


Project no. 018340

**Project acronym: EDIT**

**Project title: Toward the European Distributed Institute of Taxonomy**

Instrument: Network of Excellence

Thematic Priority: Sub-Priority 1.1.6.3: "Global Change and Ecosystems"

# C5.08 (formerly D5.2.1.2): Operational function model for revisionary taxonomy

Due date of component: Month 16
Actual submission date: Month 16

Start date of project: 01/03/2006 Duration: 5 years

Organisation name of lead contractor for this component: 11 RBGK

Revision 2 [final]

| Project co-funded by the European Commission within the Sixth Framework Programme (2002-2006) | | |
|---|---|---|
| Dissemination Level ( "X" in the relevant box) | | |
| **PU** | Public | X |
| **PP** | Restricted to other programme participants (including the Commission Services) | |
| **RE** | Restricted to a group specified by the consortium (including the Commission Services) | |
| **CO** | Confidential, only for members of the consortium (including the Commission Services) | |

**Contents**

## Introduction

One of the central goals of Workpackage 5 is the creation of an internet platform for cybertaxonomy, a collection of online resources designed to enable the scientific community to practise taxonomic work online. In any project involving the design and creation of new software platforms, the first effort must be to fully understand the existing work processes that the software aims to facilitate. To this end, an investigation was undertaken into the current work practises in the field, captured under deliverable 5.08.

The investigation sought to develop a model of the work of revisionary taxonomy across the EDIT partnership, through gathering information direct from taxonomists who were currently working in the field. Due to the complexity of this task, and the broad scope of the information sought, it was decided that the most effective method of gathering this information was by direct interview with the taxonomists concerned.

## The Interviews

The purpose of the interviews was to gather as much information as possible about the particular activities that made up the work of a revisionary taxonomist. The interviews, largely one-to-one, took place over two half-days in the taxonomist's place of work.

In order to ensure that the information gathered was representative of the whole spectrum of taxonomic work, a series of criteria were drawn up. These criteria were used to select a group of taxonomists who, taken together, covered the full range of diversity of taxonomic practise in the EDIT partnership. These criteria can be found in Appendix 1 of this document.

Over the course of the past year some twenty interviews have been conducted, covering fourteen institutes in ten different countries. All of the identified taxonomic groups have been covered, which between them also satisfy the criteria covering environmental conditions and scientific techniques. The models deriving from these interviews can be found on the EDIT website at this address:

http://dev.e-taxonomy.eu/trac/wiki/RevisionaryModels/

### Interview Methodology

A detailed account of the interview methodology has already been published online and can be found here:

http://dev.e-taxonomy.eu/trac/attachment/wiki/RevisionaryModels/interview_review.pdf

# The Models

## The combined approach to modelling

The aim of the modelling phase was to present, in a clear and simple format, the information gathered during the interviews regarding the practise of revisionary taxonomy. Process models are particularly suited to displaying information in this way. They clearly illustrate individual steps and activities and the relationships between them, and the direction of work flow is easy to see at a glance. This clarity is essential to the usefulness of the models, and maintaining this clarity places limits on the amount of information that can be captured. It would be possible to build models that represented every possible course of action throughout the revisionary cycle, but this was decided against, for two reasons.

The first, as mentioned, is that the resulting models would have been too complex to be of any practical use. The aim of the modelling stage is the communication of clear non-technical information. Detailed process models, though information-rich, do not lend themselves well to this purpose. It is recognised, for example, that in real-life activities are rarely completed entirely in discrete steps. The searching of literature is an activity that can be returned to at any stage in the revisionary process. However to represent every possible path of action would render the models unusable.

The second reason is that the models themselves were never needed as actual system designs. A in-depth activity model of the real-life processes of revisionary work would only be of use if the goal of this stage was to build a software system that replicated these processes exactly. This is not the case; workpackage 5 seeks to provide software to facilitate the processes of taxonomy rather than reproduce them in an online form, and as such a precise model of these system was not required.

Though the detailed information gathered in the interviews was not required for system design as such, it was still valuable data about existing work practises; the modelling phase was an open investigation which drew no early conclusions about the existing systems for taxonomy, or the best direction for workpackage 5 to take in augmenting them. The information was therefore captured in associated text descriptions of the models. Text allows for the recording of much more detail than a work-flow model could every usefully convey. The individual steps can be described in much more detail here without compromising the clarity of the models.

So a balance has been struck; the models have been kept relatively high level, and are accompanied with richer text descriptions that capture all the details. It is recognised that in reaching for this balance, decisions have to be reached about which information to include in the models and in which to omit, judgements that are essentially subjective. I have attempted to mitigate this in two ways. Firstly, by applying consistency between the interviews, as demonstrated by the development of the interview framework cited above. Secondly, by ensuring that all the models were developed with the taxonomists themselves and "signed off" through review afterwards. I encouraged all the scientists to think of the process model as their work as much as mine, and to contribute to the development of the model itself, not just to transmit facts in the interview.

