*Journal of Insect Conservation* **7**: 33–44, 2003.
© 2003 *Kluwer Academic Publishers. Printed in the Netherlands.*


*Practical conservation*

# Maintaining data integrity in insect biodiversity assessment projects

Simon J. Grove
*Biology and Conservation Branch, Division of Forest Research and Development, Forestry Tasmania, GPO Box 207, Hobart, Tasmania 7000, Australia (e-mail: simon.grove@forestrytas.com.au; phone: +61(0)362338141; fax: +61(0)362338292)*

Received 24 October 2002; accepted 23 March 2003

*Key words:* data integrity, databases, biodiversity assessment, quality control

## Abstract

The success of projects involving assessment of insect biodiversity depends on many things, but one which is often overlooked is the maintenance of data integrity. This is an issue best considered from project conception, through the design phase to the completion of the sample, specimen and data processing phase. This paper considers some guiding principles and details some logical steps that will help avoid loss of data integrity.

## Introduction

The literature on insect biodiversity assessment methodology is overwhelmingly composed of papers discussing the virtues of particular sampling or analysis techniques. Papers reporting the outcomes of insect biodiversity studies likewise tend to dwell on these aspects. Yet between the sampling design and analysis phases lies a lengthy phase of sample, specimen and data processing whose methodology is seldom discussed in the same detail, despite its recognised importance for project success (Cranston & Trueman 1997; Oliver *et al.* 1999). Failure to maintain quality control and data integrity during processing is likely to be a common cause of poor project performance relative to initial expectations, but this may not be apparent from the literature because studies failing to live up to expectations are seldom published. This paper considers some ways in which this source of poor performance can be avoided, through adopting protocols that encourage the maintenance of data integrity throughout the processing phase. It also considers aspects of the design phase that can ease the maintenance of data integrity during the processing phase. It is particularly aimed at researchers new to the field of invertebrate biodiversity assessment, so they may avoid learning the hard way about the importance of maintaining data integrity.

To get a feel for the possible ramifications of failing to maintain data integrity, consider the following scenarios. All are fictitious, yet each contains elements that the author has encountered through peripheral involvement in a range of projects over the past few years.

> We had a great project looking at insect biodiversity across all our forests, which the department was going to use to plan reserves. It was a huge effort using all those different traps, and must have cost the government a fortune. When we got the samples back, we sorted them to ordinal level. It took ages to count all those flies and beetles, but once we had done so we weren't sure what to do with the results. Then we farmed some taxonomic groups out to contractors, but not all of them reported back and those that did just gave us species lists with no indication of which species were found in which sample. We didn't keep other groups separate after counting them but put them all back into the sample pots. We've probably still got all the samples (minus a few taxonomic groups – not entirely sure which) somewhere in the basement, assuming they survived the flooding when the pipes burst. However, we haven't had the time, money or expertise to do anything with them, and I'm the only person left

here who remembers anything about what went on – I keep meaning to write up what we did. They've since decided to select reserves on the basis of their vegetation and birds instead.

I know we found ten jewel-beetle species at Site A, and fifteen at Site B. However, I can't actually tell the planning department if Site A really has fewer species than Site B, because our sampling intensities differed and we only made a note when a species was first detected, not which samples it occurred in and at what abundance. I think the planning department has now given up on us ever telling them anything useful about insects at these sites.

We have just heard that, in addition to the ubiquitous stag-beetle species X, there is the possibility that our study Site A, which is scheduled for clearfelling, is within the restricted range of the newly recognised and closely related sibling species Y which has just been listed as nationally endangered. We have a voucher collection of species X *sensu lato* from Site A, and none of them is identifiable as species Y. However, we threw away the rest of our specimens so we cannot say whether or not we have collected species Y from Site A. It now seems likely that Site A will be clearfelled later this year.

Dr Hottereaper presented us with a huge bill for going through our true bug collection and compiling a database of the records we needed to do our modelling of species ranges. We didn't think it would take her so long because we had already thrown out the specimens that had been attacked by mould and museum beetle, leaving only about five thousand, most of which had data-labels. But it seems some of the labels had faded so badly she was no longer able to read them, while others would only have made sense to the person who hand-wrote them years ago (who has now retired). None had any useful sample codes or accession numbers that would have speeded up databasing.

How were we to know our vacation student was no good at databases? He always looked busy – if a little bored – typing in all that information off the specimen labels prior to us throwing the collection out to make room for another office. It was only after he went on sick-leave (with repetitive strain injury) that we found out he was entering everything into a flat spreadsheet. I must admit we were a little suspicious when the computer was only able to find ten records of *Blobiopsis obscurellus* in the entire Northern District, because I'm fairly sure there would have been at least five times that many in the collection. How were we to know he had spelt the species name ten different ways? And I wish we could now go back to the collection to find out whether all those records for inner-city Launchester weren't just due to the software auto-completing an entry beginning with La, like Lake Sinclair.

I was hoping that our long-term aphid monitoring project would throw some light on whether there had been any noticeable effect of climate change in recent years. The department has been collecting specimens using sticky traps every month for decades. However, we haven't always used the same trap design, and for a while we were using two designs simultaneously but didn't record which sites had which design. Nobody now remembers what the old traps were like – whether they were any more or less efficient than our current one. We also took the decision some years ago to merge samples into three-month blocks before data entry because this was adequate for our main interests. And we merged all the non-pest species into an 'other aphids' category because they were marginal to our main interests. Unfortunately this means we can't analyse perhaps 70% of our species, and our choice of three-month blocks means we can't detect any change in seasonality even though farmers are telling us they're needing to spray several weeks earlier now.

## Some guiding principles for the processing phase

The problems exposed in the above scenarios could have been avoided if those involved had kept to a few key principles.

