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"

WP 5.2.1.1 Draft functional model and bottleneck report for rev. taxonomy

Due date of deliverable: Month 09
Actual submission date: Month 13

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

Organisation name of lead contractor for this deliverable: 11 RBGK

Revision [draft]

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

Introduction

As of March 2007, 14 models have been produced from interview, eight of which have been reviewed and posted on the EDIT wiki. The interviews have been targeted towards the criteria for selecting candidates, as agreed at the modeller's meeting in Berlin during September 2006. These were; geographical coverage, taxonomic group (2 interviews each), environment of taxonomic groups, techniques used, and size of organisation. Details of these criteria appear in appendix 1. So far, progress is as follows:

- Six out of the eleven European countries making up EDIT (including Russia) have been covered
- Eight out of nine taxonomic groups have been covered at least once
- Three out of four of the specified environments have been covered
- All three working techniques have been covered
- Both large and small institutions / departments are believed to have been covered (though some more clarity on the dividing line between the two would be helpful)

At least 4 more interviewing trips are being arranged at the moment. Once complete, these interviews should cover all of the target criteria and complete the interviewing requirements of this stage

Main findings so far

The interviews have indicated that between the different disciplines, methodologies and institutions, there exists a common framework for revisionary taxonomy. In producing the draft unified model I have included only those practises that have been observed in all the interviews so far. As the modelling phase has progressed and common activities have become clear, the form of the unified model has become apparent, and the later models are based around this. Earlier models may therefore diverge from this common form.

Outside of the common framework depicted in the unified model, methodologies differ between individuals and taxonomic group. This is largely to be expected; examination of specimens for example, will involve very different techniques between fields. The unified model and accompanying information appear later in this report.

The combined approach to modelling

Process models of real-life activities have a number of advantages over textual descriptions. Models illustrate the broad movement of work flow much more clearly than text, and in a much more digestible format. Individual steps and activities are easy to establish, and the relationships between the activities are clear at a glance. This clarity is fundamental to the usefulness of the resulting models.

Text, on the other hand, allows for the recording of much more detail than can ever be usefully expressed in a model. The various activities can be described in much more depth. A balance has therefore been struck between producing clear, understandable models, while at the same time presenting all the information gathered in the interviews. Through discussion and agreement with the interviewees, it was not felt necessary to always record every conceivable possibility of work-

flow, providing the resulting model accurately represented the general procedural flow of a taxonomic revision.

The information gathered in each interview has been represented with both a process diagram and an accompanying textual document, which, alongside the model, provides a full, detailed picture of the work processes under study.

Identifying bottlenecks

I addressed the identification of bottlenecks in two ways. Firstly, direct questioning about which areas of their work taxonomists themselves regarded as bottle-necks, defined as activities which are either time-consuming, or cause delay to their work. The twin definition is relevant, as certain activities may be time-consuming but not seen as bottlenecks, either because a swifter method cannot be envisaged, or because the activity is performed by someone else. A good example of this is the acquisition of loan specimens, which though sometimes taking months to complete, is often not seen as a bottleneck. The activity also serves as a useful example of another tendency I observed, which was to accept a process as the status-quo, simply the way things are, and so not to consider it a bottleneck worth reporting. I made a particular effort to encourage interviewees to put aside this tendency and talk freely.

Secondly, I attempted a more quantitative exercise; asking taxonomists to calculate for each activity, both the actual time elapsed between starting and completing the activity, and the amount of work required by the taxonomist for each activity. I had hoped this would provide definite figures for activity time which we could use for further analysis. Unfortunately, placing even estimates on activities proved to be a difficult exercise. Often the precise duration of an activity was difficult to put a figure to. In addition, the necessity of establishing the structure of the model first and foremost meant that this exercise had to be given a lower priority in the limited time available for interview. As a result, the data gathered was insufficient to draw any general conclusions. Further information gathering efforts may be more successful in collecting data on this topic (see *Further investigations*).

So, from direct questioning in the interviews, the following activities were explicitly mentioned as particularly time consuming:

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

I would suggest that two groups can be observed here. First, activities which take up time because the activity simply requires a lot of work. Bottlenecks 3, 4 and 5 can be seen to fall into this category. These tasks are time-consuming in their very nature. The second category contains activities which involve waiting for something, usually the completion of an external activity. Bottlenecks 1 and 2 sit largely in this category, involving waiting for requests to be processed by other departments and institutes. I see no reason why it would not be possible to streamline both types of activity with internet-based tools, but the distinction may be important when considering how to proceed.

