing Research in a Marketing Environment*. pp. 98–100. Irwin, Homewood, Illinois, USA.

Dixon, B.R., Bouma, G.D. and Atkinson, J. 1995. *A Handbook of Social Science Research*. Oxford University Press Inc., New York, USA.

Ellis, F. 2000. *Rural Livelihoods and Diversity in Developing Countries*. Oxford University Press, Oxford, UK.

ELDIS guide. http://www.eldis.org/statistics/index.htm.

Fahamu. *Writing for Effective Change*. http://www.fahamu.org.uk

Lenin, V.I. 1961. *The Development of Capitalism in Russia*. 1908 from V.I. Lenin, Collected Works. Fourth English edition. Foreihn Languages Publishing House, Moscow, Russia.

UNDP. 1998. *Zimbabwe Human Development Report*. United Nations Development Programme, Rome, Italy.

Wye College. 1995. *Research Methods and Data Analysis*. Wye College External Programme, University of London, London, UK.

4.2 Spatial data and geographic information systems

Thomas Gumbricht

- **Understanding the spatial context of an agricultural and resource management problem will probably be an important part of solving it, so can not be ignored in your research**
- **Geographical information systems (GIS) allow you to manage and manipulate spatial data**
- **Simple manipulation of data sets that has already been prepared can be learned quickly. However, using data from multiple sources for more complex tasks can be a major undertaking**
- **Many basic spatial data sets are available for Africa but poor Internet connections may limit access to them**
- **Freely available software is now sophisticated enough to be useful in many spatial research projects**

Introduction

The Earth is a sphere with an average distance to the Sun of 150 million km. The Sun radiates energy, which is received by the rotating Earth in diurnal cycles with annual modulation as the Earth completes its annual ellipse. The energy that hence reaches the Earth is mainly dissipated at the Earth's surface. It rotates the hydrological cycle, releases nutrients that feeds the ecosystems, and drives photosynthesis (all which have been largely altered by man since the Industrial Revolution). Thanks to these processes life exists and the Earth's surface has developed a 'natural' logic. In dry areas with poor resources vegetation is sparse, in valleys where water and resources accumulate the vegetation is more luxurious. If there is a trough and enough water a body of water will form. In a similar way the human landscape is also logical, with fields in fertile valleys and dwellings along the ridges. Cities have to be close to large sources of water. These logical landscapes are also evident on a much smaller scale. Most vegetation is bound to specific habitats narrowly defined by conditions of climate, soil and water; that can shift within a scale as small as one metre. At an even finer scale a human thought is also dependent on energy dissipation at interfaces – in a very well described spatial context between the synapses of nerve cells. Image analysis and location information systems are hence very important tools in medicine, sociology, anthropology, biology, ecology, geology, hydrology and many other sciences.

For a particular study the spatial information needed might only be a map – as were the descriptive studies conducted by the first European explorers. In most instances a researcher is probably more interested in extracting more information in order to test a hypothesis. This could be comparing two district-level data sets, perhaps one on poverty and one on incidence of malaria. This is easily done in a geographic information system (GIS), and you still only need a single map, with attributes (databases) on both malaria and poverty. But malaria is a vector-borne disease, and the mosquito carrying the parasite breeds in water, so proximity to water is most probably important. To test that hypothesis an additional data layer of water availability is needed. This step is a major complication that has yet to be fully taken in the case of malaria. Rivers and lakes can easily be found, and their proximity to each population group calculated. But now you ideally also want population and malaria data on village level, not just for districts. Then you realise that mosquitoes can breed in water tanks, small puddles, or even water trapped in an old bucket or boot. Now the comparison becomes almost impossible, and you need to get data on rainfall and temperature in order to calculate the daily water balance. This calculation is possible; the data are there (as you will see below), but the calculation is not a trivial task.

In general, a thematic study will need more refined data, whereas an interdisciplinary study must probably be satisfied with more generalised data. Often this is because detailed data of different origin are seldom compatible in their spatial resolution. However, the use of GIS and spatial data can be very rewarding. The first level, including a map, is almost always welcomed and very simple, it will only take a few days. The second level, comparing ('overlaying' in GIS jargon) attribute data related to the same spatial context is also quite simple, and will take a week to a month. The third level, analysing spatial relations introduces complexity, but can still be done by most standard GIS packages (and some of the freeware packages listed on page 142), but it will take some months to a year. The fourth level of integrating GIS with dynamic (time-resolved) models is quite complicated. This level will demand in-depth knowledge of both GIS and modelling, and most probably of programming as well. It will take longer than a year.

Mapping and modelling with GIS – A game of chess

Capturing the chess board

It is seldom convenient to have a sphere to portray the Earth's surface; so humans have used two-dimensional (2D) maps for at least 5000 years. Most GIS are also static 2D, even if the processes occurring on the Earth's surface are often 3D and dynamic. The most common GIS spatial data model represents space as 'vectors' (points, lines and polygons). This model is suitable for human-created objects and concepts (wells, roads, states, cities, rivers and lakes). Natural phenomena are better represented as continuous fields (elevation, land cover, vegetation density), which in GIS translate to a 'raster' or grid data model (Figure 1). For simplistic reasons, the raster model is often preferred in modelling. It is also the implicit format of satellite images or photographs.

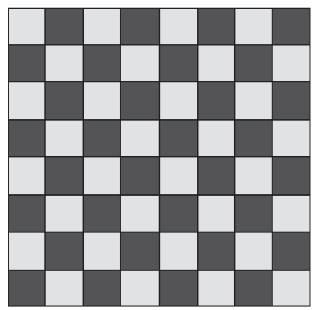


Figure 1. A raster

The raster landscape in Figure 1 is also the landscape where the game of chess takes place. Let us assume that you are unaware of the game of chess, but want to understand it by using GIS. Once you have identified the problem from a GIS perspective you must decide which data model (raster or vector) to use and how to *capture* the data. The chessboard can be captured as primary data (from satellite image or digital photograph) or from an existing analogue (secondary) source (digitising or scanning). Whatever you choose you will, implicitly or explicitly choose a certain grain size (or spatial resolution) when you capture the data (Figure 2). *Meta*-information on capture technique, resolution, and who did it should ideally always follow the GIS data, but is frequently lost on the way to the end-user.

For most spatial phenomena that are studied, you usually have a conceptual idea about the spatial patterns and dynamic processes that are occurring. If you assume that you already have some existing knowledge of chess, you can decide on a stratified sampling of data. For each square of unit distance, take one sample at a randomised point. You can further assume that there is neither an error in position, nor in the obtained value. A point is the simplest kind of vector data, and you can assign attributes to it – you can form a

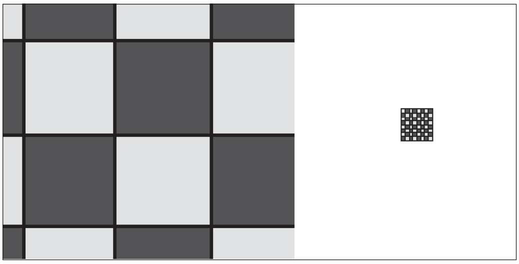


Figure 2. Two raster data sets with different grain size. The ease with which you will be able to understand chess will obviously be dependent on the resolution or grain size used to capture the chessboard

database describing the properties of this point (in this case colour, but it could also be some other capacity such as depth, elevation, or type). Then you can *manipulate* the point data by *rasterising* it to arrive at something that looks like a chessboard. If you honour the value of the measurement in each cell (or *pic*ture *el*ement – pixel) you will arrive at the correct landscape (that you happen to know in this simple case) (Figure 3).

If instead you manipulate the data by using a geostatistical interpolation function, which do not honour the observed value *per se*, you get more or less erroneous results (Figure 4).

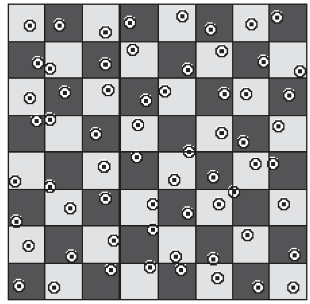


Figure 3. Sample points (vector data) and rasterised pattern

a. **b.** **c.** **d.**

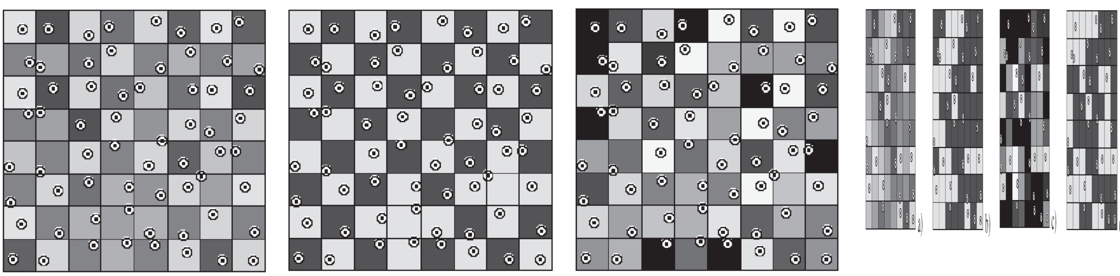


Figure 4. Interpolated 8 x 8 raster image from 64 Boolean sample points, randomly placed in each grid cell: a. Inverse distance weights (IDW) to 8 neighbours, b. Reclassification of a, c. Spline smoothing function to 8 neighbours, d. Reclassification of c. The reclassification is done as a threshold using the value 0.5. Both illustrated interpolation methods can be parameterised to get a true chessboard, that, however, demands iterations and skills, together with knowledge about the pattern of the generated surface

a. **b.** **c.** **d.**

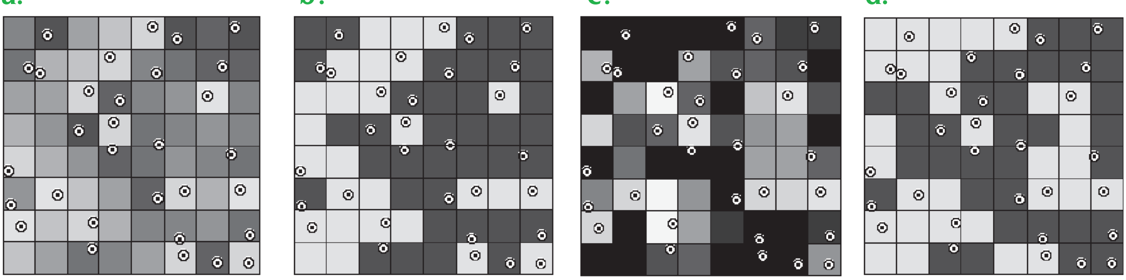


Figure 5. Interpolated 8 x 8 raster image from 31 randomly selected points (see Figure 4) a. IDW to 8 neighbours, b. Reclassification of a, c. Spline smoothing function to 8 neighbours, and d. Reclassification of c. The reclassification is done as a threshold using the value 0.5

The high cost of field and inventory work requires the fullest use of existing data and the application of interpolation methods. Hence, the sampling grid is generally much sparser than the interpolated grid (Figure 5).

Note that the interpolation of the chessboard data are truly 2D, whereas the Earth's surface is a spheroid and interpolation with different geoids and projections render different results. To choose the right projections for a particular purpose is not trivial, but is beyond the scope of this book).

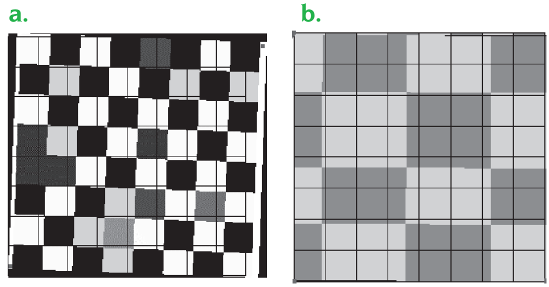


Figure 6. Schematic examples of problems with using remotely sensed data to portray the Earth's surface: a. Georeferencing and spectral properties of the observed phenomena, b. Grain size and geometrical distortions in the sensor

Primary data capture from remotely sensed imagery to GIS is an important part of the integration of GIS and modelling, also in social science for updating or downscaling census data (see below). Remotely sensed data have a definitive grain size and thus resolution. Apart from grain size, problems with sensor quality, spectral properties of the observed phenomena and georeferencing introduce errors when interpreting and classifying remotely sensed data (Figure 6).

Monitoring the dynamic game

Having established the chess-playing arena, a working hypothesis for the processes that are occurring needs to be formulated. For most dynamic phenomena an initial inductive approach is almost inevitable. Only after a set of observations is available is it possible to use coincidental data to formulate a deductive hypothesis.

Observations of natural and human phenomena are often made at regular intervals. Satellite images over an area are usually taken at the same time of day with a given interval (approximately 14 days for Landsat), as are many climate station data, water flow and water quality measurements. Measuring the chess game every morning at 09.30 (cheapest because it is outside the coffee room) always gives the same result (Figure 7).

But if you work late one evening and chance to look at the chessboard, and suddenly see

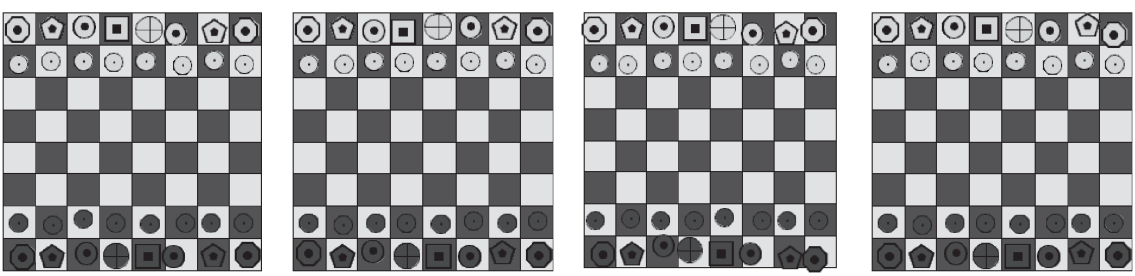


Figure 7. Observations of a chess game on four occasions. At first the game is apparently static. Only with a more detailed scrutiny it is revealed that the players actually are shifted a little between each observation. However, as we have no hypothesis or information of sub-cell pattern or process we neglect this as observation error

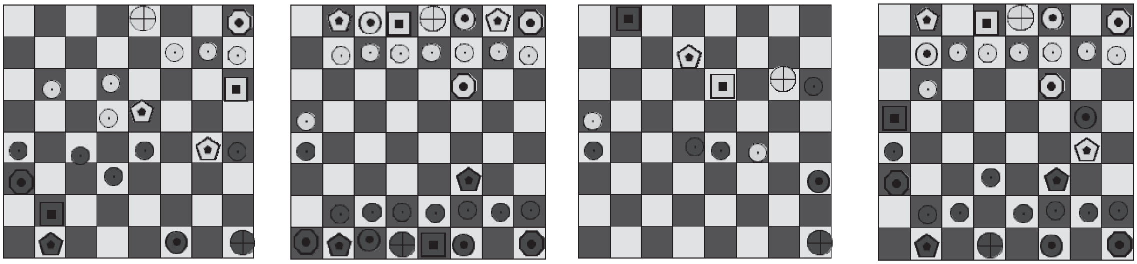


Figure 8. A series of temporally random observations of the chessboard

something has happened, you realise that there is obviously another time scale to the daily one (weekly, monthly or annual). So you start to observe the game regularly when you are working late. A rather erratic series of observations turns up (Figure 8).

Because of the strange observation angle (from above or 'nadir' in remote-sensing jargon) the visualisation of the players is poor, and it is difficult to distinguish the actors. However, a few hypotheses on their roles can be put forward:

1. One species (Bishops) seems to be bound to a certain feature type, or habitat, (namely black or white) in the playing ground.
2. The smallest and most common species (Pawns) seems only to be able to move in one direction like water downhill.
3. The species in the corners (Rooks) seem to be the most home-bound.

After some months of random observations hypotheses 1 and partially 2 are corroborated whereas 3 is falsified. After several years of fund-seeking the observations can be transformed into intense evening campaigns. With observations down to 10-minute intervals some of the players rules crystallise themselves, however the role of the knights escapes a robust formulation. Finally, a sensor connected to a real-time observation can capture the full sequence of activities, and the role of each player can be formulated.

Modelling the full game

The identified role of each player leads to a surge in modelling the game, mostly by using a rule-based (rather than statistical) approach where the roles of each player can be unambiguously defined. The formulation of initial (setting at start) and boundary conditions (edges of the playing arena) are straightforward. The application of an object-oriented approach for each player is favourable; a certain actor can only do a certain action, which cannot be done by another actor.

However, even though the game is spatially defined, it is not possible to use the toolbox of any commercial GIS to play the game. And, only a few softwares have architecture open enough to allow the GIS game to be programmed to them, but with great difficulties. With a customised GIS it is possible to create a graphical user interface (GUI) that can help to set up initial and boundary conditions, and even to allow the set-up of the players' positions in the middle of a game, and the use of that as an initial condition. This leads to the development of an intermediate coupling of the chessboard and the game simulator through their sharing a common file format. It is a bit cumbersome to use and never reaches widespread use to improve the social awareness of the game.

For the game itself, the combination of such advanced machine-learning as artificial neural networks and faster computers, mean more alternative game outcomes can be foreseen after each activity (draw). Finally, one computer (Deep Blue) succeeds in winning the game. This is now more esoteric interest among the chess community, but the general public, policy- and decision-makers are unaware of this development.

Implications

A game of chess always aims at checkmate – which is unambiguously defined, as is the role of each player. The rules of the game show no evolution, neither in space, nor over time. If you change the extent of the arena, the role of the players or the outcome for checkmate to an unknown event, the computer would have little chance of winning. In a transient social or natural environment that is how the evolutionary game is played. In the simple case of chess there are only two scales that are of importance, that of a cell and the whole board. Furthermore, the game as such has no influence on the arena. In a landscape all discretised scales are arbitrarily chosen, the real landscape is a continuous nested hierarchy: but some scales have dominance-generating spatial architectures and temporal cycles, entrapped by key stone species and related processes. This also leads to the conclusion that the processes are forming the patterns rather than the other way around – and that the systems has feedback loops at various scales. All those aspects can be disregarded in the special (and simple) case of the chess game.

The general conclusion that can be drawn is that modelling in GIS is hampered by several shortcomings, that care must be exercised when using distributed data for modelling, and that the quality of many GIS integrated models is poor. They are also poor because they have poor GUIs, fail to visualise the results, and hence do not reach the intended user community. In order to secure high-quality GIS-integrated models the following issues need to be considered:

- Close co-operation between GIS model researchers in general, and particular among
  - researchers studying the same phenomena but adopting different methods and/or scales
  - researchers, planners and decision-makers
- Up- and down-scaling, and nesting models of different resolution
- Spatial and temporal domain, grain size and sampling intensity when integrating data from various sources
- Strategies for sampling spatial phenomena to get representative data
- Selection of spatial interpolation methods and spatially correlated error tracking and tagging
- Methods for evaluating the influence of error and error propagation on model performance, and error visualisation for communication information on uncertainty
- Integration of remote sensing into GIS models
- Integration of temporal processes into GIS (3D- and 4D-GIS)
- Integrated systems that support a complete digital data flow from data collection with mobile field GIS (Global Positioning Systems, GPS) to visualise and exchange results via networks
- Formulation of versatile criteria for evaluating the prediction power of GIS-related environmental models
- Compilation of high quality, accessible (shared) databases to be used as back-drops to evaluate the predictive power of different GIS-related environmental models
- Establishing baseline and framework data
- Development of guiding GUIs that can lead the user to select the best method for the formulated problem and the available data

- Development of friendly interfaces that promote the dissemination of GIS and integrated models to domain experts, planners and managers.

Using GIS in Africa

Studies involving spatial dependence and GIS in Africa are hampered by lack of data and computer resources, and poor knowledge and communications infrastructure. However, with the growth of geoinformatics over the Internet, global and continental-scale data are becoming increasingly available. Together with more powerful free GIS and remote-sensing software, there is a good chance that the data and software needed for many studies are available, either directly or via map algebraic modelling and other manipulations applied to available GIS data in combination with satellite imagery. The global trend in adopting remote-sensing data for spatial studies is strong in traditionally data-poor regions. Free high-resolution satellite images [Landsat Thematic Mapper (TM) and Enhanced TM (ETM)] are now available for the whole African continent. Access to this data in Africa, however, is often illusive due to poor Internet connections. The global data sets derived from satellite data (including land cover) are seldom adjusted for continental needs, leading to semantic discrepancies and interoperability problems when merging data sets. Local knowledge is mostly disregarded. Further, studies employing global data in Africa are often esoteric, and seldom used for policy or management inside Africa.

GIS and remotely sensed (RS) databases for Africa

In this chapter spatial databases have been divided into framework databases and field databases. **Framework databases** are base maps holding mostly information on anthropogenic-derived features – e.g., political boundaries and infrastructure, but they sometimes also have more object-oriented physical themes like elevation contours and hydrography. These databases are typically object-oriented and in vector format. They can be used to create simple thematic maps. Framework databases available for Africa typically contain data at district level, and hence simple descriptive statistical analyses (population density, travel distances, etc.) can be done at a level based on this data. Framework data can seldom be used directly for advanced analyses and modelling (environmental studies). Environmental studies demand field data, usually in raster format for such parameters as population density, soil classes, drainage, elevation, temperature, and precipitation.

Framework databases for Africa

The foremost baseline framework database for Africa (and other parts of the world) is the Digital Chart of the World (DCW). DCW is a 1:1 million scale thematic map developed by the Defense Mapping Agency (DMA) and compiled by Environmental Systems Research Institute, Inc. (ESRI). For large parts of Africa these base maps are the largest scale maps available, either due to lack of other data or to the larger-scale maps being classified. Themes in DCW include political boundaries, populated places, roads/railroads and other infrastructure, hypsometry, hydrographical data, and rudimentary land coverage.

Based on the DCW, ESRI has assembled a more easily accessible database and has also developed a more field-oriented World Thematic Database. Several other GIS software producers have also established databases based on DCW (and additional sources mentioned below) for bundled delivery. The most comprehensive probably being the Mud-Springs Geographers – AWhere Almanac Characterisation Tool (ACT). This and other software tools (listed on **page 142**) are a very good way to learn GIS using data over Africa. AWhere-ACT is especially powerful for analysing climate data (supplied with the software) for agriculture and natural resource management applications. In many cases the software and bundled data are free for use in Africa by non-profit organisations.

Several recent efforts in creating more-detailed (large-scale) regional framework databases for Africa have been made. The most comprehensive is probably the Africover project by the Food and Agricultural Organization of the United Nations (FAO). This database also includes detailed land cover derived from combinations of Landsat ETM data and topographic maps. The agencies of the UN have also initiated an attempt to create a common depot for their GIS data – which has led to the Data Exchange Platform for the Horn of Africa (DEPHA) (see **page 140** for a more complete list of framework data sources available).

Field databases for Africa

Elevation

For Africa the elevation data in DCW (contour lines and spot elevation data) together with generalised 3-arcsecond digital terrain elevation data form the primary source for the global 30-arcsecond (approximately 1 km) GTOPO30 elevation database released by the United States Geological Survey (USGS) in 1996. The data in GTOPO30 have been hydrographically corrected and resampled to a 1-km grid, to create the HYDRO1k database. From the hydrologically corrected HYDRO1k Digital Elevation Model (DEM) seven derivative themes have been extracted: flow directions, flow accumulations, slope, aspect, compounded wetness indices, stream-lines and basin areas. Several individual countries have better elevation databases. The next elevation data set covering the whole of Africa will be the Shuttle Radar Topography Mission (SRTM) database (90 m resolution), expected to be released during 2004.

Land use/cover

Two global land cover data sets covering Africa in 1-km resolution are presently available. The latest is derived from TERRA-MODIS (Moderate Resolution Imaging Spectroradiometer) data (2000/2001) and was created by the University of Boston. MODIS has also been used to create a global tree cover database in 500-m resolution available from the University of Maryland. The older land cover is produced by the USGS from NOAA–AVHRR (Advanced Very High Resolution Radiometer) data (1992/1993). It exists in several versions useful for different applications and also includes monthly vegetation data from April 1992 to March 1993. The Africover database mentioned above is superior to these global databases but does not yet cover the whole continent.

Climate and vegetation

The United States Agency for International Development (USAID), as part of the Famine Early Warning System (FEWS), continuously provide 10-day composites of vegetation density (Normalised Difference Vegetation Index – NDVI) derived from NOAA–AVHRR in 8-km resolution covering the whole African continent. The data set goes back to 1981 and is archived and disseminated by the USGS. It can be retrieved from the African Data Dissemination Service (ADDS). Thermal Meteosat images together with 760 ground precipitation stations are used to estimate precipitation over Africa as part of USAID FEWS. Processing is based on 30-minute image intervals for cloud top temperature combined with the ground data and derived fields of humidity, winds and DEM. The data extends from 1995 and are archived and disseminated by USGS (ADDS webpage) as 10-day composites. More coarse resolution databases that cover climate together with scenarios of climate conditions under various assumptions of human impacts on the climate are available from the Climate Research Unit (CRU), University of East Anglia, UK, either directly via the Internet, or from the Intergovernmental Panel on Climate Change (IPCC) as a CD.

Population

The best and latest population figures are the 1-km resolution Landscan project data for 2000, 2001 and 2002 from the Oak Ridge National Laboratory, USA. These figures are created from census data and downscaled using intelligent interpolation (using relations such as light at night, slope, or elevation, which correlates strongly with population density). The Center of International Earth Science Information Network (CIESIN) hosted by University of California, has compiled global population data for 1990 and 1995. The data has an original resolution of 5 arc-minutes (approximately 10 km), but for Africa the data mostly represent averages for larger regions. United Nations' African population figures for selected countries covering the second half of the 20th century are available from Central African Regional Program for the Environment (CARPE) (see **page 140**).

Soil map

FAO has produced a Digital Soil Map of the World (DSMW) in 1:5 million scale. Soil classes are given as polygons, with derived characteristics attributed. The soil map is only available as a CD. For some regions FAO also has a 1:1 million scale soil map.

Satellite imagery

Remote sensing (RS) data are increasingly important for creating and updating both physical/biological and socio-economic databases. Access to RS data is constantly improving thanks to: lowered prices, declassification of historical high-resolution data, a new generation of multi-sensor satellites (TERRA and ENVISAT) that are now operating, improved computing power and better software–user interfaces.

For national to continental studies NOAA-AVHRR and TERRA-MODIS data and their derivatives are the most easily accessible. Other data of similar resolution that can be easily accessed include the European Space Agency (ESA) ERS-2 satellite and its ATSR 7-band sensor (which can be downloaded from the Internet in near real time), and the SeaWiFS 6-band sensor.

Full coverage, high-resolution Landsat TM and ETM data are now also freely available for the whole of sub-Saharan Africa via the University of Maryland. Landsat E(TM) composites in Mr-SID compressed formats of the whole globe are more easy to download and available from NASA. To find all available Landsat MSS, TM and ETM scenes, and other satellite data sources use the NASA Earth Observing System Data Gateway.

The original TERRA–MODIS and NOAA-AVHRR scenes that were used for the land use/cover classifications (see above) are all freely available as composites from University of Maryland (TERRA–MODIS) and USGS (NOAA–AVHRR). The Africa NOAA–AVHRR tiles for vegetation are also available from the International Centre for Insect Physiology and Ecology (ICIPE). Additional, raw, NOAA–AVHRR data are available via the NOAA Satellite Active Archive on the Internet, or from USGS at the cost of reproduction.

Georeferenced time series point data for Africa

Time series point data on climate (weather station data) and hydrology are available via a variety of web pages. As this is outside the main scope of this chapter we recommend the ICIPE data server as a source of archived weather station data from around Africa.

Data accuracy and merging

Most of the older global and regional databases are not quality labelled, neither for positional error nor attribute accuracy, thus potentially leading to large problems in interpretations and applications. Even if most data that can be downloaded are georeferenced, their accuracy often does not allow mapping at higher resolution than 1:1 million. The dataset with highest spatial accuracy is the NASA geocover (downoadable as MrSID images – see **page 141**), which is within 50-m and can hence be used for geocorrecting other data sets. Another problem is that the semantics used, for instance, for the land use/cover maps are not coherent with those used in different parts of Africa. This is also due to a poorly developed unanimous semantic cover of natural geography in Africa. Semantic inconsistencies lead to information loss and prevent sound conclusions being drawn.

For most spatial studies, it is necessary to merge data. Most satellite images must be georeferenced to a projection that fits the geographic location (and the framework data) before they can be used for analysis and further studies. This is not a trivial task. Field data sets continuously vary over time and space on different scales. A satellite image is already an aggregation of the land surface over the pixel size. Field data, including socio-economic data are often up- or downscaled, or aggregated. The quality difference between data sets of the same origin but presented at different scales is seldom reflected in the metadata. It is extremely important to know the timing of acquisition, grain size and scaling of field data when analysing, interpreting, and applying such data. For most of the global data sets available this is seldom a problem. However, most local users are ignorant of the problem and secondary data sets derived from such sources often lack meta-data.

Points to remember

- Increased data availability and the ease with which distributed data layers are created from point and line data, and remote sensing, have led to a widespread coupling of GIS and remote sensing to existing (non-topological) cause-effect models in, for example, hydrology and erosion studies, and to updating and downscaling land cover and population density maps
- Data availability for Africa has now reached a point where it is possible to do such studies, often with freely available data
- The major problem for the individual researcher is in accessing the data, and in acquiring the skills of GIS and RS needed to 'massage' the data into a coherent database
- The free GIS software programmes available today are powerful enough for you to learn GIS, and to create basic databases
- The bottleneck for using GIS for research in Africa is poor Internet access and poor GIS skills
- If you want to use GIS you should download the necessary data or order it via CD/DVD (usually possible for a small fee), and you should learn GIS by using one of the listed free software programmes
- As most of the software programmes have very similar interfaces, learning one means that learning a second becomes an order of magnitude easier. So going from DIVA-GIS to ArcView is very simple (also because they share data formats).

Expert systems

GIS and RS (or geoinformatics) have developed from being tools for data storage and presentation to also include analyses and modelling. Overlaying two or more thematic maps (see Figure 9) is a simple but often illustrative means of identifying relations in spatial

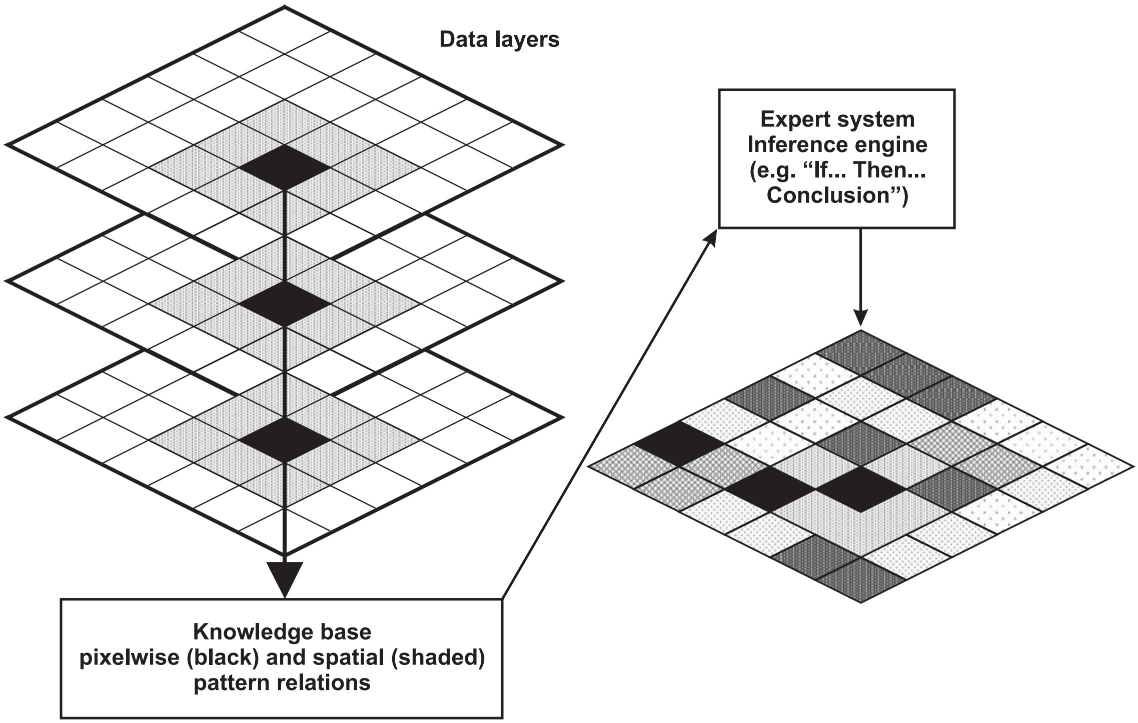


Figure 9. Schematic structure of an expert system approach for spatial data analysis

patterns. More advanced analyses include using map algebraic formulae combining several thematic layers. Such 'expert system' approaches are widely used to rank vulnerability of natural resources, food security, or water availability. One example is the DRASTIC method (Depth to groundwater, Recharge, Aquifer media, Soil, Topography, Impact of rootzone, Conductivity) for groundwater vulnerability analysis, where each of the seven factors has a physical related value. Development is towards more advanced expert systems including object-oriented methods, and considering ancillary and multi-temporal data, and spatial relations (Figure 9). Expert systems are like the game of chess – unambiguously defined with a set of strict rules. Expert systems are thus said to be data- or forward-driven. However, GIS is also becoming a decision-support system (DSS), e.g., for ill-structured (localisation) problems. Used as a DSS GIS becomes more of a tool for discussions and illustration of decision alternatives. Formal methods have been developed to involve various stakeholders in such discussions, including multi-criteria evaluation (MCE). In contrast to expert systems based on predefined rules and weights of physical parameters, DSS are related to different stakeholders perceptions, and as the aim is to reach a solution (for allocation of land use/development, water or nature protection), the method is said to be goal-driven.