*Principle 1. Recognise that good science does not come cheap, and processing specimens and data carry a significant cost*

Each stage described in this paper has budget implications – not just the sampling. Later activities will frequently consume more resources than the fieldwork (Cranston & Trueman 1997), and may depend on a higher (and therefore more costly) level of expertise, which must be factored into project budgeting. Financial resources are usually limited, so it is often better to plan an affordable small project rather than

waste resources on half-doing an unaffordable big one. It may be possible to postpone the later stages of processing and analysis if funds run out, but only if all necessary steps have been taken to ensure there will be no loss of data integrity in the mean time.

*Principle 2. Be consistent and methodical*

Without adequate quality control and forethought, even small projects can descend into chaos: samples can become detached from sampling data, specimens can go astray, databases can become out of sync with the material they refer to, or corrupted. Developing systems that enable one to be methodical about each stage of processing is therefore important for maintaining data integrity.

*Principle 3. Document and database with a view to making downstream processing and analysis easier*

When one is engrossed in a particular project activity, it is easy to forget how easy it is to loose track of the details. For instance, why were there three plots at Site A but only two at Site B? or which traps were damaged during the last sampling period as opposed to just empty? or how far through the sample series was the alcohol replaced? or who were all those flea-beetles sent off to? or what did that visiting expert say was the name of species X? Unless these details are recorded accurately at the time, there is a risk they might never be, with clear consequences for loss of data integrity.

*Principle 4. Leave any lumping of samples or taxa to the analysis phase*

It is always possible to merge data but impossible to split it beyond the level at which it was first recorded. In insect biodiversity assessment work, pitfalls to avoid are prematurely lumping taxa, samples or developmental stages.

Biodiversity data become increasingly versatile with increasing taxonomic resolution (Doledec *et al.* 2000). Many papers are published in which sites, samples or treatments are compared on the basis of ordinal-level abundance data. However, whilst there are exceptions (reviewed in New 1996), few of these offer much in the way of insights into the processes at work, because so many processes operate at the species level. Thus it is good practice to sort material to the species or morphospecies level if at all possible. If needs be, the data can be lumped up to higher taxonomic levels at a later stage. By contrast, it is impossible to split data stored at a higher taxonomic level into its component species. A difficult situation to be in is where some of the material has been sorted to species level while the rest has not (e.g. Lucanidae species 1–3 vs. Staphylinidae spp): they cannot readily be analysed together without lumping everything to the highest common level.

Similarly, it is good practice to treat all samples as separate entities throughout processing unless there is very good reason to lump them before processing. Keeping them separate allows greater flexibility in analysis – for instance, to detect levels of between-sample variation or species turnover, or to estimate species richness through examining the rate of species accumulation by sample (Brose 2002; Cam *et al.* 2002).

An additional issue arises where material of a single species comprises mixes of developmental stages. This is a situation that could arise with insect taxa that undergo incomplete metamorphosis (e.g. bugs, grasshoppers), and with many other invertebrate taxa such as millipedes and molluscs. When databasing such material it is important not to lump them if subsequent analyses will treat all material as having equal status (generally assuming that data refer just to adults). For instance, it would be invalid to treat abundance data for bugs on an equal footing with abundance data for beetles if the former included juveniles but the latter did not.

*Principle 5. Curate and archive project material and data on the premise that it has long-term scientific value*

Many present-day researchers assume their samples and specimens will have little value beyond their own study, but by using durable materials, informative data-labels and well-constructed databases that value can endure and grow. If material ends up in recognised institutions, it may contribute to a much greater body of scientific knowledge than initially conceived. At the very least, it allows future researchers to revisit the original study in the light of new knowledge and understanding.

## Nine steps to maintaining data integrity

The above guiding principles should operate throughout the design and processing phases (Figure 1). Each stage in these phases is discussed in more detail below in relation to the issue of maintaining data integrity.