Further investigations

The interview-based approach has been very useful for gathering high-level information about the work of revisionary taxonomists. The scope of the subject matter has been broad, encompassing all activities related to revisionary work. Gathering such a wide range of information required an intensive, flexible approach that could only be achieved through interview. The production of process models also required a face to face approach. There are, however, natural limitations to the interviewing approach:

- **Interviewing time is limited.** It was suggested in the modeller's meeting of September 2006 that two half-days would be the maximum time we could reasonably expect a taxonomist to commit, and this has proved to be the case. There is only so much detail that can be gathered in this time, and though follow-up by email and phone can be useful to clarify details, it is not reasonable to expect much further commitment from the taxonomists after this time.
- **Omissions will occur.** Even in the most focussed of interviews detailed will be missed.
- **Sample size is limited.** Interviewing is time-consuming. Only a relatively small number of taxonomists can contribute.

As the interviews have continued, areas of taxonomic work have been highlighted which would benefit both from a more targeted investigation, and from exposure to a much wider audience of taxonomists. These are either areas where a complete list of specific information is sought, or where it is difficult to gather information through the retrospective view of an interview. I have identified three such areas:

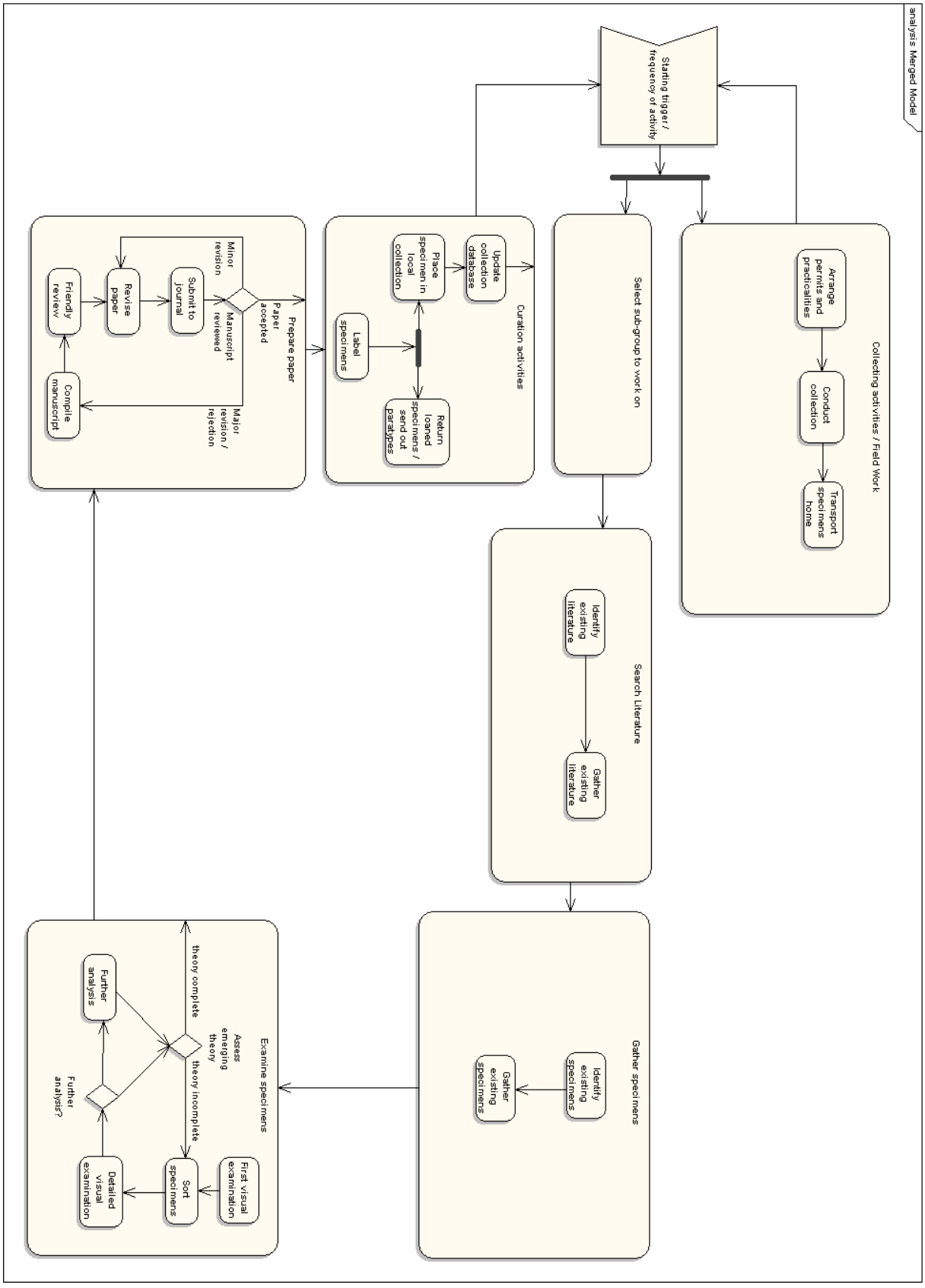
1) Software applications
2) Websites
3) Bottlenecks

Data on software applications and websites takes the form of a list (see tables 1 and 3 in this report). This type of information can easily be compiled independently of an interview. The difficulties collecting retrospective information on bottlenecks suggests that an ongoing monitoring exercise may have more success in this area – for example by taxonomists keeping some sort of simple log of their activities.

