

newsletter 2

WP 2: Providing ENBI's Communication Platform (II)

WP 5: Report on assessment of existing electronic rich data sources

WP 6: Co-operation of pan-European databases on biological collections and specimens

WP 6: Become a GBIF data provider with ENBI !

WP 7: Observational survey data

WP 8: Grid-based federated query processing and collaborative networked environments

WP 9: Powerful new version of Linnaeus II

WP 11: Access to biodiversity information systems in major languages of the European Community

WP 13: Making non-European biodiversity data in European repositories globally available

WP 13: BeBIF and the European Network for Biodiversity Information (ENBI)

WP 13: Macaranga and Mallotus (Euphorbiaceae) species of Borneo

newsletter 1

newsletters

updated on: 09-Nov-2004

ENBI WORK PACKAGE 2

Providing ENBI's Communication Platform (II)

From: Marisa Esteban and Francisco Pando , Real Jardín Botánico (CSIC), Madrid, Spain

Announcing ENBI Forums forthcoming activities: 3 rd ENBI e-conference and 2 nd ENBI Workshop

ENBI Forums announces two main events for the first half of 2005, where managers of GBIF nodes, scientists and biodiversity data stakeholders in Europe will discuss and will exchange opinions, as in last ENBI Forums meetings, on issues of common concern for contributing to the GBIF implementation both in Europe and worldwide.

- *Third ENBI Electronic Conference* will be held from February 8 th to 23 rd . The concrete topics to be discussed in both events are not decided yet, but they will be announcing shortly to all ENBI community . You can find up-to-date information and subscribe to the e-conference at <http://www.enbi.info/forums/e3/index.php>.

- *Second ENBI Forums Workshop* will be held in May and this meeting will be derived from the issues raised during the third ENBI e-conference. You can find updated information and programme, as well as subscribe to the workshop at <http://www.enbi.info/forums/ws2/index.php>

(All suggestions for topics or improvements in the organization of both events will be welcomed. Please contact us: [enbiforums @ma-rjb.csic.es](mailto:enbiforums@ma-rjb.csic.es))

ENBI Second Electronic Conference: Data Validation and Restrictions in the GBIF network

The second ENBI *e-conference* took place from 9 th to 24 th March 2004 , under the title "Data Validation and Restrictions in the GBIF network". One hundred fifteen people participated to this conference and 47 messages were posted on the discussion list along of the e-conference period, apart from the 11 commented questions daily posted by Nicolas Bailly, moderator of the e-conference.

The *e-conference* was divided into three sections, the first dealing with how to assess and communicate "quality indicators" on data; the second with possible restrictions that may be applied for protection of sensitive data, and to avoid doubtful data that may be misused, and, finally, how protect Intellectual Property Rights (IPR) of data and portal providers. Those topics were further discussed in a two days third session.

Full contents of this second ENBI *e-conference* are available at ENBI Forums (WP2) web pages (<http://www.enbi.info/forums/ec2/index.php>) and ENBI Community Library at CIRCA (<http://circa.gbif.net/Public/irc/enbi/econ2/library>).

First ENBI Forums Workshop: Promoting Access to Biodiversity Information: balancing needs of users and providers

The first ENBI Forums workshop took place from 25 th to 28 th May 2004 , under the title "Promoting Access to Biodiversity Information: balancing needs of users and providers". The topic for this First ENBI Forums Workshop derived from the issues raised during the First and Second ENBI e-conferences: Open-access for biodiversity information (8 - 28th September 2003) and Data Validation and Restrictions in the GBIF network (9 - 24th March 2004), respectively. The workshop gathered together 29 participants from 15 countries. Worthwhile to mention was the strong participation from "EU new member countries". Their presentations proved to be an eye opening experience, and illustrated quite well the different approaches and needs of these countries in relation to GBIF. The workshop also provide the means for mixing and exchange ideas among experienced participants in GBIF and not-so-experienced. The participation of Donald Hobern --GBIF officer for Data Access and Database Interoperability-- was especially relevant . Donald Hobern provided first-hand information on the architecture of the GBIF network, future directions and related issues. Besides, His input in the discussions was very helpful and important factor in the overall success of the meeting.

The presentations and reports of the workshop are available at ENBI Forums (WP2) web pages (<http://www.enbi.info/forums/ws1/index.php>) and in ENBI-CIRCA Community Library at (<http://circa.gbif.net/Public/irc/enbi/ws1/library>).

A road map of ENBI-CIRCA Virtual Community Library

CIRCA is the main web-based collaboration environment used by ENBI. It enables to share all contents and discussions of the ENBI community through different sections called 'Interest Groups' (IGs) <http://circa.gbif.net/Public/irc/enbi/Home/main?index> . At present, the ENBI-CIRCA space comprises 344 members, and 16 IGs. Every

IG has a directory of members and a library, with documents, images, links, etc. Although not all these documents are ready to be put in the public domain (internal administrative documents, preliminary reports of work packages, rough copies of current studies...), there are already an important collection of public documents and work package deliverables accessible from ENBI Community Library IG.

A road map of this ENBI Community library follows:

- About ENBI (*Document*)
- ENBI brochure
- ENBI & GBIF (*Document*)
- ENBI-GBIF meeting January 2003
- ENBI e-conferences (*Links*)
 - First ENBI e-conference: Open-access for biodiversity information (10 - 28 September 2003)
 - Second ENBI e-conference: Data Validation and Restrictions in the GBIF network (9 - 24 March 2004)
- ENBI history (*Documents*)
- ENBI meeting 23/03/01 Brussels
 - i. List of Participants
 - ii. Summary of conclusions + tasks
- ENBI meeting 02/07/01 Stockholm
 - i. Participants Stockholm meeting
- ENBI meeting 17/09/01 Warsaw
 - i. Warsaw agenda
 - ii. Warsaw participants
- ENBI Pre-call
- First ENBI presentation ICSEB Patras, Greece , 2002
- Proposal part B
- Proposal part C
- ENBI Newsletters (*Document*)
- ENBI Newsletter 1
- ENBI on-line deliverables (*Links*)
- Lepindex: Demonstrator of an on-line card archive database
- Biodiversity databases and database projects
- ENBI Glossary
- ENBI reports (*Documents*)
 - WP5: Report on the assessment of existing electronic rich data sources ('Species Bank')
 - WP7: "Observations on observational data "
 - WP8: Final Report on Distributed Information Management Requirement Analysis
 - WP9: Analysis report on common interoperability tools, and on user requirements for common access points
- ENBI web sites (*Links*)
- ENBI
 - ENBI Forums (work package 2)
 - Observational Survey data (work package 7)
 - Multi-lingual Access to European Biodiversity Sites (work package 11)
- ENBI workshops (*Links and documents*)
- Making Species Databases Interoperable-An advance workshop