Whether studying natural or social science, GIS can be very useful, and there is a plethora of methods, models and techniques that you can apply to analyse or present data that deals with spatial relations. But it is critical that you formulate a sound hypothesis and use adequate data of sufficient quality. To avoid mistakes a parsimonious approach, and rigourous meta-data description is essential. This will make it easy to update and eventually publish your data and results.

Resource material and references

Burrough, P. and McDonnell, R.A. 1998. *Principles of Geographical Information Systems*. Oxford University Press, Oxford, UK.

Fotheringham, S. and Wegener, M. 2000. *Spatial Models and* GIS. Taylor and Francis, London, UK.

Goodchild, M.F., Parks, B.O. and Steyart, L.T. 1993. *Environmental modeling with* GIS. Oxford University Press, New York, USA.

Goodchild, M.F., Steyart, L.T., Parks, B.O., Johnston, C., Maidment, D., Crane, M. and Glendinning, S. 1996. GIS *and environmental modeling: Progress and research issues*. GIS World books, Fort Collins, Colorado, USA.

Framework databases for Africa

Geography network
http://www.geographynetwork.com/

ESRI downloadable data
http://www.esri.com/data/download/index.html

Data Exchange Platform for the Horn of Africa: UN organisations Geo data depot.
www.depha.org

Digital Chart of the World (DCW): Basemaps for all the countries of the world.
http://www.maproom.psu.edu/dcw/

Food and Agriculture Organization of the United Nations (FAO)
Very good land cover maps over East and Central Africa
www.africover.org

Global GIS database – Digital Atlas of Africa
http://webgis.wr.usgs.gov/globalgis/

Field databases for Africa

GTOPO30 global topographic data
http://edcdaac.usgs.gov/gtopo30/gtopo30.html or
http://www.ngdc.noaa.gov/seg/topo/globe.shtml

Hydro1K HYDRO1k Elevation Derivative Database
http://edcdaac.usgs.gov/gtopo30/hydro/index.html

Landscan population data Oak Ridge National Laboratory
http://web.ornl.gov/sci/gist//landscan/index.html

Global landcover from NOAA-AVHRR (1992–1993 data)
http://edcdaac.usgs.gov/glcc/glcc.html

MODIS land cover from Boston University (2000–2001 data)
http://duckwater.bu.edu/lc/mod12q1.html

MODIS Global Vegetation Continuous Fields from 500m MODIS data 2000–2001
http://modis.umiacs.umd.edu/vcfdistribution.htm

CIESIN (Center for International Earth Science Information Network) Colombia University
Including climate data and global gridded population data from 1990 and 1995
www.ciesin.org

CARPE (Central African Regional Program for the Environment)
http://carpe.umd.edu/products/

Satellite imagery and related data

University of Maryland Global Land Cover Facility
A very good source of free Remote Sensing (Landsat (ETM) and TERRA MODIS) scenes
http://glcf.umiacs.umd.edu

USGS Land Processes Distributed Active Archive Center
http://edcdaac.usgs.gov/main.html

USGS Earth Explorer
http://earthexplorer.usgs.gov

USGS NOAA-AVHRR used to create global landcover – 93 original scenes (1992 to 1996)
http://edcdaac.usgs.gov/1KM/1kmhomepage.html

Africa Data Dissemination Service (United States Geological Survey – USGS)
http://edcw2ks21.cr.usgs.gov/adds/data.php

NASA Earth Observing System (EOS) Data and Information System
http://edc.usgs.gov/

NASA Earth Observing System Data Gateway.
http://edcimswww.cr.usgs.gov/pub/imswelcome/ or
http://redhook.gsfc.nasa.gov/~imswww/pub/imswelcome/

NASA global Hydrology and Climate Centre (Weather satellite data)
http://wwwghcc.msfc.nasa.gov/GOES

Goddard Institute for Space Studies
http://www.giss.nasa.gov/data/ and http://xtreme.gsfc.nasa.gov/

National Geophysical Data Center
http://www.ngdc.noaa.gov/

MrSID images (excellent geocorrected – can be used for georeferencing other spatial data)
https://zulu.ssc.nasa.gov/mrsid/

ICIPE (International Centre for Insect Physiology and Ecology) Africa Data Bank
(Including remote sensing data and weather station data Over Africa)
http://informatics.icipe.org/databank/

Japan Aerospace Exploration Agency
(Free JERS-1 radar images over most of Africa can be ordered on CD)
http://www.eorc.jaxa.jp/eorctop.htm

SRTM (Shuttle Radar Topography Mission)
http://www.jpl.nasa.gov/srtm/

Visible Earth – NASA site with preprocessed satellite images of Earth
http://visibleearth.nasa.gov/

Microsofts image database (terraserver)
http://terraserver.com/ or http://terraserver-usa.com/

Digital Globe (very high resolution data sets over selected cities)
http://archive.digitalglobe.com/

Geocommunity spatial news (incl Landsat viewer)
http://spatialnews.geocomm.com/

Free software sources

Dynamic Maps (A free ware GIS with many predefined functions for natural resource management)
http://www.skeinc.com/ or via www.africover.org

DIVA-GIS (A fully functional GIS developed by the International Potato Center)
http://diva-gis.org/

Arc-Explorer (Light-weight GIS by ESRI that also produced Arc-Info, Arc-View and Arc-GIS)
http://www.esri.com/software/arcexplorer/index.html

ERViewer (viewer for many image formats)
www.ermapper.com

WINDISP (A simple freeware for image processing from FAO)
http://www.fao.org/WAICENT/faoinfo/economic/giews/english/windisp/dl.htm

Mud Springs Geographers (A fully functional GIS bundled with free GIS data for Africa)
http://www.mudsprings.com/home.aspx

Mapmaker basic (Light-weight GIS freeware)
www.mapmaker.com

SILVICS (Satellite image processing for forests)
http://eurolandscape.jrc.it/forest/silvics/

GRASS (Advanced GIS and image analysis for UNIX or Linux)
http://grass.itc.it/

Microdem
http://www.usna.edu/Users/oceano/pguth/website/microdemdown.htm

4.3 Designing experiments

Richard Coe

- **Experiments are a central part of the scientific method because they allow you to test cause-effect hypotheses**
- **Many students learn about experiments in the context of studies of small field plots, but the key principles of experimental design are equally important in all studies**
- **All aspects of the design of an experiment depend on its objectives, so the objectives have to be carefully and thoroughly developed**
- **The details of the design of a good experiment will balance theoretical optimality with practicality**
- **Every experiment should have a written protocol that can be shared with others, so the design can be improved before the experiment starts**

Experimenting as part of research

Experimenting is a part of everyday life. In an informal way you experiment when you check whether your tea is too hot to drink, whether the bus is less crowded if you leave for work earlier or whether your supervisor approves of your style of writing. Within formal agricultural research, experimentation has long been the key tool. To many people 'agricultural research' is synonymous with field plot experiments. If you visit an agricultural research station one of the main things you can see are small plots for comparing different crop varieties or different management techniques. Much of the current theory and the methods for carrying out experiments were developed in the context of agricultural experiments, most notably by R.A. Fisher, a geneticist and statistician working at Rothamsted Experimental Station in UK.

Today field plot experiments on research stations are not the only avenue for agricultural research, but the ideas and methods of experimentation are still central to good research. Why?

Experimentation is concerned with the 'testing theory' step of research **(Chapter 4.2)**. Theories which help in problem solving often describe what will happen if a change is made:

- If we use this new variety of maize there will be less damage from stem borers
- If we substitute dairy meal with calliandra fodder milk production will not be reduced
- If we train farmers in pest management they will be able to grow cabbages more profitably
- If communities are better informed they will be more effective in managing common grazing.

Now in order to test your theory the obvious thing to do is to make the change and observe whether the predicted outcome occurs. This is the basis of experimentation, and the reason it is so important.

There are situations in which it is impractical or unethical to experiment, in these situations other ways of testing theories have to be found. It is not feasible to experiment if your prediction is:

- If the average annual temperature rises by 2°C then maize production in Kenya will drop by 15%

You could test the prediction by setting up simulation models **(Chapter 4.8)** that describe the relationship between production and temperature. But those models will themselves be based on theories tested by experiment. Here is another well known prediction made some years ago:

- Regular smoking will lead to an increased chance of lung cancer and other diseases.

It was not possible to test this by experimentation as that would have involved taking a group of people and requiring some of them to smoke. This theory was tested largely by surveys **(Chapter 4.4)** which are distinct from experiments. In a survey you observe what is happening without making deliberate changes. Thus the effect of smoking was investigated by comparing the health of people who smoke with those who don't, and a clear correlation emerged. The limitations of the study design are clear: if the smokers have a higher rate of lung cancer we can not be sure the smoking **causes** the lung cancer. Perhaps there is some unknown factor that tends to lead people to both smoke and get lung cancer. This is a problem of the survey approach to investigation, and means that theory testing is harder using surveys than using experiments. In the case of the health impacts of smoking, various possibilities for such factors were suggested (diet, genetics), and then eliminated by surveys which controlled for them, each providing evidence in support of the theory. However there will always be people who think of one more factor that could be the explanation. This would not be the case if the theory could be tested by a well designed experiment.

This chapter summarises the key decisions that have to be made if you are to conduct a well designed experiment. A prerequisite is understanding the principles and language of experimental design, described in the next section.

Basic ideas of experimental design

Think of the simple problem of determining whether the new maize variety 'Boreproof' is less susceptible to stem borer damage that the commonly used variety M512. That is the **objective**, and the objectives of the experiment will determine all other aspects of the design.

You could plant a field of Boreproof and measure the stem borer damage. But with what will that be compared? The objective requires checking it has less damage than M512. You need to make a **comparison** and so plant a second field with M512. The two varieties being compared are the two **treatments**. They are compared on fields, so the two fields are the **experimental units.**

How effective will this design be? What might you conclude when you get the data? If 50% of the plants in the M512 field are damaged, but the Boreproof field only has 20% damage has the theory been confirmed? Hardly! Any agriculturalist will tell you that stem borer damage can vary greatly between fields, as well as between different parts of the same field. So an alternative design is suggested. Have several fields each of Boreproof and M512. Suppose the results show the fields of Boreproof as having damage levels of 20%, 10%, 40% and 30%, while M512 has 50%, 60%, 40% and 35%. The **replication** of the fields allows you to check the consistency of the results. These results show a tendency for Boreproof fields to have less damage than M512 (the mean is 25% compared with 46%) but the results are not very convincing, with some M512 fields having less damage than some Boreproof fields.

A third alternative design is tried. Since you know that the pest pressure will vary between fields irrespective of the varieties being grown, maybe you can increase **precision** by growing both varieties in the same field. Make the experimental unit a **plot** (say 10m x 10m) of maize, with two plots in each field. Then put M512 on the left-hand plot and Boreproof on the right in each pair. You can now compare the two treatments within each field, and differences between fields become less important. Using the fields in this way is described as **blocking**, with each field being a **block.**

The results from this design are shown in Table 1.

Table 1. Results of a simple experiment

	Stem borer damage (%)	
Field	M512	Boreproof
1	50	20
2	20	10
3	30	20
4	60	30
5	60	40
6	20	5
7	0	0
8	40	10

Are these results more convincing? They certainly show consistency: Boreproof had less damage than M512 in every field except field 7, which has no stem borers anyway. But look carefully at the way the design was described. The M512 was always placed on the left-hand plot. Maybe the difference in stem borer damage is nothing to do with variety, but due to some other consistent difference between left- and right-hand plots. Maybe the wind blows from the left, bringing the pests or stressing the plants. You may know that is not the case, but could have trouble convincing others. And you can never be sure that there is not some other systematic difference between left- and right-hand plots. The solution is to **randomise** the allocation of treatments to plots. In field 1, toss a coin to decide whether Boreproof or M512 goes on the left-hand plot. Then randomise again in field 2, and so on. In field 1 you might end up with Boreproof on a plot with less stem borer damage for reasons unconnected with the variety, but over the whole experiment you can be sure that the only systematic difference between plots with Boreproof and plots with M512 is indeed the variety.

The basic ideas of experimentation described above should also help you understand studies which, though they involve comparison, are not experiments and can not demonstrate cause. For example, suppose a study showed that farmers in Central district have less stem borer damage in their fields than farmers in West district. They also have higher adoption rates for the Boreproof variety. This study involves comparing districts, but is not an experiment as the differences in adoption of Boreproof were not imposed by the researcher. It is also common to make comparisons over time, for example, by comparing stem borer damage levels before and after introduction of Boreproof in Central district. In such studies the change may be devised by the researcher, but it should only be considered an experiment if other features are present, e.g., some other districts in which Boreproof was not introduced, random allocation of the introductions, and some replication.

Diverse applications, common principles

The simple example in the previous section explains the basic ideas and terminology in the context of a 'classical' agricultural experiment – a variety trial. This section shows the correspondence with three other studies of different types. Each discipline tends to produce its own language and standard practices and it is important to recognise the commonality between them, and to make sure that you really understand the logic of the design. The same topic is discussed elsewhere in different contexts.

The examples in Table 2 are all different. One investigates field plots, one animals, and two people. Of the last two, one focuses on individuals and the other on communities. The practicalities of carrying out each of these studies will be quite different, but the fundamental logic of the design is the same in each case. The social sciences do not use the terms **treatment** and **unit** but the rationale for their approach is similar to research in the natural sciences. The roles of comparison, replication, randomisation, and controlling variation are the same for all of them. It is common for these aspects to be forgotten in studies involving people, particularly community-based studies like the last one in Table 2.

Table 2. Four examples of experiments

Objective	Treatments	Units	Measurement
Determine if Boreproof is more resistant to stem borer than M512	1. Boreproof maize 2. M512 maize	10m x 10m plots of land	Percentage plants damaged by stem borer
Find the effect on milk production of substituting dairy meal with calliandra fodder	1. Base diet + dairy meal 2. Base diet + calliandra 3. Base diet + 50% calliandra + 50% dairy meal	Dairy cows for 2 weeks of third month of lactation	Milk production in the second week
Check whether training in pest management allows farmers to produce cabbages more profitably	1. No training 2. Attendance at farmer field school on pest management	Farmers for whom cabbage production is a main enterprise	1. Farmers' knowledge 2. Profitability of cabbage enterprise
Evaluate the effect of community information and organisation on common grazing management	1. No intervention 2. Information provided 3. Information provided and village 'grazing committees' facilitated	Villages in areas where common grazing is degrading	1. Community views on grazing problems 2. Range quality

Design decisions

Now you understand the basics of designing experiments you can start thinking about the design of your experiments. You may need one substantial experiment, several smaller related experiments or possibly no experiments. If you are going to experiment then there are many decisions you will have to make about details of the design. How can you make those decisions? There are several sources of help:

- The fundamental principles of experimental design – the outlines above and more details in other texts
- The more practical ideas in the following sections, and in other texts
- Papers and reports describing similar experiments that others have done
- Other researchers who have worked on a similar topic (maybe in a different region) or used a similar method
- Your observations of other experiments
- Your imagination
- Pilot studies in which you try out techniques and arrangements before committing yourself to an expensive or long-term experiment.

There is no single correct way to design your trial, but there will be plenty of ways that are wrong – designs which will not lead to valid conclusions meeting your objectives. Even if you design a trial that will give valid results it may be inefficient – not give you as much information as possible for the time and effort spent. Avoid these scenarios by:

1. Thinking.
2. Using all the sources of help listed above.
3. Showing your design to others and getting their comments.
4. Envisaging the data your design might produce and the way in which you would then interpret it. Some researchers sketch out the tables and graphs they would use in the

analysis of the data, then making sure the design will generate the required numbers to complete them.

5. Thinking of the practical as well as the theoretical requirements. You have to manage your trial (set it up, look after it), cope with the travel requirements, have enough time and equipment to measure all the plots, and so on. And you have to be able to afford it!
6. Iterate. Start with a possible design, think through the consequences then go back and revise it until you have something sound.
7. Thinking.

In the following sections the main ideas you need to make decisions on each of the key points are described together with some of the common mistakes that you must try to avoid.

Objectives

All aspects of the design depend on the objectives. Therefore you must get the objectives right! Objectives must be:

- **Clear.** If the objectives are vague it will not be possible to decide on the rest of the design
- **Complete.** Often the statement of objectives is incomplete so that the experiment can not be designed.
- **Relevant.** In applied research, experiments are made to help solve real problems and fill knowledge gaps in the process. The objectives of the experiment must be relevant to solving the problem. It must be clear how you will be a step nearer solving the problem once you have the results from the experiment
- **Reasonable.** The objectives must be reasonable given current understanding of relevant phenomena and other observations. Avoid objectives that contain elements of alchemy or wishful thinking
- **Capable of being met by an experiment.** Some research questions do not need an experiment. Two problems which often arise here are:
  - objectives that require a survey rather than an experiment
  - objectives that require two or more experiments rather than a single one.

Make sure that the objectives fit in well with the overall strategy of the project. You have to be able to explain what the next step will be after the experiment is completed.

Common mistakes to avoid

1. Objectives which are too vague. The objectives in Table 2 all fall into this trap! Real experiments would need to have objectives that made it clear, for example, what sort of base diet is to be fed to the dairy cows, how much training in pest management should be given, or where the rangelands are located.
2. Objectives which just say ‘the objective is to compare the treatments’. Treatments should be a consequence of the objectives, not the other way around.
3. Loading too many objectives into a single experiment, so no design can be found that meets all of them. For example, in trials in farmers’ fields, understanding details of biophysical processes usually requires a high degree of uniformity, and hence the researcher taking control. Eliciting farmers’ assessments of the technologies requires them to have a free hand. Thus the two objectives will probably not be met in a single trial.

Treatments

There are four ideas you need when choosing treatments:

1. **Comparison and contrasts.** Experiments involve making comparisons. The exact comparisons that meet the objectives can be defined as contrasts, i.e., the numerical expression of the comparison. Make sure your experiment has all the treatments needed to make all the comparisons implied by the objectives.
2. **Controls.** 'Controls' or 'control treatments' are the baseline treatments against which others are evaluated. In the stem borer experiment M512 might be considered the control.
3. **Factorial treatment structure.** Many experimental objectives require looking at several 'treatment factors'. For example, in the stem borer experiment you may also want to look at the effect of sowing date (early, mid, or late). Then the experiment might have 6 treatments (Boreproof sown early, mid, or late and M512 sown early, mid, or late). Factorial treatment structures are important for two main reasons:
   - they tell you about **interaction** – such as whether the difference between Boreproof and M512 depends on when they are sown
   - if there is no interaction they give information about both factors with the same precision as would be obtained if the same amount of experimental effort went into investigating just one of them. This is the 'hidden replication' described in textbooks.
4. **Quantitative levels.** Some experiments require varying a quantity that could have many different levels, such as sowing date or amount of fertilizer applied. Choosing the levels to use as treatments in the experiment depends on the exact objectives and what you already know about the response to varying it. Generally fewer rather than more levels are needed, and there is rarely a reason for using more than 4 different levels.

Some common mistakes to avoid

1. Including extra treatments 'because they might be interesting' rather than because the meet a clear objective.
2. Missing suitable controls, so, for example, the new varieties are grown but there is nothing against which to assess them.
3. Thinking 'control treatment' means 'do nothing' or 'zero input', even though those might be appropriate in some cases. Control treatments are just treatments needed make the required comparisons, so you may have two or more controls corresponding to different objectives.
4. Using too many levels of a quantitative factor. Using 10 levels, say 0, 10, 20, 30, 40, 50, 60, 70, 80, and 90 kg N/ha will give more information about response to fertilizer than just using 0, 20, 50 and 90 kg N/ha. But if you can make a total of 20 observations (e.g., if you can only afford 20 plots) then 2 replicates of those 10 different treatments will almost certainly give less information about response to N than 5 replicates of the latter set of 4 levels.

Another key idea is **confounding**. Suppose that in the stem borer experiment M512 was sown on 20 March but Boreproof was not sown until 2 April, because there was a delay in procuring the seed. When the stem borer damage is observed to be less in Boreproof we can not conclude that the variety is resistant. The difference in damage may be due to the different sowing dates or different varieties. Sowing date and variety are said to be 'confounded'. **Treatments must be defined in a way that does not confound different effects.**

Units

With crop experiments, decisions have to be made on size, shape, orientation and arrangement of plants within the plot. There are a few guidelines based on theory:

- Many small plots often give more precise results than a few large plots taking up the same area
- Long thin plots often give more precise results than squarer plots.

These guidelines have to be modified by practical considerations:

- Plots have to be large enough to manage (sow, weed, spray, harvest) in a way that represents what a farmer could do
- Plots have to be large enough to take measurements, allowing for the possible disturbing effects of destructive measurements during the experiment **(Chapter 4.5)**.
- Borders may have to be left around each plot to make sure that anything happening on one plot does not influence what goes on in the next plot.

These considerations, particularly the last point, often overrule the theory.

If the units are not plots of land but animals, people or communities then there are often more decisions to make and few general guidelines. **Base the design of the experimental unit on the experience of others who have done similar experiments.** What did they use as the unit? What problems did they have? How will your experiment differ from previous ones? Does that imply any changes in unit?

Think of the experiment with factorial treatment structure, with two varieties (Boreproof and M512) each sown early, mid, and late. A common design for this type of experiment is the **split-plot.** Large plots are defined and the early, mid, or late sowing date allocated randomly to each one. Then each large plot is divided into two, with M512 and Boreproof randomly allocated to the two halves. A split-plot design can have practical advantages, for example, you are less likely to disturb the early sown plots when sowing the later ones. However it does have disadvantages. There are two sorts of plot (large plots and split plots). This complicates the analysis because variation between both types of plot has to be considered. The precision of a split-plot trial is generally less than for that of alternative of random allocation of all treatment combinations to the smaller plot. **Don't use a split plot design unless practical considerations require it.**

Some common mistakes to avoid

1. Plots are often too small, so it is not possible to manage or measure them realistically. An extreme example occurs when measuring labour. It is not possible to estimate the labour required for a task such as weeding if the plot is very small, because weeders will not work at the same rate per unit area as they would in a larger plot.
2. In situations other than annual crop experiments, interference between plots can be hard to see, but can seriously bias results. Water and insects can move from one plot to the next. Tree roots can grow into neighbouring plots. If the unit is a farmer, he or she may talk to another farmer and influence results in an unplanned way.
3. Some researchers seem to believe that experiments with factorial sets of treatments have to be done in split-plots. This is untrue.
4. Split-plot designs are useful but overused.

Replication

Replication (having several units of each treatment) is important for four reasons:

1. **Estimating precision.** The uncertainty in an average is estimated by the variation between the observations being averaged.
2. **Increasing precision.** Calculating an average over more values from a replicated experiment will increase precision since the calculated value will be closer to the true value.

3. **Insurance.** More replications in an experiment will provide some insurance against things going wrong with one or two replicates. Without such insurance an experiment may be rendered useless by, for example, goats getting into the field, or some participating farmers dropping out of the study.
4. **Replication.** This can increase the range of validity of results if a comparison is repeated under a range of conditions.

There should be enough replicates to satisfy all the above reasons for replicating:

1. **Estimating precision.** Look at the error d.f. (degrees of freedom) from the analysis of variance, 10 d.f. can be considered a reasonable minimum. Much more than 20 has no particular advantage.
2. **Increasing precision.** If you have an idea of the precision you need and the variation in your experimental material then it is possible to estimate the number of replicates needed. Details are in books, and software is available to help.
3. **The number required for insurance must depend on the risks.** A long-term trial in a risky environment (e.g., one that might be burned in the dry season) may be worth insuring, by adding replicates. A short-term trial that can easily be repeated if something goes wrong is not worth insuring.
4. **Increasing the range of validity.** Suppose the stem borer trial had some replicates on sandy soil, some on loam and some on clay soil. Then you could be more confident that the results were generally valid than if the experiment had only been done on sandy soil. The importance of this will depend on the objectives.

Common mistakes when planning the number of replicates

1. Using the ‘usual’ number of replicates (in field experiments this is often 4, for some unknown reason) rather than rationally selecting the number of replicates.
2. Forgetting the ‘hidden replication’ in experiments with factorial treatment structure.
3. Insisting that all treatments have the same number of replicates. In the absence of any other information this is sensible, but it is not necessary. Suppose you decide you need 10 replicates of Boreproof and M512, but only have enough Boreproof seed for 8 replicates. That does not mean you have to also use 8 replicates of M512. It may be sensible to continue with the 10 replicates.
4. Forgetting that in split-plot and similar types of experiments there is more than one type of unit and each has to be replicated sufficiently.
5. Estimating the number of replicates needed, finding it is too large to manage or afford, and proceeding with far too small a number. The experiment will not meet its objectives! If you can not afford to meet the original objectives then modify them rather than carrying on with an experiment that will almost certainly not be useful.
6. Assuming that sub-samples from one plot are really replicates.

The last point is particularly important and is a common problem in student projects. Suppose you measure stem borer damage by selecting 10 plants at random from the 10m x 10m plot and measuring the damage on each one. 10 plants from one plot do not tell you the same thing as 10 plants from different plots. If different treatments are applied to different plots, it is variation between plots which is important, not variation between the plants within a plot. The parallel mistake in the community experiment would be confusing the information from responses of several people in the same village with responses from several different villages.

Site

The site(s) for the experiment will be determined by the objectives. It has to be representative of the problem area, both on a large scale (for example, in the same agro-ecozone) and on a small scale (for example, having the appropriate soil type and previous management). The site also has to be practical. It should be:

- Accessible
- Secure
- Large enough

An experiment will have to be made at more than one site if any of the following apply:

1. The problem area is too variable in key characteristics for a single representative site to be found.
2. You are unsure of the key environmental (biophysical, social or economic) characteristics that may determine the outcome of the experiment, so cannot be sure they are represented by a single selected site. Getting consistent results from several sites will give you confidence that these results really do apply to a wider area.
3. The objectives of the trial require conditions to be compared that cannot be controlled as treatments, such as soil type, rainfall or soil depth.

   Cases 1 and 3 require sites to be selected in the same way that single sites are selected. There is an argument in case 2 for sites to be chosen by random selection, but that is rarely practical.

The same considerations apply when experiments are carried out with farmers and communities. Do not simply choose the villages or farmers in which last researcher worked, but look carefully at the objectives and decide on which characteristics it is important to have represented.

Some mistakes to avoid

1. Choosing a site, such as a university research farm, for its convenience rather than its suitability in meeting the objectives.
2. When working with farmers and doing experiments in farmers' fields, biasing experimentation to wealthier farmers and more fertile fields. There are techniques to avoid this.

If practicality requires you to do either of these things, then you need to be upfront and clear about how you expect them to influence your results.

Blocks and allocation of treatments

Once you know where the experiment will take place and what the units or plots are, you can define a set of units for the experiment. If the units are field plots then you could mark out the field into plots, avoiding places that are clearly unsuitable (a patch with surface rocks, an old termite mound, the strip adjacent to large trees). If the units are not plots you can do the equivalent:

- **If the units are farmers.** Produce the list of farmers who are willing to take part and meet the criteria determined by the objectives. (Make a special note of the potential bias of using only farmers who are willing, or able, to spare the time to take part)
- **If the units are tracts of rangeland.** Map out their location and negotiate their use with the communities who look after them
- **If the units are villages.** Contact those villages that meet the criteria determined by objectives.

Next, determine which treatment will be applied to each unit. Random allocation should be used. Random allocation does not simply mean 'mixed up'. Avoid any possible bias by using an explicit random process. For example, use pieces of paper with treatment names put into a 'hat'. The number of pieces of paper for each treatment will be the number of replicates. Then decide the treatment for the first unit by drawing a paper from the hat without looking, again for the next and so on. There are computer programs to help with this.

The precision of almost every experiment can be improved by blocking. Whatever units you have, you know they will vary. Some variation is predictable. Try to arrange the units into homogeneous groups, each of which will become a block. Table 3 gives some suggestions on characteristics that might be used to block different types of unit, but what is suitable for your trial will depend on the objectives of your trial. For the stem borer experiment, the level of stem borer damage in the previous season may be a good characteristic to use to group units into blocks. However it would be irrelevant if the trial was about N leaching or weeding regimes.

Table 3. Possible factors to use in definition of blocks

Units	Characteristics used in blocking
Field plots	Soil type Previous crop yield Slope Weeds present
Animals for dairy experiment	Weight Previous milk yield breed
Farmers in pest management experiment	Education level Length of time growing cabbages Size of farm
Villages in community resource management experiment	Ethnic group Presence of community organisation

If:

1. Every treatment will have the same number of replicates and
2. Every block has the same number of units and
3. The number of units in a block is equal to the number of treatments.

Then the best design is to put exactly one replicate of each treatment in each block. **The allocation of treatments within a block should be random.** This is the **randomised-block design.**

If the blocks are not all the same size, or the number of units in each block is not equal to the number of treatments, then you will have an **incomplete block design**. Take care when deciding which treatments go into each block. Software is available to help you with this.

Some common mistakes to avoid

1. Assuming blocking is only useful in field experiments. The idea and terminology was developed in the context of field experiments, but it is just as important in all other experiments, though researchers often fail to block experiments with people or experiments carried out in laboratores and nurseries. **Blocking gives extra precision at little or no cost and is almost always worth doing.**

2. Assuming blocks have to be the same size and equal to the number of treatments. Incomplete block designs can be very useful. If you have to get a bit of help designing and analysing them it will be worth it.

Management

In a field experiment, 'management' means preparing the land, sowing, weeding and all the other agronomic practices needed to raise the crop. In other types of experiments there are equivalent management activities. The management of an experiment is often not considered part of the design, yet it can have a large impact on the success of the trial.

1. Decide whether the objectives demand that you manage the experimental material to a very high level (e.g., zero weeds) or a realistic level (e.g., farmers' weeding practice). The first may be appropriate if you are studying processes such as water or N uptake, and don't want weeds to obscure results. The second will be appropriate if you are evaluating technologies and want them to represent farmers' systems.
2. Avoid confounding treatments with management differences.
3. Aim for uniform management. Often the difference between a successful and a failed trial is in how well the crops (or animals or people) were managed, and whether this was done uniformly. You can improve uniformity by, for example, training fieldworkers and monitoring the way they execute operations.

Measurement, data management, analysis and reporting

The measurement and analysis of the trial will also have implications for design, and have to be thought through at the design stage. Details are in **Chapters 4.5** and **4.7.** You will also have to plan a scheme for looking after the data, details are in **Chapter 4.6**.

Writing it up – the protocol

The protocol is the written description of everything that will be done in the experiment, starting with the objectives and going right through to the analysis, interpretation, and use of the data.

- A protocol should be written for every experiment.
- The protocol must be shared **before** the experiment starts to get input from others. There are always other people with expertise that will help increase the quality of your protocol. You should share the protocol with at least:
  - scientists at your location (who understand the local context and constraints)
  - scientist in the region (who understand what else of relevance is happening in the region)
  - a subject-matter specialist (who should be aware of relevant developments around the world)
  - other students
  - your supervisor
  - a biometrician
- The protocol must be sufficiently detailed for someone else to take over the experiment part way through, or to make sense of the data at the end of the experiment, even if you are no longer around
- The protocol should be kept up-to-date. It is not just a plan, but a record of exactly what you actually do.

The protocol must be securely archived so that information about the activity can be found in the future.

