1 Planning a research programme

William J. Sutherland

Centre for Ecology, Evolution and Conservation, School of Biological Sciences,
University of East Anglia, Norwich NR4 7TJ, UK

Introduction: reverse planning

This chapter was written in response to my frustration when talking to people who had worked hard collecting data but who had missed opportunities and largely wasted their time as a consequence of poor planning. Planning relates to any research programme, not just those involving carrying out a census.

In thinking about my own research, collaborating with colleagues and advising students I have devised a means of planning research projects, called reverse planning (see Table 1.1). In designing a research programme it is best to think backwards, starting with the question and then considering how to answer it. This method seems clumsy, but many projects either do not have a sensible question or have a sensible question but the planned research will not answer it. Many projects are clearly unrealistic: it is better to identify impossible projects before starting them.

I recommend this process of reverse planning, and especially the key stage of producing graphs, to many students and others each year. Many find this difficult, but this is only because designing projects is difficult. It is better to struggle with the planning than to struggle with poor results.

An advantage of this explicit planning process is that consulting others becomes easier. Most people go into the field with only a hazy impression of what they will do, how they will analyse it and why they are doing it. A pile of potential graphs provides an excellent basis for discussing projects. It is easy for others to make helpful suggestions and point out flaws. Similarly, planning the time allocation and designing data sheets before going into the field also allows others to make useful comments.

I have described this as a straightforward logical process and recommend working through these steps in order. In practice, and especially once experienced, many stages can all be tackled simultaneously. Experienced researchers usually do not carry out this exercise explicitly, but in my experience it is clear that they have subconsciously run through these stages. Even if one is experienced, it is still usually worth running through these stages, starting with the possible figures, then considering the questions, protocol and time to see whether there are any ways of improving the work.

Ecological Census Techniques: A Handbook, ed. William J. Sutherland.
Published by Cambridge University Press. © Cambridge University Press 1996, 2006.

Table 1.1. *How to use reverse planning*

1.	What is the specific question?
2.	What results are necessary to answer the questions?
3.	What data are needed to complete these results?
4.	What protocol is required to obtain these data?
5.	Can the data be collected in the time available?
6.	Modify the planning in response to time available
7.	Create data sheets
8.	Start and encounter reality

What is the specific question?

Much of the skill in designing a research project is in selecting a suitable question or questions. It should be worthwhile (e.g. of conservation importance or of academic interest), have a clear objective and be practical in the time available. Questions such as 'What is the population ecology of X?' and 'How are development and conservation linked?' are too broad. More suitable questions are those for which there is a specific answer, such as 'What is the population size of X?', 'What are the habitat requirements of X?' and 'Is there more of this species in one habitat than another?'

What results are necessary to answer the questions?

The question or questions will be answered by a series of results. The next stage is to plan the analyses that can answer the stated question. The usual objective of any project is to collect a pile of graphs and tables. To create a PhD thesis you need only a pile of, say, 40–80 interesting graphs or tables, and then just need to write some text to describe the figures, explain the methods used to collect them, describe what the results mean, introduce the issues and then bind it. To write a research paper you need only 3–15 tables or figures (check with the papers in the journal you would like to publish in). Most papers have remarkably few figures, so if a research paper is the objective then it makes sense to concentrate on collecting data for the few critical figures. Thinking about the figures makes planning the data collection much easier. If you are not going to use the data for some such analysis then why are you collecting them?

Good experienced researchers tackle all the stages in Table 1.1 simultaneously and effortlessly explore different combinations of questions and analyses.

I thus suggest concentrating upon the graphs and tables required before starting any project. I strongly suggest sketching the likely presentation. This forces you to think deeply about what you are trying to achieve. The figures can be crudely scribbled with the axes labelled and possible output sketched in. These usually take me about 20 seconds each to draw once I have the idea (the hard bit).

There are five main ways of presenting data.