- i. The meeting programme
- ii. A map of the Reading campus
- iii. Information for Participants
- iv. Making Species Databases Interoperable-An advance workshop – Announcement and preliminary programme
- Techniques and challenges for digital imaging of biological type specimens
- i. Draft programme
- ii. Report
- First ENBI Forums Workshop (Pruhonice, 25-28 May 2004)
- Workshop Report on: Multi-lingual Access to European Biodiversity sites
- WP13: Making non-European biodiversity data in European repositories globally available - Report of first workshop
- Introduction to ENBI-CIRCA (*Link and documents*)
- Guidelines for IG leaders
- Introduction to CIRCA for end users
- Introduction to CIRCA for Interest Group leaders

A clickable version of this road map is available on line at <http://www.enbi.info/forums/ig/repository.php>

ENBI WORK PACKAGE 5

***Report on assessment of existing electronic rich data sources
(‘Species Bank’)***

From: Pamela Harling

Available in the CIRCA library at

http://circa.gbif.net/Public/irc/enbi/comm/library?!=/enbi_reports/

The Global Biodiversity Information Facility (GBIF) has the objective of working towards a complete electronic catalogue of the names of all known species. It is working closely with Species 2000 and ITIS (Integrated Taxonomic Information System) to put together the Catalogue of Life. This report reviews the possibility of linking this catalogue in some structured way to the multitude of electronic data sets relating to species. A review is given of how the species bank architecture will be distributed or federated, so that different sectors of the biodiversity information community can contribute their knowledge to a coherent system.

The different types of rich data available on the internet and main sources are outlined. The main classes of rich data are: descriptive structural data, images and sounds, distribution data, bibliographic data, ecological data, biological data, human dimensions, conservation and inter-relations between taxa.

The main sources of rich data currently available in digital form are: global species databases, pan-European databases, regional biodiversity databases and specialist sector databases.

The final section of the report is a survey of information currently available on the web for four major groups – the European lichens, legumes, butterflies and fishes. This information is presented in a series of supplements, giving screen shots of relevant web pages.

ENBI WORK PACKAGE:6

Co-operation of pan-European databases on biological collections and specimens

Malcolm Scoble

There are two major components in WP6: developing standards and best practice for digitising type-specimens; and organising, integrating and broadening the network of European collection databases.

An account of the second of these main tasks is provided by Walter Berendsohn elsewhere in this newsletter. But the main points to note from a workpackage coordinator's perspective is that as the network grows it is already having a binding effect on the community of data providers. This effect will increase as more databases are networked. Most successful developments have a history, and it is pleasing to see that this history lies within EU-funded projects that have stretched through BioCISE, ENHSIN and BioCASE to ENBI. It is the determination of Walter Berendsohn through this line of projects that has been the driving force behind what has been achieved.

A successful workshop to examine standards and best practice in digital imaging of type specimens was held in Stuttgart in March 2004 (for report see). This meeting was held jointly with GBIF, a fact that underpins a fundamental theme of ENBI, which is to add European value to the international biodiversity position.

How does Workpackage 6 fare in the broader context of ENBI, specifically with respect to coordination and sustainability of the tasks? The distributed network of databases has a coordinating effect by its very existence: the more databases that are linked to the system, the greater will be the degree of coordination. The momentum that has been gained through the growing linkage means that we may be optimistic about the sustainability of the network. This point is underscored by the close connections between this project with GBIF, as reported by Walter Berendsohn.

The planned document on standards and best practice for databasing and digital imaging of type specimens, will also have a coordinating effect. It should, however, be noted, that the considerable interest being shown in this topic across the taxonomic/biodiversity community means that techniques will continue to develop – probably rapidly. The importance of the field to the biodiversity informatics community is such that it is almost inevitable that interest and activities will be sustained in the future.

ENBI WORK PACKAGE 6

Become a GBIF data provider with ENBI !

From: Javier de la Torre and Walter G. Berendsohn

Work package 6 of ENBI is titled "Co-operation of pan-European databases on biological collection and specimens" and led by the Natural History Museum in London . It comprises efforts to contribute to data capture and imaging of specimens through exemplary action and through development of standards and best practices. Another focus is networking of specimen information. The task of the Botanic Garden and Botanical Museum Berlin-Dahlem (BGBM) within the workpackage is to extend the GBIF network of access to specimen data in biological collections.

Up to now a total of 129 databases covering more than 250 collections were connected to the network through 32 data providers in 9 countries. In addition, 7 providers did already install the software and another 9 have already agreed to connect their databases. This result in more than 4 million records already made available to GBIF. Data access to connected collections is possible through portals developed by the GBIF secretariat, BioCASE , by national GBIF Nodes, and by individual providers.

The BGBM's tasks in ENBI are synchronised with those in other projects, most notably the BioCASE project (Biological Collection Access Service for Europe), which is coordinated in Berlin . The European networking efforts are based on software developed by BioCASE and one of its predecessors, ENHSIN (European Natural History Specimen Information Network, an EU infrastructures project lead by the Natural History Museum in London, 2000 to 2003). The BioCASE software, which succeeded the ENHSIN prototype, was significantly improved with respect to user-friendliness and documentation during the first months of the ENBI workpackage. The entire software is open source and available on the web through the BioCASE website.

