



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”

D5.18 (formerly D.5.3.1.2) Statement of Requirements

Due date of deliverable: Month 13

Actual submission date: Month 16

Start date of project: 01/03/2006

Duration: 5 years

Organisation name of lead contractor for this deliverable: **10 NHML**

Revision [final]

Project co-funded by the European Commission within the Sixth Framework Programme (2002-2006)		
Dissemination Level		
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)	

D5.18 (was D.5.3.1.2) Statement of Requirements

System Specification for software to deliver the European Virtual Library of Taxonomic Literature (E-ViTAL)

1. Introduction	p.2
2. User Requirements	p.3
3. Technical Requirements	p.4
4. Options for software solution	p.5
5. Criteria for selecting software solution to provide E-ViTAL	p.7
Acknowledgements	p.8
References	p.8
Appendix: Online Survey Results	p.10

1. Introduction

1.1 Background

Compiling a bibliography is a core component of all taxonomic work, whether published or online. Active research requires the tracing of pertinent references, actual sight of the content, retention of copies for further study (or retention of links to the source, either online or in an institutional library) and management of lists of references. When a researcher produces a taxonomic publication, a customised reference list has to be built, and formatted to fit the publisher's specifications. The task of gathering literature has been identified as one of the bottlenecks impeding taxonomic work in the Work Package 5.2.1.1 Draft functional model and bottleneck report for rev. taxonomy [1].

The Internet Platform for Cybertaxonomy, being put together under Work Package 5, will need to provide for the bibliographic needs of its users. In particular, it can help with the creation of reusable information resources and, it is to be hoped, will help reduce duplication of effort when it comes to constructing bibliographies. Platform components can take the form of software *applications* (desktop or web-based) for human users or (web) *services*. A service may provide data (a *data service*), some kind of transformation (a *transformation service*), validation of data (*validation service*), or other data processing. The platform will be a collection of interacting components which may be combined and assembled according to the task in hand.

1.2 Scope

This report sets out to present the views of the user community. The resulting 'wish-list' may well prove to be beyond the scope of the EDIT project, but the information gathered may usefully inform other projects such as the Encyclopedia of Life [2] and the Biodiversity Heritage Library [3], both of which are likely to play an important role in furthering the aims of EDIT. The report tries to avoid any preconception of what E-ViTAL might deliver or how it might be organised, in order to capture the stated needs of potential users.

The principle outcome from this report is a set of criteria (Sect. 5) that can be used for assessing the suitability of candidate products.

1.3 Methodology

Initial discussions and 'brain-storming' sessions were carried out in conjunction with scientific staff and library staff within the Natural History Museum, London. This was then extended to the online discussion sections of the WP5 & WP6 websites. The author also attended the WP6 Data Commonalities Meeting 26-27th April. This input led to the production of a questionnaire, which was made available online using the SurveyMonkey service [4] between the 3rd and 11th of May 2007. The questionnaire was announced on the TAXACOM [5] and TDWG [6] list-servers and attracted a total of 64 responses. Responses were received from 19 countries, 10 of which were within the European Member States.

Technical Requirements were compiled by the Data and Digital Systems Team in the Department of Library and Information Services at the Natural History Museum, London.

2. User Requirements

Users see E-ViTAL as providing a resource both for bibliographic discovery and reference management. Ideally they would like E-ViTAL to cover all aspects of a taxonomist's literature needs, including the storage of fully-indexed, full content articles. Only one of the features suggested in the Questionnaire was considered as 'not required' (the ability to sort references by 'type of publication') – all other features were deemed to be either mandatory or desirable.

The stated requirements of users challenge the assumption made in Deliverable 5.3.1 'Existing Digital Library activity, principles and standards' [7] that "Work package 5.3 will not primarily be involved in the creation and management of new digital objects, but instead focusing upon providing new means of access to resources as created and managed by other entities

Key requirements are to be able to populate reference lists, share reference lists and to detect and resolve duplicate references. There is a need to deal with citations at an article level (which most library online catalogues do not) and, for some users, at a page level - that is they wish to know on what page a species description or figure occurs.

Equivalent, and thus duplicate, references arise mainly through differences in the way names of authors are set out, such as transliteration from non-European scripts and handling of forenames prefixes and suffixes. Translated titles and non-standard abbreviation of journal titles also cause problems. Occasionally an article (especially early ones) may be re-printed in different publications. Respondents to the questionnaire supported the use of Globally Unique Identifiers (GUIDs) to help identify equivalent articles and to act as pointers to online content.

Users want not just access to information about literature but access to the literature itself. Ultimately, researchers need to view the content of scientific articles or, at least, find out where it is possible get access to a copy (either as a physical object in a library or in an online repository). The best that E-ViTAL can probably do is to provide links to online resources (though it is interesting to note that Google Books [8] provides a limited guide to library holdings). The Biodiversity Heritage Library will, in due course, provide a freely-accessible resource for original content.

Users should be able to annotate references and apply 'tags' to categorise references.

Bulk loading and export of reference lists to allow integration with personal reference managers was considered desirable, as was the ability to search, sort and filter reference lists in a variety of ways. Full text indexing that would allow searching for terms contained in both the title or page content (where available) was considered mandatory by a majority of respondents.