## The Merged Model

### Creating a unified model

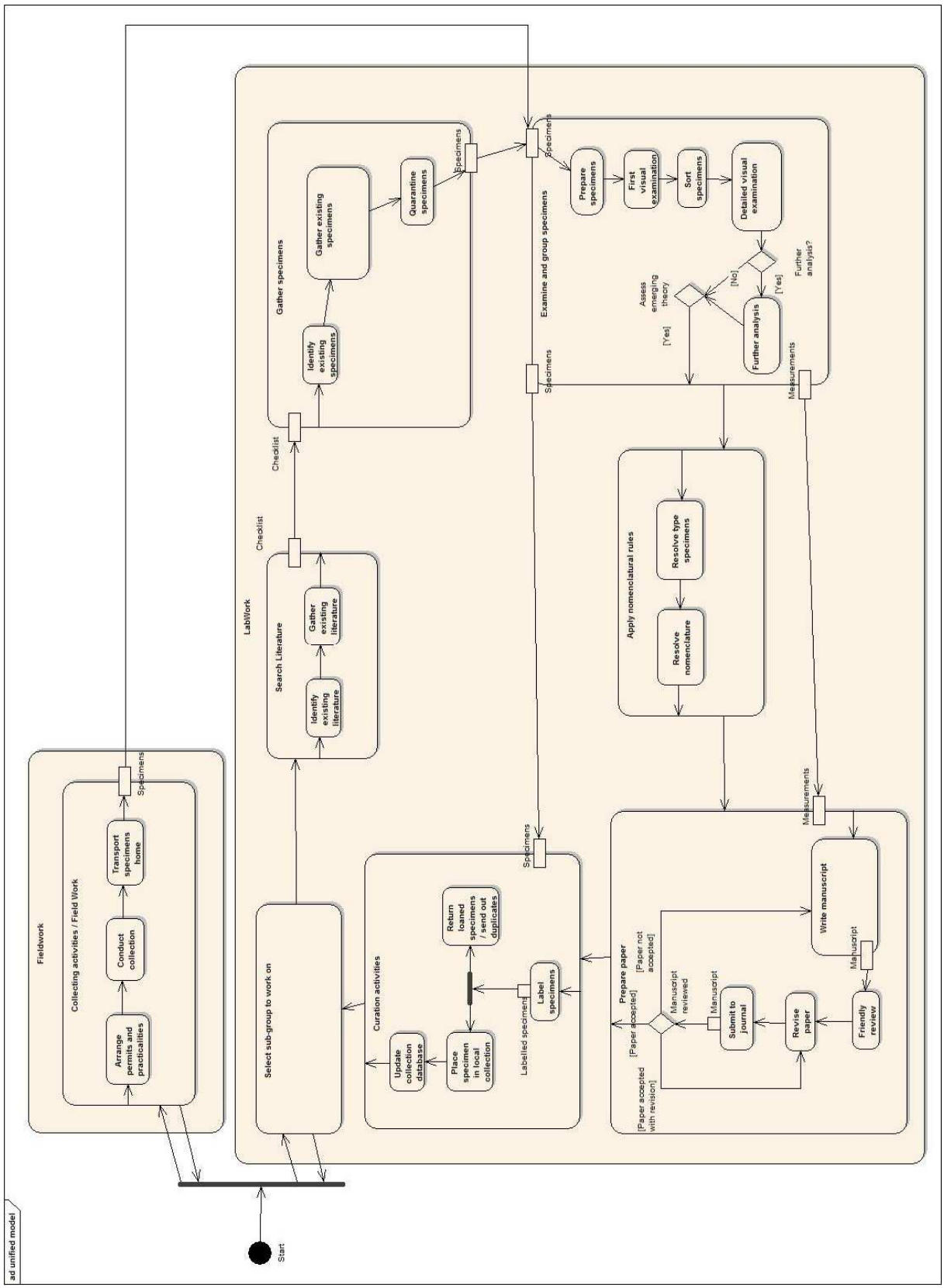
Once the interviews were underway, it became clear that a common structure was emerging. Despite the differences found between the various branches of taxonomy, the process of producing a revisionary work followed certain common themes. To take a very high level example; all revisionary projects involve the gathering and examination of type specimens. It became clear from identifying these common processes that it would be possible to develop a merged model to represent taxonomy as a whole. This merged model contains all processes that have been found to be present in all of the individual models. The order of events is that which was found in the majority of cases; differences at the individual level were resolved into a general case.

The emerging common structure fed back into the design of later models. With the commonalities being confirmed by a number of scientists, it was possible to put forward a structure for discussion rather than approach each set of interviews from scratch. As a result, the later models converge closely around the common form, whereas earlier models will be found to be more distinctive.

One particular aspect of the merged model worth highlighting here is that collecting activities are described as a parallel process. Collecting trips vary widely according to the nature of the specimens sought and the habitat in which they are found. Though some taxonomists did undertake collecting activities solely to focus on a particular group, this was unusual. Collecting trips involve much effort in terms of organisation and can be expensive, so there is a tendency to avoid focussing on narrow goals. Most trips are be arranged to cover a geographical area rather than an a taxonomic group, and so specimens from outside the taxonomist's specific areas will be collected, both for the institutes' collection and for future projects. Many trips are arranged as joint ventures, as collaborations between group of individuals and institutes. For all of these reasons, most trips are not seen as being directly related to any one revisionary project, and are considered a parallel process.

The next two sections contain the merged process model and a full description of the activities described therein.