The BGBM's tasks in ENBI are synchronised with those in other projects, most notably the BioCASE project (Biological Collection Access Service for Europe), which is coordinated in Berlin . The European networking efforts are based on software developed by BioCASE and one of its predecessors, ENHSIN (European Natural History Specimen Information Network, an EU infrastructures project lead by the Natural History Museum in London, 2000 to 2003). The BioCASE software, which succeeded the ENHSIN prototype, was significantly improved with respect to user-friendliness and documentation during the first months of the ENBI workpackage. The entire software is open source and available on the web through the BioCASE website.

GBIF recognises and incorporates both the BioCASE technical communication protocol as well as the data definition (ABCD) used by BioCASE, they form parts of the GBIF information infrastructure. ABCD (Access to Biological Collection Data) is a comprehensive proposed standard aiming to cover all data items relevant to collections. It make it possible to create specialised networks while maintaing overall compatibility of data sources. ABCD is formulated as an XML schema and was developed by a CODATA/TDWG Task group with the same name. After incorporating many comments from different collection communities a new version is now available and proposed for acceptance as a TDWG (Taxonomic Databases Working Group) standard.

Within ENBI and as a task of a forthcoming networking activity in the SYNTHESYS project (running until 2009), the BGBM is providing helpdesk functions, creating proper documentation, preparing workshops and contacting as well as visiting data providers with the aim of extending the network. If you have a database with specimen or observation data, your institution has a web server, and you wish to become a participant in the global and European networks, ENBI offers technical assistance. We can also offer to organize local workshops to adress several data providers at once.

Mobilizing data is a currently the critical point in the success of the GBIF initiative and and consequently of the European Network of Biodiversity Information. We welcome any communication regarding contacts to people or institutions that are interested in connecting their databases. We will do our best to make the process easy and successful.

www.biocase.org
provider/ENBIhelpdesk/

[http://www.biocase.org/
provider/ENBIhelpdesk/](http://www.biocase.org/provider/ENBIhelpdesk/)

Email: j.torre@bgbm.org

Phone: +49(30) 838-50284

BioCASE: www.biocase.org

ABCD: [http://www.bgbm.org/
TDWG/CODATA/Schema/](http://www.bgbm.org/TDWG/CODATA/Schema/)

ENBI WORK PACKAGE 7

Observational survey data

From Marja Vieno and Tuuli Toivonen, University of Turku

Primary data on biodiversity lies not only in the biological collections in various museums and institutes and private collections throughout the world. In addition to these collected material there is a huge amount of biodiversity data - mainly on species level - in a variety of observational databases. These data can extend back tens of years and be spatially extensive, thus providing an opportunity for monitoring trends in biodiversity. Observational datasets are, however, very scattered and diverse in their content. The way observations are made, recorded, stored or disseminated varies - depending largely on the purpose for which they are made. The majority of them are based on the work of volunteers having a keen interest on the observed species or group, and the results either end up in private note books hidden in cellars or end up as part of national or even international monitoring programmes which are used for instance in decision and policy making, scientific work, land use planning, conservation work just to mention a few. Owing to scatteredness of and access difficulties to these data, the full potential of the observation databases can only be estimated.

To tackle this whole issue of observational data the main objectives and deliverables have been formulated as follows: We aim to collect a European network of organizations, institutions and specialized interest groups that hold, analyze and present spatial, observation -based biodiversity data, secondly to document standards and best practices for processing and making available observational biodiversity data, thirdly to analyze observational biodiversity information needs and products to support the European Biodiversity Strategy and parallel activities, fourthly to define methodologies linking taxonomic, collection and specimen databases with the observational data and finally to prepare a document that will demonstrate the applicability of combined GIS and remote sensing approach as a means for spatial biodiversity data analysis and presentation.

WP7 has its own web-pages available at: <http://enbi.utu.fi>

WP7 has participated in the implementation of the GBIF demo project for year 2003. On its website, the demo introduces GBIF and the questions related to utilization of collection and observation data available in the GBIF network.

The demo presents four tours, that

- * Combine observational and collection databases interactively to interactive map services.
- * Highlight the possibilities of geographic visualization of species distributions and spatial queries of databases
- * Give user oriented examples on the power of data combination and visualization

The tours can be browsed in two ways: hands-on tours direct the users to Internet based map services that present georeferenced species data together with environmental data layers and appropriate tools. Quick-look tours give a possibility to see the examples with step by step pages, without accessing the map service.

GBIF demo site is available at <http://gbifdemo.utu.fi/>.

ENBI WORK PACKAGE 8

Grid-based federated query processing and collaborative networked environments

from: Ersin C. Kaletas, H. Afsarmanesh and O. Unal

Within the WP8, the *Co-Operative Information Management* (CO-IM) group of the University of Amsterdam (UvA) is investigating the application of Grid-based federated query processing and collaborative networked environments (Virtual Organizations) to the support of integration and interoperability among distributed biodiversity nodes.

In this respect, a number of concepts, technologies, and paradigms have been investigated to address the information management challenges in ENBI. Among these, **federated database architecture** provides the necessary mechanisms to access and integrate data from distributed, heterogeneous, and autonomous centers. The **Grid technology** is emerging as a promising solution for sharing and management of high performance computing, storage and networking resources, especially in the scientific areas. Finally, the **Virtual Organizations** (VO) paradigm is proposed as solution to emerging collaboration needs of organizations from a variety of domains, by regulating the collaboration through a set of rules and policies agreed by all the partners.

Each of these technologies and paradigms offers a large number of benefits to the biodiversity domain and meets many information management challenges and requirements in this domain. When used together, they would provide even higher value to further support the biodiversity conservation activities. The combination of these technologies, paradigms, components, and supporting tools constitutes a **Collaborative Information Management System** (CIMS).

The results of the work carried out by UvA in ENBI - WP8 are reported in a number of deliverables; namely **D8.2a** (PUBLIC LINK TO THE DELIVERABLE WILL COME HERE) discusses the potential of Grid, while **D8.2b** (PUBLIC LINK TO THE DELIVERABLE WILL COME HERE) discusses the potential of VOs. These two deliverables are available to the biodiversity community at CIRCA. The third deliverable **D8.2c**, which is at the internal review process at the moment, brings together all the above mentioned technologies under a CIMS framework, and provides a CIMS integrated with a Virtual Laboratory environment as an example.