The bibliographic component of the Internet Platform for Cybertaxonomy will need access to other components that provide taxonomic indexes and classification schemes to help navigate and filter the reference lists. This would include the automatic recognition of taxonomic synonyms to assist with query expansion.

The full results of the questionnaire are presented in an Appendix at the end of this report.

3. Technical Requirements

The component, or components, that constitute the E-ViTAL, whether specially commissioned or utilising existing products, will need to interoperate with other components of the Internet Platform for Cybertaxonomy through the Common data Model that is being developed. To this end it may also be appropriate that bibliographic data can be wrapped to conform to widely used standard formats like EndNote, to taxonomic databases with highly atomised literature coverage such as the Berlin Taxonomic Information Model [9], as well as to the reference modules available for the Drupal content management system.

Ideally, the E-ViTAL will be both flexible and scalable and will certainly need to be sustainable beyond the life of the EDIT project.

Components will need to work on all types of personal computers and the common operating systems and web browsers. Any software required on the client side should have a foolproof installation package and there should be a mechanism for supplying known users with updates as required.

In order to share data and promote interoperability, the system should adhere to recognised standards for storing and passing bibliographic metadata and Globally Unique Identifiers for literature at the item and article level. An appropriate choice between one or more of: DOIs [10], LSIDs [11], Handles [12], OpenURL [13], SICI [14] and BICI [15] will need to be made. A good overview of implementing persistent identifiers is given by Hilse and Kothe (2006) [16]. It is likely that DOIs will be utilised within the Biodiversity Heritage Library (Neil Thomson, pers.comm.)

If an interface is provided, which systematically queries third-part resources such as bibliographic datasets then it is likely that it would need to support the commonly used Z39.50 [17] or OAI-PMH [18] protocols to enable searches to be performed on compliant databases (or sets of databases in aggregate).

If it is intended to link to full text online resources then whilst support for OpenURL is likely to be essential as part of that process, the ability to direct users to appropriate resources in the context of their local environment will also be paramount and will require a resolver for the

OpenURL. This could require a database of institutional subscriptions to be maintained by a participating member.

4. Options for software solution

How might the needs of users be met within the scope of the EDIT project? It is not the purpose of this report to make recommendations for a single product, but it is pertinent to look at the range of applicable architectures and compare possible solutions within each area against requirements.

Deliverable 5.3.1 'Existing Digital Library activity, principles and standards' defines a virtual library as a means to providing access to resources created and managed by other entities. It is clear that users are hoping for E-ViTAL to be a resource in its own right, although it is also may be beyond the scope of E-ViTAL to store copies of articles, perform the services of a personal reference management system, or provide taxonomic tools (although these last two may be picked up in other components of the Internet Platform for Cybertaxonomy).

A number of users have commented on the need to avoid duplicating, in E-ViTAL, services and resources that were already available elsewhere. Indeed, it could be argued that a 'do nothing' option is perfectly valid and that E-ViTAL could serve simply as a resource discovery service providing a directory of software and websites that assist with bibliographic activities. Certainly, reference management can be undertaken using a product such as Endnote; references may be discovered through sites such as Zoological Record [19], TL-2 (for plants) [20], Google Scholar [21] and Google Books; bibliographies may be shared through services such as citeulike [22] and Connotea [23]; original content viewed through online journals and, for earlier material, through projects such as Biodiversity Heritage Library, Animalbase [24] and Gallica [25]; and the ParaCite Toolkit [26] can be used to parse online bibliographies and the content of articles.

What is true is that currently no single product meets all the bibliographic needs for taxonomic study, and it is also true that many of the more effective solutions have to be paid for.

The online questionnaire came out in favour of a web-based service (83%) rather than a portal to distributed services (15%) or a peer-to-peer network (2%). However, whilst it could be possible to build a generalised site catering for all taxonomic groups, it will also be necessary to support communities who are developing their own specific sites (at the time of writing, exemplar sites are being constructed using the 'scratchpad' software provided by WP6, which is based upon the Drupal open-source content management system [27], and which already include a bibliographic module) [28]. Similar exemplar sites are also being developed in the CATE project [29].

It will be interesting to see to what extent co-operative effort within the taxonomic community acts to overcome perceived impediments to access to and sharing of information. We have a situation where a growing number of relevant journals now make their content available online, but only through subscription, and where early and out of copyright literature is being made freely available on sites such as Animalbase, Gallica and Biodiversity Heritage Library. However this leaves a large and important body of in-copyright literature accessible only through application to an academic library - who can supply photocopies for personal research. Requests for help in obtaining literature occasionally appear on e-mail groups such as Taxacom and it is likely that researchers will share electronic copies of articles amongst themselves, rather than

loading the content onto public websites. This practice may encourage the development of peer-to-peer file-sharing services within the community.

Apart from access to past literature, E-ViTAL will also need to keep abreast of new developments in publishing, which includes a growing trend for individuals and institutions to 'self-archive' and make copies of their own publications available online.

Certainly, there is likely to be an 'information divide' between taxonomists employed by universities, museums and herbaria, who enjoy ready access to institutional libraries and subscriptions to electronic journals, and amateur workers who do not have access to such resources. Since Work Package 2 is seeking to increase the participation of amateur taxonomists, Work Package 5 should attempt to address this divide.