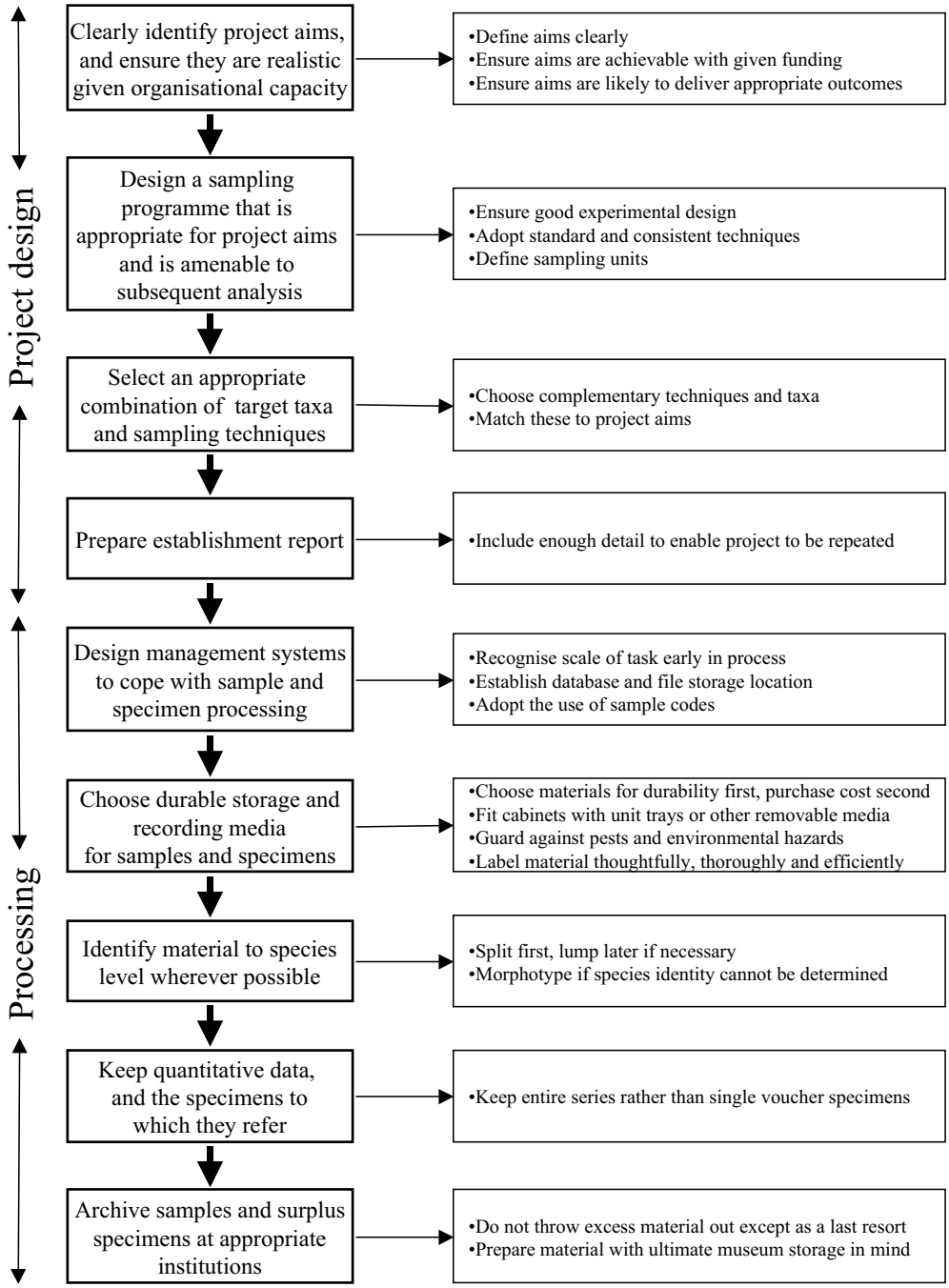


*Figure 1.* Key stages in project design and processing that can ease the maintenance of data integrity.

*1. Clearly identify project aims, and ensure they are realistic given organisational capacity*

It is easy to underestimate the costs – in time, money and expertise, involved in seeing a project through to completion. Trying to process samples, specimens and data on a shoestring is asking for trouble and is a common cause of loss of data integrity. Insect biodiversity is immensely complicated, and any project that investigates it has the potential to become immensely complicated too. In general, projects are most likely to succeed if they have clearly definable and narrowly

focussed aims (Yoccoz *et al*. 2001). For instance, rather than starting a project to look at the impact of forestry on insect biodiversity, it would be prudent to restrict it to something more achievable, say, the impact of clearfelling on litter-dwelling beetle biodiversity.

*2. Design a sampling programme that is appropriate for project aims and is amenable to subsequent analysis*

Good experimental design is the key to a good sampling programme (Scheiner & Gurevitch 1994; Underwood 1994; Margules *et al.* 1998; Deutschman 2001). As well as requiring consideration of spatial arrangement, replication, controls, etc., experimental design involves standardising on one or a few repeatable techniques, on what constitutes a sampling unit, and on sampling frequency (New 1996; Oliver *et al.* 1999). All of these issues can affect data integrity. Consistency throughout the sampling period is important.

It is important to choose sampling programmes that will mean it is possible to relate the resultant data to sampling effort. No technique is fully repeatable, but some are more so than others. Passive sampling techniques such as pitfall trapping and Malaise traps are fairly repeatable, whereas more active techniques like sweep-netting or hand collection are less so (though some have tried to make them more so: Nummelin & Borowiec 1992; Schilthuizen & Rutjes 2001). Standardising on sampling units is equally crucial. In many cases, it is appropriate for each trap to be a single sampling unit. This leaves open the option of merging data from suites of samples if necessary, without losing sample integrity. If a sample unit consists of an array of pitfall traps, the contents of which are merged before sorting, it is important that there are always the same number of traps in the array. Additionally, the trap design should remain constant for the duration of the project, to the extent that it ensures equality in sampling efficiency. Standardising on sampling frequency involves establishing a programme for regular sampling, say, every week or month. If traps operate continuously, then the contents should be collected at regular intervals if there is any intention of analysing the data to look for time-related patterns. It is not easy to detect seasonal patterns in species richness or assemblage composition when one set of samples represents one month's continuous sampling while the next set represents two months: you cannot just divide the data by two.

*3. Select an appropriate combination of target taxa and sampling techniques*

The entomological literature is full of papers discussing the relative merits of particular taxa (Andersen *et al.* 2002) or techniques (Southwood 1978; Disney *et al.* 1982; Hill & Cermak 1997; Stork & Hammond 1997; Grove 2000; Meades *et al.* 2002) for particular purposes. No one approach is universally suitable; all have their weaknesses even in the best of real-life situations because of variations in trappability of different insect taxa and under different conditions (Muirhead-Thomson 1991; Vennila & Rajagopal 1999), and because so many insects are taxonomically intractable. The important thing is to ensure that the sampling technique(s) chosen are optimal for sampling the chosen target taxa, which are in turn appropriate for the project aims. If the aim is to survey the entire insect fauna, then a range of techniques will be required. Usually, projects have a narrower aim which can be met through focussing on selected taxa and sampling techniques (Richardson *et al.* 1999). For instance, pitfall traps are appropriate for sampling a wide cross-section of ground surface-active insects, and would be a suitable technique if one were interested in monitoring the impacts of changes in litter quality, but they reveal nothing about canopy structure. Malaise traps suspended in the canopy might be a better bet for such a study, but then one would have to target different taxa – perhaps flies or wasps. By the same logic, the presence of the occasional fly or wasp in a pitfall trap would be no reason to include data on flies in analyses of pitfall trap data.