ENBI WORK PACKAGE 9

Powerful new version of Linnaeus II

from: Peter H. Schalk

The ETI Bio-Informatics Center released a powerful update (version 2.5) of its Linnaeus II content management system for taxonomic and biodiversity information. It accommodates five years of valuable feedback from taxonomists worldwide and its development was sponsored by the Netherlands GBIF Node and several international R&D projects. Linnaeus II is in use by some 1,500 taxonomists and supports the creation of taxonomic databases, computer aided identification keys (digi/polytomous, image-based, matrix-based), interactive (searchable) geographic maps, and supporting literature and glossary databases in a single, hyperlinked environment. The new 2.5 version of the software runs under Windows XP (and 98, ME) and Macintosh OS X (and 8.6, 9.x) and is designed to support the creation of 'SpeciesBanks' in the sense defined by the Global Biodiversity Information Facility (see www.gbif.org).

The Linnaeus II software consists of three modules:

- Linnaeus II 'builder' vs 2.5: to create taxonomic information/identification systems on his/her own computer,
- Linnaeus II 'runtime' vs 2.5: to publish completed Linnaeus II systems in the form of 'e-monographs' on mass produced CD-ROMs/DVD-ROMs.
- Linnaeus II 'web publisher' vs 1.0: to publish and use Linnaeus II information systems on the Internet.

Scientists at universities, research institutions and taxonomic institutions can download the Linnaeus II builder free of cost from our website:

<http://www.eti.uva.nl/Support/Download/Linnaeus.html>

The only requirement is to register as an 'ETI Partner', which entails that the licensee intends to contribute to creating digital taxonomic information systems. The user license allows the licensee to work with the Linnaeus II software for personal / institutional use. Those who are already partner are kindly asked to re-register.

ETI offers its partners a (free) e-publication programme for completed electronic monographs built in Linnaeus II format. These are published as part of the World Biodiversity Database CD-ROM Series. More than 80 CD-ROM have been published. A list of published Linnaeus II e-monographs can be found on our website: (<http://www.eti.uva.nl/Products/CD-catalogue.html>) and can be purchased at: (<http://www.etiis.org/>). Linnaeus II e-monographs are made available against the nominal costs of production; no profit margin is included. Authors receive royalties, retain full ownership of their data, and are entitled to (re)publish their data in any form at any time. The license to mass produce CD-ROMs using the Linnaeus II run-time version can also be available to third parties. Please contact ETI for details.

Exciting is the new development of the Linnaeus II 'Web Publisher' tool, which supports the full functionality of the Linnaeus II runtime software in a web environment. Resulting websites can be customized with features such as knowledge tests (quizzes) and a thumbnail index. The 'biodiversity information system' on the website of the Netherlands ' GBIF Node: http://www.nlbif.nl/Services/search_species was built using the beta version of the Linnaeus II WP tool. The web publisher tool was developed to meet the requirements of a growing number of taxonomists and assists the users to 'publish' their taxonomic information system directly on the Internet themselves. For user licenses to run the Linnaeus II Web Publishing software on local server(s) please contact ETI for details (info@eti.uva.nl).

In case you experience any problems with the download or if you have any questions please do not hesitate to contact us.

Dr. Peter H. Schalk
Director ETI Applied Informatics
University of Amsterdam
Mauritskade 61
1092 AD Amsterdam
The Netherlands
Phone: 31.20.5257239
Fax: 31.20.5257238
e-mail: pschalk@eti.uva.nl
<http://www.eti.uva.nl/>

ENBI WORK PACKAGE 11

Access to biodiversity information systems in major languages of the European Community

from: Bernd Ueberschaer

WP 11 is part of ENBI, an EC supported Thematic Network accommodating 65 European institutes representing 24 countries. ENBI's main objective is to establish a strong network that will identify biodiversity information priorities to be managed at the European scale. In that context, it was considered to be an important and worthwhile effort to provide the technical framework how to enable access to biodiversity information systems in major languages of the European Community (French, Portuguese, Spanish, Dutch, German, Italian, Greek). From surveys it was known that language matters in which biodiversity information are available in the Internet, thus it was supposed that translation into major languages of biodiversity information systems will help users such as decision makers, politicians, manager and the public in general from countries where English is not the native language to access biodiversity information in their own language.

Globalisation urges multi-lingual web content....

Although English is the language of globalisation, it is estimated that by 2050 probably half the world will be more or less proficient, there is no doubt that there is a need for the next decades to present Internet content in other major languages than English. At present, it is estimated that 85% of the Internet's content is in English, but about 45% of Internet users today cannot read English at all (on a global scale).

At present the Internet can be counted in hundreds of millions of pages, and it is growing exponentially at a very high rate. However, it is expected that the non-English speaking web users will soon outnumber the English-speaking users. Thus, it is no longer enough to translate local web sites only to English. In 2005, one expects the Web to reach one billion users and even 70% of them will be non- English speaking. It means that much effort has to be put into localisation of existing web sites and into the creation of new multilingual services, since it is certain that most web users prefer to be addressed in their native language, at least at the top-level pages of services no matter how flawed and error-ridden it may be, rather than to struggle to understand a foreign language text. Users, who are addressed in their own language, will stay at a site twice as long (more information about this topic in: *Still a Challenge: Machine translation (MT) in the 21 st century*. ENBI newsletter No. 1, download at <http://www.enbi.linguaweb.org/Downloads/ENBInewsletter1.pdf>).

The request for information in FishBase showed a considerable increase in user hits specifically from developing countries since the first version of machine translation for selected resources (search site, species summary) was established (Fig. 1).