Finally: involve a biometrician

Biometricians and statisticians are trained in the art and science of experimental design. They may not understand all the practical constraints and opportunities in your particular study. But they will be able to help with such technical details as choosing the number of replicates and allocating treatments to blocks. They will also be able to spot flaws in the logic of the design, and help you make sure it will really meet the objectives.

Many researchers only consult an biometrician when they get stuck with statistical analysis of the data. That is too late. Get a biometrician's advice early, thereby guaranteeing a good design for your study.

Resource material and references

Appendix 11. ICRAF. 2003. *Genstat Discovery Edition and Other Resources.* World Agroforestry Centre (ICRAF), Nairobi, Kenya. On CD.

Coe, R., Franzel, S., Beniest, J. and Barahona, C. 2003. *Designing On-Farm Participatory Experiments: Resources for Training.* World Agroforestry Centre (ICRAF), Nairobi, Kenya. 135 pp.
Available from www.worldagroforestycentre.org

Mead, R., Curnow, R.N. and Hasted, A.M. 2003. *Statistical methods in agriculture and experimental biology.* Third edition. Chapman and Hall, London. 472 pp.

Robinson, G.K. 2000. *Practical Strategies for Experimenting.* Wiley Series on Probability and Statistics. John Wiley and Sons, Chichester, UK. 282 pp.

Schroth, G. and Sinclair, F.L. 2003. *Trees, Crops and Soil Fertility – Concepts and Research Methods.* CABI, Wallingford, UK. 437 pp.

Stern, R.D., Coe, R., Allan, E.F. and Dale, I.C. (Eds.) 2004. *Statistical Good Practice for Natural Resources Research.* CABI Publishing, Wallingford, UK. 387 pp. (In press).

4.4 Designing surveys

Erica Keogh

- **A survey is a well organised, reliable observation of what is going on in the world, that can be used to show the current status, compare different situations and identify relationships between variables**
- **The same principles of good survey design apply whatever the subject, whether the observations are of people, land, plants, animals or institutions**
- **Design of a survey requires choosing the unit of study, defining the population of these units, selecting a sample of units to measure and designing a measurement tool**
- **A successful survey requires good management of the planning, fieldwork and resulting data, not just application of sound statistics**

What is a survey?

A survey is an observation of what is going on in the world at a particular point in time, but we use the term 'survey' in those situations where:

- Data collection is well defined and organised
- Data collected can be shown to be 'representative' and reliable
- Data are competently interpreted
- Resulting information is utilised while the data is of current value.

In research new empirical information about the world can be collected in two ways – by surveys and experiments. Surveys can be distinguished from experiments by the fact that surveys observe what is there. They do not deliberately make changes to observe the effect. Experiments impose planned changes (the treatments) in order to measure the effects they cause, whereas a survey will investigate one or more characteristics of a population. A survey is distinguished from a census in that a census attempts to cover the entire population while a survey attempts to cover a pre-determined portion (or sample) of the population.

Some books, courses, and researchers imply that surveys are only used to study people. People, households, and villages are commonly the 'objects' studied in a survey. But surveys can be used to study just about anything! In agricultural research you might have to carry out a survey of crops, soils, weed populations, or farm animals or the trees, or sediment in the rivers. Many of the principles of survey design and execution are the same for all types of survey.

Survey information can be collected in an extremely structured manner, or may be more informal, or a mixture of the two approaches, or something in between. Whatever the 'tools' used to collect the information, one thing must be made clear – it is essential to maintain consistency throughout the exercise and to avoid errors arising from inadequately prepared tools.

Why conduct surveys?

Many situations present problems into which you can gain insight by the collection and analysis of survey data, thereby allowing you to:

- Determine existing conditions
- Monitor change over time
- Evaluate new projects

- Forecast future needs.

A survey may seek to

- Describe existing conditions
- Establish relationships between different quantities
- Make comparisons
- Test hypotheses,

or a mixture of all of these.

A survey may target a large sample or a small one – the size will be determined by the sampling methods used and will have an impact on the future use of the results.

Example 1

a. Problem 1. Increased elephant damage has been reported in some villages. Are the elephants moving along normal migration routes or are they roaming more widely than before?

Here you would be interested in **describing existing conditions** and, possibly, trying to **make comparisons** with conditions in previous years. The research could be extended over a period of time, thus **monitoring** the situation over a number of years.

b. Problem 2. Is infestation of maize fields by *Striga* worse when fields are suffering from soil erosion?

In this case you could do a preliminary investigation into whether there is a measurable **relationship** between *Striga* infestation and the level of soil erosion. Alternatively, if there is prior information about this relationship, you can **test the hypothesis** that this relationship exists and is quantifiable.

Types of surveys

There are various ways to distinguish one type of survey from another, but perhaps in the present setting it is best to provide examples which will illustrate the wide variety of studies that are possible.

- Street interviews to assess public opinion about price increases of seed maize
- Household interviews to measure food production and consumption for monitoring food security
- Field observations to estimate earworm infestation in the current maize crop
- Field observations and community discussions to quantify the effects of elephant damage to crops
- Household interviews to gauge the effects of HIV/AIDS on labour availability for household agricultural activities
- A case control study to compare old and new tillage practices in different communities
- An enumeration of tree species in quadrats within a specified area for assessing biodiversity.
- A study to estimate soil fertility prior to land preparation
- A study of a sample of records from the meteorology department to track rainfall patterns over the last 50 years
- An investigation of sections of river banks to determine silting levels arising from gold panning.

From the examples you should realise that a survey may entail interviewing people, or collecting specimens, or measuring items, or studying records, or a combination of one or more of these activities. Thus, the type of survey you are planning dictates what measurement instruments you will be using (e.g., a questionnaire for interviews or a tape measure to check the area planted), and also the sampling scheme (the rules for choosing exactly which things will be measured) you will be using. This matching of 'tools' to the type of study is one of the classic features of surveys, with each survey having a unique set of instruments and methodology for efficient data collection.

Example 2

Referring back to the problems introduced in Example 1, some of the terminology you are going to meet when designing and implementing surveys can be illustrated.

	Problem 1	Problem 2
Population	Farmland in area reporting increased elephant damage	Maize-growing areas in western Kenya
Unit	Village	2m x 2m quadrat
Sampling scheme	30 villages selected at random	10 villages selected at random 10 fields selected at random in each village 2 quadrats per field, placed 1/3 and 2/3 of the way across the field from the entrance.
Measurement tools	Questionnaire for village meeting Visual assessment of damage	Counts of *Striga* plants visible in a quadrat Visual assessment of soil erosion in the field

Setting up a survey

An effective survey encompasses many activities, which must all come together to provide a useful and timely report. The actual planning for a survey is as important as its implementation, and the amount of work involved in the planning should not be underestimated. The efficient and successful management of a survey depends to a great extent on a thorough understanding of the population, of the survey topic, and on having well structured administrative backup available throughout. Available resources will often dictate planning decisions, but it is essential to aim to maintain the quality of all procedures by adopting a 'global' viewpoint, i.e., **by considering the impact of each decision made at a particular stage, on the whole project, thereby achieving balance and consistency throughout.** Some examples of surveys have been given. Next we look at the details of survey design and implementation.

Planning the survey

First and foremost you should specify the objectives of the survey.

- What should be accomplished by the survey?
- What should be measured in order to reach your goals?
- What is the analysis plan for the measured variables?

Ask yourself such questions as **'Why do we need to collect this data?'** Many surveys are multi-purpose, information on more than one topic will be collected, and you need to have some ideas of the precision required in the various areas to be studied. **Define the objectives as simply as possible and ensure they are not self-contradictory,** e.g., will analysis of one variable confuse or assist in the understanding of another? The target and study populations (see later) need to be carefully defined in conjunction with those population characteristics to be studied.

Familiarise yourself with all possible sources of existing knowledge from previous studies. Such information can be used not only to identify gaps and thus emphasise the need for the present study, but also to provide checks on possible sources of bias, to help avoid duplicating work already competently carried out, or to improve estimates previously obtained. It is also important to identify all possible secondary users, i.e., those who may have use for your data in the future. Such users can be of great help with planning, avoiding conflict, and suggesting alternative approaches. You may also be in the situation where your research project is but a small part of some on-going larger research project – in this case it is essential to:

- Maintain contact with those implementing the larger project
- Receive information about results being obtained from other sections of the project
- Ensure your project fits in with the overall larger objectives
- Provide timely feedback on your progress to all other players
- Work with others as part of the larger team.

The flow chart shown in Figure 1 illustrates the phases of a survey, each of which needs careful planning right from the beginning.

Right from the beginning, it is essential to:

- Be aware of all resource limitations
- Be able to identify, for each task:
  - Who is going to be responsible
  - How much it will cost
  - How much time it will take.

Surveys involve large amounts of documentation, all of which have to be prepared in advance and tested for ease of usage. Sometimes you will need to recruit persons who can assist you at one stage or another.

Timetables and budgets

Organising the timetable and fixing the budget are major components of survey preparation. The time and funding that are available are the major factors determining the scope and extent of your study. **It is extremely useful to use a Gantt chart (e.g., Table 1) for timetabling, since it enables you to maintain an overall view of all that the study will entail.**

There will always be time restrictions to be adhered to in any survey project. It is better to over-estimate, allowing lee-way for unforeseen happenings. It is often at the beginning of the data entry and processing stages that delays occur, but these can be minimised by adequate pre-testing checks in advance. Allow for realistic staff turnover, which may result in delays. Add contingency amounts of time allowing for weather effects, breakdown of equipment, or errors. Identify critical activities which, if delayed, will hold up other activities, and try to foresee possible alternatives. Previous research in the same area can prove useful in the time-planning context since from this you can identify what may have gone wrong before. **Asking the experienced workers is most useful since they have the first-hand experience you wish to know about. It is essential that the data be analysed while it is still relevant; so the report can be published within a realistic time after data collection.** Decide whether an initial overview preceding full analysis will be of benefit and plan accordingly.

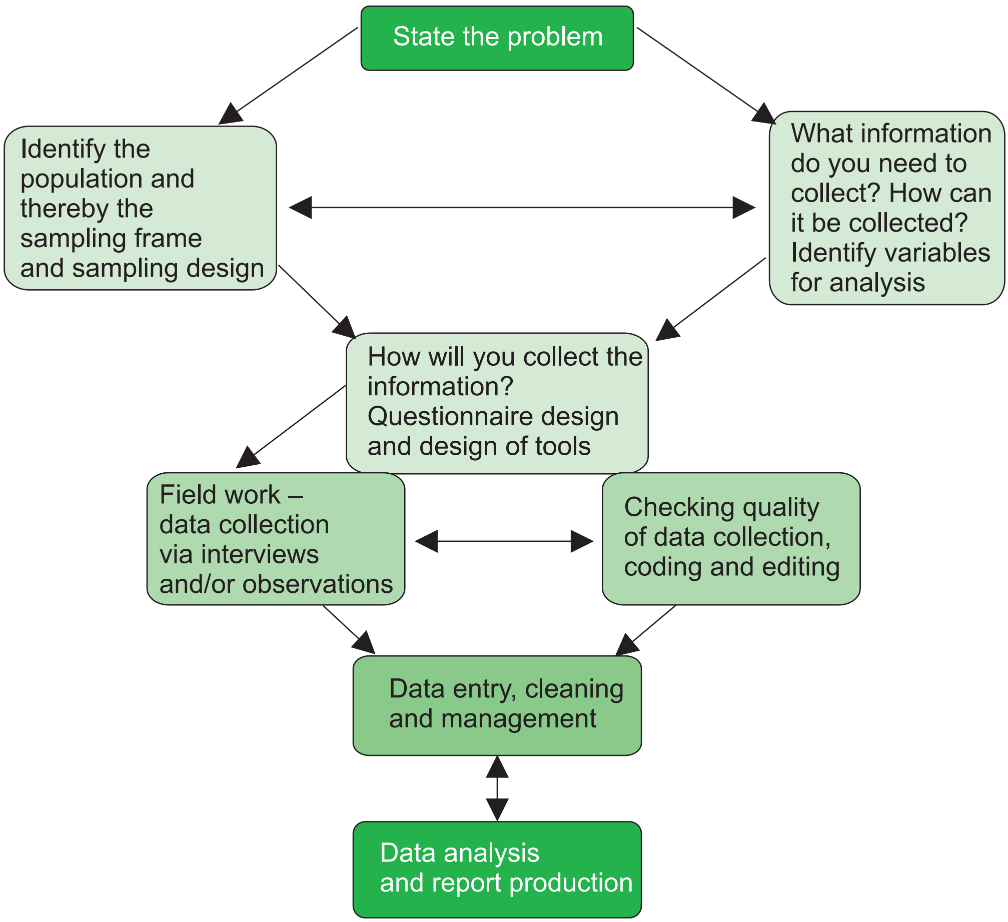


Figure 1. Steps in carrying out a survey

Table 1. Example of a draft timetable for a crop management survey

	Week number																
Task	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**15**	**16**	**17**
Consultations with communities /publicity	•	•				•	•						•	•	•	•	•
Questionnaire design and testing	•	•	•														
Sampling design and sample selection		•	•	•	•	•											
Design of data entry			•	•													
Data analysis planning			•	•	•												
Field staff recruitment		•	•	•													
Training of enumerators and pilot					•	•											
Printing of tools (questionnaire)					•	•											
Fieldwork and checking						•	•	•	•								
Data entry and validation						•		•	•	•	•	•					
Data cleaning and analysis						•					•	•	•	•	•	•	
Production of graphs and tables														•	•	•	
Report preparation					•			•	•	•		•		•	•	•	•
Archiving				•	•											•	•

Source: United Nations (2004)

Budgeting

Hand-in-hand with timetabling for the survey, is the survey budgeting. This is probably the **most difficult task of all since the survey design is totally dependent on the budget**, and *vice versa* – so which comes first?

The main components of a survey budget

- **Publicity and information** – including meetings, agreements, workshops
- **Wages and salaries** – including contingency planning for ill-health, adverse weather, inflation, resignations, after-hours working, field allowances
- **Transport and communications** – including phone, fax, postage, and e-mail usage, fuel, hire charges, bus fares
- **Meals and accommodation**
- **Equipment and consumables** – including hardware and software, printing equipment, clip boards and note books, maps, files
- **Printing and duplicating** – a major component of the budget
- **Hidden costs** – equipment usage.

Errors

A survey requires and combines the techniques of sampling, design of tools, data collection and data analysis, and **the accuracy of the methods employed will determine the quality of the information finally produced**. In any survey there are many potential sources of error which may be broadly classified as sampling errors and non-sampling errors.

Sampling errors. These are errors arising because, by chance, the sample is not fully representative of the population. Such errors can be estimated and are a random result of the sampling procedures. Broadly speaking, the larger the sample size, the smaller the sampling errors.

Non-sampling errors. This category includes all of those errors which can arise from other sources:

- Variation between data-collection personnel
- Inadequate tools
- Inadequate sampling frame
- Data-entry errors
- Coding errors
- Non-response
- Errors in response
- Effects caused by the way questions are worded.

Each of these can give rise to bias which is often not measurable. **Bias** means that the results based on the survey are not, even on average, the same as those that would have been derived from a total census of the population, but consistently over- or under-estimate quantities.

Increasing the sample size, so as to reduce sampling error, can very well increase non-sampling errors due to resulting poorer-quality enumeration and lower levels of supervision. **Sampling and non-sampling errors and their relative magnitudes must be considered simultaneously when determining sample size.** Often only sampling errors are mentioned since non-sampling errors are usually not measurable, and sometimes unknown. **You must remember that you will be making many measurements on your sample and that the precision of estimates is likely to vary from factor to factor.**

Sampling

The theory of sampling is covered adequately in many texts and only some brief notes are made here. There are two inter-related decisions to make: the type of sampling and the sample size. **Decisions** depend on blending the theoretically optimal with what is really practical, in the light of the survey objectives.

Decisions about sample size must be taken in the global context of the project and must include consideration of the following factors:

- Available resources
- Objectives of the study
- Sub-groups, within a population, that you wish to study
- Practical constraints
- The precision needed
- Homogeneity of the population.

Following are definitions and examples illustrating them.

The population

The **target population** is defined as all those units in which you are interested. The **study population** is defined as all those units that you can reliably identify. Ideally these two populations should coincide but, unfortunately, this is often not the case, particularly when the population consists of people.

Units

When you implement your survey, you are going to be dealing with a **unit,** i.e., you are going to interview a **person**, or generate discussions with a **group of people**, or count the number of *Striga* plants in a **quadrat within a field.** These are the units of study. In many studies there is a hierarchical arrangement of units. We measure things on people, but also record something about the household they are in, the village in which the household is found, and the district where the village is located. This hierarchy may be used in sampling, even if measurements are taken at only one level.

Sampling frame

The **sampling frame** is a 'list' of all the items from which you are going to select your sample, noting that you need a separate frame for each level in the hierarchy of units. Careful construction of the frame is needed since, as mentioned above, unexpected errors can easily arise if the frame is out of date, or if it has inaccurate or duplicate records, and so on. If the frame is inadequate we say it exhibits **over-coverage** or **under-coverage** – these terms simply reflecting the non-match of study and target populations (Figure 2). If you are sampling households, then the ideal frame is a list of all the households. If sampling fields or rivers, for example, the sampling frame may be a map or aerial photo, i.e., an implicit list.

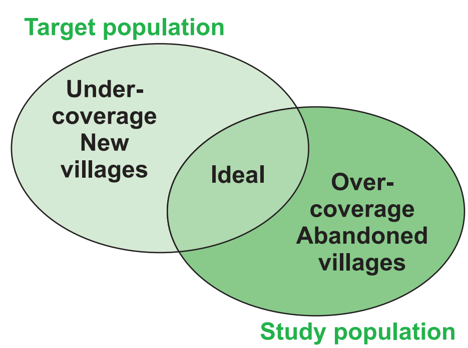


Figure 2. Target and study populations

Example 3

Refer back to Problem 1 in the previous examples. Not only do you want to observe and measure the actual damage in the fields, but you also will wish to interview the villagers and discuss with them their methods for protecting their crops. Another aspect of interest will be gender differences in managing crop damage. Suppose the district authorities provide you with a map on which the locations of villages in the study area are marked. Your first stage of sampling will be to select villages and thus your **target population** is all villages in the area, whilst your **study population** is all villages marked on the map provided to you. If the map is out of date it may mark a village which, no longer exists because its inhabitants moved out to another area 2 years ago precisely because of high rates of crop damage. We call this **over-coverage** since that village would potentially be selected into the sample (according to the map) and yet it does not really exist. Conversely, if a new village has been formed, with some inhabitants of one village moving away and making it their own new settlement area, then this village may not be marked on the map at all and so will not be available for selection into the sample. We call this **under-coverage** since that village is not (and yet should be) available for sampling. Both of these situations will give rise to **non-sampling errors** that cannot be measured and you may never know they exist.

Different approaches to sampling

There are various approaches to sampling and each survey will entail its own unique sampling design. Sampling texts will provide you with the theoretical details and what is aimed for here is to provide a 'feel' for knowing which approach to select. Firstly, two ways of selecting a sample are described.

The simplest type of sampling is known as **simple random sampling**. This involves assigning a unique identification (ID) (e.g., a number) to each item in the sampling frame and then randomly selecting the number of units you require, e.g., numbered pieces of paper placed in a hat and randomly selected one after another. Another very useful method is **systematic sampling**. Here the sampling units are listed in some order that bears no relationship to the topic under study, e.g., listing names in alphabetical order. A starting point is then randomly chosen, and thereafter the sample is determined using what is called the **sampling interval**. This method is often applied to a situation when you have a map or a grid of an area, which can be sectioned into **cells** each of which is then numbered and a systematic sample is easily selected. This latter approach is often referred to as **sampling in space.**

Next, there are a number of ways of classifying the population in different ways before carrying out the sampling, whereby aiming to make use of existing knowledge of the population to ensure the sample is an adequate representation of that population.

Stratification of the population is an approach used when the population is heterogeneous and can be subdivided into homogeneous sub-populations, each of which will be of interest in themselves. Random samples are then selected from each sub-population or strata. In **cluster sampling** the population is divided into clusters that are groups of sampling units which are not similar and one cluster can exhibit the whole range of variability of the population. Using simple cluster sampling, divide the population into clusters, then select a random sample of clusters and investigate each study unit in each selected cluster. In **multi-stage cluster sampling** you can select a sample of clusters but then, within each cluster, further select a random sample of study units.

Clearly it is often useful to classify the population in more than one way, and thus you can use techniques of stratification and clustering together. The final selection of the units you are going to study is usually done using either simple random sampling or systematic sampling. The following examples will clarify these notions.

Example 4

Referring back to Example 3 – recall we have a map (hopefully up-to-date) showing the location of villages in the area of interest. As noted before, the villages marked on the map will be the items in the sampling frame for the first stage of sampling. The actual people resident in each selected village, i.e., the households, will represent items in another sampling frame, for a second stage of sampling. Focus for now on this selection of households and recall that you are interested in the gender dimensions (of head of household) of crop protection from animal damage.

Discussions with the district officials and some initial contact with communities in the area will provide you with information about the overall picture of wild animal marauding in the district. You discover that in one area the main proponent of damage are elephants, with lesser damage caused by baboons and jackals, whilst in another part of the district with a different vegetation type, there was apparently an influx of Quelea birds which caused terrific damage during the past month. The remainder of the district suffers little from large animal damage, with only baboons and jackals causing any measurable loss. Obviously, it will be of interest to study the whole district, even though the elephant damage is only restricted to one area – by looking at the whole district you would hope to be able to compare and contrast areas with different levels of elephant damage.

On the basis of the above observations you decide that there should be three strata within the district. Within each stratum, villages can be randomly selected – probably using a systematic sample from the map. This will constitute the first stage of sampling.

The second stage of sampling, that of households within villages, can be approached in a number of ways. Firstly, for each selected village, the village head could be asked to prepare two lists of names of heads of households, a male list and a female list, and a random sample of households can be drawn from each list. Alternatively, you could, with the assistance of the village head, draw up a map showing all households in the village, marking each one as male- or female-headed. Within each group of male and female household heads, each household will be given a number and then, for each gender group, a systematic sample of households can be selected.

Heads of households can be interviewed using a prepared questionnaire to extract demographic details and obtain estimates of crop damage that has occurred in the past two seasons. In addition, focus group discussions can be held with key informants in each village in order to obtain in-depth information and opinions on the issues of crop damage and ways to reduce it.

The process described in Example 4 is called **multistage sampling** – in other words you sample at various stages of the population hierarchy, ensuring that at each stage you select an adequate sample from each sampling frame. **Issues of sample size at each level of sampling will need to be discussed and finalised with someone who understands the theoretical aspects of random sampling.**

Example 5

Now let us take Example 4 further, and address the issue of data collection from the fields and storage places of the villagers. This will constitute a third stage of the multi-stage sampling process. One approach is to use the selected sample of households as a starting point for selection of actual sites for measurements of crop damage. Each selected household will have a number of fields under cultivation. Depending on the size of area under cultivation, it may be sensible to sample areas within fields for exact measurements, or another approach could be to sample whole fields from those under cultivation. Sampling areas within fields can be done by mapping the field, dividing it into plots of the size to be examined, and then randomly selecting a sample of plots or quadrats – this can be easily carried out once the map is drawn up. Additional considerations that must be taken into account when planning this third sampling design include the type of crop planted, direction(s) from which animals invade, location of water sources, and any other factors that may have a bearing on crop damage.

An alternative approach would be to ignore the sample of selected households and begin afresh, requesting the community to draw up a map of all planted fields, including crop and animal access information, location of water sources, etc., as above. Planted areas may then need to be clustered before sampling quadrats within each cluster.

Selection of storage places for recording types of storage and amount of damage can again be approached in several ways.

Sample size

Decisions on sample size depend on a number of factors, including:

- What is required in terms of precision of variables measured?
- Just how much variability is there expected to be in each item to be measured?
- The practicalities – how big a sample can you actually deal with, in terms of both time and resources?
- What sub-groups of the population are really of interest? You need to decide on the sample size for the smallest sub-group of interest to ensure that the sample for this sub-group is adequate for realistic estimation
- Which variable should be used to calculate sample size?

A good way to think about sample size is in terms of obtaining a **confidence interval**, i.e., what width of confidence interval will be acceptable for decision-making, based on the survey results? The width you need will be used to determine the sample size, for each sub-group of the population. Expressing the results in terms of confidence intervals helps in interpreting the results more realistically. **If the confidence interval is too wide then no meaningful conclusions can be made.** As mentioned earlier, **the larger the sample size the narrower the confidence interval** – but increasing the sample size is likely to increase the cost and non-sampling errors. **Managing a large sample survey requires extensive resources and personnel if quality is to be maintained**, and it is only the large agencies who can afford this type of survey. But if the sample size is too small, then once again the quality of the estimates is at stake, and results will not be meaningful.

It is not true that the fraction (*f*) of the population sampled greatly influences the accuracy of the sample. The information in a sample of 50 from a population of 10,000 (f = 50/10000 = 0.5%) is much the same as that in a sample of 50 from a population of 100,000 (f = 0.05%), other things being equal. The sampling fraction is not something to consider when fixing the sample size, and aiming for a 10% sample or a 5% sample is not logical. The only exception to this is when f starts to get large – say over 20%.

You should be constantly aware that **each survey study planned and implemented is a unique case and thus 'standard' sample sizes do not exist. You** should familiarise yourself with previous research, but only use it to provide guidelines for your own study. The sample size used last time may or may not be suitable for the current study and you should make your own considerations and do your own calculations, rather than assuming that those used in a previous study were suitable.

Designing measurement tools

The design of the measurement tools needs to be done in conjunction with:

- Formulating and stating the objectives
- Planning which variables to collect
- Deciding how to analyse the information collected
- Consideration of time and resources available
- Bearing in mind the eventual report to be produced.

It is all too easy to imagine that you will be able to collect vast amounts of information from each unit studied – the reality is that it is usually not possible or desirable. Hand-in-hand with development of measurement tools, must be the preparation of the data analysis plan and drafting the outline of the final report.

For each item of information collected, there should be one or more corresponding sections in the analysis plan.

If your survey involves communications with groups of people, then you have to be aware of the time you are going to demand from them to assist you in your data collection. Even if your data collection only means laying out and measuring quadrats in someone's fields, they are going to need to accompany you to do this and you have to be able to rely on them.

People are busy and, in addition, many other researchers may be demanding their time. Thus deep thought should be given to the design of the data-collection tools and, for each variable selected for study, you have to ask yourself 'What useful information am I going to get from this?'

Questionnaires

When communicating with people, either via a structured interview or via focus group discussions, or by any other means, it is wise to lay out a questionnaire ahead of time and to know in advance the type of answers you can expect from each question. Questionnaire design is extremely important – when you are interviewing people, you are assuming that:

- Everyone has the same understanding of each question
- Each question does have an answer
- Each question can be relatively easily answered
- Each question should be relevant to your study
- The question is not 'leading' the respondent towards a particular answer.

Remember that sensitive questions can upset people, which will lead to inaccurate information being provided. Good questionnaire design can only come with experience and it is wise to always ask for assistance.

Questions can be classified as **open** or **closed.** An **open question** is one for which any answer is accepted and recorded in full. A **closed question** is one in which you supply pre-determined **response categories** into which each and every response should fit. Thus, **response categories** should be:

- Non-overlapping, i.e., mutually exclusive
- Exhaustive
- Permit an overview of the situation
- Neither too many nor too few
- Placed in a logical order.

Open questions provide more information than do closed questions but they are correspondingly harder to analyse and **wherever possible it is best to use closed questions. Focus group discussions (using open questions) are extremely useful for finding out general information and situations on the ground.**

Finally, remember that you should place your questions in a sensible and logical order so that the interview/discussion will flow.

Other data-collection tools

If your survey involves measuring one or more items you will need to prepare for this in advance of data collection. Usually, you should keep in mind the data entry format you will eventually use, since if the formats for both are similar it makes the data entry easier and less prone to errors. Thus, you should design a **spreadsheet** which can be imitated in **data-entry format**, containing clearly defined rows and columns in which values can be recorded, and including a column for comments that will remind you of the circumstances of the collection at data entry time. As with questionnaires, **it is important to collect only the information you really need and which you can really use.** Site details – date, place, time, methods, personnel – can be coded, but must be part of each record. **It is also essential to draw up a protocol for data collection** – this will be a series of detailed instructions and a description of how to actually go about collecting the information required, e.g., how to lay out the quadrats and how to count *Striga* within each quadrat. **The purpose of the protocol is to ensure consistency in data collection methods and implementation**, particularly if more than one person is to be involved in the exercise.

Testing the tools

Once you have designed your basic data-collection tools, the next step is to test them. We call this first testing exercise a **pre-test**, and it serves not only to see whether the tools are suitable, but also to gauge the responses to be expected and thereby to refine, adjust, and further develop the tools. This testing should be carried out using a small sample of units that will not be involved in the main survey. Those people testing the tools should be experienced researchers so that they can react properly to needs which will be highlighted.

After pre-testing it is 'back to the drawing board' again to prepare the final draft of the tools, and the final draft of the data-analysis plan. Mobilising resources and personnel will also be undertaken during this time, until finally you are ready to conduct the **pilot study.** The purpose of the pilot study is to test not only the measurement tools, but also to act as part of the training for personnel, and to test the data entry and data analysis plans. Usually the pilot study is carried out in the same area as where the intended study will take place, using a sample (from each sub-group of the population) that will not be involved in the final study – thus **the sampling design must be completed and known prior to the pilot study.** All personnel to be involved in the final study will also be involved in the pilot study – in some ways it is like a 'mock' study of all procedures. When the pilot is complete you can finalise the measurement tools and reproduce them in bulk as required.

The time between the pilot study and the main study should be as short as possible so that all personnel remain in the correct frame of mind for the main study.

Fieldwork and data collection

The pilot study is part of the preparation for fieldwork. Once the pilot study is complete and the tools finalised and reproduced, you should start the main study as quickly as possible. The training of personnel should be thought of as an on-going exercise – before, during and after the pilot study personnel will be becoming more and more familiar with the measurement tools and all the needs of the survey. During the initial start-up period of the survey it is wise to meet with all personnel on a daily basis so as to maintain standardisation of data collection. The team leaders should be moving from one person to another to:

- Check that each is collecting the information in the required manner
- Check through collected data
- Pass on the completed forms for further checking
- Code and data entry
- Liaise with other team leaders.

Once team leaders are satisfied that their members are acting as expected, the teams can disperse, but the team leaders should continue close monitoring and liaison with each other.

Data management

A survey generates a huge amount of data and thus it is essential to be absolutely organised for every aspect. Data collection forms should bear unique ID numbers which, by means of codes, will enable the data manager to know exactly where that data was collected, and by whom. The team leaders should check each completed form in the field and, if there are problems, the person who collected the information will have to return to the site and repeat the process. Once the team leader is satisfied with a form, he/she will pass it on to the data manager. If any coding is to be done it is now that it should occur – for instance, categorisation and consequent coding of the content of open questions can take place at this time. Thereafter the form is ready for data entry. Often data will be entered twice – **double data entry** – an approach that is recommended since it nullifies errors of entry – many statistical packages offer this facility. Those doing data entry should have been involved in the planning so that they are aware of the survey objectives, familiar with the measurement tools, and thus in a position to spot inconsistencies and/or errors on the data forms – in this way **cleaning** of the data begins even at the data-entry stage.

Full-scale cleaning of the data usually takes place once all data has been entered and the data files merged into one. **Cleaning** involves examining each variable in turn, looking for **outlier values** and **inconsistencies**, particularly in respect of other variables that provide complementary information. Once the data is pronounced clean the data analysis plan can be put into action and results obtained for input into the final report. Additional **recoding** may take place during the data analysis, e.g., merging categories of responses for more realistic analysis.

Data storage

The importance of **backing up** your data files cannot be emphasised too often. At least three copies of each of the following files should be kept, preferably on CD's

- Original data entry files, obtained before cleaning
- Cleaned data files
- On-going analysis files
- Records of comments on data collection
- Records of progress on data collection
- Records of coding and recoding
- Tables and other results of the analysis plan
- Reports.

All of this information will, eventually, feed into the data archive that should be set up on completion of the survey.

Reporting

The survey report should follow the phases of the survey. **Each phase should be reported upon fully, including both good and bad aspects.** Full details of measurement instruments, training instructions, field reports, coding procedures, cleaning procedures, and the data analysis plan, should be available as appendices to the main report. **Don't forget to report on the non-sampling errors!**

Resource material and references

Appendix 11. ICRAF. 2003. *Genstat Discovery Edition and Other Resources.* World Agroforestry Centre (ICRAF), Nairobi, Kenya. On CD.