*4. Prepare an establishment report*

It should never be assumed that all the necessary detail about the early stages of a project will be remembered by those involved in setting it up, even if its total duration is only weeks or months. The longer a project continues, the more likely it is that someone else might have to take on where others left off. It is good practice to produce an establishment report while all the detail is fresh in the mind. The report should outline the project's aims, and should detail study sites, sampling programmes, locations of files, names of key personnel, and as many aspects of downstream sample and specimen processing as will assist later interpretation and analysis. It should be detailed enough to enable someone to repeat or replicate the study. In other words, it will document many of the elements outlined in this

paper. The report should be stored in a suitable archive so that its existence will be known to anyone involved in the project at a later date.

*5. Design appropriate management systems to cope with the processing of samples and specimens*

Sampling programmes can generate vast amounts of material. Consider a simple study, comprising ten sites, each sampled using arrays of ten pitfall traps operating over a year and collected monthly. This study would generate $10 \times 10 \times 12 = 1200$ samples. Each sample will need to be readily identifiable from the moment it is collected. Even before specimens are removed it may need curating, such as removal of debris and transferring to alcohol. Even if only a single taxon is to be removed from each sample, the study will generate a further 1200 sample entities, all of which will require curation separate from the sample residue. If, on further sorting, each taxon sample is divided into an average 10 separate species samples, that makes 12 000 further sample entities. Thus the simple study has generated 14 400 separate sample entities.

It is wise to have systems in place from the outset to keep track of this material. In the simple example above, it would be feasible to set up a database (or at the very least a spreadsheet) knowing in advance that there will be 1200 samples to populate it. Relational databases have numerous advantages over flat spreadsheets for data storage and manipulation. They store data much more efficiently by avoiding superfluous replication of data, and they reduce the risk of making mistakes in data entry because they can be set not to tolerate spelling mistakes, multiple entries of supposedly unique records, etc.

For people or organisations likely to end up managing information on large numbers of projects with wide geographic and taxonomic coverage, there are several biodiversity database management systems designed for the job which are available commercially (e.g. Biota, http://viceroy.eeb.uconn.edu/biota) or even for free (e.g. BioLink, http://www.biolink.csiro.au/; Specify, http://usobi.org/specify/; Biotica, http://www.conabio.gob. mx/informacion/biotica_ingles/doctos/acerca_biotica.html). These systems attempt to cover a wide range of possible uses, so can have hundreds of data-fields in dozens of linked tables, making use of barcode technology and automated production of labels, GIS capabilities and the ability to store digital images. Simpler, 'home-made' systems can be equally or more effective for smaller-scale ventures, and have the advantage of being easily modified to address specific sampling issues or analyses. If they grow too big, they can be electronically transferred to one of the proprietary systems at a later date. The relational database design employed by the author (Figure 2) is presented here purely to illustrate the main elements required; there are undoubtedly other, probably better, designs but they will share at least some of these elements. The main elements are a series of tables covering the

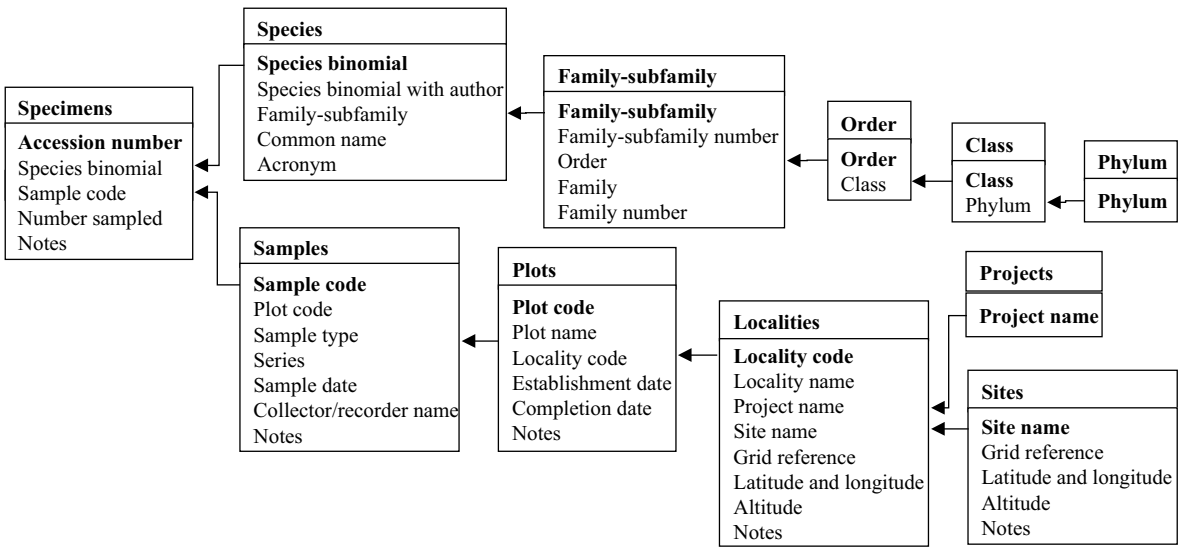


*Figure 2.* Example of a simple biodiversity database, based on that of the author, that can ease the maintenance of data integrity. The diagram represents a relational database with linked tables. The table name appears above the line in each box. Only the key fields in each table are shown. Arrows indicate direction of one-to-many relationships between linked fields.

taxonomic hierarchy, another series of tables covering a site hierarchy, a projects table, a samples table, and a specimens table. Some of the fields are coded (numeric) equivalents of text fields (e.g. family number), so that the data can be sorted numerically (e.g. taxonomically) as well as alphabetically. The specimens table houses just the important information about what species were found in what sample and how many, making data entry straightforward. All supporting information about what those species and samples refer to resides in other tables and only need be entered once. Single records can readily be entered directly into the database, whereas multiple records can more conveniently be imported via a spreadsheet. Taxonomic information (e.g. regional species lists) are often published on the Web (e.g. the Australian Faunal Directory at http://www.ea.gov.au/biodiversity/abrs/abif/fauna/afd/frames/chcklist.html), from where data can be downloaded and subsequently imported into the database. Since taxonomic understanding continues to advance, there will be a need to revisit these information sources (or others) regularly to maintain up-to-date species lists within the database. In some regions, on-line gazetteers are also available and can be a useful source of site data. Whatever design of database is chosen, it is important to backup data regularly.

