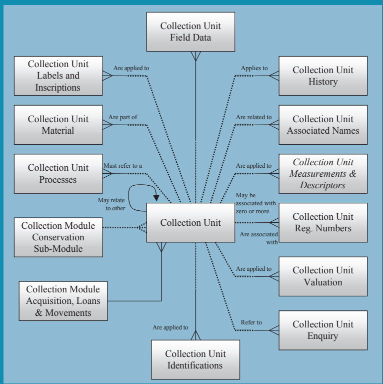
Ferrantia

ISSN 1682 - 5519

Proceedings of the first international Recorder conference

Luxembourg
2-3 December 2005



Tania Walisch (Editor)

51 2007

Travaux scientifiques
du Musée national
d'histoire naturelle
Luxembourg

Ferrantia est une revue publiée à intervalles non réguliers par le Musée national d'histoire naturelle à Luxembourg. Elle fait suite, avec la même tomaison aux Travaux scientifiques du Musée national d'histoire naturelle de Luxembourg.

Mise en page:

Romain Bei

Design:

Thierry Helminger

Prix du volume: 10 €

Rédaction:

Musée national d'histoire naturelle
Rédaction Ferrantia
25, rue Münster
L-2160 Luxembourg

tel +352 46 22 33 - 1
fax +352 46 38 48

Internet: http://www.naturmusee.lu
email: ferrantia@mnhn.lu

Echange:

Exchange MNHN-SNL
c/o Musée national d'histoire naturelle
25, rue Münster
L-2160 Luxembourg

tel +352 46 22 33 - 1
fax +352 46 38 48

Internet: http://www.mnhnl.lu/biblio/exchange
email: exchange@mnhnl.lu

Page de couverture:

1. Botanists doing field observations. Photo MnhnL
2. The main descriptive data elements associated with collection units and links to conservation and management (movements) modules.
3. Birds, butterflies, flowering plants and lacustrine fish specimens from the Albertine Rift. © NBGB and RMCA

Citation:

Tania Walisch (Editor) 2007. - Proceedings of the first international Recorder conference. Luxembourg 2-3 December 2005. Ferrantia 51, Musée national d'histoire naturelle, Luxembourg, 151 p.

Date de publication:

7 septembre 2007
(réception du manuscrit: 24 janvier 2007)

Impression:

Imprimerie Centrale, Luxembourg

imprimé sur papier FSC

ISSN 1682-5519

Ferrantia

51

Proceedings of the first international Recorder conference

Luxembourg 2-3 December 2005

Tania Walisch (Editor)

Luxembourg, 2007

Travaux scientifiques du Musée national d'histoire naturelle Luxembourg

Table of contents

Conclusion

Poster long contributions

Poster abstracts

Appendix

A CD-ROM containing a bookmarked pdf file of the complete 'Proceedings of the first international Recorder conference' (Ferrantia 51, 2007) together with the pdf files of the oral presentations held at the conference is included at the back of this book.

Preface by the editor

Tania Walisch
Information sur le patrimoine naturel
Musée national d'histoire naturelle
25, rue Münster
L-2160 Luxembourg
twalisch@mnhn.lu

Keywords: first international Recorder conference, natural sciences field and collection records, collating data, managing data

Introduction

The first international Recorder conference took place in Luxembourg on the 1st and 2nd of december 2005. The theme of the conference was *Collating and managing natural science field and collection records in Europe*.

The implementation of Recorder in Luxembourg and the collection management extension to Recorder between 2001 and 2004 did not remain unnoticed by the people involved in the domain of natural history from neighbouring countries. During this period, the biodocumentation center of Saarland, the floristic network of Germany and the Leiden Museum of the Netherlands for example were looking for a suitable data management software handling field and collection data. When they heard about Recorder through their contacts with people form the Luxembourg ministry of culture and the natural history museum, they were very interested to get to know the project in more detail.

To respond to the growing interest from organisations outside the UK, two key people in the Recorder project, Charles Copp and Guy Colling, launched the idea of an international Recorder conference. Charles Copp had designed the NBN data model on which the Recorder application was built and accompanied the Recorder project since its beginnings in the UK. Guy Colling, had been the initiator of the Recorder implementation at the Luxembourg Museum and of the development of the collection management extension for Recorder. The Luxembourg museum of natural history agreed to be the convener and organizing institution of the international Recorder conference on the condition that it got financial support from the Luxembourg national research fund (FNR).

The objective of the conference was to give an overview of the Recorder project, to show its link to other international projects, to allow networking between Recorder users and developers, and to discuss about the problems of developing, disseminating and sustainably supporting an international version of Recorder and associated software and to discuss proposals for future action.

Organisation

Organisations and people are listed in alphabetical order.

Convener

Luxembourg National Museum of Natural History

Partner organisations

Bundesamt für Naturschutz, Deutschland
Dorset softwares Ltd, U.K.

Joint Nature Conservation Council, Peterborough, U.K.
National Biodiversity Network, U.K.
Natural History Museum, U.K.
Société des naturalistes luxembourgeois, Luxembourg

Financial support

Luxembourg National Research Fund

Conference coordination

Ms Tania Walisch / Dr. Guy Colling
Musée National d'Histoire Naturelle
25, rue Munster
L-2160 Luxembourg
Phone: (00352) 46 22 40 263 or 262
Fax: (0352) 46 38 48
Email: twalisch@mnhn.lu
Internet: www.mnhn.lu

Organising committee

Guy Colling, Luxembourg National Museum of Natural History

Charles Copp, Environmental Information Management

Jim Meisch, Luxembourg National Museum of Natural History

Tania Walisch, Luxembourg National Museum of Natural History

Scientific committee

John van Breda, Dorset softwares

Guy Colling, Luxembourg National Museum of Natural History

Charles Copp, Environmental Information management

Charles Hussey, Natural History Museum, London

Tania Walisch, Luxembourg National Museum of Natural History

Conference chairman

Adrian Rissone, Natural History Museum, London.

Background and theme of the conference

Background (Copp & Walisch, 2005)

Sustainable development and the conservation and enhancement of biodiversity have been central to all European environmental and economic policy since the signing of the Rio Convention in 1992. Meeting these objectives requires ready access to scientifically valid, reliable and comprehensive wildlife and earth sciences data at local, regional, national and European levels.

National projects such as the National Biodiversity Network in the UK and European-funded projects such as ENHSIN and BioCASE have helped develop techniques and standards for biodiversity data sharing but there is still a great need for software and standards that can capture, manage and integrate the original natural sciences collection and field data. The theme of this meeting was to introduce field biologists, earth scientists and natural science museum curators across Europe to software and standards that can meet this need and discuss ways in which it could be deployed throughout Europe.

Theme (Copp & Walisch, 2005)

The conference included presentations on the development of Recorder and its associated applications (including 'thesaurus' and collections management software) in the UK and Luxembourg. Recorder has been developed at public expense and is readily available to users on a not-for-profit basis. The emphasis of the meeting was on the development and use of data standards and how the Recorder suite of applications could help build integrated systems for the collation, management and dissemination of information in the natural sciences across Europe.

The conference thus covered the development of recording and data management standards and the way they have been used in Recorder. It showed how Recorder can be used at different levels, from the individual naturalist through local (Sussex Biodiversity Record Center, UK), regional (CEDAR, Northern Ireland) and national collation centres (National Biodiversity Network, UK ; Musée National d'Histoire Naturelle, Luxembourg). Examples of working systems were demonstrated and ongoing development work was described including the internationalisation of the software to work with different languages and mapping bases and the development of a web version. Furthermore there were presentations on the international networks BioCASE and GBIF, bringing together biological observation, specimen and collection data on the web.

There were presentations on the scope and potential of the BioCASE/Luxembourg Thesaurus software for managing and controlling taxonomy and other terminology with examples of how it is being utilised in Recorder and other applications. There were also presentations on the project to create and deliver a national taxon list for the UK and associated web services in the Natural History Museum, London.

Finally the problems of developing, disseminating and sustainably supporting an international version of Recorder and associated software were discussed and proposals were made for future action.

Target audience

- Field recorders and recording scheme organisers across Europe
- Biological and geological museums looking for software to manage specimens, collections and field records
- Managers of biological records centres and geodiversity record centres
- Organisations charged with delivering biodiversity or earth sciences data and advice to planners, local and national governments
- Those involved in developing or managing local and national biodiversity networks including BioCASE and GBIF nodes
- Those involved in developing web-based biodiversity information systems
- Developers of information standards and ontologies for biological and earth sciences

Participation and outcome of the conference

71 people from 8 European countries attended the first international Recorder (Fig. 1).

At the end of the conference there was a discussion, chaired by the conference chairman, about the requirements for future development, web enablement and internationalisation of Recorder (Rissone 2007). Future actions were suggested. An informal international working group was set up with the aim of coordinating action in the first instance like internationalising Recorder, coordinating developments of modules and improving communication about an internet version of Recorder (Rissone 2007).

The Joint Nature Conservation Committee represented by Steve Wilkinson announced to give the Recorder project more visibility on the internet by creating a website bringing together information about the Recorder project and ongoing developments and hosting a forum of Recorder users.

These formal commitments as well as the informal contacts between actual and potential Recorder users and developers from various countries have yielded a number of positive initiatives in the year following the conference.

- Four large organisations have joined the Recorder community

The German Federal Agency for Nature Conservation in Bonn has chosen the Recorder software as a functional replacement of the *FlorEin* database for floristic survey data. *FlorEin* actually holds 15 million records and is linked to *FloraWeb*, a digital web flora of vascular plants in Germany (May 2007).
The Biodocumentation Center in Reden, Saarland has chosen Recorder and the collection addin for the integrated management of their earth and life sciences collection, specimen and field data.

Fig. 1: Participants at the first international Recorder conference, in front of the entrance to the exhibition building 'Naturmusée' of the Luxembourg National Museum of Natural History (J. Meisch, 2005).

The National Biological Record Center of Ireland, based in Waterfort, which opened in the summer of 2006, opted for Recorder as a tool for managing its field data.

The *Naturalis* Museum in Leiden, the Netherlands, will adopt Recorder and its collection and thesaurus extension for the management of their collections and specimens.

- The visibility and information about Recorder has improved.

The Joint Nature Conservation Council and Dorset software created a web page dedicated to the Recorder project (Recordersoftware 2006): Charles Copp EIM created a community driven documentation wiki for the Recorder and Collections Module: (EIMwiki n.d.).

- Recorder has become international.
The German Federal Agency for Nature Conservation, the Saarland Biodocumentation Center, and Dorset Software have developed an internationalisation module for Recorder. This addin permits the user to use and switch between European coordinate systems. Furthermore it allows the translation of the whole application into another language. The *Biodokumentationszentrum* is currently translating Recorder into German.

- Recorder has become interesting for new domains.

The prehistory department of the *Luxembourg National Museum of Art and History* in conjunction with Charles Copp have tested Recorder for archeological data management (Copp 2006a, Broniewsky 2007).
The Luxembourg National Museum of Natural History's department of living collections has worked with Charles Copp on an outline specification for a botanical live collections extension for Recorder *6* (Copp[2] not published). This document and a technical system design, which documents the outcome of several design sessions with *Dorset software*, will be at the base of the built of a living collections module for Recorder.

- Recorder is coming closer to the web.

Dorset software has made a proposal for the development of an *Open*Recorder on the web (Dorset Software 2006). *Open*Recorder has been created as an alternative to the full Recorder *Web* proposal. Free Open Source tools have been chosen to allow any organisation to set up biological recording and reporting websites at little cost and require only basic technical experience. Once delivered to the community it is expected that there will be subsequent additional functionality created either through the Open Source community or further costed developments. The goals of *Open*Recorder are to allow simple data entry for users who do not have Recorder installed, to allow data entered to be easily imported into Recorder and to allow simple reports to be generated, including data originally entered via Recorder (Dorset Software[2] unpublished).

For large organisations, the Recorder *Web toolkit* application would be the preferred data collection and reporting tool on the web. Several partners (Naturalis Museum, Leiden; Natural History Museum, Luxembourg with financial support from eLuxembourg) are exploring ways to finance parts of the Recorder *Web toolkit* development, at the same time looking for more partners. In a first design meeting at the *Naturalis* Museum in Leiden, they will re-evaluate the specification for the built of the Recorder *Web toolkit* (Dorset Software 2005).

Conclusion

The first international Recorder conference has been quite successful. Thanks to the well-structured programme and the very valuable contributions of all conference speakers, the conference presented the Recorder project in its full richness and complexity. Thus it showed the use of Recorder at a local, a regional and at a national level. It presented national and international efforts to bring together life sciences data on the web, and demonstrated as well as encouraged the participation of Recorder users in these projects. The many positive initiatives following the conference show that it helped to improve communication and collaboration among Recorder users. Holding an international conference has been shown to be very useful and a lot of fun. I hope that soon another organisation will host a second international Recorder conference.

Acknowledgements

I would like to thank the following people for their help. Their names appear in alphabetical order.

Guy Colling for help with organisation and content.
Charles Copp for setting up the programme and the first circular and for inviting conference speakers.
Guy Eicher for his help with financial matters.
Geneviève Faber for organising the meals and the evening reception.
Guy Frank for continuous support of the Recorder implementation in Luxembourg.
Christian Ries for setting up the web page and for help with administrational matters.
Jim Meisch for the design of the conference logo and for technical assistance during the conference.
Marie-Antoinette Djomna for help with registration.
Micky, Noël, Elsa, Alex of the Museum cafeteria for the two delicious lunch meals and for the evening reception during the conference.
Jean-Marie Zimmer for help with the abstract book and registration.

The Fonds *National de la recherche* for its financial support of the conference.

References

Broniewsky 2007. - Recorder 6 as a database for archaeological purposes: x-x, in Walisch T. (ed.), Collating and managing natural sciences field and collection records in Europe. Proceedings of the First International Recorder Conference. Luxembourg 2-3 December 2005. Ferrantia 51, Musée national d'histoire naturelle, Luxembourg, xxx p.

Copp C.J.T. & Walisch T. 2005. - First circular of the first international Recorder conference. Luxembourg 2-3 December 2005. Collating and managing natural sciences field and collection records in Europe. Available at website of Recorder conference: http://www.symposium.lu/recorder/. Accessed December 2006.

Copp C.J.T. 2006a. - Notes arising form a meeting to discuss the data management and software requirements of the Prehistory Department of the Luxembourg National Museum of History and Art, 21.02.2006. Unpublished report. Musée national d'histoire et d'art, Luxembourg.

Copp C.J.T. 2006b. - Luxembourg National Museum of Natural History. Outline Specifications for Recorder 6. Botanical Live Collections Extensions, vers.1.2. 6th November 2006. Unpublished report. Musée national d'histoire naturelle, Luxembourg.

Dorset software 2005. - Technical System Design for Recorder Web. Vers.1.1. 6th January 2005. Unpublished report. Musée national d'histoire naturelle, Luxembourg.

Dorset software 2006. - Proposal for an OpenRecorder. v.1.0. 25th September 2006. Unpublished report. Musée national d'histoire naturelle, Luxembourg.

EIMWiki n.d. - EIM hosted wiki for Recorder available at http://eim.metapath.org/wiki/index.php?title=Recorder. Accessed March 2007.

May R. & Schneider T. 2007. - Floraweb - the German web flora and the role of Recorder: x-x, in Walisch T. (ed.), Collating and managing natural sciences field and collection records in Europe. Proceedings of the First International Recorder Conference. Luxembourg 2-3 December 2005. Ferrantia 51, Musée national d'histoire naturelle, Luxembourg, xxx p.

Recordersoftware 2006. - Dorset software hosted website on Recorder available at www.recordersoftware.org. Accessed March 2007.

A. Rissone 2007. - First International Recorder Conference - Recorder now and in the future: x-x, in Walisch T. (ed.), Collating and managing natural sciences field and collection records in Europe. Proceedings of the First International Recorder Conference. Luxembourg 2-3 December 2005. Ferrantia 51, Musée national d'histoire naturelle, Luxembourg, xxx p.

Conference Programme

The duration of the talks was 25 minutes plus 5 minutes for discussion. A CD-ROM containing a bookmarked pdf file of the complete 'Proceedings of the first international Recorder conference' (Ferrantia 51, 2007) together with the pdf files of the oral presentations held at the conference is included at the back of this book.

Friday, 2nd of December 2005

Introduction and Welcome

Welcome speech by Dr. Georges Bechet, director of the Luxembourg Natural History Museum. Short introduction by Adrian Rissone, conference chairman.

Theme I: Recording and Collating

Charles Copp. – Overview of the Recorder project. Introduction to the history and development of Recorder (2002, 6 etc.)

John van Breda. – Demonstration of Recorder applications. Integrated natural science collections and observations data management.

Charles Copp. – Overview of data standards for field recording and collections management in the natural sciences. Introduction to the data model and standards; the basis on which Recorder is built.

Charles Roper. – Using Recorder for biological records in the local context: collating and creating products. Recorder 2002/6 – what it can do and how it is used – example from an active local record center.

Guy Colling. – Recorder in the Museum context. Recorder 6 and its collection management extensions. The Luxembourg project.

Hannah Betts. – Import and Export of data. Importing data from other software users and exporting data to GIS, the NBN etc.

Theme II: Terminology

Charles Copp & Guy Colling. – Creating and managing term lists and dictionaries. Introduction to the Thesaurus.

Charles Hussey & John Tweddle. – Taxonomic names and name servers. Natural History Museum Taxon List projects.

Discussion and Questions

Chairman and speakers.

Saturday, 3rd of December 2005

Theme III: Using the Internet

Examples of projects that deliver information to a variety of users over the web and in which Recorder and dictionaries play a part

John van Breda & Alistair McLean. – Web-based Recording software. The development of Recorder Web and its potential. Showcase of Nature Notebook.

Steve Wilkinson. – UK national system. The NBN Gateway – a national system.

Damian McFerran. – A UK regional system. CEDAR – a regional system.

Tania Walisch. – The Luxembourg system. Luxembourg project – building a national bio- and geodiversity web portal

Anton Güntsch & Patricia Mergen. – Linking to the BioCASE Web Portal. The BioCASE Project – a biological collections access service for Europe.

Larry Speers. – Linking to GBIF and other international systems.

Theme IV: International Recorder and related software

John van Breda. – Internationalising Recorder. The work done so far in making Recorder available outside the UK and how it can be tailored to different languages and geographic areas.

Rudolf May. – Recorder project in Germany. The German web flora (www.floraweb.de) and the role of Recorder: first results of the Recorder pilot project supported by the BfN in Germany.

Adrian Rissone, Steve Wilkinson & John van Breda. - Supporting International Recorder. Discussion of how we fund and support the expansion of Recorder use in Europe.

Conclusion

Adrian Rissone, Steve Wilkinson & John van Breda. Future action. Set agenda for next actions and targets.

Poster session at lunch breaks on Friday 2nd and Saturday 3rd of December

Raymond Aendekerk & Marc Thiel. - Grassland mapping of the commune of Niederanven.

Stefan Klein, Anton Güntsch, Markus Doering & Walter G. Berendson. – The BioCASE network: Integrating European Collection Information Ressources.

Patricia Mergen, Michel Louette, Jos Snoeks, Marc de Meyer & Danny Meirte. – The Royal Museum for Central Africa in the era of biodiversity informatics.

Susanne Rick. The FNR 'Espace et Patrimoine Culturel' Project: A computer aided cultural heritage management tool for the Grand-Duchy of Luxembourg.

Francis Rowland. – Species records, data flow and the Biological Records Center.

Introduction to the history and development of Recorder

Charles Copp
Environmental Information Management
8 The Paddock
GB-BS21 6JU Clevedon
eim@globalnet.co.uk

Recorder began life in the 1980's as a database, created by Dr Stuart Ball, for the recording of sites and species in relation to a project called the Invertebrate Sites Register. The original version of Recorder was written in a product called Advanced Revelation (AREV), which is very different in its structure from modern SQL relational databases, but did allow very large, text-based records with multiple values for fields that made it ideal for biological recording.

Stuart started to make Recorder available, in a voluntary capacity, to other naturalists and over a period of time a development committee was established to discuss ways in which the program could be maintained and improved. Key features of Recorder were a set of data entry and maintenance screens linked to embedded reference tables including a 'species dictionary' and habitat lists. Over a period of years, the application came to include a great many, useful report and mapping functions. The content of the reference tables was built up through voluntary cooperative effort.

As the use of Recorder grew, this was recognised by his employers (initially the NCC and since 1991 the JNCC) and he was able to provide some of his work time to develop and support the software in association with a Recorder Board of Management. By 1995 there were over 300 registered copies of Recorder in use and many other unofficial copies. Recorder had been updated to a stable version 3 and software support at that time was contracted out to the UK Wildlife Trusts.

In 1995 a review of Recorder was commissioned, which recommended that the software be re-developed to run under Windows and that it be linked to published standards to support the emerging National Biodiversity Network. This included de-coupling the 'species' and other dictionaries which were to become open to users of other software. A major requirement was for a standard data model that could be used to guide software design and inform data exchange.

The Systems Analysis for the rebuild of the Recorder Software (Nov. 1996 - March 1997), was undertaken by Charles Copp and dealt extensively with the nature and structure of biological records and their management. This information was used to define a modular data model from which the table structure of the new application could be defined. The data model was deliberately cast wider than the immediate confines of the Recorder Project and has proved to be very powerful.

The analysis phase of the model's development included a review of existing data models, questionnaires and meetings and theoretical analysis. A series of workshops were held all over the UK involving representatives of a wide range of governmental and non-governmental organisations as well as software developers and individual naturalists. The principal outcome of these meetings was an ambitious requirements catalogue from which the new Recorder Software was defined.

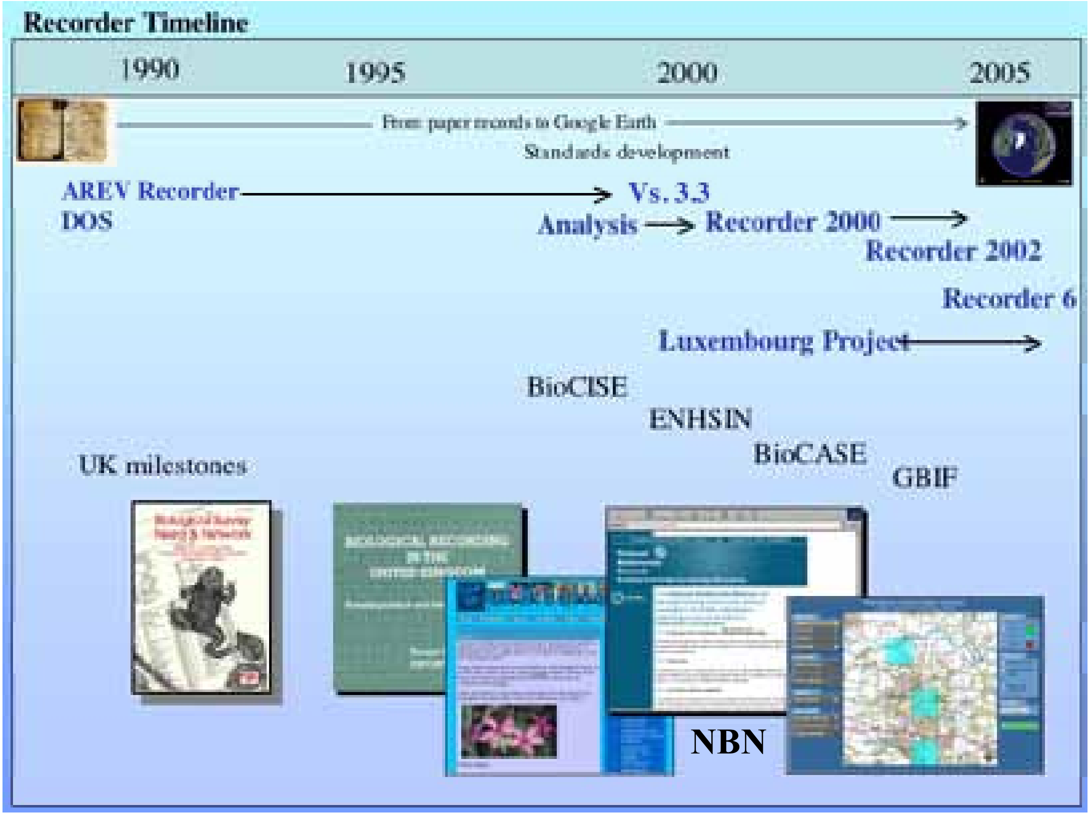


Fig. 1: The various versions of Recorder and related developments in Biological recording and Biodiversity Information systems in the UK such as the boost that came to biodiversity projects following the signing of the 'Rio Treaty' in 1992 and the establishment of the UK National Biodiversity Network in 2000.

The final data model was tested against existing surveys, databases, recording media and the user requirements within different types of organisation. The data model included modules for the various dictionaries and that for the 'taxon dictionary' was adopted by the Natural History Museum, London when they took over responsibility for managing the UK species dictionary. The data model has been revised and extended several times since its creation, including the definition of a new specimen collection management module and a more powerful Thesaurus Module. The current version of the 'NBN Data Model' was published on the NBN website in 2004 and updates are published on the Recorder website and its associated wiki. (Recordersoftware 2006, EIMwiki n.d.).

During 1997 the model was further tested through the building of a prototype 'habitats' database for the Countryside Council for Wales and in 1998, following a large scale tendering exercise, the design and build of the new software, to be called Recorder 2000, took place. The main design and build process took a year and included a design group, including Stuart Ball and Charles Copp, working with Dorset Software, who are the appointed developers.

The original intention had been to create a purely modular piece of software that would include a simple version for naturalists and a more complex version for record centres that collate data from many sources. In practice there were not the resources to do this or to make a system that would work on any database back-end and so the release version was a single application that ran on Microsoft Access.

The first version, Recorder 2000, was released in September 2000. Recorder is controlled by the JNCC but distributed through a number of appointed re-sellers who are able to offer tailor-made support, customisation and training. There does not appear to be a fixed price for the software, which is intended to be nominal and resellers find it hard to economically distribute it. Recorder continues to be supported and controlled by staff at JNCC although there is also a management board and an active web forum. (NBN Forum n.d.).

There has been continued investment in improvements to Recorder in the past 5 years and a new version, Recorder 2002, has recently been released. In the meantime a more powerful version, Recorder 6 has been developed, that works with Microsoft SQLServer and is more suitable for large record centres. During this period many other developments have taken place including the provision of 1:50,000 scale maps and boundaries of Watsonian Vice Counties and the writing of numerous 'add-ins'. The largest 'add-in' developments have been funded by the Luxembourg National Natural History Museum, who have paid for the extension of Recorder to earth sciences data and the management of specimens together with the development of a new Thesaurus Module. A further large add-in is about to be developed to enable the management of information related to living collections and a preliminary study has taken place to investigate the addition of archaeological and anthropological records.

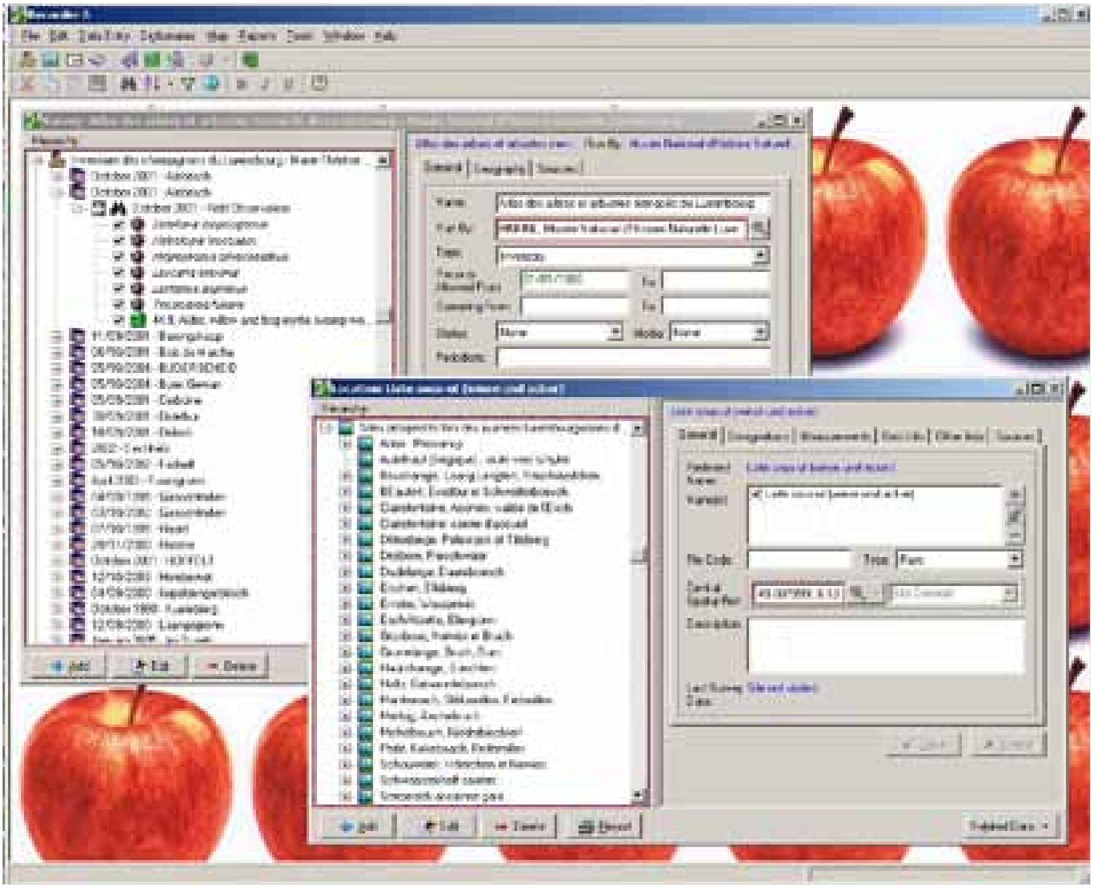


Fig. 2: The Recorder application makes use of hierarchical trees for navigating the complex data structure, providing add, edit and management functions throughout. A number of tasks, including reporting, have wizards to help the user. The illustration shows the main field records window and the Location (sites) window. Data can be dragged and dropped from one window to anther, for instance when associating a site with a field observation.

The main focus of development is now on Recorder 6 and its numerous add-ins. With the development of the new Thesaurus module we have been able to add new facilities for recording more types of field occurences and specimen types, particularly earth science specimens and stratigraphic occurrences. There is also a project to build a web interface for Recorder that will provide a toolkit capable of creating user-specific interfaces for adding and retrieving data whilst maintaining the powerful security and validation features of the main Recorder system.

There are now many hundreds of registered users but many of them are amateur naturalists who struggle to understand the complexities of the interface and its underlying data model. Recorder can provide them with very simple data capture screens and reporting facilities but this is not often implemented and can be further complicated poor documentation and by the lack of a single 'recommended checklist' in the species dictionary, although this is now changing for UK users through the work of the Natural History Museum, London.

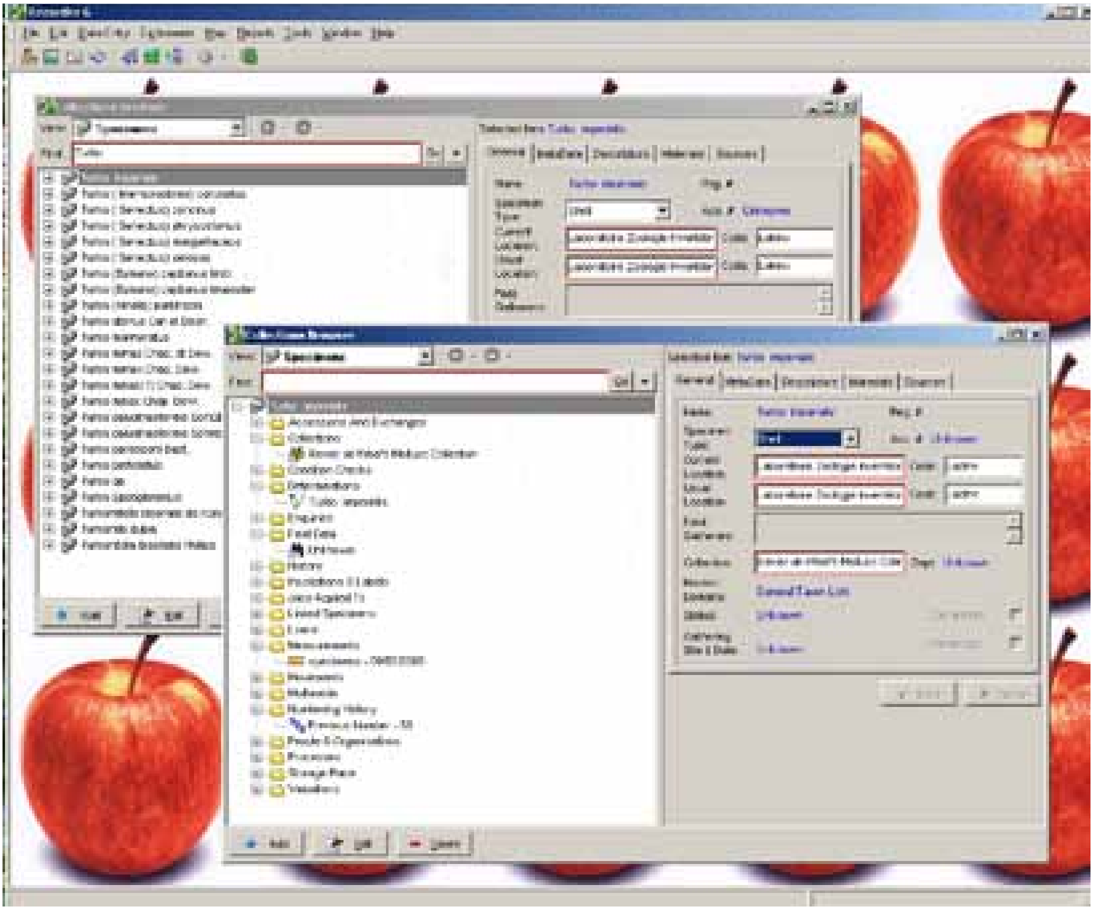


Fig. 3: The Recorder application can be extended through add-ins. The largest extension to Recorder is the Collection Module which includes the functions for storing information about specimens and collections and details of their management. The illustration shows two instances of the Collections Browser Window, one opened to show a list of the specimens identified as belonging to the Genus ***Turbo*** and the second showing one of these specimens with the many 'folders' of information that can be attached to it. The Collection Browser window allows many different views of specimens, collections, storage locations and collection management functions.

Recorder is changing and we are embarking on projects that will help us realise the aim of creating a range of modular applications that are geared to user needs but are based in standards and the ability to share data. The development of the proposed 'Web Recorder' holds the greatest promise in this direction and could be the ideal tool for the majority of amateur naturalists and for museums and record centres wishing to engage with the wider public. Recorder is also becoming internationalised and the take up across a broader geographic area will help more people to engage with projects such as GBIF and develop their own national networks. An international user base will also encourage the establishment of new mechanisms for support and development.

There is no doubt that Recorder is a very powerful and flexible tool, that offers a complete system of data capture, management, reporting and mapping across the range of natural sciences. Such potential, also means that it requires a real commitment to make the best use of it and probably its greatest weakness is a lack of accessible support and training for users. It is clear that the development of Recorder has gone beyond the scope of the original specification and the take-up of Recorder in Europe makes it difficult for the UK JNCC to continue to control and support Recorder as it did at the start of the project. It is also true that Luxembourg Museum cannot alone fund all new developments and users in Europe. We must now be actively seeking to set up a cooperative body that brings together organisations able to contribute to the ongoing development and maintenence of the software and establish Recorder as an open, community supported project.

Recorder in many ways reflects the history of our understanding of biological recording standards and the growth in electronic data sharing. It has had many years of effort and considerable sums of money spent on its development and it is a European application that suits our needs. The story is far from over, in fact it is just changing up a gear.

Demonstration of Recorder applications - Integrated natural science collections and observations data management

John van Breda
Dorset software Ltd.
Corbiere House
New Quay Rd
BH15 4AF Poole
John.vanBreda@DorsetSoftware.com

Keywords: *field observations, specimen data, collection data, data storage, thesaurus*

Presentation abstract

Recorder provides comprehensive support for storage and analysis of species and habitat based field data (figs. 1&2). However, there is increasingly a recognition of the need to underpin field data with specimen collections to guarantee the quality of the records. In general, there has been a tendency for software to specialise in either field or collections based records. This presentation introduced the integration of field and collections data provided by the Recorder application when the Collection Module addin is installed (fig. 3). This provides a system for recording field observations along with comprehensive support for data regarding the specimens that are gathered, including tracking of collection, accession, storage, movements and preservation (fig. 4). Provision of a thesaurus extends the capabilities of the Taxon and Biotope Dictionaries in Recorder to allow recording of unlimited types of bio- and geodiversity observations and specimens (fig. 5).

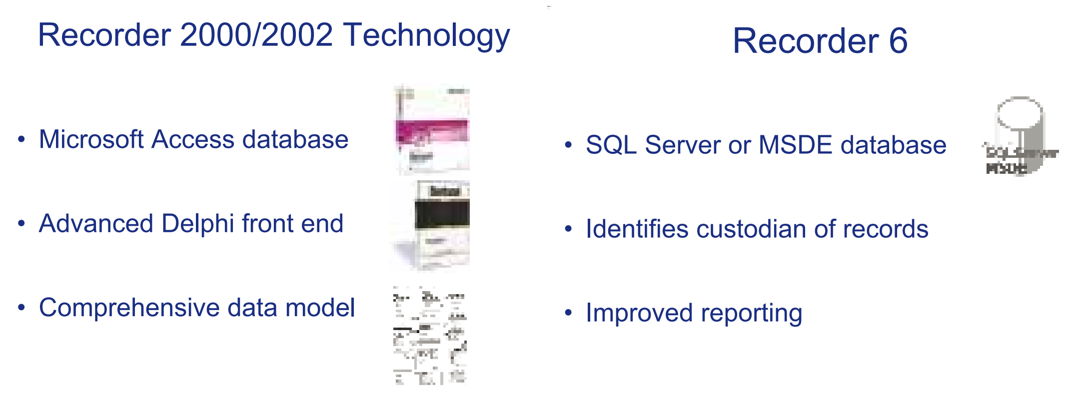


Fig. 1: Image showing the taxon dictionary, the document, the observation and the map windows of the Recorder *2002* application.

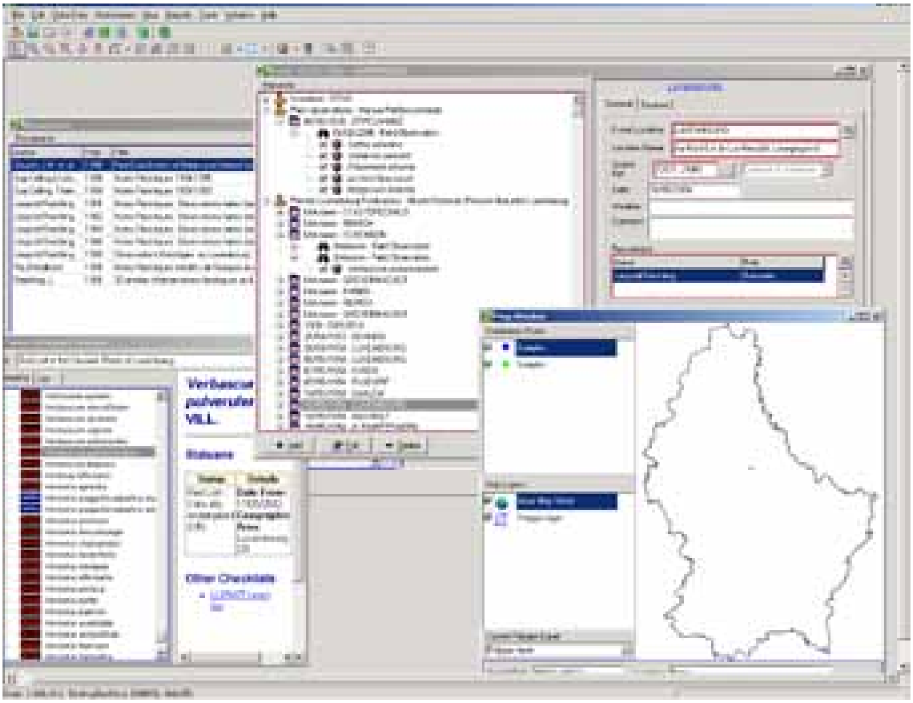


Fig 2: Recorder 2000/2002 Technology and the main changes implemented in Recorder 6.

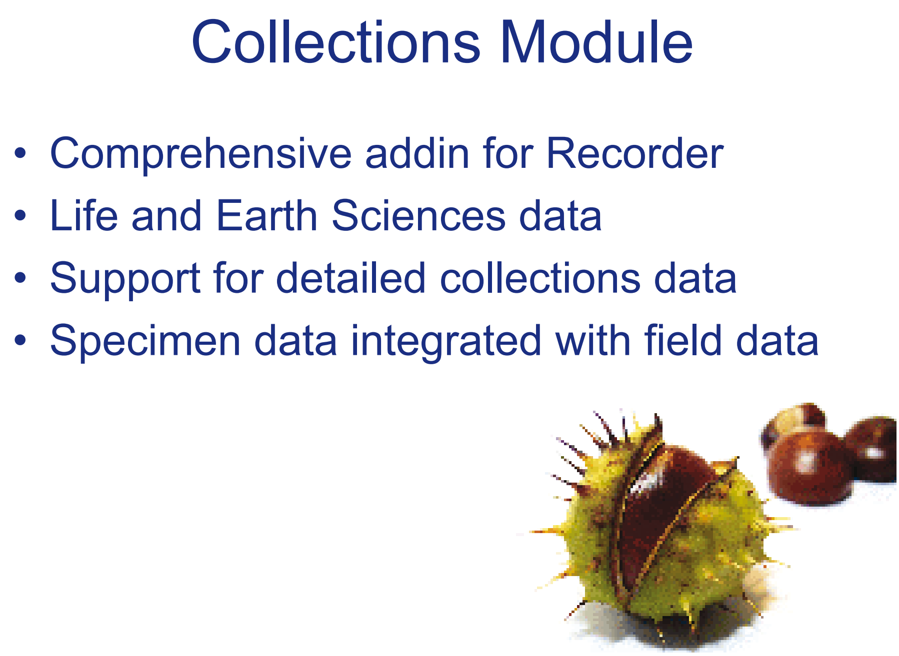


Fig. 3: Slide listing the main characteristics of the Collections module for Recorder 6.