1. Histograms are used for plotting the frequency of occurrence. A common error is to plot data for individual sites that are better expressed as histograms or tables. For example, suppose that the study measured the fish density along various sections of rivers. Plotting histograms of fish density in each section would be of little use. If the interest is in the density in each section then a table may be more useful, since the exact result can be given. If, however, the aim of the study is to show some general pattern, such as that fish density is related to weed cover, then it is usually better to have a scattergram of fish density against cover.
2. Scattergrams (plotting sample points in relation to two variables, such as the density of invasive plants in quadrats in relation to distance to the forest edge). A common error is to plot as means and errors data that are more usefully shown as histograms. For example, if one is comparing the grass heights of sites that contain a butterfly with those that do not, it is possible to present each as mean grass height $\pm$ 1 standard error. It would, however, be more instructive to give the frequency distribution of mean grass height of all sites used as one histogram above a second histogram of the grass heights of unused sites.
3. Tables. These give the exact values for the data. They tend to be used when the reader is interested in the particular cases, whereas graphs are used if the main interest is in the relationship between parameters. For example, use a table if giving the butterfly densities of various sites but a scattergram if showing the relationship between soil pH and butterfly density. Tables are also used if a number of measurements are made, for example when describing the features of a range of sites when there are too many data to be presented as figures.
4. Maps. These are rarely used for presenting data but obviously the best way of showing distribution data. Simply plotting where a species was encountered is of little use. It is necessary to describe the areas that were surveyed. For example the survey may exclude all areas above 500 m because it is known that the species definitely does not occur above this height, so these areas can be shaded in the map. The areas searched should then be shown. It is acceptable to use multiple pie charts on maps, for example to show the proportions of juveniles in various locations or the proportions of injured individuals in various locations.
5. As facts within the text. This is used for giving simple measures, e.g. ‘of the twenty individuals found dead, 75% had gunshot wounds’. In practice, the results of less exciting graphs and tables are often stated in the text, e.g. ‘there was no relationship between the number of pollinators and sward height $r_s = 0.013$, $n = 23$). However, in planning I suggest that these are still drawn.

Pie charts should be used only for showing data to children or politicians (a table or histogram is usually clearer and the data can be read off precisely). Furthermore, most pie charts have a multicoloured key and it is often a nightmare to work our exactly which shade of pink corresponds to a given slice of the pie. Three-dimensional pie charts (as produced routinely by some computer packages) should not even be shown to children or politicians. The basis of pie charts is that the angle at the centre is in proportion to the frequency represented yet this is not the case in these pointless charts. The one exception regarding the use of pie charts is that they are useful when superimposed on maps to show how proportions vary geographically. For example, the proportion of individuals that are adult or the proportion of trees that are burnt could be counted at various points and each plotted on a map.

My experience of discussing science with students and others is that drawing figures is usually a difficult and painful process. However, it is difficult only because science is difficult. My experience is that, if the person is unable to draw the figures, it is because the person is insufficiently prepared to start the research. Once the science has been resolved, drawing the graphs is simple.

The statistics usually follow from the presentation. Thus scattergrams tend to lead to regression or correlation, histograms tend to lead to *t* tests, ANOVA or Mann–Whitney tests, and tables often lead to comparisons of frequency data, such as X^2. As Chris du Feu states, ‘statistics is drawing pictures then drawing conclusions’.

If the researcher is not a competent statistician then I recommend showing the figure and table sketches to someone who is before collecting the data. It is important to know how the data will be analysed before collecting them. In a few cases the data just cannot be analysed, but more frequently the lack of planning means that the opportunities are missed and the analyses are more restricted than necessary.

What data are needed to complete these analyses?

How much data is likely to be necessary to make these analyses worthwhile? How many individuals have to be marked in order to estimate survival rate? How large an area has to be visited in order to plot the distribution? Of course, this depends upon how strong the relationship or differences are and the variation that occurs. In the absence of any other information I usually suggest that a sample of 20 for the scattergram or each histogram point is a useful, if arbitrary, level to start thinking around. If the differences are likely to be slight, or the variation between samples considerable, then this will be insufficient. If the patterns are likely to be strong and consistent then a smaller sample might well suffice. Chapter 2 shows how large samples need to be to attain the required precision. Understanding how one will analyse that data is important in planning to avoid collecting data that cannot be analysed.

The total data required are thus the data needed to complete all of the planned figures drawn in the previous stage.