There is an increasing global trend for developing virtual databases and metadatabases of invertebrate biodiversity data, in which databases from geographically separated collections are linked over the Web. An example is the Australian Plant Pests Database, which seeks to provide information on the distribution of potential insect pest species based on physical specimens in collections around Australia. For these systems to work, certain data standards must be employed by each participating institution so that data can be shared, or systems can be interrogated remotely. This is yet another reason to develop databases, using standard software, early in a project, so that the data can one day be fed into one or more of these networks without someone else having to enter it all afresh.

It is a great advantage for later data analysis to allocate every sample a unique, preferably numeric, code (or lot number). This number is shared by all specimens from a single sample (e.g. from a single pitfall trap clearing), and is unique to those specimens. Employing sample codes can help with sample processing too if labels bearing the sample code (plus all other relevant information, see below), or a barcode, are inserted into the samples at the time of collection or when they were brought back from the field for curation. If transparent sample containers are used then labels can be inserted so that they can be read from the outside. Inserting labels is anyway preferable to sticking them on the outside of sample containers, or writing directly onto the containers. These latter approaches risk loss of data integrity because labels peel off or text can become illegible due to abrasion, staining or fading.

Keeping samples from different months or sites in separate labelled boxes also helps keep track of processing, and can be used to break down otherwise overwhelmingly large processing tasks into more manageable ones. A spreadsheet, database or notebook is a useful tool for keeping track of sample processing, with each column or field representing a stage of processing that has been completed.

*6. Choose durable storage and recording media for samples and specimens*

In collections around the world there are innumerable samples and specimens whose potential scientific value has not been realised because poor-quality materials were used when they were first acquired, or because their data-labels lack even the most basic of information, have faded and become illegible, or have disintegrated. Complacency about the risk of degradation is probably the biggest cause of damage to specimens and loss of data integrity.

If material is to be wet-stored long-term, there is a need to consider the nature and quality of the preservative solution and the containers in which they are stored. Simple measures like using deionised or distilled water rather than tap-water when diluting preservatives can prolong material life. Choosing containers that resist chemical degradation, leakage and evaporation is also important. It may not be enough to assume that they will be suitable if they show no detectable change over several months. Unless they are going to be regularly curated, they may need to survive years.

If specimens are dry-mounted, then choice of pin, card and glue is important. Pins should be resistant to corrosion, card should be acid-free and glue should be water soluble. Really sharp pins are more expensive than blunter ones, but make mass-mounting series of specimens much easier because it is possible to use them to pick up points and labels by ‘stabbing’ rather than having to do this with fingers and thumbs. Using a mounting block to mount specimens and labels at standard heights up the pin is a great advantage for later visual scanning of trays of specimens. Using storage cabinets whose drawers are fitted with moveable but tight-fitting unit-trays adds greatly to the useability of a voucher collection. Precautions are necessary

against fungi and pests that might otherwise eat their way through the collection. One technique is to insert slow-release chemical repellents (e.g. naphthalene) and fungicides (e.g. thymol) in each drawer. The best drawers for this purpose are those with a moat around the periphery designed specifically for holding the repellent. An alternative technique is to periodically deep-freeze drawers to kill any pests that might have entered. Either way, vigilance is essential. It is also important to maintain a record of when each drawer was last treated or is due for additional treatment, and to have this readily available as a reminder. Storage cabinets should not be kept in hazardous environments (e.g. prone to flooding or excessive heat), or where they might be forgotten about.

If each sample is allocated a unique sample code or lot number (as advised above), then this is a very valuable item of information to put on every specimen label. It may not mean much to those not familiar with the codes, but within the confines of the project it will greatly improve the efficiency of handling and storing specimens and in databasing specimen information. Consider sorting 10 samples, each representing one of five sites sampled on one of five occasions, using one of three techniques, and each containing up to a 100 specimens belonging to up to 10 species. Having divided them up into their appropriate species, how should they be databased? If sample codes, or barcodes, have not been employed, it will be necessary to examine the detail of every label and transcribe the site, sampling technique and date information for each. If sample codes have been employed, the code is all that is needed to transcribe. It is then relatively straightforward to make up a tally sheet, whereby each time another specimen for a given species is scored it is tallied next to the appropriate sample code. If all specimens of a particular species from a particular sample are kept together on the same mount (as discussed below) then the process is simpler still. Once the tally sheet is typed into a spreadsheet the whole lot can be imported into the database. Data entry for barcoded specimens can be simpler still with the use of a barcode reader.

Besides the sample code, a label should contain basic information useful to outsiders (e.g. Figure 3). It is best to assume that each specimen could end up anywhere in the world, so latitude and longitude coordinates are preferable to a locally or nationally defined grid reference. Region/State and even country information are useful too, as well as site/locality/plot information, date of sampling, sampling technique, and name of collector/institution. Text can be abbreviated to squeeze it into the six lines which is about all that can fit on a specimen label. If this is still insufficient space (e.g. for host associations, rearing records), a secondary label can be prepared to be mounted below the primary one. Storage space is expensive, and having labels twice as big as they need to be means that storage is twice as costly as it needs to be. For pinned insect specimens, the label should be small enough so that it does not become the main limitation on how many specimens can be fitted into a given space. The size of the specimen will then often be the limiting factor for how many can be stored in a given space – but as an aside, even here there is room for space-saving. Rather than setting all specimens so that their legs, wings or antennae stick straight out, experienced collections managers advise that it is preferable to set the bulk of them with these appendages folded in close to the body so that they are still visible but take up less space. This also reduces their vulnerability to being damaged.