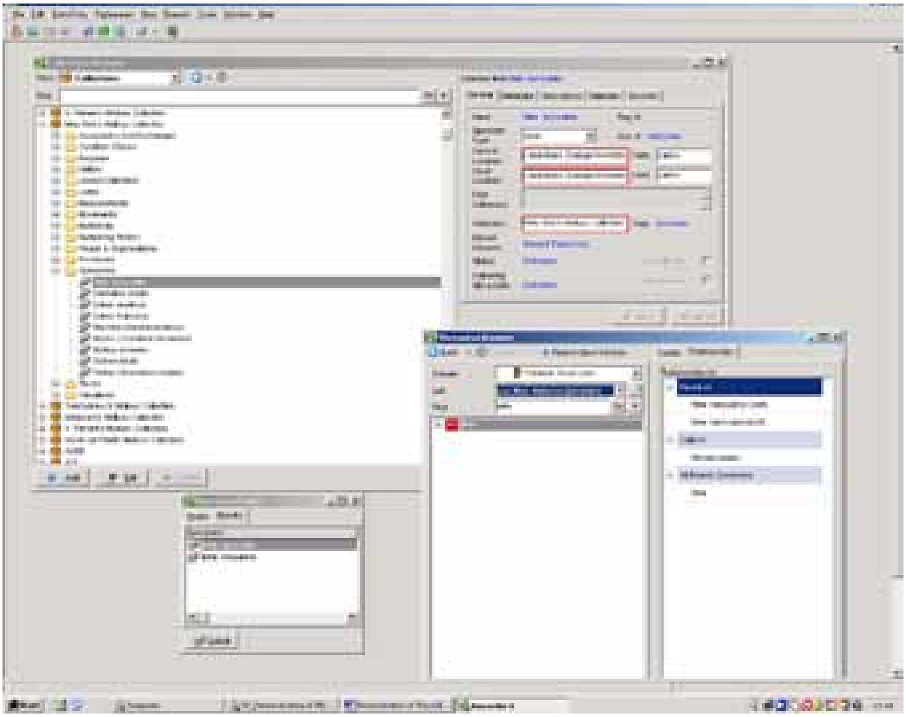


Fig. 4: View of the main functionalities implemented by the collections management addin for Recorder 6: the collections browser, the thesaurus browser and the specimen finder.

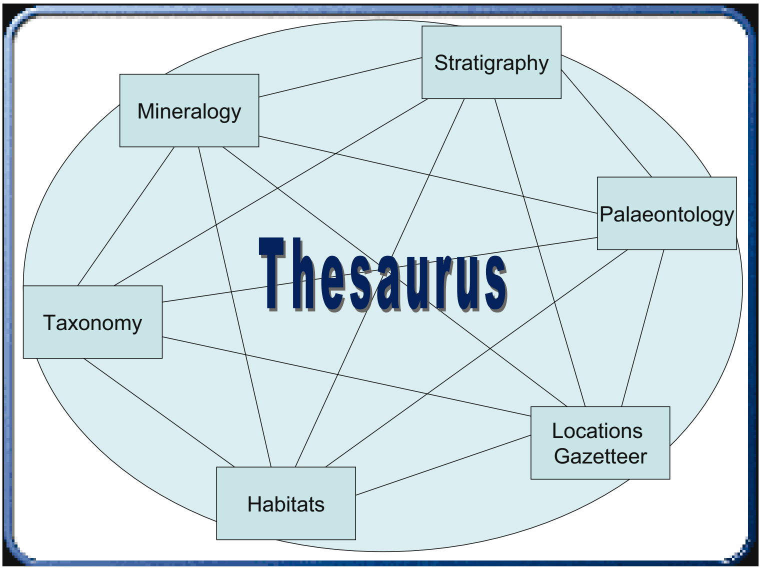


Fig. 5: The thesaurus has been developed as part of the collections module addin. It is multi-domain and allows to interrelate anything and everything. It provides a dictionary service for Earth sciences and all new term lists. It is multilingual and contains semantic information. It allows the import and export of lists from and to the taxon dictionary of Recorder.

An introduction to data standards and their implementation in Recorder

Charles Copp
Environmental Information Management
8 The Paddock
GB-BS21 6JU Clevedon
eim@globalnet.co.uk

Keywords: NBN data model, data standards, data content, data transfer, data storage, operational standards, methodologies, data quality standards

Abstract

Collating and sharing data from different sources and disseminating it in many ways, such as providing records through an Internet portal, is full of challenges. Much of the existing data can only be collated at relatively basic levels and even then there can be many problems in harmonising taxon names, geographical references and measurement systems. The electronic systems for sharing information, that are being developed, also require extensive metadata for their data catalogues and to protect the intellectual property rights of recorders. Great progress has been made in the development and adoption of standards to aid these processes although there is still much to do both in their creation and their adoption. Standards are needed to cover virtually every aspect of data collection, collation, management and dissemination. These include both Data Standards which refer to the organisation and content of information and Operational Standards, which directly relate to the data standards.

The Recorder 2000/2002 & Recorder 6 applications were designed to deliver the standards developed or adopted by the UK NBN. This includes the underlying data model and a concept-based taxon dictionary. This makes it ideal for building local and regional biological information networks. Recent developments through projects such as BioCASE, GBIF and the work of the Taxonomic Database Working Group (TDWG) have progressed the definition of a number of related world standards and these will need to be taken into account as Recorder is both extended into a specimen collections management role and internationalised for use throughout Europe.

Introduction

Collating and sharing data from different sources and re-purposing it in many ways, such as making it available through an Internet portal, is full of challenges. Much of the existing data can only be collated at relatively basic levels and even then, there can be many problems in harmonising taxon names, geographical references and measurement systems. The electronic systems for sharing information, which are being developed, also require extensive metadata for their data catalogues and for the protection of the intellectual property rights of recorders.

Great progress has been made in the development of standards to aid these processes, although there is still much to do both in their creation and their adoption. The need has been widely recognised and there is world-wide effort going into the definition and documentation of standards. This trend has been strongly influenced by the growth of the Internet and the adoption of new representational technologies such as XML and RDF which have lead to the evolution of schemas and ontologies.

In the UK the National Biodiversity Network (NBN n.d.) has developed and promotes a wide range of standards ranging from a common data model to standards for data custodianship and intellectual property rights and has championed the development of the Recorder Data Model and the Recorder application. The Taxonomic Database Working Group (TDWG n.d.), which started as an organisation primarily concerned with data exchange between botanical gardens has long since broadened its remit to all biodiversity data and now refers to itself as *Biodiversity Information Standards (TDWG)*. TDWG, working with projects such as GBIF, are now the leading organisation in the development of international biodiversity data exchange standards and is beginning to work closely with other standards organisations, such as the Open Geospatial Consortium (OGC 2007) to harmonise the delivery of all types of geospatial related data.

The Recorder 2000/2002 & Recorder 6 applications were designed to deliver the standards developed or adopted by the UK NBN. This includes the underlying data model and a concept-based taxon dictionary, which makes it ideal for building local and regional biological information networks. Recent developments through projects such as BioCASE, GBIF and TDWG have progressed the definition of a number of related world standards and these will need to be taken into account as Recorder is both extended into a specimen collections management role and internationalised for use throughout Europe.

Standards

Why we need standards

The term 'standard' is so frequently used that the meaning can become obscured, particularly as it can be used in relation to such diverse things as procedures, products, scales of measurement and data formats and may also be used to imply quality as well as content. All these things represent an agreed means of describing things which, through common adoption, can lead to efficiency, economy and reliability in the delivery of biodiversity data to a wide range of users.

Standards are needed to cover virtually every aspect of data collection, collation, management and dissemination. These include both **Data standards** and **Operational standards**.

Data standards

Data standards refer to the organisation and content of information. They include:

- **Data models** which define the scope of the information of interest to users (the domain). Data models can be expressed in numerous forms including logical data models and high level ontologies and xml schemas. The NBN data model, used by Recorder uses a multi-level, relational, logical data model and has been used to define the physical model for the Recorder database.
- The individual data attributes related to entities in the logical model or classes in an ontology may be organised into physical tables and fields to create a database. These attributes go together to form **Data Content** standards.
- Terminologies and structures (e.g. hierarchies) form **Data Classification** standards whilst the controlled terminology comes under the heading of **Vocabulary conventions**.
- The format or syntax of data in different attributes (e.g. personal names, dates and grid references or measurement units used) form **Syntax conventions**. Many format and syntax definitions are already available as ISO standards.
- Storage and transfer formats (including **Data Exchange Schemas**) describe the actual way data are represented electronically (e.g. file formats) and made available to other systems. Biodiversity and geospatial data transfer standards have been vigorously explored by TDWG and OGC.
- **Metadata standards** are used for describing the content and format of individual datasets. There are numerous metadata standards projects throughout the world including the influential Dublin Core Metadata Initiative (Dublincore 2007) and the TDWG Natural Collections Description Group (TDWG NCD 2007).

Operational standards

Operational standards deal with the way we do things and those which directly relate to data standards include:

- Methodologies for collecting and recording biological information
- Means for accrediting recorders or determiners
- Means for verifying data content

- Quality standards for data management
- Controls and arrangements for access to the data
- Arrangements for ensuring the physical security of the data

Where we need standards

We need standards to help us record better information in the first place and to ensure that the data are accurate and meaningful, not only for the original purpose for which they were gathered, but for future and often unpredictable uses. A typical example is the person who collects, for example butterfly records, who might simply enter a taxon name, place name and date onto a simple spreadsheet. As far as the person is concerned, he knows that the recorder was himself and he has mental references of all the sites he visits and the taxa he sees. He can even convey this information to his friends successfully because of shared common assumptions. Once those data are passed to strangers, however, none of this contextual knowledge is available unless it is written down. For the data to be of longer term use it is essential to be able to map it to widely used and understood taxonomic and geospatial reference systems and also to have documented information on the origin and copyright of the original records. With the growth of electronic data networks that can deliver information in many formats to anywhere in the world, the need to understand the quality and origin of information in the system becomes ever more important. The Recorder application has been built as a tool, not only for storing a

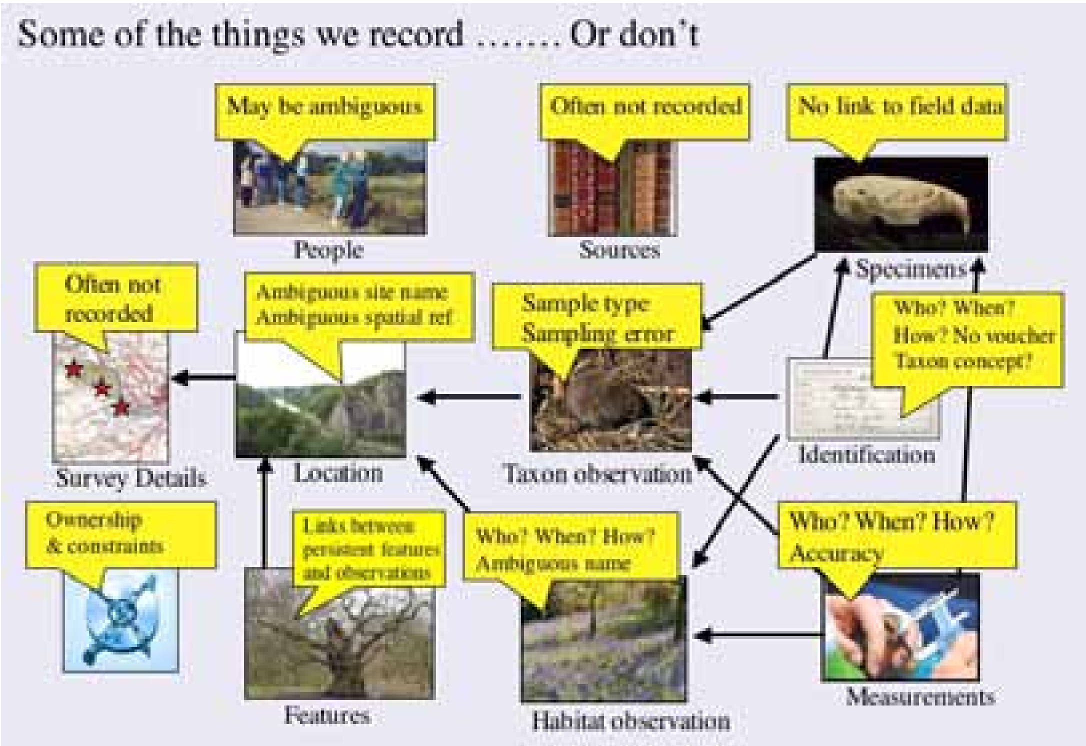


Fig. 1: There can be an enormous gulf of understanding between those who collect data and those who manage data as to what information is needed in a biological record or attached to a specimen. What is obvious and understood to the collector, may cease to be once the record is passed on. The figure above highlights some of the areas where information may commonly be lacking or ambiguous. The Recorder application has minimum record definitions and validation to help ensure that records are both reliable and re-usable.

wide range of biodiversity (and geodiversity) data but also for validating the data against standards and for managing the associated ownership and dissemination issues.

The discussion above and the items illustrated in Figure 1 indicate that standards should apply throughout every area of record collection, management and dissemination. This is a huge burden on data managers if it is not supported by freely available, relevant authority files and robust software to deliver them. The current huge growth in the building of web-based biodiversity information systems will be entirely dependent on the availability of web-based authority and validation servers providing, in particular, taxonomic and geographic names, probably associated with globally unique identifiers (GUIDs and LSIDs) (see WIKIPEDIA n.d. for articles and links to LSIDs). These web-based services will become increasingly important to all data managers.

Locally, however, for some time yet, most information will continue to be collated and managed in traditional databases where names and places are just part of the array of validation and management information that needs to be available through all the 'life stages' of the information management process (i.e. from collection, through to collation, revision, dissemination and possible deletion). This is the most important role for software such as Recorder.

Data standards in Recorder

The scope of Recorder

Figure 2 gives a simplified overview of Recorder's 'information space' for biodiversity records.

The left-hand side of the figure is occupied by the Thesaurus and dictionaries which provide names and controlled terms that are referenced by all parts of the system but especially to describe and identify observations, specimens and places. The thesaurus can also link to other information sources and provide contextual information that

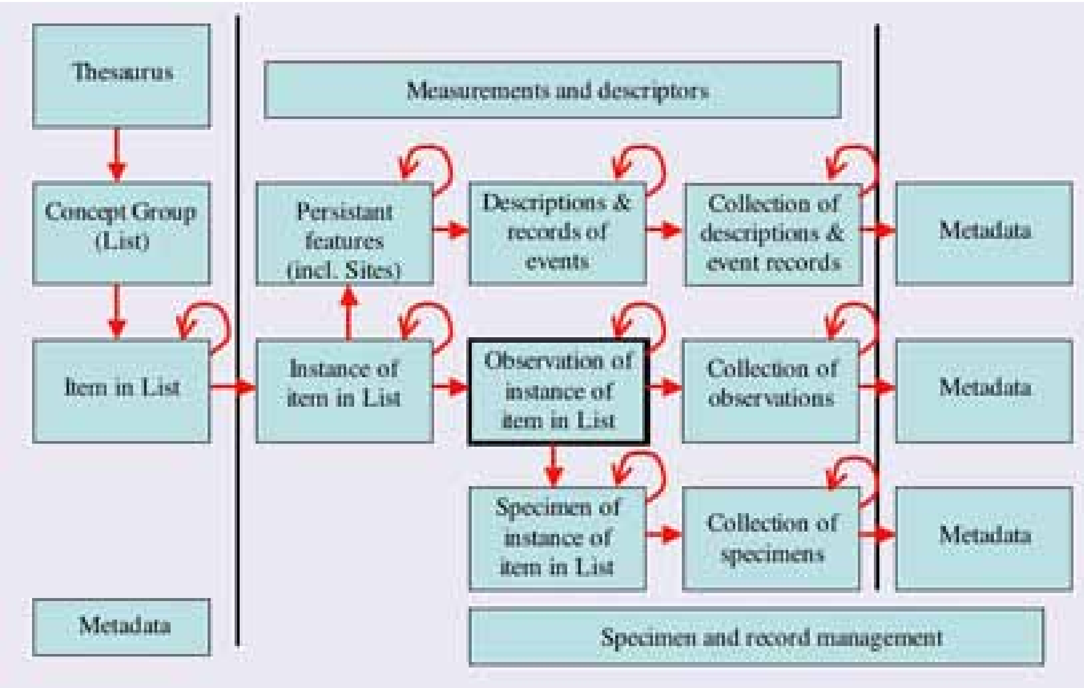


Fig. 2: A simplified overview of Recorder's 'information space' for biodiversity records.

aid in the interpretation of the managed data.

Every item in the system about which identifications, measurements, descriptions or management information is stored is an instance (or occurrence) of an item in one of the thesaurus, dictionary or location reference lists. (The ambiguity here refers to the difference between the Field Survey part of Recorder which still uses separate dictionaries and the newer Collections Module that uses a multi-purpose thesaurus). This instance may be associated with an observation (e.g. a field record of a plant) and it may also be linked to some persistent feature such as a named site.

Instances may be related to each other in many ways (temporal, spatial, ecological) and observations may also be linked and be part of a collection of observations (e.g. an organised survey). Persistent features, such as sites may be of interest to the system and have their own descriptive and other data attached to them and may also be parts of collections of data (e.g. a list of sites of wildlife importance). Observations may give rise to specimens which then become the focus of more types of identification, measurement and management information and specimens may also be aggregated into collections. Term lists, features, observations and specimens all attract their own metadata relating to scope, format, origin and ownership.

The information space managed in Recorder is, therefore, rather complex as it needs to maintain many potential types of relationships between gathering events, field observations, specimens and collections, geographic locations and features and also track a potentially vast range of descriptors, measurements, curatorial actions and management events. This is achieved through the use of the controlled terminologies in the thesaurus and dictionaries, the physical structure of the data model and the functionality built into the application.

The Recorder Data Model

Scope of the Recorder Data Model

One of the problems that still persists in biodiversity recording is that much of the collected data is held in idiosyncratic applications designed to store data in the format easiest for generating the main output. Attempting to collate and re-purpose such data can be very challenging and tends to exacerbate the idea that there are fundamental gulfs between different types of biodiversity data (e.g. observations, sites and specimens). The model presented in Figure 2, however, indicates that there are clear relationships between these types of record and that they can be harmonised into a single data model.

The realisation that it is possible to create a single model for all biodiversity data led independently to the development of two information models, one developed for Recorder (see Copp 2004, 2006) and another under the European BioCISE and ENSHIN projects (Berendsohn et al. 1999). These models agree very closely in their detail and both can be mapped to the ABCD Schema (ABCD Schema 2006) developed through TDWG and BioCASE as a comprehensive standard for access to and exchange of biological records. These efforts predate current work on defining a core TDWG ontology (TDWGwiki TDWGOntology 2006) which may become an important tool for mapping various discipline specific models in future web-based applications and portals.

Recorder has been built using a physical model that closely follows the NBN logical data model. The NBN Conceptual Model includes six key modules, listed below and illustrated in *Figure 3*;

- Surveys (field observations and gathering events)
- Locations (named places and collecting sites)
- Collections (specimens and their management)
- People & Organisations (contacts and addresses)
- Sources (publications and images)
- Thesaurus.

These top-level modules may incorporate further sub-modules, for instance, *Sources* includes references and images. At the lower level, modules are comprised of entities that group and describe the relationships of individual items of data. There are other potential main modules that could be added to extend the model (e.g. finance control) but these were not within the scope of the Recorder development.

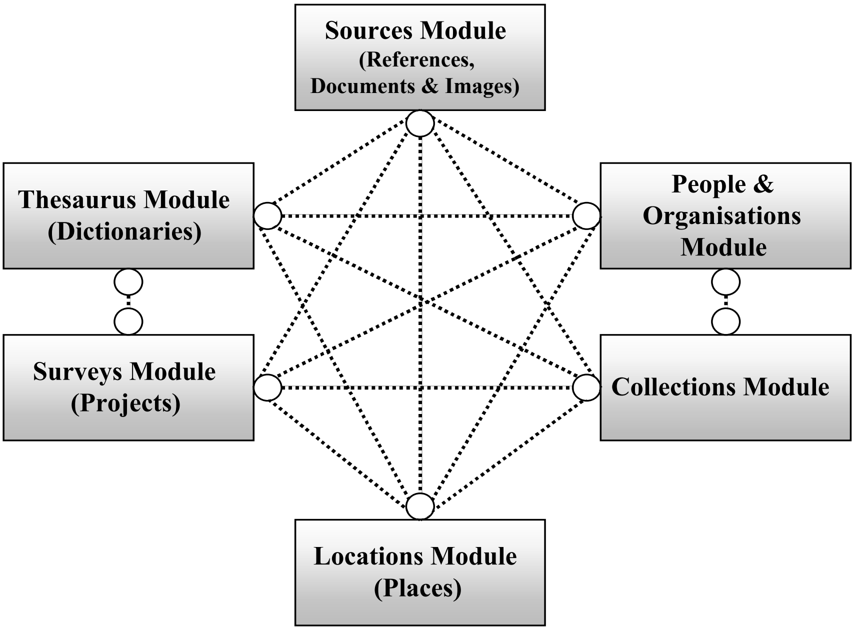


Fig. 3: The NBN Data Model, used in Recorder, consists of a number of inter-related modules, that can call each other and a number of properties (measurements, descriptors, identifications and spatial references) that can be used by any module.

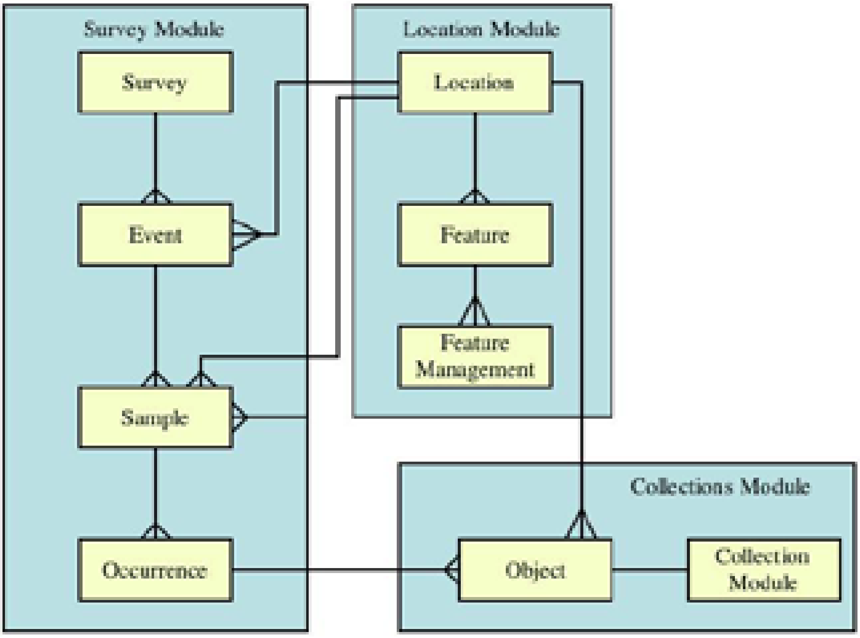


Fig. 4: Each module consists of numerous tables (simplified in this diagram) or sub-modules containing related tables. The design has been optimised for wildlife and earth science data but is readily adaptable to other disciplines although there might be a need to add facilities for recording specialist methodologies (e.g. for recording archaeological contexts).

The modules cross relate to each other, for instance, the thesaurus provides the controlled terminology that would be used in applications featuring the other modules. *People* and *sources* can be referenced from any other module. In addition to the main modules, there are a number of Common Entities or properties that can appear in any module. These common properties include measurements, identifications (determinations) and spatial (geographic) references.

The principal data modules are the **Survey**, **Location** and **Collections** modules. These are illustrated in a simplified form in Figure 4. and are the only parts of the data model described in this paper. Full details of each module and physical data models are available on the Recorder website and wiki (EIMwiki, n.d.; Recordersoftware, n.d.).

The Recorder Survey Module

The Survey Module holds all the information related to field observations and recording methodologies. The essential characteristics of the Survey Module are illustrated in Figure 5.

The structure of the Survey Module is based on the concept that every observation or find is a type of 'occurrence' that can be linked to a specific sample and sampling technique. For instance, in *Figure 5*, the sampling technique is a *Longworth Mammal Trap* and this has an occurrence record for a vole caught in the trap and for the habitat within which the trap was placed. An occurrence can be anything that is observed in relation to a sample, ranging from a plant or animal (taxon occurrence) to minerals, fossils, soil type, geomorphological feature or stratigraphic layer and much more. It is thus possible to record the intersection of many things within a single sample (e.g. linking a fossil to a biostratigraphic zone).

Any number of occurrences can be linked to the same sample and any number of samples can be linked to a survey event. In Figure 5 this is represented by a number of traps set at the same time by the same person in a single area (e.g. a trap line) in the event record. The survey event links people and samples to a location (a site or a spatial reference) and a date or set of related dates, although individual samples can have their own more specific dates and spatial references.

In Recorder, survey events must be linked to a 'survey' which is in fact any collection of metadata that states who is responsible for the data, what purpose it was collected for and the wider geographic and time coverage of any group of related records (observations and finds). A 'survey' can be at any scale, for instance, The Beagle Voyage, the 2005 Butterfly Survey for Luxembourg, A single geological mapping project, Charles Copp's Garden Bird records etc.

Figure 5 also shows that objects/specimens (here the skull of a vole) can be linked to the Survey Module through occurrences. The Collection Module also allows the more general attribution of locality data and inferred information to objects that do not have accurate provenance data.

The Recorder Locations Module

The Recorder Locations Module is illustrated in *Figure 6*. A more complete explanation of the Location module can be found in the original Recorder Systems Analysis document and in the NBN data model document (Copp 1998). The purpose of the Location Module is to provide the means of managing information about 'sites' of interest to the system i.e. ones that have or are likely to have field observations or objects linked to them or those with features subject to grading, management plans and monitoring. Sites can have sub-sites to any level and so can be used to create complex site hierarchies The Location Module is not intended to act as a general gazetteer, rather, the data collected extend that available in the gazetteer, which would be managed in the Administrative Areas dictionary or in the Thesaurus Module (depending which version of Recorder you are using).

The key entity is **Location** (Location sub-module). In Recorder the Location table includes minimal data because so much of the information about sites may be multiple (e.g. alternative names for a site) which in a relational database are stored in separate tables.

A second key entity is **Location Feature** (Feature sub-module), which can be linked to sites or sub-sites. Features can be any item of 'interest' that is regarded as a persistent feature of the location. This could be a geomorphological feature, an ancient tree, a population of wood-boring beetles and much else. The feature is the entity to which management plans, damage records or management events can be linked. There is an extension to Recorder (The Parks and Woodland Pasture application) that has extended the Feature sub-module to include hierarchical relationships between features and also extended the descriptive and measurement aspects. The current Recorder application would need some modification to enable the linking of observations to features rather than just sites and specific spatial references as at present.

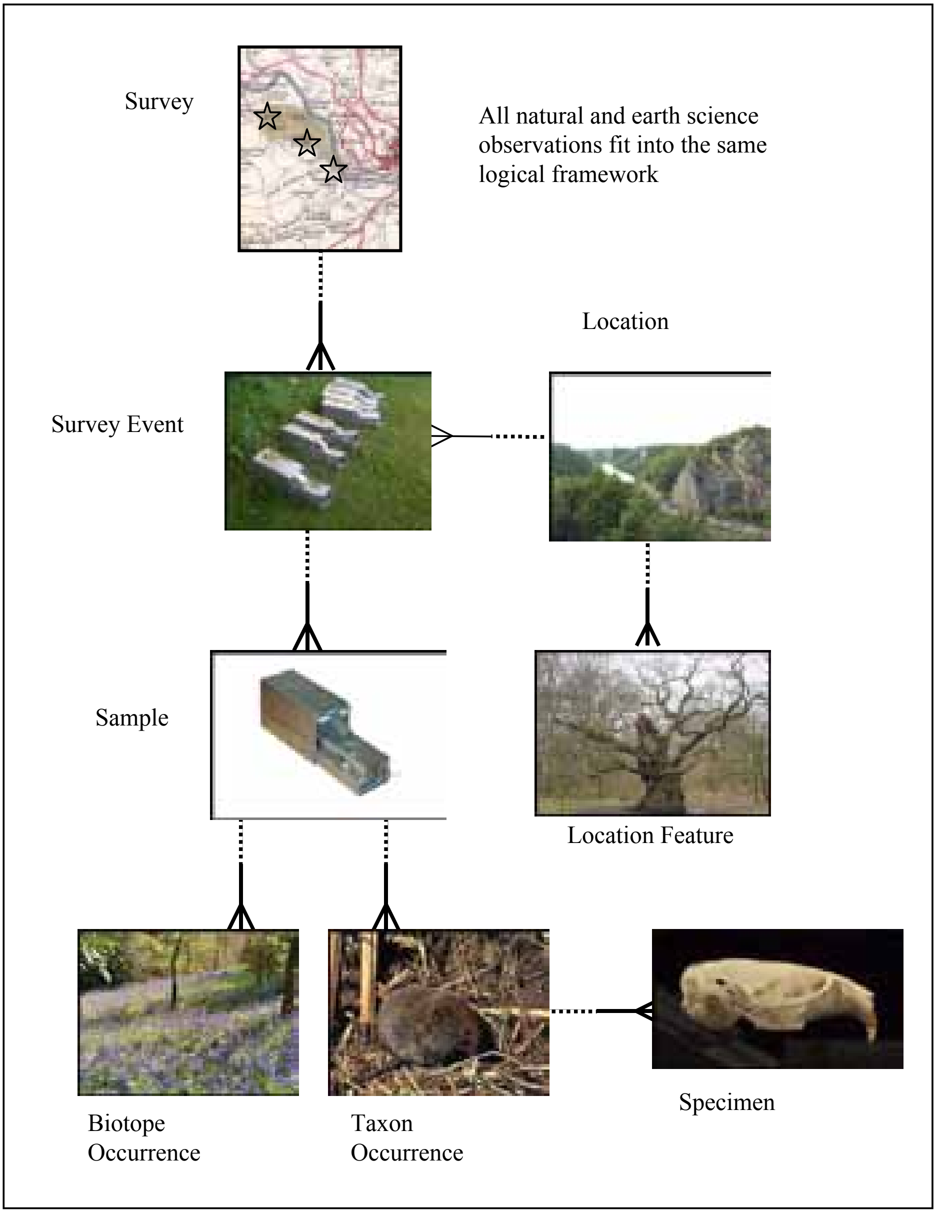


Fig. 5: A schematic diagram of the Recorder Survey Module illustrating how field observations and specimens relate to sampling events and the grouping of records into 'surveys'.

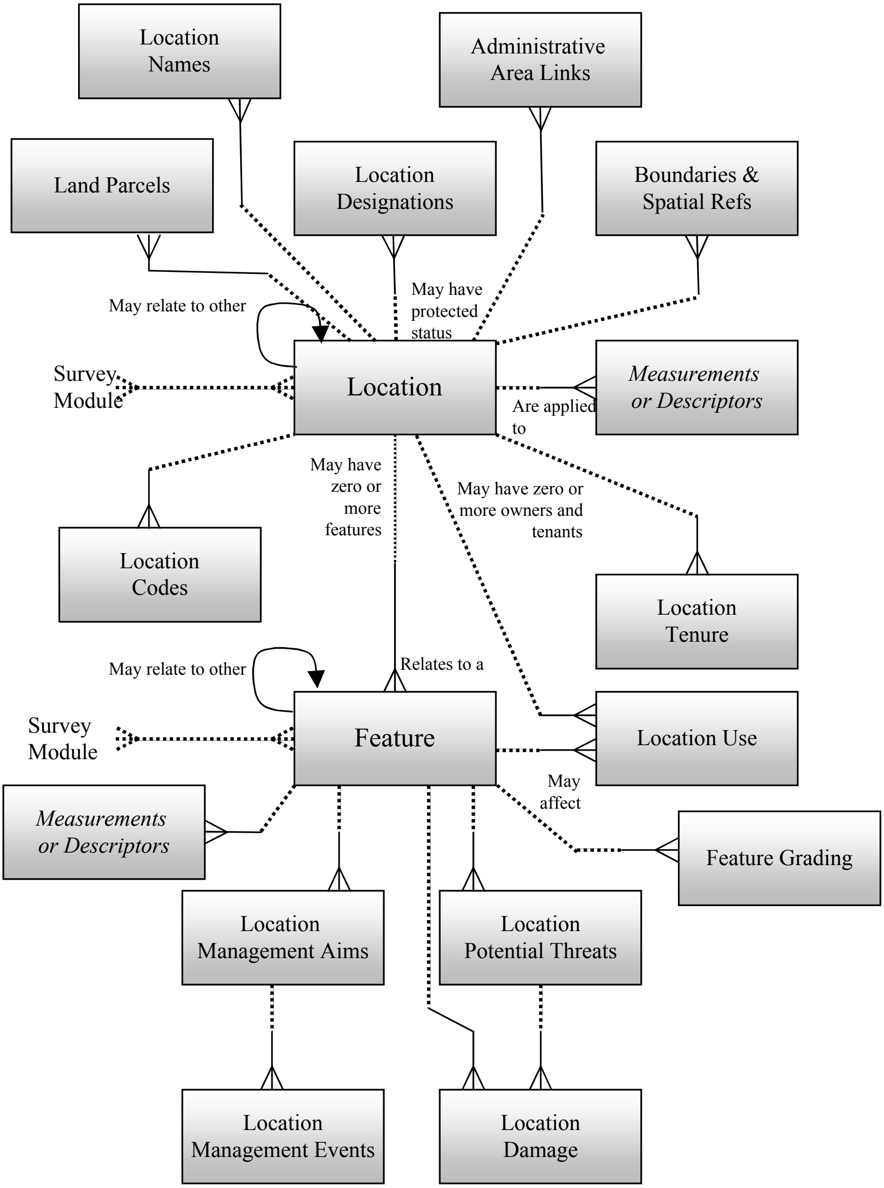


Fig. 6: The Recorder Locations Module allows extensive information to be linked to named sites and sub-sites. Digital site boundaries are stored as links to either polygons in the Recorder map module or to an external mapping system. Images, scans, documents and references relating to sites and features are all handled by links to the Sources Module.

Key: Designations = protection status or schedule. Land Parcels = numbered or otherwise identified areas of land that fall within or overlap site boundaries. Administrative Area links = allows linking site to named administrative and other areas - mainly used as a text alternative to GIS. Location Codes = Any numeric or alpha-numeric codes used to identify the site (e.g. paper file reference), site number etc. Location Tenure = Ownership and occupation of land. Feature Grading = Can be used to record grading systems e.g. for educational potential, quality assessment etc.

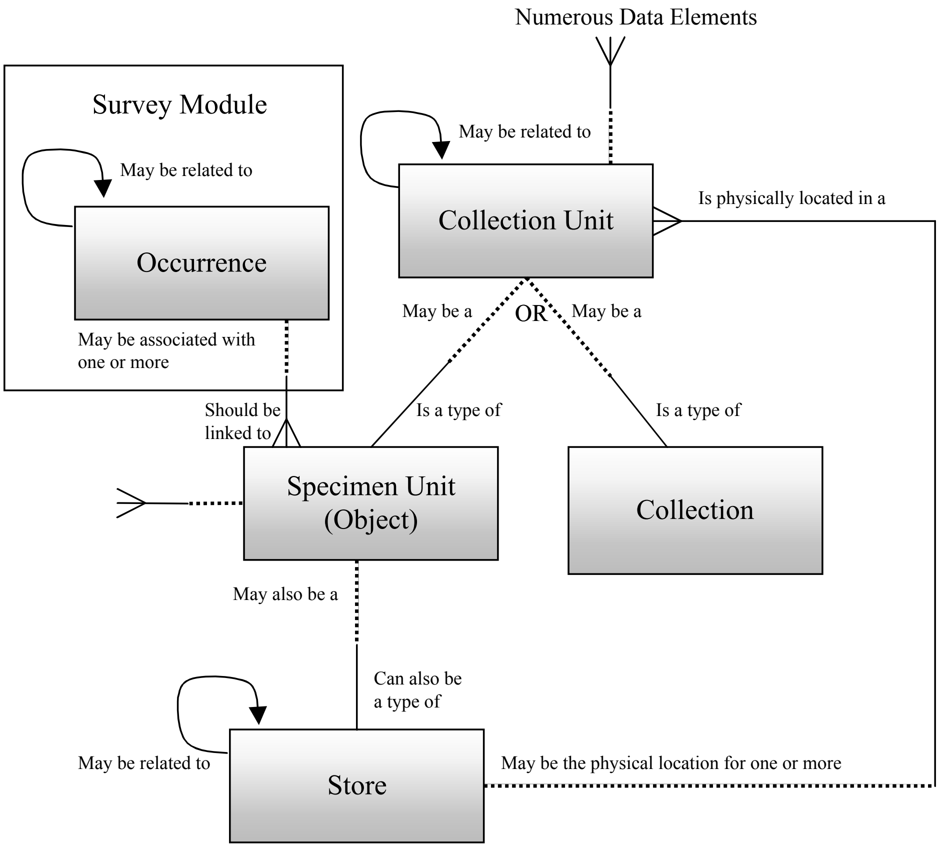


Fig. 7: Specimens, collections and stores form part of the Collection Unit entity.

The Recorder Collections Module

The Recorder Collections Module is a very large extension to the Recorder application that was commissioned by the Luxembourg National Museum of Natural History. It adds not only a number of additions to the original Recorder physical model but much new functionality in the application. For instance, the scope of Recorder was not only extended to cover specimens and collections management but to include earth sciences records. The new functionality includes a whole new means of managing term lists (Copp 2007) and a new multi-purpose data browsing window called the 'Collections Browser'.

One of the problems which has arisen, because of the way that changes to the core of Recorder are currently controlled and the extensive nature of the additions associated with the Collections module. In fact, the way data are stored as well as certain aspects of terminology control differ between the original and new parts of Recorder. Items entered using the thesaurus as a term source cannot be searched for using the original Recorder Report Wizard. These problems can be overcome using different techniques and do not undermine the strict standards conformance of Recorder but are issues that will need to be resolved as the use of Recorder spreads in Europe and particularly with increasing use of the Collection Module.

Central to the Collection Module is the concept of a Collection Unit (Fig. 7). A Collection Unit represents either a specimen (object) or a collection of specimens. Collection Units are located in stores which may themselves be a type of specimen. For instance, a single pinned butterfly might belong to a named collection housed in an antique entomological cabinet. Each of these items (specimen, collection and store) can share many common attributes, such as each may be accessioned, owned, loaned, conserved, valued, numbered, identified etc. Collection Units can have unlimited levels of hierarchy (e.g. sub-units of sub-units); collections contain specimens, specimens may be derived from other specimens, cabinets have drawers and so on. The Collection Browser window lets the user follow and explore these complex links.

The concept of a Unit was adopted as central to the BioCISE data model (Berendsohn et al. 1999) and is also a key feature of the ABCD schema. In the BioCISE/ABCD models, field observations are also treated as units whereas in the NBN/Recorder Data Model occurrences (and linked observations) are treated as part of the separate but linked Survey Module. The reason why the NBN/Recorder model separates these concepts is that it makes it simpler to use the model for defining practical applications that focus either on field data or the management of physical objects. Modelling is always a balance between abstraction and application, the separation of these concepts helps modularise the data.

The Collections Module is complex both in its logical and physical build forms and cannot be described in detail in this paper but more details are available on the Recorder wiki (Eimwiki n.d.) and will be updated regularly. Some of the types of data that may be attached to collection units are illustrated in Figure 8.

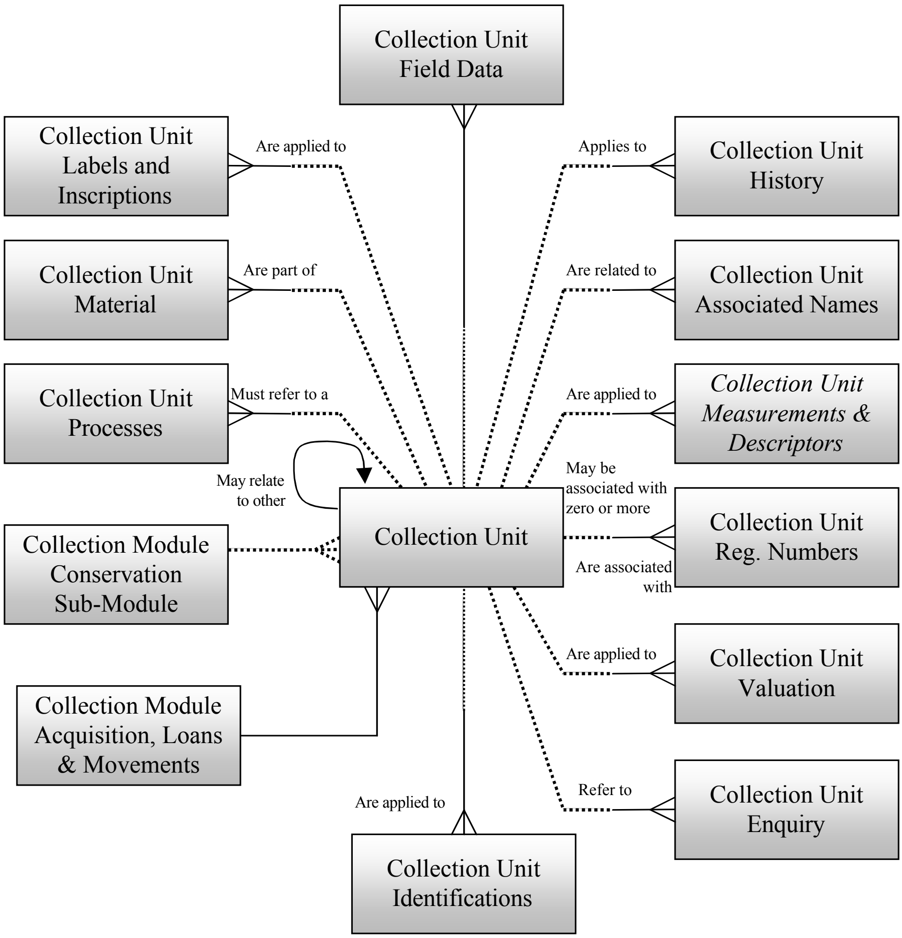


Fig. 8: The main descriptive data elements associated with collection units and links to conservation and management (movements) modules.