Alreck, P.L. and Settle, R.B. 2004. *The Survey Research Handbook.* Third edition. Irwin-McGraw Hill, Burr Ridge, Illinois, USA.

Ballar, B.A. and Lanphier, C.M. 1978. *Development of Survey Methods to Assess Survey Practices.* American Statistical Association, Washington DC, USA.

Belson, W.A. 1981. *The Design and Understanding of Survey Questions.* Gower, Aldershot, UK.

Casley, D.J. and Kumar, K. 1988. *The Collection and Use of Monitoring and Evaluation Data.* IBIRD/World Bank, Washington DC, USA.

Casley, D.J. and Lury, D.A. 1993. *Data Collection in Developing Countries.* Oxford University Press, Oxford, UK.

Errington, A.J. 1985. Sampling frames for farm surveys. *Journal of Agricultural Economics* 36(2): 251–259.

Ferber, R. 1980. *What is a Survey.* American Statistical Association, Washington DC, USA.

Hayek, L.C. and Buzas, M.A. 1997. *Surveying Natural Populations.* Colombia University Press, New York, USA.

Kish, L. 1987. *Statistical Design for Research.* John Wiley & Sons, New York, USA.

Lohr, S.I. 1999. *Sampling Design and Analysis.* Duxbury Press, Pacific Grove, California, USA.

Moser, C.A. and Kalton, G. 1985. *Survey Methods in Social Investigation.* Gower, Aldershot, UK.

Oppenheim, A.N. 1978. *Questionnaire Design and Attitude Measurements.* Heineman Educational Books, London, UK.

Scheaffer, R.L., Mendenhall III, W. and Ott, R.L. 1996. *Elementary Survey Sampling*. Duxbury Press, Kent, UK.

Thompson, S.K. 2002. *Sampling*. Second edition. John Wiley & Sons, New York, USA.

United Nations. 1964. *Recommendations for the Preparation of Sample Survey Reports* (Provisional issue), Statistical Papers, Series C, ST/STAT/SER,C/1/Rev.2, New York, USA.

United Nations. 1984. *Handbook of Household Surveys* (Revised edition), Series F, No. 31. New York, USA.

United Nations. 2004. *Technical Report: Household Surveys in Developing and Transition Countries: Design, Implementation and Analysis* (Section D, Chapter 14) (in preparation) ST/ESA/STAT/AC.85/www. New York, USA.

Internet resources

www.reading.ac.uk/ssc/develop

http://unstats.un.org/unsd/hhsurveys/index.htm

4.5 Measurements

Jane Poole

- **Measurements generate the primary data in your study, whether it is a survey or experiment**
- **You will have to measure not only the primary quantities that meet your objectives, but those data that help explain and qualify them**
- **There are always alterative ways of measuring anything. Choose the method that best meets your objectives while being practically feasible**
- **Pay attention to quality control: careless measurement can jeopardise the whole study**

Introduction

Measurement is a general term that encompasses many types of data collection. Measurements may be numbers that a scientist collects, such as yields of a crop in a field trial. But they can also be notes made of a farmer group discussion, climatic data provided by a local meteorological station, or responses to survey and interviewing questions.

Every aspect of your research study needs careful design. This includes choosing what measurements to take, when to take them and why. You must also consider how to measure and how much to measure. For large trials and surveys it may be necessary to delegate data-collection to other scientists, local extension officers or farmer representatives – you will need to decide who takes the measurements. This chapter provides general guidance on how to make these choices and highlights important issues to be considered.

What are measurements?

Measurements generate the data you need for your research. You require these data and their analyses to make your research conclusions. There are many different types of measurements and your choice of which to use will depend on the objectives of the study, and on other details of the design. The measurements needed will also determine some aspects of the design.

The following are examples of measurements that may be taken for different types of agricultural research. These examples are just a small selection of the hundreds of possible measurements you could take.

- **Laboratory trials** – chemical properties of soil and water samples, pathogen growth on petri dishes, insect mating and offspring production, eating routines of insect pests
- **On-station and on-farm field trials** – plant heights, insect pest and disease levels, crop and biomass yields, root damage of plants, farmer-participatory evaluation of varieties, labour requirements
- **Participatory research** – farmer group characteristics, farmer perception of new technologies, farmer evaluation of on-station demonstration trials
- **Biophysical surveys** – site location and characteristics, plant varieties, crop management, scientist-evaluated disease infection levels, farmers' perception of disease infection levels

- **Socio-economic surveys** – site location and characteristics, household and farmer characteristics, farmers' perception of crop management practices, farm labour information
- **General/environmental measurements** – weather data (rainfall, temperature), soil type and properties.

Types of measurement

Qualitative and quantitative

Both qualitative (farmer opinions of new technologies) and quantitative (crop yields) data require measurements. **Quantitative** measurements are necessary for many analyses and interpretations. **Qualitative** data can often add insights and explanations that are hard to capture in numbers. The distinction between the two is not always clear. Qualitative data (farmer reasons for crop failure) can be quantified after coding, (e.g., by noting whether or not 'drought' is given as reason for crop failure and then reporting the proportion of farmers who give different coded answers such as the proportion believing 'drought' to be the reason).

Example 1

An on-station researcher-managed trial was conducted to investigate sorghum varietal resistance to stem borers. Quantitative measurements were taken of the number of stem borers in the stems, stand count and crop yields. Local farmers were then invited to the station to view the different treatments and group discussions were held to elicit farmers' opinions on the performances of the varieties. These additional qualitative data provided the researchers with information about: characteristics farmers found important, the opportunities for transferring the experiments on-farm, and the likelihood for farmer uptake of the most resistant varieties.

Repeated measures

Measurements taken on the same unit (plant, plot, household) repeatedly during a study are called '**repeated measures**'. These type of data are frequently used in laboratory and field trials, e.g., plant disease levels estimated every week, the growth of a fungal pathogen on a petri dish measured every 3 days.

When you are collecting repeated measures how often should you collect the data? In some cases the answer to this question is simple, as when data are required after chemical spraying or rain, then the occasions are defined. In other instances it is up to you to decide how frequently to measure.

General guidelines when choosing the number of repeated measures to take:

- If you want to fit a (growth) curve to your data then 4–5 time points are usually sufficient
- When you don't know which time points will give you information (plant disease levels in a field trial may stay constant for some time) then you may need to take measurements regularly (once a week). Note that for the plant disease example there is no point taking measurements at the start of the trial if there is no disease present. In this case you should be checking the site regularly and then start taking measurements when the disease starts to appear, otherwise you will spend a lot of time collecting a lot of zeros!
- It is not essential that the observations be taken at equal time intervals. However, it is important to record details of each time point so that the patterns observed can be accurately plotted (time plotted on the x-axis on the correct scale).

Destructive and non-destructive

An important option to consider when taking laboratory and field trial measurements is whether they are to be made **destructively**, as when a plant is cut down to measure root sizes, or **non-destructively,** as when you simply measure plant height. If it is your trial, make sure destructive measurements will not disturb further observations. If you intend taking destructive measurements in farmers' fields make sure they understand and agree. The 'destructive' option is not usually applicable to socio-economic surveys and participatory methods of research!

Example 2

A researcher wishes to measure above-ground biomass in an agroforestry trial, over a period of 3 years. The plot size is set at 10m x 15m. He/she has several measurement options for evaluating the amount of biomass, some 'destructive' and others 'non-destructive'. What measurement(s) could he/she take? Some options are in Table 1.

Table 1. Options for measuring biomass in an experiment

Measurement	Advantages	Disadvantages
Destructively sample a few plants per plot at regular intervals	• Collect large amounts of data on biomass production	• Lower precision of yield estimates (increase plot size to overcome this) • If plant size within a plot is highly variable then a large sample is needed for a precise estimate of biomass • Time requirements are high
Destructively sample a few plants from the guard rows at regular intervals	• Collect large amounts of data on biomass production	• Guard rows may not be representative of the plot • Over time the guard rows will lose their ability to 'protect' the crop • Time requirements are high
Destructively harvest the whole plot at the end of the experiment only	• Time requirement is low • Does not require extra plot area	• Have no idea of the biomass production over the 3-year time period
Record the plant heights at regular intervals and harvest the whole plot at the end of the experiment	• Does not require extra plot area • Large sample (whole plot) can be measured • Plants can be followed over time	• The height measurements may not be representative of the biomass yields
Record the plant heights at regular intervals. A sample of plants grown close to the trial are harvested regularly	• The harvest measurements from neighbouring plants can be used to calibrate the non-destructive measurements	• Requires a lot of experience to correctly calibrate the measurements

Bulked samples

Some variables, like the soil samples and chemical properties can be measured by **bulking** together samples collected in the plot (or laboratory, site, etc.). You take N samples from a plot/location and mix them together to form a single composite sample. M sub-samples are then extracted from the composite mixture and measurements taken for each. Things to note about this type of measurement:

- The variation you observe between the M sub-samples is due to measurement error and/or poor mixing. It has nothing to do with the variation in the plot
- The closeness of the measured values to the plot value will depend on how close the value in the bulked sample is to the plot value. This is determined by the N field samples. The more samples you bulk together (i.e., N is large) the more representative of the site your composite mixture will be
- If the N field samples are highly variable, or collected in a way that introduces bias (e.g., all samples taken from one corner of the plot), then increasing the number of sub-samples you take (M) will not help.

Think carefully about the information you really need. Do you want to know how soil P, for example, varies between different samples from the same plot or how it varies between different plots? If you only need the latter then maybe M can be 1, but N may still have to be large to make sure the bulked sample really represents the whole plot.

What measurements to take and why?

Your choice of measurements (the what, when, why, how and how much, who to measure?) depends primarily on your **research objectives.** You must ask 'What data do I need to collect and analyse in order to achieve my objectives?' Careful consideration of your detailed objectives, together with the practicalities of measurement – this means the resource availability, will assist you to collect the relevant data. **Researchers often take measurements that will not help them to answer their objectives and/or take measurements which duplicate the information. This is usually because they have not given sufficient consideration to the data they need.** Collecting data you don't really need is a waste of time, and in some cases is even unethical (for example, in a household survey in which you take up the householder's time). Failing to collect data you do need will mean you can not achieve your objectives.

Measurements may be **primary responses** that are central to answering your research objectives or **variables** that help to explain them. Examples of primary responses may be crop yields (Objective: compare yields under different management methods) or disease infection levels (Objective: map the geographical distribution of the disease in a region). Primary responses are usually highly variable, with variation at every level of the design hierarchy. You therefore need to investigate the reasons for this variation and may also want to make comparisons with similar research. In the second example above the researcher collects additional data on potential sources of variation such as soil type, climatic conditions and crop management. The following are examples of measurements that may be relevant to specific research objectives.

On-station field trial

Objective – Evaluate the effects of 5 treatments on cabbage crop aphid numbers. (Detail – treatments include chemical, biological and un-treated control, 4 blocks). (Table 2)

Table 2. Suggested measurements in cabbage experiment

Primary response measurement(s)	Additional variable(s)
Plant aphid counts (every 7 days)	Yield at harvesting (to investigate the aphids' effect on yields)
	Rainfall and temperature (changes in climatic conditions may affect aphid numbers, how do these relate to the treatment effects?)
	Soil fertility measurements

Socio-economic survey

Objective – Investigate farmer perceptions of the impact of *Striga* on their maize yields. (Detail – 200 farmers interviewed in one district of Kenya).

NB. This survey could be combined with a 'researcher observed' level of *Striga* to compare farmer perception to the actual levels of infection.

- Look carefully at your research objectives
- What are the primary response measurements you should take so that you can answer your objectives?
- What are the additional variables you could measure (Table 3) that will help you to explain the patterns you observe and enable you to compare your research to similar work?

Table 3. Suggested measurements in *Striga* survey

Primary response measurement(s)	Additional variable(s)
Farmer perception of *Striga* levels	Maize management methods (that may affect levels of infection)
Farmer perception of yield loss due to *Striga* infection	Importance of maize to farmer's livelihood (looks at the impact of the perceived *Striga* levels)

So, your measurement options are determined by your research objectives. But often you will find that several different measurements could be used to answer the objectives so how do you decide which ones to use, without duplicating the information? The answer to this question depends on your research design, available resources, and practical considerations.

Research design

Almost all experimental designs have more than one level of hierarchy (villages/ farms/fields/ plots, or plot/row/plant/leaf) and you have to decide what measurements to take at each level. Different quantities should be measured at different levels of the hierarchy, for example, the wealth of a farmer is usually measured at the household level, the crop yield may be assessed for each plot, and tree height has to be measured on individual trees. Other variables may be measured at higher levels, for example, discussions with a farmer group will generate village-level variables.

The type of research you are doing also determines which measurements are appropriate. For a researcher-designed and managed trial it makes sense to take measurements on every plot and location. In a farmer-designed and managed trial measurements may only be taken on some plots. In a farmer-designed and managed varietal trial the objectives require crop yields to be measured. However, on some farms the level of crop management was very low and weeds greatly reduced yields. Yield measurements were taken on the sub-set of well managed farms and conclusions applied to this environment. The reasons for varying management input were recorded on all farms to explain the differences between the well managed and poorly managed sites. In this example measuring the yields on poorly managed plots would not have provided the information necessary to explain varietal differences.

Research resources

The type and number of measurements you can take will depend on the resources available in terms of time, money, and human resources.

It is often not possible to take as many measurements as you would like due to a lack of these resources. So, should you take small samples of many different types of measurements or fewer types with more samples? The answer depends on how precisely (i.e., the size of measurement error) you want to evaluate each type of variation. It is often possible to simplify your measurements, by using indicators and proxies, so that a larger sample can be measured. Review the following two situations and decide which measurement option you would take.

Situation 1

Conduct a biophysical survey to evaluate the levels of coffee berry disease in five coffee-growing districts. You have enough resources to sample 1000 trees. The majority of farms have around 200 trees and there are approximately 500 farms in each district (Table 4).

As an alternative to this option you could increase the number of farms to 20 and decrease the trees sampled per farm to 10 – thereby increasing the precision at the district level but decreasing precision at the farm level.

Table 4. Measurment options in coffee survey

Measurement options	Gain/loss considerations
Sample (all) 200 trees on each farm Visit 1 farm in each district	A precise estimate of disease level on each farm but you only have one observation per district and therefore no idea of variation within the districts
Sample 20 trees on each farm Visit 10 farms in each district	Estimation of variation within each farm and also within each district. Comparison of the two is also possible
Sample 1 tree on each farm Visit 200 farms in each district	A good estimate of disease levels within each district but no idea of variation within each farm

Situation 2

Carry out a farmer-managed experiment to evaluate the yield potential of 4 sorghum varieties. You have 50 farmers who are willing to participate in the trial, but you are the only scientist on the project. The crop matures on all farms in the same 2 weeks (Table 5).

Table 5. Measurement options in sorghum variety trial

Measurement options	Gain/loss considerations
Visit every farm and carry out the harvesting yourself, avoiding the edges of plots, taking into account damaged plants and gaps in the plot, etc.	1. Time – you don't have enough of it!! 2. Should the researcher control the harvesting of a farmer-managed trial? 3. The assistance provided by you may give a bias to the farmers' perception of the varieties 4. Does harvesting the whole plot at the same time (which you would have to do) accurately reflect the actions of a farmer?
Take proxy measurements such as stand count and height prior to harvest	1. Requires less of your time and can be carried out before harvest 2. May not always be a good proxy for the crop yield
Ask farmers to harvest their own plots and provide you with sorghum yields for each plot, in kg/plot or as a score such as, 'poor' to 'excellent'	1. There is less time needed for you to interview the farmers about their yields and perceptions 2. Farmers maintain 'ownership' of the trial and the trial remains 'farmer managed' 3. You do not obtain a precise researcher-controlled crop yield, although you can use the farmer evaluations to answer the objectives of the trial

- What measurements do you need to take at each level of your design hierarchy?
- What resources do you have for your research – in terms of time, money and labour?
- What use of resources will give you the highest precision for your most important measurements?

When to take measurements?

One way to categorise your measurements is using the 'Before – During – After' approach. In any research project, experiment or survey, you can take measurements at each of these three stages.

Before

Measurements taken at the start of your research can:

- Provide you with a baseline for your experimentation, e.g., soil fertility measurements of a field-trial site
- Be used to characterise the plot/farm, e.g., wealth categorisation of farmers prior to their participation in an on-farm 'uptake of technology' trial
- Assist with your design, e.g., characteristics of regional farming population used to select a representative sample for participatory work.

During

You want to collect data on 'interim' responses whilst conducting your research:

- Common measurements on crop trials include plant stand and height. Other measurements may include labour use for different operations, insect pest and disease levels etc.
- You might opt to include the use of participatory research tools, e.g., participatory rural appraisal (PRA) during on-farm experimentation
- Whilst evaluating the use of agricultural information centres or extension offices you may choose to record the daily attendance numbers.

After

Towards the end of your research there may be follow-up measurements that can help you to complete your understanding of the results:

- You could measure soil fertility levels at the end of an on-station field trial, or farmer perceptions and technology uptake at the end of an on-farm trial, for instance, do they choose to continue using one of the tested technologies?
- Data could be collected to demonstrate the impact of your research, by comparing to your baseline data.

Data collection and quality control

Taking all measurements yourself will help to maintain high data quality but may not be possible if you are collecting data over many plots, farms and locations. If enumerators are used then it is harder to ensure they are using common methods, and you must monitor their performance and follow-up on difficulties and questionable data. Problems are especially likely to arise when carrying out socio-economic surveys or using PRA methods as these require considerable interviewing skills such as probing, performing arithmetic calculations to confirm responses are reasonable, and assessing the attitudes of the farmer. In this type of research it may be better to keep the sample size smaller and conduct the interviews yourself.

Field trials

- Ensure data collectors are trained in how to take each of the measurements. Give your enumerators a demonstration of the data-collection methods
- Monitor enumerators' performance by observing at least some of the data collection
- Check through the data as soon as it is given to you and follow-up on any problems with the enumerator immediately – before their memory fades
- Remember to take photographs of significant results or events, and label them carefully for later reference.

Surveys, interviews and participatory research

- Interviews and surveys should be kept as short as possible. If your questionnaire has, for example, 50 questions then it means you are collecting a lot of irrelevant information, repeating measurements and are not focused on your research objectives!
- When conducting farmer interviews for on-farm trials, socio-economic or biophysical surveys and other participatory research it is preferable to conduct the interviews in the field. Farmers will find it easier to evaluate and quantify their observations if they can see the crop/trial in front of them
- If you are using enumerators to assist in data collection then they must be appropriately trained. Meetings to discuss the survey measurements are useful so that you can agree on common methods of data collection. To standardise the perceptions/abilities of the enumerators you should carry out a few 'mock' interviews with farmers
- As with field trials, check through the data as soon as it is given to you and follow-up on any problems with the enumerator immediately.

Even if you have other scientists, technicians and enumerators working on your study you should be spending time in the field collecting data. Only by being personally involved can you monitor the quality and understand difficulties and things that are not going as planned. You will also gain insights into the problem that you would not get simply by analysing data that someone else collected.

Record ancillary observations, comments and notes along with all your planned observations. These can be anything. Include comments on data such as, 'plot 17 did not look so well weeded as the others' or, 'Mrs Njoroge was recovering from malaria when we conducted the interview'. Include notes on things to follow up, such as, 'I could see trees on the hill top but no one mentioned these in the interview' or, 'there were many more bees on the local varieties than on the introductions'. Include ideas that occur to you as you spend time in the field – 'Maybe we have to sort out insect damage before soil fertility will make any difference'. All these things will add to your interpretation and real understanding of the research you are doing. When researchers used pencil and paper exclusively their field books were usually full of such notes. Now computers are used to capture data, often all you will find are files full of numbers, with no marginal notes and comments. This is not very useful. Such information can easily be recoded in your computer, or you may still use a notebook. But if you do not write down the notes and comments when they occur to you, you will not remember them.

Photographs can also be an important record of the ancillary information. You can use them to illustrate the points you are making in your report and presentations. They will also help you recall the field situation, allowing more relevant and effective analysis and interpretation of numerical data.

How much data should you collect?

Often too much data are collected, just in case it might be useful. Researchers believe some measurements require little cost and therefore are worth taking. However, even if you have the necessary amount of staff/labour time some 'costs' may remain hidden. For example, staff asked to collect frequent growth data, which they know are rarely used in the final analysis, may not pay sufficient attention to data quality. Additionally, computer entry of the data may take up too much of the researcher's time. **When large volumes of data are collected it is often because staff have not yet given sufficient thought to the analysis.** Careful thought about what measurements are really needed to answer the objectives of the research should help to avoid this pitfall.

Measurement tools and equipment

The tools available for taking measurements are varied and you should use the ones most appropriate to the objectives and practicalities of your research. It is unlikely that your research will require only one type of tool and it is usually a good idea to combine such rigourous studies as quantitative field trials with PRA-type methods like farmer evaluations.

- For quantitative field data standard tools such as balances and tapes are used. **Ensure that you and/or the enumerators are fully trained in the use of each tool so that data collection is of a consistently high quality**
- When collecting data from farmers other tools will be needed. **Formal individual farmer questionnaires are often used but they are not always the most efficient method of data capture.** They are an intensive method of data collection and resources may limit the number of interviews you can do. Using a range of methods from PRA and other types of social enquiry may provide you with just as much information, for less work. For example, group discussions can be used to record a single consensus measure that is often useful when collecting baseline study data. **Use of PRA methods requires intensive training and quality control of enumerators.** Many PRA methods are designed for open-ended, exploratory enquiry. They can be very useful in more-structured investigations, but take care to use them in a consistent way
- Choose tools most appropriate to your research objectives and the practicalities of your work, i.e., available equipment and resources
- This may mean combining tools used in rigourous studies, such as the tapes for measuring tree heights with PRA-type methods like farmer evaluation of the tree
- Use quality control methods to ensure measurement tools are being used appropriately and accurately.

Resource material and references

There are very few books and papers written specifically to tackle the issue of measurements. 'Survey sampling' and 'Experimental design and analysis' books sometimes contain sections or chapters on data collection methods. There are also subject-specific books with details of measurement methods, some of which are listed below. You may find that scientific or discussion papers covering research topics similar to your work provide the best 'measurement' ideas.

Appendix 8. Presentations and Style – Tips on Photography and Writing. Eric McGaw. On CD.

Appendix 11. ICRAF. 2003. *Genstat Discovery Edition and Other Resources*. World Agroforestry Centre (ICRAF), Nairobi, Kenya. On CD.

Ashby, J.A. 1990. *Evaluating Technology with Farmers: A Handbook*. CIAT Publication no. 187. Centro Internacional de Agricultura Tropical (CIAT), Apartado Aereo 6713, Cali, Colombia. 95 pp.

Ackroyd, S. and Hughes, J.A. 1981. *Data Collection in Context*. Longmans, London, UK. 155 pp.

CIMMYT Economics Program. 1993. *The Adoption of Agricultural Technology: A Guide for Survey Design*. Centro Internacional de Mejoramiento de Maíz y Trigo (CIMMYT), Mexico DF, Mexico. 88 pp. cimmyt@cgiar.org

Coe, R., Franzel, F., Beniest, J. and Barahona, C. 2003. *Designing on-farm participatory experiments*. Resources for trainers. World Agroforestry Centre (ICRAF), Nairobi, Kenya. www.worldagroforestycentre.org

Feldstein, H. and Jiggins, J. 1998. *Tools for the Field: Methodologies Handbook for Gender Analysis in Agriculture*. Kumarian Press, Hartford, Connecticut, USA. 270 pp.

Franzel, S. and Scherr, S.J. 2002. *Trees on the Farm: Assessing the Adoption Potential of Agroforestry Practices in Africa*. CABI, Wallingford, UK. 208 pp.

Lal, R. 1994. *Soil Erosion Research Methods*. St Lucie Press, Florida, USA. 352 pp.

Philip, M.S. 1994. *Measuring Trees and Forests*. CABI, Wallingford, UK. 336 pp.

Spencer, D. 1993. Collecting meaningful data on labour use in on-farm trials. *Experimental Agriculture* 29: 39–46.

Schroth, G. and Sinclair, F.L. 2003. *Trees, Crops and Soil Fertility – Concepts and Research Methods*. CABI, Wallingford, UK. 416 pp.

Internet resources

- Reading Statistical Services Centre (SSC) website (http://www.rdg.ac.uk/ssc/) contains several downloadable booklets and papers. They provide 'easy to read' discussions and advice on various aspects of experimental and survey design and analysis. 'Measurements' are often discussed within these topics.
  Examples of useful information on the SSC website:
  N. Marsland, I.M. Wilson, S. Abeyasekera and U. Kleih (2000). A *Methodological Framework for Combining Qualitative and Quantitative Survey Methods*.

DFID Good Practice Guidelines:

- Guidelines for planning effective surveys
- On-farm Trials – Some Biometric Guidelines
- Centro Internacional de Agricultura Tropical (CIAT) website (www.ciat.cgiar.org)

Online publications:

  Horne, P.M, and Sturr, W.W. (2003). Developing Agricultural Solutions with Smallholder Farmers: How to get started with participatory approaches.
  TSBF Institute of CIAT (2001). Legume Cover Crop and Biomass Transfer Extension Leaflets

- Food and Nutrition Technical Assistance (FANTA) website (www.fantaproject.org) (Downloadable: Agricultural Productivity Indicators Measurement Guide)
- International Livestock Research Institute (ILRI) website (www.ilri.cgiar.org) – check the 'Capacity Strengthening – Training Materials' page for some on-line materials.
- International Institute of Tropical Agriculture (IITA) website (www.iita.cgiar.org)
  On-line publications (a few of these need to be ordered from IITA):

IITA Research Guides:

- IRG2 – Soil sampling and sample preparation
- IRG7 – Use of maps for planning research farms
- IRG9 – Morphology and growth of maize
- IRG18 – Farm records and work planning on agricultural research farms
- IRG31 – Tips for planning formal farm surveys in developing countries
- IRG50 – Socio-economic characterisation of environments and technologies
- TE/124 – A field guide for on-farm experimentation

4.6 Data management

Gerald W. Chege and Peter K. Muraya

- **'Data management' refers to all the steps in looking after and processing your data, from observation in the field until the end of the study, and after**
- **Attention to data management is important to ensure your observations are valid, they can be processed efficiently and will remain available for follow-up analysis at the end of your study**
- **Your project must have a data management strategy that describes procedures and responsibilities**
- **Computing will be an important part of a data management strategy. If your data are simple then spread-sheets may be suitable tools for data management. There are good and bad ways of using spreadsheets**
- **If your data are complex then spreadsheets will not be sufficient and you will need to learn something about database design and use**
- **Misunderstandings over data ownership can damage projects. Make sure all ownership issues are resolved before data are collected**

Introduction

Research work, irrespective of whether experimental or survey type, generates data. Data are the resources used by scientists to make conclusions and discoveries. As in other human activities, if you plan to use resources you need to take care of them, because lack of care may have disastrous effects. For example, a computer file containing medical data collected over a number of years could become corrupted. If there was no other copy elsewhere the total value of the resource would be wiped out.

Data management can be defined as the process of designing data collection instruments, looking after data sheets, entering data into computer files, checking for accuracy, maintaining records of the processing steps, and archiving it for future access. It also includes data ownership and responsibility issues.

Data management is important for the following reasons:

- **To assure data quality.** Since conclusions are based on data, accuracy is paramount and errors resulting from wrong data entry, incorrect methods of conversion and combining numbers must be avoided
- **Documentation and archiving.** Documenting or describing data and archiving it are important so that anybody can make sense out of the volumes of rows and columns of numbers for ongoing research and future use
- **Efficient data processing.** Scientists spend a great deal of time preparing data for analysis. This includes converting data to suitable formats, merging data sitting in different files, and summarising data from field measurements. The time spent in this pre-processing step can be greatly reduced if data are properly managed.

To see why data management is important, it may be worthwhile considering how organisations manage financial and accounting data. Whole departments spend huge resources on tracking transactions to ensure quality, on keeping records to document and describe those transactions, and on ensuring the records are available for future reference, to generate invoices, or to make payments and summary accounts. Specialist accountants are trained and hired to do this. Unlike accountants, scientists are expected to perform similar tasks with research data without the benefits of training.

The key steps followed in research data management are summarised in Figure 1.

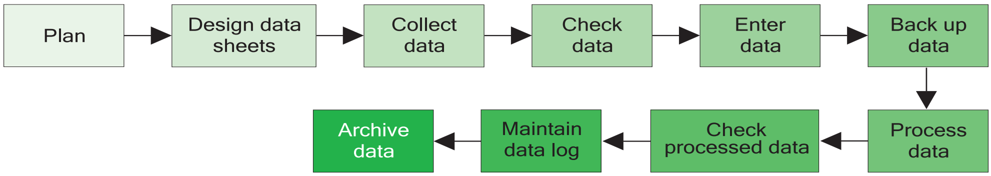


Figure 1. Data management processes

Planning for data management takes into account research objectives, resources and skills available. Appropriate field data recording sheets are designed. **Data collection** includes appropriate quality control. **Raw data** should be checked for errors. It should be entered into well organised computer files. **Captured data** must be backed up to safeguard against catastrophes. Data are processed for analysis, the results of which are checked again for any errors. Any **data processing** is logged to track data changes. Finally, data are archived for future reference possibly by other scientists.

After reading this chapter, we hope you will be better able to manage your research data. To appreciate the difficulties involved, some of the problems will be discussed. Such problems are both technical and people-oriented.

Technical problems include such issues as: lack of skills, lack of data documentation so future access is not straightforward, joint access for team projects, lack of proper design so as to meet data requests, incompatible data sets in cases where similar data are gathered at different locations or times, or files backed up on software that is no longer supported.

'Soft' or people issues include: time wastage in searching for data, re-processing old data sets, collecting data that had already been collected, and reformatting data.

Data capture

As shown in Figure 1, data capture is the activity that combines collection, checking, entry and saving data in some permanent electronic medium. You can get lots of help from specialists in carrying out the data management steps before and after data capture, but this is the one step you cannot shortcut and have to do yourself without much help. **It is the step that takes most of the resources (time and money) meant for the research; and that's why it is so critical.** The quality of the data processing that comes after this step will be determined by many factors, including which data you capture, how you lay it out, and which tools you use to do the job.

The tools

Some data capture needs can be sufficiently met by using word-processing software to publish the final results in a simple table. That's important, but it's not the main reason you enter data. It is to help you turn raw data into more meaningful results, an operation that is more difficult to archive with **word processors** than other software tools. **Databases** are the other type of tools available for data manipulation. However they are not in common use because their use is not intuitive for users who do not have much programming experience. Between the word processor and database extremes lie the **spreadsheets** that some people prefer to use for data capture because they are easy to use for data entry, **limited** manipulation and to display simple graphics. Here, the word limited is emphasised because the extent of the spreadsheet limitation is something that is under your control. Used without any discipline, a spreadsheet can be as severely limiting as a word processor; with discipline you can use it to process your data with a flexibility coming very close to what you can achieve with a well designed database application.

Disciplined use of identifiers

Most of the data types that you will need to capture will be numeric, but these will very often need a few **non-numeric identifiers** for labelling rows or columns of numbers. Some people use long descriptive names as identifiers to make it easy for humans to understand and process the data. Others will use short often cryptic codes as identifiers, so that when the data are exported to other processing environments, the codes are very useful for formulating data manipulation commands. The majority of users use a mix of the two types of identifiers in an unplanned way, thus making it very difficult to understand them and limiting the extent to which the data can be processed either in or out of spreadsheets. Table 1 shows a spreadsheet that attempts to capture both types of identifiers and is laid out in a form that is easy to export for processing.
In this layout you will notice that:

Table 1. Using long descriptive and short coded identifiers in a spreadsheet

Plot identifier	Name of the village	Size of plot in square meters	Maize yield in kg	Is the plot infected or not?	Number of insects counted
plot	**village**	**size**	**yield**	**infected**	**insects**
1	Kesen	2.5	40.7	no	0
3	Sabey	2	53.6	yes	144
4	Sabey	5	50.7	no	0
4	Kesen	8	27.6	yes	107
6	Kesen	4.5	48.7	no	0
8	Sabey	4.5	37.7	no	0
8	Kesen	7	25.8	no	0
9	Kesen	1.5	40.4	no	0
10	Sabey	4	30.6	no	0
11	Sabey	3.5	24.3	no	0
12	Sabey	4.5	59.3	yes	35
13	Sabey	5	44.6	yes	340
14	Sabey	5	56.8	no	0
19	Sabey	5	62.1	no	0
20	Sabey	1.5	33.6	yes	489

- The long names are captured in single cells and formatted in word wrapping style – instead of the more common way of using multiple cells to break the label into small displayable chunks
- The short names are all alphabetic. Avoiding other characters or alphanumerics is a good discipline since most other applications will strip them out, or replace them with codes that may change the column names to something unexpected
- The short labels are entered on the last row, just before the numbers – allowing a **data export range** that excludes the long titles to be formulated and named.