The users' wish to access page-level citations can be met when digitised and indexed content is available online. Until then most likely source would be through annotations to references provided through community effort.

The ParaCite Toolkit (mentioned above) for parsing references; specifically allows the conversion of reference lists into valid OpenURLs, converting existing metadata into valid OpenURLs, collecting metadata from references to carry out internal searches. It can also be used to extract reference lists held within documents - whether these are a reference list at the end of a digitised article or a bibliography posted to the Internet as an HTML page. COinS (ContextObjects in Spans) [30] provides a specification for embedding OpenURL citation metadata in conventional HTML. An example of the use of COinS can be found on the website of the West Midlands Bird Club (<http://www.westmidlandbirdclub.com/biblio/NBotWM.htm>) where the ISBN is linked to Wikipedia:Book sources [31] that allows searching of the catalogues of libraries and booksellers.

The requirement to build up reference lists that include articles that are not available online precludes the use of resources such as Connotea and citeulike, as they are presently configured. One way to overcome this deficiency would be for interest groups to construct web pages containing at least the metadata associated with such articles which would then have URIs that could be cited.

In addition to the Digital Library projects mentioned in Deliverable 5.3.1, there are two projects which were funded by the European Commission that set out to tackle similar information-handling issues within the Bioinformatics community and which may have produced tools that are applicable to EDIT WP5. The Online Research Information Environment for the Life Sciences (ORIEL) [32] has developed tools and procedures to promote access to and integration of a wide range of information resources in the life sciences and to enable effective linking of different types of biological information (literature, factual and multimedia databases). One of the tools worth investigating is the iHOP Literature Networking Tool [33] that builds a network of interlinked references indexed using text mining technology. Technologies developed within ORIEL feed into the second project: E-BioSci [34], a European platform for access and retrieval of full text and factual information in the Life Sciences. One of its features is to allow searches of literature full-text to be based on concepts rather than individual keywords.

There are also a number of open source collaborative web-based reference managements systems that offer some of the features that users require: Aigaion [35], Refbase [36], Wikindx. [37] and Bibster (which allows sharing of bibliographic metadata on a distributed peer-to-peer network) [38].

5. Criteria for selecting software solution to provide E-ViTAL

The previous Section shows that there is a whole range of choices for a software solution to the bibliographic component of the Internet Platform for Cybertaxonomy. However it is both possible and desirable to set out in general terms a set of criteria that help evaluate candidate products.

Essential

- Provide access from any personal computer to a single, web-based, site
- Allow entry, storage and searching of metadata relating to references that are not yet recorded in other online resources
- Provide a solution that is sustainable beyond the life of the EDIT project
- Provide for de-duplication of equivalent references
- Allow labelling of references with an acceptable form of Globally Unique Identifier
- Allow bulk import and export of bibliographic references in common formats
- Allow collaborative working and sharing of references
- Provide for searching, sorting and filtering of reference lists
- Integrate with other components of the EDIT Internet Platform for Cybertaxonomy, both at a technical level (through the Common data Model) and with respect to content.
- Provide for links to external resources

Desirable

- Allow storage of online content (images of text, PDF files, etc.)
- Provide for personalised content to enable sets of references to be accessed by individuals or groups of researchers
- Allow annotation and tagging of references
- Provide facility for comparison and merging of separate lists of references
- Provide for page-level citations
- Provide a validator for journal titles and their abbreviations
- Provide an alert service (e-mail or RSS) to tell a user when new references are added within their field of interest
- Provide a multi-lingual user interface

Acknowledgements

The author is indebted to the following people for their helpful contributions and comments: Walter Berendsohn, Irine Brake, Markus Döring, Mark Jackson, Rod Page, Dave Roberts, Bernard Scaife, Vince Smith, Neil Thomson, John Tweddle; also to all the people who took the trouble to respond to the online questionnaire.

References

1. EDIT Deliverable 5.2.1.1: Draft functional model and bottleneck report for rev. taxonomy. Available at: <http://wp5.e-taxonomy.eu/blog/about/deliverables/>
2. Encyclopedia of Life. <http://www.eol.org/>
3. Biodiversity Heritage Library. <http://www.biodiversitylibrary.org/>
4. SurveyMonkey. <http://www.surveymonkey.com/>
5. Taxacom - Biological Systematics Discussion List. <http://mailman.nhm.ku.edu/mailman/listinfo/taxacom>
6. Biodiversity Information Standards TDWG Announcement List. <http://lists.tdwg.org/mailman/listinfo>
7. EDIT Deliverable 5.3.1: Existing Digital Library activity, principles and standards. Available at: <http://wp5.e-taxonomy.eu/blog/about/deliverables/>
8. Google Book Search. <http://books.google.com/>
9. The Berlin Taxonomic Information Model v2.0. <http://www.bgbm.org/biodivinf/docs/bgbm-model/>
10. The Digital Object Identifier System. <http://www.doi.org/>
11. Life Science Identifiers. <http://www.omg.org/cgi-bin/doc?dte/04-05-01>
12. The Handle System. <http://www.handle.net/>
13. The OpenURL Framework for Context-Sensitive Services. ANSI/NISO Z39.88 -2004. http://www.niso.org/standards/standard_detail.cfm?std_id=783
14. Serial Item and Contribution Identifier (SICI). ANSI/NISO Z39.56 -1996 (R2002). http://www.niso.org/standards/standard_detail.cfm?std_id=530
15. Book Item and Component Identifier (BICI). <http://www.niso.org/standards/resources/bici.html>
16. Hilde, H-W. & Kothe, J. 2006. Implementing persistent identifiers. Consortium of European Research Libraries & European Commission on Preservation and Access. Available as PDF at (<http://nbn-resolving.de/urn:nbn:de:gbv:7-isbn-90-6984-508-3-8>)
17. Information Retrieval (Z39.50): Application Service Definition and Protocol Specification. <http://www.loc.gov/z3950/agency/Z39-50-2003.pdf>
18. The Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH) <http://www.openarchives.org/OAI/openarchivesprotocol.html>
19. Zoological Record. Thomson Scientific. <http://scientific.thomson.com/products/zr/>