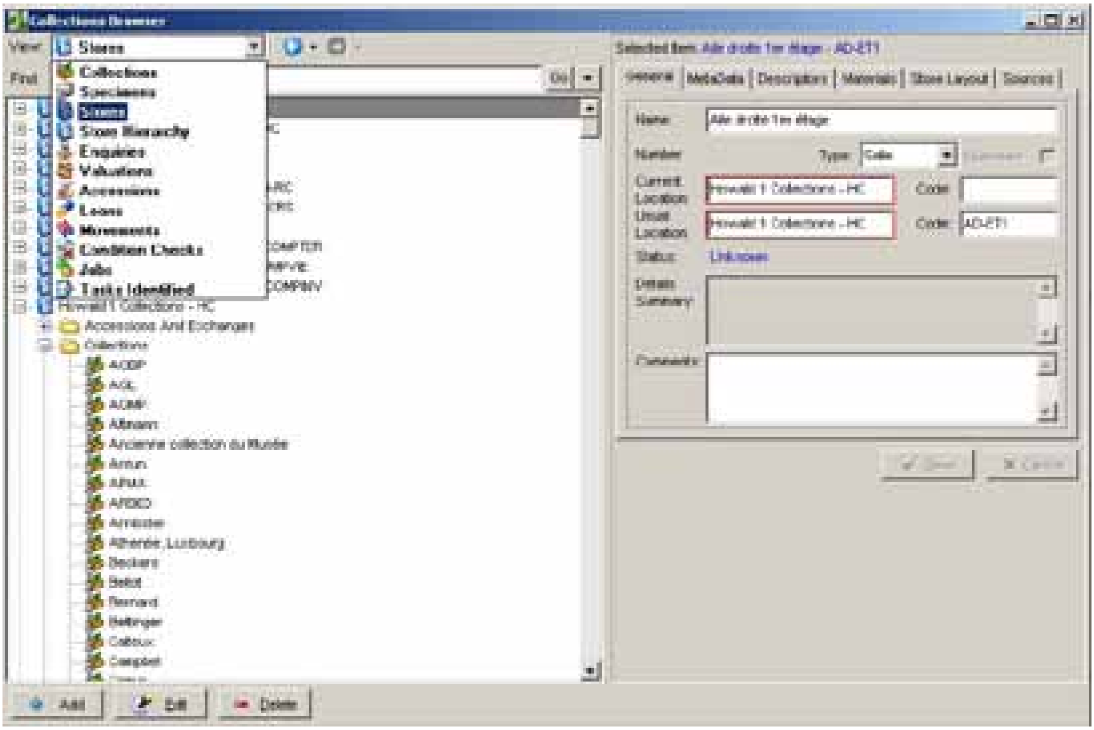


Fig. 9: The Collections Browser allows the user to explore the data model from many perspectives. In the figure the left pane shows a list of stores, one of which has been opened and the list of collections associated with that store displayed. The drop-down menu lists other ways of viewing the database.

A further key aspect of the Collections module data model is the use of the concept of movements. Many of the actions relating to specimens or collections that are documented can be regarded as movements, for instance acquisition, storage, loans, being placed on exhibition, transfer between departments and disposal. The movements are normally inward or outward and may or may not include a change of ownership

The collection module tables include many constraints and format controls to ensure data validity alongside many triggers and stored procedures that carry out further validation and processing. *Figure 9* shows the Collection Browser window which is the main way that user interact with their collection data. The Collection Browser allows many views of the data within one window. Multiple copies of the window can be opened, for instance there might be a copy open displaying stores and what collections are found in each store whilst a second copy may be opened from the first to display the details of a selected collection and so on. Users can add, edit or create reports on the data from any view. The Recorder application itself is intrinsic in delivering data validation through controls and checking functions at every stage on the various forms.

Other Recorder Standards

Recorder Terminology Standards

Recorder makes extensive use of dictionaries and the thesaurus to control the content of individual fields (e.g. taxon names, gazetteers, collection methods etc.). The use of the thesaurus is described in a separate paper in this volume (Copp 2007).

Format and Syntax Standards

The format or syntax of data entered into many different fields (e.g. personal names, dates, grid references. and measurement units used) form **Syntax conventions,** some of which may be existing international standards (e.g. ISO dates). In Recorder, syntax conventions are enforced both in the table definitions and within the data entry forms, often by calls to stored procedures. For instance, spatial references such as UK Grid References are checked not only for format (are they well formed?) but also for meaning (e.g. do they fall within the given geographic area) and on saving are also converted into latitude/longitude coordinates that can be used for reports that merge records from different sources or for sending to outside utilities for further validation.

Data Transfer Standards

Historically, most databases were built for specific purposes and not optimised for data sharing, often data could only be fully copied to databases with identical structures and this approach remains popular among closed networks of users. This situation started to change rapidly with the spread of web-based information systems, particularly through the use of XML for tagging data and the creation of Federation Data Schemas.

Sharing information between non-identical systems normally requires translating data from the structure and format of the provider database into some other format suitable for transmission or copying and which can be used, sometimes with further transformation, as input to one or more other information systems. This has traditionally been difficult to accomplish and over the years many, usually domain specific, **Data transfer standards** have been developed. The rise of XML both for creating data schemas and encoding data has enabled the development of generalised and more flexible systems. Foremost among the schemas developed for biodiversity records are ABCD (see BGBM n.d.) and Darwin Core (see TDWGwiki Core 2006).

The Recorder Project started before the development or availability of ABCD and Darwin Core, but it was decided to use XML for data transfer between Recorder systems. For this purpose a Recorder Data Transfer Standard based on an XML DTD was created. The Recorder DTD covers the whole of the original NBN model and Recorder includes an export management utility that allows the user to select surveys and data, using filters, to export as Recorder XML files. Unfortunately it was found that big datasets can create enormous XML files and the subsequent parsing and processing can be slow so Recorder also includes facilities to export and import data in the form of zipped (compressed) MicroSoft Access database tables.

Recorder cannot export data by choosing external xml templates but it is possible to make information available by means of a wrapper that is mapped to the table structure, or a snapshot view or a data cache created by an SQL query. A typical scenario would be, creating one or more cache tables that hold copies of the data that are to be made available and using a wrapper such as the PyWrapper (PyWrapper n.d.) to map the fields to the ABCD (or similar schema), so that queries could be sent to the wrapper, over the Internet using the BioCASE or TAPIR (TDWGwiki TAPIR 2007) protocols as the messaging system. In this way Recorder databases can become part of heterogenous biodiversity data networks and accessible by any number of portals or web services.

Metadata Standards

Metadata standards are models for describing the content, format and accessibility of individual datasets. The definition and scope of metadata standards has become a complex discipline because it is required in many informatics domains, including libraries and museum communities and has become a major research area because of the demands for improved 'resource discovery' and the development of the so-called 'semantic web'.

The development of Recorder began before the current fashion for ontology building and rdf-based (w3 2007) resource discovery services and therefore, is rather 'old fashioned' in its approach to metadata. Metadata information occurs throughout Recorder mainly embedded as attributes (fields) within the many tables. Most of the metadata is therefore, held in context and intended to be used that way. The metadata fields built into Recorder are extensive but the collection module also includes a thesaurus-linked metadata tab (see

Fig. 10: Collection metadata includes information on who assembled the collection, its ownership history, time coverage, related bibliographic references and many other descriptors.

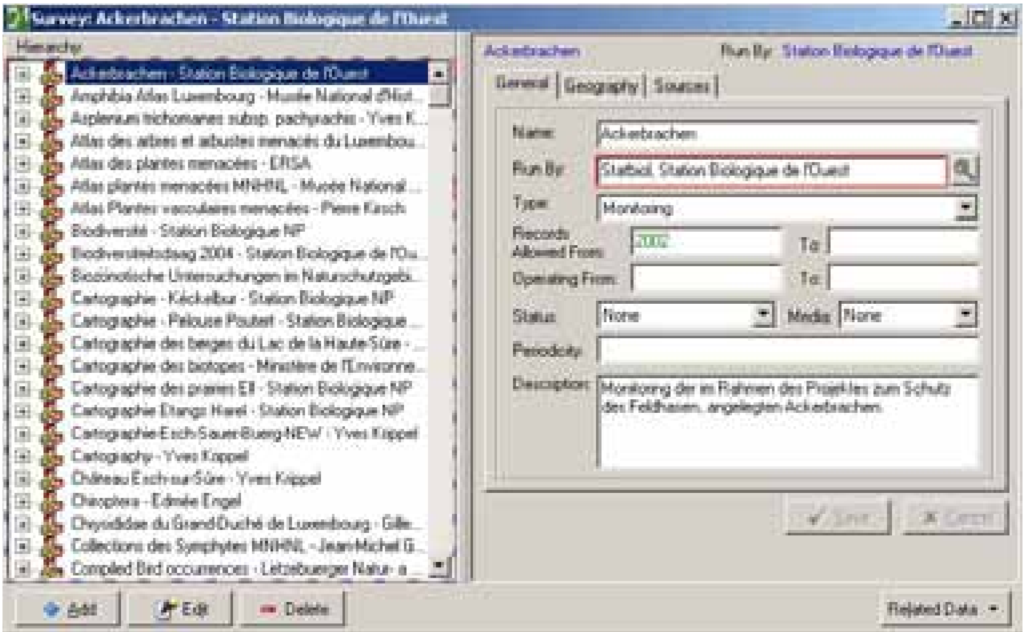


Fig. 11: The ‘Survey Window’ in Recorder showing a list of field surveys held on the database and metadata relating to the highlighted survey in the right-hand pane. The geography tab allows the user to define a bounding box for the extent of the survey or link it to a named geographic area.

Figure 10) that allows any number of user-defined metadata items to be added to most of the entities within the module. Recorder probably collects more metadata than any other comparable application. The drawback is that there are, currently, no built in functions, to extract this information for modern directory and resource discovery services.

All records of observations and objects catalogued (and also entries in the thesaurus) in Recorder have a custodian field associated with them which defines who is allowed to edit the content and acts as a pointer back to the authority for issues of record use and copyright. All records also include information relating to when and by whom they were entered and last edited. Critical information, such as taxon determinations, cannot be deleted (generally) but are multi-valued, allowing for changes of opinion and attribution (and who made the entry) to be tracked. Records may also be flagged as to their validation and confidentiality and such records can be filtered out of reports and exports.

All records in Recorder are identified by unique 16 character NBN Keys (e.g. MNHNL00000000001), which identify (in the first 8 characters) the system in which the record was originally created. This also helps in tracking records when they are transferred to other Recorder systems, as happens when collaborators submit records to a collating record centre or recording scheme organiser.

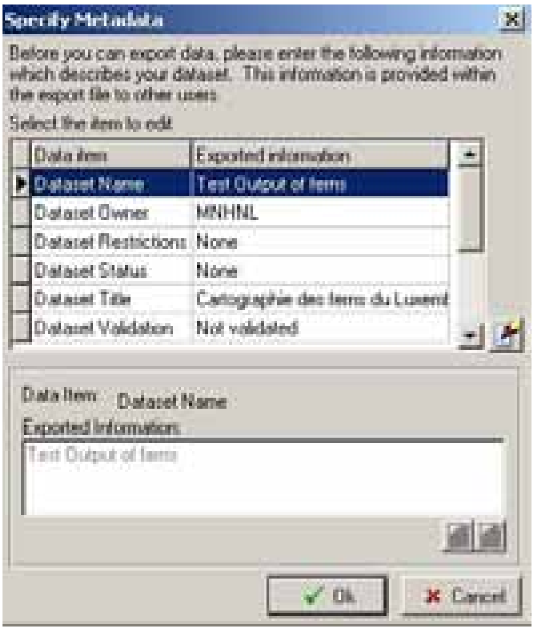


Fig. 12: The Recorder export function includes a form for specifying metadata information to be included in the XML header of the exported data file.

Groups of records such as surveys and collections of specimens have their own metadata fields that include geographic coverage, organiser, purpose of the survey/gathering event(s), original recording media and more (Fig. 11). This information has recently been extended. When the user chooses to export data, for instance to export all records for a survey, there is a special metadata screen that requires the user to enter further details relating to validation, ownership and use, which are written into a metadata header in the output XML file. In the UK NBN Gateway, data providers can also enter and manage metadata related to the data that they make available through the gateway, online, without sending the data as part of the data file.

Recorder stores extensive metadata information related to records and groups of records and provides the means to add user-defined metadata attributes. Efforts have been made throughout the development period to align Recorder's metadata fields with other metadata standards but facilities do not yet exist to gather the metadata together and to export it in standard formats or as rdf output. These could be developed if needed in the future.

Operational Standards

Recorder supports operational standards in a number of ways:

Recording and sampling methodologies:

Recorder can be used to store information related to many types of recording methodology, although the standard system does not have specific methodology input screens. Special versions of Recorder have been created with specific methodology record forms for Parks and Lowland Woods and also for Marine recording.

Verification and validation systems:

Verification and validation functions are built into the Recorder application at the attribute (field) level. There are no integral checks of records against known ranges or phenological distributions but these can be applied using external tools. The use of dictionaries and the thesaurus together with contextual checking of spatial references help to reduce transcription and data entry errors.

General operational standards

Recorder is closely allied to the UK National Biodiversity Network and also to the UK National Federation for Biological Recording (NFBR), organisations which strongly promote the development of sound operational standards for biodiversity records management and also the growth in Local Records Centres and the establishment of biodiversity records management as a profession. Among the topics on which documentation and standards have been developed are:

- Intellectual property rights and copyright controls
- Data exchange and sharing agreements
- Controls and arrangements for access to the data
- Quality standards for data management
- Arrangements for ensuring the physical security of the data
- Backup and archiving systems
- Accreditation and training of records management staff and record centre
- Means for accrediting field recorders

These issues are outside of the scope of this paper but can be followed up through the NBN (NBN n.d.) (such as in the NBN Handbook) and NFBR (NFBR n.d.) websites.

References

ABCD Schema 2006. - Botanischer Garten und Botanisches Museum Berlin-Dahlem hosted webpage of the ABCD Schema - Task Group on Access to Biological Collection Data. Available at http://www.bgbm.org/TDWG/CODATA/Schema/. Accessed March 2007.

BioCASE n.d. - The Biological Collections Access Service for Europe. Access protocol webpage available at http://www.biocase.org/products/protocols/index.shtml. Accessed March 2007.

Berendsohn W.G. et alia. 1999. - A comprehensive reference model for biological collections and surveys. Taxon 48 August 1999 pp511-562. Retrieved March 2007 from http://www.bgbm.org/biodivinf/docs/CollectionModel/

Copp C.J.T. 1998. - The Recorder Project: Systems Analysis. January 1998 unpublished report. JNCC, Peterborough. Version 16/11/2000 available at BGBM website at http://www.bgbm.org/biodivinf/docs/archive/Copp_C_1998_-_The_Recorder_Project.pdf. Accessed March 2007.

Copp C.J.T. 2004. - The NBN Data Model. Part 1: Description of the Model. June 2004 unpublished report. Available on NBN web site at www.nbn.org.uk

Copp C.J.T. 2006. - The NBN Data Model. Updated version of 2004 JNCC report. Available on EIMwiki at http://eim.metapath.org/wiki/index.php?title=Data_Models. Accessed March 2007.

Copp C.J.T. 2007. - An introduction to the Recorder Thesaurus: x-x, in Walisch T. (ed.), Collating and managing natural sciences field and collection records in Europe. Proceedings of the First International Recorder Conference. Luxembourg 2-3 December 2005. Ferrantia 51, Musée national d'histoire naturelle, Luxembourg, xxx p.

Dublincore 2007.- Dublincore Metadata Initiative® . Website available at http://dublincore.org/. Accessed March 2007.

EIMWiki n.d. - EIM hosted wiki for Recorder, available at http://eim.metapath.org/wiki/index.php?title=Recorder. Accessed March 2007.

NBN n.d. - National Biodiversity Network website available at www.nbn.org.uk. Accessed March 2007.

NFBR n.d. - National Federation for Biological Recording. Website available at http://nfbr.org.uk. Accessed March 2007.

Opengeospatial 2007. - Open Geospatial Consortium website available at www.opengeospatial.org. Accessed March 2007.

PyWrapper n.d. - Pywrapper website available at http://trac.pywrapper.org/pywrapper/ . Accessed March 2007.

Recordersoftware 2006. - Dorset software hosted website on Recorder, available at www.recordersoftware.org. Accessed March 2007.

TDWG n.d. - Biodiversity Information Standards (TDWG) website available at www.tdwg.org. Accessed March 2007.

TDWG NCD 2007. - Biodiversity Information Standards (TDWG). Natural Collections Description Group webpage available at www.tdwg.org/activities/ncd/charter/. Accessed March 2007.

TDWGwiki TAPIR 2007. - Biodiversity Information Standards (TDWG) hosted wiki. TAPIR protocol webhome available at http://wiki.tdwg.org/twiki/bin/view/TAPIR/WebHome. Accessed March 2007.

TDWGwiki TDWGOntology 2006. - Biodiversity Information Standards (TDWG) hosted wiki . TDWGOntology webpage available at http://wiki.tdwg.org/twiki/bin/view/TAG/TDWGOntology. Accessed March 2007.

TDWGwiki Darwin Core 2006. - Biodiversity Information Standards (TDWG) hosted wiki . Darwin Core WebHome available at http://wiki.tdwg.org/twiki/bin/view/DarwinCore/WebHome. Accessed March 2007.

Wikipedia n.d. - Wikipedia article and links on LSID at http://en.wikipedia.org/wiki/LSID. Accessed March 2007.

World wibe web consortium 2007 - Webpage on the Ressource description Framework available at http://www.w3.org/RDF/. Accessed March 2007.

Recorder use on a local level: collating and creating products

Charles Roper
Sussex Biodiversity Record Centre
Woods Mill
GB-BN5 9SD Henfield
CharlesRoper@sussexwt.org.uk

Keywords: data standard, Sussex biodiversity record centre, data collation, data retrieval

Abstract:

A local record centre inhabits a unique space, not only providing local and, to an increasing degree, regional decision makers with crucial biodiversity information, but also supporting the needs of local naturalists and ecologists, from where the vast majority of records in an LRC's database come from.

The dynamic nature of an LRC's data flow and varied analysis needs requires a standards based approach to data storage and retrieval. Data coming in from a huge variety of sources, each with subtle (and not so subtle) differences in form and content, require a flexible, modular and scalable framework within which to be stored. The diverse range of information sought by local authorities, NGOs, commercial organisations, members of the public and other users of LRC data require data to be extracted, manipulated and presented in a huge variety of ways. Crucially, data needs to be shared to be made most effective, and the role of the LRC is not only to collect and collate data, but also to pass them on to county recorders, schemes and societies and, of course, the NBN.

To achieve these goals, record centres require information standards to act as the foundation for their work. These standards come in the form of the Recorder software and its underlying ontology, the NBN Data Model. This contribution will show how the Sussex Biodiversity Record Centre is utilising the Recorder platform in its work to build bespoke reporting and analysis systems, with a special emphasis on the Rare Species Inventory (SxRSI); and how the standards based nature of the model allows for us to easily build into a regional, national and now European biodiversity network.

The need for local biodiversity data

The need for high-quality, up-to-date biodiversity data is on the increase as tremendous pressures are placed upon our natural environment from a variety of sources. Key decision makers need to have a clear picture of the state of the environment, and they need it in a timely manner. Local planning guidance in Britain is now taking biodiversity seriously into account (Office of the Deputy Prime Minister 2005). Busy town and country planners need biodiversity data to be presented from a local perspective, and in an easily digestible manner. Government agencies need solid local data to be available at a moment's notice in order to inform policy. The UK Biodiversity Action Plan requires data with a local flavour for the LBAP (Local Biodiversity Action Plan) process (UK BAP, n.d.). And so the list goes on. The need for locally informed data containing all the subtleties and nuances conveyed upon them through the knowledge of local naturalists and experts cannot be understated.

Local Record Centres

It is within this local arena that local record centres play their vital role. LRCs in Britain form a network of organisations dedicated to collecting, collating and disseminating species and habitat data. Local record centres provide a level of knowledge, intimacy and detail in regards to local and regional issues that organisations serving the wider national interest cannot achieve. And so local record centres inhabit a unique space; not only providing local and, to an increasing degree, regional decision makers with crucial biodiversity information, but also supporting the needs of local naturalists and ecologists, from where the vast majority of records in an LRC's database come from.

The dynamic nature of an LRC's data flow and varied analysis needs requires a standards based approach to data storage and retrieval. Data collected from a variety of sources, each with subtle (and not so subtle) differences in form and content, require a flexible, modular and scalable framework within which to be stored. The diverse range of information sought by local authorities, government agencies, NGOs, commercial organisations, members of the public and other users of LRC data require data to be extracted, manipulated and presented in a number of ways and for a range purposes. Information needs to be presented appropriately to audiences with radically different levels of knowledge; in other words, raw data needs to be converted into information that ultimately *informs*. Crucially, data need to be shared to be made most effective, and the role of the LRC is not only to collect and collate data, but also to pass them on to county recorders, schemes and societies and, of course, the UK's National Biodiversity Network and GBIF, the Global Biodiversity Information Facility.

To achieve these goals, record centres require information standards to act as the foundation for their work. These standards come in the form of the Recorder software and its underlying ontology, the NBN Data Model. Recorder affords a great deal of flexibility not only in the data it can hold but also in the ways those data can be used. This enables local record centres to mould the data stored within the Recorder framework in an almost endless variety of ways while safe in the knowledge that the data they hold will remain compatible with record centres and other organisations throughout the country, thus forming a truly distributed national database.

Sussex Biodiversity Record Centre and Recorder

The London and south east region of England is covered by eight local record centres, of which the Sussex Biodiversity Record Centre (SxBRC) is one (SxBRC, n.d.) (Fig. 1). Sussex consists of two counties – East Sussex and West Sussex – and is located on the south coast. SxBRC is a partnership organisation, with the majority of its funding being provided by its partners, which consist of local authorities, national agencies such as Natural England, conservation organisations, NGOs and the private sector. SxBRC's partners, commercial enquirers and the general public require that the centre is able to answer biodiversity enquiries in a detailed, digestible, yet expeditious fashion. It is a difficult balance to strike. Maintaining this balance has meant the development of bespoke software that works integrally with Recorder; this not only enables SxBRC to turn-around enquiries extremely quickly (typically in under an hour) while retaining high levels of detail within reports, but, by using Recorder as the core data store, also ensures the data held by the centre remain compatible with the wider community.

In order to present the most appropriate information to enquirers, data flowing into the centre needs to be filtered and checked by both computer and by human-being. In order to manage theses huge amounts of data (now standing at over 1.5 million records), and to avoid overwhelming enquirers with pages and pages of densely packed information, specialised checking software has been developed that enables two members of centre personnel to check incoming records on a weekly basis. The result of this checking and filtering forms the basis of the Sussex Rare Species Inventory and the Sussex Protected Species Register.

Sussex Rare and Protected Species Inventories

The Sussex Rare Species Inventory (SxRSI) and the Sussex Protected Species Register (SxPSR) form two of the longest running components of the standard SxBRC biodiversity report. They

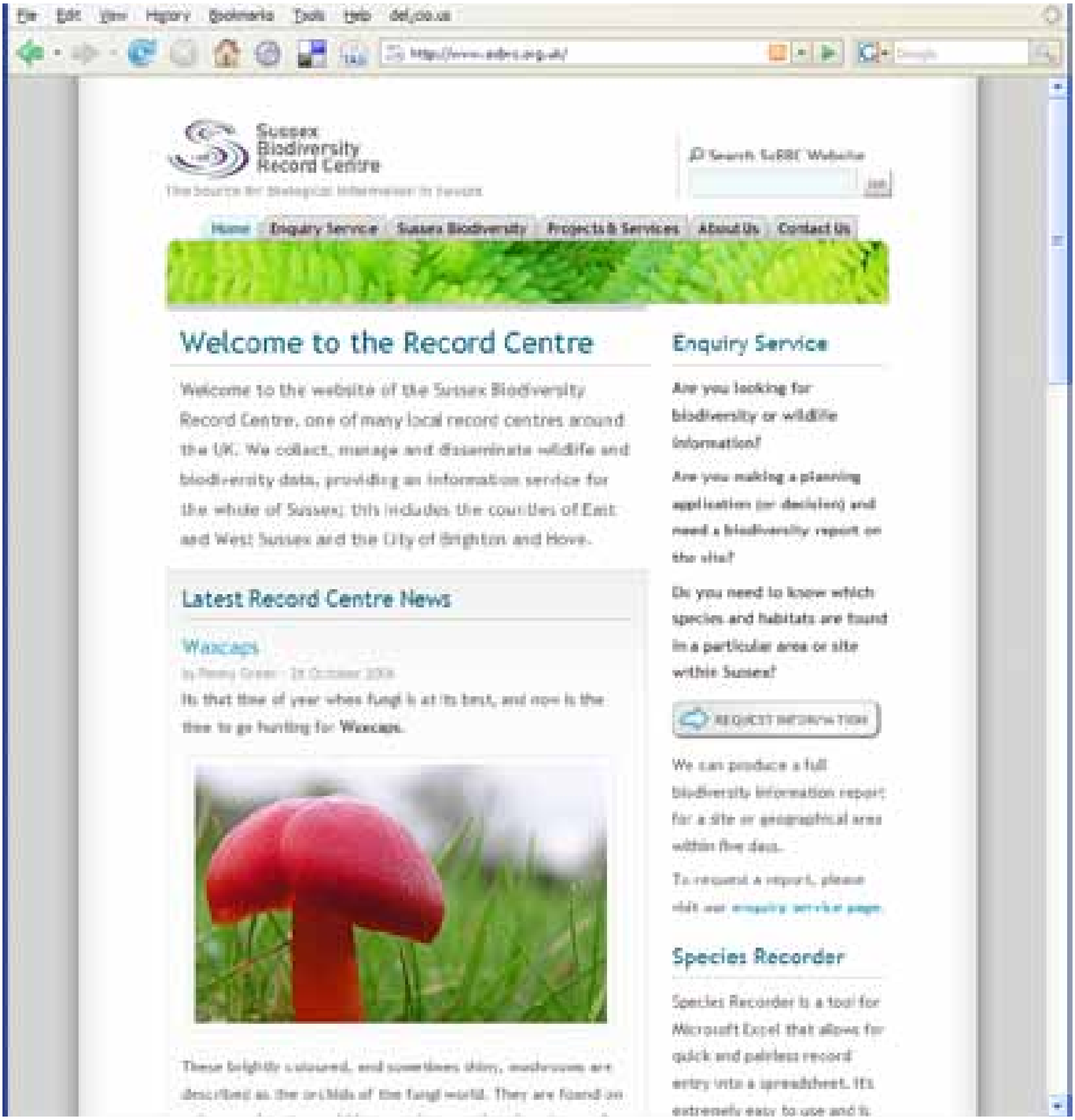


Fig. 1: Web site of the Sussex Biodiversity Record Centre

form two inventories containing select rare and protected species records within Sussex. Somewhat ironically, the rarer species tend to be more widely recorded than the more common species, particularly in protected areas and nature reserves. For example, there are currently only 198 records of the common Brown Rat, *Rattus norvegicus*, in the SxBRC database, while there are 684 records of the considerably rarer Water Vole, *Arvicola terrestris*. In order to temper this proliferation of "rare" records, the "best" record for each species within each 1 km^2 is selected for inclusion in the inventories by hand by an expert. Originally, the data were held in Recorder 3, with each incoming record checked by an expert against a set of criteria. If the criteria matched, the record would be included in the relevant inventory. It was a painstaking and laborious process.

With the introduction of Recorder 2002, SxBRC – with the help of developer Mike Weideli, one of the UK's approved Recorder experts - moved to a much more efficient system that allowed for the two experts co-ordinating the SxRSI and the SxPSR to check incoming records in a fraction of the time and with a much greater degree of accuracy. However, Recorder 2002 is limited to an approximate maximum of one million records and, with the help of the import wizard for importing huge amounts of data, this one million record barrier was quickly reached. Fortunately, the timely release of Recorder 6, with its Microsoft SQL Server database backend, has enabled SxBRC's data holding to grow well beyond the million record mark.

So while the SxRSI and SxPSR checking and reporting systems are entirely unique to Sussex, and offer information relevant to Sussex only, the data from which these inventories are formed remain compatible with other record centre's databases and with NBN standards because they exist within the Recorder framework. Recorder is designed as such that data irrelevant to those outside of Sussex will be automatically stripped out on export, leaving only the more generic, universally useful data behind.

The Future

Now that Recorder 6 has been fully integrated into SxBRC's process, the centre is developing further bespoke reporting tools. The centre is ready to start work on including records into its Biodiversity Action Plan Species Inventory (BAPSI), an inventory much like the SxRSI and SxPSR. By far the largest dataset within SxBRC's database is that of the Sussex Ornithological Society, standing at over 700,000 records. A tool for producing specialised bird reports has been developed and will soon be deployed into the centre's enquiry response service. In order to facilitate the development of this new reporting tool, and in order to ensure maximum benefit across the whole community, crucial aspects – such as the latest Birds of Conservation Concern (BoCC) list – were built directly into the taxon dictionary by John Tweddle of the NBN Species Dictionary Project. Building metadata into Recorder in this fashion, rather than keeping them externally in a separate database, ensures they become available for use by all users of Recorder. In a similar vein, further metadata has been incorporated into the taxon dictionary to enable the development of a tool suitable for producing reports for the Higher Level Stewardship agri-environment scheme (DEFRA, 2003). Again, incorporating these metadata into Recorder ensures maximum benefit to users of the software.

Finally, the Kent and Medway Record Centre (KMBRC) has commissioned Dorset Software to produce an add-in that will allow for the integration of survey metadata into Recorder on a much richer level than is currently possible. Based on the existing Sussex Environmental Survey Directory, the new add-in – called SMART, the *Survey Metadata And Reporting Tool* – has been designed to be as generic as possible and thus it is envisaged will be useful to record centres and other organisations across the country and possibly Europe. By building the tool into Recorder as an add-in, it enables the easy integration of species and habitat data, location data, and the myriad other entities Recorder is capable of storing, with survey metadata. It ensures that record centres and other individuals and organisations can easily install the system and immediately start using it with their data.

Recorder Makes It Possible

So it is in this climate of ever increasing demand and rapidly changing requirements for biodiversity data that Recorder acts as the bedrock of SxBRC's work. A combination of bespoke software, GIS and standard off-the-shelf tools such as Microsoft Office working integrally with Recorder make the goals of the centre achievable with only relatively small budgets and minimal staff. Recorder's versatility and comprehensive design has allowed SxBRC to grow rapidly and has almost certainly played no small part in helping the Sussex Biodiversity Record Centre become one of the strongest and most well respected local record centres in the UK.

References

DEFRA 2003. - Environmental Stewardship guidance. Rural Development Service. Department for Environment, Food and Rural Affairs, UK. Retrieved 23 October 2006 from DEFRA website: http://www.defra.gov.uk/erdp/schemes/es/.

Office of the Deputy Prime Minister 2005. - Planning Policy Statement 9: Biodiversity and Geological Conservation. The Stationary Office, UK. Retrieved 23 October 2006 from http://www.planningportal.gov.uk/england/professionals/en/1020432885185.html.

Sussex Biodiversity Partnership 2001. - The biodiversity action plan for East and West Sussex, and Brighton & Hove.. Website available at www.biodiversitysussex.org/. Accessed 23 October 2006.

SxBRC n.d. Sussex Biodiversity Record Centre Website available at www.sxbrc.org.uk. Accessed 23 October 2006.

UK BAP, n.d. - UK Biodiversity Action Plan. Website available at www.ukbap.org.uk/. Accessed 23 October 2006.

Recorder in the Museum context

Recorder 6 and its collection management extensions

Guy Colling
Musée national d'histoire naturelle
25, rue Munster
L-2160 Luxembourg
gcolling@mnhn.lu

Keywords: Luxembourg Museum of Natural History, museum collections, specimen data, extended NBN data model

Abstract

The Collection Module for Recorder 6 provides collectors of natural history and biodiversity objects with a system allowing them to catalogue, describe and track specimens within their collections. The system allows users to add, edit and delete data for collections and specimens that have been accessioned into the museum (i.e. have been documented as in their ownership). It also allows managing information regarding specimens and collections at various sites. As well as collections and specimens, information regarding stores is maintained for the purposes of organisation and location of specimens. Data gathered by the system will form the underlying dataset for future web-portals and front end museum software used by visitors to the museum. The system also provides facilities for rapid data entry and reporting of this data. The Collections Module is formed of a number of COM addins to the existing Recorder 6 Project.

Introduction

In the past, natural history museums were among the first biological record centres. Extensive collections were elaborated to document the living world of both the nation they were located in and from abroad. Although the main purpose of these collections is taxonomical, they also represent valuable taxon occurrence data. Moreover, many difficult taxonomic groups are collected in the field for determination purposes, which often means that the only observation data available for these groups is held within collections.

The Luxembourg National Museum of Natural History (LMNH) was the first biological record centre in Luxembourg and in the beginning mainly observation data were entered into the first version of the LUXNAT database. However, the absence of a clear data model led to the situation that there was a proliferation of 'home made databases' within the museum. As the museum integrates many different scientific sections, each developed its own system.

In 2000, the LMNH decided to adopt Recorder 2000, the previous version of Recorder 6, for the management of its observation data. However, earth sciences and natural history collections did not fit in. In 2001, Charles Copp elaborated an extended data model integrating the different aspects of museum collections (Copp 2001). The NBN data model, as used in Recorder 2000, includes the greater majority of attributes required for the recording and management of biological field data. Recorder 2000 had, however, only limited facilities for recording details of specimens linked to records and none at all for museum specimens lacking field data. Recorder 2000 was also unable to manage earth science data related to either field records or specimens. The extended data model as used in the new Collections Module for Recorder 6 handles the following extra classes of information:

1. Accessions
2. Collections
3. Specimens (all types including minerals, rocks, fossils, herbarium sheets, mounted skins, wet collections, skeletal material, shells, mounted insects and others)
4. Documents as objects
5. Images as objects [management and description]
6. Stores and storage furniture
7. Loans (In and Out)
8. Exchanges
9. Valuations
10. Conservation Checks
11. Conservation Tasks
12. Conservation Jobs
13. Materials
14. Funding Sources
15. Enquiry Form
16. Quick data entry forms for specimens [including customised forms for subject domains]
17. Thesaurus including a Stratigraphy Dictionary (including Biostratigraphy, Lithostratigraphy and Chronostratigraphy), Mineral Dictionary, Rock Names Dictionary, Fossil Names and new museological term lists

Based on the extended data model, the Collections Module for Recorder 6 has been developed during 2005-2006 for the National Museum of Natural History in Luxembourg by Dorset software in collaboration with Charles Copp.

Recorder 6 versus Collections Module

There were several advantages to enhancing an existing system (Recorder 6) rather than write a complete collections management system as a standalone product. The total cost of the project is reduced as much of the functionality and term lists required are already in the system. It also opens the possibility to join an existing community of users and support. Moreover, it was necessary for the Museum to document and manage information of field data associated to a specimen, for example where it was gathered, who gathered it and when. The Recorder 6 application, developed by Dorset Software in conjunction with JNCC can maintain this kind of information. It was selected as the underlying product for the Collections Module for the following reasons:

- It is oriented towards complete and scientifically accurate recording of biological observations.
- It supports addins, which are code objects that can be added to the existing system without rebuilding the original code.
- It was already in use at Luxembourg Museum of Natural History.

The Recorder 6 application supports biological records in a hierarchical data fashion. Observations are organised into the following entities:

- Surveys (containing information about the reason why the data was gathered, the scope and the organisers)
- Survey Events (visits to sites by a group of people at a particular time)
- Samples (samples taken during the Survey Event, e.g. field observations or traps)
- Occurrences (currently supports taxon occurrences which represent observations of an organism and biotope occurrences which represent observations of a habitat).
- Determinations (identifications made for an occurrence. An occurrence may have more than one determination).

In addition, Recorder 6 supports other data entities which are useful for the Collections Module project: documents and references, individuals and organizations and locations.

Description of the Collections Module for Recorder 6

The Collection Module for Recorder 6 provides collectors of natural history and biodiversity objects with a system allowing them to catalogue, describe and track specimens within their collections. Although the development was primarily aimed at meeting the requirements of Musee national d'histoire naturelle (www.mnhn.lu) it is a generic product which could be used by other institutes.

The system allows users to add, edit and delete data for collections and specimens that have been accessioned into the museum (i.e. have been documented as in their ownership). It also allows managing information regarding specimens and collections at other sites or owned by other individuals and organisations. As well as collections and specimens, information regarding stores is maintained for the purposes of organisation and location of specimens. At present, the system is designed for use by museum staff. Data gathered by the system will form the underlying dataset for future web-portals and front end museum software used by visitors to the museum. The system also provides facilities for rapid data entry and reporting of this data.

The Collections Module is formed of a number of COM addins to the existing Recorder 6 Project. In addition, several changes had been made to the core product. The Collections Module adds a number of data entities to the existing Recorder data model. The key entities are those for specimens, collections and stores, collectively known as collection units. The following class diagram (Fig. 1) illustrates how these entities relate to the existing Recorder Observations hierarchy:

The Collection module entities in Fig. 1 are summarised by the following statements:

- Specimens can be associated with any number of field occurrences.
- Specimens have one or more associated determinations.

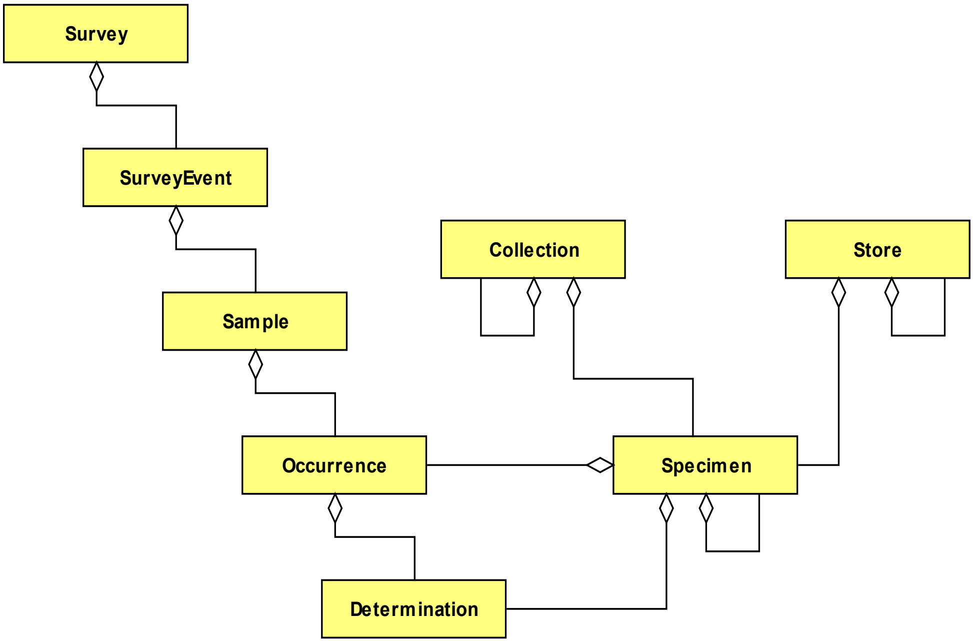


Fig. 1: Relationship of the Collections module entities to the Recorder 6 Observations hierarchy. The diamond symbols indicate aggregation, for example a collection aggregates a number of specimens. There are a number of associated entities not described in this diagram such as those concerning movements, accessions, enquiries and relationships (Dorset software, 2003).

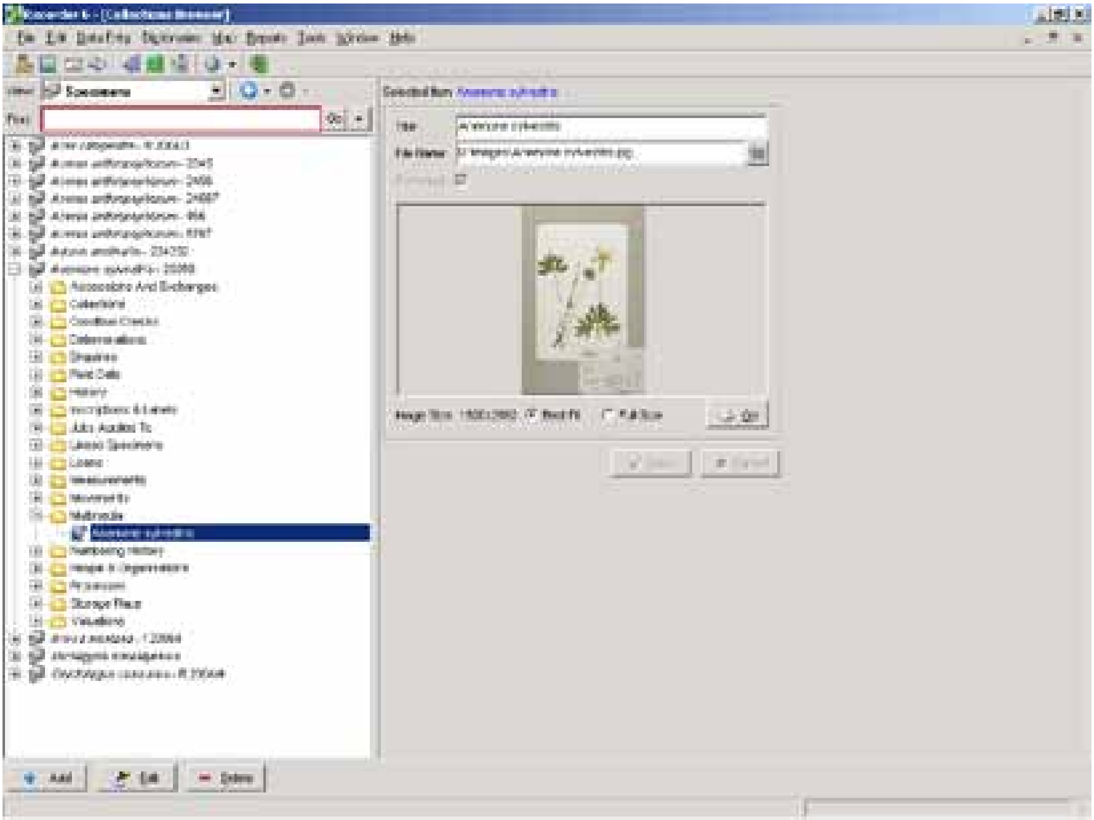


Fig. 2: Screenshot of the Collection browser component of the Collections Module.