If a database is established from the outset of the project, it becomes an easy matter to pre-print sheets of labels for every sample or specimen using the appropriate fields from the database. There are several ways of doing this. The database's report function could be used to print out rows of labels directly. A program designed specifically for producing labels, or the mail-merge function in word-processing software, could be used instead. Or the appropriate fields from a database could be exported into a spreadsheet, and manipulated there to get the desired format. This last option is the author's preferred option. It takes a bit of effort initially, but only needs doing once for each set of samples – perhaps only once for an entire project. Once the required fields have been exported into a series

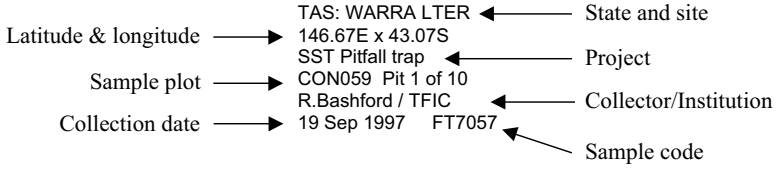


*Figure 3*. Example of a specimen label including the main elements that can ease the maintenance of data integrity.

of columns (with one row per sample), they are rearranged to condense the information content into no more than six columns. They are then transposed so that the six successive columns form the six rows of a label, with successive columns representing successive sample labels. What is done next depends on the purpose of the labels. If multiple labels are needed for the same sample, as is generally the case when preparing specimen labels, then the first step is to apply a suitably small font so that a label with six rows will not take up too much space but will still be legible. Arial 3.5 pt is suitable for this. After readjusting column widths and row heights, the rest of the page is then filled up with copies of the first six rows, with successive pages representing successive sequences of samples. If only single labels are needed, which is generally the case when preparing labels for unsorted samples, a larger font, such as Arial 10 pt, is applied. The same process of filling up the entire page with copies of the first six rows is followed, but in this case unwanted repeats are then progressively cut out so that the spreadsheet ends up being one page wide, with successive sequences stacked one below the other. In either case, the spreadsheet is given a suitably descriptive file name and header and filed in a suitably named folder alongside related files.

It is important to print labels with regard to their durability. Choice of printer/ink and paper/card both affect durability. Printer inks and printing processes vary in their durability, and many that are current now have probably not been around long enough for their long-term properties to be adequately assessed. If time is not an issue, then hand-writing labels using graphite pencil or India-ink is a proven technique for wet and dry material. In general, most laserjet inks and photocopy inks do not bond sufficiently to paper to survive long-term immersion in spirit or water. Some people recommend ironing or baking the paper after printing, to melt the ink deposit further into the paper. Nevertheless, there are sufficient stories of museum curators finding 'alphabet soup' at the bottom of their sample pots to remain cautious about relying on these inks. They may, however, be fine for labels for dry-mounted material. Inkjet inks can be better at bonding with paper, but this depends partly on the solvent used to make the ink. Sometimes the solvent itself damages the paper or the specimens.

Most office paper is not of archival quality, meaning that it yellows, disintegrates or causes the ink to fade over years or decades. It is much better to use archival, acid-free paper, or goatskin/cotton parchment, available through art supplies. Choose the thickest (highest gsm) that will fit through the printer: 120–200 gsm is generally suitable. Avoid laminated card (e.g. thick colour photocopy paper), especially for wet material, as it can gradually turn to pulp. A relatively recent alternative to paper or card made from natural fibres is 'plastic paper'. An example is Teslin® (http://www.ppg.com/chm_teslin/whatsteslin/whatis.htm), whose normal use is for printed identity cards. It is extremely resistant to chemical attack and mechanical abrasion, does not pulp, and bonds strongly with laserjet inks. In the author's experience over several months, it is suitable for wet or dry mounting, and is not unduly expensive compared with art paper. It can be ordered through office supplies outlets.

If sufficient resources are available, and if concerns remain about how well some of the new technologies may perform in the long-term, it may pay to get labels printed on archival quality paper by professional printers. They use genuine printing inks that do not suffer from the same shortfalls as the inks described above.

### *7. Identify material to species or morphospecies level wherever possible*

The cost (time and money) of identification increases with taxonomic resolution, but so too does the value of the resultant data (Figure 4). In general, it would be better to admit that one cannot put names to the 100 beetle species in a sample, and instead to call them beetle species 1–100, than to give up and hope that leaving everything in a category with the scientific-sounding name 'Coleoptera' is going to impress peers. Generally it won't (or should not), any more than a botanist would be able to impress peers by trying to publish papers in which plant communities were compared on the basis

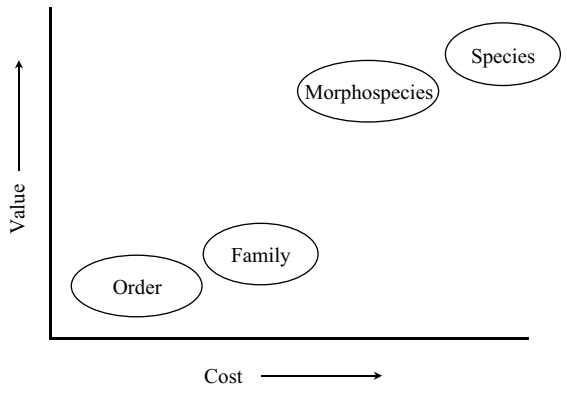


*Figure 4*. Schematic representation of the relative costs and values of identifying project material to different levels of taxonomic resolution.

of the number of Asteraceae versus Myrtaceae. There are some situations where savings can be made by only identifying material to higher taxonomic (e.g. ordinal) level (New 1996), but normally this would provide insufficient resolution for biodiversity assessment purposes. In the majority of cases, if there is insufficient time, money or expertise available to sort a proportion of the material to species level, it will often be better to exclude that portion from analyses and focus on groups that can be sorted to species level. Whilst it is best to aim for precise species identifications, much can be achieved by sorting to morphospecies, or recognisable taxonomic units (Oliver & Beattie 1996; Pik *et al.* 1999; Derraik *et al.* 2002).