## The Merged Model



**Details of the Merged Model**

The table below contains further details for each of the activities and actions described in the merged model.

| | | |
|---|---|---|
| Activity | Select a sub-group to work on | This activity doubles as a 'starting trigger' for the model. Varies widely. The direction of a taxonomist's work will be determined by their career path, personal interests, and the requirements and focus of their institution. |
| Activity | Collecting activities/ Field work | This activity refers to collecting new specimens from the field. Collections are undertaken for a wide range of reasons, amongst others:<br>- Collecting specific taxa as part of a project<br>- General collecting to explore a poorly understood area<br>- To improve the institute's collection<br>- Collaboration with a partner institute<br>- In-keeping with institutional goals<br>- Gathering material for teaching<br>- Or any combination of the above |
| Action | Arrange permits and practicalities | The permits required for collecting activities vary according to both location and the material under collection. Permits may be required for a number of reasons, for example:<br>- To enter the collecting site at all. Whatever body maintains the collecting may restrict access, for safety or other reasons.<br>- To remove material from the site.<br>- CITES restrictions may apply to certain taxa<br><br>Most taxonomists tend to organise the trips themselves, often through a collaboration with a foreign institute. |
| Action | Conduct collection | This refers to the actual collection. The activities here will vary widely according to the taxonomic group under study, and the purpose of the collection. It is hardly possible to describe a "general" collecting trip, as the methods and equipment will vary tremendously. |
| Action | Transport specimens home | Most collections will gather too much material to be taken back home by the taxonomist and will generally have to be sent back separately. Far flung collecting trips tend to use air or sea mail. Specimens may have to be treated before packaging, for example, dried, pressed, stored in alcohol. Most specimens will need to |

| | | |
|---|---|---|
| | | be accompanied by the relevant permits. This process can take sometime; items sent by sea-mail, for example, may spend months in transit. |
| Activity | Search Literature | Researching existing literature in order to acquire a full understanding of previous work on the group. This refers to past revisions of the related taxonomic groups, and also more general work. |
| Action | Identify existing literature | Sources for identifying literature are numerous of course, and include:<br>- Online search engines both subject specific, such as IPNI or GEO-REF, or general search engines like Google.<br>- Personal knowledge of the field<br>- Citations in other works. In this way one can follow a 'trail' back to the proto-log<br>- Library search catalogues<br>- Colleagues<br>- Amateur / enthusiast communities<br>- Published bibliographies of works in the field |
| Action | Gather existing literature | Acquiring a copy of the work. In many cases literature can be downloaded from the internet, usually from subscription websites such as the Zoological Record, or online journals such as Zootaxa. Some work can be found published for free.<br><br>If not available online, an inter-library loan can normally be arranged.<br><br>Many scientists also routinely distribute reprints of published work to colleagues. This is especially common in smaller fields, being much more practical. Some institutes will have a collection of reprints. |
| Activity | Gather specimens | This broad activity refers to the sourcing of existing specimens related to the group under study. |
| Action | Identify existing specimens | First the existing specimens need to be identified and located. Sources of specimen information include:<br>- Publications will usually indicate the location of at least the type specimens used in the paper.<br>- Online search catalogues such as fishbase<br>- General search engines such as Google<br>- Personal knowledge of a collector's career and the institutes they worked for<br>- Colleagues |

| | | |
|---|---|---|
| | | Specimens may not always be available however. Some will not be available for loan due to fragility or other reasons. Others may simply be lost. |
| Action | Gather existing specimens | Once identified, the taxonomist needs to physically examine the specimens. This can either be done by travelling to the institute housing the collection, as is often the case where the specimen can not be sent out, or more commonly, requesting to loan the specimen.<br><br>Institutional loan policy varies, though all will have some procedure for receiving and assessing loans, then processing the loan request. Differences may include ;<br>- Charging. Many institutes will send out loans for free, some need to apply a charge for this service.<br>- Assessing the loan. Institutes ask for different levels of information regarding the loan; details of the study, past work, references, etc.<br><br>Loaning is generally a lengthy process - typically it takes several months to receive specimens from request. This turn-around time is widely acknowledged throughout taxonomy, but it is not seen as an a real problem, just a fact of life. Other work can always be undertaken whilst waiting for specimens.<br><br>One can also submit blanket requests for all specimens of a particular taxa, or for all unidentified material that is thought to belong to a taxa. These requests obviously involve more curatorial work.<br><br>Specimens can also be found in private collections. Arrangements to view such specimens will be particular to the case.<br><br>Many scientists will use existing travel arrangements as an opportunity to visit other collections, and examine any specimens they need to. |
| Action | Quarantine specimens | Whether arriving from the field or from other collections, almost all institutes quarantine incoming specimens for a time before allowing their release into the collection. Quarantine usually involves storing the specimens in cold conditions for a number of days, in order to kill off any organisms that may have a harmful effect on the collection. Exceptions to this can include alcohol-stored specimens, which are judged to be sterile. |

| | | |
|---|---|---|
| Activity | Examine specimens | As with collecting activities, examination techniques depend entirely on the nature of the specimen, and will differ according to taxa. Examination tends to be an iterative process, with the focus becoming more detailed as the work continues. |
| Action | Prepare specimens | This refers to any methods used to prepare the specimens for examination, and commonly involves the preparations of microscope slides and the dissection of the specimens. |
| Action | First visual examination | Almost all examinations begin with an initial visual assessment of the specimens, before any other examinations take place. The goal is to acquire a familiarity with the specimens and identify the main characters. |
| Action | Sort specimens | Physically sorting specimens into groups is a common practise. The process of sorting helps to highlight the differences and similarities between specimens. Viewing the specimens in their proposed group serves to bring the emerging taxonomic hypotheses into sharper focus. |
| Action | Detailed visual examination | The process will now move onto more detailed examination, almost always using a microscope of some sort. This action examines the finer morphological features of the specimens, allowing further assessment of the emerging theory. Standard light microscopy is extremely common, with florescence and stereo microscopy also commonly available. Dissection may be performed to examine internal structures.<br><br>The recording of results is very individualised; pen & paper, spreadsheets, and statistical software all serve as first points of entry for measurements. Some do not record results at all until the theory is complete. Images are often taken at this stage, commonly without assistance. Many taxonomists prepare their own illustrations, a process that all regard as being helpful to the examination process itself.<br><br>That nature of the measurements taken will be particular to the group. Morphological features, however, are almost ubiquitous in taxonomy. |
| Decision | Further analysis? | Aside from light microscopy there are a wide range of other examination techniques potentially available. Whether further techniques are employed here, and which ones, will depend on the nature of the group |