I would therefore recommend that a second round of more focussed information gathering takes place once the interview stage is complete. The goal of this round should be to focus on these three specific areas, and any others that may be considered appropriate for further investigation. The method of data collection needs to be practical, simple and quick to complete, and the compilation and analysis of the results should be straight forward. This format may suggest a short online questionnaire, consisting of closed questions and lists, which could be distributed amongst the EDIT institutes. With regard to bottlenecks, a simple online logging facility could be provided to allow taxonomists to monitor their own activities. In this way we could gather a large body of more focussed information, from a much wider audience, in a relatively short time, and in a format much more suitable for analysis.

David Taylor, March 2007

The draft unified model



Detail of the draft unified model

The table below covers the activities of the draft unified model. All the activities referred to of course generalizations drawn from all previous models so far. Note that the model is divided into two paths, separating out collecting activities from the other processes. It is difficult to place the collecting process in the linear flow with any consensus; the logistics of a collecting trip often place it naturally outside a step-by-step process, and collections can and do take place independently of a revisionary projects. For a unified model, collecting activities are best represented as a parallel activity.

Event	Starting trigger	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: - Collecting specific taxa as part of a project - General collecting to explore a poorly understood area - To improve the institute's collection - Collaboration with a partner institute - In-keeping with institutional goals - Gathering material for teaching - 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: - To enter the collecting site at all. Whatever body maintains the collecting may restrict access, for safety or other reasons. - To remove material from the site. - CITES restrictions may apply to certain taxa 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	Select a sub-group to work on	This activity represents the process of breaking down a large body of work into manageable chunks. Many revisions will involve the study of a large number of taxa, and potentially many more specimens. Large projects generally have to be approached piecemeal. The sub-groups are generally chosen on the basis of taxa or geography, or perhaps a combination of the two.
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: - Online search engines both subject specific, such as IPNI or GEO-REF, or general search engines like Google. - Personal knowledge of the field - Citations in other works. In this way one can follow a 'trail' back to the proto-log - Library search catalogues - Colleagues - Amateur / enthusiast communities - 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. If not available online, an inter-library loan can normally be arranged. 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:

		- Publications will usually indicate the location of at least the type specimens used in the paper. - Online search catalogues such as fishbase - General search engines such as Google - Personal knowledge of a collector's career and the institutes they worked for - 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. Institutional loan policy varies, though all will have some procedure for receiving and assessing loans, then processing the loan request. Differences may include ; - Charging. Many institutes will send out loans for free, some need to apply a charge for this service. - Assessing the loan. Institutes ask for different levels of information regarding the loan; details of the study, past work, references, etc. 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. 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. Specimens can also be found in private collections. Arrangements to view such specimens will be particular to the case. Many scientists will use existing travel arrangements as an opportunity to visit other collections, and examine any specimens they need to.
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	First visual examination	Almost all examinations begin with an initial visual assessment of the specimens, before any other examinations take place. Initial thoughts are formed as to the broad taxonomic grouping.
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. 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. 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: - S.E.M. - T.E.M. - Chemical composition analysis - DNA analysis - Phylogenetic analysis 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	Prepare paper	The process of compiling a scientific paper and arranging for publication.
Action	Compile 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. - Distribution maps - Comparison tables summarising main features - A taxonomic key - A phylogenetic tree / cladogram and it's data matrix - A discussion of previous work - A discussion of the main findings and any other related work - Graphs and tables illustrating other findings - Photographic Images of the specimens, usually prepared, occasionally in the wild - Illustrations indication the main features - References and a bibliography 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: - Accepted outright. It is relatively uncommon for a paper to be accepted entirely without revision. - Minor revision. The paper is accepted subject to minor revisions. These can be presentational or concerned with the subject matter. - 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. - 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 paratypes	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. 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.

Adjuncts to the unified model

I have attempted to related the interview findings to the software categories currently 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.

List 1. 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

List 2. Main Activities identified from models :

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

Table 1. 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 on keyword. Provides basic image manipulation such as	7	Prepare paper	5

	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

Table 2. Links between Main Activities and WP5 Software Categories

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

Table 3. List of Main Activities against known websites

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

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