What protocol is required to obtain these data?

It is often sensible to devise a protocol, e.g. visit a site within the period 10 a.m. to 4 p.m. and count tortoises along four 100-m transects recording the soil temperature at depth 5 cm at the start of each, analyse the vegetation as percentage cover using a 1-m^2 quadrat at the middle of the transect, estimate the percentage of sand within 1 m^2 every 25 m along each transect. Assess the vegetation and sand cover around each tortoise located. Measure, weigh and sex each tortoise.

It is useful to write down the methods before starting. This is especially true if there is more than one observer, so that the methods can be agreed upon. Providing the full methods in the paper or report is essential in order that others can use the same methods for comparable studies.

Can the data be collected in the time available?

Determining the time necessary to collect the data seems like a boring task. It is. However, many projects are unsuccessful because they run out of time. One example is the expedition that sought to compare two sides of a mountain, but ran out of time and just compared one side of the mountain. If the project is not possible in the time available then it is much better to modify it or do something else than start and have to abandon sections that were never feasible. A common example of such a problem is to collect far more samples than can ever be identified. Apart from the possible ethical and conservation issues, such a project is less efficient than one that collects a sensible number of samples.

I suggest a four-stage process for the basic method for allocating time.

1. Decide upon the time available. This is usually the number of days available. Allow time for illness, bad weather, equipment failure, dealing with others, buying supplies etc. Be realistic.
2. List all the jobs that you would like to do and the number of times, e.g. ‘electrofish lake five times along 100-m transect’.
3. Estimate how long each would take in appropriate units for question 1 (e.g. days). Techniques include imagining doing it; estimating the time of each part; considering a similar task or speaking to others who have carried out the task. If necessary use a range of values.
4. Consider whether there is sufficient time.

Using a calendar

This is the same method as just described but the time is divided into periods. This could be hours, days, weeks or months, depending largely upon how long the project is. If using weeks (often a convenient scale), write the weeks along the top and the jobs down the side. Allocate time (e.g. days) to each job, taking into consideration when each should be completed. In reality this will take some juggling (often considerable juggling) to ensure that the work can be completed in time and that the load is distributed sensibly. Again, this juggling is the key part of the exercise.

Example. A project will be going to Sabah to examine moths. Moth sampling takes five days and is done at the start of each month. Mark–release–recapture takes eight days and should be done in as many months as possible. A vegetation survey of seven days can be carried out at any time. A *social survey of seven days should be carried out as late as possible, when you are more accepted by the community. The preliminary results will be analysed and presented to the local community. Packing will take 2 days. The project lasts from 1 January to mid March*

My answer assuming 5 working days per week and 4 weeks per month:

	Jan.				Feb.				Mar.	
	1	2	3	4	1	2	3	4	1	2
Moth sampling	5				5				5	
Vegetation surveys			2	5						
Social surveys							2	5		
Mark–release–recapture		5	3			5	3			
Present results										3
Pack										2
Totals	5	5	5	5	5	5	5	5	5	5

Calendars with divisions

It is often necessary to subdivide the calendar where there are other constraints, for example if there is a constraint that you can visit the field site for only eight days a month then divide into the calendar 'field site' and 'other'. It is then necessary to juggle the tasks to provide an efficient solution. Others studying frogs or moths may divide the time into days and nights. If various team members have different skills then you could divide for each member and allocate tasks in the most efficient way.

Example. You are going to carry out some mammal surveys and relate the results to vegetation, soil type and land use. You can start on 1 June and end on 1 August. Mammal surveys take place in the morning. You can carry out one transect per morning. Five transects are required per site. The mammal survey is a standard method and must be used. It is important to do this in as many sites as possible. Vegetation surveys take three half days. It is essential to carry out a full vegetation survey for each site with a mammal survey. It would also be interesting but not essential to carry out a bird mist netting survey per site, which takes two mornings per site. The director of the local research institute is away in June. On return he has promised to give you the data on soil type and past management. Meeting him and extracting the data will take two half days. You have been asked to collect lichens in any sites in which this is possible, which would take two half days (morning or afternoon) per site, but this is the lowest priority. The analysis of the project must be completed before leaving and will take about six half days but requires the institute data before it is worth starting. Plan your programme.