| | | |
|---|---|---|
| | | under study, the availability of the technique (and hence the resources available to the taxonomist), the level of certainty about the emerging theory, and the personal preference of the taxonomist. |
| Action | Further analysis | Other examinations include:<br>- S.E.M.<br>- T.E.M.<br>- Chemical composition analysis<br>- DNA analysis<br>- Phylogenetic analysis<br><br>The examinations may be performed by the taxonomist, or by colleagues. The strength assigned to the various results differs between scientists. Phylogenetic analysis is a good example of this - some taxonomists use this as the basis of a theory, some to back-up a developed theory, others not at all. |
| Decision | Assess emerging theory? | It is almost universally agreed that the development of the taxonomic theory and the examination are not separate events. The taxonomic theory develops with the examination, from the initial sorting of specimens through to the more detailed analyses. At some point though, a decision is made that the theory is complete and that no further examination is needed. |
| Activity | Apply nomenclatural rules | This represents the separate stage of applying the rules of nomenclature to the new specimen groups. The rules are stipulated by the ICBN (http://www.bgbm.org/iapt/nomenclature/code/SaintLouis/0000St.Luistitle.htm) or ICZN (http://www.iczn.org/iczn/index.jsp) |
| Action | Resolve type specimens | Type specimens are assigned to each specimen group. The process by which this is done can be complex, and is described in full at the above web addresses. |
| Action | Resolve nomenclature | Nomenclature is determined for each specimens group. In both zoology and botany, the guiding principle of the nomenclatural codes is priority. Again, the full process is too detailed to describe here, but all the information can be found at the ICBN and ICZN websites. |
| Activity | Prepare paper | The process of compiling a scientific paper and arranging for publication. |
| Action | Write manuscript | Prepare the various sections of the scientific paper, and compile according to the editorial guidelines of the intended journal. Typical sections include: |

| | | |
|---|---|---|
| | | - Taxonomic treatment. The basis of a revision.<br>- Distribution maps<br>- Comparison tables summarising main features<br>- A taxonomic key<br>- A phylogenetic tree / cladogram and it's data matrix<br>- A discussion of previous work<br>- A discussion of the main findings and any other related work<br>- Graphs and tables illustrating other findings<br>- Photographic Images of the specimens, usually prepared, occasionally in the wild<br>- Illustrations indication the main features<br>- References and a bibliography<br><br>The various sections will be prepared using the appropriate software, or occasionally manually; photo-plates for example. Almost all taxonomists compile the manuscript using MS Word. |
| Action | Friendly review | This is an informal review of the manuscript by colleagues, arranged to gather comment on the paper before submission to a journal. |
| Action | Revise paper | Revise the paper in the light of comments and suggestions. |
| Action | Submit to journal | The manuscript is submitted to the intended journal, usually by email, again in accordance with the editorial guidelines. |
| Action | Paper accepted? | This action is of course external to the taxonomist's work process, but important as it's results will affect the direction of the project. There are 4 possibilities:<br>- Accepted outright. It is relatively uncommon for a paper to be accepted entirely without revision.<br>- Minor revision. The paper is accepted subject to minor revisions. These can be presentational or concerned with the subject matter.<br>- Major revision. Significant changes are suggested. These may be related to the findings or other key aspects of the paper. The journal may also feel that the paper needs a different approach to fit within it's subject boundaries.<br>- Rejected outright. Also relatively uncommon, and can be related to suitability to a particular journal, or simply the quality of the paper |
| Activity | Curation activities | Many taxonomists have some level of curatorial |

| | | |
|---|---|---|
| | | responsibility in their institutes. Those that do not will still need to prepare specimens for storage and arrange for the return of loans. Curatorial work is very often performed by an assistant, with guidance from the taxonomist. |
| Action | Label specimens | All specimens need to be labelled prior to storage. This is the case for new specimens, and those subject to changes under the revision. Labels are often printed using a variety of software applications, or may be hand-written. |
| Action | Return loaned specimens/send out duplicates | Often the conventions of a collaborative project, especially one involving collecting activities abroad, will dictate that types or paratypes are sent to the collaborating institute for permanent storage there. Loans will also need to be returned. This will be possible by post, delivery by courier, or by a visit to the institute.<br><br>This step can also take many months to complete, especially if waiting for opportune travel arrangements. |
| Action | Place specimen in local collection | The newly labelled specimens is placed in the institute's collection for permanent storage. |
| Action | Update collection database | All collections will have some sort of database, be this an electronic database or a printed directory. This will need to be updated with the new information. |

# Discussion

## Bottlenecks

In addition to producing the process models, it was also decided to attempt to identify areas of the revisionary process which might particularly benefit from the introduction of online resources; activities which were seen to be an inefficient use of specialist resources, or those that involved unnecessary delay. These activities were termed *bottlenecks*.