20. TL-2 Online. the online edition of Taxonomic Literature, 2nd edition. <http://tl2.idcpublishers.info/>
21. Google Scholar. <http://scholar.google.com/>
22. citeulike. <http://www.citeulike.org/>
23. Connotea. <http://www.connotea.org/>
24. Animalbase. Early Zoological Literature Online. <http://www.animalbase.de/>
25. Gallica, la bibliothèque numérique de la Bibliothèque nationale de France. <http://gallica.bnf.fr/>
26. Paracite Toolkit. <http://paracite.eprints.org/>
27. Drupal, an open source content management platform. <http://drupal.org/>
28. Drupal Bibliographic Module. <http://drupal.org/project/biblio>
29. CATE (Creating a Taxonomic e-Science) <http://www.cate-project.org/>
30. CoinS (ContextObjects in Spans). <http://ocoins.info/>
31. Wikipedia:Book sources. http://en.wikipedia.org/wiki/Wikipedia:Book_sources
32. ORIEL. Online Research Information Environment for the Life Sciences. <http://www.oriel.org/>
33. A Gene Network for Navigating the Literature. Hoffmann, R., Valencia, A. Nature Genetics 36, 664 (2004) and - iHOP - <http://www.ihop-net.org/>
34. E-BioSci. A European platform for access and retrieval of full text and factual information in the Life Sciences. <http://www.e-biosci.org/>
35. Aigaion - A Web based bibliography management system. <http://www.aigaion.nl/>
36. Refbase - Web reference database. <http://refbase.sourceforge.net/>
37. Wikindx. <http://wikindx.sourceforge.net/>
38. Bibster. <http://bibster.semanticweb.org/>

Appendix: Online Questionnaire

This is the list of questions and responses received to the Online User Requirements Questionnaire.

E-ViTAL should:

1. Provide an online repository (source) of references relevant to taxonomic resource, assembled through community effort.

Yes	98.4%
No	1.6%

1.a If No, please give your alternative vision. **Please State - 2 responses (see below)**

2. Provide a reference management facility (duplicating certain aspects of packages such as Endnote).

Yes	74.1%
No	25.9%

3. Store textual content of articles (subject to IPR/Copyright issues).

Mandatory	26.6%
Desirable	67.2%
Not Required	6.2%

4. Link references to external source of textual content (will usually be an online Journal, or Biodiversity Heritage Library).

Mandatory	46.8%
Desirable	50%
Not Required	3.2%

5. Be built as:

A single web-based system. **46.6%**

A single web-based system + private work areas (for individuals and groups) on server. **25.9%**

A single web-based system + toolkit/client installed on users' PCs. **10.3%**

Portal to distributed databases of references (of individuals, EDIT Exemplar projects, abstract/indexes of e-Journals, institutions and commercial). **15.5%**

Peer-to-Peer architecture using individual PCs. **1.7%**

5.a It would be most helpful if you could give reasons for choice of answer. **Please State - 38 responses (see below)**

6. Provide Multi-lingual user interface (ability to swap between common languages).

Mandatory	8.1%
Desirable	53.2%
Not Required	38.7%

7. Be able to import files of references in various formats and from various applications.

Mandatory	37.3%
Desirable	49.2%
Not Required	13.6%

7a. If required, please state which packages and formats (e.g. Endnote, tab separated) **Please State - 41 responses (see below)**

8. Be able to export selected references in variety of formats.

Mandatory	47.5%
Desirable	44.3%
Not Required	8.2%

8a. If required, please state which packages and formats (e.g. Endnote, tab separated) **Please State - 45 responses (see below)**

9. Provide easy switching between formats for citation, to meet different standards required for publication in Journals.