Other descriptors

To further describe data sets, users will often go to great lengths to formulate folders and filenames that document the data. So a folder/filename like, /Western Kenya/Eva/*Striga* research/2000.xls is not uncommon. There are two problems with describing data sets like this. The first is that you lose this description if the file is copied to another folder. The second is that the folder/filename structure gets very convoluted if you attempt to cram in all available documentation. One way to get round these problems is to enter these other documentations directly into the spreadsheet, rather than coding them into folders and filenames. Entering them at the header is less likely to interfere with other spreadsheet operations than anywhere else. A good example is shown in Table 3.

The body of a data set

Data identifiers will normally be few, placed at the top of a worksheet, and are needed to provide meanings to the values that form the main body of a worksheet. The quality of a dataset and its processing efficiency is determined by how much discipline has gone into the construction of the spreadsheet's main body. Here we look at a few tips that are easy to follow and that have profound effects on your data-processing efficiency.

One value per cell: when to create a new worksheet

One general rule for capturing and storing data is the concept that data in a single cell is **atomic**, i.e., only one data item occupies one cell. It is difficult to analyse multiple data entries per cell. The solution is to create another sheet with repeating data and to link it to the first sheet. For example, suppose variable Q1 in the variable set Q1, Q2, Q3, ..., Qn, has repeating values as in Figure 2a.

a.

Qno	Q1	Q2	Q3	Q4 ...
1	2, 6, 7	5	10	20
2				
3				
4				
.....				

b.

Qno	Q1
1	2
1	6
1	7
.....	

Figure 2. One entry per cell principle: a. Error in column Q1, b. Solved by another sheet for Q1

The circled entry is in error. The solution is to create another worksheet for Q1 as shown in Figure 2b. The two worksheets are then linked, using special formulae in Excel, e.g., vlookup(...) that are more difficult to use than exporting the data to a database package like Access.

Data body: row consistency

In the body of a spreadsheet, all the rows should represent the same entities. A further addition to this rule is that each row should be so completely filled in that sorting the body in any way should not result in the loss of meaning for that particular row. Table 2 shows a data set that clearly violates this rule. It is clear that:

- Line 9 represents a new sub-plot heading, and not crop row measurements as in all the prior rows. If you sorted the rows using some order, say ascending total grain fresh weights, you would no longer be able to tell which cropping rows came from which sub-plot. The solution is simple: create a new sub-plot column and fill it accordingly. The other

Table 2. A spreadsheet body with inconsistent row entries

1		Sub-plot 1				
2	Block	Maize plot	Row	Cropping system	Total fresh grain weight (g)	Total fresh cob weight (g)
3	1	1	1	control	2291.0	528.0
4			2	control	1156.0	228.0
5			3	control	871.0	199.0
6			4	control	missing	missing
7			5	control	505.0	88.0
8						
9		Sub-plot 2				
10			1	control	571.0	147.0
11			2	control	564.0	132.0
12			3	control	430.0	113.0
13			4	control	188.0	108.0
14			5	control	649.0	236.0
15						
16		Sub-plot 3				
17			6	control	861.0	201.0
18			1	combi	lost	lost
19			2	combi	381.0	121.0
20			3	combi	536.0	143.0
21			4	combi	438.0	140.0
22			5	combi	617.0	169.0

solution of moving the second part to a different sheet is not recommended because you lose the integrating effect that allows you to analyse the data set as a single unit.

- Lines 8 and 15 are blanks, which represent entities that are different from the prior rows. The user may have inserted them for some sort of clarity, and not to indicate the end of a data set range, which is how Excel would interpret them. So, if you used your sort function, only the top part, up to the blank row, of your data set would be sorted, which is probably not what you intended.
- The case for data lines 4–7, 10–14 and 17–22 also needs attention. Without rearranging these data, it is clear to us what the implied values are in the blank entries. But this would no longer be the case if the data were sorted. This is a very common problem when users try to make a spreadsheet look exactly like the paper forms. The solution, of course, is to fill in the implied values.

Data body: column consistency

The column entries in the body of a data set should all be the same type of data. This is important to prevent errors during data conversions using some formulae, or during data exports. Some entries in Table 2 violate this tip. Note that 'missing' and 'lost' values for Total fresh grain weights in rows 6 and 18 are text data types which data processors would understand differently from the numbers. What you should put in these cells depends on the software that will be used to further process these data. For instance, for Genstat you would use the star (*); for SAS you would use the dot (.) In some cases Excel would treat these labels as 0, resulting in an incorrect result when the columns are used in calculations. Our recommendation is to leave the columns blank. Should you need to explain further why no value existed, use the comment feature. Unfortunately, comments are ignored when data are exported to environments outside of Excel, thus limiting their usefulness.

Putting it all together

Table 3 shows a spreadsheet whose preparation has considered most of the tips given in this section. It does not matter that this one has been designed for experimental data; the same tips are applicable to survey types of data.

Table 3. An example of a spreadsheet design that uses some of the tips discussed in the previous sections

Program	Domestication of Agroforestry Trees					
Project	Genetic Resources of Agroforestry Trees					
Experiment	Leucaena family trial					
Location	Kenya, Muguga					
Investigators	James Were, Tony Simons					
Start Date	5/1/1996					
Statistical Design	Incomplete block design					
Assessment	Tree growth					
Date	Feb-97					
Replicate number	Blocking id	Plot number	Leucaenea family id	Tree identifier	Tree height (cm)	Number of stems
rep	**block**	**plot**	**family**	**tree**	**height**	**stems**
1	1	1	20	1	214	4
1	1	1	20	2	252	6
1	1	1	20	3	153	2
1	1	1	20	4	183	4
1	1	2	18	1	98	1
1	1	2	18	2		
1	1	2	18	3	201	3
1	1	2	18	4	192	1
1	1	3	9	1	232	8
1	1	3	9	2	201	7
1	1	3	9	3	198	4
1	1	3	9	4	152	4
1	1	4	10	1	175	2

Data entry and validation schemes

A well designed experimental data sheet is ideal for data collection. If the design is done before field data collection, it should be printed and used by all those who are collecting data in the field. Data entry should be done as soon as data collection is complete so that any clarifications are sought while people's minds are fresh. Some initial checks should be done for obvious errors. Missing values should be carefully treated. Ensure non-available data appears as blanks in the worksheet (not as zero since zeros are included in statistical calculation). You can use a number of techniques to aid data entry and avoid transfer errors.

During data entry, and especially for long or wide lists, it is a good idea to be able to see column and row headings all the time even as you scroll through the worksheet. This is achieved by freezing or splitting the window panes. In Excel this is achieved by selecting Window ➠Freeze Panes (or Window ➠ Split) when the row below the column heading is selected. To remove this effect, select Window ➠ Unfreeze Panes.

Validation during data entry

Data should be entered quickly and in raw form to minimise the chances of making errors in transcription. You should enter all the data. Partial data entry that can be quickly analysed is not recommended. If you enter it all, you can cross check during entry to minimise errors.

To identify data records, each field should be unique, for example, plot number. Derived data should not be entered. Since computers are good at calculations, you are well advised to simply enter primary or fundamental data, thus avoiding possible errors caused by hand calculation. For field experiments, it is a good idea to enter the data as they appear in the plots (for logical mapping to the physical plot) or to use two columns to identify the location of the plot – (x, y) coordinates using some reference frame.

Drop-down lists can be used to avoid typing a sequence of data more than once and to avoid typing errors. A drop-down list is a set of data (such as crop names) from which you can chose one for entry into a cell. This is created by highlighting the data set, and selecting Data ➠ Validation➠Allow:➠ List. For Source of list give the range of the data set.

Data validation is also done on a range of cells. When new data is entered in cells with range checks, any data values outside these ranges generate error messages. For example, if values for a variable *wheat % moisture* is in the range 13 to 29, you can set the range check by highlighting the data area for that variable, then select Data➠ Validation ➠ Allow: Decimal and set Minimum as 13 and Maximum as 29.

Validation after data entry

Scatter plots can be used to spot data outliers once data have been entered. These are data values considered outside the allowed range and easily seen from a plot (see Figure 3). Line plots can also be used to spot outliers.

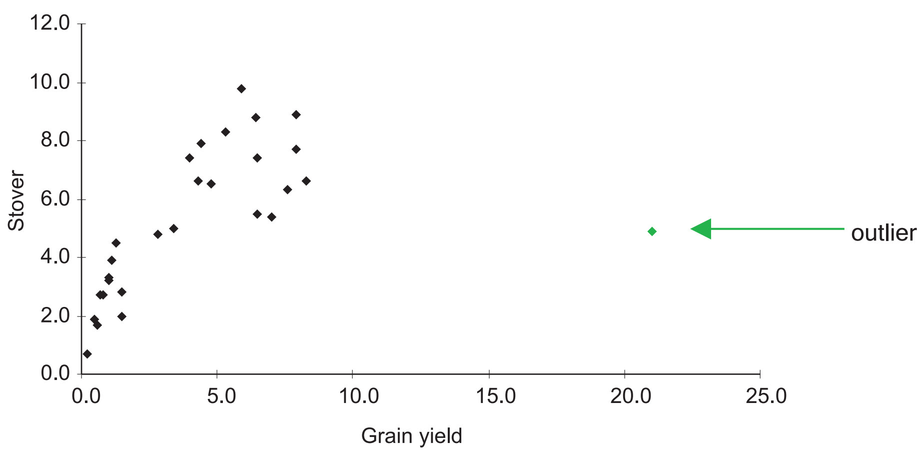


Figure 3. Scatter plot example showing data outlier (for details see Appendix 11)

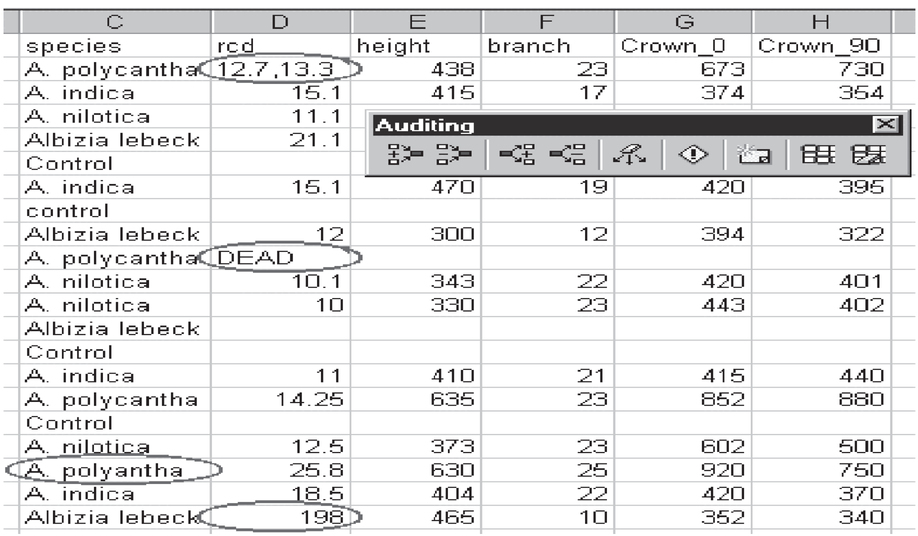
Adding comments to cells

Unusual instances occur in data capture and subsequent entry. For example, when no value is recorded for a variable, it is a good idea to indicate the causes of that anomaly. This can be done by inserting comments in the worksheet using **Insert ➠ Comment** on the subject cell.

Data auditing

For already existing data, the auditing tool allows you to check for some errors. Auditing is created by: **Select Tools ➠ Formula Auditing** ... and then click on **Show Formula Auditing Toolbar.** On the toolbar, click on the icon **Circle Invalid Data** (second from right). This draws a ring around invalid data.

An illustration of auditing, with validation rules for both ***species*** and ***rcd*** is shown in Figure 4. In the figure errors are circled for the variables ***rcd*** and ***species***. All cells in a column should have the same data type. In the case of column **D** (Figure 4), the data type is numeric. The string 'DEAD' in cell **D10** is therefore inappropriate. The entry 12.7, 13.3 in cell **D2** is also invalid (it is ringed) because of two decimal number values and so is 198 (in D21) because it is out of range. In cell C19, A. ***polyantha*** is wrongly spelled and hence ringed. You can see how the auditing tool helps you to spot data entry errors before processing.



C	D	E	F	G	H
species	rcd	height	branch	Crown_0	Crown_90
A. polycantha	12.7,13.3	438	23	673	730
A. indica	15.1	415	17	374	354
A. nilotica	11.1				
Albizia lebeck	21.1				
Control					
A. indica	15.1	470	19	420	395
control					
Albizia lebeck	12	300	12	394	322
A. polycantha	DEAD				
A. nilotica	10.1	343	22	420	401
A. nilotica	10	330	23	443	402
Albizia lebeck					
Control					
A. indica	11	410	21	415	440
A. polycantha	14.25	635	23	852	880
Control					
A. nilotica	12.5	373	23	602	500
A. polyantha	25.8	630	25	920	750
A. indica	18.5	404	22	420	370
Albizia lebeck	198	465	10	352	340

Figure 4. Auditing of existing data (see SCS–University of Reading: Disciplined Use of Spreadsheet Packages for Data Entry)

Saving and protecting files

Once data have been entered, it is important to use a good **file naming** and **saving strategy** so you can easily refer to the same data in the future. Files and where they are stored (directory structure) should be named with sensible names that suggest the research type. Choose a directory structure that best describes the structure of the files. Data files should not be mixed with program files. A common scheme is to use the **directory** \USER for storing all data files. Each research location should have its sub-directory and each experiment within that location should have its sub-directory. For example, 'Eucalyptus tests' at

'Kakamega' might have data stored in the directory \USER\KAKAMEGA\EUCALYPTUS. All files related to this experiment would be stored in that directory. Such files might include field data, reports, charts, documentation and statistical results for the Kakamega eucalyptus tests.

It is also a good idea to document the workbook using Excel's *summary information* which shows the *title, subject* and *keywords* for the workbook. This is particularly important if many workbooks are likely to be used. Summary information is achieved by clicking on **File ➠ Properties.**

It is important to **save** and **backup** your data regularly. Hard disks get corrupted for a variety of reasons. The computer could get stolen, burned or simply fail to work. Backing up can be done on diskettes, tapes or CD-ROMs using a CD writer. For diskettes, keep at least two sets (a copy may take several diskettes) in different places besides the one on hard disk. To avoid having several versions of the same data, it is advisable to create a **master copy** (with the same file name) which is updated and backed up every time there is a change of data, and keep it away from the computer.

It is a good idea to **protect data files** from unintended changes and when they occur, to keep track by highlighting them. This is achieved in Excel by selecting: **Tools ➠ Track Changes or Tools ➠ Protection** (this allows you to set up rules for file protection).

When to use an advanced tool

MS-Excel is a powerful tool when used properly with single data files. When multiple data files are used, it becomes difficult to maintain data in those files. Often you end up with several files of the same information, and it is hard to keep them consistent with each other. Querying these files is not easy and becomes cumbersome and inefficient if you have several of them.

Relational databases were designed specifically to handle many related data files and are optimised to allow efficient data querying, ensure data integrity and sharing of data between different individuals when necessary. A database allows you to have data of the same subject in a table, and then, assuming there a number of different data sets, have each on a different table in a database. Because research is usually on related objects, the next step is to relate the different objects. This ensures **referential integrity** so that changes to a data item are made through a controlled environment. During your analysis phase, you should use a statistical tool like SPSS, **Genstat (Appendix 11)**, or SAS. The database data can be exported to the specific analysis tool of your choice.

Designing a database

If you are handling several related worksheet files, the logical step to take is to use a database for these data. The key issue in a database approach in capturing research data is the initial design of the base tables. Each data object (called **entity** – like a plant), has characteristics that define that object. For example, a plant has leaves, height, maturity stage, shoot size at a given time, description, etc. These are the **properties** of the plant object. In Access-speak, the plant entity's properties are the **fields** or **attributes** and each has a **data type.** (sample data types are integers – identified as **integer** and **long integer**, decimal point numbers – identified as **long** and **double**, and **text**, amongst others).

Database design means defining each of these attributes with their data types, selecting one or a combination of attributes as a **unique identifier** for the plant object (called the

primary key) and then repeating this process for other entities. After this, you can create **relationships** between the different entities. For more complicated databases some real design work is necessary (which includes **normalisation**). An example is the relationship between an employee and her dependants shown in Figure 5 (both employee and dependant are entities). This type of relationship is called one-to-many – one employee can have several dependants.

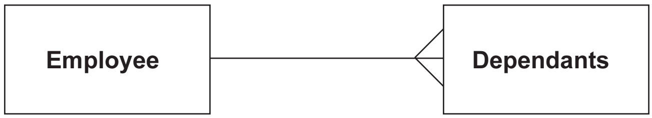


Figure 5. Relationship between an employee and her dependants

Once the fields for each entry are choosen you can define a table to hold the data. The table design screen in Figure 6 shows the design of a person-level table. Names for the fields and their data types are defined. Once the table is created you can enter data via the datasheet or the spreadsheet view. This is shown in Figure 7. The datasheet resembles the Excel worksheet.

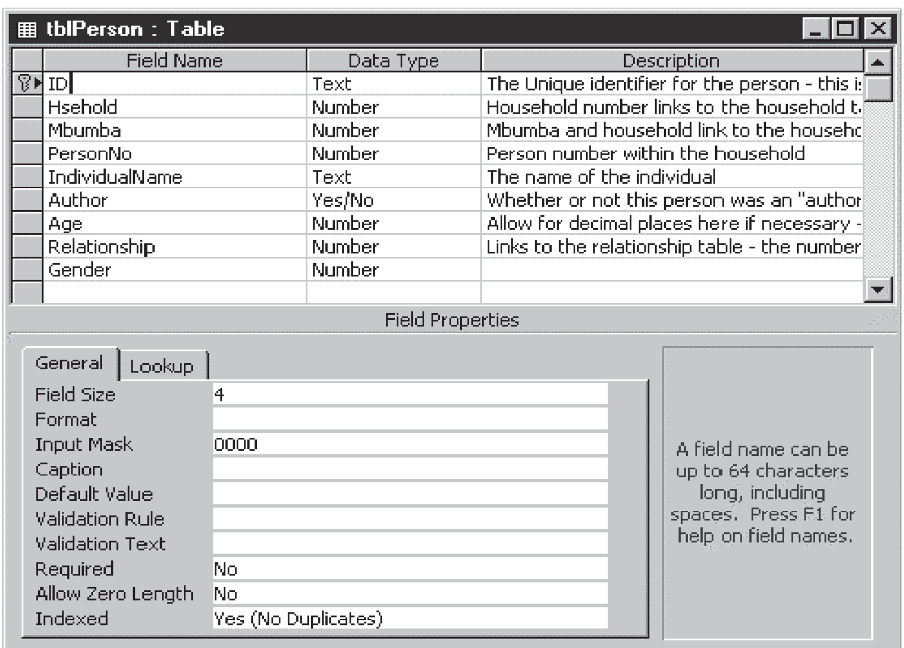


Figure 6. Table design in MS-Access for a person-level table

As with the use of a spreadsheet, it is important to use a database package '**with discipline**'. With minimal discipline, defining the number of fields and their data type is enforced, but you should normally do more than the minimum.

tblPerson : Table

ID	Hsehold	Mbumba	PersonNo	IndividualName	Author	Age	Relationship	Gender
2101	1	2	1	Mai Mazinga	☐	55	1	Female
2102	1	2	2	Mercy	☑	18	3	Female
2103	1	2	3	Tokozani	☐	15	3	Male
2104	1	2	4	Charity	☐	1	6	Female
2105	1	2	5	Unknown	☐			
2205	2	2	5	Mr Nangwale	☑	40	1	Male
2206	2	2	6	Martha	☐	31	2	Female
2207	2	2	7	Enifa	☐	11	3	Female
2308	3	2	8	Frank Filipo	☐	30	10	Male
2309	3	2	9	Femia	☐	27	2	Female

Record: 1 of 94

Figure 7. 'Datasheet' view of person-level data

Selecting data for analysis

There are two alternative ways of getting Excel data into a database. The first is by **importing** it into the database. This leads to two copies of the same data set and can be a major source of data inconsistency when changes are made in the database but corresponding changes are not made on the worksheet. A good practice in data management is to designate either the worksheet or the database the **master copy** so that data changes are only done on the master and all data sub-sets are extracted from the master for analysis.

An alternative to importing Excel data into a database is **linking**. Here, the database and the spreadsheet use the **same datasheet copy.** Changes made in the database or in Excel are reflected in this copy. **This option is the best in terms of data management.** You will have no worries about managing multiple data sets if you use this method.

Using queries for calculations

Databases are designed to store **fundamental data. Computed data** is not fundamental since it can be derived from other data. To get computed data in databases, you can use a **query** and create a new field where a formula for the derived data is entered. For example, to compute the average height of a tree in a certain experiment, assuming individual heights are stored in a field called height, you would enter the following in a blank field of the QBE grid – **avheight:sum([height])/count([height]).** Once computed, this field cannot be updated manually with data, unlike in Excel where a formula can be overwritten with data.

Using queries to select records

In Excel, you can select rows of data by using **Data ➠ Filter** and specifying a criterion. This is quite simple but the resulting rows of data must be copied to another sheet before use. Both the filtering and copying of results are manual processes and prone to error, and the results are a duplicate of the original data. As observed, creating copies of the same data is a bad data management practice. In databases, you can get the data sets you want by creating a query and setting criteria (e.g., all trees with a height between 20m and 40m). It is possible to have complex queries using **and** with **or** logic combinations. The important point is that the **query** is stored in the database (not the results) and you simply execute it whenever the dataset is required. Similarly, queries can be created to select fields, link multiple worksheets so that data sets can be extracted from all of them, and also to check data validity.

Data archiving

Archiving is the process of storing data for future use. The user of archived data is not necessarily the person who did the experiments, or carried out a survey. Indeed a well archived data set can be used by others to derive new relationships in the data or to compare primary data with secondary data. Funding agencies may even be attracted by the possibility of archiving data from the findings of a proposed project.

The process of archiving data requires three basic principles:

1. The data about the project rather than the results of the study itself (sometimes called meta-data – description of the data itself) should be archived.
2. The description of why data was collected should be archived.
3. You must archive a description of the data files – their types and structure.

The latter makes future retrieval easier. The first point makes it possible to easily understand the rationale of the data-collection exercise, while the second gives additional information on the procedures and processes of data collection. This means a future researcher would be able to replicate the experiment or survey for scientific validation of the findings. Data files need to well structured, in the majority of cases they are computer files.

Backups of computer files should be made regularly with a strategy of keeping a master copy far away from the archival site. This ensures continuity in case of natural hazards like fires, floods or earthquakes.

To summarise, a good data archive should be/have:

- **Accessible:** hence easy to access by many users who have commonly available software
- **Easy to use:** so that the field data collection forms, and what will be entered into the computer are similar
- **Clearly defined variables:** the units of measure and codes used (labels for names of variables) should be as clear as possible
- **Consistency:** of names, codes, units of measurement, and abbreviations
- **Reliable:** archive should be as free from errors as possible
- **Internal documentation:** documentation should be complete with regard to: procedures for data collection, sampling methodology and sampling units used; the structure of the archive (how different files are related); a list of all computer files in the archive; a full list of variables and notes on how to treat missing values; summary statistics for cross-checking the information in the archive; and any warnings and comments that need to be observed for data usage
- **Confidentiality:** ensure that the data remain confidential if this is required by the sources
- **Complete:** if possible you should store copies of: the data capture field forms; the data management log-book; a description of derived/calculated variables.

Storage and access to the archives is also an issue for you to consider. A good archive includes information on how to get into the archive with rules of use and replication to other third parties. The storage medium for computer archives is in most cases, hard disks. With the new pervasiveness of the Internet, access to the archives is mainly by downloading archive files. This is true for different types of data including text, graphics, maps, photos, and audio and video material. Using diskettes and/or CDs as distribution media is of course an option. The other medium is good old printouts sent by mail for long-distance access.

To show the seriousness of data archiving and its place in research, it is now possible to publish peer-reviewed **data papers** (http://esa.sdc.edu/Archive/E081-003/main.html).

Data ownership issues

Any data set must belong to somebody. **Ownership** in the wider context means who can access data for reading only, updating, deletion or creation. In the scientific community, a data set does not necessarily belong to the individual who generated it. It generally belongs to an institution that can ensure continuity through hosting the data and providing access to it to individuals, or other institutions.

If data are generated by more than one institution, for example, through collaborative research, then they belong to the participating institutions. If data are generated by publicly funded projects, then they are public property, held in trust by an institution for the public. There is need to recognise **intellectual property** for scientists who may generate data. If scientists generate data using public funds, they must use the data for the purpose for which it was intended.

There are data ownership issues that need to be agreed on by several different parties. Institutions have an obligation to give **public access to data**. They do this by adopting some policies and procedures that must be communicated to all the interested parties.

It is important to balance the rights of individuals who collect data with the need to ensure future public access to data. Two approaches are used:

1. **A time limit** is set beyond which a scientist can not claim ownership to data. For example, 1 year for field research is typical, because this gives the scientist some time to carry out the analysis and publish before the data become public property.
2. Data owned by an institution may be released to a researcher if a good case is made for that access. Acceptable reasons may include: to check analysis, to improve on analysis, to correct an analysis, to analyse new questions using the same data, for integration with other data, and for meta analysis.

The **subjects of a data set**, for example, the people interviewed or the farmers who took part in an experiment, also have some rights. These may be set by common values or may be determined by law. The minimum all subjects can expect, and that all researchers should ensure, are:

1. That all personal information will be kept confidential.
2. That the data will only be used for the intended purposes, and the people concerned agree to this before participating.
3. That results of the study are made available to all people participating.

Data management strategy

A **data management strategy** is a set of policy guidelines developed for an institution to help it cope with different facets of data management. A data management strategy requires the following:

- **Commitment.** This is important for all stakeholders of a project including project managers and researchers. It is a pre-requisite for a successful data management strategy since it will enable commitment of resources to develop and maintain the strategy
- **Skills.** These are required by all players to do what is necessary for the data management strategy. They include data entry skills for junior members of the team, form design skills, data entry and validation skills, and data archiving skills amongst others

- **Time.** This is necessary for a good output. Funding agencies have recognised that besides the final research output (analysis results), data is also an important output achieved by archiving. Clearly enough time ought to be devoted to data management in the overall research timeframe
- **Financial resources.** It is important to include data management within the project proposal, otherwise the necessary tasks it involves will not be done.

Key components of a strategy

The four key components of a strategy are:

1. **Transformations and their products** – these are the steps in research data management.
2. **Managing meta-data** – the process of defining and managing descriptions about the data.
3. **Data management plan** – this is the overall plan of the strategy and how steps in the strategy can be measured for performance.
4. **Data management policy** – these are principles that guide structure and contents of meta-data and the strategy plan.

1. Transformation

This describes the entire data management cycle starting from problem definition, formulation of research objectives/hypothesis, development of data capture tools, data entry using some validation rules, selection of data for analysis, the actual data analysis, management of results and finally publication of findings. This is a cycle because it is possible to go back to any point in the process in case there are errors.

2. Managing meta-data

Meta-data is a description of the data to be handled in a research project. It can be used to describe data sets, enable effective management of data resources, and to enable other researchers to understand the data sets of a project.

The key areas of a meta-data are:

i. **Title.** The name of the data set or the project
ii. **Authors.** Names of researchers (principal researcher and others) with addresses, phone, e-mail, and web contacts
iii. **Data set overview.** Introduction to the data set, location of data, time of experiments/ survey, and any references
iv. **Instrument description.** Brief description of data capture instrument with references
v. **Data collection and processing.** Description of how data were collected, computed values, and quality control procedures
vi. **Data format.** Structure of data files and naming conventions, codes (if used), data format and layout, version number and date.

Meta-data description must be done for every project.

3. Data management plan

A plan shows how data will be recorded, processed and managed for the duration of the project. It includes roles for staff, back-up procedures, quality control checks and how to handle errors, procedures for managing the data management strategy e.g., discussions in meetings, procedures, software upgrades, methods of creating archives, and how the archive will be maintained.

4. Data management policy

These are policy statements that guide data management to ensure consistency. They are high-level objectives of data management. A typical policy consists of the following objectives:

- Establish and distribute high quality data sets
- Standardise quality control procedures
- Ensure data and other project materials are archived and reviewed regularly
- Reduce time between data collection and analysis
- To maintain data securely
- Facilitate data access and usability through improved meta-data.

The **policy** includes the **roles and responsibilities** of individuals in the project, specifies the data owner, the data custodian (a manager in charge of the data management process), data user (individual with access rights to the data), security administrator and an information systems group.

Conclusions

Data management in research is very important. It is the entire process encompassing project initiation, through all the phases up to the time a paper is published as a result of that research. For quality results, all the phases as described in the strategy above must be managed according to the stated principles. The scientific community now accepts **data papers** for publication, in addition to the traditional research output documents. Data archives are a rich source of information so your data should be archived following the guidelines discussed above. These should give you enough reasons to embrace research data management and practice it. After reading this chapter, you should have sufficient information to better manage your research data.

Besides the referenced material, additional sources of relevant information are provided below.

Resource material and references

Appendix 9. Muraya, P., Garlick, G. and Coe, R. 2003. *Research Data Management.* World Agroforestry Centre (ICRAF), Nairobi, Kenya. On CD.

Appendix 11. ICRAF. 2003. *Genstat Discovery Edition and Other Resources.* World Agroforestry Centre (ICRAF), Nairobi, Kenya. On CD.

Ecological archives. A *Publication of the Ecological Society of America*: (http://esa.sdc.edu/Archive/E081-003/main.html)

Coe, Richard. 2001. *Audit Trail in Research Data Management.* (Draft Report), World Agroforestry Centre (ICRAF), Nairobi, Kenya.

Coe, Richard. 2001. *Issues in Data Ownership and Access.* (Draft Report), World Agroforestry Centre (ICRAF), Nairobi, Kenya.

Statistical Services Centre (SSC), University of Reading, UK. http://www.reading.ac.uk/ssc.

Case Study No. 6 – Good practice in data management (2000)

The role of a database package for research projects (November 2000).

Project Data Archiving – Lessons from a Case Study (March 1998).

Good Practice Guidelines (2000).

4.7 Analysing the data

Susan J. Richardson-Kageler

- **Analysing the data means turning the raw observations into summaries that can be interpreted**
- **Appropriate methods for analysis depend on the objectives, the study design and the nature of the observations**
- **Exploratory and descriptive analysis, that displays the main patterns in the data with summary tables and graphs, will allow you to tentatively meet many of your analysis objectives**
- **Formal or confirmatory analysis will add information about uncertainty and allow you to disentangle complex patterns**

Introduction

This chapter summarises the key points to look out for when you analyse your data, and the dos and dont's. It is assumed that you have taken a statistics course at some stage. If you need more details of the statistical techniques that are used in this chapter, then refer to the relevant text books (see Resource material and references at the end of the chapter for some of these texts).

Students often get stuck when they have to start analysing their data. This can happen to you for a number of reasons. Perhaps you are scared of the analysis stage. You have approached the research confidently, enthusiastically collected the data and now find that you do not know how to proceed with the analysis. Sometimes the problem is made worse because you have left the analysis till the last moment. The analysis should not be considered the final stage in the research process, but should be done as soon as data become available. A good researcher will think about the analysis at the research proposal stage. In the research method section of your study proposal you should include an analysis plan with descriptions of the possible tables, figures and methods you will use. This will help you to think about analyses early and will ensure that the analysis will be possible and will meet your objectives.

Descriptive methods are more important in real analysis than their emphasis in many statistics courses would suggest. Often the most elegant analyses and main results are obtained from well thought out summary tables and skilfully designed graphs. They require little more than common sense and a clear idea of what you are trying to find out. So, even if you are unsure about formal statistical methods, you should be able to start your analysis.