I approached this in two ways. First, direct, open questioning as to which areas the taxonomists themselves considered bottlenecks. From this direct questioning, the following activities were explicitly mentioned as being particularly time consuming:

1) Gathering type specimens
2) Gathering literature
3) Entering label data
4) Preparing illustrations
5) Preparing plates

From this, it can be seen that the bottlenecks fall into two categories; activities that involve delay, and activities that simply take time. The distinction may be important in the future when considering whether it is possible to facilitate these activities through the cyberplatform.

From the five processes identified above, activities 3, 4 and 5 are by their nature time-consuming. Entering label data, preparing illustrations and preparing plates are tasks for the individual, and take hours to complete. All of these can be assisted by computerised resources. Labels can be printed, and in bulk, direct from a collections database. Illustrations can be produced with the assistance of a scanned image, and the subsequent measurements taken by computer. The inclusion of an illustration in a scientific paper can also be aided using computer applications. Photo-plates can also be produced digitally, indeed it is rarer now for these to be physically cut and pasted together, though the practise still occurs.

Activities 1 and 2 however involve waiting for other events to occur; i.e. the arrival of loan material or literature. Loans applications in particular can sometimes take months to be completed. The reasons for this have not been investigated in depth, but the comments of taxonomists (who are often also curators) reveal some possible reasons:

- The processing of specimens for loans is a physical activity; specimens must be retrieved from the collection, packaged and sent, and this requires the time of the collection curator, who will have many other duties.
- In some cases specimens will be out on loan and must be recalled.
- Large loan requests may be delivered in stages to reduce risk

Generally, the delay in processing loan applications tended to be accepted as the status-quo and unavoidable. Having conducted no detailed investigation into this I would not like to pass an opinion on whether this is the case, but it is certainly true that a quicker turn around for this process could reduce the duration of a project by several months.

The gathering of literature was also seen as a lengthy process. This again was attributed to the time taken for the literature to be sent as an inter-library loan. Perhaps this activity is the most obvious for online assistance; scanned papers can be sent instantly. The problem is that not all

literature exists in digital format, though I am aware that there are several projects already in operation to improve this.

## The Questionnaire

This 'soft' information was easy to gather, and though intuitively correct was still subjective. To complement this a second approach was attempted, in the form of a questionnaire designed to gather quantitative data on the time taken up by the various activities. This was first attempted during the interviews themselves, and proved a difficult exercise. The necessity of establishing the model structure first and foremost meant that questionnaire had to be given a lower priority in the limited time available for the interviews. Placing precise figures on activities which may be fragmented over weeks or months required time that was often not available. As well as this a further complication was that during the early interviews the merged model has not yet been created, and the time estimates were related to activities that may not have been suitable for a combined analysis at a later date. For these reasons it was decided to circulate the questionnaire as a separate exercise after the interviews were finished, and the merged model was in a revised and stable form.

After the interviews were complete, the questionnaire was re-designed to reflect the merged model. The specific figures sought were those of *time taken* versus *elapsed time* for each of the activities described. Time taken refers to the amount of time, in days, that the taxonomist spent working on the activity. Elapsed time refers to the time that passed while the activity was taking place. So for example, it might take 1 days work to request a specimen loan, but 40 days for the loan request to be completed. Using these, we would be able to see quantitatively which activities involved delays, and which were the most efficient.

Rather than attempt to calculate average or typical figures, the taxonomists were asked to choose a recent project and use time estimates from this.

The questionnaire is reproduced in Appendix 2.

### Questionnaire results

At the time of writing, the full set of questionnaire results have not been submitted, so the result printed here are interim figures. These figures are a compilation of the responses from four different taxonomists. The full results will be presented on the EDIT website in due course. From the results in so far, the following calculations have been made :

1) Average figures for each activity
2) From the set of averages, the ratio of working time divided by elapsed time

The results below have been divided into activity and sub-activity, then ordered by ratio. All figures are measured in days :

| Activity | Sub Activity | Average Time taken | Average Elapsed time | Ratio |
|---|---|---|---|---|
| Select sub-group | | 10.5 | 320 | 0.03 |
| Curation activities | | 27 | 570 | 0.05 |
| Gather specimens | | 30 | 245 | 0.12 |
| Search Literature | | 34 | 255 | 0.13 |
| Apply nomenclatural rules | | 50 | 260 | 0.19 |
| Examination and grouping | | 155.5 | 555 | 0.28 |
| Collecting activities | | 47 | 127.5 | 0.37 |
| Prepare paper | | 185.5 | 266 | 0.7 |
| | | | | |
| | Update collection database | 5.5 | 245 | 0.02 |
| | Arrange permits | 1.5 | 45 | 0.03 |
| | Label specimens | 10.5 | 245 | 0.04 |
| | Receive and quarantine specimens | 10 | 145 | 0.07 |
| | Transport specimens home | 3 | 32.5 | 0.09 |
| | Identify literature | 12.5 | 125 | 0.1 |
| | Gather specimens (loan or visit) | 12.5 | 122.5 | 0.1 |
| | Return loans/send duplicates | 5.5 | 45 | 0.12 |
| | Friendly review | 5 | 40 | 0.13 |
| | Place specimen in local collection | 5.5 | 35 | 0.16 |
| | Gather literature | 21.5 | 130 | 0.17 |
| | Resolve type specimens | 22.5 | 130 | 0.17 |
| | Resolve nomenclature | 27.5 | 130 | 0.21 |
| | First visual examination | 30.5 | 122.5 | 0.25 |
| | Sort specimens | 32.5 | 130 | 0.25 |
| | Prepare specimens for examination | 2.5 | 10 | 0.25 |
| | Submit to journal | 1.5 | 6 | 0.25 |
| | Further analysis | 35 | 140 | 0.25 |
| | Detailed visual examination | 55 | 130 | 0.42 |
| | Identify existing specimens | 7.5 | 17.5 | 0.43 |
| | Manuscript reviewed | 31.5 | 55 | 0.57 |
| | Revise paper | 12.5 | 20 | 0.63 |
| | Conduct field collection | 42.5 | 50 | 0.85 |

From these figures we can draw a measure of the efficiency of the various activities. Activities at the head of the two lists represent those which involve the most delay, i.e. waiting for other events. These would seem to be most open to improvement in terms of efficiency.