My answer. Since the objective is to collect as many mammal surveys as possible and these can be carried out only in the morning, I have cut the bird surveys. The lichen collection can be included without affecting the main objectives so I have included this.

	June								July							
	1		2		3		4		1		2		3		4	
	am	pm	am	pm	am	pm	am	pm	am	pm	am	pm	am	pm	am	pm
Mammals	5		5		5		5		5		5		5		5	
Vegetation		3		3		3		3		3		3		3		3
Institute										2						
Lichens		2		2		2		2								
Analysis												2		2		2
Totals	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5

Gantt charts

These are used where a number of individuals are involved as a way of allocating work that has to be carried out in a sequence.

	March																
Task	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15		Person responsible
Obtain aerial photos	×	×															Isabelle
Pitch camp	×	×															Carlos/Andrew
Decide on sites			×	×													All
Cut transects					×	×											Andrew
Cut mist net rides					×	×											Jenny
Count fish							×	×	×	×	×	×	×				Isabelle
Decide on trap location					×												Carlos
Create traps						×	×										Carlos
Check traps								×	×	×	×	×	×	×	×	×	Carlos
Mist net birds							×	×	×	×	×	×	×	×	×	×	Jenny
Botanical survey														×	×	×	Andrew

This can be combined with allocating time as in the previous methods if each time period is, say, a week and the number of days devoted to each task is then given.

Modifying the planning in response to time available

Allocating time is usually a painful process, since most people have more ambitious plans than are realistic. In practice this is likely to involve rethinking the previous stages. It is this rethinking that is central to the planning of the project. This rethinking usually involves slithering between the various stages until the questions, figures, protocol and time allocation are all compatible. In extreme cases it might be clear that it is not possible to answer a sufficiently worthwhile question in the time available. The options are to start an inevitably doomed project or design another project that might work.

Creating data sheets

Many projects involve data sheets and their creation is often a critical stage in clarifying thinking. The creation of the data sheet and protocol often go together and the protocol then involves filling in the data sheet in a systematic way. Data sheets make it much less likely that any data are missed. This makes it more likely that analysable data will be collected rather than a set of unsystematic observations. It is sensible to create a draft data sheet and test and then alter it before adopting it. It is useful to use a range of forms, e.g. one for the habitat of each site, one for the laboratory soil analysis and a third for recording pollinator behaviour. Completed data sheets do not need to be taken into the field and so are less likely than notebooks to become lost and can even be photocopied and stored in a separate building to reduce the risk of fire or theft. Using abbreviations saves time in the field but can lead to confusion and doubt over the accuracy afterwards. One technique is to confirm any unusual data, for example by underlining to distinguish them from errors.

Quantifying the variation in the field looks simple but actually requires great skill. For example, sites may differ in cattle densities, slope, tree height and the amount of bare ground and the abundance of a rare orchid. The first decision is how each can be measured. One approach is to divide into classes, e.g. high, medium or low cattle density, but this is a poor technique, since it is often not repeatable and is difficult to analyse statistically. It is thus much better to quantify these as continuous variables. The skill is to find simple ways of obtaining sufficient precision. This is a balance between methods that produce data that are too inaccurate to be useful and methods that are more time consuming than the results obtained justify. It is tempting to quantify all variables but this takes time and reduces the sample size or increases the cost, so it is necessary to balance their use against time or cost.

A frequent mistake is to collect data that are too detailed. Experts commonly advise that a huge amount of data is necessary to assess a population or a physical feature with sufficient accuracy and thus a proposal to compare say, twenty sites, is completely impractical. This is often because the expert has a different agenda. The expert is often seeking a precise measure for each site, but for the comparison it may be sufficient that the ranking is reasonably correct. For example, the grazing intensities might vary considerably between sites and you may wish to have some relative measure that will reveal whether light or intense grazing has any consequences but the expert might be thinking about how to provide a precise assessment of grazing pressure for a given site.