The following are the usual steps in the analysis of data that are considered in this chapter:

- Define the analysis objectives
- Prepare the data
- Descriptive analysis
  - Tables
  - Graphs
  - Summary statistics
  - Identify oddities
  - Describe data pattern
- Confirmatory analysis
  - Adding precision
  - Improving estimates

- Interpretation
  - Understand the results
  - Combine new and old information
  - Develop models
  - Develop new hypotheses

As you progress through your research and analyse the data as you go along, you will find that the data analysis is **iterative**, this means it is not a simple matter of following straight through the process outlined above. You will need to stop and revisit previous steps as new information is discovered. Even though you analysed the data as you progressed through your work, you may need to re-organise your data and reanalyse, so you must leave plenty of time to complete the analysis and write up the results after data collection. Remember, it always takes longer than you expect to get the tables, figures and analyses compiled and written up effectively.

In this chapter, two examples are used. The first is a survey which investigates farmer's perceptions to, and use of, planted fallows. A questionnaire was administered to 121 farmers who had experience with planted (improved) fallows grown with or without rock phosphate fertilizer in Western Kenya.

The second example is an experiment to evaluate whether the pumpkin (*Cucurbita maxima* L.) variety Flat White Boer can be used as a smother crop when planted at the same time and intercropped with the long-season maize variety PAN86 at the University Farm, Mazowe, Zimbabwe. This evaluation is done by comparing sole maize, sole pumpkin and a pumpkin-maize intercrop.

Analysis objectives

Two key points for this section:

- Analysis objectives are determined by, but more specific than, the overall research objectives.
- Analysis objectives will evolve during the research as you gain insights and experience.

When the research proposal is put together at the beginning of the research, it is usual to state an **aim** and a set of **objectives** in the introduction. **Analysis objectives** often need to be stated separately from objectives of the research work. This is because the analysis objectives are dealing with the **specifics** of the analysis. The original objectives may appear vague in comparison as they often do not specify precisely which variables are to be analysed and how they are to be processed. The analysis objectives will determine such specific things as:

- What the relevant variables are and to which level they are summarised (see later in this chapter in the section on Preparing for analysis)
- The specific comparisons that will be made
- The relationships between variables that will be investigated.

For the pumpkin-maize intercrop example, the analysis objectives include:

- Comparing the maize grain yield between sole maize and pumpkin-maize intercrops
- Comparing the weed density of the sole maize, sole pumpkin and pumpkin-maize intercrop
- Determining how the maize yield depends on pumpkin and weed cover.

The analysis objectives should be refined as you proceed through your data collection. Your experience in the field will give you ideas and insights you did not have when you planned the research. It is a good idea to keep a notebook and record your observations and ideas as you go along. Things often happen for which there is no place on your data entry field sheets. You often find that when you come to complete the writing up that you cannot remember these important things that have occurred during the research process. Even go so far as to keep a notebook beside your bed at night. This will help you to sleep better as you can write down the things that you think of, and so you won't have to keep yourself awake to make sure that you remember your ideas in the morning! Examples for your notebook include the fact that one of your test animals broke out and ate your neighbour's vegetables, or ideas for more informative graphs and data summaries.

Preparing for analysis

Some key messages for this section include:

- Data must be well organised – see **Chapter 4.6**
- Data must have been checked – see **Chapter 4.6** – but remember that more mistakes will become apparent as you do the analysis
- Some data preparation is necessary once the analysis objectives are clear
  - – Summarise to the right level
  - – Use suitable format for the software.
- The preparation of data for analysis starts with the project proposal and ultimately fulfils the objectives of the research. From the start, think through how the data are going to be entered onto the computer and which software packages are going to be used.

The stages involved in the preparation of the data for analysis are:

- Raw data entry and checking
- Organisation of the data to the form needed for the analysis to meet the objectives
- Archiving of the data so that it remains available.

Once the data are entered, check the data entry using simple data analysis. This includes transforming and plotting the data, summaries which show extreme values (minimum and maximum values, trimmed means), boxplots, scatterplots, tables of the data in treatment order, frequency tables of coded data, ANOVA and plots of the residuals.

After checking the data entry, or sometimes before it is entered, the data must be summarised to the appropriate level for data analysis. To do this, you will need to recognise the correct data structure. You need to decide at which level you will do the analysis to satisfy the objectives of the study. If data have been collected on several individuals per house, do you want to analyse the data about individuals or about households? If data have been collected at sub-plot level, for example, 10 plants have been randomly sampled from each plot, you may need to calculate a summary statistic for these 10 plants to be used in an analysis at the plot level. The data may be complicated and involve many sites over several districts. Some of the objectives may be fulfilled by summarising data at the site level, clustering similar sites into groups and then comparing the sites across these groups. A survey on disease in coffee involved many districts, farmers and trees. The huge data matrix looked like a nightmare, yet the first objective was satisfied by simply calculating the proportion of farms in each district that had the disease. The next objective required calculating the same thing for two different coffee varieties.

You may also need to calculate new variables from those you have measured. For example, you might have measured fresh weight and moisture content, but later discover that you need an analysis of dry weight. The whole process of understanding your data structure and deciding on the summaries or variables or units to be used needs to be continuously related back to the objectives of the study so that the analysis does not become misdirected. You may need to revisit this step again after carrying out part of the analysis as you may realise then that the summaries you have calculated may be inappropriate, or some analyses have indicated patterns that will be best examined at a different level or with different variables.

The results of this stage are data sets that are in the correct form to answer the research objectives. If this stage was not done in a statistics package, then the data should now be ready for transfer to a statistics package for analysis. Software for handling data entry, modification and analyses should all be **compatible** and includes:

- Database management software
- Spreadsheets
- Statistics packages
- Word processors.

Compatibility means that you can move the information from one package to another. For example, you may want to add your data to your final report as an appendix. You should be able to copy the data into a data-entry package such as Excel, and paste them into the word-processing package in which you are writing your project report, e.g., Word. Further, the graphs generated in a statistics package such as Genstat or Minitab, can be copied from the statistics package and pasted into your project report in the word-processing package.

Exploring and describing the data

Important points to remember are:

- Descriptive analysis uses tables and charts of summary statistics to show the main patterns in the data
- It also reveals unusual or surprising observations
- Preliminary reports and conclusions can be based on the descriptive statistics.

The aim of exploring and describing the data is to find out what the data has to tell you. The data can be split into two parts:

Data = pattern + residual

Pattern is the underlying structure or shape of the data, in which your primary interest lies. Knowing the pattern should mean that you satisfy the objectives. An experimental pattern is often the result of the treatments that you have applied. The pattern is summarised by descriptive statistics, e.g., the mean of the treatment. **Residual** is the remaining, unexplained variation. There should be no pattern in the residual part of the data. If there is pattern in the residual part of the data this indicates that some effect has been forgotten, perhaps due to the layout, treatments or measurements. **The ultimate aim of the data analysis is to describe the pattern.**

Data analysis starts by exploring and describing the data. This is the point at which you begin to understand what is really happening. When this step is carried out effectively, you can make subjective conclusions about the research and write a preliminary report. Students sometimes forego this step and go straight to the confirmatory analysis (see the next section). The **confirmatory analysis** in agricultural research often includes an **analysis of variance (ANOVA).** ANOVA can be used in a variety of circumstances: both as a descriptive tool and in inference where it is used to identify which parts of a model are important.

Example

One student who went immediately to an ANOVA missed out the most important finding of his two and a half years of research. He was investigating the effect of different diets on the fat in ostrich meat. He collected the data and then carried out an analysis of variance. He did not plot any graphs, calculate any summary statistics, or check the residuals resulting from the analysis of variance. Later, when the residuals were examined, it was found that there was at least one outlier generated by each analysis of variance. On examination of these outliers, it was found that the birds concerned all came from the same farmer. He was raising them so they had consistently lower fat in their meat than any of the other ostriches. Further investigation of this farmer could reveal he used a diet that will answer the aim of the research – which was to develop a diet for ostriches which results in low body fat. The farmer was doing exactly what was required as an outcome, he was already producing birds with low body fat and high muscle yield – but the student had missed the point entirely. The lesson from this is that **data analysis is not complete without a proper investigation of the pattern before carrying out a confirmatory analysis.**

The preliminary analysis of the data (exploration and description) should reveal the following:

- Structure/shape of the data and pattern as related to the objectives
- Outliers or unusual observations
- The need to modify the data
- Patterns suggesting new questions and the data analyses.

Methods used to explore and describe data

- Descriptive statistics
- Tables
- Graphs

All of these must correspond to the objectives of your research. These methods will carry you a long way through the analysis when added to the formal statistical techniques that you will use.

Table 1. Numbers of people interviewed by village

Count of village	
Village	Total
Eb	4
Ed	7
Ei	18
Ek	1
El	12
Em	13
Es	4
Et	1
Eu	7
Ey	4
Lu	7
Mu	4
Ny	17
Sa	18
So	1
Sr	3
Grand total	121

The process of describing data sets is probably best illustrated with examples. The data for the first example, the survey of farmer's perceptions, were first entered into an Excel spreadsheet. When starting the data analysis of surveys it is common practice to start with a series of tables summarising the data. The **Pivot Table** and **Pivot Chart** facility in Excel is possibly one of the most powerful and useful tools that Excel provides (Table 1). Table 1 is a one-way table that summarises the numbers of people interviewed by village. It is usual to start the analysis of survey data by describing the demographics of the population interviewed.

The summary would be more informative if more information was included. For example, it could include gender (Table 2). This table is now a two-way table. However, examination of Table 1 reveals that there are a number of villages with few respondents. At some stage in the analysis it may be worth combining the smaller villages into like groups. Similar groups can be established using common

Table 2. Numbers of people interviewed by village summarised by gender

Count of village	Gender		
Village	F	M	Grand total
Eb	2	2	4
Ed	3	4	7
Ei	11	7	18
Ek	1		1
El	8	4	12
Em	1	12	13
Es	2	2	4
Et		1	1
Eu	4	3	7
Ey	1	3	4
Lu	3	4	7
Mu	4		4
Ny	12	5	17
Sa	12	6	18
So	1		1
Sr	1	2	3
Grand total	66	55	121

Table 3. Summary of the use of natural fallows by gender

Count of natural fallow	Gender		
	F	M	Grand total
No	31	22	53
Past	5	6	11
Still	29	26	55
Unknown	1	1	2
Grand total	66	55	121

sense, for example, villages close together, or by using a data-driven method such as **cluster analysis**. The clustering could be based on variables decided upon after examining the objectives of the research.

But also think about the analysis objectives: Do you actually need to know about differences between different villages? Maybe 'village' is only recorded as part of the logistics of data collection, and need not appear in your summary tables. **The point is: the tables should relate to what you need to know.**

One objective in this survey is to assess the use of fallows in the past. This is given in Table 3.

For the intercropping experiment, the data were first entered into an Excel spreadsheet. It was noted, while entering the data, that a mistake was made in carrying out the experiment. One of the plots that should have been Treatment 3 was accidentally assigned Treatment 6. This is not the end of the world, and the data can still be analysed with slight adaptations to some of the methods. As these data are in Excel it is possible to carry out some analyses using Excel. The output shown here is not the default from Excel. The Excel output showed many decimal places. This has been modified to give one decimal place for each average. Displaying too many decimal places is a common mistake made by students – often because they just copy and paste the output from one package into their document.

A quick examination of Figure 1 reveals that the treatment with the highest weed biomass was the pumpkin-only crop with no weeding. The lowest weed biomass in the sole maize and sole pumpkin crops was for those weeded at 3+5+8 weeks. All the pumpkin–maize intercrops that were weeded showed similar weed biomass levels. It is not clear if these are different, however it appears that the best return for effort is to weed the intercrop at 3 weeks. Bar

Table 4. Two-way pivot table of average weed biomass for the three crops and four weeding treatments

	Weeding				
Crop	None	3 weeks	3+5 weeks	3+5+8 weeks	Average
Sole maize	83.5	79.3	28.3	6.8	51.4
Pumpkin–maize intercrop	87.6	6.2	3.5	5.5	24.2
Sole pumpkin	162.2	30.9	40.8	3.1	59.3
Average	111.1	35.6	23.7	5.1	44.2

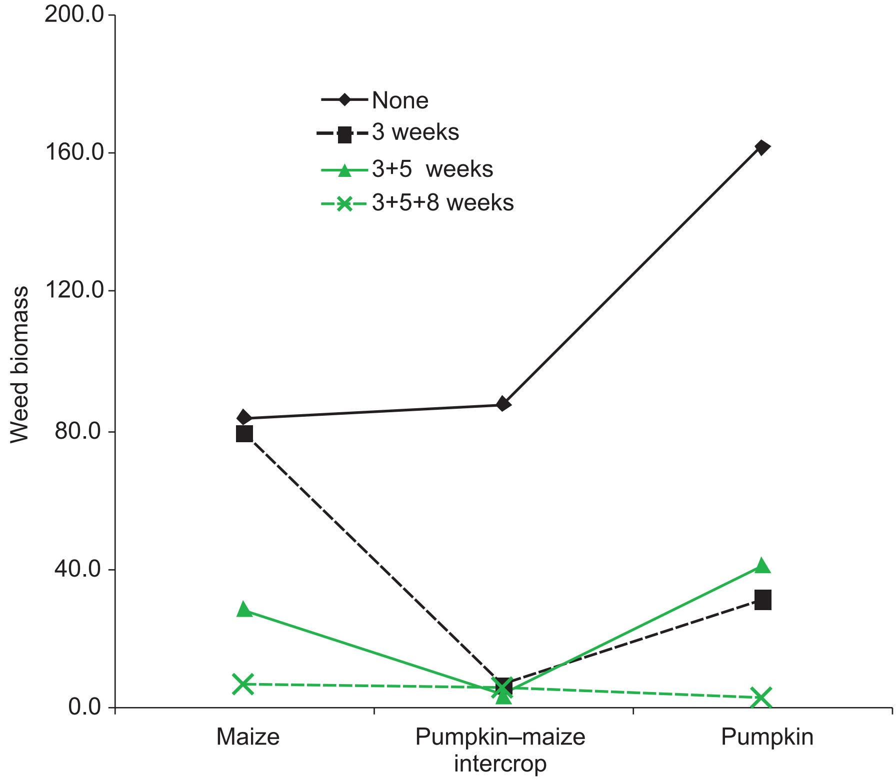


Figure 1. Graph of data in Table 4

charts can be useful displays when presenting results because they give a quick visual display of what is going on that most people intuitively understand. However, the x-axis would be the crop treatments, the plots would be each weed biomass and separate lines could be used to join each weed treatment for the three crop treatments. (Figure 1). Note that the lines connecting the points highlight which ones are from the same weeding treatment. They do not suggest that there is a weed biomass for some intermediate treatments.

So far, the analysis has summarised the data using **means**. Other summaries such as the **minimum and maximum values, trimmed means, standard deviations** and **standard errors** can be calculated. These summaries are useful when dealing with larger data sets, as they may give indications of outliers, but they are not useful when dealing with small data sets like the pumpkin-maize intercropping data. Take care in your use of a spreadsheet, it may contain statistics calculated in a way that you are not sure about. For example, the way the spreadsheet deals with missing values, or whether the spreadsheet is calculating a population or a sample standard deviation. If in doubt, take the data into a statistics package.

Table 5 shows a printout for the summary statistics calculated on the pumpkin-maize intercrop. This is not suitable to be presented in your report as an unmodified printout for a number of reasons. First, notice that the variable name 'Treatment' has been reduced to eight characters by the statistics package. Next, unhelpful treatment numbers rather than informative names are given. Most importantly there are statistics given by this printout that may not be appropriate, or that you may not even understand. The statistics may not be

Table 5. Descriptive statistics of the weed biomass of the pumpkin-maize intercrop data (see text for explanation as to why this table should not be presented in this form)

Descriptive Statistics

Variable	Treatmen	N	Mean	Median	TrMean	StDev
Biomass	1	3	83.5	84.5	83.5	27.8
	2	3	79.3	98.4	79.3	58.5
	3	2	28.30	28.30	28.30	4.92
	4	3	6.81	5.44	6.81	5.84
	5	3	87.6	73.3	87.6	36.3
	6	4	6.24	7.03	6.24	2.27
	7	3	3.527	2.960	3.527	1.136
	8	3	5.49	7.32	5.49	3.24
	9	3	162.2	160.0	162.2	18.9
	10	3	30.92	34.89	30.92	9.14
	11	3	40.83	38.03	40.83	6.01
	12	3	3.14	1.73	3.14	3.10

Variable	Treatmen	SE Mean	Minimum	Maximum	Q1	Q3
Biomass	1	16.1	55.2	110.8	55.2	110.8
	2	33.8	13.7	125.9	13.7	125.9
	3	3.48	24.82	31.78	*	*
	4	3.37	1.79	13.21	1.79	13.21
	5	20.9	60.7	128.8	60.7	128.8
	6	1.14	2.91	7.98	3.87	7.81
	7	0.656	2.785	4.835	2.785	4.835
	8	1.87	1.75	7.39	1.75	7.39
	9	10.9	144.4	182.1	144.4	182.1
	10	5.27	20.48	37.41	20.48	37.41
	11	3.47	36.73	47.72	36.73	47.72
	12	1.79	1.00	6.70	1.00	6.70

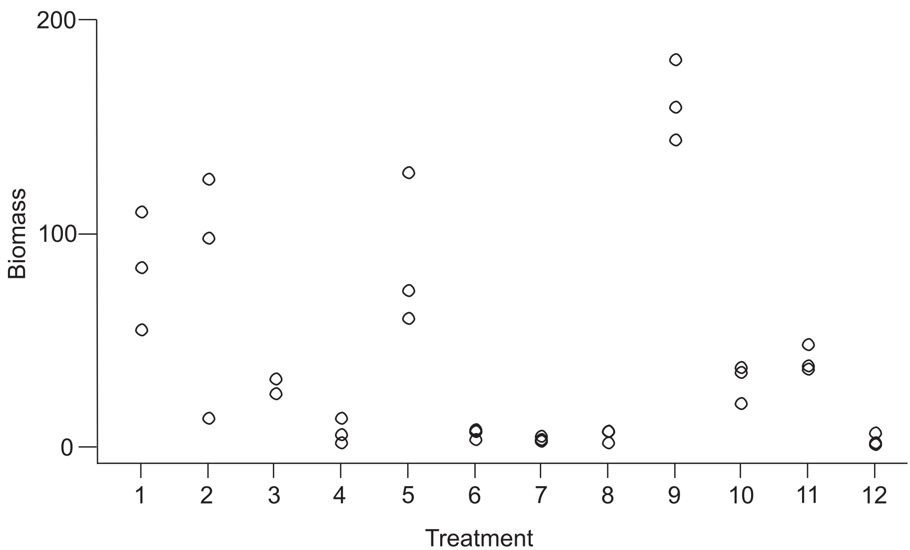


Figure 2. Dotplot of the weed biomass for the pumpkin-maize intercrop experiment showing each of the treatments

appropriate for the data structure and the statistics may not answer the research objectives. The student is also allowing the statistics package being used to have undue influence on the analysis and the presentation of the results and to distract him/her from the objectives of the research. This is also an extremely untidy table which is difficult to read! Tables are better if they don't go over a row for each treatment, and 'treatment number' would make more sense if it was replaced by a text label.

Figure 2 shows a type of exploratory graph (dotplot). **Dotplots** show the spread of the data. Notice how the points for Treatments 1, 2, 5 and 9 are more spread than those of the other treatments. The graph is still labelled with unhelpful treatment numbers rather than names but maybe that does not matter. This is an example of a graph which is important to you in analysing the data – it shows that some treatments are much more variable in weed biomass than others. That is something you may need to take into account in your analysis. But, unless it relates to a key analysis objectives, you will not need to include this graph in your report. Hence, its inelegant layout is not a problem.

Other plots like **boxplots** and **stem-and-leaf plots** are available and should be tried. If the data have two related variables, for example, yield and amount of fertilizer applied, **scatterplots** should be plotted to check whether a **regression line** can be fitted.

The analysis shown so far should be repeated for each response variable in a data set. You can see that by this stage you could already write a fair amount on the data patterns and subjectively suggest results. For example, in the intercropping example you could suggest which is the best crop and weeding regime to use. But there are some severe limitations to this analysis. Two important ones are:

1. Only simple patterns can be investigated. You can look at how y varies as x varies by plotting y against x. But what if there are several x's, all to be considered simultaneously?
2. There has been no consideration of the uncertainty in any of the summaries that are used to interpret the data. Yet we know there is variation in the observations, so there is uncertainty in the results.

The formal analysis addresses these problems. But, you should note that although Excel is useful for the descriptive analyses described so far, Excel is not good for more formal analyses and modelling. Use a reputable statistics package such as Minitab, Genstat or SAS.

Formal analysis and statistical modelling

Key points for this section include:

- Complex patterns involving several variables at the same time can be investigated by fitting statistical models to the data
- Much of the formal statistics taught in introductory courses aims at providing information about the precision of estimates used to interpret the data
- Formal statistical analysis should never be an end in itself, but part of data interpretation.

The next stage in the analysis after describing the patterns and identifying any outliers is to confirm them. This is done by fitting **models** and carrying out a **confirmatory analysis** of the subjective results. Some of the confirmatory analysis may also be used to look at pattern and residuals, which means the exploratory data analysis stage (and the data checking) is not yet completed.

When doing 'research' you will usually wish to generalise from your data to some wider population. This is what statistical inference is all about. So you are likely to find that descriptive stuff is insufficient. You are after all doing a research degree. If there is no generalisation, then there may be no research – and you might not get your degree!

It is at this stage that you will need to get some idea of the precision and accuracy of your results. **Precision** is the closeness of the data points to each other. It is often measured using **variance**, whereas **accuracy** is the closeness of the data points to the true population value.

Regression

Statistical models are mathematical representations of the pattern in data. The simple regression model is often taught in basic statistics courses as it illustrates many important concepts. A regression is the fitting of a straight line to data to describe and predict the relationship between two variables. These variables consist of a response variable or **dependent variable** and the variable to which it responds, the **independent variable.** In the following regression example (Figure 3) the relationship between inorganic soil nitrogen and crop yield was investigated.

Checking the nature of the relationship

Before any model is fitted you must investigate the **nature of the relationships** between the variables. With a regression model an attempt is being made to fit a straight line to the relationship, consequently you must check first if there is a straight line relationship. Any model can be fitted to any data, but this doesn't mean a relationship actually exists. One of the biggest errors you can make is to fit a model without checking that the model is sensible for the data.

Fitting the line

In Figure 3 there appears to be a straight-line relationship between the two variables, although there is some variation about this line. After carrying out the regression analysis it was found that in the model, the regression line is:

Yield = 1.23 + 0.142 inorganic soil nitrogen

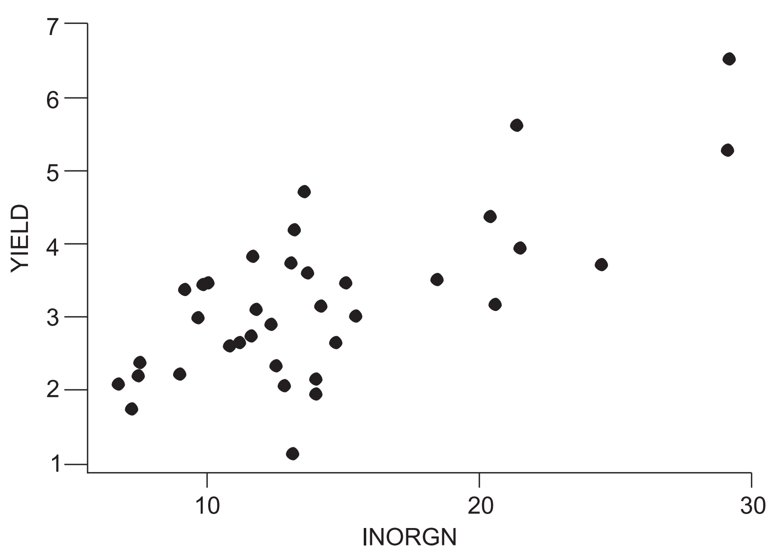


Figure 3. Relationship between soil inorganic nitrogen and crop yield

Another common error is to include all the regression output generated by computer software in the text of the report. If you really need to include the output, it is best put into an appendix with the key points summarised in the text. This is also a common mistake when carrying out the ANOVA and can be dealt with in the same way.

Your work is not complete when you find the model or regression line. You must still check to see if you have fitted an appropriate model and if each of the parameters should be in the model. This is done using the residuals and carrying out a number of significance tests. You must indicate in the methods section of your report that such checks have been carried out, or the reader will assume they have not been done and question the validity of the models that you have produced (see any good statistical text or ask a statistician for details on how to check residuals).

Part of exploratory data analysis is checking patterns of the residuals – there should be no patterns if you have picked up the pattern and structure of the data correctly. However, you often cannot check the residuals until you have actually fitted a model because the residuals are the result of fitting the model.

Confirmatory analysis

Having checked that a model is now possible you need to look at the confirmatory statistics supplied in the output. This is the **statistical inference** part of the data analysis. There are several concepts you need to understand before you can do the next part of the analysis. These are **estimates (point and interval)** and **tests of significance.**

It is the ideas of **inference** and **modelling** that students are usually taught in university statistics courses, but find difficult. So you may have to review the key ideas of estimation, confidence intervals and significance testing. Often this has been covered in your statistics course, but in ways that are difficult to relate to your needs in analysing your research data. This is perhaps because your statistics course was too theoretical. It may not have included realistic examples. Some courses still do not integrate the use of the computer with the discussions of the concepts of statistical modelling. Also, you may not have been very interested at the time, perhaps because you had convinced yourself that the ideas were difficult.

There are now many resources that you can use to review the ideas you need without using too much mathematics. The references at the end of this chapter are for students who need to review such ideas. But don't leave this too late in your thesis writing. The later you leave it, the more pressure you will be under to finish the thesis, so you will not be able to concentrate on reviewing ideas that are not central to the needs of a current chapter.

If learning statistics was a problem for you, then remember it might also have been a problem for your supervisors. They may be hoping that you will be more comfortable with statistical concepts than they were! Even if they now like statistics, be aware that some supervisors may cling to one or two favourite methods of analysis. These may not be the only methods that can now be used to process your data.

You should understand these ideas, because **you should not do analyses that you do not understand. The rule remains to analyse the data in ways that are dictated by the objectives of your study.** For example, suppose you use a method called principal components in your analysis. You do not necessarily need to understand all the formulae that underlie this method. But you must be able to explain (perhaps in an oral examination) why you have used this method and how the results have contributed to your understanding of the data in relation to the objectives of the analysis. It is not sufficient to say:

- 'It is the common method that everyone seems to use
- My supervisor said I should use this method
- An article I found in an international journal used this method.'

These are all sensible reasons, but it is your research and your data. You must be able to explain why each method is appropriate for your own work. This is rarely a big problem, but it can loom large, because you may not feel confident about the topics, and hence feel that you are unable to decide how to proceed. In such cases it is good if you encourage communications, perhaps the statistician and your main supervisor could meet to discuss their differences. If they meet, you may find they are discussing general issues of principle, and straying from your well defined problems of how to analyse your data. If so, then the key is (as always) that the analysis must help in the objectives of your study. In the end you may have to make some compromises (see Table 6, and the section on ANOVA).

Analysis of variance (ANOVA)

Another type of modelling that you will commonly come across in agricultural research is the **analysis of variance**. As the name suggests, it allows you to determine how much of the variation in the response can be attributed to different treatment factors or other effects. For further details on ANOVA, you can refer to any good experimental design text book (Mead *et al.*, 2003, or the Statistical Services Centre, University of Reading website).

Investigating pattern in ANOVA

The ANOVA can also be used for further investigation of pattern, by using it to generate plots of the two factors and the responses and tables of means and examination of the residuals resulting from the fitted model. This means the data are examined using more than one source of variation at a time (combining the two factors instead of examining just one). In Figure 1, for example, you can see the effect of weeding, of the crop and of any interaction (non-parallel lines).

Confirmatory analysis in ANOVA

Figure 1 shows us that weeding treatments 2, 3, and 4 generally give a much a lower weed biomass than weeding treatment 1, where no weeding was done. The plot also shows an interaction – the lines are not parallel. Such results are often presented with **significance levels** (**P-values**). Make sure you know what these mean and how to interpret them. You should not be using such statistics if the implications and assumptions on which they are based are not clear.

The same is true when it comes to presentation of results. For example, many supervisors (and journal editors) insist on placing 'letter values' adjacent to the numbers in tables of means, as in Table 6a. Here a supervisor insisted that the results of a multiple comparison test are included. He always does this, and it was the key in his thesis, which was in the same area as this work. A statistician may well have other ideas on how the results are best presented, perhaps as in Table 6b.The statistician does not see how these tests help in the analysis, and proposes that just the standard error of the differences between the means is presented instead, while also matching the layout of the table to the structure of the treatments. You need to understand enough of the statistical ideas to choose between these (and other) presentations, and to defend your choice in front of supervisors, examiners and editors.

Table 6. Tables of mean weed biomass for the pumpkin–maize intercrop experiment

a. Treatment	Mean weed biomass (g/m^2)[1]
1. Sole maize with no weeding	83.50b
2. Sole maize with weeding at 3 weeks	79.30b
3. Sole maize with weeding at 3+5 weeks	28.30a
4. Sole maize with weeding at 3+5+8 weeks	6.81a
5. Pumpkin–maize intercrop with no weeding	87.60b
6. Pumpkin–maize intercrop with weeding at three weeks	6.24a
7. Pumpkin–maize intercrop with weeding at 3+5 weeks	3.53a
8. Pumpkin–maize intercrop with weeding at 3+5+8 weeks	5.49a
9. Sole pumpkin with no weeding	162.20c
10. Sole pumpkin with weeding at 3 weeks	30.92a
11. Sole pumpkin with weeding at 3+5 weeks	40.83a
12. Sole pumpkin with weeding at 3+5+8 weeks	3.14a

1. Means with the same letter are not significantly different (5% LSD)

b.	Weed biomass (g/m^2)		
Weeding	Sole maize	Intercrop	Sole pumpkin
None	83.50	87.60	162.20
3 weeks	79.30	6.24	30.92
3+5 weeks	28.30	3.53	40.83
3+5+8 weeks	6.81	5.49	3.14

Average SED = 18.5

There is a need to examine the appropriateness of the model, and the model fitting itself can lead to further data exploration and understanding of the results. In the pumpkin–maize intercrop, the ANOVA table showed some significant effects, but the analysis of the residuals showed that the model was not appropriate (the residuals showed inconsistent variances across the treatments and the normal probability plot of the residuals was not a straight line). If you only fit the model for the significance levels to test your null hypotheses you will miss the real information in your data. You might produce significance levels with no understanding of what really happened. The question is, you found a significiant result, but so what? You have missed the patterns and information in the data and you have yet to prove that the 'significant' model is actually appropriate.

Mixed modelling

ANOVA can be difficult to apply to situations where there are multiple sources of random variation – such as those between villages, between farms within villages and between plots within farms. An approach to modelling, called **mixed modelling** is now available to deal with these situations. This is an important statistical development for analysis of many field studies, both survey and experiment. See Allan and Rowlands (2001) for further information.

The methods touched on briefly in this chapter are the main methods that have been used in agricultural research in Africa.

Note that the value of any statistical method depends on what you want to find out, and the nature of the data and the research design that generated it.

It does not depend in any way on whether your data came from a research station or from farms, whether the study was participatory, or whether it relates to the biophysical, social or economic aspects of a problem.

Making sure you satisfy the research objectives

Important concluding points to remember:

- Revisit objectives and make sure they have been met
- Check that all the analyses you include in your thesis really contribute to those objectives
- Don't work in isolation.

Once the analysis appears to be complete, you need to revisit your objectives and make sure that they have been fulfilled. Remember, the whole process of data entry and analysis should have been iterative and should aim to meet the objectives of the research. The analysis that you ended up doing may not be the same as that planned in the original project proposal. At this stage you need to revisit your objectives and relate them to the results you have obtained.

A way to start this process is by laying out the original proposal objectives, tables, graphs, outputs, descriptions, conclusions and interpretations in front of you. Lay out the results in the same order that you lay out the methods. The sequence should be logical and should not jump around from topic to topic. Now try the 'so-what' test. Check that every item of statistical analysis you are going to report actually contributes to the conclusions that you have reached. Check that the conclusions match the original objectives. Are your conclusions really conclusions? Make sure you haven't read into the data something that you would like to see there. This is really easy because you have been so close to the whole project it is difficult to divorce yourself from it and see it objectively. Now relate and interpret your results to the literature. At this, as in all stages of the research process, it is important that you don't become tempted to work in isolation. Talk to your colleagues, get help and give seminars periodically so that you can get feedback from those around you.

Remember, don't waste your data by only carrying out significance tests and that you don't have to be a hot shot statistician to get really good information out of your data.