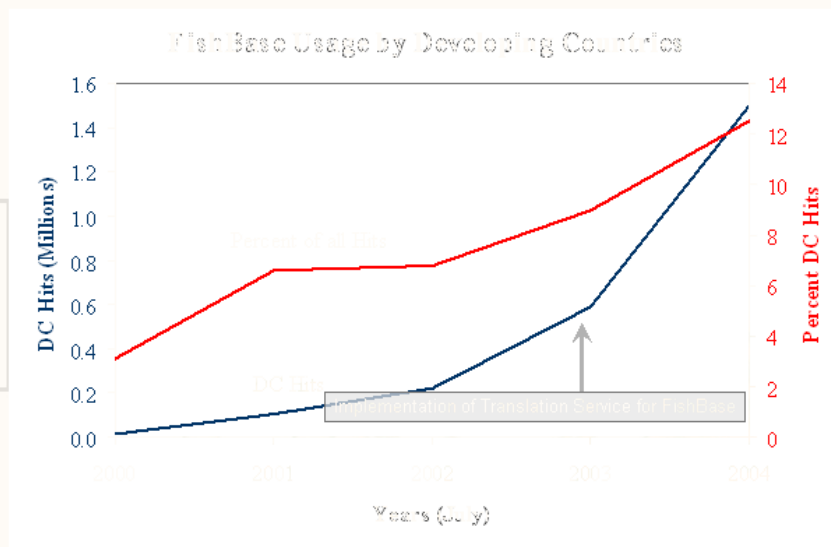


Fig. 1: Substantial increase in hits to FishBase from developing countries (DC) since a preliminary translation service from English into 7 other major European languages was established in November 2003.

Interactive Machine Translation: No alternative choice for dynamic website content.

The approach how to deal with translation of Internet resources depends on their nature. Since Internet content is a very dynamic issue, manual translation is hardly an option, specifically for sites which have typically a dynamic content with many information being updated in short intervals. Global information systems such as FishBase (www.fishbase.org) are a typical example for those dynamic sites. Consequently, apart from some static translation, machine translation was chosen as the major mean for the compilation of translation techniques and strategies in WP 11.

The Systran® translation engine (<http://www.systransoft.com>) is being considered to be the "State-of-the-Art" system at present for machine translation and was in favour for the realisation of the multilingual access in ENBI. However, since the desktop version of Systran® is useful for the basic evaluation of the process of machine translation, it is not appropriate for the translation of website content "on the fly". For those purpose (the user open a website, wants to see the content in an other language, clicks on the related link, e.g a flag in the desired language which is shown in the page, and a translated page is returned to the user within seconds) the company Systran® offers SYSTRAN@Links, which is a turnkey website translation solution. SYSTRAN@Links transforms standard websites and content applications into interactive multilingual hubs. SYSTRAN@Links offers all major European, Asian and Russian languages. However, the fee for this commercial service (from c.a. 25,000 US \$) is out of reach for the most biodiversity websites which are in the public domain and maintained by non-profit organisations.

EC-MT-Department has unlatched access to their translation service for WP 11

The EC has developed its own MT-System since the 70's, based on the Systran® engine (EC-Systran®) for many language pairs. Since this service can basically be used free of charge for non-profit projects in the European dimension, the ENBI project requested permission to be connected to those system, and ENBI and the EC-Translation Department came to an agreement which allows the ENBI project to use their system in order to establish a Website "on the fly" translation for selected biodiversity information systems in the Internet. However, at the date of commencement of ENBI, this service was not yet established and it took more than a year of further technical developments to establish this service. Some months ago, the MT-Department has eventually opened access for WP 11 to their translation service for websites. Although the access is still restricted, the access to real-time translation from the EC-Systran System it is being considered as a major breakthrough for the goals of WP 11. Moreover, facilitated by a very helpful and efficient personal contact between the EC MT-Department (Cameron Ross) and the WP 11 project manager (Bernd Ueberschär) the MT-Department has started to implement the special dictionaries which were compiled from WP 11 translation partner in 7 languages on the basis of missing translations in general dictionaries which were tested for the content in certain biodiversity information systems (e.g. FishBase). Since the results of the machine translation depends to a large extent on the availability of correct translated terms (e.g. "stock" in the business world has a complete different meaning from "stock" in relation to biodiversity information), three major dictionaries with specific terms are being hand-coded now from the technical team in the MT-department (noun, proper noun, verb, adjective etc.) and tapped to the EC-Systran® System under consideration of an assigned category (e.g. "Biology", "Fisheries", "Environment" etc.). These three major dictionaries "Biology", "Morphology" and "Distribution" contain more than 3000 specific words which do not appear in general dictionaries; in addition the system is being fed with about 20,000 scientific names which are not to be translated, but advises the translation engine to pay attention to the original name.

Realistic online testing under consideration of specific dictionaries for specific categories has begun.

The access to the EC-MT system either for single words, sentences or whole paragraphs and complete websites (just the given URL is necessary) has opened now the opportunity for realistic testing procedures, which is going to be practised now by all translation partner. This online-testing procedure is very helpful to monitor the effect of growing special dictionary resources in the categories which are needed for the translation of biodiversity information as well as to monitor the impact of source text structure (... "telegraphic style ruins the machine translation...") on the translation results. With this option, specific demands of each target language in relation to the source text in English can be elaborated from the home office of each translation partner.

Once the EC-Service is open for the public and implemented into the various information systems, one single click only on the flag is needed to get the translated website displayed within seconds on the users' screen (Fig. 2). European biodiversity web sites can avail of this service in the future by showing a "Translate" button or flags on their pages (presumed specific dictionaries are delivered). The cooperation between ENBI and the EC-MT-Department has good prospects concerning the improvement of MT for biodiversity information in the Internet. As considered above, the quality of MT considerably depends on the customized activities. Since ENBI has created special biodiversity dictionaries which are going to be integrated in the machine translation service of the European Commission, a good result can be expected after some revisions were applied. Nevertheless, machine translation is not considered to be comparable with manual translation, keep always in mind what was expressed by Brian Garr, (IBM):

"Machine translation is a viable technology that can have good value. You just need to set your expectations properly so you get the most out of it."