Although data sheets avoid many of the problems associated with notebooks, they still have risks. I know of a case in which data that had taken two years to collect were stolen from a car, another in which all data were kept in a black bin bag to keep them dry but thrown out with the rubbish and another in which a departing researcher had all his field data in a briefcase, which was then stolen. In the last case the depressed biologist was later approached by a small boy selling peanuts wrapped in his data sheets. He was able to buy most of the sheets back at considerable cost. One way of reducing this risk is to add the data to the computer as soon as possible but still keep the original data sheets plus separately stored back-up computer files.

Types of data sheet

There are three main types of datasheet, although there can be some overlap.

Single event sheets. These are completed on one occasion, for example for a single survey.

Continuous data sheets. Each new observation is added. For example, for each new individual caught and marked. Date and location are usually added each time.

Updated data sheets. These are usually based around a site, nest or individual. Much of the data is recorded once, but future visits are used to record extra data. Data may be added with each visit.

Example A. Make repeated visits to a series of lakes counting the waterbirds on each and record the diet and nesting success of fish eagles. Lake data sheet to describe the habitat (single event sheet); count data sheet (continuous data sheet); fish eagle diet data sheet (continuous data sheet); nest record card (updated data sheet).

Example B. Carry out a series of point counts recording habitat and collect data on all orchids found and detailed data on flowering time and seed set for the rare species. Data sheet for each point count giving habitat (single event sheet); orchid data sheet (continuous data sheet); rare-orchid data sheet (Updated data sheet).

Box 1.1. **Tips for creating data sheets**

- Place boxes around everything that has to be filled in, especially if others are completing the forms.
- Make the box size appropriate.
- Rearrange the sheet so that space is used sensibly. It may be possible to fit multiple copies (e.g. more than one survey) onto one sheet.
- Order the data on the sheet in a logical order that relates to a sensible sequence in the field.
- Consider how the data will be typed into the database. The data sheet should mirror the database.
- Consider analysis. For example, should data be continuous or categorical?
- Show to others for suggestions of improvements both to the structure of the sheet and to the methods being used.
- Think about what should be done for ambiguous and unusual cases. Write down these criteria.
- If several observers are to use the sheet, consider including an example record at the top of the columns.
- Test and modify. Use all of the following three approaches.

(1) Imagine being in the field and think how the data would be collected.
(2) Test in the field. If necessary, this can be carried out on unrelated systems. There might not be any rainforest nearby, but going into the local wood and considering how the data will be collected will reveal many problems.
(3) Modify after carrying out in reality. It is thus ideal, if it is possible, to modify the data sheets after an initial field visit.

- Leave a column for notes in which unusual events can be recorded. However, it is important that almost everything is quantified and there are not essays written on each.
- Decisions should usually be made in the field (e.g. whether the individual is within the study area). This often requires making definitions (e.g. if the individual is heard but not seen, does it count?).
- Include a box for the observer's name and usually have boxes for date and time started and ended.
- Ask for exceptional cases to be underlined (to distinguish them from errors).

Start and encounter reality

Having planned your project, you have ensured that your project seems feasible and practical and will answer useful questions. Once the data collection starts it is usually clear that the plans are insufficient. It might be impossible to collect certain data. Some data might be far more time consuming to collect than expected. Occasionally some data are much easier to collect than expected. You might notice something really interesting. Under such circumstances rigorously sticking to the plans can be ineffective. However, spontaneous changes of plans rarely work. The sensible solution is to run through the entire reverse planning process again, quickly, and modify it appropriately. Having already worked through a plan in detail and undergone the painful process of planning, it is usually much quicker to change. These decisions can be made almost instantaneously. For example, you notice that some of your study individuals are limping. You then decide to categorise each individual as to whether limping or not and measure snare density in transect in each 1 km^2 of forest. You will correlate the two and plot histograms of percentage limping and snare density at various distances from the forest edge. The transects will take another four days but will replace the vegetation survey that turned out to be more complex than planned because the herbarium specimens proved difficult to use.

Acknowledgements

Thanks are due to Rosie Trevelyan for discussions and suggesting the term reverse planning and to Chris du Feu, Jeremy Greenwood and Rob Robinson for useful suggestions.