Resource material and references

Appendix 10. Coe, R., Stern, R., Allan, E., Beniest, J. and Awimbo, J. 2002. *Data Analysis of Agroforestry Experiments*. World Agroforestry Centre (ICRAF), Nairobi, Kenya.

Appendix 11. ICRAF. 2003. *Genstat Discovery Edition and Other Resources*. World Agroforestry Centre (ICRAF), Nairobi, Kenya. On CD.

Allan, E. and Rowlands, J. 2001 *Mixed Models and Multilevel Data Structures in Agriculture*. Statistical Services Centre, The University of Reading, UK. http://www.reading.ac.uk/ssc/

Jones, A., Reed, R. and Weyers, J. 1998. *Practical Skills in Biology*. Second edition. Longman, UK. 292 pp.

Mead, R., Curnow, R.N. and Hasted, A.M. 2003. *Statistical Methods in Agriculture and Experimental Biology*. Third edition. Chapman and Hall, London, UK. 472 pp.

Muzamhindo, N. 1999. *Analysis of Experimental Data for Maize Crop at University of Zimbabwe Farm*. BSc. Honours Project, Department of Statistics, University of Zimbabwe, Harare, Zimbabwe.

Stern, R.D., Coe, R., Allan, E.F. and Dale, I.C. (Eds.). 2004. *Statistical Good Practice for Natural Resources Research*. CABI Publishing, Wallingford, UK. 387 pp.

The Research Support Unit of the World Agroforestry Centre (ICRAF) have some training materials and other guides on analyses. www.worldagroforestrycentre.org/research support.

Velleman, P.F. and Hoaglin, D.C. 1981. *Applications, Basic and Concepts of Exploratory Data Analysis*. Duxbury Press, Boston, Massachusetts, USA. 354 pp.

4.8 Models

Catherine Wangari Muthuri

- **A model is a simplified representation of part of the real world. In this chapter we discuss models that can be described mathematically**
- **Models are based on theory. In research models help to test theory by making predictions that can be compared with observations**
- **Models also allow the implications of research results to be explored by making predictions for new situations**
- **Each model is built for a specific purpose. A model that is useful for one job may be inappropriate for another task on a similar topic**
- **Models vary in scope from the simple, which you can put together and use very quickly, to the complex that may take much of your project time to develop and use**
- **Computing tools designed for the job can make modelling feasible for students who are not specialists**

Introduction

Modelling can mean many things in research, and models of one sort or another play a crucial role in much research. Experience shows that the role and use of models is rarely explained in research methods courses. The result is that many students have only a vague idea about what models can and should be doing for them. Modelling is often regarded as the domain of specialists who sit hunched over computers, not of agricultural researchers who want to solve real problems in the field. The result is that much research is less effective than it might be. The aim of this chapter is to start to fill that gap.

The chapter is divided into three major parts. The first shows you how models are a natural part of the research process. This is to help you develop your ideas from the general 'models are everywhere' to the main focus of the chapter, which is concerned with **mathematical** or **simulation** models. The second part discusses your options if you plan to do some mathematical modelling. Finally, details of the steps you need to follow to construct, use and test simple models are described, using examples where modelling tools have been applied in research studies in Kenya. Research findings can be enriched by the use of simulation models and this is an attempt to encourage you not to shy away from using modelling tools just because you don't like maths!

Model types

Models are everywhere

You may not be aware of them, but you are using models all the time. They come as physical models in all shapes and sizes from dolls, miniaturised cars and aeroplanes and globes, or as visual representations in maps or pictures. They may be presented as verbal or mental models, or in more abstract arithmetic or algebraic form, in nearly all we learn. **A model is just a simplified representation of part of the real world.**

Physical models have been used for centuries in research. Engineers use models of boats to study their stability and resistance to movement through the water. In biological research one species is often said to 'model' another; in the early stages of medical research monkeys and mice are used to model man, because they represent *some* aspects of human physiology well. The images we carry in our minds, i.e., mental models, are simplified representations of complex systems. We use them constantly to interpret the world around us and we usually do not realise that we are doing so.

None of these models involve the complete similarity of real world and model, but similarity in key features. A model is useful if it behaves in a realistic way for your problem. The scale model of a ship may be useful for investigating its stability in the water, but it will be useless for determining the profitability of operating the ship. Different models of the same phenomenon are useful for different things. Take a 1-ha farm as an example. A map of the farm (a visual scale model) might be useful when the farmer is planning the location of different crops. Physical models of the landscape, built up from clay and painted, can be used to examine the interaction of the farm with neighbouring farms and other land areas. Numerical input–output models help in making investment decisions. Detailed numerical topological models can be used to understand water flow and erosion on the farm. Each of these is a 'model of the farm' and each is useful for its own purpose, but inadequate for other purposes.

Models in the research process

Research involves developing a theory of the real world and testing it with observation, then perhaps using it to explain and predict further phenomena. Models are representations of the theory and hence a fundamental part of the research process. Whether the model needs to be formalised and described mathematically depends on whether the predictions of the theory can be worked out without formalisation.

Models can be used in two steps of research:

1. In generating hypotheses or predictions, that will suggest the observations of the real world that need to be made.
2. In assessing the extent to which our theory (as captured in the model) explains the real-world observations.

If the model and observations agree then there is nothing in the data to suggest the theory is not a good description of the real world. But of course we might have collected data that does not test the theory in ways that are interesting! An important part of research design is planning observations that do discriminate between models which are fit and unfit for their intended purpose.

If the model and observations do not agree then you can:

- Question the model structure and assumptions, and revise it
- Question the data: perhaps it is not really relevant to the model you have chosen
- Abandon the line of research.

Mathematical models

This chapter is about the mathematical models that are used in agricultural research. If the relationships and rules that make up the model are sufficiently well specified, then they can be written down mathematically and produce numerical results. In very many models the basic mathematical relationships and rules are simple (such statements as 'volume = mass/density 'or 'yield is zero until after flowering'). Complex patterns of results often emerge because of the many interacting components, rather than because there are some complex mathematical ideas embedded in the model. This is important. **It means you do not have to be a mathematician, or even very good at using mathematics, to make effective use of models in your research.**

A **mathematical model** is a set of equations that represent interconnections in a system, and can be worked out either by hand or by using a computer. The equations are written in terms of mathematical objects that correspond directly to physical quantities. If these objects change as part of the phenomenon they are generally called **variables** while if they are fixed they are generally called **parameters.**

Typically a model will consist of formulae that link some responses or quantities of interest with inputs, or the things that affect them. For example, a simple model of soil moisture changes is illustrated in Figure 1.

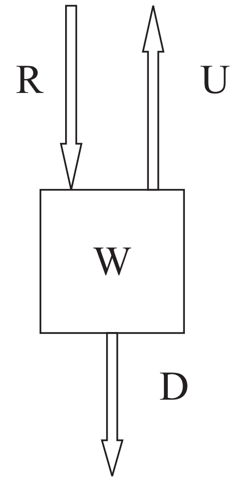


Figure 1. Simple model of soil moisture

The soil moisture (W) at time t is W_t. Rainfall is R_t, uptake by plants is U_t and drainage is D_t. The model can be written mathematically as:

$W_{t+1} = W_t + R_t - U_t - D_t$

If we know the initial conditions (W_0) and the values of R, U and D then we can calculate W at any time. The model is simplistic. It ignores soil evaporation. That will not be a problem if the model is being built for applications in which soil evaporation can be ignored, but would be a major deficiency in other cases. The model also requires inputs that might be hard to measure (U and D). For some purposes you might be able to predict U, by adding another part (some more components) to the model. For some purposes you could take:

$U_t = c.P_{t,}$

where P_t is the potential evapotranspiration and c is a 'constant'. This model might well be useful for studying the effect of day-to-day changes in P on W. However it is still too simplistic for longer-term studies, as c will probably not be constant, but will change as the crop grows and matures. The value of c may also depend on W, with the plants able to take up less water when the soil is drier. It is easy to see how this process can quickly lead to models of ever-increasing complexity, even though each step involves simple and realistic relationships.

Part of the skill in modelling is in choosing the components to model, including the things which will be necessary but not putting in everything you can think of.

Conceptual and empirical models

Models can either be **empirical** (data-driven) or **theoretical** (theory-driven or conceptual). An **empirical** model is based mainly on data. It may be used in statistical analysis of study results and to predict within domains of 'similar' conditions to the empirical base. It does not explain a system. For example, a fertilizer response curve is an empirical model. It can be developed from observations on the yield of crops with different amounts of fertilizer, and used to predict the yield at any fertilizer level. However, it does not explain why the yield response is the way it is. An empirical model consists of one or more functions that capture the trend of the data. Although you cannot use an empirical model to explain a system, you can use such a model to predict behaviour. We use data to suggest the model, to estimate its parameters, and to test the model. An empirical model is not built on general laws and is a condensed representation of data. However many statistical or empirical models are built on elements of an underlying theory, for example, we construct the input variables in a regression model based on a theoretical understanding of factors that should determine the response.

A conceptual, theoretical or 'process based' model includes a set of general laws or theoretical principles. If all the governing physical laws were well known and could be described by equations of mathematical physics, the model would be physically based. However, all existing theoretical models simplify the physical system and often include obviously empirical components. Thus the distinction between conceptual and empirical models is not clear-cut. And again, it is the modellers job to use something appropriate for the task, rather than to assume that one approach has more intrinsic value than another.

Roles of models

Models play several roles including:

- **Exploring** the implications of theory. It may not be possible to see the implications of theories that involve several interacting components without calculating what happens in different conditions. Used in this way, models provide insights and add creativity
- **Prediction** or forecasting tools help users make sensible educated guesses about future behaviour. These can be used in planning, scenario analysis and impact analysis
- **Explaining** observations and generating hypotheses
- **Training** so that learners can carry out 'virtual experiments', exploring the result of making changes.

In research models can help answer such questions as:

'Can I construct a theory that explains my observations?'
'Is my hypothesis credible?'
'What new phenomenon does my theory help to explain?

Used for prediction, models can answer such questions as:

'Given the model, what will happen in the future?'
'Given the model, what's going on between places where I have data?'
'What is the likelihood of a given event?'

How to model

You have three options if you decide to use simulation models in your work. You can use an already existing developed model, modify an existing model or develop a new model altogether.

Using an already developed model

Hundreds of models relevant to agricultural research have been developed and described and are available to you. A few have to be purchased. Many are available free to researchers and can be down loaded from web sites or obtained from the authors.

The advantages of using a model that someone else has developed include:

- **Time saving.** Some of the hard work has already been done
- **Recognition.** Some models have been widely used and described. Their value is already recognised so you will find it easy to justify their use
- **Support.** You will find documentation, examples and maybe technical assistance in using the models.

However there are also disadvantages, compared to the alternatives of developing your own models. These include:

- You may not find a model that actually describes the phenomena in which you are interested at the right level of simplification
- The available models may require inputs that are not available to you
- You may not fully understand how the model is constructed (the theory on which it is based)
- The model may not run on any computer available to you, or in the way you need for your research.

If you are considering using a model, then select it by:

1. Determining exactly what you want to do with it. You will only be able to decide if candidate models are suitable when your task is clear.
2. Searching literature and the Internet for references to models that tackle your problems, and asking experts in the field.
3. Evaluating each possible model against your requirements. If you end up with more than one candidate then choose the simplest.

Modifying an already existing model

You may well find that no available model meets your requirements but that some come close. Therefore it may be desirable to modify a model. Modifying it may mean anything from changing the way input files are handled to adding to or changing some of the underlying theory. Often modification will mean adding a description of further components and processes to address a specific situation.

If you plan to adapt or modify an existing model, all the points above about selecting it apply. In addition you will have to be able to:

- Get access to the original computer code and description of the theory behind it
- Understand them fully
- Know how to modify it for your needs.

The computing skills you need will probably be more than those you need to just run an existing model.

Some models are much easier to adapt than others. If they were originally designed and produced with adaptation in mind then the task may be straightforward. If they were not built to be adapted the task of modification may be all but impossible.

Adapting a model takes longer than using an already existing model. You need to go though all the steps in the modelling process that are discussed later in this chapter. This implies that the exercise becomes a major component of your research. It therefore demands that you have sufficient skills and are familiar with the language of the packages and software used.

Developing a new model

The third option is to develop your own model. Situations that necessitate developing models include those when:

- The outputs generated and inputs required are not catered for in the existing models
- Exiting models are too clumsy or complicated, or have a poor track record
- You are working in an area where no existing models can be found.

Given the novelty of most research, the last is likely to be the case.

Building and using your own models could be:

- Something that takes a few hours, if you are simply looking at a few interacting components and are familiar with a suitable computing environment

- Something that takes most of your 3 years as a PhD student!

More likely it will be somewhere between the two. The steps in developing a model are outlined below. The most critical are the first ones: **defining useful** and **realistic objectives.** You will probably be most successful if you start with simple objectives. Reduce the problem to its simplest objectives, and work on the simplest model that will meet those. This might be a model with no more than two interacting components and simple rules describing them. Yet even these models can give insights into your theory and observations that are not apparent until the model is formalised.

Steps in modelling

The steps involved in the modelling process are summarised in the flowchart (Figure 2). However, developing any useful model will be an iterative process – you will certainly have to return to early steps, for example, if you are looking again at the interactions in your model when it does not seem to give sensible predictions.

The model-building process can be as enlightening as the model itself, because it reveals what you know and what you don't know about the connections and causalities in the system you are studying. Thus modelling can suggest what might be fruitful paths for you to study and also help you to pursue those paths.

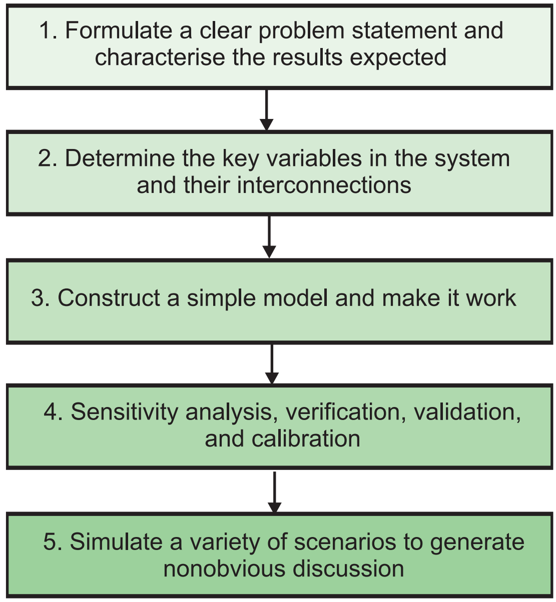


Figure 2. A summary of the steps involved in the modelling process

Formulate a clear problem statement and characterise the results expected

As with all other aspects of research, what you do depends on what you want to find out. Setting realistic and detailed objectives for your modelling will determine the whole nature of the task. It will help you decide on the following important characteristics of models:

1. Will the model need to be deterministic or stochastic?

In **deterministic** models the future state of the system is completely determined (in principle) by previous behaviour. In **stochastic** models the system is subject to unpredictable, random changes. These models involve probability and statistics. If you are interested in risks, your model will have to use stochastic components.

2. What timescale is appropriate?

The **timescale** of the processes in question determines the timescale of the models. Depending on the time taken for the processes under question to reach an equilibrium or to be felt, useful decisions on what to include/exclude in the model can be reached. For example, when looking forward 100 years, you need to ignore daily/monthly or seasonal variations of the parameters in question. Such variations can be ignored in a long-term model but could be important in a short-time model. Examples of scales and typical times are:

- Metabolic (enzyme-catalyzed reactions; seconds to minutes)
- Epigenetic (short-term regulation of enzyme concentration; minutes to hours)
- Developmental (hours to years)
- Evolutionary (months to years).

3. Does the model need to be spatial?

All agriculture takes place in a spatial context, but only some problems require you to specifically describe spatial interactions. Think of the problem of modelling small farms. If you want to describe economic inputs and outputs of the farm you need to know that there are crops, animals and trees, but it may not matter where on the farm they are. If you want to model nutrient flow between tree and crop plots, then their location matters and the model you use will have to be explicit about that. Many of the management decisions made by small-scale farmers living in heterogeneous environments make use of spatial variability on their farms, such as growing different crops on different patches of land, abandoning part of their land, or focusing their efforts only on those patches with the highest returns to investment of labour or inputs. Most of the current models in agriculture do not handle spatial variability well, if at all. There is a clear need to develop existing models further, or to construct new ones, in order to address this limitation. Unfortunately, the structure of many existing models does not facilitate transformation to spatially explicit versions, as their linear nature restricts them to being run in sequence many times, in order to simulate each patch of land in turn. This makes it difficult to simulate simultaneous interactions between patches of land (e.g., soil, or water flow down a gradient). In circumstances where spatial variability is a key factor affecting the study it is advisable for you to explore using a model that takes this into consideration.

Determine the key variables in the system and their interconnections

In this step you need to determine the key variables in your study that will be represented by variables in the computer model. **Key variables** are the few most important, significant factors that affect the system and their relationships. The cause- and-effect connections in the real system will be represented by interconnections in the computer model. Adding more and more interconnections makes the model complex, though by design, models should be a simplification of the system under study. A determinant of model usefulness is therefore the ability of the modeller to leave out unimportant factors and capture the interactions among the important factors.

Note that a model is:

- Too complex when there are too many assumptions and relations to be understood
- Too simple when it excludes factors known to be important.

Constructing a model

Building a model is an interactive, trial and error process. A model is usually built up in steps of increasing complexity until it is capable of describing the aspects of the system of interest. **Note: It will never 'reproduce reality'.**

The appropriate tools you need to construct a model depend on the complexity of the model. The simplest tools may be **paper and pencil.** Others may use **spreadsheets,** while the more complex models may require **dedicated modelling software** that uses its own language. The simplest mathematical model takes the form of **equations** show how the magnitude of one variable can be calculated from the others and **spreadsheets** like Excel are adequate for the task.

More complex computer simulations use special software that allows the building and testing of a model. There are software products available that make building and running some types of models very easy even if you know nothing about computer programming. Investigate such software as STELLA and ModelMaker before trying to write your own code in lower-level computing languages. They make the job of developing and running your own models very much simpler!

The development of the simple soil water model outlined in Figure 1 is shown here to give you an idea of what is involved. The model represented in Figure 1 is drawn in STELLA. In Figure 3a. STELLA uses four main types of building blocks:

Stocks. These are stores of 'stuff', represented by rectangles. They may describe water, money, people, biomass,... whatever you are modelling.

Flows. These are the movements of material into and out of stocks, represented by broad arrows. The arrow can be thought of as a pipe, with a tap on it to regulate the flow. Sources and sinks of the material are represented by 'clouds'.

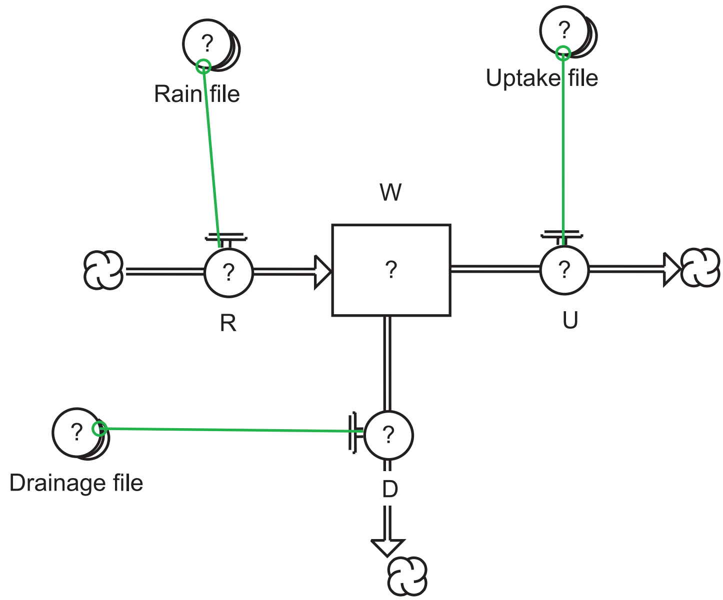


Figure 3a. Simple soil water model in STELLA

Converters. These are represented by circles. They hold values of constants and formulae used to convert one type of material to another.

Connectors. These narrow arrows show the logical connections between components in the model. The equations describing the model must be consistent with these connections.

The stock of soil water (W) has an inflow of rain (R) and outflows of uptake (U) and drainage (D). The actual values of these are read from data files. The model is completed by filling in a formula or other details in each location marked by '?'. The model can then be run.

In Figure 3b the uptake is now calculated as c.P, where P is the potential evapotranspiration (PeT), also read from a file. It should be clear from this that modifying the model requires little more than adding components to the diagram. The real challenge of course is deciding *how* to model uptake, not changing the computer code – this is why software such as STELLA is so important. The final step (Figure 3c) shown here displays two more changes that the modeller thought would help. The drainage is now calculated (because there was no measured data available) and the uptake now depends on both the crop biomass and the soil water. The latter involves keeping track of the biomass growth, a second stock in the model. Many physiologists would be uncomfortable with a single 'type' of biomass, and start differentiating it into, say, roots, stems, leaf and grain. Then you need to add components that describe what the partitioning depends on. Similarly the soil scientist would like to have several soil layers, each with different hydraulic properties. The model can quickly become complex. The value of software such as STELLA is that it allows you, as researcher, to think about what constitutes a sensible model for you, rather than worrying about computer codes.

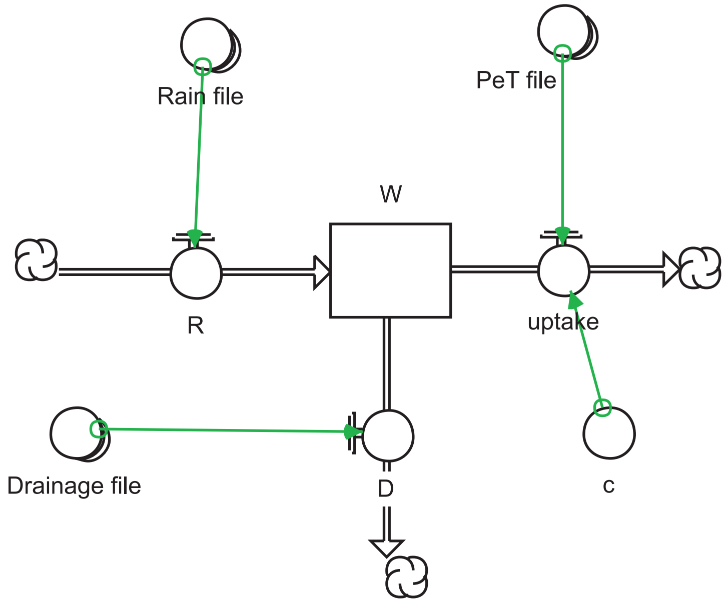


Figure 3b. Simple soil water model with uptake modelled as c.PeT

Sensitivity analysis, validation, verification and calibration

Sensitivity analysis

Through sensitivity analysis, you can gain a good overview of the most sensitive components of the model. Sensitivity analysis attempts to provide a measure of the sensitivity of other parameters or forcing functions, or sub-models to the stated variables of greatest interest in the model. It helps you to

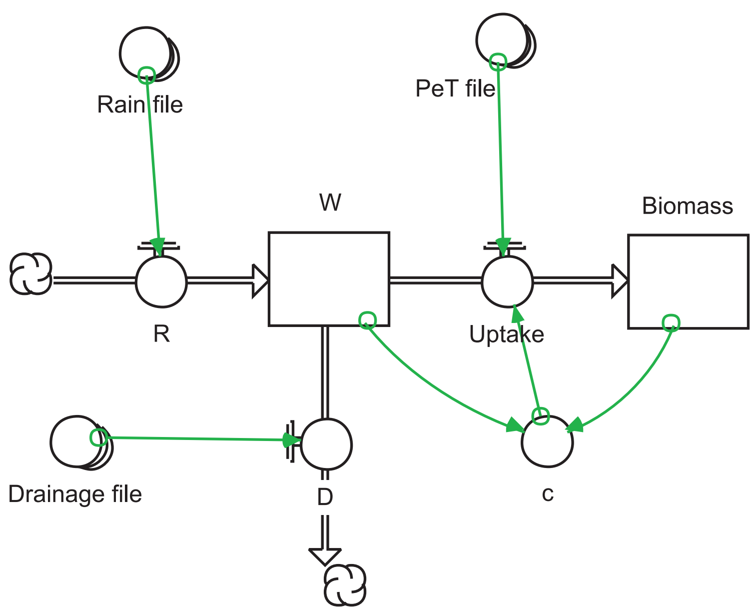


Figure 3c. Simple soil water model with uptake depending on both crop biomass and soil water

systematically explore the response of the model to changes in one or more parameters, to see how sensitive the overall model outcome is to a change in value. This **sensitivity** is always dependent on the **context** of the setting of other parameters, so you should be careful about the conclusions you draw. Some parameters only matter in particular types of circumstance. Others, however, seem to always matter, or to matter hardly at all. This type of model analysis is used to see which parameters should get priority in a measurement programme. You must be provided with affordable techniques for sensitivity analysis if you are to understand which relationships are meaningful in complicated models. This is equally true whether you are using an already developed model, modifying a model or developing one.

Validation, verification and calibration

In general, **verification** focuses on the internal consistency of a model, while **validation** is concerned with the correspondence between the model and the reality. **Calibration** checks that the data generated by the simulation matches real (observed) data, it can also be considered as tuning of existing parameters.

These steps can be among the most conceptually difficult. No model is universally 'valid' in the sense that it will give 'correct' predictions in all circumstances. There will always be discrepancies between observed and predicted values. These discrepancies can be made smaller by calibration and by making adjustments to the model. However this does not necessarily increase the usefulness of the model in either: explaining your observations of the real world, or making predictions about behaviour in the real world.

Simulate a variety of scenarios to generate non-obvious discussion

Simulation models have been used widely in Kenya to address various problems. Three examples are given to help you see how they can be used.

Soil fertility management in western Kenya: Dynamic simulation of productivity, profitability and sustainability at different resource endowment levels

A farm economic–ecological simulation model was designed to assess the long-term impact of existing soil management strategies, on-farm productivity, profitability and sustainability. The authors developed a model that links biophysical and economic processes at the farm scale. The model, which runs in time units of 1 year, describes soil management practices, nutrient availability, plant and livestock productivity, and farm economics. It concluded that low land and capital resources constrain the adoption of sustainable soil management practices on the majority of farms in the study area. Previously it had been assumed that low-input organic methods were suitable for the poorest farmers. For more details, see Shepherd and Soule (1998).

Modelling leaf phenology effects on growth and water use in an agroforestry system containing maize in the semi-arid Central Kenya using WaNuLCAS.

The three tree species under study were *Grevillea robusta* (evergreen), *Alnus acuminata* (semi-decidous) and *Paulownia fortunei* (deciduous). The inputs included climate data, soil data, calendar of events, crop and tree parameters, agroforestry zones and layers, and leafing phenologies. The scenario outputs included soil water balance, tree and crop biomass and stem diameter. WaNuLCAS model simulations demonstrated that altering leaf phenology from evergreen through semi-deciduous to deciduous decreased tree water uptake and interception losses but increased crop water uptake, and drainage rates in all the species. It was therefore concluded that deciduous tree species would compete less with crops and be more advantageous in increasing stream flow than evergreen trees. Phenology had not previously been a major consideration in determining tree selection For more details, see Muthuri (2003).

Modelling the benefits of soil water conservation using PARCH; A case study from a semi-arid region of Kenya.

The PARCH model was used to simulate maize grain yield under three soil/water conservation scenarios: 1. a typical situation where 30% of rainfall above a 15 mm threshold is lost as runoff, 2. runoff control, where all rainfall infiltrates, and 3. runoff harvesting, which results in 60% extra 'rainfall' for rains above 15 mm. The study showed that runoff control and runoff harvesting produced significant maize yield increases in both the short and the long rains. Previously runoff control was justified more for erosion benefits than increased crop production. For more details, see Stephens and Hess (1999).

Conclusions

The success of models developed by physicists and chemists has led to the rapid development of modern technology, the conquest of many diseases resulting in increased life expectancy, and the improvement of human lives on earth. But, no matter how successful a model has been, scientists realise there may be aspects of the world that the model fails to explain, or worse, predicts incorrectly. Nevertheless, creating and using models is one of the most powerful tools ever developed. But, there is a need to revise and improve models as new information is discovered.

Further resource material and references

There are many books, journals and articles on models. Most tend to be specialised and specific to certain models or application of models in specific areas of specialisations. To understand some basics on what models are, and how you can build a model, three books are listed below particularly useful.

Appendix I. The Craft of Research. Paul L. Woomer. PowerPoint on CD.

Appendix II. ICRAF. 2003. *Genstat Discovery Edition and Other Resources*. World Agroforestry Centre (ICRAF), Nairobi, Kenya. On CD.

Anon. 2003. Why the analysis of wide range of physical phenomena leads to consistent and successful results when applying the BSM concept and models? http://www.helical-structures.org/Applications/why_successful.htm

Ford, A. 1999. *Modelling the Environment. An introduction to system dynamics modelling of Environmental Systems*. Island Press, California, USA. 401 pp.

Jorgensen, S.E. 1994. *Fundamentals of ecological modelling*. Elsevier, London, UK. 628 pp.

Matthews, B.R. and Stephens, W. 2002. *Crop–Soil Simulation Models: Applications in Developing Countries*. CAB International, Wallingford, UK.

Muthuri, C.W. 2003. *Impact of Agroforestry on crop performance and water resources in semi-arid central Kenya*. PhD Thesis. Jomo Kenyatta University of Agriculture and Technology (JKUAT). 289 pp.

van Noordwijk, M. and Lusiana, B. 2000. WaNuLCAS version 2.0: Background on a model of water, nutrient and light capture in agroforestry systems. International Centre for Research in Agroforestry (ICRAF), Bogor, Indonesia, 186 pp.

Shepherd, K.D. and Soule, M.J. 1998. Soil fertility management in Western Kenya: dynamic simulation of productivity, profitability and sustainability at different resource endowment levels. *Agriculture, Ecosystem and Environment* 71: 131–145.

Soto, R. 2003. Introducing system thinking in high school. The connector (Connecting system thinkers around the world) 1(5). http://www.hps-inc.com/hps/zine/sepoct03/jake.html

Stephens, W. and Hess, T.M. 1999. Modelling the benefits of soil water conservation using the PARCH model—a case study from a semi-arid region of Kenya. *Journal of Arid Environments* 41: 335–344.

Internet resources

Ecological models http://www.wiz.uni-kassel.de/ecobas.html

CERES crop models http://www-bioclim.grignon.inra.fr/ecobilan/cerca/ceres.html

FALLOW model at http://www.icraf.cgiar.org/sea/AgroModels/FALLOW/

FLORES model at http://www.cifor.cgiar.org/flores/ An example of model building in participatory research

PARCHED-THIRST at http://www.cluwrr.ncl.ac.uk/projects/tanzania/modelling.html

WaNuLCAS model at http://www.icraf.cgiar.org/sea/AgroModels/WaNulCAS/

STELLA software; High performance Systems Inc at http://www.hps-inc.com/

Powersim software; The business simulating company www.powersim.com/

Vensim PLE. Vantana Systems Inc. www.vensim.com/

Management Unit of the North sea Mathematical Models (MUMM) (2003) http://www.mumm.ac.be/EN/Models/Development/Ecosystem/how.php

Model Maker: available from www.modelkenetix.modelmaker/index.htm

4.9 Where is the participation?

Richard Coe

- **Effective projects will involve participation of stakeholders in all stages of planning, implementation and evaluation**
- **Participation in a research study, both who and how, should be determined by the objectives of the study**
- **Participation of farmers or others in a research study is no reason to forget the elements of research design that will allow you to reach valid conclusions**

'One hour of life, crowded to the full with glorious action, and filled with noble risks, is worth whole years of those mean observances of paltry decorum, in which men steal through existence, like sluggish waters through a marsh, without either honour or observation.'

Sir Walter Scott

Part 2 of the book made a strong case for 'participation' – involving those who will use the research in the process. So why has Part 4 of this book, which focuses on research methods, not included a specific section on these methods?

The answer is simple: appropriate method and levels of participation are needed in ***all*** stages of a project. The subject of this book is sound research, and the principles of methods needed to do good research are much the same whoever is participating. 'Participatory methods' for both the social and natural sciences are widely discussed and described in the literature and they have been referenced in the appropriate chapters. Thus, just as we do not describe all the methods available for experimental design, or analysing data or building models, so this book does not provide a comprehensive review of the research methods which are available when collecting and analysing data in a participatory research process. There are specialised texts to help you with this.