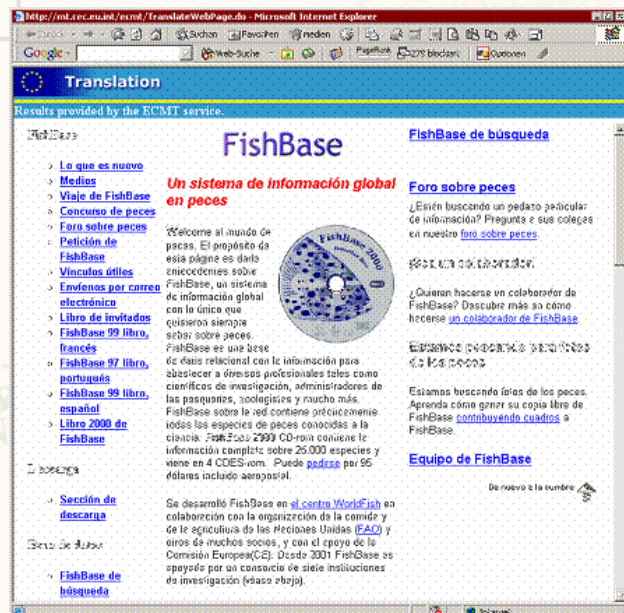


Fig. 2 : Homepage of FishBase translated from English into Spanish by means of EC-Systran® machine translation (no further editing applied). The headline indicates the ECMT-Service.

This is the first time, that machine translation is going to be implemented for biodiversity information systems in the Internet. So far this project was quite successful, and the ENBI consortium strongly recommend to use the experience of this project and to take the chance to consider machine translation for other systems beyond the trial systems of this project (e.g. FishBase). In the next years, the languages of Eastern Europe will be added to the EC-MT system, and work has already begun on Czech and Polish, and with those new EC-member countries, the demand for MT will increase tremendously.

Other News from WP 11:

Explore the options to build a prototype biodiversity glossary that can be used by European biodiversity web Sites

Four, partially multilingual, public domain glossaries in the Internet, related to the subject biodiversity, were "deep-linked" to FishBase. Search in 4 additional Internet glossaries beyond the genuine glossary of FishBase is now feasible. Tapping of other glossaries is being considered beyond the termination of ENBI. Other online glossary approaches as e.g. Wikipedia, a free-content encyclopedia in many languages, will be evaluated as an option for a glossary portal on biodiversity terms in the Internet. Manual translation of the FishBase glossary of terms from English into German is an ongoing effort.

Explore the options to significantly improve access to biodiversity information through vernacular names

Common names of species have a long history, far exceeding that of scientific names, and often, common names are more stable than scientific names. Since the public as well as biologists use those names frequently, a standardized and easy access to common names in major languages in the Internet is desired in the modern world of electronic information systems (Fig. 3). The ENBI project suggested to explore the options for an improved and standardized access to common names in the Internet. As a result and as a first practical respond in that context, it was agreed to enter all common names from the FAO compilations on common names into the Species 2000 system. The NGO "FIN" (FishBase Information and Research Group, Inc.) has the necessary technical skills to deal with this task and was contracted from WP 11 to encode, in a joint effort with the FishBase team (WorldFish), the FAO lists in 8 languages (English, German, Dutch, Spanish, Portuguese, French, Greek, Italian). This task was started in August 2004 and will be completed in February 2005. Any information system in the Internet should be able then to link to those Interface to common names beyond fish.

About 10,000 visitors per month

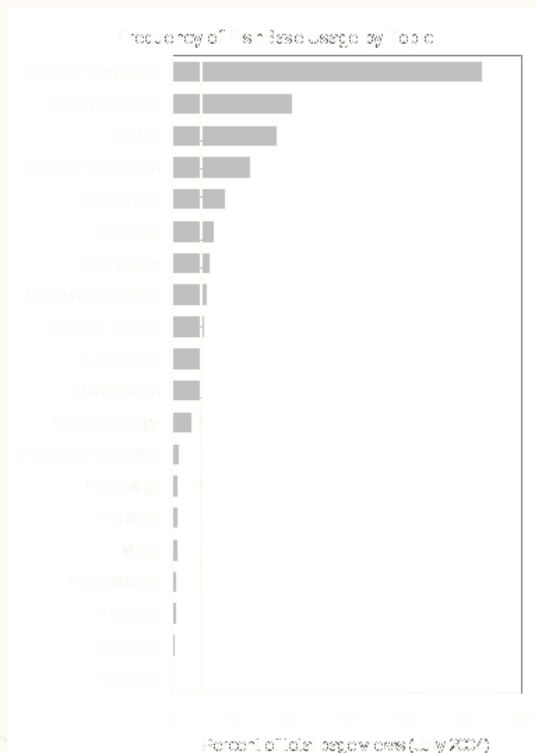


Fig. 3: User request for resources in FishBase. The "common names" search interface is the second most used resource in FishBase.

2nd Translation Workshop conducted at IfM-GEOMAR in Kiel , Germany

A second interim workshop on "Multi-lingual Access to European Biodiversity Sites" was successfully conducted from the 20. – 21. September 2004 at the Leibniz-Institute of Marine Research, Kiel , Germany , with translation partners from six European countries. The workshop had a rather pragmatic approach to machine translation, taking advantage from the online access to the EC-MT-Systran® translation system. Major issues were to test the effects of source text variation on the translation results, to compile guidance for source text in various languages and to monitor the effects of the user dictionaries which are being implemented into the EU-MC Systran® system. A detailed workshop report will be available for download soon at the WP 11 website www.enbi.linguaweb.org .

ENBI WORK PACKAGE 13

Making non-European biodiversity data in European repositories globally available

From: ISABEL CALABUIG

Work package (WP) 13 is the element of the ENBI programme that deals with making non-European biodiversity data in European repositories globally available. Although the European Network of Biodiversity Information serves as a framework for the scientific bio geographical coherence of Europe, there is scope within the activities of ENBI to reach beyond this region. Notably, because many European countries have a colonial past, a large share of global biodiversity information, especially from developing countries in the tropics resides in databases of European institutions (including museum and botanic garden collections). With a still very fragile infrastructure for taxonomy in tropical developing countries and a very limited number of well-trained taxonomists, there is a great need for high quality collections-based information among local (para)-taxonomists, conservationists and other users of floristic and faunistic information. Data sharing may take the form of making available (visible, downloadable, analysable) on the Internet, current names, digitised images and collecting data of selected specimens, for which there is an urgent information need in the developing countries.