Mandatory	18.3%
Desirable	53.3%
Not Required	28.3%

10. Index all text – allowing search of terms contained in title or page content (where available).

Mandatory	51.7%
Desirable	40%
Not Required	8.3%

11. Allow users to tag references using keywords and thus find related articles or to be used to filter reference (Could include taxonomic and geographic names, or to flag new taxon, descriptions, figures, ecology, etc.) by:

Community tagging

Mandatory	20%
Desirable	63%
Not Required	18%

Controlled terminology (pre-built)

Mandatory	21%
Desirable	70%
Not Required	9%

Hierarchical classification system (i.e. for subjects, taxonomy, geography)

Mandatory	39%
Desirable	49%
Not Required	12%

12. Allow users to annotate references (i.e. express opinions on content).

Mandatory	11.3%
Desirable	58.1%
Not Required	30.6%

13. Provide ability to sort references by: **(NOTE: this question was incorrectly formatted, so that only one instance of each option 'mandatory, desirable, not required' could be chosen. This means that the scoring did not work as expected)**

author	Mandatory 90%, Desirable 9%, not Required 2%
year	Mandatory 11%, Desirable 82%, not Required 7%
title	Mandatory 7%, Desirable 73%, not Required 20%
Journal/Publisher	Mandatory 20%, Desirable 40%, not Required 40%
type (monograph/book, journal, symposium proceedings, etc.)	Mandatory 0%, Desirable 23%, not Required 77%

13a. other **Please State - 31 responses (see below)**

14. Provide ability to filter results of a search by:

author	Mandatory 91%, Desirable 9%, not Required 0%
year (single)	Mandatory 65%, Desirable 33%, not Required 2%
year (range)	Mandatory 57%, Desirable 38%, not Required 6%
title	Mandatory 60%, Desirable 32%, not Required 8%
Journal/Publisher	Mandatory 48%, Desirable 40%, not Required 12%
type (monograph/book, journal, symposium proceedings, etc.)	Mandatory 28%, Desirable 38%, not Required 34%
keywords (could include taxonomic names)	Mandatory 84%, Desirable 14%, not Required 2%

14a. ...or other **Please State - 8 responses (see below)**

15. Not be limited to 'online' articles (i.e. avoid constraints inherent in Connotea) but allow incorporation of references only occurring in print form.

Mandatory	68.9%
Desirable	29.5%
Not Required	1.6%

16. Provide for merging and de-duplication of reference lists which may contain variants in reference/author/journal citations, whilst retaining language variants of a reference (e.g. retaining title in original language and translation/transliteration).

Mandatory	37.7%
Desirable	57.4%
Not Required	4.9%

17. Assign/handle Unique Identifiers (GUID/DOI) for individual references and articles.

Mandatory	49.2%
Desirable	39%
Not Required	11.9%

18. Provide for page-level, as well as article-level or page-level citations.

Mandatory	25.5%
Desirable	54.5%
Not Required	20%

19. Provide for query expansion through knowledge base of:

taxonomic synonyms	Mandatory 61%, Desirable 35%, not Required 3%
orthographic variants of authors names	Mandatory 45%, Desirable 52%, not Required 3%
multilingual translation (i.e. input search term in English and searches for equivalent term in French, German, Russian titles)	Mandatory 22%, Desirable 66%, not Required 12%

20. Allow creation (and storage) of 'sets': collections of citations tailored (selected from entire reference holdings) to meet needs of a user group (research team, taxonomic group).

Mandatory	15.3%
Desirable	69.5%
Not Required	16.3%

21. Permit user to save personal search criteria (and results of searches).

Mandatory	27.1%
Desirable	61%
Not Required	11.9%

22. Build up an Authority File (data warehouse) of all submitted articles, from which user can select and download (or mark) articles for incorporation in a (personal) bibliography.

Mandatory	31%
Desirable	60.3%
Not Required	8.6%

23. Provide a validator for Journal titles and abbreviations.

Mandatory	45%
Desirable	48.3%
Not Required	6.7%

24. Extract references from within an article (e.g. system would need to index reference list at end of article - assumes access to source).

Mandatory	8.8%
Desirable	70.2%
Not Required	21.1%

25. Ability to display linkages – i.e. which articles are cited by which others.

Mandatory	10.3%
Desirable	69%
Not Required	20.7%

26. Record 'Date of Entry' (of each reference), so that user can check for updates.

Mandatory	28.6%
Desirable	51.8%
Not Required	19.6%

27. Include an alert system so that user can be notified of new references relating to their area of interest.

Mandatory	21.7%
Desirable	66.7%
Not Required	11.7%

28. Include directory of other Bibliographic portals and resources (e.g. Intute, Biosis, Google Scholar, e-Journals).

Mandatory	26.8%
Desirable	58.9%
Not Required	14.3%

29. Provide ability to include external resources in a search:

Catalogues of relevant libraries	Mandatory 21%, Desirable 67%, not Required 12%
Scientific image collections	Mandatory 24%, Desirable 67%, not Required 9%
EDIT microsites (exemplar projects)	Mandatory 8%, Desirable 74%, not Required 18%
Sources of abstracts, where accessible (i.e. Biosis, Zoo Record, Index Kewensis)	Mandatory 36%, Desirable 59%, not Required 5%
Google Scholar (and similar search engines)	Mandatory 29%, Desirable 64%, not Required 7%
Biodiversity Heritage Library	Mandatory 36%, Desirable 60%, not Required 4%

30. Collate references cited in Manuscripts submitted through EDIT WP6.

Mandatory	12.2%
Desirable	57.1%
Not Required	30.6%

31. Provide Taxonomy Tools (classification hierarchies, informal group names) to help users locate names in references.

Tools built into site	Mandatory 33%, Desirable 47%, not Required 20%
Link to external sources	Mandatory 22%, Desirable 65%, not Required 14%