*8. Keep quantitative data, and the specimens to which the data refer*

Generally, one cannot be sure that one has sampled the entire species-pool at any site. Comparing species richness among sites then relies on knowing something about relative sampling intensity. Additionally, there is a need to be able to estimate rates of species accumulation with successive samples, because the relationship is seldom linear. Since the number of species in a single sample depends on the number of individuals sampled (but not linearly), it is important to know this too. Failing to keep quantitative data (i.e. recording species X from sample Y but not recording how many) is a common cause of loss of data integrity. Social insects such as termites and ants (Longino *et al.* 2002) present a slightly different problem (when is an individual an individual?), which makes it generally inadvisable to carry out analyses using abundance data for mixes of social and non-social taxa.

Another cause of loss of data integrity is failing to keep track of all the specimens of a species from a given sample. For instance, imagine that 10 specimens of species X were recorded from sample Y, of which only one was put aside in a voucher collection for species X, the rest being discarded. Later on, it is noticed (perhaps by a successor to the initial investigator) that the voucher collection of species X actually comprises two species: perhaps the species has been split following taxonomic work which was not published at the time of the first sort. It is possible to sort out the vouchers but impossible to say which of the specimens that were discarded belonged to which of the two newly recognised species, so it is not possible to make use of the quantitative data except by re-lumping.

Both these situations emphasise the importance of maintaining a comprehensive reference collection of identified specimens. If the identities of different specimens have been determined by different people, it is appropriate for each specimen to have its own 'det' label, comprising the species' name and the name of the person who identified it. Alternatively, all members of a given species can be stored in a single location (e.g. unit tray) which is itself labelled with the species name. It is best to avoid using non-project-specific names for morphospecies such as Curculionidae sp 01. If a morphospecies is initially identified only by a code-name, it is advisable to choose a code-name that is unique to the specimens to which it refers, for instance by including a reference to the project in the code-name. For example, it is clear from the code-names for Curculionidae SST sp 01 and Curculionidae LD sp 01 that they are not being treated as one species but as two, each derived from a separate project. If it is later decided that Curculionidae SST sp 01 is the same as Curculionidae LD sp 01 but its true identity is still unknown, then they can be merged and given a new name that perhaps refers to the entire reference collection (e.g. Curculionidae TFIC sp 01). All associated records should then be changed in the database accordingly. If each specimen has a sample code or accession number, then it is relatively quick to note down which specimens require identity updates in the database. If each has an additional barcode label, data management can be easier still.

It takes time to prepare specimens for storage, and space to store them. These ought to be considered justifiable project costs and budgeted for. Even if space or time is really limited, it is advisable never to discard all material: vouchers should always be retained for future reference. If identifications are certain, or if the integrity of the work does not depend on the need to revisit the specimens, it may be worth only retaining a few vouchers for each species and discarding the rest. However, one should first investigate whether anyone else would like the spare specimens: for research, teaching or for their own reference collections (see below).

More often than not, there will be some lingering doubt about the identity of some species collected. In such cases, it is best to keep all material, unless it is clear that nobody is going to want to make use of it in future. If knowledge advances and the 'species' in question are split, lumped or redefined, it will be possible to reallocate all specimens to the new species, and update the database accordingly. If there are multiple specimens of one species from a single sample, it may not be necessary to mount them all individually. If they are small enough, one option is to mount one but to

put the rest in a transparent gelatine capsule, boxes of which are available from pharmacists. The number of specimens can be written on the outside of the capsule using a permanent ink-pen. The capsule can then be mounted on the pin beneath the single mounted specimen. Another option is to keep the extra specimens in vials in a separate storage area, which can either be in taxonomic or sample code order. It will still be necessary to add a label to the mounted specimen to indicate that further specimens are stored elsewhere.

*9. Archive sample residues and surplus specimens at appropriate institutions*

Only consider throwing material out as a very last resort. If sample containers, pins, labels etc have been chosen with thought for the future, then there are many museums, regional and national collections that may be willing and eager to receive surplus specimens and sample residues – but most will be reluctant to accept poorly curated material. Some of them may do no more than store them; others may have active systematics departments able to make ready use of the material (Funk & Richardson 2002). Cultivating a good relationship with key people at these institutions early on in the life of a project serves two purposes. It will clarify their own needs in relation to curation standards before different standards are adopted; and it may result in useful feedback on what they found in any samples subsequently sent them.

## Conclusion

Following these nine steps would not guarantee a successful project outcome. However, it should greatly limit the opportunities for loss of data integrity, which is a precursor to a successful project. It is worth comparing the scenarios painted at the beginning of this paper with the one that follows, and asking how one would like one's own projects to turn out:

> Our project went really smoothly, despite some data-gaps due to traps being destroyed by rats and a cyclone. Because we kept good records, we were able to work around these in our analyses – in fact, they showed that if we were doing a similar project again, we could get away with only half the sampling effort. And because we were able to demonstrate that we were serious about our data, we got a lot of support from taxonomic experts and a top-up grant from the research council. I think the Board would have liked our project to have come up with a different answer, but they recognised its scientific underpinning and have taken steps to implement our management recommendations. They have also increased and guaranteed our budget for the next three years, so we can extend our assessment into a longer-term project to monitor the impacts of the mitigation and restoration measures that are to be introduced. I even heard the CEO talking about the importance of beetles for ecosystem health on the TV last night!

## Acknowledgements

I thank Geoff Monteith and colleagues at the Queensland Museum for sharing their ideas on sample and specimen processing when I first started my doctoral research. I also thank them as well as Dick Bashford (Forestry Tasmania) for helpful comments on a previous version of this manuscript. Mention of trade names or commercial products does not constitute endorsement or recommendation for use.