- Specimens can aggregate other specimens (e.g. a rock can contain an ammonite).
- Stores can aggregate specimens (e.g. a tray can contain a butterfly specimen).
- Stores can aggregate other stores (e.g. a room contains cabinets, a cabinet contains shelves).
- Collections aggregate specimens.
- Collections can aggregate other collections (e.g. a shell collection is a sub-collection of the zoology collection at a museum).

The scope of the Collections Module can be summarised by the following components:

Collections Browser (an addin screen designed to allow the user to navigate through collections, specimens and the related data, Fig.2). The screen is also used for data entry of individual records.

Thesaurus Browser (an addin screen allowing the user to navigate the Thesaurus. The Thesaurus is a repository of terms and their relationships used throughout the Collections Module. This supersedes the dictionaries and term lists models used in Recorder 6 although to minimise the impact on the Recorder system the existing dictionaries and term lists are not converted into the Thesaurus. The Thesaurus provides terms in domains such as Fossil Taxi, Soil Types, Rock Types, as well as lists linked to specific controls in the application.

Quick Entry (addin screens that facilitate rapid data entry of specimens and observations data, Fig. 4).

Multiple Maps (allows the user to maintain more than one base map simultaneously in Recorder, Fig. 5).

Reports (includes a number of standard reports available for entities within the Collections Module, Fig. 6).

Specimen Finder Addin (screen that allows users to build queries to locate specimens in the system by drag and drop,Fig. 7).

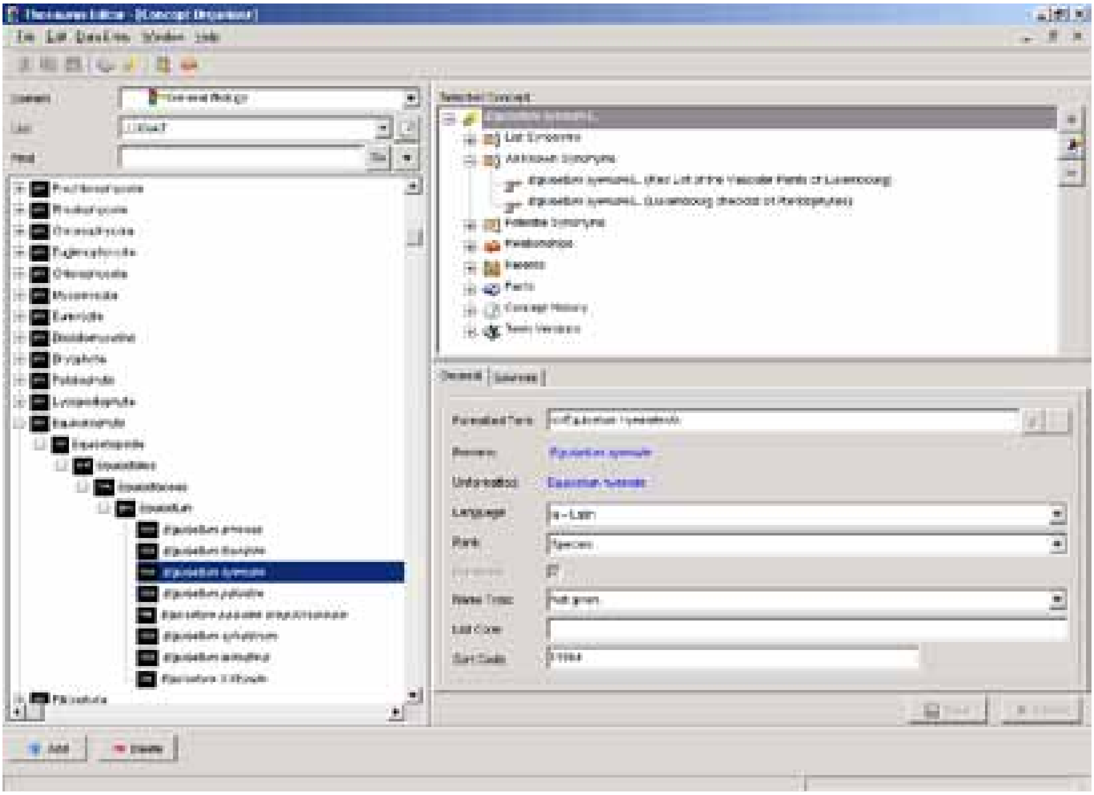


Fig. 3: Screenshot of the Thesaurus browser component of the Collections Module.

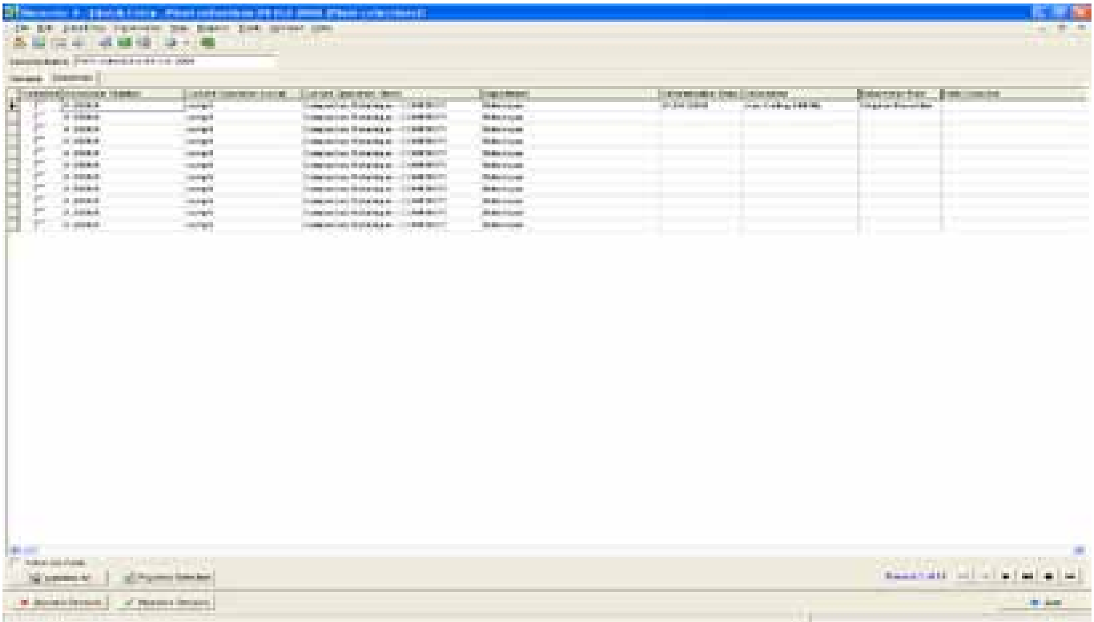


Fig. 4: Screenshot of the Quick entry component of the Collections Module.

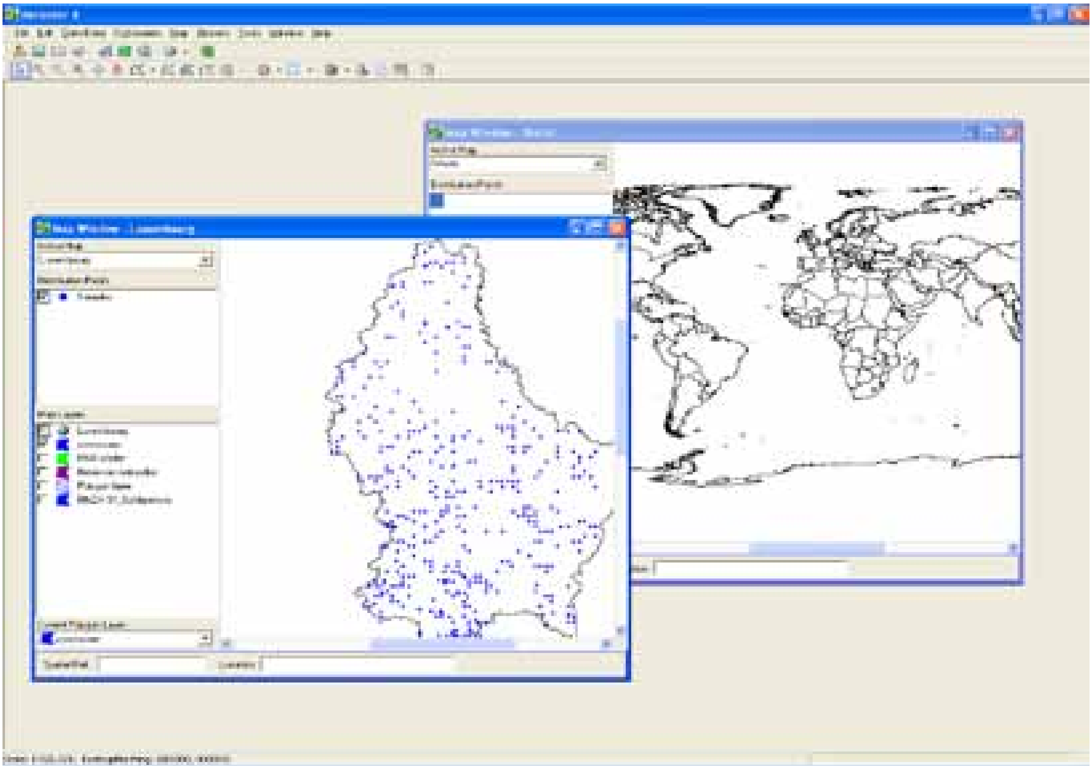


Fig. 5: Screenshot of the Multiple maps component of the Collections Module.

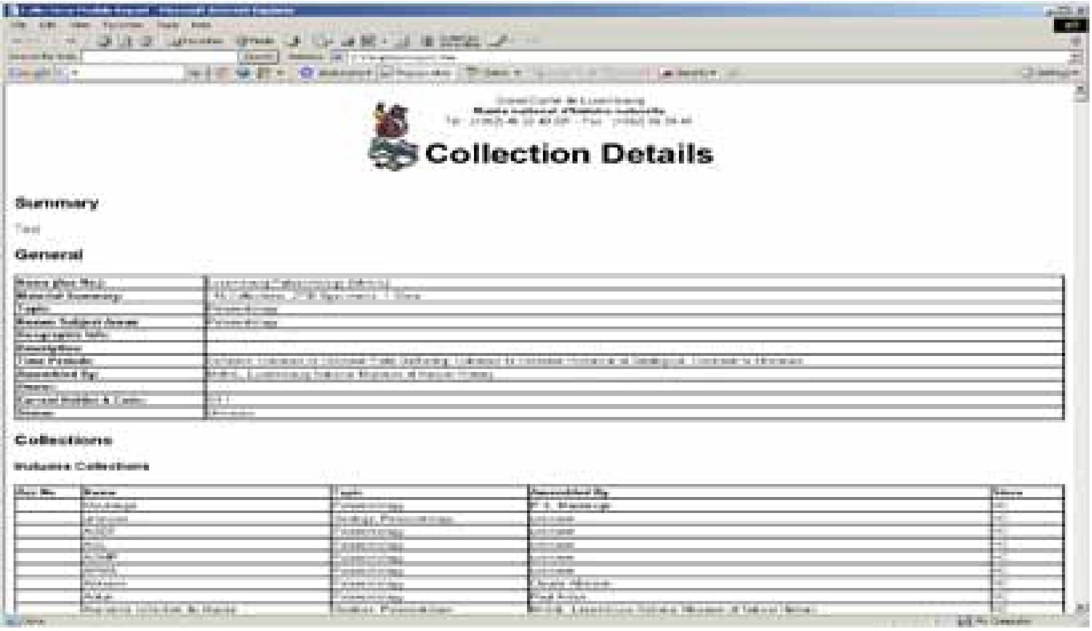


Fig. 6: Screenshot of the Reports component of the Collections Module.

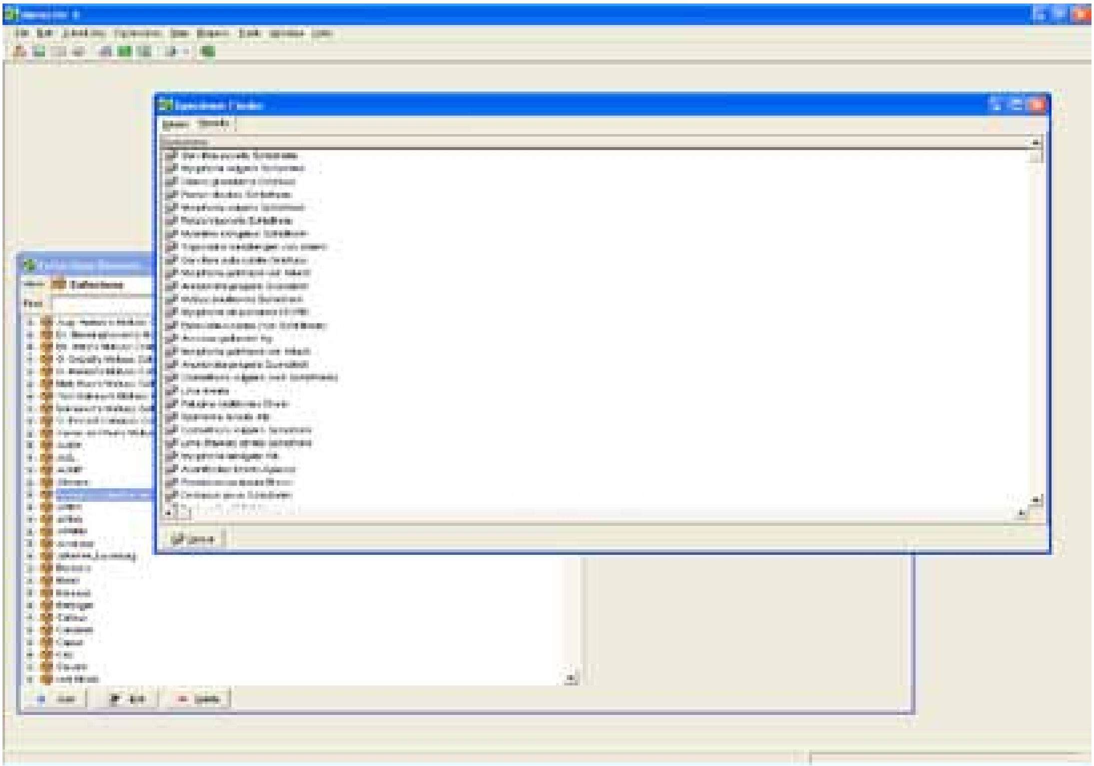


Fig. 7: Screenshot of the Specimen Finder component of the Collections Module.

References

Copp C.J.T. 2001. - Development of a Collections Management Module for Recorder 2000. Unpublished report. Musée national d'histoire naturelle, Luxembourg. Available at http://www.mnhn.lu/recherche/db/recorder/luxmus.pdf

Dorset software 2003. - Technical System Document of the Collections Module for Recorder 6. Unpublished report. Musée national d'histoire naturelle, Luxembourg.

Import and export of data

Importing and exporting data from different software systems - the NBN experience

Hannah Betts
Joint Nature Conservation Committee UK
Monkstone House
City Road
GB- PE1 1JY Peterborough

Keywords: import data, export data, GIS, NBN Gateway, Recorder

Presentation abstract

Most data managers have to move data from one software package to another on a fairly regular basis. Whilst in theory the process should be straightforward, I am also sure that many of us have seen what appears to be small and easy transformation tasks descend into a huge gnarled mess (Fig. 1). Whilst Ad Hoc transfers from historical data sources into Recorder 6 will never be problem-free; regular transfers of data can be managed to make them easier, quicker and result in a better reflection of the original source.

The issues affecting three types of data transfer were presented in the talk:

- Import into Recorder 6 from data sources such as spreadsheets (Tab. 1)
- Two-way transfer between different Biological Recording Software
- Export of Recorder 6 data into other packages (such as GIS software) for analysis and reporting purposes

Each type of transfer was investigated for:

- Specific problems as well as issues they have in common (Tab. 2)
- Useful techniques and tools and how they can be utilised in Recorder 6

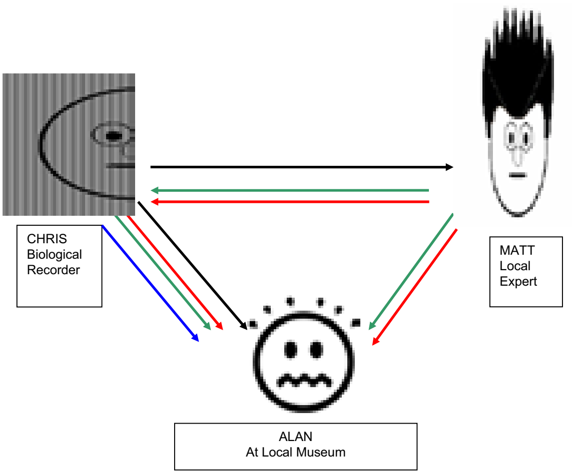


Fig. 1: The schema illustrates the possible data transfers between recorders and shows that it is essential to keep control of the different versions of the data. Within Recorder, version control is handled through a combination of methods. Thus every record has a globally unique identifier (GUId) composed of the site ID and a record key. Furthermore every record has a custodian who is the only person allowed to edit it. Finally every record has a last edit date.

Table 1: Example of a flat table structure holding simple data for import into Recorder via its import wizard. The file holds information of who observed what, where, when and how.

Observer	Species	Sex Sex Abundance	Grid Reference	Date	Record Type
Hannah Betts	*Lutra lutra*	2 males	SP01945263	01/01/2000	Field Observation
Hannah Betts	*Lutra lutra*	1 female	SP45273816	02/01/2000	Field Observation
Hannah Betts	*Lutra lutra*	1 female	SP04657384	03/01/2000	Field Observation
Hannah Betts	*Lutra lutra*	1 male	SP02738452	04/01/2000	Field Observation
Hannah Betts	*Lutra lutra*	1 male	SP04657384	05/01/2000	Field Observation
Hannah Betts	*Lutra lutra*	1 female, 2 juveniles	SP05728462	06/01/2000	Field Observation
Hannah Betts	*Lutra lutra*	1 female	SP01945263	07/01/2000	Field Observation
Hannah Betts	*Martes martes*	1 adult	SP04657384	01/01/2000	Field Observation
Hannah Betts	*Martes martes*		SP45273816	02/01/2000	Field Observation
Hannah Betts	*Martes martes*		SP02738452	03/01/2000	Field Observation
Hannah Betts	*Martes martes*	1 male, 1 femle, 1 juvenile	SP01945263	04/01/2000	Field Observation
Hannah Betts	*Martes martes*	1 adult	SP45273816	05/01/2000	Field Observation
Hannah Betts	*Martes martes*	1 adult	SP04657384	06/01/2000	Field Observation
Hannah Betts	*Martes martes*	1 adult	SP04657384	07/01/2000	Field Observation

Bad Data Sets

Table 2: Examples of 'bad' data sets for import into Recorder. Frequently encountered problems with records are the lack of appropriate metadata like for example explanations of codes or survey techniques used (b). These data sets also lack standard terminology (a), (b).

(a)

Species	Sex Sex Abundance	Grid Reference	Date
Lutra lutra		Point 1	Sunday 01/01/2000
Lutra lutra	1	Point 2	Monday
Lutra lutra	1	Point 3	Tuesday
Lutra lutra	1	Point4	Wednesday
Lutra lutra	1	Point 3	Thusday
Lutra lutra	1 female, 2 juveniles	Point 5	Friday
Lutra lutra	1 female	Point 1	Saturday
Martes martes	1 adult	Point 3	Sunday 01/01/2000
Martes martes		Point 2	Monday
Martes martes		Point 4	Tuesday
Martes martes	1X, 1V, 1Z	Point 1	Wednesday
Martes martes	1 adult	Point 2	Thusday
Martes martes	1 adult	Point 3	Friday
Martes martes	I only cought a glimpse as i was going a round the coner but i am fairly sure it was an adulte male	Point 3	Saturday

(b)

Species	01/01/2000	02/01/2000	03/01/2000	04/01/2000	05/01/2000	06/01/2000	07/01/2000
L	SP01945263	SP45273816	SP04657384	SP02738452	SP04657384	SP05728462	SP01945263
	2 males	1 female	1 female	1 male	1 male	1 female, 2 juveniles	1 female
M	SP04657384	SP45273816	SP02738452	SP01945263	SP45273816	SP04657384	SP04657384
	1 adult			1 male,1 female, 1 juvenile	1 adult	1 adult	1 adult

An introduction to the Recorder Thesaurus

Charles Copp
Environmental Information Management
8 The Paddock
GB-BS21 6JU Clevedon
eim@globalnet.co.uk

Keywords: thesaurus editor, thesaurus browser, linking term lists, classifications, cross-disciplinary ontologies

Abstract

A thesaurus is a key enabling component for effective data retrieval from large data networks such as those being created under collaborative projects such as GBIF, BioCASE and the UK NBN.

The thesaurus, which has been built into the Luxembourg Collections Management software (Recorder 6 extension) was originally designed for the European-funded BioCASE project (Nov. 2001- Jan. 2005). The thesaurus is designed to manage multiple related term lists covering all aspects of natural science collections and field observations including taxonomy, habitats, gazetteers, collecting methodologies and stratigraphy.

The thesaurus is being populated with classifications and term lists derived from existing sources and also terms entered by users of the database (e.g. imported from existing registers). The thesaurus can place these terms in a meaningful context with other equivalent terms and both broader and narrower categories useful for maximising or refining the number of returns to queries. Moreover the thesaurus enables a multilingual approach that is very important in a European context.

The present version of the Luxembourg thesaurus runs on an MS SQLServer Database as part of the Recorder 6 Application. There is also an associated Thesaurus Editor program for managing lists. The Editor allows the management of both complex dictionaries and simple term lists in a very flexible manner. New lists and terms can be added and modified through the editor interface or imported from spreadsheets.

The structure of the thesaurus model is such that it can be applied to any discipline and it has great scope for managing cross-disciplinary ontologies that will be of great value in building better user interfaces in the future, especially for the web. This will create a significant resource that will need a long-term management strategy and provision made for its continued maintenance and custodianship. Our current thinking is towards ways of enabling collaborative creation of thesaurus term lists and the linking of lists through a web-based application.

Introduction

Recorder is a database application that began, in the United Kingdom, as a system for storing simple field observations about plants and animals. Over a period of 16 years it has been extended to include specimens and collection management data and broadened in scope to cover earth sciences and more types of biological information.

The need to provide controlled lists of taxa via a taxon dictionary has been part of Recorder since the first versions, as has the functionality to build lists of accepted site names. Over the years a number of other term lists, including a biotopes dictionary, were developed to provide a means of controlling the entries into various fields. With the development of Recorder 2000 and subsequently, Recorder 6, with the addition of the collection module and internationalisation, it has been necessary to extend the number of term lists managed by the application and also to adapt to the needs of multi-lingual users and developments such as 'concept'-based taxonomy. The Recorder Thesaurus was created as a means for handling these many types of term lists, including ontologies, from any discipline, to serve as the core technology for controlling and interpreting the information managed within the developing Recorder system.

Why do we need controlled terminology?

Data entry

We need to control the terms that we use in databases for a number of reasons, these include:

- Standardised data entry helps avoid spelling and transcription errors
- Standardised terms express common meaning and enhance the clarity of the data recorded
- Standardised terms make it easier to reliably retrieve data

It is commonly thought, that controlled terminology can be solved simply by providing the right lists of terms, be this a list of recording methodologies or a national taxon list. Unfortunately, this simplistic approach can only work in limited situations. As the database scope increases not only in its topic coverage but in the number of organisations using it or where there is a desire to exchange or supply data then one is faced with issues of language, synonymy, old terms that must be maintained (e.g. from specimen labels) and the need to add new terms.

Some of the most difficult problems arise with the naming of taxa. The use of a single approved list of taxon names works well for making new records such as field observations of species but becomes a problem when dealing with historical records and museum specimens. Specimens identified at different times from different works may bear taxon names that have subsequently been revised or changed in scope (e.g. through lumping or splitting) and this information needs to be retained whilst also allowing for search and retrieval using currently accepted names. In Recorder this development was marked in the release of Recorder 2000 by a move from a single taxon lookup list to a multi-list taxon dictionary. The Taxon dictionary model developed for Recorder 2000 was adopted and extended by the Natural History Museum, London for the management of taxon lists that are, among other things, being used to create the UK species list.

With the development of the Thesaurus in Recorder 6, the various dictionaries were incorporated into a new physical model that was specifically designed to be able to accurately map the relationships between terms and subtleties of meaning as are found in concept-based taxonomy. In addition to single domain relationships, the thesaurus was further developed to manage cross domain relationships including temporal precedence, for instance it can map different types of stratigraphic terms (biostratigraphical, lithostratigraphical, chronostratigraphical) to one another including temporal overlaps (e.g. where a named rock type partially falls within two time zones).

The aim is that we use controlled terminology not in a restrictive sense but as a means of allowing the reliable capture of both the original data associated with a specimen and its actual meaning. This is not a simple thing to accomplish, as it requires both extensive and reliably cross-linked lists in the thesaurus and good data with the specimens!

Data retrieval and interpretation

The purpose of a database is not just to store information, it must also be possible to retrieve the information. How do you know what to ask for and if you do, how do you know that you have found all the relevant records? It all depends on the use of words and the quality of the indexing. This is a common problem with all databases, once data are entered you cannot easily see what is in there. The data can be simply scanned for a single table with a handful of rows but with large relational structures holding thousands or millions of records it is impossible to directly know what is in there. We are removed from the information and have to use various tools to 'fish' for it. If we know what was put in we then have some idea of what to ask for and some measure of the success of retrieval but otherwise we are working in the dark. Simple word indexes are not enough because there may be many alternative terms used, including those with a broader or narrower context and there will be inevitable typographic errors. There is also the problem that the terms and syntax used by the cataloguer or indexer may not match those of the enquirer, and our system must meet the needs of both.

The understanding of words and their semantic relationships is not trivial but is essential if we are to address the needs of the broadest spectrum of users. Whilst the database remains the province of the specialist curator, there is at least the strong likelihood that they will know what terms to use to retrieve their data but this changes if the database is also the source of information for wider dissemination such as through a public website. Our retrieval system must equally be able to meet the needs of specialists who know precisely what they want (e.g. the location of cultures of a specific strain of bacterium) and the child who wants to find out about dinosaurs. For this reason we need a thesaurus to manage the relationships of technical classifications to each other and to common language.

In data retrieval the thesaurus can play an essential role by:

- Providing links to synonyms and overlapping terms
- Allowing Improved indexing (e.g. many terms are indexed under one preferred term)
- Helping users expand or narrow queries
- Providing links to related terms
- Providing links to other information systems (e.g. through web services)
- Providing the potential to build knowledge-bases by adding supplementary information to explain terms and their relationships (ontologies)
- Improving the potential to exchange information through term and concept mapping
- Documenting relationships between terms

A thesaurus therefore, should be the key enabling component for effective data retrieval from large data networks such as those being created under collaborative projects such as GBIF, BioCASE and the UK NBN. As Recorder use spreads in Europe, especially amongst organisations using it for collections management, the need to control and share common term lists will become a growing problem and more work will need to be undertaken to develop these multi-lingual multi-level term lists.

The Structure of the Recorder Thesaurus

Overview

The Thesaurus model is exceptionally powerful. It can be used to store everything from simple non-hierarchical term lists to complex multi-facetted information nets. It can deal with taxonomic concepts and many subtleties of term relationships including temporal and spatial relationships. The thesaurus, which has been built into the Recorder 6 Collections Management module was developed (by John Van Breda and Charles Copp) from the thesaurus model originally designed for the European-funded BioCASE project (Copp 2002, 2003a, 2003b).

The Thesaurus Model is primarily concerned with the relationships of terms (e.g. taxon names, place names, habitat names) with the concepts that use them. The same term may be used many times in many different contexts and with differences of **meaning**, Lists of terms such as taxonomic revisions, habitat classifications and gazetteers are, therefore, regarded as collections of concepts that use terms in specific ways; for this reason, in the thesaurus model, lists are referred to as **Concept_Groups**.

Concept_Groups (Lists) may be simple lists, hierarchical lists or complex polyhierarchical classifications. The terms, stored in the **Term** table can be referred to concept_groups associated with many different disciplines (e.g. taxonomy, biotopes, stratigraphy, gazetteers etc.) and so the thesaurus includes entities for **Subject_Areas** sub-divided into **Domains**.

The physical table model used in the Recorder Thesaurus is shown in Figure 1. The most important tables are highlighted in red. A more complete description of the Thesaurus Model and the attributes used in each table can be found in the description of the NBN data model (Copp 2006).

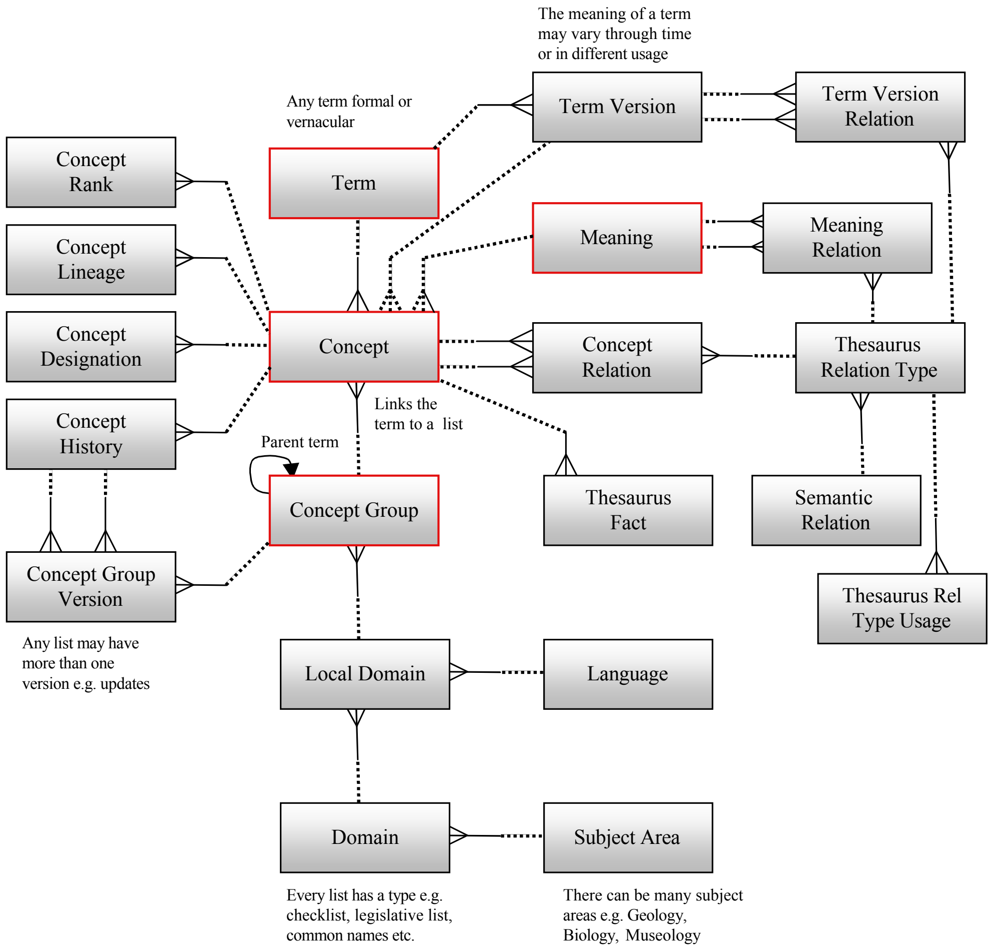


Fig. 1: The physical model for the Thesaurus database in Recorder 6 Collections Add-in and Thesaurus used in the BioCASE Project. The tables outlined in red are the ones most used in queries.

Subject Areas and Domains

The Recorder Thesaurus is divided into **domains** of terms, such as botanical taxonomy, chronostratigraphy, European biotopes, UK gazetteers etc. The **Subject_Area** groups domains into higher level categories (e.g. Geology, Taxonomy, Gazetteers etc.) solely to assist in organisation of the domains for management or when they are presented to the user (Fig. 2).

Each domain may contain **Local domains,** which define the geographic and lingual scope of the data present for that domain in the current system. For example, within the Biotopes domain there might be a UK (English) local domain and a Luxembourg (English) local domain, indicating that lists are available for biotopes in the UK and Luxembourg, all in English. Local domains also define the hyperlinks used to build search URLs for submission to web portals relevant to that domain (e.g. for linking to contextual information such as distribution maps).

Domains also define the classification ranks that are available for attachment to concepts within the domain which supply information regarding the ordering of the concept within the group's hierarchy, for example taxonomic ranks in the taxonomy domain and stratigraphic ranks for the stratigraphy domain.

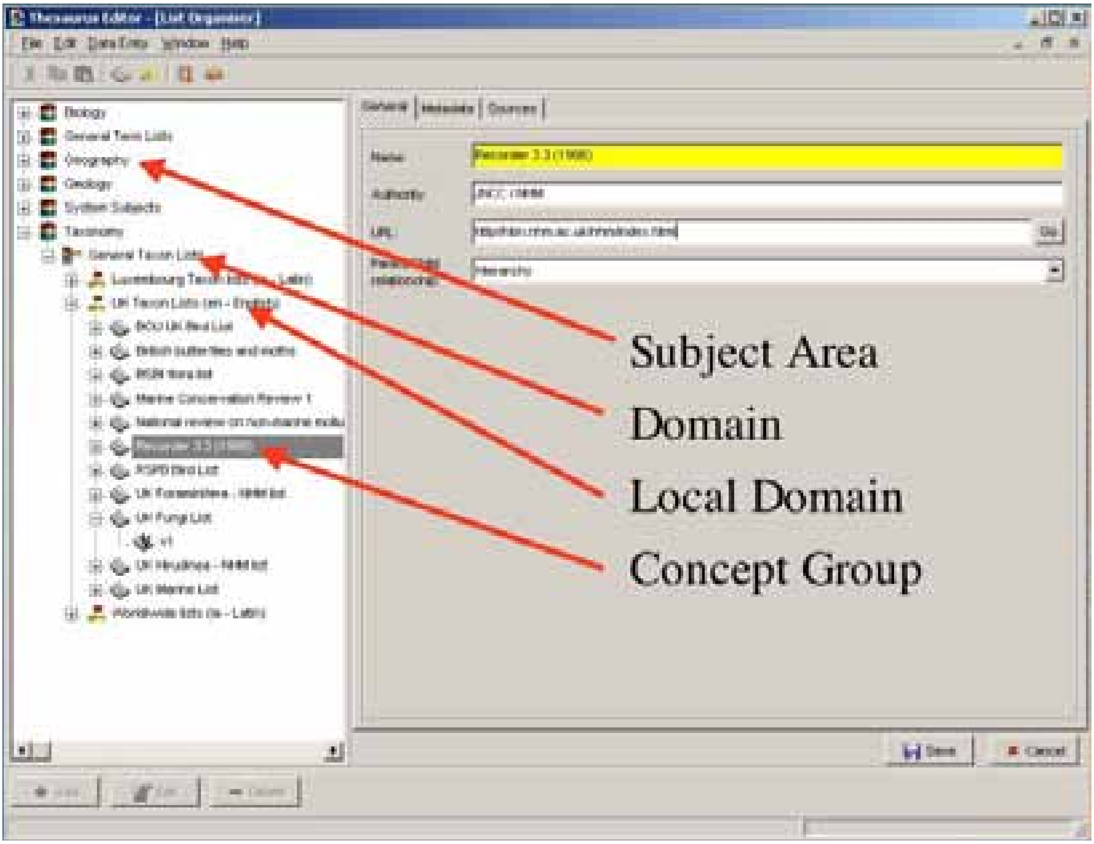


Fig. 2: The relationship between Subject Areas, Domains, Local Domains and Concept Groups as seen in the Recorder Thesaurus Editor (also known as ***Thesaurus Rex***)

Terms and Term Versions

Individual terms along with their language and whether they are formal (controlled) or informal (common language) terms are stored in the term table. Taxon names are, by convention, recorded as "latin" and "formal" terms. A term can be a single word or several such as a whole taxon name. The same term may be referenced by many different concept_groups (lists) and may even be used in different subject domains (e.g. plant names, animal names and fossil names may be homonyms).

Many terms have only one meaning but others such as taxa and geographic place names may be used in several contexts, for instance, through the process of taxonomic revision (e.g. lumping and splitting) or redefinition of boundaries (e.g. Germany has had several different spatial extents). It is thus necessary to have a **Term_Version** entity so that the right 'meaning' of a term can be linked to its use in a specific concept group (term list). Taxon names are stored in the Term table without their author name which is stored as an attribute of a term version and linked to a specific Meaning Key.

Concept Groups (Term Lists)

Domains contain **concept groups** (Fig. 2); each concept group contains a number of **concepts** as a single list or a hierarchy. A term list, such as a taxon checklist or gazetteer of place names, is therefore, referred to as a Concept_Group. Concept groups can have **concept group versions** that reflect published amendments made to the list or hierarchy. Concepts may be linked to concept group versions, as terms may be added, deleted and added again during various revisions of term lists.

Concepts and Meanings

The **Concept** table is the primary place through which other systems are linked to the Thesaurus. A concept defines the term used, the meaning of the term used, and the **concept group** that the concept belongs to. Each **concept** refers directly to a single **term** in the term table Where a term may have more than one variation (e.g. different author strings) the concept may also refer to a key in the Term_Version Table.

Every concept that has exactly the same meaning is related to the same record in the **Meaning** table. The Meaning Key allows a concept group to include many terms for the same point in a list (e.g. current preferred name, English name, French name, synonyms etc.). Only one term can be the preferred term for each language. The meaning key also works between concept groups, therefore it is possible to list all synonyms for a concept (regardless of its concept group) by listing those terms that share the same meaning key; subsequent translations either into other lists or by language are then possible by selecting for terms with the same meaning.

Meanings may be related to other meanings and the relationships have semantic information understood by the system. For instance this might be used to indicate that one meaning of a term was derived by splitting from an earlier meaning of the term. This allows groups of synonyms to be related to each other with a single record and also related to other groups. When the user views a concept in the Thesaurus Browser, the relationships listed are those related directly to the concept plus those related to the group of synonyms that share the same meaning and also potential synonyms derived from other uses of the same term.

Hierarchical relationships of terms and term synonymies are functions of individual versions of Concept Groups. For instance, it is quite common for 'competing' taxonomic checklists to have quite different higher level relationships for taxa. The position of terms within lists can be handled in a number of ways. If the concept group is hierarchical (as in taxonomy) each concept (term) can store a pointer to its 'parent term' in the hierarchy. Concepts may also be assigned a rank in a list. The hierarchical relationships pertinent to a domain are listed in a separate **Concept Rank** table that, among other things, can store the order and sort positions of hierarchical terms (e.g. phyla, classes orders, Eons, Periods, Stages).

Concepts may also have a specific sort code which enables non-alphabetic sorting of lists if available. Hierarchical relationships within a list can therefore be handled by a combination of declaring the term's rank, declaring the immediate parent term or using a sort code for the terms.

The **Concept_Lineage** table defines the ancestry of the concept when viewed within the concept group hierarchy (it lists the parent and grand-parent concepts). The Concept_Lineage table may contain more than one record for a concept if it, or any of its ancestors, have more than one parent in the hierarchy. This allows any sort of hierarchy or network, including polyhierarchical ones, to be quickly navigated or displayed. When navigating term lists in the theasurus browser, it is possible to ask to see the parent hierarchy for a term and where multiple possible parent hierarchies are detected, these are displayed for the user to choose from. The creation of the concept lineage record is done automatically by the thesaurus software and is internal to the application.

Each concept can also be related to any number of other concepts in any domain, allowing complex ontological information to be defined such as predator-prey and biotope associations. The relationship types that can be used (and added by users) include semantic information about the relationship (e.g. boolean relationships) that can be provided to the system for processing. Concepts may also have a rank and a type, and concepts can have any number of designations (e.g. protected status such as SSSI or inclusion on CITES lists) and facts attached to them.

The **Concept_History** table lists the concept group versions that any individual concept applies to within the concept group such as when it first appeared in a list and when it was removed (e.g. in legislative schedules of protected species).

Facts about terms, such as a taxon's association with given biotopes or other taxa, are stored in the **Thesaurus Fact** table and can be linked to the concept, term version or meaning. If the fact is true regardless of the list a term might appear in (e.g. The Death Cap (Amanita phalloides) is always poisonous) then it is linked to the Meaning. When there is information about a term which is specific to an individual list (e.g. a taxon description from a specific checklist) this is linked to the Concept. Facts can also be cascaded down to concepts further down the hierarchy, therefore if mammals are warm-blooded then the fact can also apply to cats and dogs without repetition.

The Thesaurus as an information system

The Recorder thesaurus was created to manage multiple related term lists covering all aspects of natural science collections and field observations including taxonomy, habitats, gazetteers, collecting methodologies and stratigraphy. The potential to use the thesaurus as a semantic management tool for terms and associated facts from all disciplines was first mooted in the European BioCASE Project (BioCASE n.d.), but was not realised within the project, however, on a more modest scale, aspects of it can be achieved within Recorder implementations.

The job of the thesaurus is to provide terms for controlled data entry and to place those terms in a meaningful context with other equivalent terms and both broader and narrower categories useful for maximising or refining the number of returns to queries. The structure of the thesaurus also allows the addition of contextual information and links to other information sources which can be used to guide the choice of terms or to aid in the interpretation of data extracted from the database (Fig. 3). For example the thesaurus might include links from taxon lists to information sources such as the UK NBN or GBIF, such that a list of species records from the database might be viewed in relation to distribution maps obtained from linked web services and also locally stored context (and local language) information.