32. Any other requirements that you think essential. **Please State - 18 responses (see below)**

33. Any other requirements that you think desirable. **Please State - 8 responses (see below)**

Submitted responses:

1a. *Provide an online repository (source) of references relevant to taxonomic resource, assembled through community effort. If No, please give your alternative vision. - 2 responses*

- E-ViTAL should serve as a portal for online taxonomic references. Now when we need a 17th, 18th or 19th Century taxonomic reference, we have to go one-by-one to all the available web sites, such as the Latin publications site, Gallica, New York Botanical Garden library, Missouri Botanical Garden library, the Madrid botanical garden, Smithsonian library, the Hunt library, etc., until we find what we need. A one-stop portal for all of these sites would be a tremendous aid. Also, I fear that duplication of effort may start creeping into the digitization of taxonomic works. I already know about several digital copies of some important botanical works.
- My answer really depends on a number of factors: would you be charging? How could one search it? Could everyone, even folks outside Europe writing about plants outside of Europe contribute or? No answer until I know what you have in mind.

5a. *Nature of build. (Most helpful if you could give reasons for choice). - 38 responses*

- [portal] To provide for additional portals to specimen based research.
- [web-based] Seems better retrievable
- [web-based] There are too many portals on www; sometimes, duplicate information is useless, it's a mess
- [web-based + private area] I'm not sure what all the options actually mean.
- [web-based + private area] This option is the most universally available with added options for groups.
- [web-based + private area] It's more simple and useful (and keep one private area to keep the information that the users consider necessary).
- [web-based + private area] Experience with botanical databases has shown me that distributed databases are never maintained equally at separate sites.
- [portal] Many important early botanical works are already digitized. What we need is easy access to finding those works and to avoiding duplication of effort.
- [portal] harvest/cache model is appropriate to avoid duplication of existing effort
- [web-based + private area] individual specialists and groups already have specialist bibliographies (on a taxon eg) that could contribute and be built upon
- [web-based + private area] I am interested in basal poidis, triticeae, and intermountain region. N
- [web-based + client] Simple and convenient
- [web-based] simplicity of use
- [web-based] This is how I find literature databases on the web most helpful. Everything else is too complicated.
- [peer to peer] Best taps into existing data
- [portal] An integrative interface to many different sources would be user-friendly but at the same time integrate more content than a single web-based system could.
- [web-based] it shouldn't depend on my department setting something up
- [web-based + private area] I envision a system in which references are available to all and taxon-specific lists can be edited as wikis for specific working groups.
- [web-based] keep it simple
- [web-based] Build standalone server but support federation of databases

- [web-based] KISS-Keep it simple and Stupid! Everybody has his own reference managing system on his own computer, so why duplicate it, once more.
- [web-based] A single web-based system would ensure the high quality of all data and not a diverse source of conflicting quality standards
- [web-based] I usually prefer easy to use systems with as less complexity as possible. A single web based system seems to be the most user-friendly option to me.
- [web-based + private area For integration with other initiatives e.g. EOL
- [portal] Sustainability.
- [web-based + private area possibility 2 of 5 would be the optimal to guarantee misuse
- [web-based] provide textual information only like that done in Botanicus - scanned documents with links to protologues of all types at BM
- [web-based + private area Should link to separate web-based community repositories of bibliographic information such as those from the EDIT Scratchpads.
- [web-based] to be available for every single computer on earth
- [web-based] at best something between a single pure web based system (avoid clients for data editing as in SysTax - this always causes problems with actualisation of client) and a peer to peer network
- [portal] Do one thing and do it good. Other functions are covered by other software, such as CMSs, Endnote. The essential thing is the access to digital references down to the level of individual pages etc.
- [web-based] focus resources on key infrastructure and content
- [web-based + private area access from all over the world with all personal data.
- [web-based] best if anyone could use it (no need for installation of special software or library connections)
- [web-based] easy access from everywhere
- [portal] sustainability of project and work results
- [web-based] the easier the better. Least time and money should be spend on the system it self, it should go to content. That is also why there should on NO CIRCUMSTANCES money should be spent on multi-linguality.
- [portal] More quick and easy to achieve than a single web-based system although the latter is more desirable

7a. Be able to import files of references in various formats and from various applications. If required, please state which packages and formats (e.g. Endnote, tsv). - 41 responses

- Excel, Access or similar spreadsheet or database
- tab and pipe delimited
- Endnote and tab separated seem enough
- endnote, ris format
- EndNote
- Endnote
- If you can bring in delimited ASCII, you can work from any output
- Endnote
- tab separated
- Endnote/RefMan RIS, MARC, TDWG Lit std (when exists)
- Endnote
- Decided to delete my answer but cannot
- Endnote