It also makes sense however to focus efforts on those activities in which improving efficiency would generate the greatest benefit, that is, those activities which take the longest to complete. For example, it would be more important to target a moderately efficient activity lasting 100 days than one which is very inefficient but takes only a day to complete. Even if the latter could be completely automated, the saving in terms of time could never be greater than a single day; the same time could be saved from a 100 day activity by reducing the time taken by just 1%.

To this end I have re-ordered the above figures by elapsed time:

| Activity | Sub Activity | Average Time taken | Average Elapsed time | Ratio |
|---|---|---|---|---|
| Curation activities | | 27 | 570 | 0.05 |
| Examination and grouping | | 155.5 | 555 | 0.28 |
| Select sub-group | | 10.5 | 320 | 0.03 |
| Prepare paper | | 185.5 | 266 | 0.7 |
| Apply nomenclatural rules | | 50 | 260 | 0.19 |
| Search Literature | | 34 | 255 | 0.13 |
| Gather specimens | | 30 | 245 | 0.12 |
| Collecting activities | | 47 | 127.5 | 0.37 |
| | Label specimens | 10.5 | 245 | 0.04 |
| | Update collection database | 5.5 | 245 | 0.02 |
| | Write manuscript | 135 | 145 | 0.93 |
| | Receive and quarantine specimens | 10 | 145 | 0.07 |
| | Further analysis | 35 | 140 | 0.25 |
| | Sort specimens | 32.5 | 130 | 0.25 |
| | Gather literature | 21.5 | 130 | 0.17 |
| | Detailed visual examination | 55 | 130 | 0.42 |
| | Resolve nomenclature | 27.5 | 130 | 0.21 |
| | Resolve type specimens | 22.5 | 130 | 0.17 |
| | Identify literature | 12.5 | 125 | 0.1 |
| | First visual examination | 30.5 | 122.5 | 0.25 |
| | Gather specimens (loan or visit) | 12.5 | 122.5 | 0.1 |
| | Manuscript reviewed | 31.5 | 55 | 0.57 |
| | Conduct field collection | 42.5 | 50 | 0.85 |
| | Return loans/send duplicates | 5.5 | 45 | 0.12 |
| | Arrange permits | 1.5 | 45 | 0.03 |
| | Friendly review | 5 | 40 | 0.13 |
| | Place specimen in local collection | 5.5 | 35 | 0.16 |
| | Transport specimens home | 3 | 32.5 | 0.09 |
| | Revise paper | 12.5 | 20 | 0.63 |
| | Identify existing specimens | 7.5 | 17.5 | 0.43 |
| | Prepare specimens for examination | 2.5 | 10 | 0.25 |

There are a number of ways to look at this data. To address the first point, are there activities that are both time-consuming (in terms of elapsed time) and also have a low ratio of time taken to elapsed time as calculated above? Comparing the two sets of figures it can be seen that the labelling of specimens, and the receiving and quarantining of specimens all feature highly on both lists. It might be remarked here that labelling and loaning of specimens were both explicitly mentioned by taxonomists when invited to volunteer activities they considered bottlenecks.

Second, why is there such a large difference between the figures for time taken and those for elapsed time? I would suggest two explanations for this. The first is as mentioned above; that these activities are subject to external delays over which the taxonomist has no control. It would seem likely that a number of activities fall under this category; the loaning of specimens, arranging for permits, the quarantine of specimens. A second explanation is that the activity was not externally delayed, but broken up over a longer time period for practical reasons. It has already been explained that in boxing activities as discrete events, the models simplify real-life. Activities can be fragmented over a longer time period simply because they are approached in a

piecemeal fashion for convenience, or because their component tasks are addressed as they arise, rather than postponed and dealt with *en masse*. Specimens may be labelled individually once their examination is complete, for example, or parts of the manuscript updated as the project continues.

Obviously, not all of the activities described here will be suitable for assistance through online resources. It is difficult to see for example, how the procedure for processing specimen loan requests can be addressed by the cyberplatform, especially when most loan requests are processed electronically as it is. It may not be possible to provide online assistance in all cases. In addition, it is recognised that the figures and analysis described here are necessarily limited. However, I hope these figures can serve as a starting point for debate and an indication as to where the efforts of EDIT may be best placed.

## Conclusions

The modelling work has been an interesting exercise, and though it has revealed much about the practise of taxonomy in Europe, there are clearly areas that should be subject to further investigation. The limitations on this round of information gathering has meant that the data gathered on bottlenecks has been necessarily brief, and it further quantitative research here may be of value. Nevertheless, clear trends have emerged, and common difficulties identified. To my mind, this suggests that measures to improve practise in these areas should be also be implemented at a high level. The acquisition of specimens and literature stand out as areas where improvements could be made.