A more recent development that has come about through the need to improve data retrieval from

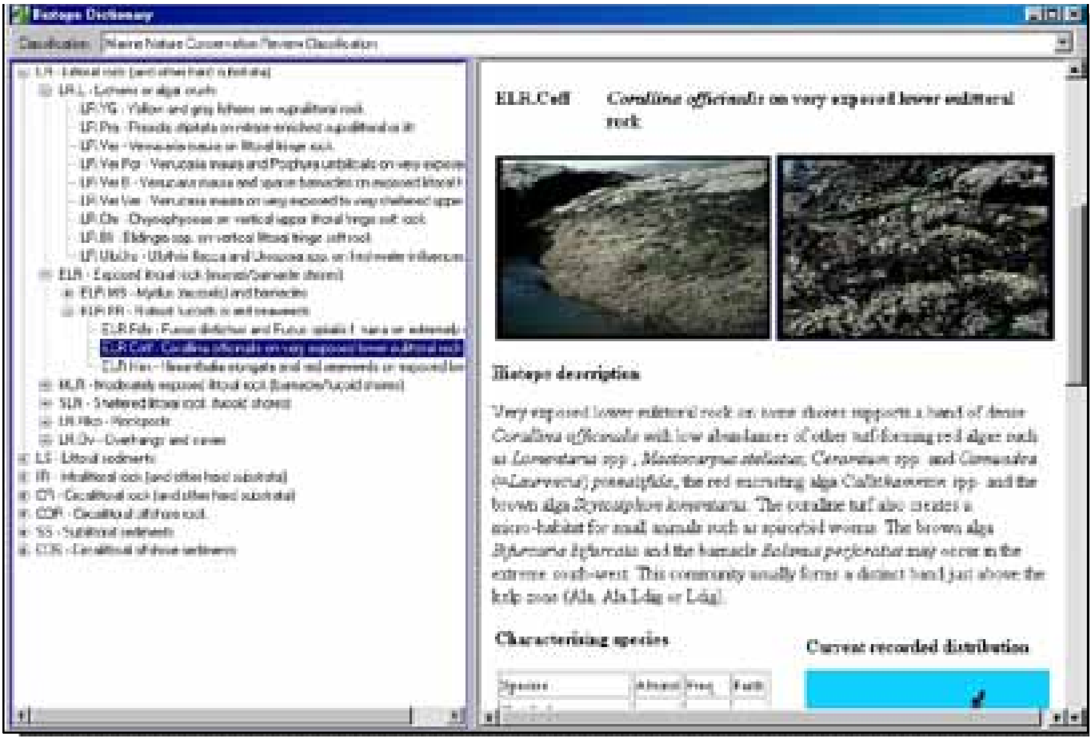


Fig. 3: The terms in a dictionary or thesaurus list can be enhanced by the inclusion of explanatory text and images. These may be stored locally or derived through weblinks to other information systems. The illustration shows an explanation of a habitat type found in one of the Biotope classifications held in Recorder 2000.

complex systems such as the world wide web is the creation of ontologies. Ontologies capture knowledge about concepts and describe their relationships and are often used to capture 'rich metadata' about sources of information such as research papers or images to aid retrieval. An ontology is essentially a hierarchical structure of meanings and properties related to ideas, which may be anything classifiable such as, species names, types of artifact or even database schema concepts. Ontologies may even be cross discipline, linking ideas from many fields of study. The thesaurus was designed to be able to manage ontological information and this could become an important role for it in the transition from using Recorder for data capture to using it as a source for information dissemination systems.

Managing the Thesaurus

The present version of the Recorder thesaurus runs on an MS SQLServer Database as part of the Recorder 6 Application. The database needed to manage the term lists and map their relationships is complex and it is not easy to work directly with the data tables without the aid of software such as the Recorder Thesaurus Editor (Fig. 4). The Editor allows managing the dictionaries of Recorder 6 in very flexible manner and new lists can be imported from spreadsheets or CSV files.

A major part of the work involved with any implementation of Recorder in a new country or dealing with new types of collection, is the establishment of the term lists in the thesaurus. This includes everything from locally used taxon lists and place

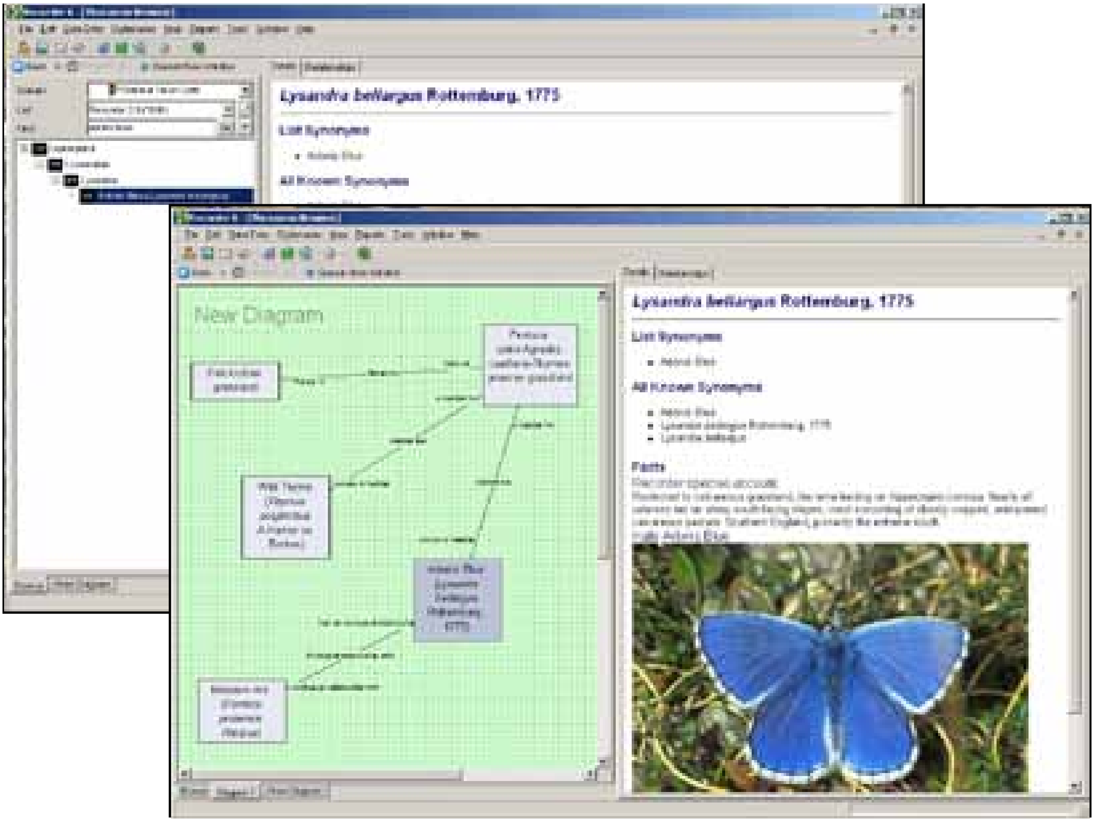


Fig. 4: A Thesaurus as an information system - The Recorder Thesaurus allows linking of facts and images to terms and also the recording of relationships with items in other domains and classifications. The illustration shows an example of a butterfly, displayed in the Recorder Thesaurus Browser, that not only includes synonyms but relationships with habitats and through the habitat, links to other species.

name lists to lookup lists for recording technique, preservation and much more. The work might include translation of existing lists (by adding local language terms as new term synonyms and making them the preferred term) or importing whole new lists. The problem can be acute when attempting to import large sets of data derived from old labels where taxa and place names may not match those in modern lists.

Once the main development is complete and the thesaurus is in use, it can be updated and new lists added as the need arises. However, all added terms will need checking for validity and for linking to existing terms (e.g. as synonyms, broader terms and narrower terms). It can be very time consuming to manually check or relate every item so strategies and software are needed to aid this process. Certainly, one of the big problems encountered in the work so far, is that of importing new taxon lists and trying to spot all of the possible synonyms and links to existing lists. Work is currently under way to develop software tools to make this task easier.

Dynamic lists obtained from outside sources (e.g. Species 2000, 2007) need monitoring for updates and new lists will also be identified from time to time for inclusion. Any links to other on-line thesauri such as gazetteers will need monitoring to check for changed links.

Creating a thesaurus on the scale of the Recorder Thesaurus also creates a responsibility and a resource requirement for its longer-term upkeep and development. Upkeep implies maintaining the relationships developed with both list suppliers and ongoing access for thesaurus users. The thesaurus can become an enabling technology for a whole network of Recorder users across Europe but thought needs to be given to how we can jointly develop all of the term lists that we need and how to tackle the responsibility for their upkeep and maintenance.

Ongoing use of the thesaurus will develop and refine its value for relating terms and informing database queries. The structure of the thesaurus model is such that it can be applied to any discipline and it has great scope for managing cross-disciplinary ontologies, that will be of great value in building better user interfaces in the future, especially for the web. This will create a significant resource that will need a long-term management strategy and provision made for its continued maintenance and custodianship. The resources and effort required to maintain and develop the thesaurus can be justified by making provision for its long-term availability to partners and by developing collaborative ways for making its content more accurate and more widely available. One possibility is to develop a web-based version of the thesaurus that enables distributed editing and access and has web service links to other term list suppliers.

References

BioCASE n.d. - The Biological Collections Access Service for Europe website. Available at www.biocase.org. Accessed March 2007.

Copp C.J.T. 2002. - The BioCASE Thesaurus: Logical and Physical Models Deliverable 9. Workpackage 4. The BioCASE Project.

Copp C.J.T. 2003a. - The BioCASE Thesaurus: Proposed Final Logical and Physical Models Deliverable 25. Workpackage 4. The BioCASE Project.

Copp, C.J.T. 2003b. - Creating and managing a thesaurus for accessing natural science collection and observation data: 149-164 in SCOBLE, M.J. (ed.), ENHSIN The European Natural History Specimen Information Network. The Natural History Museum, London.

Copp, C.J.T. 2004. - The NBN Data Model. Part 1: Description of the Model. June 2004 unpublished report. The NBN Trust, Newark. Available on NBN web site at www.nbn.org.uk

Copp C.J.T. 2006. - The NBN Data Model. Updated version of 2004 JNCC report available on EIMwiki at http://eim.metapath.org/wiki/index.php?title=Data_Models. Accessed March 2007.

Species 2000 2007. - Species 2000 website and 'Catalogue of Life' available at http://www.sp2000.org/. Accessed March 2007.

Compiling and managing the Taxon Dictionary for the Recorder software package

John C. Tweddle, Charles Hussey
The Natural History Museum
Cromwell Road
GB- SW7 5BD London
j.tweddle@nhm.ac.uk
c.hussey@nhm.ac.uk

Keywords: taxonomic databases, biological recording, Recorder, National Biodiversity Network, Species Dictionary, Taxon Dictionary, nomenclators

Abstract

Taxonomic names play a key role in biological studies and access to a wide range of biodiversity information relies upon their careful assignment and interpretation. In the Recorder biological records software, naming information is held within the Taxon Dictionary module. All observational and specimen data are associated with the name strings of the organisms concerned, and this information needs to be accurately stored and maintained if the data are to be interpreted correctly. In this paper we introduce the structure of the Taxon Dictionary and discuss possible methods by which it can be compiled, maintained and distributed. Building upon our work for the National Biodiversity Network Species Dictionary Project we then present a worked case study detailing how nomenclature is being managed for Recorder within the UK, including challenges encountered and solutions adopted. Although specifically intended to help those looking to compile or manage naming within Recorder, many of the observations made will have wider relevance to other taxonomic databasing projects.

Introduction

Taxonomic names play a key role in biological studies. They identify which organism a piece of information relates to (uBio n.d.) and allow us to link between resources that hold data about the same taxon (Page 2005). Access to a wide range of biodiversity information, from primary observational data to museum specimens and legislative listings, therefore relies upon the careful assignment and interpretation of taxonomic name strings (e.g. Chapman 2005a).

The application of naming is not always straight forward, however. Taxonomy is a dynamic field and advances in scientific knowledge can result in changes to the nomenclature. New insights may cause a species to be reassigned to a different genus, whilst the circumscription of what constitutes a particular species can change over time (Hussey et al. 2006). For example, one or more species may be subsumed into another ('species lumping'), or what was originally considered to be a single species may be separated into two or more ('species splitting'). Different experts can also have varying interpretations of what defines a particular species (see uBio n.d.). Misspellings (Chapman 2005b), homonyms and regional differences in vernacular names present other potential sources of confusion.

This means that an organism can be known by a variety of different names: a currently accepted scientific name, obsolete scientific names (synonyms), one or more vernacular names and misspellings of any of these. The exact concept of what is meant by a particular name may also differ between datasets, or over time. In the context of Recorder, it is important that observational and specimen data are explicitly tied to the name strings of the organisms as supplied, and that this information is carefully stored and maintained. In order to facilitate the collation and exchange of records from different sources, it is also necessary to ensure that the usages of names are correctly mapped.

The Recorder biological records software allows naming information to be stored in a precise, standardised manner and is underpinned by the modular and extensible UK National Biodiversity Network data model (Copp 1998). This data model comprises a series of modules and dictionaries with organism naming held within the Taxon Dictionary (Copp 2000). It is therefore imperative that the Taxon Dictionary in Recorder contains accurate representations of the names that are required by the user.

In this paper we discuss the Taxon Dictionary within Recorder 2002 and Recorder 6. We introduce the structure of the Taxon Dictionary and consider methods by which it can be compiled, maintained and distributed to users. Building upon our work for the UK's National Biodiversity Network we then present a worked case study of how nomenclature is being managed for Recorder within the UK, including challenges met and solutions adopted.

The structure of the Taxon Dictionary in Recorder 2002 and Recorder 6

A list of species names is vital for activities concerning biological recording or biodiversity (Copp 2000). In Recorder, this function is provided by the Taxon Dictionary. It holds the names themselves as well as information on the various checklists in which they occur, associated designatory data and taxon fact pages. The Taxon Dictionary incorporates rigorous NBN data standards and allows naming to be stored in a standardised fashion (Copp 2000). It comprises a series of related tables and, although there are some minor differences, it is employed in a broadly similar way in both Recorder 6 and Recorder 2002. The following section provides an introduction to the Taxon Dictionary, with Figure 1 presenting an overview of the principal data holding tables. Readers looking for more detailed table descriptions are referred to Copp (2000).

Taxon names are contained in the TAXON and TAXON_VERSION tables and referenced by a unique identifier: the TAXON_VERSION_KEY. Rather than being held as a single string, names are stored as three separate elements: ITEM_NAME and AUTHORITY in the TAXON table, and ATTRIBUTE in the TAXON_VERSION table. ITEM_NAME holds the organism name itself (e.g. *Lutra lutra*) and AUTHORITY contains the full naming authority, including year where applicable (e.g. Linnaeus, 1758). Every different combination of taxon name and authority requires an entry in TAXON and is assigned a unique TAXON_KEY. The TAXON table holds both scientific and common names and contains fields that identify the type (formal scientific or vernacular) and language of each name entry.

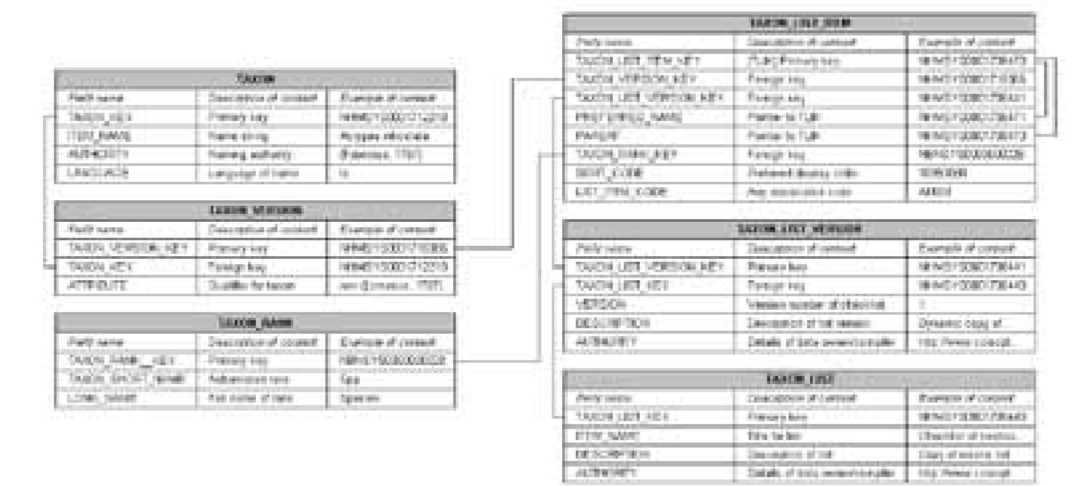


Fig. 1: Simplified relationship diagram for the Taxon Dictionary in Recorder (some fields and tables are omitted).

As previously stated, a given name and authority combination can be used in different ways by different authors and at different times. To identify these varying uses, each meaning of a name is assigned its own unique record and TAXON_VERSION_KEY in the TAXON_VERSION table. In most cases a name will only have one interpretation within Recorder, so will have a single TAXON and corresponding TAXON_VERSION record. Where a name has several possible interpretations, the same TAXON entry will link to multiple TAXON_VERSION records. The ATTRIBUTE field stores any qualifying information that can be used to identify and distinguish these differing uses of a name (e.g. *sensu stricto, sensu lato, sensu* Smith, *nec* Smith, *auct. brit*.).

One of the most important aspects of Recorder's Taxon Dictionary is that every name is stored in association with one or more checklists. A checklist can be any defined grouping of taxa, such as a taxonomic list or a legislative schedule of protected species (Copp 2000). The structure allows many different checklists and versions of checklists to be held, along with a mechanism for translating between them (but see Challenges encountered and solutions found, page 72). Although it would be possible to store all of the names required in a single list, there are benefits to maintaining separate lists that are suitable for different purposes. Relating names to the lists in which they occur allows individual lists to be recreated and compared. This makes it different from virtually all other taxonomic database projects and provides the flexibility to allow competing classifications to be presented and data to be stored against different lists for different purposes (Copp 2000). Information concerning the various checklists in which each taxon name occurs is stored in the TAXON_LIST_ITEM, TAXON_LIST and TAXON_LIST_VERSION tables.

The data needed to recreate an individual list are stored in the TAXON_LIST_ITEM table. Every record in every list that has been added to the Dictionary has an entry in this table, linked to the relevant taxon name via the TAXON_VERSION_KEY. If wished, synonyms and a hierarchy can be included for each list using the PREFERRED_NAME and PARENT fields. A sort code can also be added so that the names in the list can be presented in a preferred order. The TAXON_LIST_VERSION_KEY of this table identifies which version of which list the data originate from, with full metadata for the list being stored in the TAXON_LIST and TAXON_LIST_VERSION tables. The use of the TAXON_LIST_VERSION table means that it is possible to include different versions of the same checklist. These could be versions tailored to suit different uses, or different editions of a dataset, such as an annually compiled taxonomic list. It is possible to add such multiple editions as entirely new versions, or as amendments to an existing version.

Information concerning the designatory status of a taxon, for example its national or international threat status, is contained within the TAXON_DESIGNATION table. This links directly to the TAXON_LIST_ITEM table (rather than to the TAXON or TAXON_VERSION table), thus allowing the legal or other conservation status of a species to be associated with specific lists. The TAXON_FACT table can be used to store any additional information about an organism, such as a factual description, image or sound file. Recorder 6 also contains a TAXON_GROUP table that assigns every TAXON_VERSION_KEY to an informal group (e.g. 'mammal'; see Challenges encountered and solutions found, page 72).

The primary keys for each table follow National Biodiversity Network naming conventions. Each code is a 16-character string comprising a six-letter prefix and a running series of ten digits (e.g. NHMSYS1234567890). The prefix identifies the organisation responsible for creating the record and thus provides a way of tracking and identifying ownership. Although not shown in Figure 1, each table also includes four fields that enable additions and changes to table content to be tracked (ENTERED_BY; ENTRY_DATE; CHANGED_BY; CHANGE_DATE). These can be used to aid database management and identify which records to include within an update.

It is worth noting that although Recorder 6 and Recorder 2002 contain broadly equivalent Taxon Dictionary structures there are some minor field differences, full details of which can be provided on request. Note also that the Thesaurus extension of the Collections Module allows naming to be stored in a broadly analogous fashion, but as a component of the more generic Dictionary Module. Readers interested in learning more are referred to Copp (2004a, b) and other papers in this volume.

Compiling and maintaining the Taxon Dictionary

Possible approaches for building a Taxon Dictionary

As currently sold, copies of Recorder for UK use are supplied with a ready-built Taxon Dictionary containing a wide range of taxonomic, legislative and designatory checklists. If Recorder is to be used outside of the UK, however, it will be necessary to create and add lists that are relevant to the area concerned, along with appropriate base-maps. In this situation Recorder can be supplied with selected UK checklists as examples of how the data should be formatted, but depending upon their wider suitability it may be advisable to begin afresh with an empty database, once familiarisation is complete.

Creating a new Taxon Dictionary can be a labour intensive task, but as naming provides access to all record and specimen information that is subsequently stored, it is important that time is taken to compile it accurately. If starting from scratch, one should attempt to ensure quality by entering only completely formed authoritative names, together with their authorities.

There are three principal methods through which the information for the Taxon Dictionary can be sourced, compiled and maintained. They each have both benefits and disadvantages and the approach adopted should be the one that is most appropriate for the project concerned. Factors to consider include user requirements, whether appropriate datasets already exist, access to taxonomic expertise, and the time and resources that are available. Irrespective of the approach chosen, it is important that users are included in the process so that you can be sure that you are meeting their needs.

The first possible approach is for the Taxon Dictionary to be compiled locally by each user. In this case a copy of Recorder will hold a unique version of the Dictionary that is tailored for the particular needs of the organisation concerned. This can be a very efficient method to employ if suitable naming information is available, especially if the information is already held by the host organisation. For example, it may be possible to migrate the naming from a pre-existing electronic or paper-based system. This is also a good approach to adopt if the naming required is highly specialised and the taxonomic expertise is available in-house. Offset against this, there is the need for the time, taxonomic expertise and technical ability to compile and physically add the information to Recorder, and the likelihood of duplication of effort by other Recorder managers. It can also make data exchange with other Recorder users harder, as the TAXON_VERSION and TAXON_LIST_ITEM keys associated with a given name will differ in each copy of Recorder.

The second approach is to create and manage the Taxon Dictionary using a dispersed network of Recorder users. Different members of the community contribute their expertise and serve to validate and moderate the content through the process of peer-review. This ensures that the naming suits their needs and is both accurate and up to date. The system can work well providing that the user network contains the necessary taxonomic and technical expertise. Any conflicts in opinion also need to be rapidly resolved so that the peer-review system does not break down.

The final option is via a central coordinating organisation that is responsible for managing the taxonomy and providing updates to users. The organisation manages a core Taxon Dictionary which it compiles and updates with the help of taxonomic experts. There are a number of benefits to this approach. The principal one is that whilst Recorder users can of course provide input, they do not need to worry about the task of compiling the Taxon Dictionary themselves. The coordinating organisation can include taxonomists and data sources from beyond the Recorder user network, and is responsible for ensuring that the data are checked for accuracy and standardised. This can reduce the number of errors, removes the possibility of duplication of effort and facilitates data exchange. It is a system that can work well providing that the coordinating organisation has the long-term resources to manage and maintain the database.

To be efficient, both the second and third options require a centrally held master Taxon Dictionary from which regular updates are provided to the user network (although the second option could also operate through users agreeing to share

Dictionary checklists). This inevitably means that there will be a time delay between a suggested change and the dissemination of amended data to each locally held database.

Types of names required

The Taxon Dictionary should aim to contain all of the names that will be required, the range and types of which will differ according to the user base. Table 1 contains examples of the kinds of name that are used by a representative range of people. As can be seen, although there is often overlap, there are also differences and one naming source is unlikely to satisfy the requirements of all groups. In particular, the naming required by biological recorders, curators and research taxonomists can be quite dissimilar.

Biological recorders generally require a source that lists the current authoritative names and recording aggregates (groups of species that cannot be easily separated in the field). These may be taken from a national or regional inventory, or checklists compiled by recording schemes or societies devoted to the study of particular groups of organisms. It may also be desirable to add the naming that is used in popular field guides, and to include synonyms to help with collation of older data. A full list of synonyms will not usually be necessary, just those that have been in frequent use in the country concerned. Vernacular names should not be overlooked either, as many people are unfamiliar with scientific names (Hussey et al. 2006). Indeed, for some taxonomic groups such as vertebrates, vernacular names are often used in preference when discussing or exchanging information. Strict taxonomic treatments (e.g. monographs) may not always meet the naming needs of recorders. If there is a public interface, then informal names (for both species and higher taxa) become vital to open up access to the information held.

Museum curators will require names to be entered exactly as they appear on specimen labels, even if they are obsolete or misspelled. They will also need to store multiple identifications and to relate their verbatim names to synonyms and current names. Research taxonomists need full taxon concept information and comprehensive coverage of scientific names, both current and synonymous, but not recording aggregates or vernacular names.

Besides strict taxonomic lists, it is probable that other types of checklist will also be helpful. Examples include legislative and designatory listings, such as CITES, the Berne Convention and Red Data Books. These are particularly useful for reporting purposes and are integral to the activities of Records Centres and government decision makers. However, the inclusion of international lists in their entirety may introduce unwanted exotic species, such as gorillas and penguins, into your Taxon Dictionary! Supporting information such as taxon fact pages can also add value to the database.

Sources of names

Compiling a checklist requires considerable taxonomic expertise and can be very labour intensive. It therefore helps if there are existing name sources that can be accessed and used to both populate and validate the Taxon Dictionary. Checklists may be available within the host organisation, but it will often be necessary to search for an external naming source. This could be a published monograph, an on-line database, or a dataset held by a different organisation. Liaising

Table 1: Types of name required by different user groups

Type of name	Biological recorders	Consultants	Decision makers	General public	Museum curators	Taxonomists
Current scientific name	Y	Y	Y	Y	Y	Y
Previous scientific names (synonyms)	Y	Y	-	-	Y	Y
Recording aggregates	Y	Y	-	-	-	-
Mis-spelt names	-	-	-	-	Y	Y
Vernacular names	Y	Y	Y	Y	Y	-

with taxonomic specialists is the simplest way to locate the most up to date resource for a particular taxonomic group.

When investigating a prospective data source, as well as checking that it contains the breadth and types of names required, it is also important to ensure that the geographical scope will meet the needs of users. Biological recorders will require the names of organisms that can be found in the study area and perhaps also surrounding regions. The latter will help cater for observations of wide-ranging or invasive species. Museum curators on the other hand frequently need datasets that are global in coverage.

Appendix 1 lists a selection of potential on-line sources for national, European and global taxonomic checklists. Legislative and designatory lists are freely available from various websites and a web-search on the name of the resource concerned is usually sufficient to locate a copy.

Assuring data quality

The quality of data added should be as high as practically feasible. Incorporating the best available validated checklists is a very good start, but it is a fact of life that errors will creep into large databases, whether through mistakes in transcription or unspotted errors in the original data sources.

Checking for errors or otherwise verifying the accuracy of the naming data can be done at the data entry stage (Chapman 2005b), or later on if expertise is not initially available, or the dataset is too large to systematically check (Hussey et al. 2006). In this respect, access to taxonomic expertise and authority files for taxon names can be of significant help for data validation (Hussey et al. 2006). It is possible to compile a series of routines that check for errors and inconsistencies (see Challenges encountered and solutions found, page 72 and Chapman 2005b). Edwards (2004) observes that making data available for public (i.e. user) scrutiny is an efficient way to detect any remaining errors, and this is supported by our experience working with Recorder in the UK. Users should of course be made aware of the quality of the taxonomic checklists that are included via comprehensive metadata. In Recorder this can be stored in both the TAXON_LIST and TAXON_LIST_VERSION tables.

Importing data

Once a checklist has been compiled and validated, it will need to be imported into the Taxon Dictionary. Whilst limited editing of existing names is possible via the Edit Taxon Details function, the import of new lists requires the underlying database to be worked on directly. Possible methods for the physical addition of data are discussed in more detail in Challenges encountered and solutions found, page 72. Note that the Thesaurus Editor in the Collections Module extension contains additional tools to simplify the compilation and management of the Taxon Dictionary (EIMWiki n.d.).

When it comes to building checklists, anything that can be done to prevent 're-inventing the wheel' should be considered. Sharing checklists between different Recorder projects can potentially save a great deal of duplicated effort. In this context, it may be worthwhile establishing a central repository where different Recorder managers can store summary metadata for the various taxonomic checklists that they hold (and have permission to share with other Recorder projects). This would facilitate re-use of high quality datasets that have already been converted to Recorder format and can thus be imported easily.

Maintaining the Taxon Dictionary and providing updates to users

Following initial compilation, any Taxon Dictionary will need to be maintained either via regular updates or *ad hoc* changes. Amendments and additions will be necessary to track changing usages of a name, to add the details of newly recorded or collected species, to correct errors, and to incorporate new naming resources. The names attached to specimens and observational records may also change following redetermination events (Hussey et al. 2006). The method used to update the Taxon Dictionary will obviously depend upon how the project is being managed (see Possible approaches for building a Taxon Dictionary, page 68), but it is important to ensure that a simple mechanism exists whereby Recorder users can feed back any comments that they have. This need for maintenance should be factored into project plans and budgets.

Updating the Taxon Dictionary is simplest in situations where users log on to a single, centrally held copy of Recorder. In this case, it is only necessary to update one copy of the database and this could be undertaken by anyone with write access. In many cases it will be desirable to build programmatic tools (e.g. using SQL or VBA) to simplify the most common maintenance tasks (see Challenges encountered and solutions found, page 72 below). If updates to the Dictionary are supplied by an external organisation, however, it will be necessary to run an automatic or manual data transfer. Whichever approach is taken, it is good practice to create a back-up of the original database prior to making any changes. It may also be prudent to keep a series of archive copies of the Dictionary from back-ups taken either annually or every 6 months.

Providing updates to a dispersed group of users poses a number of additional considerations. In the UK, each registered user is provided with a local copy of the Recorder database with the core Taxon Dictionary supplied to them by a central body. Updates to the core Taxon Dictionary therefore need to be distributed to all users (e.g. as downloads from the Recorder website). Issues that must be addressed include how to notify users that an update is available, and the frequency and format of the provision of updates. The simpler the mechanism the better, as users may not have the time or technical expertise to perform lengthy procedures. Version control is a potential problem and it will be necessary to clearly identify different editions of the Taxon Dictionary. To help users successfully update their copies of Recorder, a customer support element will also be required.

Individual users can, of course, always add their own naming data and assign local TAXON_VERSION_KEYs though these can hinder data transfer. If these names are subsequently incorporated into the Taxon Dictionary as part of an official update, then they can be converted to centrally compiled keys using the Merge Data Items tool.

Case study: managing the Taxon Dictionary for Recorder in the UK

Approach adopted

In the UK the Taxon Dictionary is managed by a central coordinating organisation. This role was initially fulfilled by the Joint Nature Conservation Committee (JNCC), but since 2000 development of taxonomic content has been led by the Natural History Museum as part of their broader National Biodiversity Network (NBN) Species Dictionary Project.

The Species Dictionary originated as the Taxon Dictionary within Recorder 3.3, but is now managed as an independent product and forms a contribution towards the UK's NBN. It aims to provide a freely accessible standard reference for the names of all organisms found in the UK, along with associated biodiversity-related information. A key component of this will be a fully comprehensive, maintained master list of all UK taxa. The nomenclatural data that are being collated can be accessed via a free website (http://www.nhm.ac.uk/nbn/) or can be provided upon request as a Microsoft Access database to NBN partners. Web Services that allow direct interrogation of the data are also planned for the future. The Species Dictionary aspires to meet the naming requirements of people working in the field of UK biodiversity, including government agencies, NBN partners, biological records centres and wildlife societies. It is thus a focal point for taxonomic information and needs to fulfil multiple purposes. There is a growing user base and, besides Recorder, clients of the Species Dictionary include the Global Biodiversity Information Facility (GBIF), the NBN Gateway and the UK's Environment Agency. Users play a key role in guiding the evolution of the project by recommending resources that they would like to see added and providing details of any errors that they have spotted. This helps ensure that priority is given to the preparation of resources for which there is genuine demand.

The Species Dictionary is a database that builds upon the rigorous NBN Data Model (Copp 1998).

Its physical structure closely follows that of the Recorder Taxon Dictionary (as described in section The structure of the Taxon Dictionary in Recorder 2002 and Recorder 6, page 66), but a number of amendments and additions have been made to cope with the broader needs of the project. Enhancements that are relevant to managers of Recorder are outlined in section Challenges encountered and solutions found, page 72. A data warehouse approach has been adopted, with all datasets stored and managed in a single, centrally held database. This approach has been found to make it particularly easy to locate errors within and between checklists, compared to federated data sources. The Species Dictionary is a work in progress and relies upon input from a network of over 100 taxonomic specialists and users. At the time of writing it offers up to date, comprehensive coverage of the naming for approximately 75 % of the taxonomic groups that occur in the UK. This naming is held in over 200 individual checklists and the database currently contains 237,000 different names.

The Taxon Dictionary that is provided to users of UK copies of Recorder is based upon a subset of the above information. The Natural History Museum have responsibility for compiling and maintaining the taxonomic data, but JNCC coordinate the physical provision of Taxon Dictionary updates to Recorder users. A copy of the Species Dictionary database is sent to JNCC on a periodic basis, who extract new and changed data and transform them to fit the structures of the Taxon Dictionaries employed in Recorder 2002 and Recorder 6. JNCC then post the prepared updates as self-executable files to the Recorder (http://www.recordersoftware.org/) and NBN (http://www.nbn.org.uk) websites, from where they can be downloaded by owners of Recorder software.

Utilising naming data that are already being compiled as part of another project provides an opportunity for the application of taxonomic standards across projects and promotes the use of standard recommended checklists within the biodiversity community. Both of these aspects greatly enhance storage and exchange of high quality data.

Challenges encountered and solutions found

Many of the challenges that have been encountered whilst compiling the Species Dictionary are of wider relevance to those wishing to create and manage a Taxon Dictionary for Recorder. Selected examples are outlined below.

Sourcing data can be a very time-consuming process, particularly for less well-studied groups of organisms, where checklists may not yet exist and expertise can be hard to locate. To satisfy the varied needs of different user communities, and in the absence of an existing comprehensive UK species list, we have needed to source and add a large number of different checklists. The inclusion of such a diverse range of datasets is desirable in many ways, but the large choice of lists may confuse Recorder users, and data entry and report compilation in Recorder would be simplified if it was possible to use only a single species list. This is not currently possible within the UK context, but could be a sensible route for other Recorder projects.

The task of adding checklists to the Species Dictionary can take considerable effort, depending upon how much data manipulation is required and the size of the dataset. We have found that most of the people who supply us with checklists either do not have the time, or technical skills, to manipulate their datasets into the optimum format for data import. As we rely upon voluntary input, we accept checklists in whatever format data providers are happy to supply and then convert the data to fit the Dictionary data model ourselves. To maximise the efficiency and accuracy of data import we have developed and adopted a standardised technique (Fig. 2). The data are initially manipulated to fit a defined spreadsheet format. This spreadsheet is then imported into the working Microsoft Access database by a largely automated process using a series of our own VBA routines. The routines populate all appropriate database tables and are run via a user-friendly interface. Comprehensive checks are built in to ensure data quality and search for likely problems, such as formatting inconsistencies and potential homonyms (see below). Carrying out quality checks at each step of the process ensures that data are standardised as far as possible and likely sources of error minimised. Other tools have been

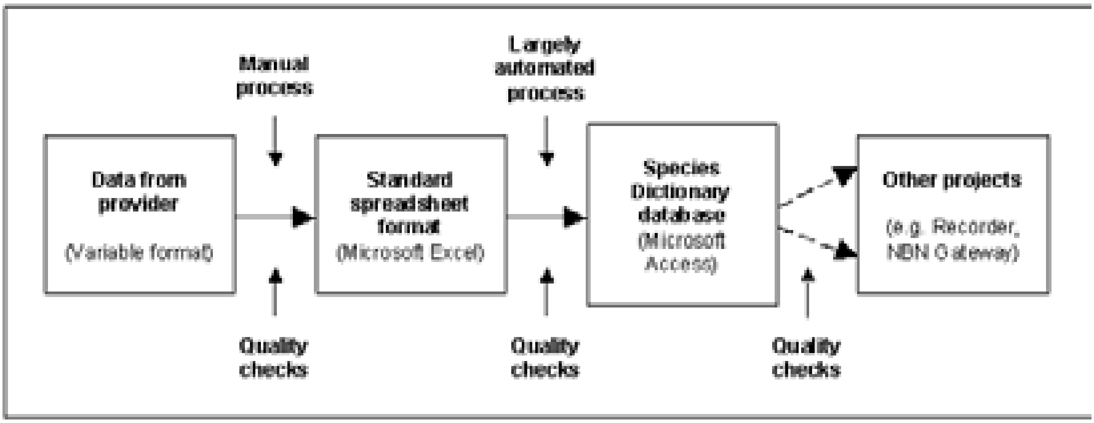


Fig. 2: Flow diagram of the processes used to add data to the Species Dictionary.

developed to manage existing database content.

Homonyms, in the context of taxonomy, are names that can apply to more than one organism and whilst they occur occasionally in vernacular names (e.g. the English name Redshank can be a bird, a moss or a flowering plant) they are more of an issue for scientific names, particularly genera. They can present a real challenge, particularly in resources that contain names from more than one Kingdom. It is important that the correct version of the name is selected, both when adding a checklist and storing data, otherwise the meaning of a record can be dramatically altered. We check for over 5,400 known generic homonyms as part of our import routine. To further help detect homonyms and to aid interpretation and filtering of data, we have devised a series of informal taxon reporting groups. All names in the database are assigned to an OUTPUT_GROUP that identifies the type of organism (e.g. terrestrial mammal, liverwort, millipede). A further two-tier system of INPUT_GROUPs can be used to aid searching that uses broader categories (e.g. mammals, lower plants, invertebrates - terrestrial and freshwater). This has required altering and expanding the existing NBN Data Model. Table 2 provides an example of a homonym; note how the output groups can help select the relevant name and that the problem is resolved if naming authorities are employed.

Intellectual Property Rights (IPR) are another aspect that must be considered. When including a list within Recorder it is necessary to ensure that the IPR of the data owner is not infringed. In order to protect the legal rights of both the data providers and the Natural History Museum, we have designed a Data Collation Licence for use with the Species Dictionary. It allows data owners to clearly specify how their datasets can be accessed, and by whom, and ensures that both parties agree what the data can and cannot be used for. In most situations, we have found that checklist compilers are very happy to encourage the wider use of their data for not-for-profit purposes. To protect IPR and ensure that datasets can be reconstructed as intended, we have adopted a policy of entering naming data exactly as provided. Any errors that are spotted are reported back to the data owner for checking. This confirms that the changes are

Table 2: Example of a homonym showing how the use of naming authorities and informal groups can remove confusion

Name	Authority	Rank	OUTPUT_GROUP
Ctenophora	(Grunow) D.M. Williams et Round, 1986	Genus	diatom
Ctenophora	Blackwell, 1870	Genus	spider
Ctenophora	Hatschek, 1888	Phylum	comb jelly (Ctenophora)
Ctenophora	Meigen, 1803	Genus	insect - true fly (Diptera)

valid and helps the data provider to improve the quality of their dataset. However, there are cases where it is not appropriate to correct mistakes. For example many published checklists (particularly legislative lists) contain misspellings that are now in widespread use.

Aggregating so many different datasets into the Species Dictionary has shown that the naming data currently being used for biodiversity purposes are extremely messy. Spelling variants and typographical errors are surprisingly frequent. Scientific names can also be presented with or without naming authorities or, in the case of plants, with inconsistent authority abbreviations. When genuine taxonomic synonyms and vernacular names are also considered, it becomes clear that the same organism can be known by a confusing variety of different name strings. It is desirable to retain these variants in order to be able to reproduce published lists accurately and to capture names that may have found their way onto other on-line resources.

If all of the names used for a taxon occur in the same list, they can be stored as synonyms and related to the same preferred name in the TAXON_LIST_ITEM table. If, however, the names occur in different lists then the original NBN Data Model does not allow them to be adequately linked, or be combined for reporting purposes. To solve this problem we have created a new table called NAMESERVER. The NAMESERVER table has a simple structure as shown in Table 3. Its main role is to map every name string to a correctly formed, currently recommended scientific name (via TAXON_VERSION_KEYS). We have also included three flag fields that help to characterise the names. The TAXON_VERSION_FORM field indicates whether a name is correctly formatted and spelt (including whether the authority is complete). The TAXON_VERSION_STATUS field shows whether the name is a recommended name or a synonym, and TAXON_TYPE flags whether the name is a scientific or vernacular name.

Thus the name-server links all the different names that have been applied to an organism, including scientific names, vernacular names and their misspellings. It can subsequently be used as a resource finding and query expansion tool that ensures that all relevant data are returned for an organism when a search term is entered, regardless of which particular name string the data are attached to. This query expansion facility can help determine whether a search on a name string that returns no results arises because no data are available for the organism concerned, or because data are present, but stored against a different name. The above functionality fits the criteria for what is now being termed taxonomic indexing (Patterson et al. 2006). The ability to map equivalent names has proven to be a most important and much appreciated feature, and the name-server has now been introduced within Recorder 6.

Finally, perhaps the biggest challenge that we have found concerns the ongoing maintenance of the taxonomy. To remain current, the datasets need to be amended to reflect changing taxonomic understanding and incidences of new species.

Table 3: Structure of the NAMESERVER table

Field name	Description of content
NAMESERVER_ID_KEY	Primary key
INPUT_TAXON_VERSION_KEY	The TAXON_VERSION_KEY of the name to which the following data fields relate.
TAXON_VERSION_FORM	Flags completeness/correctness of name. Options are [W]ell-formed, [I]ll-formed, [U]nverified.
TAXON_VERSION_STATUS	Flags whether name is a [R]ecommended name, a [S]ynonym, or [U]nverified .
TAXON_TYPE	Flags whether name is a [S]cientific or [V]ernacular name.
RECOMMENDED_TAXON_VERSION_KEY	The TAXON_VERSION_KEY that corresponds to the recommended scientific name of the INPUT_TAXON_VERSION_KEY

Keeping track of the relevant literature and biological record data can be a major undertaking if, as in the Species Dictionary, a comprehensive species listing is required. To approach this we are establishing a voluntary network of experts, with different people responsible for monitoring change in each taxonomic group. This process has highlighted how difficult it can be to detect instances where a name has been interpreted differently across datasets, or has changed in meaning through time. To reliably detect, store and map this information requires an in-depth knowledge of both the taxonomy concerned and the datasets that are held within a particular copy of Recorder. Careful mapping of equivalent uses and separation of dissimilar interpretations of a name is an issue that is becoming more and more relevant now that projects are starting to aggregate biodiversity data (e.g. observational records) from a variety of sources.

References

Chapman A. 2005a. - Uses of primary species-occurrence data. Retrieved 24 August 2006 from Global Biodiversity Information Facility website: http://www.gbif.org/prog/digit/data_quality

Chapman A. 2005b. - Principles and methods of data cleaning. Retrieved 24 August 2006 from Global Biodiversity Information Facility website: http://www.gbif.org/prog/digit/data_quality

Copp C.J.T. 1998. - The Recorder Project Systems Analysis. Unpublishedreport. JNCC, Peterborough.