- Endnote, <symbol> separated
- Standard reference programs, tab separated and dbf-legacy files
- endnote, reference manager
- Endnote, BibTeX, xml
- endnote, bibtex, tab separated
- endnote, tab separated
- not just Endnote-formats but also internet-resources such as Kew Taxonomic Database
- CSV
- No proprietary formats. Tab separated text file, Excel at best
- endnote, referencemanager, tab or whatever separated *.txt, *.xml, etc....
- Endnote
- Endnote, MS Reference Manager , tab separated
- Endnote, OpenURL, DOI dereferencing
- Endnote
- EndNote (or ProCite)
- Should talk to the other web repositories. e.g. via OAI protocol. Endnote exports to any format. Also BibTeX
- papyrus, endnote, tab delimited, etc
- Endnote
- tab delimited export format, MS ACCESS format, DBASE format with the field conventions of Reference Manager and Endnote (both are quite similar), not only TAG-format as provided by both
- Endnote, tab separated, downloadable formats, upcoming TDWG standards
- Reference Manager
- tab separated (Mandatory), Endnote & ReferenceManager (desirable, not mandatory since both are supposed to export to tab separated csv)
- tab separated
- EndNote, RefWorks, tab separated, -- as many as possible
- endnote, tab separated, txt
- Endnote and other key proprietary software
- tab separated, Excel files, reference manager
- tab separated

8a. Be able to export selected references in variety of formats. If required, please state which packages and formats (e.g. Endnote, tsv). - 45 responses

- Excel, Access or similar spreadsheet or database
- tab and pipe delimited
- Endnote and tab separated seem enough
- endnote, ris format
- EndNote
- tab separated
- Endnote
- Excel file, Access, tab separated
- Endnote
- tab separated

- as above
- Endnote; tab-separated
- Endnote
- Endnote, <symbol> separated
- tab separated
- endnote, reference manager
- Endnote, BibTex, xml
- endnote, bibtex, tab separated
- tab
- a format tha can be imported in various packages (tab separated ?)
- Endnote, ASCII, etc.
- CSV
- BibTex
- Tab separated text file. As above too many programmes with too many formats to provide for that. TextTab is universal, also to very other programmes, such as Mathematica etc etc
- endnote, referencemanager, tab or whatever separated *.txt, +.xml, complete reference, etc....
- Endnote, .doc
- Endnote
- Endnote, some form of XML or RDF (cf. Connotea)
- Endnote
- EndNote (or ProCite)
- Endnote, bibTex
- all possible
- Endnote
- see question 7
- Endnote, tab separated, upcoming TDWG standards
- tab/csv
- Reference Manager
- tab separated (Mandatory), Endnote & ReferenceManager (desirable, not mandatory since both are supposed to import from tab separated csv)
- tab separated
- EndNote, tab delimited
- tab separated, txt. endnote
- tab separated
- Endnote etc
- tab separated or Excel files, reference manager
- tab separated

13a. Provide ability to sort references by: other (please state). - 31 responses

- taxa
- species names described in publication
- all of the above
- mandatory for ALL (survey limited marking to one)
- all of the above should be mandatory but I could not click all the buttons on
- Why can I only mark 3 options here?
- is it possible to sort by any of these. Not only by one?

- radio buttons screwed up - can't select more than one as mandatory
- author and year mandatory
- All the above
- Key words eg, taxa
- It is not possible to select mandatory for all suggestions in your form, that is what I wanted to tag
- All mandatory - form broken
- taxon
- author+year+title=mandatory. Other = desirable (was not possible to check all boxes)
- all of these
- and year, all three mandatory... (I can't select more than 2 for some reason, nor choose more than one as mandatory)
- sorting is essential to retrieving the information quickly. So the references should be sortable to all these fields, also in sequence: author, then Year, then Journal
- 'Comment: there are problems in this document to mark more than one possibility as mandatory'
- keywords
- author year and title ALL mandatory!
- referenced taxa
- all of the above
- taxon
- author, year, title and journal are all mandatory, but I cannot input this in this sheet
- Only one choice per column; all of the above or at least the first four mandatory.
- ## there is a bug in the radio button grouping ## sorting by all is highly desirable
- all of the above, including reverse chronological
- taxon treated
- for all categories to check inconsistencies
- [question does not work, one can only cross one per column]

14a. Provide ability to filter by: other (please state). - **8 responses**

- All the above
- specimens cited in text
- all
- Georeference coordinates with error bounds
- see my remark above
- page number; image of original present
- Boolean search function, please!
- keywords only if taxonomic names are meant

32. Any other requirements that you think essential. - **18 responses**

- Source of data about the reference (e.g., taken directly from the reference itself, from a secondary source, etc.). (Basically some way to gauge the reliability of data about the reference.) 2) Exact publication date, if known, in addition to nominal year of publication.
- Not at this moment.