First feasibility study of Work Package 13 goes online:

One of the deliverables listed within WP 13 is a number of feasibility studies that are meant as test cases for how to share biodiversity data between European repositories and the countries of origin. The final evaluation of proposals for such feasibility studies took place at a workshop in Tervuren, Belgium, 30-31 October 2003, hosted by the Royal Museum for Central Africa. Participants at the workshop included the WP co-ordinators, partners, invited experts, and associates. A very important aspect of the workshop was for experts on biodiversity data-use from developing countries to discuss with the European data providers the proposals in order to ensure the highest possible quality and usefulness of the projects. Besides the selection of feasibility studies the workshop had the purpose also to discuss issues such as user needs in data sharing. Four feasibility studies were selected to be carried in 2004, and we are happy to announce that the first of these is now completed and accessible on the internet:

Project title:

Quantification of ecosystem disturbance in the West Malay Archipelago with the aid of Mallotus and Macaranga plant species (Euphorbiaceae)

Responsible ENBI partner:

Peter C. van Welzen, Nationaal Herbarium Nederland - Leiden Universiteit Branch

URL: <http://www.nationaalherbarium.nl/macmalborneo/index.htm>

The website includes features such as:

- Descriptions and photos of all Mallotus and Macaranga species of Borneo
- A dichotomous key (with 1000 pictures of characters) to the species
- A spot character key
- Literature references
- Links to the Malesian Euphorbiaceae website and to the Euphorbiaceae specimens in the NHN-Leiden.
- Calculation of the type of disturbance by numbers of individual plants per plot

Announcement on URL details etc. on the three remaining feasibility studies will follow as they go online by the end of this year. The titles of these upcoming projects are:

- Prototype Image Server to Integrate the Martius Herbarium and the Digital Flora brasiliensis – by the National Botanic Garden, Meise, Belgium

- Providing Access to Albertine Rift Biodiversity Data – by the Royal Belgian Institute for Natural Sciences, Vertebrate Department.

- Afrotropical Ceratitidine Fruit Flies, A queryable website on recognition, distribution, and host plants – by the Royal Museum for Central Africa, Department of African Zoology, Tervuren, Belgium.

The feasibility studies will be evaluated and responses from users, adjustments to the sites and continued activity discussed at a second workshop ultimo 2004/primo 2005.

Report finalised on existing information on user needs from programmes, institutions, and other sources

Draft report available for ENBI WP13 members at:

<http://circa.gbif.net/Members/irc/enbi/wp13/library?l=/reports&vm=detailed&sb=Title>

Final public version submitted to ENBI administrator, Amsterdam , and being uploaded shortly .

This ENBI report deals with **non-European user needs for biodiversity data in European holdings** with the aim to provide information on the needs for biodiversity information by various non-European user groups. The report focuses on what type of biodiversity data end-users would like electronically shared and what kind of data output they would prefer.

Important questions include:

- Who are the various user groups (e.g. scientists, policy makers, pest management agencies)?
- What are their specific needs/ requirements (e.g. species identification, species lists) for biodiversity data?
- What types of biodiversity data can fulfil their needs?

Focus will solely be on non-European user needs, as work package 12 of ENBI will deal with user needs for biodiversity information by European users. The report will not deal with technical parts of electronic data sharing such as data standards, protocols, software etc. since other ENBI work packages (WP8 "Data management in large distributed biodiversity database systems", WP9 "Interoperability and common access" and WP10 "Generic analysis tools and data mining") deal with these issues.

Report finalised on existing models of active data use

Draft report available for ENBI WP13 members at:

<http://circa.gbif.net/Members/irc/enbi/wp13/library?l=/reports&vm=detailed&sb=Title>

Final public version submitted to ENBI administrator, Amsterdam , and being uploaded to the ENBI CIRCA library shortly.

This ENBI report (Deliverable 13.4) deals with **active use of biodiversity data** regarding how end-users use existing biodiversity databases and how those databases present end-users with biodiversity data. The report aims to:

- Analyse various aspects of some existing biodiversity databases to provide useful information for new data sharing projects,
- Acquire information on how end-users use existing databases to illustrate the variety of purposes biodiversity databases can be used.

In addition, the report was initially supposed to deal with existing data-sharing activities that share biodiversity data electronically with the country of origin in order to extract the lessons learned from existing data-sharing ("repatriation") activities. However, as the Outreach and Capacity Building Programme of GBIF planned to produce a "white paper" on data sharing simultaneously with the launch of WP13 of ENBI, we focused solely on active data use in order not to duplicate efforts. Instead, WP13 collaborated with the Outreach and Capacity Building Programme of GBIF on all three areas (user needs, data use and data sharing), especially on data sharing. Thus, for information on data sharing consult the GBIF report "Study on Data Sharing of Countries of Origin" presented below.

Report on existing mechanisms of electronic data sharing - GBIF report on Data Sharing with Countries of Origin

A report commissioned by GBIF and prepared by CRIA is available at: <http://www.gbif.org/prog/ocb/sdco>

The report presents the experiences of primarily developed-world institutions in the process of sharing data with countries of origin of specimens the institutions house. The ENBI work package 13 coordinators collaborated with the GBIF programme officer for outreach and capacity building on the planning of this important report.

ENBI WORK PACKAGE 13

BeBIF and the European Network for Biodiversity Information (ENBI)

From: Patricia Mergen

Important collections of the three major Federal Scientific Museums, *the Royal Museum for Central Africa (RMCA)*, *National Botanic Garden (BR)* and *the Royal Belgian Institute for Natural Sciences (RBINS)*, have been chosen as pilot projects for the ENBI Work Package 13, Making Non-European biodiversity data in European repositories globally available, feasibility studies.

More precisely the following projects have been selected:

- Providing Access to Albertine Rift Biodiversity Data,
- Afrotropical Ceratidine Fruit Flies (Diptera: Tephritidae) a queryable website on recognition, distribution, and host plants
- Prototype image server to integrate the Martius Herbarium and the Digital *Flora Brasiliensis*

The three project leaders decided to work under common GBIF standards and part of the EU funding of these three projects were attributed to BeBIF to fulfill some IT tasks related to these projects. Data concerning these projects have already been delivered to GBIF (<http://www.GBIF.net>) and can also be consulted on our national Portal and GIS server (<http://www.bebif.be>). Digital images and related information concerning African Butterflies are already partially available. For each species a male and female specimen is showed in a recto and verso view (Figures 1).