Another area of interest is that of the non-professional communities. Some branches of science seem to have much more involvement from the amateur taxonomist than others, often related to the ease with which specimens can be collected or observed. Although I was unable to interview any amateur taxonomists as part of this investigation, I became aware that in many branches of taxonomy they are considered a valuable resource who can provide significant resources and expertise, and of course, are as free and able to publish recognised revisions as the paid professional. Those scientists who maintained close relationships with the amateur community almost invariably benefited from this, and were able to pass on their skills and knowledge of good scientific practise. I would certainly recommend that the amateur community is not left out from the continuing work of EDIT, and have been pleased to see that there has been involvement in some areas already.

That it was possible to develop of a unified model for revisionary taxonomy is encouraging for the goals of workpackage 5. Though the practise of taxonomy differs between individuals, institutes, and the many branches of the science, it can be seen that there is a wider framework within which the common activities fit. As a result, it would seem that the development of a platform of tools and resources that can be used by the whole taxonomic community is an attainable goal. The identification of the key components of revisionary work lends itself well to the modular nature of modern software platforms, which is a key feature of the vision for the internet platform for cybertaxonomy.

This investigation has shown that many of the difficulties faced by scientists in this field are common to the wider community, and as such, need to be addressed with common solutions. Taxonomists across Europe benefit greatly from involvement with the wider taxonomic community, and the goals of integration promoted by the EDIT project can only lead to further improvements and the practise of a more streamlined science.

David Taylor, July 2007

## Acknowledgements

The interviews I conducted for this work were not a trivial exercise. Each interview took up at least two half-days, along with a follow-up questionnaire, and prior arrangements. As well as this, I am aware that the process of recalling and recording the minutiae of one's work can be an arduous one. All of the scientists I spoke to had taken time out from their many other duties as taxonomists, teachers, editors and departmental heads, and I would very much like to thank all of them for the time and effort they volunteered. Without their efforts, and the positive and engaging attitude that all displayed, this exercise would not have been possible. Certainly I feel that this part of the EDIT project has achieved it's goals, and I hope it was a worthwhile experience for all concerned.

## Appendix 1. Criteria for selecting candidate for interview:

1) A representative geographical coverage of the EDIT partnership

2) Taxonomic group (at least 2 interviews for each)
    a. Vertebrates
    b. Entomology
    c. Terrestrial invertebrates
    d. Marine invertebrates
    e. Plants
    f. Fungi
    g. Lichens and moss
    h. Algae
    i. Palaeontology
    j. Parasitology

3) Environment of taxonomic groups
    a. Marine
    b. Terrestrial
    c. Soil
    d. Freshwater

4) Techniques used
    a. Traditional
    b. Molecular
    c. Numerical

5) Institution/ Department size
    a. Large
    b. Small

## Appendix 2. Institutes visited for interview

| Institute | Specialism | Date of interviews |
|---|---|---|
| Royal Botanic Gardens, Kew | Botany | October 2006 |
| NHM, London | Entomology | October 2006 |
| HNHM, Budapest | Phycology, Paeleobotany | October 2006 |
| Zoological Museum, University of Copenhagen | Arachnology, Polychaetes | November 2006 |
| Botanical Museum, University of Copenhagen | Mycology | November 2006 |
| Royal Botanic Gardens, Edinburgh | Botany | January 2007 |
| National Museum of Natural History (Naturalis), Leiden | Paleo-malacology | February 2007 |
| Zoological Museum, University of Amsterdam | Ichthyologist | February 2007 |
| Comenius University, Bratislava | Acarology | March 2007 |
| Slovakian Academy of Sciences, Bratislava | Lichenology | March 2007 |
| Stuttgart Natural History Museum | Palaeontology | March 2007 |
| MIIZ, Warsaw | Entomology | April 2007 |
| Royal Museum for Central Africa, Tervuren | Ornithology, Arachnology | May 2007 |
| ZINRAS, St.Petersburg | Nematodes | June 2007 |

# Appendix 3. The Questionnaire

| Activity | Sub-activity | Working time | Elapsed time |
|---|---|---|---|
| Collecting activities / Fieldwork | Arrange permits<br>Conduct field collection<br>Transport specimens home | | |
| Select sub-group | | | |
| Search Literature | Identify literature<br>Gather literature | | |
| Gather specimens | Identify existing specimens<br>Gather specimens (loan or visit)<br>Receive and quarantine specimens | | |
| Examination and grouping | Prepare specimens for examination<br>First visual examination<br>Sort specimens<br>Detailed visual examination<br>Further analysis | | |
| Apply nomenclatural rules | Resolve type specimens<br>Resolve nomenclature | | |
| Prepare paper | Write manuscript<br>Friendly review<br>Revise paper<br>Submit to journal<br>Manuscript reviewed | | |
| Curation activities | Label specimens<br>Return loans/send duplicates<br>Place specimen in local collection<br>Update collection database | | |

# Appendix 4. Software

As separate efforts to address the use of software in taxonomy are underway in workpackage 5, the list below makes no particular analysis of software use but does provide a indication of the range of software systems currently being utilised. I have attempted to related the interview findings to the software categories proposed by WP5. Below is a list of all software encountered during the interviews, along with the model activity during which it is used, and the WP5 category to which it belongs. I have also included any web-sites used. Web resources seem to be at least as important as software applications for the revisionary process.