## References

Andersen, A.N., Hoffmann, B.D., Muller, W.J. and Griffiths, A.D. (2002) Using ants as bioindicators in land management: Simplifying assessment of ant community responses. *J. Appl. Ecol.* **39**, 8–17.

Brose, U. (2002) Estimating species richness of pitfall catches by non-parametric estimators. *Pedobiologia* **46**, 101–7.

Cam, E., Nichols, J.D., Sauer, J.R. and Hines, J.E. (2002) On the estimation of species richness based on the accumulation of previously unrecorded species. *Ecography* **25**, 102–8.

Cranston, P.S. and Trueman, J.W.H. (1997) Indicator taxa in invertebrate biodiversity assessment. *Memoirs Museum Victoria* **56**, 267–74.

Derraik, J.G.B., Closs, G.P., Dickinson, K.J.M., Sirvid, P., Barratt, B.I.P. and Patrick, B.H. (2002) Arthropod morphospecies *versus* taxonomic species: A case study with Araneae, Coleoptera, and Lepidoptera. *Conserv. Biol.* **16**, 1015–23.

Deutschman, D.H. (2001) Design and analysis of biodiversity field experiments. *Ecolog. Res.* **16**, 833–44.

Disney, R.H.L., Erzinçlioglu, Y.Z., Henshaw, D.J.C., Unwin, D.M., Withers, P. and Woods, A. (1982) Collecting methods and the adequacy of attempted fauna surveys, with reference to the Diptera. *Field Studies* **5**, 607–21.

Doledec, S., Olivier, J.M. and Statzner, B. (2000) Accurate description of the abundance of taxa and their biological traits in stream invertebrate communities: Effects of taxonomic and spatial resolution. *Archiv für Hydrobiologie* **148**, 25–43.

Funk, V.A. and Richardson, K.S. (2002) Systematic data in biodiversity studies: Use it or lose it. *Systematic Biol.* **51**, 303–16.

Grove, S.J. (2000) Trunk window trapping: An effective technique for sampling tropical saproxylic insects. *Memoirs Queensland Museum* **46**, 149–60.

Hill, C.J. and Cermak, M. (1997) A new design and some preliminary results for a flight intercept trap to sample forest canopy arthropods. *Aust. J. Entomol.* **36**, 51–5.

Longino, J.T., Coddington, J. and Colwell, R.K. (2002) The ant fauna of a tropical rain forest: Estimating species richness three different ways. *Ecology* **83**, 689–702.

Margules, C.R., Austin, M.P., Davies, K.F., Meyers, J.A. and Nicholls, A.O. (1998). The design of programs to monitor forest biodiversity: Lessons from the Wog Wog habitat fragmentation experiment. In *Forest biodiversity research, monitoring and modelling* (F. Dallemeier and J.A. Comiskey, eds.), pp. 183–96. Carnforth: Parthenon.

Meades, L., Rodgerson, L., York, A. and French, K. (2002) Assessment of the diversity and abundance of terrestrial mangrove arthropods in southern New South Wales, Australia. *Austral. Ecol.* **27**, 451–8.

Muirhead-Thomson, R.C. (1991) *Trap responses of flying insects: The influence of trap design on capture efficiency*. London: Academic Press.

New, T.R. (1996) Taxonomic focus and quality control in insect surveys for biodiversity conservation. *Aust. J. Entomol.* **35**, 97–106.

Nummelin, M. and Borowiec, L. (1992) Cassidinae beetles of the Kibale Forest, western Uganda; comparison between virgin and unmanaged forests. *African J. Ecol.* **30**, 10–17.

Oliver, I. and Beattie, A.J. (1996) Invertebrate morphospecies as surrogates for species: A case study. *Conserv. Biol.* **10**, 99–109.

Oliver, I., Dangerfield, J.M. and York, A. (1999) When and how to conduct a biodiversity assessment of terrestrial invertebrates. In *The other 99%: The conservation and biodiversity of invertebrates* (W. Ponder and D. Lunney, eds), pp. 8–18. Mossman: Royal Zoological Society of New South Wales.

Pik, A.J., Oliver, I. and Beattie, A.J. (1999) Taxonomic sufficiency in ecological studies of terrestrial invertebrates. *Austr. J. Ecol.* **24**, 555–62.

Richardson, B.J., Azarbayjani, F.F., Burgin, S. and Richardson, S. (1999) Arboreal arthropod biodiversity in woodlands: Effect of collection procedures and geographic distance on estimates of diversity found on two species of *Melaleuca*. *Austr. J. Ecol.* **24**, 544–54.

Scheiner, S.M. and Gurevitch, J. (1994) *Design and analysis of ecological experiments*. London: Chapman and Hall.

Schilthuizen, M. and Rutjes, H.A. (2001) Land snail diversity in a square kilometre of tropical rainforest in Sabah, Malaysian Borneo. *J. Molluscan Studies* **67**, 417–23.

Southwood, T.R.E. (1978) *Ecological methods, with particular reference to the study of insect populations*, 2nd edn. London: Chapman and Hall.

Stork, N.E. and Hammond, P.M. (1997) Sampling arthropods from tree-crowns by fogging with knock-down insecticides: Lessons from studies of oak tree beetle assemblages in Richmond Park (UK). In *Canopy arthropods* (N.E. Stork, J. Adis and R.K. Didham, eds), pp. 5–26. London: Chapman and Hall.

Underwood, A.J. (1994) On beyond BACI: Sampling designs that might reliably detect environmental disturbances. *Ecologi. Applic.* **4**, 3–15.

Vennila, S. and Rajagopal, D. (1999) Optimum sampling effort for study of tropical ground beetles (Carabidae: Coleoptera) using pitfall traps. *Current Sci.* **77**, 281–3.

Yoccoz, N.G., Nichols, J.D. and Boulinier, T. (2001) Monitoring of biological diversity in space and time. *Trends Ecol. Evol.* **16**, 446–53.