Copp C.J.T. 2000. - The NBN data model and its implementation in Recorder 2000. Unpublished report.The NBN Trust, Newark.

Copp C.J.T. 2004a. - The NBN Data Model, Part 1 Description of the Model. Unpublished report. The NBN Trust, Newark.Available on NBN web site www.nbn.org.uk. Accessed March 2007.

Copp C.J.T. 2004b. - The NBN Data Model, Part 2 Physical Implementation of the Model. Unpublished report. The NBN Trust, Newark. Available on NBN web site www.nbn.org.uk. Accessed March 2007.

Edwards J. L. 2004. - Research and Societal Benefits of the Global Biodiversity Information Facility. BioScience 54: 485-486.

EIMWiki n.d. - EIM hosted wiki for Recorder, available at http://eim.metapath.org/wiki/index.php?title=Recorder.. Accessed 18 September 2006.

Hussey C., Wilkinson S. & Tweddle J. 2006. - Delivering a name-server for biodiversity information. Data Science Journal 5: 18-28. Retrieved 18 September 2006 from http://www.datasciencejournal.org/.

Page R. D. M. 2005. - A taxonomic search engine: federating taxonomic databases using web services. BMC Bioinformatics 6: 48. Retrieved 18 September 2006 from: http://www.biomedcentral.com/1471-2105/6/48.

Patterson D. J., Remsen D., Marino W. A. & Norton C. 2006. - Taxonomic indexing - extending the role of taxonomy. Systematic Biology 55: 367-373.

uBio n.d. - Universal Biological Indexer and Organiser website available at: http://www.ubio.org. Accessed 15 September 2006

Appendix 1: on-line sources for national, European and global taxonomic checklists

National

1. Austria: Nature Web
http://www.natureweb.at
2. Belgium: Biodiversity in Belgium
http://bch-cbd.naturalsciences.be/belgium/biodiversity/biodiversity1.htm
3. France: INPN
http://inpn.mnhn.fr/
4. Italy: Fauna Italia
http://faunaitalia.it/
5. Netherlands: Netherlands Biodiversity Information Facility
http://www.nlbif.nl/
6. Romania: Romanian Species Information Center
http://mybiosis.org/nature/
7. Slovakia: Databank of Slovak Fauna
http://www.dfs.sk/
8. Spain: Fauna Ibérica
http://www.fauna-iberica.mncn.csic.es
9. Sweden: Artportalen (Species Gateway)
http://www.artportalen.se/
10. United Kingdom: NBN Species Dictionary
http://www.nhm.ac.uk/nbn/
11. United Kingdom: Nature Navigator
http://www.nhm.ac.uk/naturenavigator/

European

1. Biological Collection Access Service for Europe (BioCASE)
http://search.biocase.org/synth-ui/
2. Bird Names Translation Index
http://www.mumm.ac.be/~serge/birds/
3. Delivering Alien Invasive Inventories for Europe (DAISIE)
http://www.europe-aliens.org/
4. European Nature Information System (EUNIS)
http://eunis.eea.europa.eu/species.jsp
5. Euro+Med PlantBase
http://www.emplantbase.org/
6. European Network for Biodiversity Information (ENBI)
http://www.enbi.info/ see also: Multilingual Access (Workpackage 11) at http://www.enbi.linguaweb.org/
7. European Register of Marine Species
http://www.marbef.org/data/erms.php
8. Fauna Europaea
http://www.faunaeur.org/
9. Flora Europaea
http://rbg-web2.rbge.org.uk/FE/fe.html
10. Multilingual Animal Glossary of Unveiled Synonyms (MAGUS)
http://www.informatika.bf.uni-lj.si/magus.html. Also http://www.agroweb.bf.uni-lj.si/nomenklatura-multilingual1.htm. Multilingual Glossary of Common Names of Animals. Mammals and Birds.
11. Species 2000 Europa
http://sp2000europa.org/
12. Society for the Management of European Biodiversity Data (SMEBD)
http://www.smebd.org/

Global

1. algaeBASE
http://www.algaebase.org/
2. Amphibian Species of the world
http://research.amnh.org/herpetology/amphibia/index.php
3. Avibase - the world bird database
http://www.bsc-eoc.org/avibase/avibase.jsp
4. Biosystematic Database of World Diptera
http://www.sel.barc.usda.gov:591/diptera/names/searchno.htm

5. FishBase
http://www.fishbase.org/
6. Global Biodiversity Information Facility (GBIF)
http://www.gbif.org
7. Index Fungorum
http://www.indexfungorum.org
8. Index to Organism names (ION)
http://www.organismnames.com
9. Integrated Taxonomic Information System (ITIS)
http://www.itis.usda.gov/
10. International Plant Names Index (IPNI)
http://www.ipni.org/
11. Mammal Species of the World (MSW)
http://nmnhgoph.si.edu/msw/
12. National Center for Biotechnology Information Taxonomy Browser
http://www.ncbi.nlm.nih.gov/Taxonomy/
13. Orthoptera Species File Online
http://osf2x.orthoptera.org/O/OSF2X2Frameset.htm
14. Species 2000
http://www.sp2000.org/
15. The Global Lepidoptera Names Index
http://www.nhm.ac.uk/research-curation/projects/lepindex
16. Universal Biological Indexer and Organiser (uBio)
http://www.ubio.org/
17. w^3 TROPICOS
http://mobot.mobot.org/W3T/Search/vast.html
18. World Biodiversity Database
http://www.eti.uva.nl/tools/wbd.php
19. World Odonata list
http://www.ups.edu/x6140.xml
20. World Spider Catalogue
http://research.amnh.org/entomology/spiders/catalog

A consideration of the web-based future of Recorder

John van Breda
Dorset software Ltd.
Corbiere House
New Quay Rd
GB-BH15 4AF Poole
John.vanBreda@DorsetSoftware.com

Keywords: Biological recording, Recorder, Internet, Web, Open source

Abstract

The Recorder application, associated modules and tools provide a platform for storage, retrieval and analysis of biological records and collections management data. Whilst widely used by Local Record Centres in the United Kingdom, as well as by recording schemes and other organisations, its uptake by amateur biological recorders has been limited to those both technical and keen enough to install such a complex application. The application is also currently limited to Windows users. However, the demographics of the UK biological recording community are such that there are large numbers of amateurs producing valuable data in a non-digital or unstructured digital form. This leads to delays and inaccuracies between data collection and the data being made available in a useful manner.

A web based solution allows a wider audience to record their observations digitally since there is no need for application installation on each machine. In addition, support costs are reduced since the application is centralised on a single web server and a database server.

Introduction

The Recorder application suite provides a client-server application for biological recording and natural history collections management on a local area network. In this format it is widely used by Local Record Centres in the United Kingdom and is also used by the MNHN museum in Luxembourg. It is being adopted by Naturalis, the national natural history museum of the Netherlands. The software's strengths lie in the enforcement of data standards at the point of data entry and validation through the interface and, notably, through the NBN data model that it supports. The quality of the data model ensures the useful lifespan of the data entered through the application is maximised, even if the application itself is replaced or updated.

This paper considers the need for a solution that uses internet technology to allow Recorder application functionality to reach a wider audience.

Historic Considerations & Development

A web based development approach was considered during the original procurement of Recorder 2000 in 1999. This approach was rejected because of the limited number of possible data contributors who possessed a web connection at that point in time. Further consideration was given to a cross-platform desktop application and database independence, but these options were rejected on the basis of increased costs. It was decided that the only financially viable approach would be to develop a Windows desktop application using Delphi and Microsoft Access 97, but to structure the code in such a fashion that future changes to the database platform would have a minimal impact on the system and incur a minimal cost.

In 2003 the low scalability of Microsoft Access, which in practice limits the Recorder database to around 1.25 million species observations, was reconsidered and a decision was taken to change the back-end database to a scaleable client-server RDBMS. Microsoft SQL Server 2000 was selected over Oracle because of the availability of an option to use MSDE, an effectively free SQL Server version for databases up to 2Gb in size. MySQL was also rejected as, at the time, support for a procedural language and stored procedures on the server was limited or absent. Recorder 6 has the option to use MSDE for data storage at no additional cost. Professional users or those requiring large datasets are able to upgrade to one of the full versions of SQL Server for support of theoretically unlimited datasets. Recorder 6 was developed initially during 2003 with the single most significant change being the migration of the data to SQL Server 2000.

Key factors supporting the need for a web based version

Providing Incentives for Recording

Online environments provide opportunities to incentivise users to contribute data. Many recordings are made on paper, in local spreadsheets or databases but are not contributed to the wider community because the user does not believe they will be of value. This is often not the case, particularly in under-recorded taxonomic groups. By allowing the user to immediately review their recordings alongside national datasets, contributors are able to appreciate the value of their observations and are more likely to contribute further data.

Limitations of Hardware

Currently, Recorder is used widely in Local Record Centres (LRCs) within the UK. Uptake of the application by amateur recorders, who form the backbone of biological recording in the UK, has been limited to those technical and keen enough to install a complex application on their computers. A large number of recorders are limited to recording on paper, leading to important records being lost and also additional data entry effort for Local Record Centres and other collating organisations. In addition this approach does not enforce any data quality control at the point of data collection, resulting in possible problems with the data later. Whilst any new PC sold will have sufficient performance to run Recorder 6, many potential users only have access to an older computer with limited memory, disk space and processor performance. This is particularly true in larger institutions, where the cost of updating the PCs for an entire network of users may be huge. They are therefore forced to continue with old desktop technology not capable of running Recorder 6. As Recorder is supplied with a substantial Taxon Dictionary database in the UK the disk space usage is also a limiting factor.

Data Sharing

In today's business climate there is a need to allow organisations to specialise and to share their data with other organisations on a near real-time basis. Web services allow resources held at any location in the world to be accessed from any other location in a platform independent manner. Biological Recording organisations are now requiring up to date and accurate digital species and habitat dictionaries. These requirements are not well supported by the isolated client-server and desktop environment in which Recorder currently operates.

Connection Availability

The adoption of web technologies in households and businesses within the UK and across Europe since 1999 has been rapid. Between January and April 2006 57 percent of UK households could access the internet, 69 percent of which were via broadband (UK National Statistics, 2006a). The percentage of internet users with broadband connections has risen from 0% at the end of 2001 to 72.6% in June 2006 (UK National Statistics 2006b). As some of the households with no web connection do not have a home PC, it can be assumed that a large majority of households with a home PC also have web connectivity. Connections have moved away from the cost per minute charging model of the early internet years towards packages where a single monthly charge covers all internet use.

This enables serious use of web applications without incurring additional line usage costs and indicates that consideration must now be given to providing a strong web based platform for biological recording and analysis within Europe.

System Requirements

Usage Scenarios

Currently, Recorder 6 is used in the following scenarios amongst others:

- Amateurs recording data for personal use, with possible contribution to national schemes or Local Record Centres
- National Recording Schemes, charities and other organisations collating data for a particular purpose
- As a conduit for data eventually contributed to the NBN Gateway, GBIF or BioCASE
- Local Record Centres collating data in order to provide a biodiversity information service to town planners, scientists and the general public

The following usage scenarios are envisaged which benefit from a web based version of Recorder:

- Amateurs and the general public contributing directly to national surveys and recording schemes where all data is available for public access
- Amateurs and the general public contributing directly to national surveys and recording schemes where access to data is controlled by user logon and varying levels of access rights
- Internal usage by staff at recording institutions allowing remote entry of data to the central database. This also applies to volunteer collaborators who are not actually employed by the institution
- Providing online education on the importance of biological recording, or the provision of online biodiversity information such as dynamic distribution maps and reports for education and public awareness
- Allowing town planners, scientists and the general public to directly query the data and produce their own reports and maps on demand. However, care must be exercised particularly where raw data is used to facilitate decisions made by authorities without a full understanding of the data. As such the role of Local Record Centres in providing a value added data analysis service is not diminished by the existence of a web based tool, as professional interpretation of the data is still required.

It is clear that the requirements of the general public and amateur users differ quite substantially from the professional scientific, conservation and planning users of the system. A web application that allows simple and rapid data entry, along with the ability to run predefined reports and draw limited distribution maps, will help to engage the public in the issues of biological recording and increase the ease with which data can be gathered from the community. However, this application would not meet the needs of many of the professional groups involved or even some amateurs. These users require flexible and rapid data entry as well as powerful reporting and analysis. Furthermore, integration with other applications and uses becomes critical, for example integration with GIS systems already in use for spatial analysis and web service links between data providers and consumers within and across organisations.

User Interface

Having identified the target audience for a technical solution, the fundamental requirements for the user interface are the same for web based and desktop solutions. The interface must be clear, intuitive, consistent and accessible. As large volumes of data are involved, the interface must allow rapid data entry and provide facilities for import and upload of existing datasets. The complexity of biological recording means that many tasks require a rich and dynamic interface which is easily achieved in a desktop application but difficult to achieve within a web environment. However, the current rate of development of web technology, including the mainstream acceptance of technologies such as AJAX (Asynchronous JavaScript and XML), allows modern web interfaces to be advanced, rich and useable. A basic

HTML interface requires that each user interaction is submitted to the server and the entire web page is refreshed to reflect the required change to the page. However, AJAX allows asynchronous operations and updates can occur on just one part of the web page without refreshing the entire page. This simple achievement removes one of the major limitations of web interfaces and is leading to the proliferation of rich and usable web sites that is currently occurring. Despite this, it is likely that for advanced use, onsite users will continue to benefit from a desktop client-server version of the application for the near future.

Web developments incur additional development costs over their desktop counterparts, particularly when full consideration is given to support for multiple browsers and writing truly accessible code. Whilst the additional testing required is not insignificant, it is directly related to the additional target audiences reached by supporting multiple browsers and accessibility.

Development Considerations

Business Model

Implementation of a sustainable business model is essential to the success of a web based version of Recorder. The following aspects require investment of time and/or money, without which the product is likely to fail:

- Infrastructure for source code management, including disk storage, backup and source code version control.
- Infrastructure for community users, including a web presence, download and support options.
- Product development. As new web technologies arrive so quickly, it is imperative that modern products are kept up to date.

The Recorder 6 desktop application project was managed by contracting out the software development process to a commercial company, Dorset Software. The resultant code and IP rights are owned by JNCC, who are primarily responsible for the ongoing support of the application. The fees charged for the software cover production of installation media and a limited amount of support for installation, but do not cover ongoing product support and development. It is clear that any development of a web version of Recorder needs a sustainable model which includes support and development during the full lifespan of the product.

Technical Approach

There are many approaches that can be taken to web development of a project of this nature, particularly when factors including the development tools together with the funding model are taken into account. This paper discusses just 2 of the possible approaches.

The first approach considered is development of an open source web application with the following characteristics:

- Developed using open source tools. A LAMP (Linux Apache MySQL PHP) based project is considered most likely to succeed, since there are many developers with some experience of these technologies who could contribute to the project.
- Ability to deploy onto a low cost hosted web server. LAMP hosting services are generally cheaper than other technologies such as Windows IIS, Oracle or SQL Server. By ensuring the running costs are minimised, the availability of the product for use by the many organisations operating on low budgets is maximised.
- Utilise an existing open source Content Management System (CMS) to provide site content and design tools. This allows the administrator of each web site to design a fully featured site around the core tools provided by the Recorder web application. In effect the Recorder web application code will consist primarily of modules or PHP scripts which exist within the infrastructure provided by the CMS. A default installation includes a basic website template that can be modified by the administrator of the site.
- In order to initiate the project successfully, consideration should be given to an initial investment to develop the core components of the system, enough to provide a simple but

usable application. This will help to garner the interest of the available open source community to ensure a viable ongoing developer team.

- The initial development is likely to be simple, including: a basic data entry system (fig. 1); ability to exchange data with desktop Recorder; and simple reporting and distribution mapping facilities.

The second approach is to contract out the complete development of a fully functional and powerful web based version of Recorder. This development has the following characteristics:

- Selection of developer tools not on the basis of cost, but on the basis of development speed and capability. Tools such as ASP.Net are considered, utilising SQL Server as the database to facilitate simple data sharing and compatibility with desktop Recorder.
- Development using a highly designed, modular and service oriented architecture (Fig 2). This facilitates the reuse of modules and integration with other systems. For example the solution includes a web service for accessing taxon dictionary data. This web service can be replaced with a wrapper service written to allow taxon dictionary data to be used from another existing data provider application.

To summarise, the first approach results in a usable application with a significantly lower initial investment. The second approach results in a well designed, fully featured and scalable application which integrates well with other systems but requires a higher initial investment.

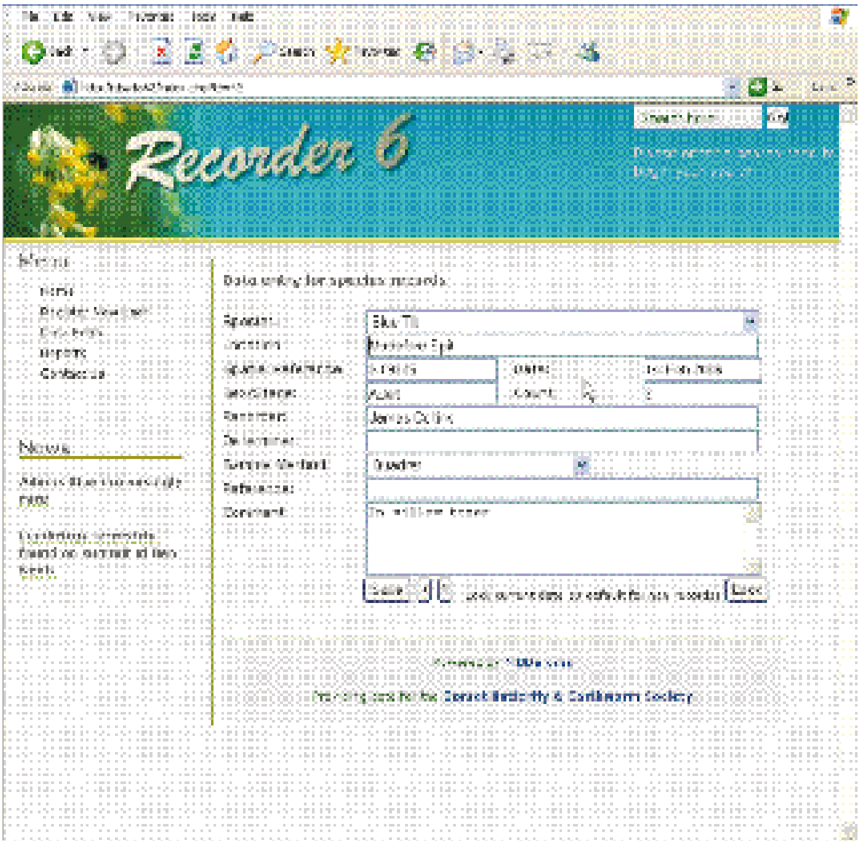


Fig. 1: Example of a data entry page for an open source web application. (Dorset software[1], unpubl.)

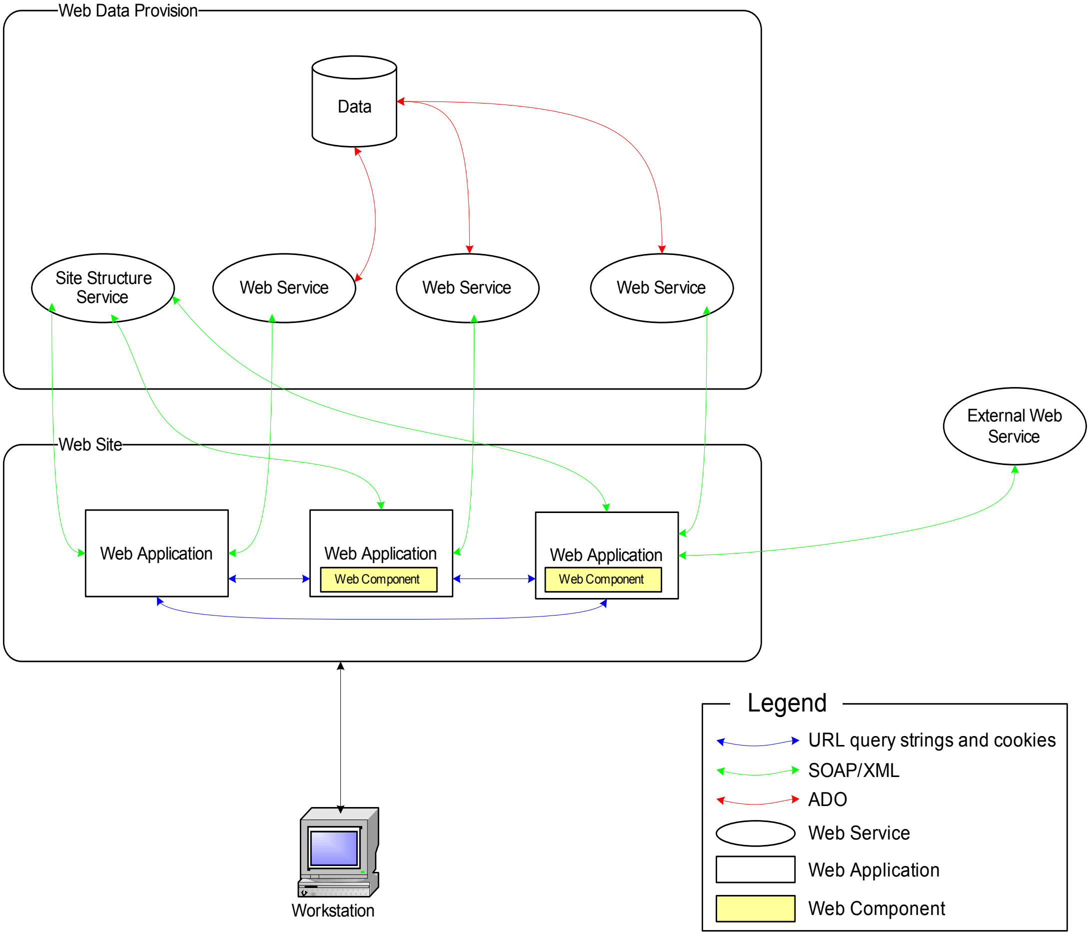


Fig. 2: A generalised version of the Recorder Web toolkit architecture design and the technologies used for communication between each component (Dorset software unpubl.[2])

References

Dorset software 2006. - OpenRecorder. Proposal v1.0. 25th September 2006. Unpublished report. Musée national d'histoire naturelle, Luxembourg.

Dorset software 2005. - Technical System Design for Recorder Web. Vers.1.1. 6th January 2005. Unpublished report. Musée national d'histoire naturelle, Luxembourg.

National Statistics 2006a. - Uk National Statistics. Omnibus Survey and Survey of Internet Service Provider. Statistics on Internet Access. Retrieved November 2006 from http://www.statistics.gov.uk/cci/nugget.asp?id=8

National Statistics 2006b. - Uk National Statistics Quarterly Internet Connectivity (June 2006). Retrieved November 2006 from http://www.statistics.gov.uk/downloads/theme_commerce/InternetConnectivity_Quarterly_back_data.xls

The UK NBN Gateway. A national system

Steve Wilkinson
Joint Nature Conservation Committee
City Road
GB-PE1 1JY Peterborough
Steve.Wilkinson@jncc.gov.uk

Keywords: NBN Gateway, United Kingdom

Within the UK there is a large and very active network of naturalists who independently collect records on range of species and, to a lesser extent, habitats. The total number of records is estimated to be within the hundreds of millions. Clearly, easy to access this vast resource could significantly enhance many aspects of conservation direction setting and delivery. The NBN Gateway (www.searchNBN.net) was developed in an attempt to improve the availability and accessibility of these data (Fig. 1).

The Gateway is a website that provides a range of simple outputs, such as distribution maps or a list of species for a geographical area, on the basis of an underlying repository of biodiversity data (Fig. 2).

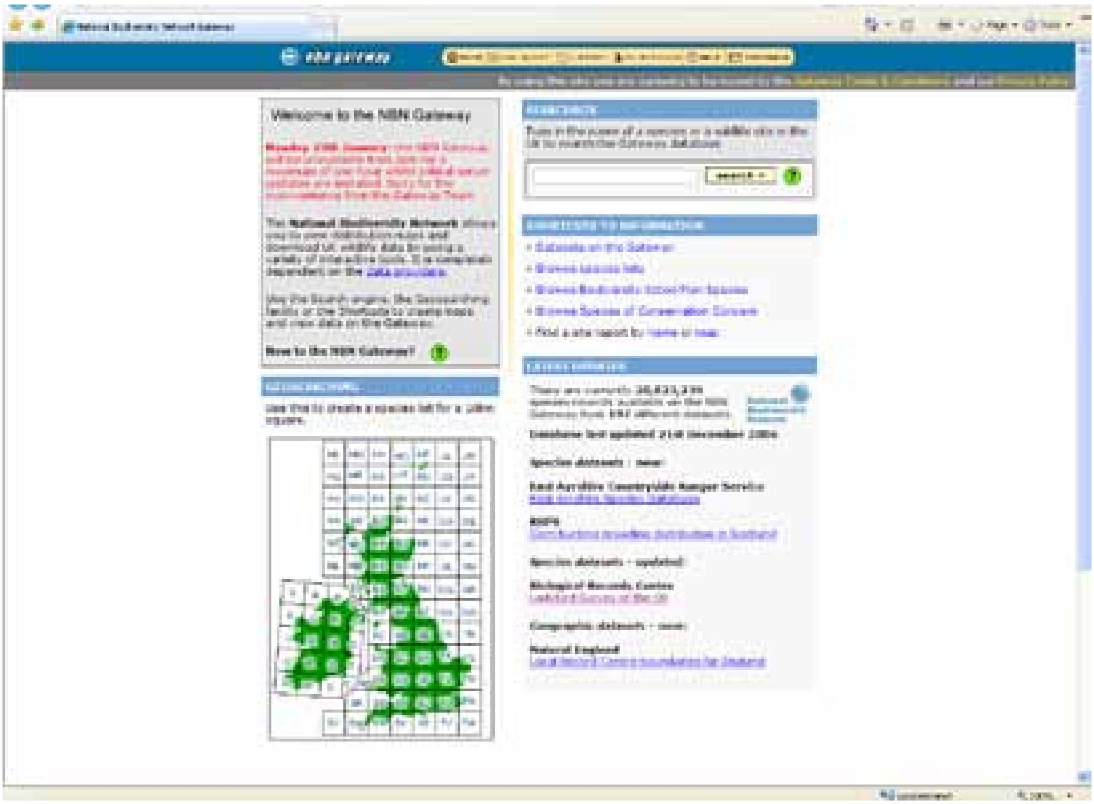


Fig. 1: The NBN gateway home page.

Encouraging the data providers to contribute data to this was obviously paramount to the projects success and thus considerable efforts were focused on catering to the requirements and needs of this group.

The primary concern of the providers was that they would lose control over their data (i.e. in many cases they wanted to limit who could access it). To address this, a comprehensive set of access controls were developed which allowed them to not only control who could access the data but also the resolution the data was made available at and what attributes of the data were exposed. In addition the sources of the data were clearly acknowledged at all times. Much of the functionality of the website was also developed to meet the needs of the data providers, attempting to produce on-line versions of products they were familiar with such as national distribution maps (Fig. 3).

The project has been very successful and now approximately 19 million species records are available through the system and the funding partners are now exploring ways in which these data can be applied within conservation in the UK. In addition, where the data providers are prepared to make their data more widely available they are also exported to the GBIF network.

The Gateway is now entering a new phase of development with the creation of a suite of "web services". In essence these will allow a user to seamlessly integrate the data and functionality available through the Gateway into their own websites and applications. The importance and potential applications of these is discussed.

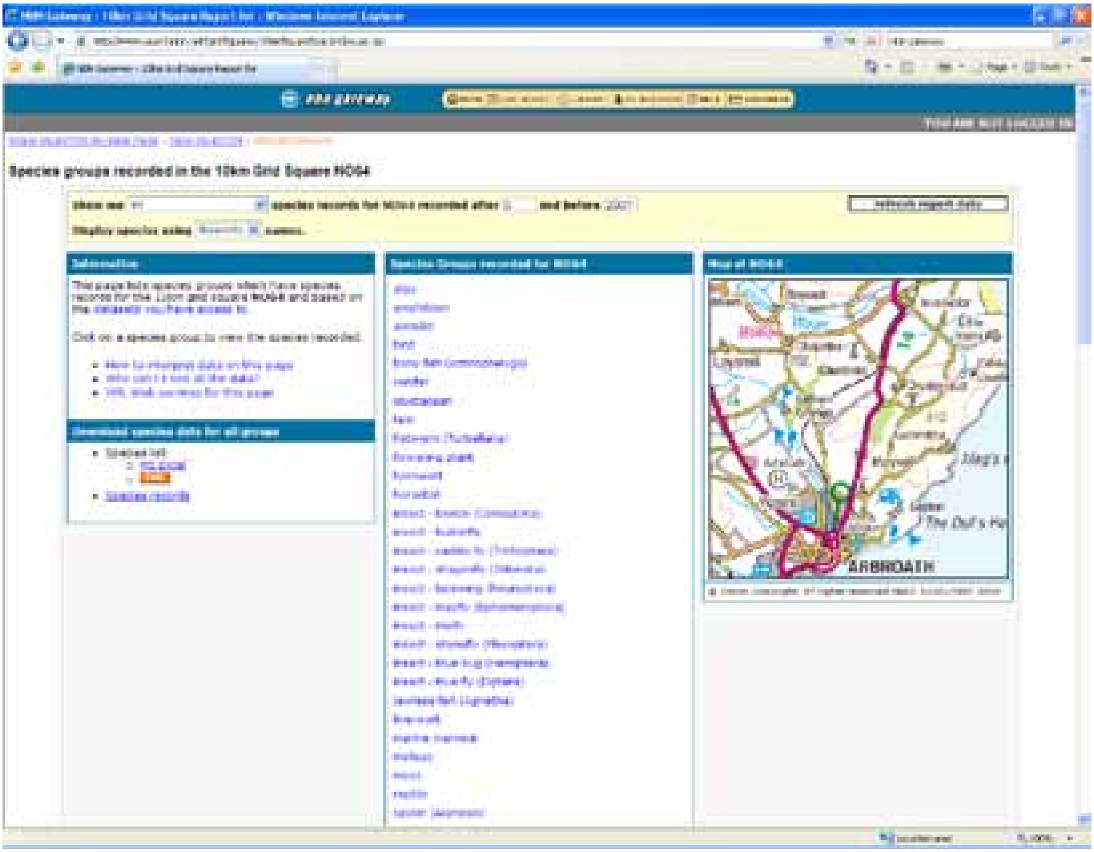


Fig. 2: The Gateway provides a list of species for a geographical area, on the basis of an underlying repository of biodiversity data.

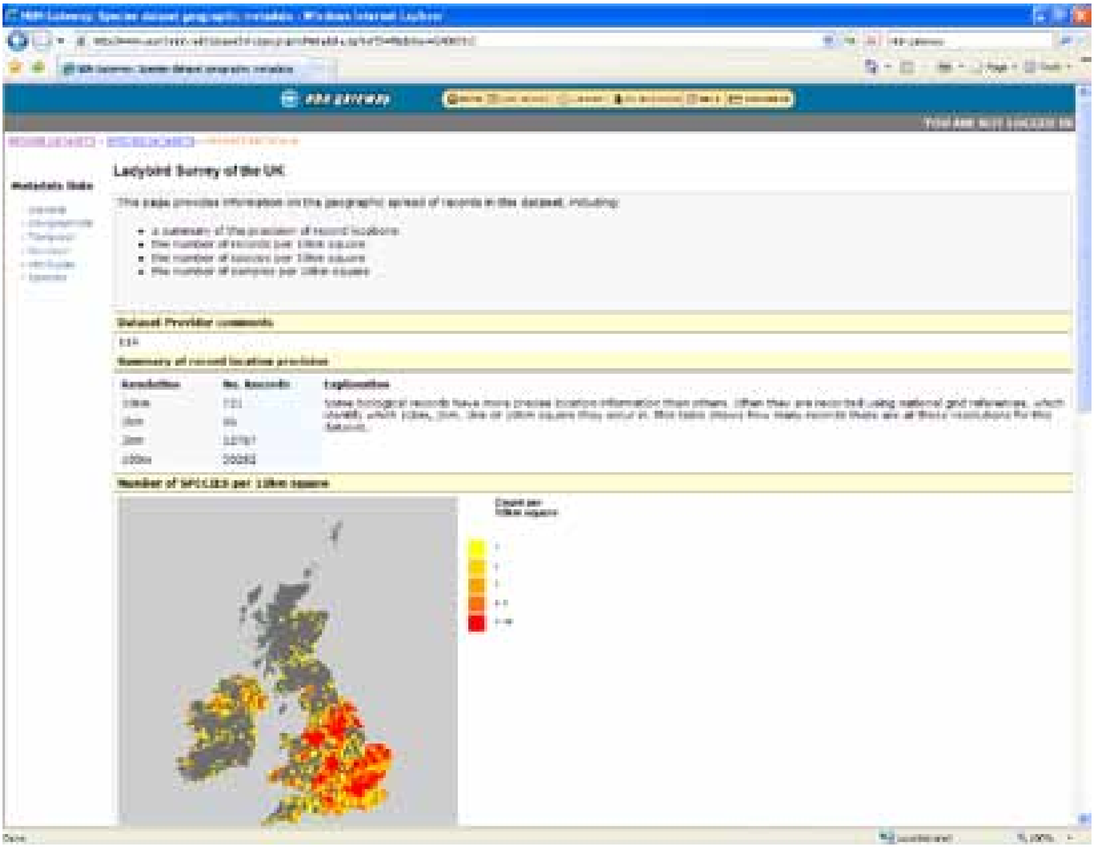


Fig. 3: A distribution map from a national ladybird survey of the UK.

CEDaR – a regional system

Damian McFerran
Centre for Environmental Data and Recording Northern Ireland (CEDaR)
Ulster Museum, Botanic Gardens
IE-BT9 5AB Belfast
Damian.McFerran@magni.org.uk

Keywords: regional biological recording, environmental data, CEDaR, Northern Ireland

Presentation abstract

The Centre for Environmental Data and Recording (CEDaR) is the Local Records Centre for Northern Ireland. Through the financial support of the Environment and Heritage Service (EHS, An Agency within the Department of the Environment for Northern Ireland) and funding in-kind from the Ulster Museum, CEDaR was established at the Ulster Museum, Belfast, Northern Ireland

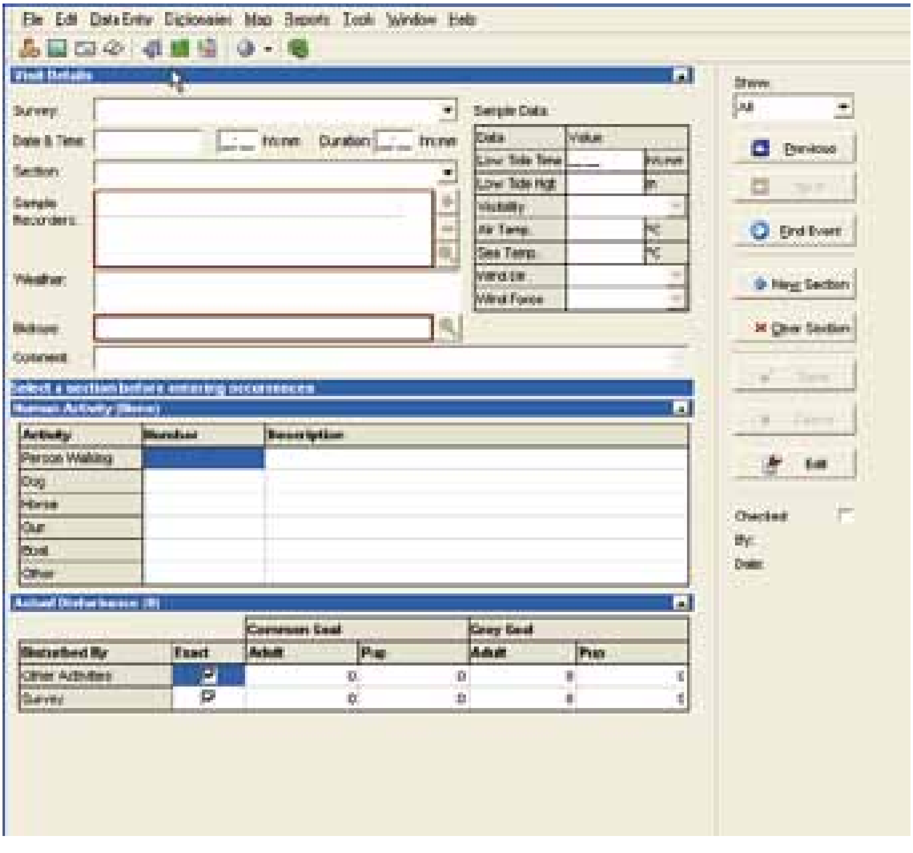


Fig. 1: View of a special addin in Recorder 6 which is used by CEDaR to enter Seal Count observation data.

Fig. 2: Overview of biological recording projects carried out by CEDaR.

in January 1995. The core function of CEDaR is to collect, collate, manage and disseminate environmental records that relate to Northern Ireland, its geology and associated coastal waters. The information obtained by CEDaR is collated on several databases, of which Recorder 6 is but one — approximately 1.5 million records are collated on Recorder 6 (Fig. 1). A marine and earth sciences database are also used by staff. Information has been made available by a number of individuals, groups and societies and Government and non-Government organisations. This collection of data suppliers forms the Environmental Recorders' Group.

CEDaR is currently approaching the end of the second year of its Five-Year Business Plan. The Plan states that the overall aim of the CEDaR project is to develop an effective and efficient mechanism for the delivery of environmental data and recording activity in support of the activities of the Northern Ireland Biodiversity Strategy. In order to fulfil this aim, activities are delivered through the following eight subject areas — Data Management; Web Management; Database Management; National Biodiversity Network (NBN) Developments; Environmental Recorders' Group; Access to Data; Initiatives and Publicity and Marketing.

CEDaR has also established a close working relationship with colleagues in National Parks and Wildlife Service (NPWS), Dublin, Ireland. This partnership has facilitated, for example, The Ground Beetles of Ireland on the Web (2000),

Figure 3: Example of a website available through HabitasOnline: Priority species of Northern Ireland.

DragonflyIreland (2000-2004) and more recently LichenIreland (2005) (Fig. 2).

We are particularly indebted to EHS for their core financial support and The Heritage Council, Kilkenny, Ireland for their contribution to key projects. The staff continues to develop and manage numerous biological recording projects (Fig. 2). A selection of projects launched since 1995 are available through the HabitasOnline suite of web sites (Fig.3). The creation of these sites has been assisted, managed and encouraged by staff of the Sciences Division, Ulster Museum (www.habitas.org.uk).

In 2006 a Records Centre was established in the Republic of Ireland. This development should facilitate further collaboration in the establishment and management of all-Ireland recording projects. Indeed, the significance of the all-Ireland component of DragonflyIreland was commented upon in the joint ministerial foreword of the book to summarise the project, *The Natural History of Ireland's Dragonflies*. Therefore, it is both conceivable and within our collective interests that, within the next five to ten years, we will witness the collation and dissemination of comprehensive and current Irish data sets for key species and groups. This advance should be welcomed by all with an interest in the conservation and management of our species and habitats

The Luxembourg System – Building a national bio-and geo-diversity node

Tania Walisch
Musée national d'histoire naturelle Luxembourg
25, rue Münster
L-2160 Luxembourg
twalisch@mnhn.lu

Keywords: Natural heritage data basing, Recorder, biodiversity, geo-diversity, national node

Abstract

The Luxembourg National Museum of Natural History (LNMNH) has the mission to assemble and manage observation and collection data on biological and geological diversity and to make this information available to the public. The Museum hosts a MSSQL server Recorder 6 database, which functions as the central node for a network of institutions, associations and naturalists. They digitize their data using satellite installations of Recorder 2002 and contribute their Recorder data via the NBN data transfer model to the central database. The Museum curators are involved in the digitization of occurrence and specimen information held at the Museum. Further commitments of the Museum are the production of standardized taxonomic lists, the release of new Recorder versions, data security and access to the data held in the central database. There are currently two ways of accessing the data via the Web. Thus the Luxembourg natural heritage portal (LUXNAT) extracts species observation data to a general type of user and the BioCASE portal (Biological Collection Access Service for Europe) provides researchers with access to biological occurrence and collection data. In addition to presenting existing tools for data access and data entry, I will try to identify deficiencies of the existing and propose alternatives with a special emphasis on of web technologies. I will thus promote the idea of an integrated natural heritage information system for Luxembourg based on a Web enabled Recorder database.

Background

Studying and documenting the natural heritage are among the primary commitments of the Luxembourg Natural History Museum (LNHM). To meet this obligation the Museum gathers, assembles and conserves natural heritage information and promotes and participates in its digitization. Furthermore it provides access to the assembled information.

Natural Heritage information comprises biological records relating to fauna, flora and biotopes as well as earth science records on fossils, minerals and rocks.The Museum documents the Natural heritage through the inventories of occurrences in the field, as well as collections of specimens.

Since it's beginnings in 1854, the Museum has been involved in gathering and conserving specimens from Luxembourg and from abroad. A great number of volunteer naturalists have since contributed to the Museum collections through donations of their private collections, catalogues and notes. The scientific value of any of these natural history specimens is not only determined by the object itself, but also by the quality of the information associated with it, for example on an attached label or in a catalogue. Essential bits of information are the name and/or geographical position of the gathering site and date, the name of the collector and of the determiner.

In the 1980's, some curators within the LNHM started to make use of the emerging information technologies in order to improve the management of natural heritage information. Over the past decades the efforts to digitise and structure information have differed in the various departments of the Museum and have led to a considerable number of databases and data formats. Thus the complexity of database structures ranges from simple Access or Excel tables to a quite complex relational database for biological field records called Luxnat. Luxnat has become the largest database at the Museum and currently holds about 500,000 biological field records. It has however several shortcomings. It does not allow the professional management of natural heritage collections (loans, storage, thesauri, etc). It is based on a complex, partly redundant database structure and it does not provide easy data transfer nor direct contributions of external collaborators.

In the year 2000, a new department was created, whose tasks were the digitisation, management and access of natural heritage information. After a few months of searching for a database solution enabling the professional management of natural heritage information, the unit decided in favour of Recorder 2000, an application developed by the Joint Nature Conservation Committee in the UK. Many reasons motivated this choice, the most important being the quality of the underlying NBN data model, the open flexible built allowing new functionality to be added through the use of COM technology and the integrated NBN data transfer format using XML. Moreover the relatively easy installation on standalone computers gave the possibility to scientific collaborators to enter their data at home and to send extracts of their data to the centralised Recorder database held at the Museum on a regular basis.