## WP5 Software categories :

1 - Bibliographic
2 - Geographical
3 - Taxonomic
4 - Descriptive
5 - Communication
6 - Publication

7 - Image processing
8 - Fieldwork
9 - Phylogenetic
10 - Specimen access
11 - Nomenclature

## Main Activities identified from models :

Collecting activities
Search Literature
Gather existing specimens
Examine specimens
Prepare paper
Curation activities

**List of all tools used against software categories and main activities:**

| Software | Use | Main Software category | Main Activities | Model |
|---|---|---|---|---|
| Adobe Illustrator | Assembling plates | 7 | Prepare paper | 2 |
| Adobe Photoshop | Preparing illustrations and photographs | 7 | Prepare paper | 2, 4, 7, 10, 13 |
| Arc Explorer | Distribution maps | 2, 6 | Prepare paper | 13 |
| Arc View | Distribution maps | 2, 6 | Prepare paper | 13 |
| BG-BASE | Database | 10 | Curation activities | 9 |
| BIOTA | SQL database used to store images and data. Links images to species, specimen, collection etc. Allows collaborative work over the internet. Can function as a specimen management tool. | 10, 5 | Examine specimens | 5 |
| CANOCO | Statistical analysis | 7 | Examine specimens | 4 |
| Coral Draw | Preparing illustrations and photographs | 7 | Prepare paper, Examine specimens | 4, 6, 14 |
| DELTA | Descriptive Language for Taxonomy. Produces computer generated descriptions and keys. | 4 | Examine specimens | 5 |
| DIVA | Distribution maps | 2, 6 | Prepare paper | 9 |
| EndNote | Bibliographies and references | 1 | Prepare paper | 1 |
| ESRI ArcView GIS 3.2 | Maps | 2, 6 | Prepare paper | 2 |
| FileMaker DB | Label Printing | 11 | Curation activities | 12 |
| Filemaker Pro | Maps | 2, 6 | Prepare paper | 2 |
| Google earth | Distribution maps | 2, 6 | Prepare paper | 9 |
| Henning86 | Phylogenetic analysis | 9 | Examine specimens | 2 |
| Illustrator | Mounting plates; Making line illustrations | 2, 6 | Prepare paper | 10 |
| Image Pro Plus | Preparing and analysing images | 2, 6 | Examine specimens, Prepare paper | 3 |
| IMATCH | Image management tool. Stores digital images with metadata, allows searching | 7 | Prepare paper | 5 |

| | | | | |
|---|---|---|---|---|
| | on keyword. Provides basic image manipulation such as contrast adjustment. | | | |
| intkey | Reading interactive keys | 3 ? | Examine specimens | 9 |
| LUCID | Compiles interactive keys | 3 ? | Examine specimens | 9 |
| Maclade | Stores character measurements and produces nexus file of data | 9 | Examine specimens | 6, 12 |
| MESQUITE | Stores character feature in a data matrix. Provides phylogenetic and multivariate analysis. Uses a range of methods such as parsimony, maximum likelihood etc | 9 | Examine specimens | 5 |
| MS Access | Collection and specimen management, label printing | 10 | Prepare paper, Curation activities | 1, 6, 10 |
| MS Excel | Comparison tables, graphs, personal database | Used in most activities | Used in most activities | All |
| MS Word | Preparing and compiling the final paper and other word processing | Used in most activities | Used in most activities | All |
| PADME | In house collection database (MS Access) | 10 | Curation activities | 8, 9 |
| Paint Shop Pro | Preparation and manipulation of images | 7 | Examine specimens | 3 |
| Paradox DB | Database for storing information on species, specimens, locations and literature. | 10 | Curation activities | 13 |
| PAST | Statistical analysis | 6 ? | Examine specimens | 10 |
| Paup | Performs the actual PG analysis | 9 | Examine specimens | 6 |
| PAUP | Phylogenetic analysis | 9 | Examine specimens | 12 |
| Powerpoint | Presentations | 5 | | 9 |
| SPOT | Preparation and manipulation of images | 7 | Prepare paper | 3 |
| SPSS | Statistical analysis application | 6 ? | Examine specimens | 12 |
| Syntax 2000 | Statistical analysis | 6 ? | Examine specimens | 4 |
| WinClada | Setting up cladogram | 9 | Examine specimens | 2 |

## Links between Main Activities and WP5 Software Categories

| Main activities | WP5 Software Categories |
|---|---|
| Examine Specimens | Specimen access,<br>Communication,<br>Image processing,<br>Descriptive,<br>Phylogenetic,<br>Geographical,<br>Publication,<br>Taxonomic? |
| Prepare paper | Image processing,<br>Geographical,<br>Publication,<br>Bibliographic,<br>Specimen access |
| Curation activities | Specimen access,<br>Nomenclature |

## List of Main Activities against major websites

| Main activities | Websites used |
|---|---|
| Search Literature,<br>general background information | Google |
| Search Literature | GEO-REF |
| Search Literature | The Zoological Record |
| Search Literature | Google Scholar |
| Search Literature,<br>general background information | ISI Web of Knowledge |
| Search Literature,<br>Gather Specimens,<br>general background | Index Herbarium |
| Search Literature,<br>Gather Specimens | Index to Organism Names (ION) |
| Search Literature,<br>general background | ZooTaxa |
| Search Literature,<br>Gather Specimens | Fishbase |
| Search Literature,<br>Gather Specimens | World Wide Catalogue of Spiders |
| Search Literature,<br>Gather Specimens | BioSystematik Database of World Diptera |
| Search Literature,<br>Gather Specimens | Index Fungorum |
| Search Literature,<br>Gather Specimens | Sylloge Fungarum |
| Search Literature,<br>Gather Specimens | IPNI |