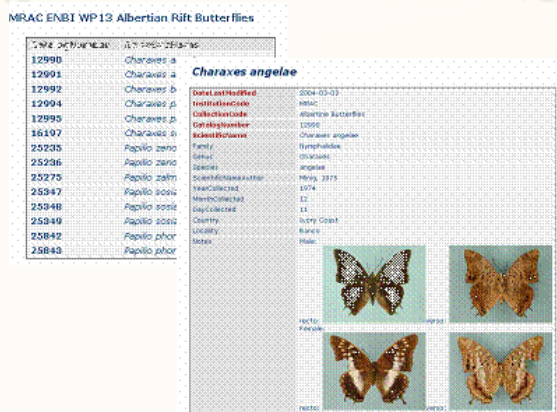


Figure 1 : African Butterfly specimens part of the “ Providing Access To Albertine Rift Biodiversity Data” project (data provided by Prof. Ugo Dall'Asta, MRAC, Belgium)

Information about cichlid fish from the Lake Tanganyika were our first collection connected to GBIF using the Access to Biological Collection Database (ABCD) Schema. Sampling locations along the shore of Lake Tanganyika can be nicely visualized on our GIS server (Figure 2).

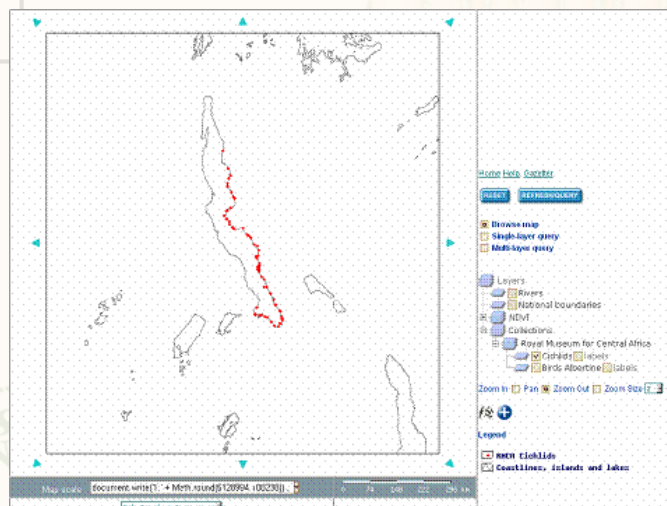


Figure 2: Sampling points of Cichlid fish along the shore of Lake Tanganyika part of the “ Providing Access To

Albertine Rift Biodiversity Data" project (data provided by Prof. Jean Snoeks, MRAC and Prof. Eric Verheyen, RBINS, Belgium)

Concerning the fruitfly project all the data are databased, they are now checked by the specialists and images will also be provided (data provided by Prof Marc de Meyer, MRAC).

For the *Flora Brasiliensis* project, the taxonomy of the 8 pilot families has been reviewed by Prof H. Foerther of the Botanic state Collection of Munich. The digitizing process of the relevant herbarium sheets from the National Botanical Garden of Belgium and from Munich is also achieved. The Missouri Botanical garden has also provided the scanned images of the plates from the concerned *Flora Brasiliensis* volumes. The collected information will now be reviewed by selected Brazilians experts of these 8 pilot families (Alismataceae, Bignoniaceae, Cactaceae, Clusiaceae, Croton, Onagraceae, Rutaceae and Simaroubaceae) for data quality control.

The images in .tiff format have in order to allow extensive zoom in without quality losses each a size up to 200 mega. As it is inconvenient to upload such big files the Botanical garden of Meise has develop a image viewing system where the original images are split in several tiles (Figure 3). The user sees first a small Jpeg images of the whole sheet and only the asked tile of the image is downloaded for zoom in. This enhances the performance of the system greatly.

We are hosting all the images files on our server and have installed the tool designed by the Botanical Garden at the Belgian Node. It will also be possible to use this tool not only for the Martius collection but also for the herbarium sheets concerned by the Albertian Rift project. With the agreement of the garden and some adjustment it may also be possible to use this tool for the other specimen digital images concerned by these three projects (Butterflies, Cichlids and Fruit flies)



Figure 3: Examples from the image viewer designed by the National Botanical Garden of Brussels (Image viewer designed by Alain Vander Velde from the National Botanical Garden)

The taxonomical review of the 8 families has been done in Microsoft Word by the specialist, but we managed so far to transform these files in .html and to split these files by Family and Genus, put hyperlinks to the images when barcodes were cited as well as links to the *Flora Brasiliensis* plates and scanned texts. This information is now online at <http://martius.bebif.be/martius/> with a restricted access for the time being, as the content will be reviewed by Brazilian specialists of the concerned families. Please fill free to contact us at BeBIF if you wish an early access. The primary data retrieved directly from the herbarium sheet (like label information) is currently in the process to be encoded in a for that purposed designed Microsoft Access templates.

As digital imaging is a major part of these projects, the technical aspects were presented at the ENBI WP6 workshop on "Techniques and Challenges for digital imaging of biological type specimens" in Stuttgart, Germany in March 2004.

ENBI WORK PACKAGE 13 SLIK

Macaranga and Mallotus (Euphorbiaceae) species of Borneo

From: Ferry Slik

We have recently opened a webpage that was made with funding from ENBI - WP13. The website contains keys, spot characters and descriptions to all *Macaranga* and *Mallotus* (Euphorbiaceae) species of Borneo. With funding provided by NLBIF, an interactive method to calculate the disturbance level of lowland dipterocarp forest in Borneo based on *Macaranga* and *Mallotus* counts in plots of 0.3 ha was included as well. This method was especially developed as a rapid assessment tool to quantify forest disturbance, but can also be used to monitor forest recovery after disturbance.

The website address:

<http://www.nationaalherbarium.nl/MacMalBorneo/index.htm>

FERRY SLIK
Nationaal Herbarium Nederland
Plantdiversity of the Indopacific and Tropical Asia
Van Steenisgebouw
Einsteinweg 2 room B101
PO-Box 9514