This answer will not satisfy everyone, so it is elaborated below. In the discussion I distinguish between participation in a ***project*** and participation in the ***research studies*** that make up the project. The role of participation at the two levels can be different, and the extent to which, as a student, you can have some influence over them is also different.

Projects

Think of the cowpea project described in **Chapter 2.3**. Pests were found to be an important constraint in cowpea production, so the project aimed to find ways of overcoming them. However it is conceived, a project to tackle this problem will have elements of refining understanding of the problem, devising and testing possible solutions, promoting widespread use of the solutions and assessment of their impact, with iteration and cycling around these elements. It therefore seems natural that farmers should be involved in all the stages –

- Who better understands the occurrence and impacts of cowpea pests than the farmers growing the crop?

- Who can test and evaluate solutions, but farmers who will have to use them?
- Who can evaluate their impacts, but the farmers who feel them?

These are the pragmatic reasons for participation in problem-solving projects.

Other reasons are also often given, reasons which might be described as ideological, some discussed in **Part 2**. Development workers all over the world believe that 'development' involves giving people more control over their lives and resources – indeed for some this is the definition of development. It is a right of farmers not to be *told* that their problem is cowpea pests, and told what to do about it, but facilitated in understanding for themselves their problems and solutions which suit them. Projects that take this approach are more likely to lead to sustainable solutions, that continue to have impact past the end of the project and the departure of researchers.

So, if the cowpea project leaders are aware and agree with these arguments they will set up a project which involves the farmers in all stages, with farmers working collegially with researchers, each bringing their own expertise and knowledge to the table. But try to do this and many further questions arise. For example:

1. Just who should participate? Farmers, or maybe others with an interest, such as cowpea consumers and traders. Almost certainly the interests of all parties will not coincide.
2. Who decided that cowpea pests was the problem to work on? If you ask farmers their problems you are more likely to get responses such as 'Paying school fees' or 'Getting a job', rather than 'Pests on the cowpeas'.
3. The participatory approach requires intensive engagement in villages, meaning the project can only involve a few of them. But the problem covers large areas. How can you get the participation of all cowpea farmers?

There are no simple answers to these types of questions. Each project has to find ways that are best suited to the circumstances, but this has to be done knowing and understanding the many options and approaches available.

Point 3 above is one of the main reasons why projects need a sound research component. If your objectives only stretch to helping in those villages or households with which you are immediately engaged, then maybe you do not have to pay much attention to research methods. However there are few instances in which this is the case. Every project wants to generate information that will be useful beyond its own bounds. The only way to do this with known reliability is to use well planned research methods.

As a student joining a project to undertake thesis research you may not have been involved in planning the overall project and the approach to participation adopted. However you need to understand what the approach is and the reasoning behind it. And you must be prepared to challenge it if necessary.

Research studies

Most projects will involve a number of specific research studies which contribute to its overall strategy. How should participation be built into these, and what participatory methods are appropriate? The answer is the same answer that we give to most other research methods questions: it depends on objectives. If the objectives are specified clearly enough then they should guide you. And, referring back to the previous section, there should have been appropriate participation in setting the objectives for each component study.

A single project may well have component studies with different levels of participation, which are mutually agreed by all concerned. An example of a project introducing high-value trees in a mountainous region identified (among others) three component studies:

1. Farmers wanted to plant a cover crop between newly established trees but did not know whether beans or sweet potato would be most suitable. They agreed they could investigate this themselves without involving a researcher.
2. They had consistent difficulty germinating one species and asked the researcher to investigate. It turned out to be a problem of fungal attack which could be controlled by keeping moisture levels low. Farmers got involved again with testing ways of managing the moisture level in the nursery.
3. They wanted to know how well different species were adapted to different altitude zones. This they agreed had to be a joint effort, with farmers growing the trees but the researcher coordinating across altitudes to make sure comparable methods were used and results compiled.

Don't make the mistake of getting locked into a single mode of participation for all studies! Franzel and Coe (2003), for example, explain how different objectives in testing agricultural technologies can lead to differing balances of farmer and researcher involvement in design, implementation and assessment of the trial.

If your study will require an experiment, there are many ways in which stakeholders may participate. They could have been involved in the whole project, setting objectives, approaches and priorities, including identifying the need for that experiment. They may set objectives for that particular experiment, define details such as treatments and management, choose methods of assessment and evaluation, and plan for taking the work to the next step. Some objectives require little more from the researcher than facilitating farmers evaluating alternatives that they design, using criteria that they choose. In the cowpea project, this might be the case if the objective is simply for farmers in participating villages to determine if the new varieties are of value to them. However if the objectives require you, as a researcher, to come up with generalisable and defendable conclusions, then you will need to pay attention to the experimental design.

Chapter 4.3 on design of experiments described the reasons why experiments are important, the elements of the design that will lead to valid results, and ways of making the study efficient. These do not depend on who is participating or why. They depend on the logic of replicability and reliability of the results. Thus it will be important to ensure that the experimental design takes account of this, while not compromising on participation. For example, just what are the treatments being compared? The new varieties will be grown and compared with existing ones. Researchers and farmers together can agree on which control varieties should be included. Not every farmer needs to test all the new varieties or all the controls – that sort of 'balance' is *not* a requirement of a sound experiment. But the design will have to include sufficient replication of the various comparisons that are important – that *is* a requirement.

If some farmers grow cowpea monocropped and some intercropped, then comparing the new varieties under these two conditions may be added to the objectives. This will increase the range of treatments needed. Your approach to participation may still have farmers select the actual treatment combinations they test. What do you do if no one wants to test some of the treatments? Find out why. Maybe farmers know something you don't and recognise that those treatments cannot work. Or maybe after discussion some farmers will decide that testing them is a good idea. Or maybe you have to add some of you own plots to look at those treatments. All these things can add to the practical complexity of carrying out the experiment and analysing the data. But they do not alter the underlying requirements for a good experimental design.

The same is true for surveys, the other way of collecting data. Here it is maybe even more important to recognise the elements of sound design when using participatory methods. Many of the tools used in participatory research are actually survey tools, so think about their use in terms of survey principles. For example, focus group discussions can be very valuable in understanding local conditions, problems and opportunities, particularly when linked to such techniques as resource mapping and wealth ranking. They give real insights into the villages with which you are working. And they give the participants themselves insights into their own situation. If that is your objective, then you can be flexible in how and where the tools are used But do you want to know how broadly applicable those results are, and the extent to which they are representative of a larger population? If so, then use sampling techniques to select the villages in which you work, and use the ideas of survey management to make sure that information is really comparable across different villages.

The final point to make about the use of participation in a research study is 'Beware of packages!'. There are many guides to participatory research that present a packaged set of tools and processes, based on what the author has found to work. But that author's objectives and circumstances will not be the same as yours, so do not expect to be able to follow the same steps and use exactly the same tools. You have to learn to pick and choose the methods and tools that meet your objectives. As a simple example, we have found that participatory tools for matrix ranking to assess alternatives, based on the traditional mbao game, can be used effectively to evaluate an on-station, researcher-designed experiment.

Resource material and references

Appendix 3. Designing Research Around Client Needs. Paul L. Woomer. PowerPoint on CD

Appendix 11. ICRAF. 2003. *Genstat Discovery Edition and Other Resources*. World Agroforestry Centre (ICRAF), Nairobi, Kenya. On CD.

Chambers, Robert 1992. *Rural Appraisal: Rapid, Relaxed and Participatory*. Institute of Development Studies (IDS) Discussion Paper 311, UK.

Franzel, S. and Coe, R. 2003. The balance between researcher and farmer involvement in technology testing. In: *Designing Participatory On-farm Experiments: Resources for Trainers*. Coe, R., Franzel, S., Beniest, J. and Barahona, C. (Eds.), pp 52–63. World Agroforestry Centre (ICRAF), Nairobi, Kenya.

There are many African universities involved with participatory research and you are encouraged to consult their publications. Two university institutes in the region with particular strengths in, and experience with participatory research include:

Centre for Applied Social Sciences, University of Zimbabwe, Zimbabwe
http://www.cassuz.org.zw/

University of Western Cape, South Africa
http://www.uwc.ac.za/

The online 'Participation reading room' at the Institute of Development Studies (IDS) has a large amount of material on participation in many different contexts
http://www.ids.ac.uk/ids/particip/information/readrm.html

The World Bank also maintains extensive online resources on participation
http://www.worldbank.org/participation/participation/participation.htm

PLA notes is one of the leading periodicals on particpatory work, available from the International Institute for Environment and Development (IIED), UK
http://www.iied.org/sarl/pla_notes/whatispla.html

The PRGA program of the CGIAR is continually developing new material
http://www.prgaprogram.org/

IIRR is an example of an NGO producing guides on doing participatory work
http://www.iirr.org/

If all else fails, resort to ELDIS, the most comprehensive online development library with many documents on participation.
http://www.eldis.org/

5 Where to from here?

Kay Muir-Leresche

- **The more people able and determined to contribute to sustainable development the better**
- **Your individual ideas and actions do count. You can make a difference**
- **Creativity and adaptability are essential criteria for successful economies in a rapidly changing and global environment**
- **You can contribute to both the poor and to your own advancement with imagination**
- **Experience and a track record are important for getting jobs — you may need to start off in a menial position or doing voluntary work to establish your credentials**
- **It is possible to be an entrepreneur even without capital**
- **Blend modern and traditional, indigenous and conventional**
- **Be proud of your heritage, understand the limitations and grasp the opportunities**

'...the initiation of all wise and noble things comes...,generally at first from some one individual.'

John Stuart Mill
Representative Government

'...there will be no injustice in compelling the philosophers who grow up in our state to have a care for the others.'

Plato
The Republic VII

Where to from here?

You have finished your thesis, you have had it examined and you have undertaken your corrections — it is bound and you have a copy which you proudly present to your family.

What now?

Remember when you started out how unsure you were – 'How will I ever manage that?', you thought? Well you have – and you have grown in the process. You have developed skills and most important of all, you have gained confidence. After the strain of producing your thesis, you may be feeling a bit flat and uninspired. You probably need to renew your enthusiasm and burning desire to contribute.

Confidence, the ability to use your own initiative and the inspiration and determination to make a difference, are the most valuable of all resources any country can have.

The more people there are determined and able to contribute to sustainable development, the better. You only have to look around Africa, or the world, to see that it is not necessarily rich mineral resources, nor rich agricultural land, nor rich coastal waters that make countries better able to provide for their poor. It is their human capital. It is their commitment to success and their ability to respond quickly – to be creative and adaptable – so that they can take advantage of changing technology, institutions and social relations. It is essential for us to start to take charge of our destiny; to succeed in developing our countries. We Africans want our children to grow up in an environment where they are able to chart their own course and do not feel hopeless. We need to be able to move away from corrupt practices that obtain short-term advantages. We need to earn our incomes by providing goods and services which in turn will develop Africa. We need to be able to take control of our development at the personal, village, national, and regional levels.

Finding work that provides you with money and prestige are common goals. They are important. We need that money to live, to repay our families who invested in our education and to provide for our futures. Social standing can be important to many people – but remember that fashions change and what is prestigious today may not be in 10 years' time. Happiness, however, is not limited to wealth and fame. There is considerable personal satisfaction from contributing to society. If you can make a lasting difference to the lives of the poor, to the development of your country, or even to one student or one farmer or one village, you will be able to look back when you are 80 and say – Yes, I did make a difference!!

These goals do not have to be mutually exclusive.

Example 1

Nyasha is hired by Norsk Agricultural Chemical Co. to promote the sale of fertilizers and pesticides. She earns her salary by selling to conventional markets and using the established recommendations. Perhaps she remembers that when she was doing her graduate research, many small-scale farmers could not afford to apply fertilizer at the recommended rate. She has heard of someone who has adapted the established recommendations to more closely suit small-scale, poor farmers. So in her spare time, she contacts them and then draws up a marketing strategy that would provide farmers with access to this new information. She has to persuade the company that, although these recommendations are for lower fertilizer use, they will make it possible to sell to many more farmers.

In this example the sales agent used her initiative and commitment to change things for the better. She also advanced her career. You can all do this. In every job you do, it is possible to make the world a little better for the future. You need to believe in your own power. You need to learn to be a self-starter and to be prepared to put in that extra effort.

You need to take an ethical stand. Do not allow your valuable skills to be used to further corruption and the cheating of your fellow citizens. Do not contribute to the degradation of the environment and the impoverishment of future generations. We owe it to our children to leave a better world than we found, and you can make a difference.

Employment

The formal sector

Remember that employment does not necessarily mean working for someone in return for a salary. Employment means using your skills and labour to produce output that will have financial and other rewards. Professional jobs for new graduates in Africa have become increasingly difficult to find, despite the considerable shortage of skills. This is because for some 50 years, governments employed new graduates. They would obtain practical experience and learn to operate in the working world that gave them credibility and led to formal employment in the private sector. Decentralisation, declining government budgets and reduced investment in research, extension, and education have all contributed to shrinking these opportunities in most African countries. At the same time, the private sector is reluctant to hire untested graduates. In most countries very strict legislation makes it difficult for employers to release staff once they have been hired and as a result they are very risk-averse in their employment policies. You are required to be much more innovative than your parents in seeking employment.

You need to get together some evidence of your ability. Take a copy of your thesis and of a few other projects or papers you have produced. Ensure that you include the extra-curricula activities with which you have been involved and any positions of leadership or trust which you may have held. Speak to the people you are going to use as referees and be sure they are happy to do this. Provide them with a copy of your CV so that it is easier for them to write the reference.

Find out about the company before you go for an interview. See where you think you would fit in. At the interview you should not be arrogant but you must make an opportunity to be able to tell them how you think you could contribute to their organisation. For example, if the job involves selling tractors, you might mention your contacts from your home area who may be interested clients – or mention your experience working in a garage during one of your vacations. If it is project management and budgeting, you could mention your role in the university agricultural student society. If you don't have anything specific you could offer, at least be sure you understand what the organisation does and show that you have thought through how you could play a role within it.

Something else

If you are unsuccessful in obtaining formal sector employment, you should seriously consider voluntary service as a stepping stone. Most prospective employers would be prepared to provide you with basic transport and food costs. If you cannot find a company to hire you even on these terms, then prepare a research proposal and contact relevant NGOs, government research departments or even churches. Do not be ambitious for a high financial reward even when you are contacting an international agency. Remember this first 'job' is more to establish your credibility and gain experience than to provide an income. Impress the prospective benefactor with the fact that you are prepared to sacrifice in order to get ahead in the future and to contribute to your society. You need to realise that the world does not owe you a living and that you have to be creative in getting that first job. Once you have experience, if you prove yourself, it will be much easier to move up the ladder.

For many African students this is difficult. Their families have invested resources in the graduate's education and now they expect that person to start to contribute to the family. Prepare your family. Show them your strategy ahead of time and I am sure you will find them much more understanding.

Even if no-one is prepared to take you on, even as a volunteer, you may then have to go and take a much more menial job. Look at it positively as a stepping stone and be constantly on the look-out for how you can contribute to the success of the organisation for whom you are working. It is surprising how many highly successful people have started in very menial positions.

Increasingly in Africa the best way to get ahead is to become an entrepreneur yourself. How you go about this will depend on the contacts and resources you have. If you are able to raise capital then you can be more adventurous. If you cannot raise any capital then start very small. Identify a need and provide for it even in a very small way.

Example 2

Tapiwa realised that there would be no bread available in the following year. He knew that urban workers would need to have convenience food that they could afford. He went to his aunt in the rural areas and asked her to provide him with some sweet potatoes and promised to repay them when he harvested his own crop. He went and read up all the literature on sweet potatoes and learned what he could about their preferred soil types, mineral requirements and the most ideal moisture conditions. He could not afford fertilizer but he approached the people in his street and asked them if they would put all their vegetables and other wet refuse into bags for him. He would collect it and this would reduce the unpleasantness of such refuse left out on the road for days. He also collected newspapers and on a vacant lot he made a compost pit. As a result he had a bumper harvest of high-quality sweet potatoes that fetched a high price because of the need he had identified. In due course he became a successful market gardener, bought his own plot and was able to employ workers.

Agricultural research and our commitment to sustainable development

Most of students with post-graduate degrees will eventually go into work that involves research. The ultimate goal of research is to search for the truth. Thus, we make a moral commitment when we undertake research and we need to honour that. We must avoid any fabrication or falsification of information and data. We must be sure that we set the highest standards for ourselves and that we maintain our integrity. This will require our research to be as objective as possible. It will mean that we need to closely supervise the collection and entry of our data. Most important of all, we must remember that the work we produce will be used to affect the lives of people who live at the margin. A small error can tip them into very serious poverty or even starvation. At the same time we need to respect their abilities, their privacy and their needs. We need to listen to them and to try to establish research, policies, and implementation strategies that empower the disadvantaged. The results may be slower but they will be more sustainable.

We need to find ways in which we can adapt some of the modern technologies so that they can be used despite the constraints facing the users. We need to adapt traditional norms and values so that they can accommodate new technology. There is much scope for blending modern and traditional, conventional and indigenous and of finding ways to commercialise, improve and extend the use of traditional commodities. We need to find better ways to use our resources so that we do not endanger our environment. Be proud of your heritage, understand the limitations but grasp the new opportunities. Graduates must take a pride in creativity and in their ability to make something different by using both the old and the new. Technology is changing constantly and global competitiveness requires the ability to innovate rapidly (Porter and van der Linde, 1996).

Resource material and references

Foster, Michael B. 1942. *Masters of Political Thought: Plato to Machiavelli.* Harrap & Co., London, UK.

Lancaster, Lane W. 1959. *Masters of Political Thought: Hegel to Dewey.* Harrap & Co., London, UK.

Porter, M. and van der Linde, C. 1996. Ending the stalemate. In: *Business and the Environment*, Richard Welford (Ed). Earthscan, London, UK.

Contributors

Gerald W. Chege is a Kenyan with a PhD in Parallel Computing from York University, UK. He is presently Assistant Professor and Coordinator of the Information Systems and Technology Department, United States International University, Nairobi. His main area of interests are: computer networks, database technology; systems development, and Internet technology.

Richard (Ric) Coe is an Applied Statistician from the United Kingdom. He gained an MSc in Biometry from the University of Reading, where he continued as a lecturer for 10 years. During that time he was involved in a number of training and research projects in Africa and Asia. In 1990 he moved to the World Agroforestry Centre (ICRAF) in Nairobi, Kenya. There he is Head of the Research Support Unit that provides technical support and training in research planning and design, data management and analysis to all ICRAF projects and partners. His interests are in making research for development as effective as possible through the use of sound methodology, and increasing capacity in Africa to do this. He has taught courses at several universities in the region and has worked with hundreds of graduate students on their research projects.

Tony Greenfield, a graduate in Statistics from London University with a PhD in Experimental Design from Sheffield Hallam University, was formerly Head of Process Computing and Statistics at the British Iron and Steel Research Association, Sheffield, and Professor of Medical Computing and Statistics at Queen's University, Belfast. He is a Visiting Professor to the Industrial Statistics Research Unit (ISRU), at the University of Newcastle-upon-Tyne and is past President of European Network for Business and Industrial Statistics (ENBIS). While at Queen's University he established a course in research methods for the medical faculty. His publications include *Research Methods: Guidance for Post-Graduates* (Editor and co-author), first published by Edward Arnold in June 1996, second edition in June 2002. This book is used in some English universities in courses for postgraduates who intend to proceed to research degrees.

Thomas Gumbricht is a Swedish Hydrologist working with the World Agroforestry Centre. He holds a PhD in Land Improvement and Drainage from the Royal Institute of Technology (KTH), Sweden, and prior to his arrival in Kenya was Head of Geoinformatics at the Department of Earth Sciences, Uppsala University, Sweden. His main interests are systems ecology and hydrology; using geoinformatics as a platform for understanding and modelling processes on a landscape scale.

Sue Hainsworth has been editing all her working life. After graduating in Agricultural Sciences from Nottingham University she edited the Tropical Pest Management Journal and wrote the first three titles in the Tropical Pest Management Manual series on bananas, groundnuts, and rice. After time in Rome with the Food and Agriculture Organization of the United Nations (FAO) and the International Plant Genetics Resource Institute (IPGRI, then IBPGR) in 1983 she joined the International Crops Research Institute for the Semi-Arid Tropics (ICRISAT) and rose to become Manager, Publications before leaving to start her own Editorial and Publishing Services in 1998.

Erica Keogh is a Zimbabwean, holding an MSc in Statistics (University of Zimbabwe, 1987). She has been employed as a lecturer at the University of Zimbabwe since 1980 but is currently on long leave while she engages in a long-term consultancy with the UK's Department for International Development (DFID) monitoring their Humanitarian Relief Programme in Zimbabwe. Since the early 1990's she has become increasingly involved in applications of statistics and has had extensive experience in the design and implementation of surveys in both rural and urban areas, focussing mainly on issues of poverty and related aspects of social change.

Eric McGaw, an American national, has lived and worked in the developing world for over 30 years. He graduated from Rockford College with a degree in Fine Arts, and completed post-graduate work in Education at Boston State College, USA. After serving in the Peace Corps in El Salvador, he worked as a university professor, a deep sea diver, a freelance writer and editor, and a communication specialist in Colombia, Brunei, Singapore, the Philippines and India. He has travelled widely throughout Latin America, Asia and Africa. Currently, he is employed as Head of Communications at ICRISAT located near Hyderabad, India.

Kay Muir-Leresche is a Zimbabwean with 23 years' developing-country experience in agricultural and natural resources higher education, research, training, and policy analysis. Her main focus since 1981 was as an Economics Lecturer in the Faculty of Agriculture, University of Zimbabwe. When she left in 2002 she held the Professorial Chair in the Department of Agricultural Economics and Extension with major responsibility for the supervision of doctoral candidates. She taught both undergraduate and graduate programmes and supervised student dissertations and she served on the Faculty Higher Degrees Committee for 15 years.

Peter K. Muraya is a Kenyan with a BSc in Electronic Computer Systems Engineering from Loughborough University, UK. Presently he is Data Management Specialist with the responsibility of leading the World Agroforestry Centre's initiative to bring management of research data to agreed standards in all regions and projects. He was earlier involved with the development of simulation models for agroforestry systems. His main area of interest is in conceptualising, design and implementation of data management methods and software tools.

Liliosa Maveneka has a BSc in Mathematics and Botany and an MSc in Agricultural Economics. She also is an Associate Member of the Institute of Chartered Secretaries and Administrators. She has worked as a Registrar in the Faculty of Agriculture and as a Senior Administrator for the University of Zimbabwe for 15 years. She is a consultant working on HIV/AIDS impacts, in issues related to water allocation and pricing and in providing assistance in accessing Internet data to post-graduate students and researchers.

Catherine Wangari Muthuri is a Lecturer in the Department of Botany, Jomo Kenyatta University of Agriculture and Technology (JKUAT), Kenya, where she earlier gained an MSc in Botany (Plant Physiology). Catherine submitted her PhD thesis on the 'Impact of agroforestry on crop performance and water resources in semi-arid Central Kenya' for examination in September 2003. She carried out her research work for 3½ years at the World Agroforestry Centre. For the past 10 years, Catherine has been involved in teaching, administration and research at JKUAT. Her research interests include environmental plant physiology in agroforestry and non-agroforestry systems with particular emphasis on drought stress and the application of agroforestry models in research.

Joseph Opio-Odongo is a Ugandan holding a PhD in Rural Sociology from Cornell University, USA. He is one of the United Nations Development Programme's (UNDP's) Environmental Policy Specialists, out-posted to Nairobi to provide technical backstopping to UNDP Country Offices in sub-Saharan Africa on policy and programme development. He previously served as a Sustainable Development Advisor at the UNDP Country Office in Uganda after some years of teaching at Makerere University in Uganda and Ahmadu Bello University in Nigeria. His research and teaching experience has been mainly in the fields of agricultural and rural development. His research and development interests include policy analysis, empowerment of civil society, sustainable development, organisational development, science and technology policy, lobbying and advocacy, and the codification and application of indigenous knowledge and technology.

Bharati K. Patel has been working in the Food Security Division of The Rockefeller Foundation in Africa for the past 10 years. As an Associate Director she ran the Forum on Agricultural Resource Husbandry a competitive grants programme designed to encourage and support research on agricultural resources. The programme supported the staff in ten Faculties of Agriculture in Kenya, Malawi, Mozambique, Uganda, and Zimbabwe in their training of graduate students. A Zambian with a primary degree in Botany and a PhD in Nematology from the Waite Institute in Australia, Bharati also worked for the Zambian Agricultural Research System for 20 years where she rose to become the first woman Director of Agricultural Research in Africa. She also worked in ICRISAT prior to her assignment with The Rockefeller Foundation.

Aleya Pillai is an Indian, with a Commercial Arts Diploma from Jawaharlal Nehru Technological University, Hyderabad, India. She worked as Art Director from 1984 to 2001 with various Indian advertising agencies, the last one being Mindset in collaboration with Saatchi & Saatchi. During that time she designed several award-winning annual reports, calendars, press ads and logos. She is no newcomer to cartoons, having used them to get across scientific concepts in the past. Aleya also enjoys cartooning and oil painting in her spare time.

Jane Poole is a British national and holds an MSc in Biometrics (Applied Statistics) from Reading University, UK. She has recently returned to the UK after 6 years of working in Africa, where she provided biometrics support to scientists and research students at the World Agroforestry Centre and CAB International (Africa Regional Centre), both based in Nairobi, Kenya. Jane currently works at the UK Forest Research Agency with scientists covering a wide range of disciplines: from forestry, entomology, and pathology to ecology and environmental research. Jane has wide experience in experimental design and analysis and in small and large-scale biological and socio-economic surveys. She enjoys working with students and scientists from many disciplines, learning about their research and working with them as a member of the research team throughout their projects.

Sue J. Richardson-Kageler is the Biometrician in the Faculty of Agriculture, University of Zimbabwe, where she previously taught in the Statistics Department. One of the major components of her job is advising undergradutes and postgraduates on the design of their experiments. A Zimbabwean by nationality, Sue has a DPhil in Ecology which examined the changes in woody plant diversity brought about by large herbivores and an MSc in Biological Computation from the University of York.

Jayne Stack is a Senior Lecturer in the Department of Agricultural Economics and Extension, University of Zimbabwe and has more than 20 years experience in development training, development programmes and research in Africa and Asia. She has taught research methods at undergraduate and postgraduate level and contributed to the development of distance-learning courses in research methods and data analysis for Imperial College, London. Jayne has a wide interest in development issues ranging from crop marketing to agricultural policy reform, household food security and livelihood analysis. Her research work aims to make a difference in the lives of the poor and to contribute information that will enhance livelihood security of vulnerable households.

Paul L. Woomer is a researcher working with the Sustainable Agriculture Centre for Research Extension and Development in Africa (SACRED-Africa), a Kenya-based NGO. One of his major interests is the adaptive research process where different potentially useful technologies are compared, combined and refined to suit the needs of individual farmers. He has written or edited four books and published over 90 papers or chapters in international journals and multi-authored books. Paul was raised in the Hawaiian Islands where he developed a keen interest in tropical crops and ecology. He attended the University of Hawaii, where he obtained a BSc in Agronomy and a MSc and PhD in Soil Science. He previously worked with NifTAL-MIRCEN, TSBF-UNESCO, The Alternatives to Slash and Burn Consortium and the University of Nairobi. Paul has lived in Kenya since 1990 and has visited or worked in 18 different African countries.

Acronyms and abbreviations

ACSS	African Crop Science Society
ADDS	African Data Dissemination Service
ACT	almanac characterisation tool
AFRENA	Agroforestry Research Network for Africa
AHI	African Highlands Initiative
AI	appreciative inquiry
AKIS	Agricultural Knowledge and Information System (World Bank)
ANOVA	analysis of variance
AVHRR	Advanced Very High Resolution Radiometer
CARPE	Central African Regional Program for the Environment
CBO	community-based organisation
CD	compact disk
CGIAR	Consultative Group on International Agricultural Research
CIESIN	Center of International Earth Science Information Network
COSOFAP	Consortium for scaling up options for increased farm productivity in Western Kenya
CRSP	Collaborative Research Support Program (USAID)
CRU	Climate Research Unit (University of East Anglia, UK)
CTA	Technical Centre for Agricultural and Rural Cooperation (the Netherlands)
DCW	Digital Chart of the World
DEPHA	Data Exchange Platform for the Horn of Africa
DEM	digital elevation model
DFID	Department for International Development (UK)
DMA	Defense Mapping Agency
DRASTIC	depth of groundwater, recharge, aquifer media, topography, impact of root zone, conductivity
DRC	domestic resource cost
DSMW	Digital Soil Map of the World (FAO)
DSS	decision-support system
ENBIS	European Network for Business and Industrial Statistics
ESA	European Space Agency
ESRI	Environmental Systems Research Institute, Inc.
ETM	Enhanced Thematic Mapper
FAO	Food and Agriculture Organization of the United Nations
FEWS	Famine Early Warning System (USAID)
GDP	gross domestic product
GIS	geographic information system
GPS	global positioning system
GUI	graphical user interface
IARC	international agricultural research centre
IBPGR	International Board on Plant Genetic Resources (now IPGRI)
ICIPE	International Centre for Insect Physiology and Ecology
ICRAF	World Agroforestry Centre

ICRISAT	International Crops Research Institute for the Semi-Arid Tropics
IDS	Institute of Development Studies (UK)
IDW	inverse distance weight
IPCC	Intergovernmental Panel on Climate Change
IPM	integrated pest management
INASP	International Network for the Availability of Scientific Publications (UK)
IPGRI	International Plant Genetic Resources Institute
ISSER	Institute of Statistical, Social and Economic Research
KARI	Kenya Agricultural Research Institute
KEFRI	Kenya Forestry Research Institute
LFM	logical framework matrix
MAP	methods of active participation
MCE	multi-criteria evaluation
MODIS	Moderate Resolution Imaging Spectroradiometer
NARES	National Research and Extension Services
NARO	National Agricultural Research Organisation (Uganda)
NARS	national agricultural research systems
NDVI	Normalised Difference Vegetation Index
NGO	non-governmental organisation
OECD	Organization for Economic Cooperation and Development
PAR	participatory action research
PI	principal investigator
PRA	participatory rural appraisal
R&D	research and development
RELMA	Swedish-supported Regional Land Management Unit
ROM	read-only memory
RS	remote sensing
SACRED	Sustainable Agriculture Centre for Research Extension and Development in Africa
SAS	statistical analysis system
SPSS	statistical program for social sciences
SRTM	Shuttle Radar Topography Mission
SSA	sub-Saharan Africa
TEEAL	The Essential Electronic Agriculture Library
TM	Thematic Mapper
TSBF	Tropical Soil Biology and Fertility
UNDP	United Nations Development Programme
UNESCO	United Nations Educational, Scientific and Cultural Organization
UNSO	United Nations Statistical Organization
USAID	United States Agency for International Development
USGS	United States Geological Survey
VCD	visual compact disk
WARDA	West Africa Rice Development Association
WBS	work breakdown structure
WFP	World Food Programme
WRI	World Resources Institute

Appendices on the CD

Appendix 1. The Craft of Research. Paul L. Woomer. PowerPoint

Appendix 2. Innovation, Problem Solving and Operational Research Strategies. Paul L. Woomer. PowerPoint

Appendix 3. Designing Research Around Client Needs. Paul L. Woomer. PowerPoint

Appendix 4. Preparing and Refining a Research Proposal. Paul L. Woomer. PowerPoint

Appendix 5. Stapleton, P., Youdeowei, A., Mukanyange, J. and van Houten, H. 1995. *Scientific Writing for Agricultural Research Scientists*. WARDA/CTA, Ede, The Netherlands.

Appendix 6. Publication as an Output of Science. Adipala Ekwamu. PowerPoint

Appendix 7. The Art and Ups and Downs of Scientific Publication. Adipala Ekwamu. PowerPoint

Appendix 8. Presentations and Style – Tips on Photography and Writing. Eric McGaw.

Appendix 9. Muraya, P., Garlick, G. and Coe, R. 2003. *Research Data Management*. World Agroforestry Centre (ICRAF), Nairobi, Kenya.

Appendix 10. Coe, R., Stern, R., Allan, E., Beniest, J. and Awimbo, J. 2002. *Data Analysis of Agroforestry Experiments*. World Agroforestry Centre, Nairobi (ICRAF) Kenya.

Appendix 11. ICRAF. 2003. *Genstat Discovery Edition and Other Resources*. World Agroforestry Centre (ICRAF), Nairobi, Kenya.

The African Crop Science Society
P.O. Box 7062, Kampala, Uganda

ISBN 9970-866-00-1