



Directions for Deliverables of EDIT's First Reporting period (covering Month 0 – March 2006 to Month 11 – February 2007)

The deliverables are often written reports but can also take another form, for example the completion of a prototype, etc. In such cases the deliverable should nevertheless also be documented in a written record of the achievement of the deliverable in addition to being listed as an achieved deliverable in the Periodic Activity Report, including any available supporting material (e.g. photos of the prototype, the report of the conference....).

Any delay in the submission of a deliverable must be reported in the Periodic activity report, in the section **“Section 2 - Workpackage progress of the period”**, where both the due date and the actual submission date (or the foreseen date, if the deliverable is not yet submitted) are reported.

Please note that the following front page is a standard provided by the EC, all requested information on this page must be filled in.

Max. 2 pages (front page excluded) per deliverable in “Garamond” 12 points. As far as possible please do not change the lay out of the standard front page.



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"

M5. Criteria for descriptive tools agreed and published

Due date of deliverable: Month 15
Actual submission date: Month 15

Start date of project: 01/03/2006

Duration: 5 years

Organisation name of lead contractor for this deliverable: MNHN 2

Revision [draft, 1, 2, ...]

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)	
CO	Confidential, only for members of the consortium (including the Commission Services)	

OBJECTIVES OF WP 5.6

The aim of WP5.6 is to identify, test and document available descriptive tools for taxonomists, and to collaborate with WP5.2 to integrate these tools as components of the Internet Platform for Cybertaxonomy. We have selected 7 to-be-tested descriptive tools. This selection was based on 4 main criteria:

- Availability (ease of obtaining and installing the software)
- Validation of suitability (i.e., is it really a taxonomic application?)
- Documentation
- Features provided (especially whether the software is for identification or analysis alone, or includes an editor tool to collect and to manage descriptive data, i.e. it doesn't need another system to edit the data)

In the present document, we will define which criteria the taxonomic tools will be assessed in the coming months.

Thus, the following table gives a non exhaustive list of criteria on which taxonomic tools could be evaluated.

Some criteria are impersonal facts (e.g. OS compatibility) but other ones result from users impressions (e.g. quality of using interface).

We structure the criteria in 4 main categories:

- I. Technical properties: "Can I use it on my computer?"
- II. Knowledge representation: "Is it adapted to represent my data?"
- III. Functionalities: "What can I do with the tool?"
- IV. User interface: "Is it easy to use and easy to obtain help and information?"
- V. Other questions

This document will be adapted and modified according to the first taxonomists interviews feedback.

I. TECHNICAL PROPERTIES
<i>A. GENERAL QUESTIONS</i>
1. CONFIGURATION REQUIRED
Operating system (Windows, Mac, Linux)/Multiplatform Is the taxonomic tool able to run on multiple platforms?
Application(s) required Does the tool need other software to be installed on computer? (e.g. Java, Access etc.)
2. HOW TO GET THE PRODUCT?
Distribution: CD, downloadable online
License List the license type (free, GPL, commercial etc.)
Price (for commercial software)

3. INSTALLATION AND USAGE
Installation Can the tool be easily installed (e.g. “plug and play”, portable)?
Is documentation or technical support available for installation? What about the quality of this documentation?
Does the system run offline, online or both?
<u>B. FORMATS</u>
1. NATIVE FORMAT
Type of storage files (XML, text files, binary files, database etc.)
Which character sets are supported (ANSI, ASCII-OEM, Unicode)?
Other software able to load this native format Is the native format of the tool adopted or recognized by other tools?
2. IMPORTING FORMAT
Is the tool able to import other formats (CSV, DELTA, SDD, other XML schema etc.)?
For each importing format what are the constraint or limits to be compatible with the native format?
3. EXPORTING FORMAT
Is the tool able to export a large set of formats (CSV, DELTA, SDD, other XML schema etc.)?
What are the limits of the export for each type of format (e.g. Xper2 doesn't export images and character dependencies in CSV format)?
II. KNOWLEDGE REPRESENTATION
<u>A. OBJECTS THAT CAN BE REPRESENTED</u>
1. TAXA
Are taxa represented as a simple list or hierarchical list?
Can specimens or parts of objects be described separately? Can stages, sexes, etc. be described separately?
Can the scope of a description be restricted to a geographical area or season (“Swiss trees in winter”)?
2. CHARACTER
Does the tool support character applicability rules (i.e. controlling plus dependent characters)? Are both “applicable-if” and “inapplicable-if” (positive/negative) rules supported? Can the controlling character be only categorical (qualitative) or also quantitative?
Character groups Support available for descriptive concepts (character groups)?
Are categorical (qualitative) data supported?
Is it possible to represent quantitative data? Which statistical measures (min, max, mean, mode, unknown range, confidence intervals, variance, standard deviation, std. Error, sample size, etc.) are supported?