- As user and co-developer of a plant-based database, I know it is helpful to have a word-bank or nomenclatural list of taxa that a user can work from. It is much easier to search for something if you have it spelled correctly! Even if the name is not currently accepted, it is still an entry point into the data. Perhaps E-ViTAL could tap into IPNI and other sources. This is an expansion of your option 31, above.
- The questions suggest that an indexing much like BIOSIS is being considered. Constructing a new 'BIOSIS' would be a waste; it already exists! The planning should be along the lines of providing a portal that in which a single search for a taxon, group of taxa, or entity returns information on the myriad of sites now available. Its other function is to coordinate the efforts to insure that they do not duplicate work and their data is compatible and can be searched from a single site.
- I had trouble answering Q13. The HTML wouldn't let me accurately express my preferences.
- Don't reinvent the wheel. There are great reference management tools and well working reference search engines and databases (eg. PubMed, Web of Science etc). What would be helpful for me is a depository of pdf files of old (pre-pdf publication of journal articles) and rare publications that don't exist online otherwise. Absolutely essential would be good search options and import and export compatibility. For search options a system like Web of Science with access to reference lists and search through them would be great.
- The citeUlike model may be useful - allows adding of own reference library. (<http://www.citeulike.org>)
- Information about availability of the full text / abstract only and copyright information
- If the E-ViTAL hosts any page images ensure these have a resolvable URL
- Since working in Nairobi, I do appreciate the simple User Interfaces of Google very much, this keeps access fast and efficient. Keep programming simple so that after EDIT there is a long term perspective. Also watch out, that the E-ViTAL passes the 'Bus Test'. Do link and coordinate with the - fortunately - many websites for zoological and botanical literature like Animalbase, national Libraries of France and Germany and may other links I collected (fhaas@icipe.org for questions)
- This survey suggests the potential for a project that, in my opinion, is far too big and unwieldy. I suggest focussing on a core set of functions, and leave the rest to projects that do this sort of thing much better. I think the two most important things this project could do are:
 1. provide an OpenURL service that will resolve a bibliographic citation (or a GUID)
 2. provide GUIDs (e.g., SICIs and BICIs) that cost nothing for articles, books, and book chapters, and provide a resolver for these GUIDs.
 3. provide a service to parse bibliographic citations (along the lines of Paratools).
 If you have 1, then you have a web service that people can use. For example, Connotea users can specify an OpenURL resolver to try and find an article online. By default it is CrossRef's service, which relies on DOIs. Imaging having a service that also supports literature that uses other GUIDs (such as Handles and SICIs). If you have 2, then everything in this project has a URI and therefore exists on the web (or, at least it's metadata does). This is one of the motivations behind the bioGUID project. If you have 3, then you have a service that would be useful in automatically integrating bibliographies from a range of sources (e.g., parsing lists of the literature on the web, or in bibliographies of papers). In terms of a portal, etc., in an age of Connotea this seems wasted effort -- they do this sort of thing well, why duplicate effort? Community tagging, etc., relies of communities, and Connotea already has one. The other lesson to learn from Connotea is that relies on existing web services, such as CrossRef's OpenURL resolver (to extract metadata associated with a DOI), and NCBI's PubMed service. By using

these services, the developers could concentrate on other things. What I am arguing is that if you develop services, the 'portal' (or whatever) becomes either (a) easy, or (b) largely irrelevant because you can use existing tools.

- I represent EoL. We have tools that will be provided to BHL for taxonomic indexing - and these can be extended to this project. Not only will this provide a taxonomic indexing service, but the indices can be placed in the context of any classification and reconciliation of alternative names will be an integral part of the indexing service. For more information eol@mbi.edu.
- How would the publishers of journals and monographs react to such a project? It would be essential to them to convince them that they get enough value selling access to printed and electronic contents to as many central and specializing libraries as possible.
- The number one priority for working taxonomists is the original literature describing taxa. This is crucial for those without access to world class libraries. So the priority is to digitize this literature not create search engines or literature databases. Just the raw published data please. Your priority at BM should be to make all descriptions of your types available with their images on your web site. Look to the models of Tropicos and Botanicus and AntBase - this is what working taxonomists need. Perhaps this is why these types of databases were created by practicing taxonomists, not computer jocks. Again please, I cannot emphasize enough that BM needs it's types databased with images on the web, not a computer bibliographic search engine.
- Any system must talk to those in use by the content providers (at least those using EDIT tools) to harvest bibliographic references at article level from these sites.
- Field conventions (content of each field, not only field type and length) should be published, so that local databases can adopt / convert to this format to open them the possibility to equalize data
- Boolean search ability A function that brings up lists of synonyms
- include varieties and forma

33. Any other requirements that you think desirable. - 8 responses (7 shown)

- Links to member/specialty/contact info lists of various organizations (e.g., American Society of Plant Taxonomists), so one can look up a specialist in one's group of interest. These lists are often helpful in tracking down the authors of articles or finding a specialist to discuss the literature
- For plants, incorporate the Taxonomic Literature ed.2 database, and ask IAPT (Vienna) about it. IAPT need to update and extend this data, which is not happening
- Many of the features I've ticked relate to a portal, although I personally think developing a portal is likely to be a mistake, or at least a lot less important than developing the underlying services
- Integration with EOL.
- The tool should act as a gateway to online resources linking the content providers (e.g. the Scratchpads) to the full text citations where it is possible to do this. Achieving this is beyond the likely scope of the project, but to my mind, this is the long term vision for E-ViTAL.
- Please have a look to: <http://www.gbif2.mwn.de/> <http://www.myriapoden-info.de/MyriaLit/index.html> we will be pleased to be informed about your progress
- Return results for author names resembling the name entered if the name is spelled strangely when entered and brings up no results on its own

Countries of respondents to questionnaire

Country	No. of responses
Australia	3
Brazil	3
Canada	3
Colombia	1
England	2
Estonia	1
France	1
Germany	15
Greece	1
Hungary	1
Iceland	1
India	2
Italy	1
Kenya	1
Netherlands	2
New Zealand	2
Scotland	2
Spain	1
Sweden	1
UK	6
USA	11
not declared	3