In order to make Recorder 2000 usable in Luxembourg, an addin for the Luxembourg geographical coordinate system (Gauss-Luxembourg) had to be developed and the taxon dictionary and location term lists had to be adapted to the Luxembourg situation. As JNCC had designed Recorder for the management of biological field records only, the Museum took the initiative to develop an additional collection module, which could manage life and earth science specimens and collection data. This major development including testing took three years and was finished in 2004. It was financed through the eCulture project by the Ministry of Culture and Higher Education, which had also financed the built of the web portal providing access to natural heritage information held at the Museum's Recorder database.

Concurrently to implementing Recorder 2002 for biological record information in Luxembourg, the Luxembourg Museum developed its role as a national node for biological collection data for the European funded BioCASE project.

This historical and more recent background explains why LMNH holds the largest datasets on natural heritage information in Luxembourg and it is recognised as the national node for bio- and geo-diversity information in Luxembourg.

Developing a central data node on a national level is not a trivial undertaking especially in a small country. In fact the complexity of the task and the kind of technical developments needed are basically the same as in a larger country, whereas the number of staff, the infrastructures and financial resources are generally less important than in a larger country. In the following part I will attempt to present a possible vision of a natural heritage information system for Luxembourg, in which the Museum deals with data assembly and conservation at the national level and ensures a platform for data exchange and dissemination. I will present what has been achieved so far and what remains to be done, with a special emphasis on the usefulness of web technologies.

A central node of bio- and geo-diversity information

First of all the national node should facilitate and improve the networking of data contributors and data users. Figure 1 shows a diagram of actors and data flows of a possible information network in Luxembourg.

Actors

The Museum is represented as the central node for natural heritage data from different sources (Fig. 1). This central node functions like a technical platform, assembling, but also giving access to natural heritage data. The Museums role is also to control and validate natural heritage information it receives and to ensure its long-term preservation.

The backbone of biological recording are volunteer naturalists, mostly having acquired the status of scientific collaborators of the Museum and as such are encouraged to share their data. Regional

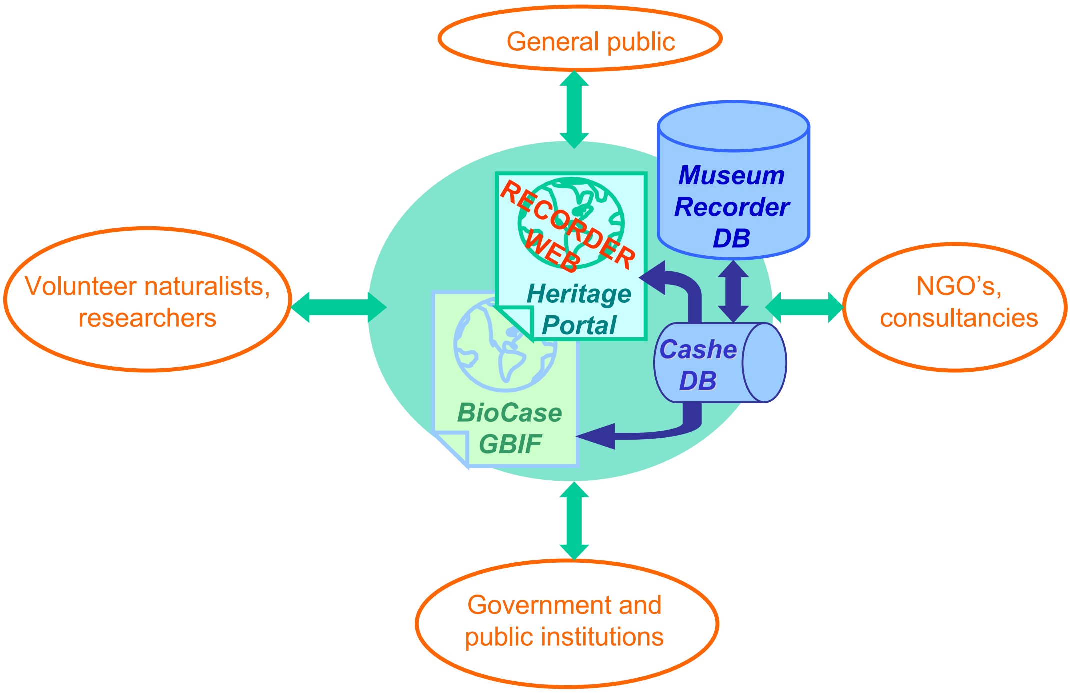


Fig. 1: Diagram of actors and data flows of a possible information network in Luxembourg.

biological stations dealing with practical nature conservation work and monitoring have built up considerable biological databases. They are backed by the communes of a region and the ministry of the environment. Private consultancies, or nature conservation NGO's also contribute to biological recording in Luxembourg, often they do this as part of a nature conservation related project, which they have contracted from a public institution.

Government institutions need biodiversity information for land planning and management and are responsible for more than 80% of data requests to the Museum's natural heritage database. They also produce new taxon or biotope occurrence records as part of nature conservation or restoration projects or environmental impact studies.

Researchers in the domain of biodiversity often need information about taxon and biotope occurrences as the basis for their research study. In many cases they also record new occurrence data, as part of their research study.

Unfortunately a large part of the natural heritage information, especially records dating back to the last century, only exist in paper reports or publications or in a poorly structured digital form. Therefore various institutions, upfront the Museum, have embarked on a large digitisation and data standardisation effort of existing data, using the Recorder application and its underlying NBN data model as a central tool (Copp 2000).

Recorder - the essential tool

Recorder with its collection and thesaurus extensions, has become the essential instrument for the collation and management of natural heritage information in Luxembourg. In 2000 the Museum started off with a Recorder 2000 server installation using an Access database, which was updated to Recorder 2002 a few years later and eventually to Recorder 6 and MSSQL server in 2005. It actually represents the central database for a number of institutions like regional record centers, associations and naturalists who collect species and biotope occurrence records. They all together represent a network of distributed data providers who have installed their own copies with unique licence codes of Recorder 2002. The NBN data transfer schema allows the secure and non-redundant data transfer according to a Document type definition (DTD). This generally allows the exchange of data among Recorder databases and, more particularly, the data flow to the central Museum database. The Museum produces and manages the taxonomic and biotope lists as well as other term lists and is responsible for the release of new Recorder versions. Thus in the present situation every single user needs to update his Recorder installation each time a maintenance upgrade or a change in term lists is released.

The distributed Recorder copies allow for decentralized data input and for easy, reliant and non-redundant data transfer. This allows data holders to participate in a networked data system, but can hold a certain number of difficulties for the individual data providers. The complexity of Recorder often puts off people who wish to record simple observation data. There are people wanting to contribute, but who don't have the appropriate hardware to install and run Recorder. Moreover, single users sometimes lack the necessary tools or technical skills to ensure backups on a regular basis, thus risking the loss of data.

The need for a web-enabled Recorder

A web-based Recorder, where people would enter data via the web into a central database, using entry skins adapted to their domains of interest and to the level of data complexity they need, would greatly facilitate and speed up data entry for everyone involved in biological recording and last not least it would avoid updates and hardware problems at the user end.

This should incite more people to contribute their data even enable school kids and the general public to get involved. In 2005 we therefore designed, together with experienced people form the JNCC, the Sheffield Museum and Dorset software, what we called a Recorder Web toolkit (Dorset software 2004). The toolkit allows organizations to set up their own Recorder Web for their recording community and link it to their Recorder database (van Breda 2007). The idea was welcomed by other data centers outside Luxembourg, and we have since attempted to promote a corporate built of Recorder Web (EIMWiki n.d.).

An alternative to the full Recorder Web proposal based on Microsoft technologies would be a development based on open Source tools. Like the Recorder Web toolkit, Open Recorder would allow any organization to set up biological recording and reporting at little cost and require only basic technical experience. The approach for Open Recorder would be an initial development coded in php script, where the data would be stored in MySQL using a simplified data model and import/export tools (Dorset software 2006). OpenRecorder would use an Open Source Content Management System like for exemble ModX, providing the user with the ability to restructure and template the site to meet their requirements, as well as integrate other functionality such as blogs and RSS news feeds (Dorset software 2006). Once delivered to the community it is expected that there will be subsequent additional functionality created either through the Open Source community or further costed developments. However the long-term aim would be to develop the more powerful Recorder Web toolkit as the preferred data collection and reporting application for large organizations.

The Luxembourg Natural Heritage Portal LUXNAT

Presentation

The Luxembourg natural heritage portal (Luxnat 2003) is one way of accessing the data held at the Museum's Recorder database. It allows the user to enter a latin or, if available, a french taxon name in a free text field or to select a name from a systematic hierarchical tree. The user also has a geographical search option by entering a location or a place name in a free text field or selecting a ten or five kilometer grid square on the map of Luxembourg, which yields a list of taxon occurrences and their most recent observation date, ordered by phylum. The portal delivers a distribution map for a given taxon (Fig. 2). Furthermore it shows an image and a short description of the taxon, extracted from the taxon facts in the Recorder database if available. The geographical precision of an observation dot on the map corresponds to one square kilometer.

Fig. 2: Example of a result screen for a given taxon from the Luxembourg natural heritage portal (http://luxnat.mnhn.lu).

Technical specifications

The Luxembourg natural heritage wrapper has a 3-fold architecture composed of a backend, a front-end and a middle-end. The backend represents the data base part, the middle end holds the access programs to the database like specific algorithms, transaction management, data storage management, and the connection pools for performance optimization. The front-end assures the presentation layer. It also detects the terminal type, manages navigation, calls up the middle end and generates pages in the format asked for by the terminal. This three-fold architecture allows for the independence of the middle part, which holds a number of specific technical developments. The application is installed in an apache Tomcat servlet container. This engine generates html pages, which are eventually dealt with by the Apache server for web diffusion. Tomkat uses Java SE Development Kit (JDK) resources. The described unit is open-source (Cap Gemini Ernst & Young 2004).

The middle end contains a number of web services, which communicate with the front end through a SOAP protocol. For security reasons the web services are not directly linked to the real Recorder 2002 Acccess database. In fact the relevant tables from the taxon dictionary and taxon occurrences elements of Recorder are exported to a temporary MSAccess database where they are filtered into views. A final set of views containing ready-processed and relevant information are then exported to an MSSQL server database, which the Luxnat wrapper accesses and queries for information it needs. In 2004 the Museum installed Recorder 6 on a new server and migrated its Access database from Recorder 2002 to an MSSQL server database of Recorder 6. The temporary views needed by the Luxnat wrapper had to be generated form Access tables linked to the Recorder database via ODBC.

Context

The natural heritage portal is the result of a pilot project, which was part of a larger project called eCulture, targeted towards improving the digitization and web accessibility of cultural and natural heritage information. The screens are compliant with the graphical design policy for web pages of the Ministry of Culture. At the start of the project, the development of Recorder 6 and the collection and thesaurus extensions weren't finished yet. Therefore a certain number of compromises were necessary. The wrapper had to be connected to the Access database of the Recorder 2002 database of the Museum. It extracted only field occurrence data of taxa, neglecting biotope occurrences and omitting attribute information like for instance abundance, stage of development or observation method. Collection and specimen data were left aside as well. The geographical accuracy of a site was leveled to one square kilometer to avoid any misuse. The wrapper was targeted towards one general type of user, familiar with basic taxonomical concepts.

There has not been a follow up project of the pilot wrapper. Nevertheless the pilot natural heritage portal has been functioning now for more than two years and despite the obvious shortcomings it has become a rather useful tool for scientific collaborators and researchers wishing to get a general overview of the distribution of species in Luxembourg or to determine data deficiencies concerning the inventory of certain species.

Access to detailed data

So far all requests for detailed natural heritage data have to be made in writing to the Museum. The processing of requests takes a long time and represents a considerable workload for the natural heritage information department at the Museum. In fact prior to sending the data for a given project, the Museum asks the consent of each scientific collaborator having provided data for the given area. We are currently elaborating a data access policy and a licence system , which encourages data providers to give unrestricted data access for research and nature conservation projects (Deontology wiki 2005).

In order to facilitate access to detailed data for professionals like town planners, foresters, nature conservation agents the Museum plans to build a GIS web module for Recorder 6. This module will integrate open source mapping software. It will manage three user levels: a standard user, an advanced one and a specialist user. Users are setup with advanced access for specific geographical zones (e.g. polygons of nature reserves) and for a limited duration of time (e.g. one year). Advanced access will provide maximum geographic accuracy. The users will be able to add layers to the map and define a polygon area. The polygons will be used to create a dataset for a report. Occurrences will be depicted on the map by an icon or a filled grid square. Attributes of the icon like shape, colour and transparency relate to attributes of the occurrence: the membership to a phylum, the age of its recording, its size and its positional precision. Clicking on a distribution point on the map will display underlying data. The GIS web module for Recorder should be developed as part of an integrated Recorder Web toolkit for bio- and geo-diversity (Dorset software 2004). Thus it should not be limited to the specific needs of nature conservation actors, researchers or naturalists but it should also stimulate the interest of school kids and the general public. The user could for example navigate through images of different insect and plant families and look up the distribution of selected species or look for a picture of a specimen of the same species conserved in the Museum collection.

The BioCASE portal and the need for a database Cache

The Museum participated in a EU funded project called BioCASE, which has created a web-based information service providing researchers with unified access to biological collections in Europe (Guentsch 2007). The central database provides

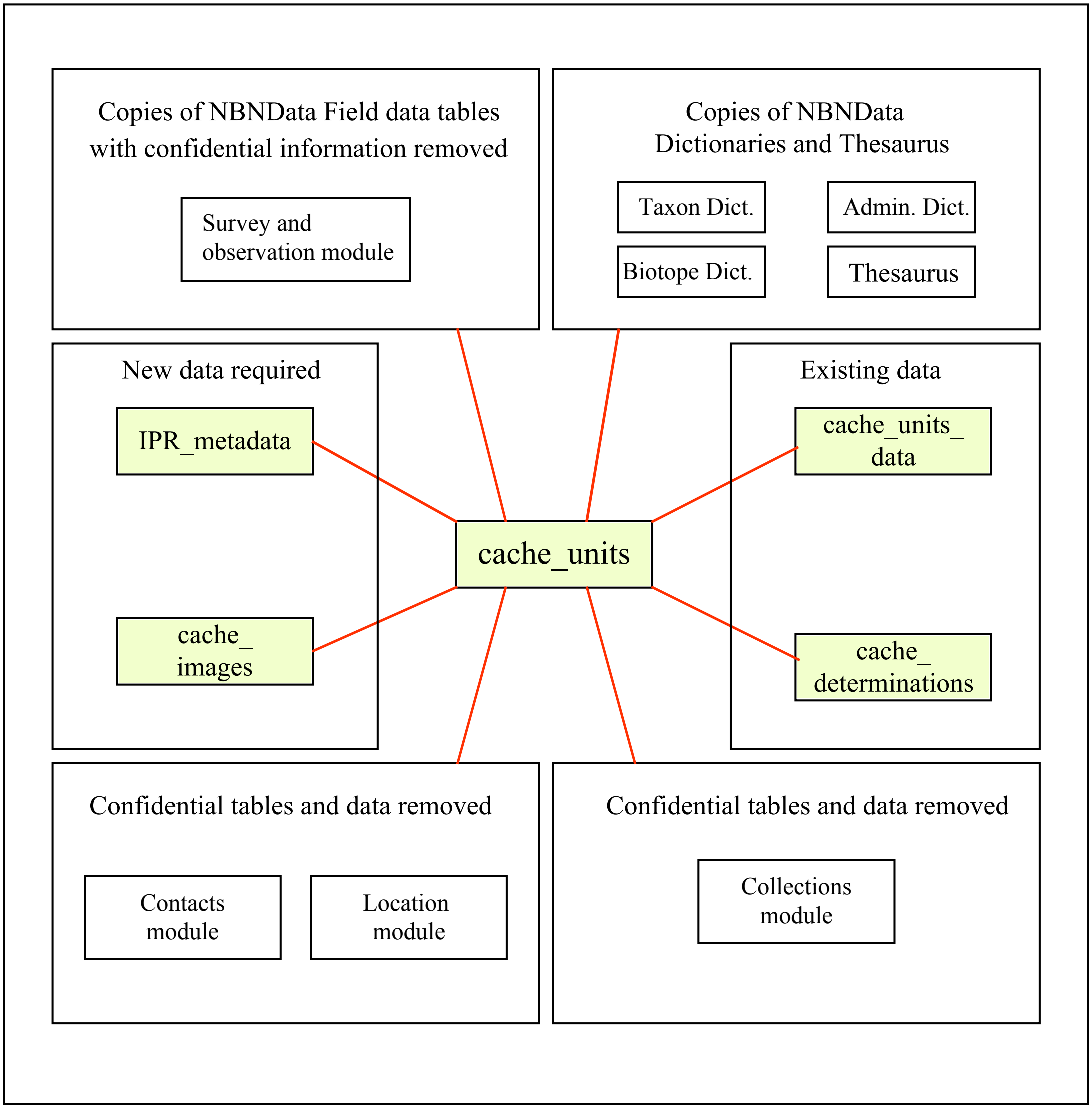


Fig. 3: Relationship between the cache tables and standard Recorder Modules (Copp 2005)

data to the global biodiversity information facility (GBIF n.d.).

The Recorder database could be directly connected to the wrapper. However the NBN model, underlying Recorder, has been optimized for flexibility in storage and handling a very broad range of data, is not an efficient structure for rapid delivery of information (Copp 2000, Copp 2005). It was therefore necessary to link the BioCASE wrapper to simple cache structures with minimal table joins and ready-processed information. These tables hold the most relevant and non-confidential data in order to enable rapid data retrieval to web-users. They are highly redundant, but this is not a serious issue for data delivery. Recorder 6 is currently running on an MSSQL Server database and all functions, views and SQL scripts of the database cashe were written for that environment.

The cache_units table is the main cache table against which web queries are made (fig. 3). Its contents are strongly de-normalised and include concatenated data from related tables. For instance, measurements and field recorders are formatted and concatenated in the cache_units table but multiple values, are also stored in linked relational tables (e.g. **cache_units_data**) or linked to the original database tables through foreign keys stored in the cache_units table (Fig. 3.) (Copp 2005).

Identifications are handled slightly differently. The identification (name) given for a unit is the preferred name used in the database (preferred flag in original record) and where there are multiple determinations against a specimen or field record these are formatted and stored in the **cache_determinations** table. The principle is to cut the number of links in queries to the minimum (Copp 2005).

The Cashe_Units_data table includes measurements and descriptors from the taxon_occurrence, biotope_occurrence, occurrence and collection_unit_data tables.The Cache_images table will come to hold details and locations of images and thumbnails which will be accessible from a linked website and linked to data in the cache_units table. The Cache_ipr_metadata table will hold IPR statements relating to use of information and images available on the web site. Retrieval of data can be further improved through the judicious use of indexes (Copp 2005).

Finally the database cashe was not only written for the BioCASE wrapper but was meant to become the main access point for any new web interfaces querying the Recorder database.

References

BioCASE n.d. - Biological Acccess Service for Europe. Website available at: www.biocase.org. Accessed October 2006.

Cap Gemini Ernst &Young 2004. - Manuel d'Installation et maintenance pour le Wrapper MNHN. Unpublished report. Luxembourg, Musée National d'Histoire Naturelle.

Copp C.J.T. 2005. –Mapping data from recorder 6 and the Collections module to a web database cashe. Unpublished report. Luxembourg, Musée National d'Histoire Naturelle.

Copp C.J.T. 2000. - The NBN data model and its implementation in Recorder 2000. Unpublished report. The NBN Trust, Newark.

Dorset Software 2004. - Technical System Design of Recorder Web. Unpublished report. Luxembourg, Musée National d'Histoire Naturelle.

Dorset Software 2006. - Proposal for the initial scope of the OpenRecorder development. Unpublished report. Luxembourg, Musée National d'Histoire Naturelle.

EIMWiki n.d. - EIM hosted wiki for Recorder available at http://eim.metapath.org/wiki/index.php?title=Recorder. Accessed 13 October 2006.

GBIF n.d. - Global Biodiversity Information Facility. Website available at: www.gbif.org. Accessed October 2006.

Guentsch A., P. Mergen, & W.G. Berendson, 2007. - The BioCASe project – a Biological Collections Access Service for Europe: 100-102, in Walisch T. (ed.), Collating and managing natural sciences field and collection records in Europe. Proceedings of the First International Recorder Conference. Luxembourg 2-3 December 2005. Ferrantia 51, Musée national d'histoire naturelle, Luxembourg, 156 p.

LUXNAT n.d. - A Luxembourg Natural Heritage portal. Website available at http://luxnat.mnhn.lu. Accessed October 2006.

Deontology wiki 2005 – PBWiki hosted wiki for a natural heritage data access policy for the Luxembourg national museum of natural history. Website available at: http://deontology.pbwiki.com. Accessed October 2006.

van Breda J. 2007. – A consideration of the web based future of Recorder: 81-86, in Walisch T. (ed.), Collating and managing natural sciences field and collection records in Europe. Proceedings of the First International Recorder Conference. Luxembourg 2-3 December 2005. Ferrantia 51, Musée national d'histoire naturelle, Luxembourg, 156 p.

The BioCASE Project - a Biological Collections Access Service for Europe

Anton Güntsch & Walter G. Berendsohn
Botanic Garden and Botanical Museum Berlin-Dahlem
Department. of Biodiversity Informatics and Laboratories
Königin-Luise-Str. 6-8
D-14191 Berlin
BiodiversityInformatics@bgbm.org

Patricia Mergen
Royal Museum for Central Africa,
Department of Zoology,
Leuvensesteenweg 13
B-3080 Tervuren, Belgium
patricia.mergen@africamuseum.be

Keywords: BioCASe, BioCASE, ABCD, collection databasing, collection networking

Abstract
The implementation of world wide data networks for interchange of and access to biological collection information was hindered by the diversity of collection management systems using a variety of different operating systems, database management systems, and underlying data models. Several initiatives have tackled this problem by defining data standards and Internet protocols as well as by providing software modules enabling collection database holders to link their collections to international networks without having to modify the implementation of their systems. In the course of the BioCASE project such standards and software have been developed focusing on rich collection data. Both provider software packages and portal implementation tools are available, and several portal implementations have proven their stability, including the Global Biodiversity Information Facility (GBIF).

Background

During the last few decades, an increasing number of institutions holding biological collections started to build electronic inventories. Apart from primarily scientific uses in most cases this was to facilitate and document daily activities such as managing accessions in botanical gardens and loans and label printing in herbaria. Available collection management software often did not meet the specific needs of institutions. Consequently, many individual applications have been developed almost independently using very different information models, database systems, and database interfaces resulting in a vast diversity of existing systems.

With the emerging web technologies and increasing awareness of the immense value of unified collection information several initiatives are aiming at building data networks making individual and local data sets available for the international scientific community (Berendsohn 2003). The Species Analyst network (http://species-analyst.net/) was built using the z39.50 protocol (a standard developed for the library community). The exchange format for content data here used was named the Darwin Core (http://darwincore.calacademy.org/) and consisted of a relatively simple set of elements considered to be adequate for most types of collections. A 5th Framework European Union project, the European Natural History Specimen Network ENHSIN (http://www.

nhm.ac.uk/science/rco/enhsin/) followed a similar approach in using a relatively simple element set but used XML technologies (http://www.w3.org/XML/) from the beginning (Güntsch 2003). The ENHSIN pilot network provided access to 7 distributed collection database (3 herbaria and 4 zoological collections) and served as a prototype for the implementation of the Biological Collection Access Service for Europe. BioCASE (http://www.biocase.org) is a comprehensive information network giving access to biological collection and observation data of any kind using advanced XML technologies and the fine-grained element specification ABCD (Access to Biological Collection Data, http://www.bgbm.org/tdwg/codata/). Finally, DiGIR (Distributed Generic Information Retrieval, http://digir.sourceforge.net/), succeeded the Species Analyst network and is now the most widely implemented XML based protocol used in conjunction with the Darwin Core.

All networks and their underlying technologies are built on two basic requirements: first, primary data should stay with the data owner rather than being exported into a central data repository to ensure that information holders have full control over the publication of their data and that updates are available almost immediately. Secondly, it should not be necessary to migrate collection information to a new database system to become compliant with the respective network architecture so that database holders can stay with their existing systems.

The most prominent biodiversity data network GBIF (Global Biodiversity Information Facility) is supporting both BioCASe and DiGIR protocol which means that any collection database using one of the respective software packages and registering the installation is accessible through the GBIF. By autumn 2006 about 180 providers were registered and about 100 million collection units were accessible through the GBIF portal (http://www.gbif.net).

Protocol and data specification

Biodiversity information networks and data networks in general rely on data providers being understood by all data consumers (e.g. portals or individual applications) and *vice versa.* With a growing number of different systems, query languages, and response structures participation becomes more and more difficult because communication software has to be implemented and installed for every consumer – provider pair. Figure 1 shows the worst case for this approach each circle depicting a software installation for an individual consumer – provider agreement.

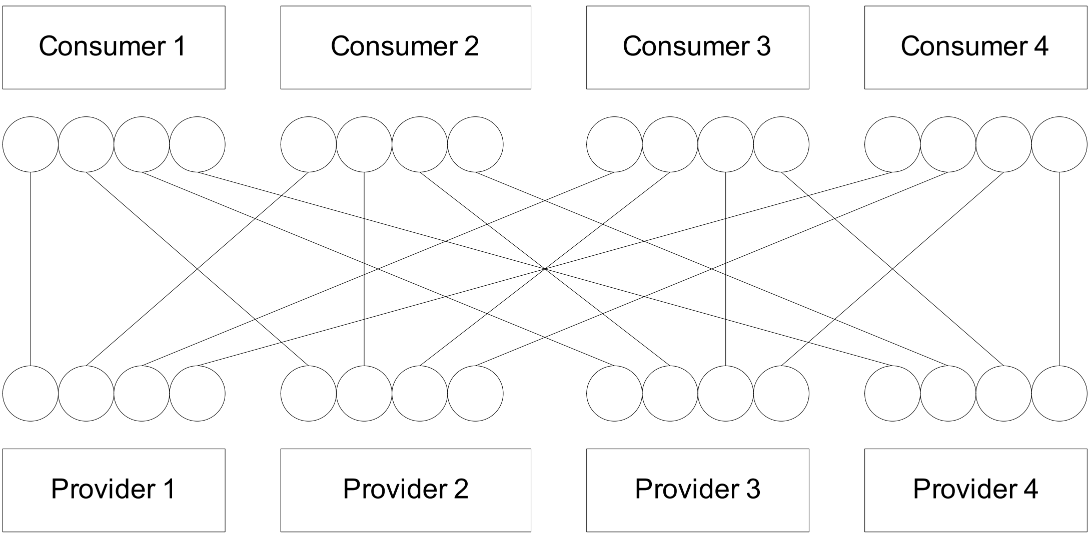


Fig. 1: data network with individual agreements between providers and consumers.

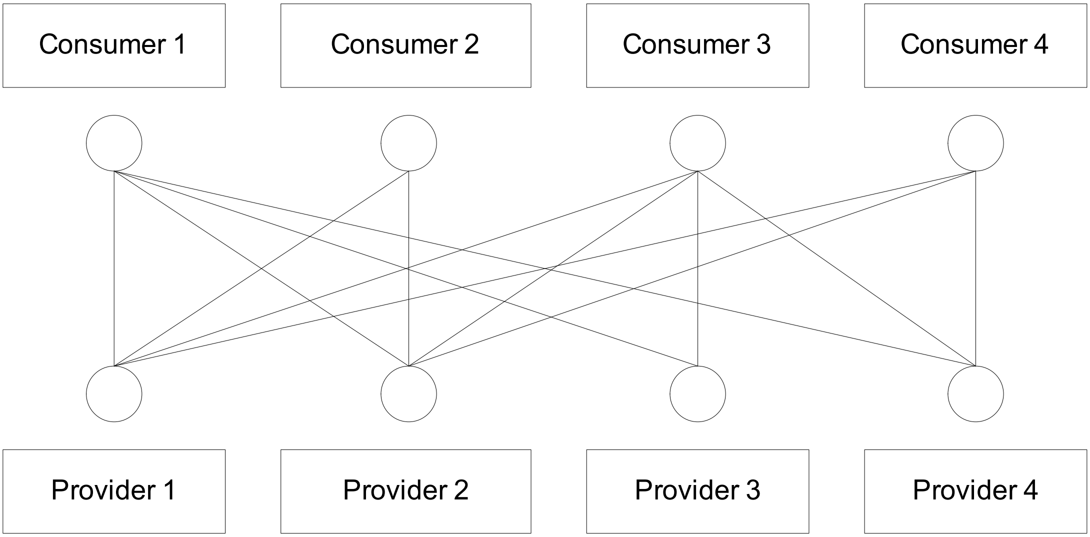


Fig 2: data network based on a single agreement between all participants.

Obviously, such a data network can be greatly simplified if all participants agree on a common query and response system so that each participant has to implement only one software component establishing the information flow for all other relevant provider or consumer nodes (Fig. 2).

Such an agreement has to be achieved at two levels, the protocol level and the data level. The network protocol defines the structure of queries and responses within the network whereas the data specification defines the terms to be used and their meaning so that, for example, a data provider knows that when asked for a *FullName* the full scientific name including an author string has to be returned. The BioCASE project provided accurate definitions for both levels based on XML, the extensible markup language.

The BioCASe protocol (http://www.biocase.org/dev/protocol/index.shtml) is a sound specification for both query messages sent by consumers and responses sent by providers of a network. In its current version 1.3 three basic query types are defined:

- **Capabilities:** returns the set of data elements a network participant is capable to provide.
- **Scan:** returns the set of distinct values for a given element (similar to an SQL distinct search).
- **Search:** returns matching records for a given search pattern (similar to an SQL select search).

Responses are returned as XML documents as specified with the data definition so that the network component receiving the response can rely on the structure when parsing and processing results of a query. In contrast to other protocols, the BioCASe protocol is capable of handling nested response documents with repeated elements so that for example multiple identifications for a single collection unit can easily be processed.

Although the BioCASe protocol can potentially cope with any data element definition, network participants have to agree on one or more common definitions. For exchange of biological collection data a joint CODATA (http://www.codata.org) and TDWG (http://www.tdwg.org) initiative with support from GBIF and BioCASE has developed a comprehensive XML schema (http://www.w3.org/XML/Schema) for Access to Biological Collection Data (ABCD). ABCD provides a common definition for content data from living collections (e.g. zoological and botanical gardens), natural history collections (e.g. herbaria), and observation datasets (e.g. from floristic or faunistic mapping). It also offers detailed treatment of provider rights, IPR, and copyright statements. In many cases, it defines elements for both highly atomized and less structured data to encourage potential providers to take part in information networks even if their collection databases are less atomized or not normalized. Where possible ABCD incorporates semantically identical elements in existing standards for collection data such as HISPID (Croft 1992), BioCISE model (Berendsohn 1999), and the ENHSIN and Darwin Core element sets.

With the flexibility of the protocol and the comprehensiveness of the data definition schema BioCASe protocol and ABCD together form a solid basis for the implementation of biological collection data networks of any kind.

Linking collection databases with standard software

With accepted standards for both protocol and content specification available, data providers have to implement only a single software module to link their collection database to international biodiversity networks such as GBIF and BioCASE. However, the majority of collection database holders lack the resources to program such a module themselves. In any case, it is much more efficient to provide a generic software system for this purpose which only has to be configured for the specific parameters of data providers.

The BioCASE project has developed a generic and flexible tool for linking up data providers to BioCASe compliant information networks (Döring & Güntsch 2003). The PyWrapper (http://www.biocase.org/provider/default.shtml) is a CGI script working on almost any operating system platform (e.g. Windows, Linux, MacOS). Prerequisites on the provider's side are a collection database holding the data, a publicly accessible web server (e.g. IIS or Apache), and the installation of the open-source CGI software Python and the PyWrapper on that server. The PyWrapper includes database modules for most of the relevant database management systems such as PostgreSQL, MySQL, Oracle, and Microsoft® SQL-Server and also allows connecting to smaller systems such as Microsoft® Access and even spreadsheet based systems using Microsoft® Excel.

To set up the system no programming is necessary but three things that are specific to the local collection database have to be configured:

- **Database connection:** the PyWrapper needs to "know" which database is to be linked and how it can be accessed. For this, the name of the database (or data source) as well as a valid username/password combination has to be entered in a configuration file.
- **Data structure:** the individual data structure of the collection database has to be declared to the system by listing the relevant tables holding the data which are to be published, their primary keys, and foreign keys to other tables. Additionally the primary key belonging to the representation of a collection unit (or observation) in the database has to be declared to enable the system to perform counts on collection units which in turn enables client software to page through huge sets of returned records.
- **Element mapping:** database attributes ("fields") have to be mapped to elements of the content schema (ABCD) so that the system knows that for example an attribute *LatinName* in the name table of the provider's local database corresponds to the *FullScientificNameString* attribute in the ABCD schema. With this step the local element set and its semantics is mapped to a semantics which is internationally understood and which can be processed automatically by Internet portals and scientific applications.

All three configuration steps can be carried out without implementing a single line of programming code just by modifying configuration files accompanying the BioCASe software. Experiences with linking up databases have shown that the time needed for configuring usually ranges from two hours to one day mainly depending on the complexity of the underlying collection database data structure. Often, providers simplify the process by migrating the data they wish to make accessible cyclically to a simplified data structure, which is then easier to map. This may also improve system performance and security.

To make configuration even simpler a graphical User interface has been developed to facilitate the setup process for database holders (Fig. 3). The tool allows establishing the database link, declaring the internal underlying data structure, and mapping database attributes to ABCD schema elements by picking from a list ordered by relevance according to pre-selected collection categories. The tool can be configured to work with content schemas other then ABCD.

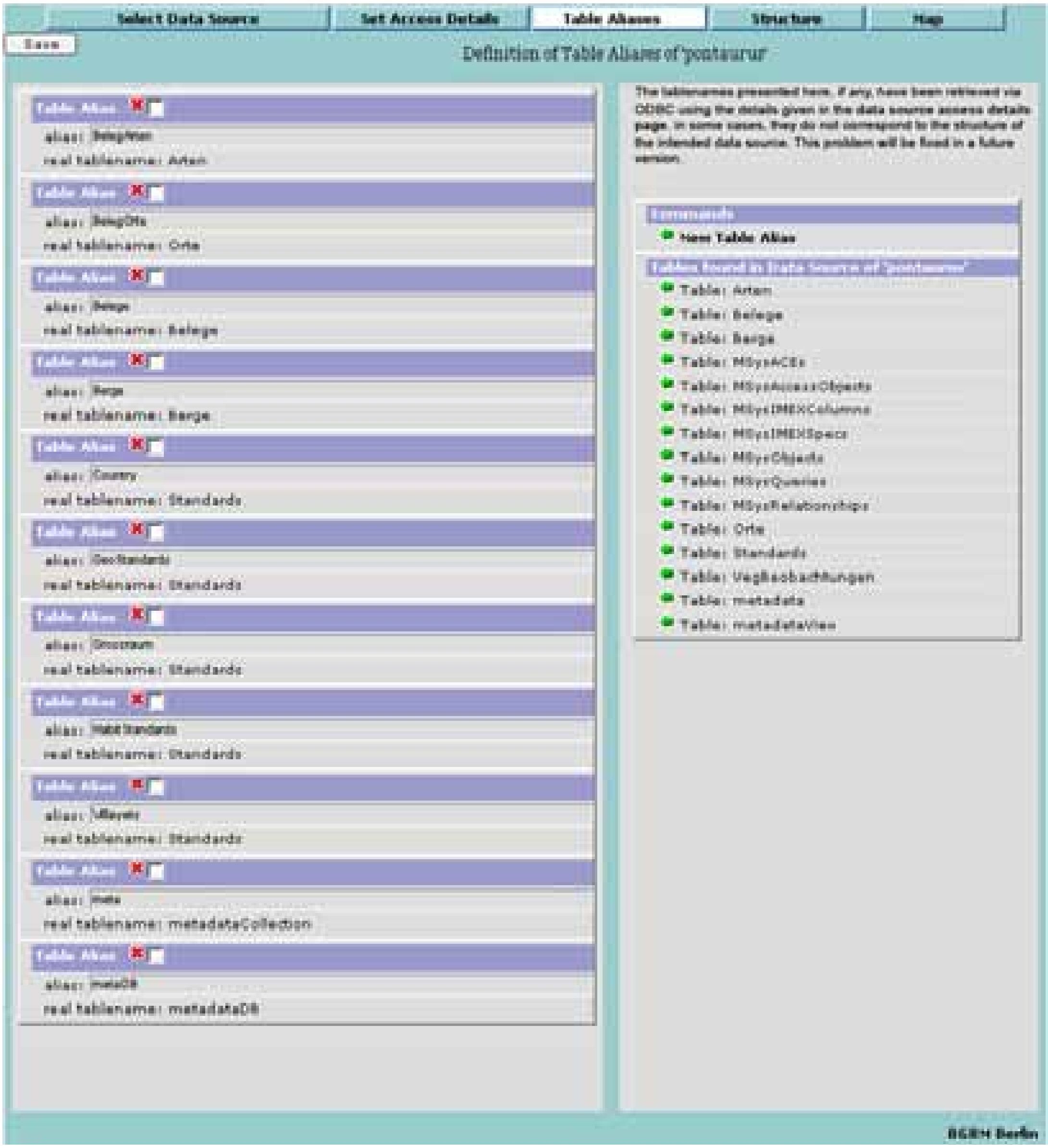


Fig. 3: screenshot of the BioCASe configuration tool.

Outlook

With the availability of standardized and internationally accepted protocols and content schemas for biodiversity data as well as appropriate and stable software for data providers it is for the first time possible to build information systems such as data portals processing all primary data presently available. This situation will generate a new generation of applications for example for the prediction of species distributions based on millions of observations and collection records distributed all over the world. In turn these applications will convince collection holders not yet linked or even not yet computerized to put more effort into the digitization and availability of their collection.

References

Berendsohn W.G., Anagnostopoulos A., Hagedorn G., Jakupovic J., Nimis P.L., Valdés B., Güntsch A., Pankhurst R.J. & White R.J. 1999. - A comprehensive reference model for biological collections. Taxon 48: 511-562. Retrieved October 2006 from http://www.bgbm.org/biodivinf/docs/CollectionModel/

Berendsohn W. G. 2003. - ENHSIN in the context of the evolving global biological collection information system. In: Scoble, M. (ed.): ENHSIN, the European Natural History Specimen Information Network. The Natural History Museum, London.

Croft J. R. (ed.) 1989. - HISPID - Herbarium Information Standards and Protocols for Interchange of Data [version 1]. Australian National Botanic Gardens, Canberra.

Döring M. & Güntsch A. 2003. - Technical introduction to the BioCASE software modules. 19th annual meeting of the Taxonomic Databases Working Group (TDWG 2003), Oeiras, Lisbon 2003, Abstract.

Güntsch A. 2003. - The ENHSIN Pilot Network, in Scoble, M. (ed), ENHSIN - The European Natural History Specimen Information Network. The Natural History Museum, London.

Linking to GBIF and other international systems

Larry Speers
Global Biodiversity Information Facility (GBIF)
Universitetsparken 15
DK-100 Copenhagen
lspeers@gbif.org

Keywords: GBIF, international biodiversity information facility, databases

Presentation abstract

At the regional, national and international levels, observational data is increasingly being relied on as the basis for monitoring biodiversity change. Unfortunately, the vast majority of this critical data is under-utilized because potential users are not aware of its existence or the data cannot be easily accessed. The value of quality primary species occurrence data is in its use and any data that potential users are not aware of or cannot access is of little or no value. Prior to the development of electronic information management systems, paper based processes made it extremely difficult and expensive to manage, share and collate primary observational data. However, advancements in information management technologies now allow us to efficiently collect data for one purpose and through sharing it across the internet, re-utilize it many times, often for purposes other than those for which it was originally collected (Fig. 1).

It is clear that the current barriers to the exchange and re-utilization of this critical data are not technological but are the consequences of fundamental funding policies and organizational design structures that instead of encouraging or supporting data sharing often reinforce processes that actively discourage these activities. Fortunately, many progressive funding agencies and other government programs are now recognizing the cost benefits and public good aspects of increased availability and sharing of biodiversity data

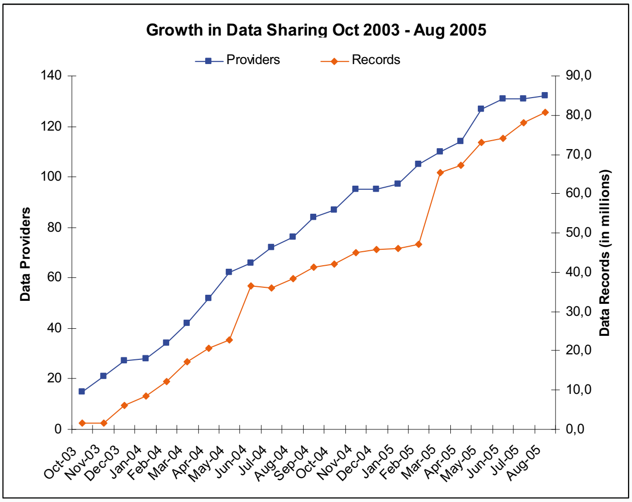


Fig. 1: Growth rate of GBIF data sharing.

and are building into their funding strategies, policies that support and encourage the sharing and open access to biodiversity data. The Global Biodiversity Information Facility (GBIF) provides a fundamental infrastructure for the global indexing and exchange of the kind of structured primary species occurrence records that the Recorder 2002 software is designed to manage (Fig. 2). In this presentation, the benefits of sharing data through the GBIF network were discussed and the process through which this is accomplished was explained (Fig. 3)

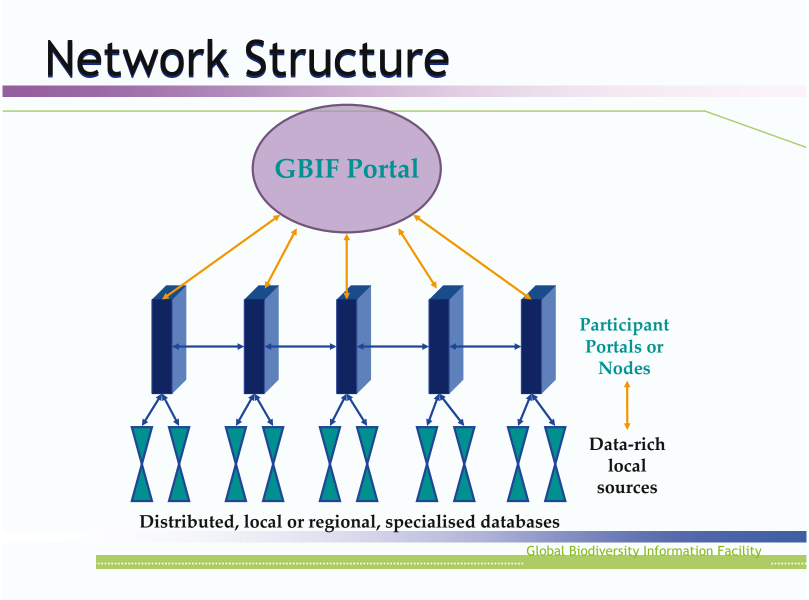


Fig. 2: A schema of the network structure of the GBIF portal.