<p>Polymorphism Is support for polymorphism in the case of categorical (qualitative) data available? If so, is it possible to code the frequency of a state (rarely, sometimes, often, etc.)? In what form is it possible (system-defined frequency categories, user-define frequency categories, quantitative values or ranges)?</p>
<p>Expressibility extension Can the standard expressibility (i.e.based on a controlled vocabulary) be extended through free-form text comments on categorical or quantitative data?</p>
<p>Missing data Is it possible to explain reason why descriptive data are missing (« coding status », e.g. data unavailable, inapplicable, not interpretable, out-of-scope/not-to-be-coded, withheld for legal reasons)? Which coding status values are supported?</p>
<p>Other types of characters/descriptors Is free-form text (“text character”) supported? Are molecular sequences supported? Which other forms of data are supported?</p>
<p><u>B. MEDIA SUPPORT</u></p>
<p>Images Is it possible to add a picture of the taxa? (None, 1 image, several images) Is it possible to add a picture of the characters? (None, 1 image, several images) Is it possible to add a picture of the character states? (None, 1 image, several images) Is it possible to add a picture of the character groups? (None, 1 image, several images)</p>
<p>Links to external resources, web files, references, texts, images Does the tool allow to link taxonomic concepts to external resources?</p>
<p>III. FUNCTIONALITIES</p>
<p>Allows editing and identification Does the tool allow edition and/or identification and/or other functionalities (e.g. descriptive data tools like Mesquite support editing and phylogenetic analysis, but not identification)?</p>
<p><u>A. EDIT KNOWLEDGE BASE</u></p>
<p>Editor for single item/taxon (multiple characters)?</p>
<p>Editor for multiple items/taxa (single characters)?</p>
<p>Character x item matrix editor?</p>
<p>Are revisions in the character or character states automatically reflected in the knowledge base?</p>
<p>Are tools to reorganize the data provided (e.g. move state values to a different character for all items)?</p>
<p>Are tool for data checking (redundancy, etc.) available? Check consistency of the description (in respect to logical dependencies between descriptors)</p>
<p>Support for multilingual character sets/How is it accomplished? Original DELTA supported multilingual data, but at a great cost (need to synchronize separate files manually). Other tools integrate this into a database</p>
<p>Possibility to read the same knowledge base from several computers concurrently? Possibility to edit the same knowledge base from several computers concurrently?</p>

Can bibliographic references be added to descriptive data in the tool, on the methods etc. ?
<i>B. TO IDENTIFY</i>
Knowledge management: uncertainty management (uncertainty occurs in both data collection and identification) Is it possible to qualify data with uncertainty (probably, perhaps, etc.)?
Does the tool explain how the result was reached? If no taxa remains in an identification attempt, can the tool report taxa that do not fully satisfy all criteria? Does it explain which observations are not fulfilled for a given taxon? »
Multi-access key Is it possible to choose the sequence of characters during the identification or is it an imposed process?
Description of the identified taxon Is it possible to have access to a complete descriptive card of the identified taxon?
Possibility to sort characters (groups or hierarchical views, filtering by groups, father/son characters)
Does the tool support data set creators in providing for anticipated misinterpretations by users (e.g. Can a plant be found, even if the flower bracts are misinterpreted as petals (as in dogwood)?)
IV. USER INTERFACE – DOCUMENTATION – AVAILABLE APPLICATIONS
Online help Does the tool provide an online help? How do you estimate the quality of this help? Does the tool provide offline help? How do you estimate the quality of this help?
Available languages Is the tool available in several languages? Which one?
Friendliness of the user interface Is it easy to do the job with the “tool interface”? Do we find functionalities quickly?
Loading time How much time does it take to install it? How much time does it take to launch it? When using it, does the tool have latency times of more than 1 or 2 seconds for some operations?
Dynamic community (mailing list, blog, etc.) Is there a website which provides news, dynamic forum and mailing list?
Number of known users (for commercial packages that are sold and accounted) Which other users use the tool?
Known knowledge bases created with the tool (examples)
Can the major operations be run from other applications without human user intervention? Which scripting languages or interoperable object models are supported?
Configurable Can the user modify toolbars and create keyboard short cuts?

Known bugs
Possible future developments Can new features be expected to become available soon?
V. OTHER QUESTIONS
What do the users expect?
What do they consider the biggest shortcomings of a program they use for their work

Elise Kuntzelmann and Régine Vignes