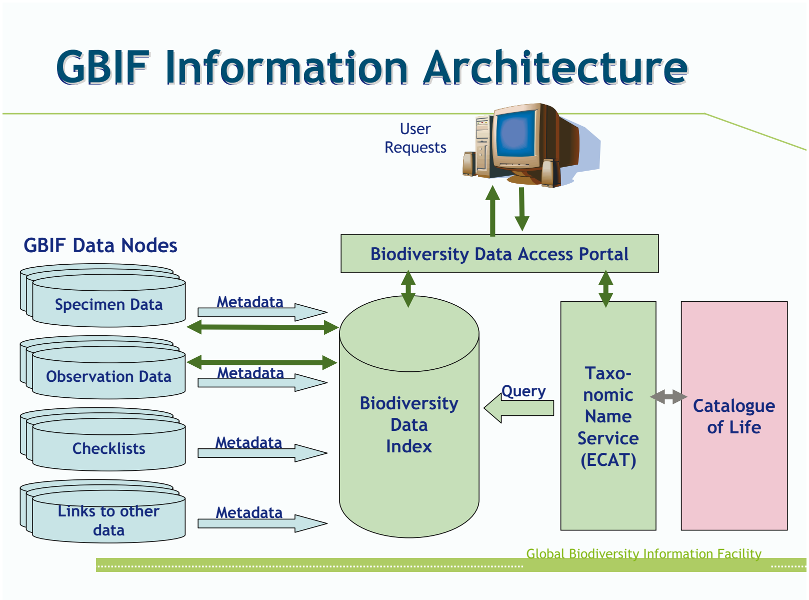


Fig. 3: GBIF information architecture.

Recorder Internationalisation - Technical Considerations

John van Breda
Dorset Software Ltd
Corbiere House, New Quay Rd
GB-BH15 4AF Poole
john.vanbreda@dorsetsoftware.com

Keywords: Internationalisation, Localisation, Languages, Spatial Reference Systems, Checklist Management, Unicode

Abstract
Enabling Recorder to be used in languages and locales other than English provides many benefits through increased user community size, shared data and reduced overall costs. There are 2 steps involved, internationalisation (preparation of the application for translation) and localisation (translation implementation). As well as translation, the issues considered include taxonomic and other checklists as well as mapping and spatial reference systems.

Introduction

Allowing Recorder to be used in countries other than the UK has many benefits. Overall development costs are reduced as each country benefits from developments implemented for others. The greater user community results in a better support network. Data sharing opportunities, including the ability to share taxonomic checklists where appropriate, reduce the maintenance burden of managing a biological recording product.

Internationalisation

The process of preparing an application for use in more than one locale is termed internationalisation. This process involves the following general steps:

- Ensuring that all application language texts are held in a form that can be swapped for locally translated ones as required. In practice, for a Windows application this typically involves utilising a resource file to hold the language specific data, which can be swapped for translated versions.
- Ensure that all date, time and number handling code is written to correctly use separators defined by the system locale. For example, code is written to separate dates with a '/', '-' or any other character defined as the preferred separator by the system. Internationalised code does not rely on any one of these characters.
- Implementation of a suitable tool to allow the language resource file to be updated and translated into each required language.

In addition, the Recorder project has the following internationalisation requirements:

- Enabling a local base map file to be used.
- Enabling a local spatial reference system to be used for data entry and mapping.
- Enabling local spatial grid requirements to be fulfilled. For example, 10km square distribution maps are used in the UK, whereas the equivalent grid cell in Germany is 10 by 6 minutes on the Bessel projection.
- Enable local species lists and other controlled terminologies to be defined.
- Preparation of a locale specific installation CD.

Although well designed modern applications are often internationalised from the outset, Recorder development started in 1999 when internationalisation techniques were less well defined. Therefore code written during the initial development does not separate language text into a resource file. Recent code changes, however, do represent language text in a form that is suitable for translation.

The grid based mapping within the current version of Recorder is dependent on distribution maps being drawn over 10km squares although the planned implementation of Recorder in Germany includes a modification to support other grid systems including the German 10 * 6 minute grid. This project also includes the integration of an internationalisation and translation tool into the Recorder application (Korzh n. d.).

Other aspects of internationalisation of the Recorder product are already implemented and have been used in Luxembourg (Fig. 1).

Localisation

Once internationalised, an application is prepared for use in a specific locality through the process of localisation. This involves the following general task:

- Provision of translated versions of all snippets of language text.

In addition, the Recorder product requires that the following tasks are undertaken in order to prepare the product for use in a locale:

- Provision of local base map files (for example, an outline of the country or region). Recorder is able to import widely available base map outlines from ESRI Shape file format (*.shp).

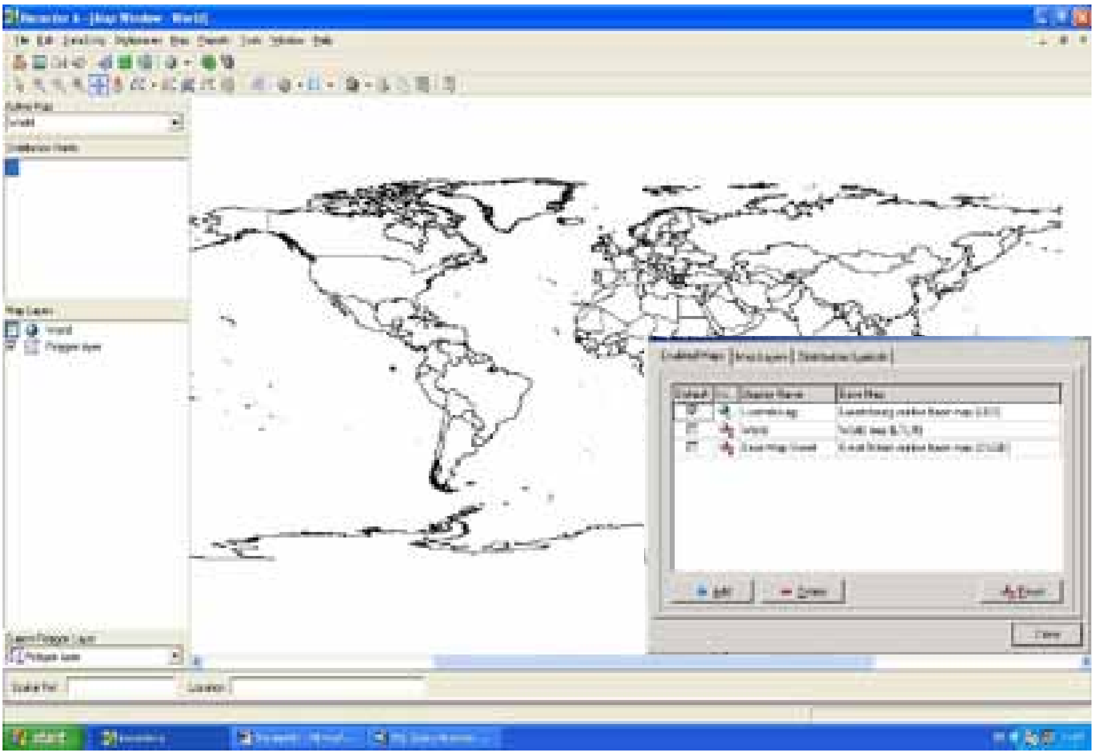


Fig. 1: The map option window in Recorder 6 allows to chose among different base maps.

- Preparation of an addin that is able to convert between Latitude/Longitude (WGS84) spatial references and the local spatial reference system.
- Preparation of local species lists and other controlled terminologies, to provide a modified initial installation database for the locale. Recorder itself does not include tools suitable for large scale authoring of species lists, but there are several possible approaches to this problem. One method is to use the Collections Module addin to prepare species lists which can then be imported into Recorder's Taxon Dictionary module.

Further Issues

The following internationalisation issues need to be considered:

- The user interface controls in Recorder, such as data entry boxes and labels, do not support Unicode. This means that characters used in international strings are limited to those within the ASCII character set. It is possible to implement Unicode versions of controls for Recorder though the development effort required to do this is significant as all controls need to be replaced by alternate versions.
- When species checklists and other controlled terminologies are created in different countries, it may not be possible to exchange observational data between countries without further work. For example, a dictionary entry for a species created in one locale will have a different primary key to a dictionary entry for the same species created in a different locale. Although tools are available for automated detection of possible matches between dictionary items, these tools are based on simple text matching and require considerable manual intervention to ensure accuracy.

References

Korzh n.d. - Korzh Localizer website available at http://devtools.korzh.com/localizer/. Accessed November 2006.

FloraWeb – the German web flora and the role of Recorder

Rudolf May
Bundesamt für Naturschutz
Konstantinstrasse 110
D-53179 Bonn
Rudolf.May@BfN.de

Thomas Schneider
Klinkerstr. 92
D-66663 Merzig

Keywords: floristic mapping, mapping software, floristic database, web flora

Abstract

The German Federal Agency for Nature Conservation (BfN) maintains FloraWeb – a digital web flora of vascular plants in Germany which is freely accessible over the internet. The core of FloraWeb is a database with distribution data from floristic surveys carried out from 1950 until now, containing about 15 million records of plant observations. Most of this data has been entered using a software program developed and freely distributed from 1990 – 2000 called FlorEin. Since this DOS-based software has not been modernized, the British Recorder software has been chosen to serve as a functional replacement. In a pilot phase basic adaptations of the software are funded by BfN to provide a first functional German version.

Introduction

When we draw a timeline of milestones (Fig. 1), it can be shown that floristic mapping in Germany has always involved the use of information technology available at the time. In the late seventies until 1986 mainframe computing has been used for data capture and the plotting of distribution maps. The goal and the obtained result was the publication of the first German distribution atlas (Haeupler & Schönfelder 1988). In 1988 the data was transferred from mainframe files to an oracle database (version 5) at the Federal Research Centre for Nature Conservation and Landscape Management (BFANL), which in 1993 was reorganized to give rise to the German Federal Agency for Nature Conservation (BfN). From 1989 to 1997 a floristic mapping centre was funded by the Federal Ministry for the Environment, Nature Conservation and Nuclear Safety (BMU) with BFANL/BFN as project executing organization. The main tasks of the floristic mapping centre were the support of data capture and collation from numerous regional mapping projects and the establishing of a data flow from the regional projects into the central German floristic database at the BfN (Bergmeier 1992). Fig. 2 gives a schema of the organisation structure and the data flow that was established from the beginning of the 1990ties. The success of this project can be demonstrated looking at the increase of database records from 2 to 10 million records from 1990 to 1997. It is largely owed to the development and free distribution of a software program that was called FlorEin (Fig. 3), which in conjunction with the rapid spread of personal computing enabled the local floristic researchers as well as the regional floristic mapping projects to set up and maintain their own floristic databases. The main funding period ended 1997, although some data harmonization and correction processes had been funded until the end of 2000. As a major result of the ten years of data collection, a digital flora of Germany, FloraWeb, has been built up in the past 5 years which is freely acces-

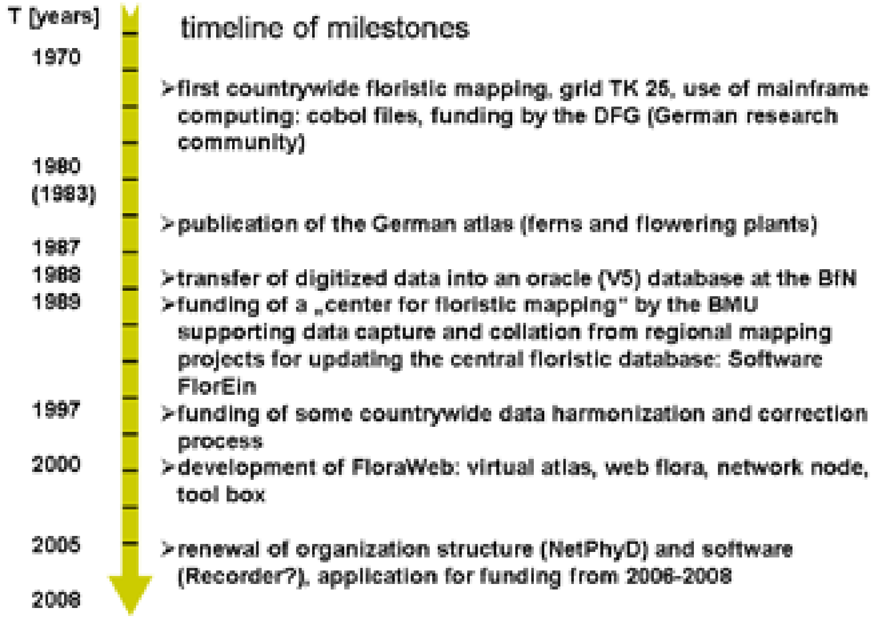


Fig. 1: timeline of milestones in floristic mapping and the use of information technology in Germany

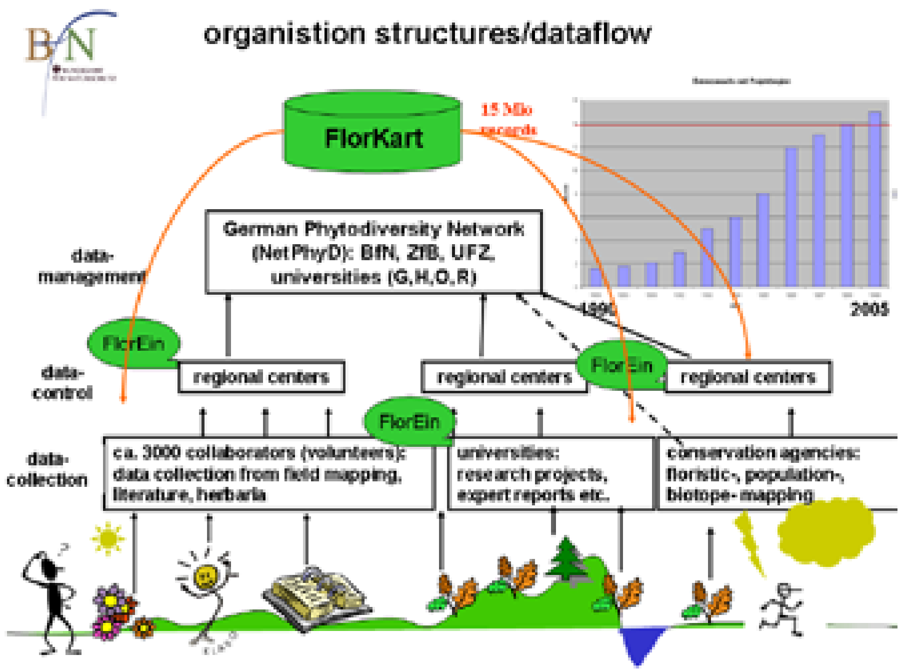


Fig. 2: Schema of data flow and organisational structures of floristic mapping in Germany in the 1990ies.

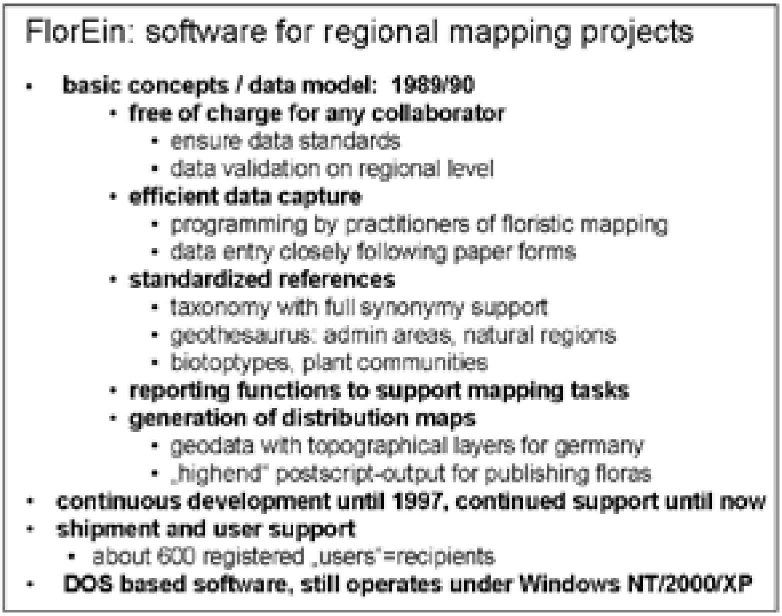


Fig. 3: Some aspects of the FlorEin software as a tool for data capture und data flow of floristic mapping data

sible over the internet (http://www.floraweb.de). The German floristic database with about 15 million records of species occurrences of vascular plants in Germany is the core dataset on which this information system is built.

Components of FloraWeb

FloraWeb is a web-based information system on German vascular plants and vegetation types. In the course of setting up the database on the distribution of plant species as originally planned, additional information related to biological, ecological and plant functional traits was integrated through projects the BfN was cooperating with and/or partially funding. Another area of additional information are the red list data plus legal status information that the BfN is responsible for. A long term project that was coordinated by the BfN was the development of a European map of vegetation types. The German part of this map was used for displaying the potentially natural vegetation of Germany in an interactive web mapping application. Fig 4 gives an overview of different sources of information and data that were used to build FloraWeb.

FloraWeb incorporates different functions: it serves as a data warehouse, dynamically produces web pages with information on plants and vegetation and provides evaluation tools for interactively querying data and displaying results in web pages and web-mapping tools. Some general features are:

- it is based on existing and published, but not always easily accessible or even combinable data
- this data is made publicly available via web pages
- the content is dynamically produced from databases
- the previous characteristics allow the data to be accessed via wrapper interfaces by information brokers like GBIF (Global Biodiversity Information Facility, http://www.gbif.org) BIG (Federal Information System on Plant Genetic Ressources, http://www.big-flora.de) or PortalU (German Environmental Information Portal, http://www.portalu.de).
- we aim at a broad spectrum of users, such as scientists, specialists in governmental or non governmental organizations, teachers, students

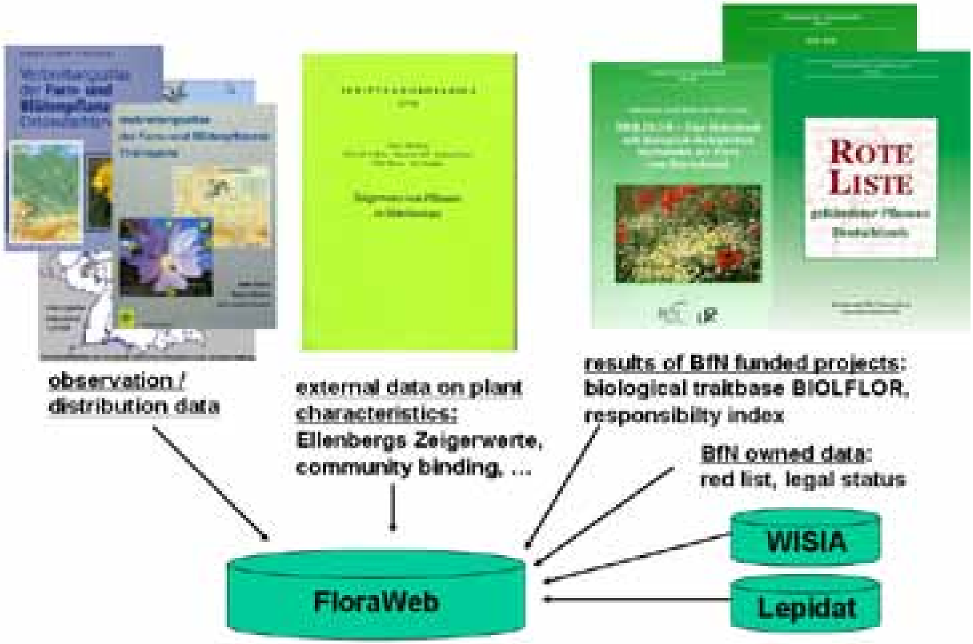


Fig. 4: Sources of information and data used to build FloraWeb.

and last but not least by the interested public.

For example in FloraWeb we have fact sheets for each of about 3500 vascular plant species with a thematic compilation of the underlying data from various sources. We generally try to reduce the number of technical terms and provide explanations in popup windows wherever they are essential. Of course at least one image of high photographic quality is provided.

The species fact sheets can be accessed via a text search for scientific or vernacular names , by an alphabetic index or by queries on free combinations of plant characteristics like traits, red list values or family names. For example a user can produce a query like 'give me all species which are endangered or critically endangered within the family of the orchids, which grow on calcareous substrates and have an atlantic distribution range" by selecting the appropriate combination of characters from list boxes. The results of these queries are displayed in lists, where the scientific names link to the complete fact sheet for each species.

The most powerful evaluation tool in FloraWeb is FloraMap, a web mapping tool which can be used for displaying the distribution of single species or for any group of species that is fed into the tool. The map content is dynamically generated from the distribution database at runtime, so the maps are always as up-to-date as the database is. Distribution maps in FloraWeb and FloraMap are grid maps based on the German grid system, where each grid cell is 10 x 6 minutes in latitude/longitude. For species groups, which can be fed into the system from the results of the query described above, or accumulated by name queries within the tool, a map is produced with classified numbers of species records for each grid cell. By this means the combination of species with specific traits can be used to produce maps showing for instance the occurrence of warm, open grassland on calcareous grounds or the areas where we have remainders of bogs in Germany (e.g. see Fig. 5).

Besides displaying distribution data, FloraMap can also be used for data entry of new plant observations. When zooming in, we provide map layers up to a scale of 1 : 25000 with reduced resolution images of the German ordnance survey map

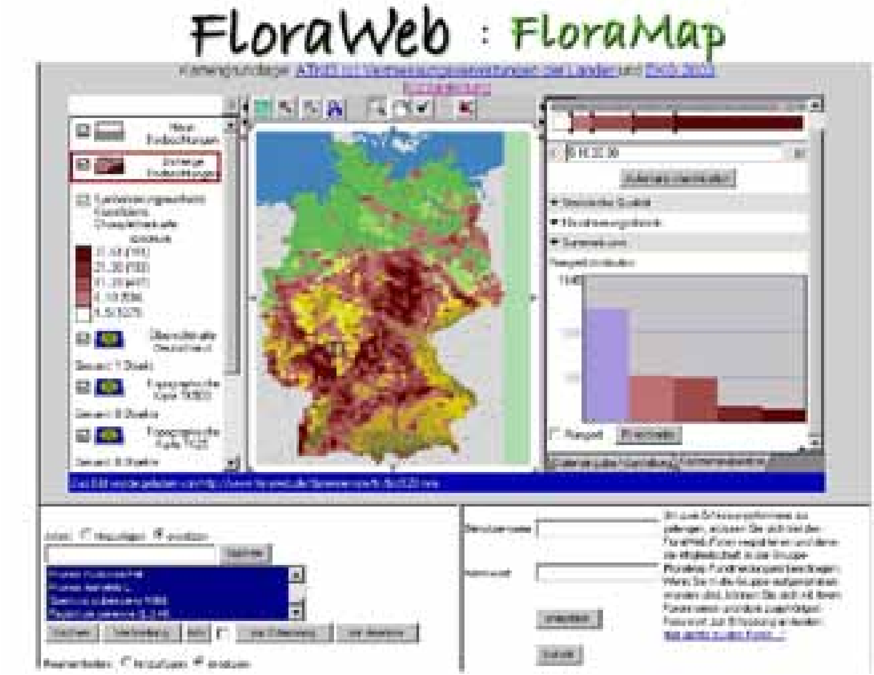


Fig. 5: Coincidence maps for species groups produced by FloraMap.

sheets. On these maps new observations can be marked as coordinate points or grid cells, and all relevant and obligatory information for capturing a new observation record can be entered into a form provided in another html frame. This concerns the *Where* (coordinates or grid cell ids from the map), the *Who* (the user that has logged into), *When* (a date that must be entered), and *What* (the species that have been selected from the reference list). Additional information on the naturalisation status and population size can also be entered (see Fig. 6).

With this tool we provide means for

a) quickly entering new observations that may have been made casually, like on the Sundays walk, or on a field excursion,

b) collecting new records for specific observation programs, like in an early warning system for neophytes.

It is not meant and will not fulfill the needs for regional mapping projects like doing complete floristic surveys, which the old software FlorEin has been used for.

Role of floristic mapping software

As already shown above, the software FlorEin, which was developed in the 90ties and distributed freely to single volunteer or professional collaborators and coordinators of regional floristic mapping projects had a great impact on the mobilisation of floristic mapping data. The existence of this software combined with the fact that it perfectly fitted the aims and needs of floristic mapping projects, providing means of handling collected data by the collector himself and enabling the production of regional floras "with a keystroke" pushed existing projects and helped to trigger new ones.

Another effect of the widespread use of the software was the effect it had on data quality and data harmonisation and standardisation, by providing standardized references in the area for taxon names, geographical references and termlists. Data exchange and data collation was much simplified by providing common data structures implemented in dbase-files used by the Clipper compiled software.

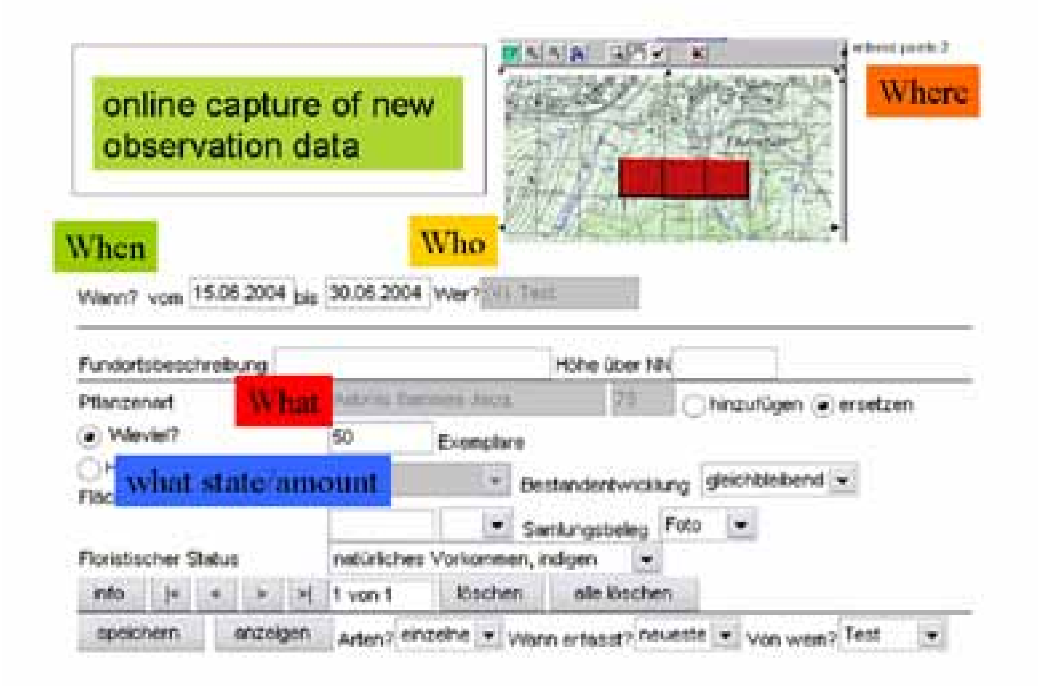


Fig. 6: Entering new observation using the FloraMap tool.

Today we still have a number of projects which use the FlorEin software, even if it's operating system DOS is out of date. Although it is operable in all current versions of the Windows operating system, it doesn't provide any features that users today would expect and are used to. In order to ensure that the data flow that has been established by providing this software will not slow down, the BfN needs to provide a functional replacement of the FlorEin software that will run on and use all features of modern Windows operating systems. Such a functional replacement would have to keep the existing level of functionalities such as efficient data capturing, standardized references, reporting functions for supporting the mapping tasks, generation of high quality distribution maps for publications and atlas productions. Demand for new or enhanced functionalities has been issued in the areas of export/import, GIS-support, handling of taxonomic concepts and management of collection data. Redesigning FlorEin would in fact have meant to start from scratch, allowing for an adapted workflow and the integration of new possibilities provided by a graphical, multitasking user interface. A survey of existing modern software solutions for species recording led to selecting the English Recorder software as the optimal choice for the task.

Some major points that triggered this decision were:

- existence of a highly sophisticated, elaborated and well documented data model, which is suitable for all groups of organisms,
- the software is fully developed and well approved,
- there is a large number of users, an open dialogue and support forums,
- the software has implemented comprehensive techniques of data exchange, caring for transport of metadata and intellectual property rights,
- an open interface for the development of add-ins based on active-x components is implemented and well documented, so that the software is open for independent customization needs,

- many of the new or enhanced functionalities that FlorEin was lacking are already implemented in Recorder.

The initiative for providing a new recording software for the German floristic mapping projects luckily coincides with the initiative of the Joint Nature Conservation Committee in Great Britain to look for partners in further supporting and maintaining the Recorder software and the Luxembourg initiative of adding a collection management module as an add-in, which is donated to the public domain. In 2006 a pilot project is being funded by the BfN to make the first necessary adaptations for providing a German version of the Recorder software. These first adaptations include the translation of the user interfaces and the implementation of a spatial reference add-in which supports all German base map projections. Of course all specific taxonomic and geographical reference lists as well as term list will be added or replaced. When the basic adaptations of the pilot phase have proven successful, further add-ins could provide specific functionalities like rapid data entry in FlorEin style. A main project phase for implementing such features, but also for supporting the roll out of the software by providing a help desk and giving support transferring existing data sets into the Recorder system is planned for 2007 until 2008. By then we hope that Recorder-*D*, the German version of Recorder, will have become a new standard for floristic mapping software in Germany.

References

Bergmeier E. (ed.) 1992. - Grundlagen und Methoden floristischer Kartierungen in Deutschland. Floristische Rundbriefe, Beiheft 2, Göttingen, 146 p.

BIG n.d. - Federal Information System on Plant Genetic Ressources, website available at http://www.big-flora.de. Accessed December 2006.

Floraweb n.d. – Floraweb website available at http://www.floraweb.de. Accessed December 2006.

Gbif n.d. - Global Biodiversity Information Facility. Website available at http://www.gbif.org. Accessed December 2006.

Haeupler H. & Schönfelder P. 1988. - Atlas der Farn- und Blütenpflanzen der Bundesrepublik Deutschland, Stuttgart, 768 p.

May R. 1990. - Die Datenbank der Florenkartierung: Weiterentwicklung der Datenstrukturen – Einbindung von PC's in ein integriertes EDV-Konzept. Floristische Rundbriefe 23(2): 160-169.

May R. 1994a. - Datenbank FLORKART: Entwicklungsstand, Import- und Ausgabefunktionen. Floristische Rundbriefe 28(2): 84-94.

May R. 1994b. - Die Datenbank der Floristischen Kartierung: Ein Beispiel für die dezentrale Erhebung und zentrale Zusammenführung von raumbezogenen, naturschutzrelevanten Informationen: 155-157 In: Kremers, H. 1994, Praxis der Umweltinformatik. Band 5 Umweltdatenbanken, Marburg.

May R. & Subal W. 1992. - Einsatz von EDV für Kartierungsprojekte. Floristische Rundbriefe, Beiheft 2: 87-93.

PortalU n.d. - German Environmental Information Portal. Website available at http://www.portalu.de. Accessed December 2006.

First International Recorder Conference - Recorder now and in the future

Adrian Rissoné
Department of Palaeontology, Natural History Museum
GB-SW7 5BD London
A.Rissone@nhm.ac.uk

Keywords: Recorder, field observations, natural history collections, species occurrence, recording biodiversity, internationalisation

Abstract

Since 1992 biodiversity and conservation have gained a high profile in decision and policy making throughout the world. The consequential need for the acquisition and dissemination of biodiversity data has encouraged the development of a variety of software solutions for data capture, management and exchange. Integration of field observations, species check lists and natural history collections data has proved challenging but products and standards are slowly emerging.

Since its origins as a bespoke working tool for recording biological field observations in the United Kingdom the Recorder product has developed into a stable, versatile system with a growing user base in Europe. The current version (version 6.9.3, December 2006) is compatible with Microsoft SQL Server 2005 allowing theoretically unlimited capacity (up to 20 million records per server). Moving (in Recorder 6) from MS Access to SQL Server solved many size, stability and other problems. During development of Recorder 6 numerous minor enhancements have been introduced, funded by the existing user community.

Recorder has recently been extended by the Musée national d'histoire naturelle de Luxembourg to incorporate natural history collections data handling, making it a complete system for the recording of field-based information. The Luxembourg Collections Module is a major development – probably almost as big as the core product – and for the first time allows Recorder to manage Earth Science data – stratigraphy, rocks, palaeontology. An important part of the development is the implementation of a hierarchical thesaurus capable of, for example, handling Common Names in multiple languages.

The First International Recorder Conference, hosted by the Musée, introduced Recorder to a wider European and international audience and addressed the questions of future development, web enablement and internationalisation.

Introduction

The author, invited by the Conference organisers to chair the meeting as an independent observer, was able to see the historical development of the Recorder product from an unbiased perspective. It was immediately clear that interest in field recording software within Europe is increasing as the need to record species occurrence and biodiversity intensifies. Recorder is one of several software solutions to biological field data recording to be developed world-wide over the past few years. Each product has originated in a particular project or organisation and has, inevitably, been designed to satisfy local requirements. Unusually, Recorder has proved to be both fit for its original purpose and also extensible. The two factors (community interest and extensibility) combined led to the need for an international conference to showcase the current product, identify future requirements and influence a development path.

The major part of the 2-day conference was devoted to discussion and demonstration of the way Recorder has developed so far and been used throughout its user base. The majority of the presentations are included in this volume, representing a fascinating and unusually complete history of the development and practical application of a software system. Throughout the Conference the emphasis was on integration and management of data, at several levels and in various working environments, and the dissemination of those data throughout Europe. Examples of working systems were demonstrated and current development work was described. Day 1 included presentations on the origins, technologies and standards underlying Recorder, recent developments in the area of collections data management, the integration of term lists, dictionaries and taxonomies, and the provision of a thesaurus extending the capabilities of Recorder to allow recording of unlimited types of biodiversity observations and specimens. The morning of Day 2 introduced Internet applications for data capture, web presentation and data exchange, including a presentation of the Luxembourg integration of field data and collections data through a web portal. During the first part of the afternoon the potential of internationalisation of Recorder was explored.

Discussion

As anticipated, the widely differing needs of the community prompted more questions than answers and the meeting progressed naturally to a summary session where the primary future requirements and direction of the product were debated (fig. 1). These were broken down into the following topics:

- Functionality (Recorder "Classic")

The majority of users represented at the conference were of the opinion that the core product includes most of the functionality they require. A notable exception would seem to be Geographic Information System (GIS) integration. With any product of this type the user interface is critical and the developers recognise that ongoing development is required. Such development must, however, be in the context of Recorder being a crucial field tool for the product "owner" - the Joint Nature Conservation Committee (JNCC) of the United Kingdom. The subjects of ownership, and the consequential constraints on development, are discussed elsewhere in this paper. Less controversial would be improvements to the uploading and downloading of species lists to facilitate targeted usage by various users. There was some discussion on the extension of Recorder to other domains (e.g. Earth Sciences).

During the period following the Conference international interest in the product has continued to grow. Recorder is newly in use in the Falkland Islands and the National Parks and Wildlife Service in Eire, and in the Channel Islands (Jersey) with an add-in developed to support Jersey Transverse Mercator.

- Functionality (Recorder Web)

A commonly expressed limitation of Recorder is the lack of a web interface. Delegates were reminded that Recorder started as a working tool with specific aims. Recorder Web shares the same data structure but has a different set of tools and benefits, and its cost model is different. At the moment it has only been designed and prototyped, under the direction of the Luxembourg Natural History Museum. Although interest in the web version is growing there remains uncertainty as to the relationship between "classic" and "web" versions, and its role in the broader picture.

The Recorder Web Toolkit may be good candidate for open-source development (particularly given its current stage of development). Due to the high cost of a full Microsoft environment solution, a more realistic approach could be an Open Source solution allowing record centres and recording schemes to support online data entry at very low cost. Without a champion, development of a web product is likely to progress slowly, given the already demanding roadmap for the existing Recorder product.

Since the close of the Conference the Musée national d'histoire naturelle de Luxembourg has issued a Call for Partners as part of a bid for support for development from eLuxembourg, a funding body for web applications in the public sector (http://www.mnhn.lu/recherche/recorder_web.asp).

- Internationalisation (Recorder*International)*

There was lively discussion of the possibility of "international" versions of Recorder. The difference between "internationalisation" and simple multi-lingual versioning (which can be achieved, albeit with some message translation problems) was emphasised. There was a strong spirit of co-operation but many delegates could foresee difficulties of scale and essential customisation, especially geographic mapping differences and local terminology. It was thought

Fig. 1: A panel of representatives of users and developers of Recorder guided the summary session where the primary future requirements and direction of the product were debated. From left to right: Guy Colling (MNHNL, Luxembourg), John van Breda (Dorset software, Poole), Charles Copp (EIM, Clevedon), Steve Wilkinson (JNCC, Peterborough), Charles Hussey, (NHM, London). The chairman of the session on the right hand side was Adrian Rissone (NHM, London). Photograph by G. Marson (2005).

that large-scale funding was more likely to be obtained for a multi-partner European initiative but that it might be difficult to get commitment from national and regional institutions. One issue is defining the applicability of Recorder in varying national frameworks. Internationalisation has been emphasised by the development of the Luxembourg Collections Module – much of the new interest in Recorder is directed here, with several institutions across Europe clearly looking for an effective solution at a reasonable cost. It was felt that ideally there should be a suite of applications within a Recorder product family so as to suit differing application requirements.

The important issues of core functionality and bespoke customisation could limit and delay development of an international product. Ownership is the primary factor – the JNCC has a requirement to safeguard the core product for its original purpose in the United Kingdom and to ensure that extended functionality and bespoke customisation does not compromise its operational requirements. The primary focus for JNCC is to make the package successful in the UK, with as much uptake as possible. This requires a series of minor improvements but no additional functionality in the short term. Uptake on a wider international scale has a much lower priority for the JNCC at the moment. Taking Recorder into an Open Source development environment could possibly offer a solution. It is not yet clear whether Open Source development could help with difficult internationalisation problems such as different mapping grids in different countries.

- Ownership and Management

It can be seen from discussion of the "classic", "web" and "international" options, that ownership, control of development and long term support are the issues of most concern to delegates, and may limit the pace of development for the foreseeable future. It has already been noted that the JNCC has a strong imperative to retain control of at least the core product. Equally, the success of the Luxembourg Collections Module Extension shows how the core product can be used as the basis for valuable new applications. How such development can be supported and maintained is, as yet, unclear.

Delegates agreed that at the very least there should be an Interest Group. The majority of delegates indicated support for the concept of a Steering Group. Others proposed the formation of a not-for-profit Foundation, which would take over development of the product suite, authorise versions and approve add-ins. This would require the JNCC to accept partial or complete transfer of ownership (something it may be unable or unwilling to do), and there are other contractual and intellectual property issues. For example, some code libraries are owned by Dorset Software. The Luxembourg Natural History Museum has made a significant investment in the Collections Module – they are happy to share product but not provide support and maintain evolving product. The more formal the management of the product becomes, the more important become clear responsibilities, and mechanisms for managing costs and charging users.

There is a danger of attempting to push forward new development before being certain of the potential use of Recorder in the broader community. It will be difficult to attract new users, or public development funding, for a product that has existing commercial competitors unless it can be demonstrated to be superior, better fit for purpose, or more cost effective. It may be wise to carry out a survey of alternative products that a prospective customer might consider. Especially in the area of Collections there are alternative products already available, fully usable. What features and benefits of Recorder make it unique, or more attractive than the rest?

Attracting even a few additional partners in the short term would enable development to take place, perhaps without the need to obtain large-scale funding, but these partners may well have their own particular requirements and their own agenda for development. There should be a drive to get potential partners together to look for common ground. At the same time the Recorder community should make efforts to ensure that potential partners know more about the current product.

Who leads in taking things forward? The Recorder community needs to become much more visible. Even in the United Kingdom there has been no real champion (but the JNCC is set to address this). A proper management structure is needed to carry forward the outcome of this Conference, to discover user needs, publicise the product, advertise success by placing icons and links wherever Recorder data appears.

- Development Options

In theory there is a strong argument for development going towards Open Source, the advantages being full modularisation, lower unit costs, faster development and community-led feature and content development (via a Wiki). On the other hand, successful Open Source projects require a large technically-competent community, and redevelopment from the existing product cannot be quick or cheap (for example, the dependency on SQL Server stored procedures). Also, Recorder uses proprietary components (the mapping system) although integration of Open Source GIS would be attractive.

It may be that for the moment Open Source must remain a goal for the future and that a line should be drawn under what has been achieved so far, add-in code (at least) then becoming Open Source.

Whichever direction is taken, there must be a strong emphasis on standards during development and "best practice" during application. A Steering Group should consider the appointment of a "standards officer" to advise and monitor as required. Equally important are the encouragement of a good user group and effective training.

- Maintenance and Support

This may be categorised into:

1. Support for the core product
2. Support for existing extensions (Luxembourg Collections Module and bespoke add-ins)
3. Support for code add-in in future
4. Distribution, Installation, Training, User Support

No matter how future development is achieved there will always be an ongoing support cost. Some sort of formalised user group would seem to be essential. It was argued by some delegates that the Foundation model would provide the best option overall, using service agreements and centrally funded staff (or a commercial organisation) contracted in. It is conceivable that a Foundation could protect existing interests (both ownership and commercial interests) but provide a sustainable framework for all of the above categories within a relatively simple user service agreement. If the Foundation model is unacceptable then a Steering Group could set up a framework to which the existing stakeholders would commit, providing facilities and resources as required under a slightly different service agreement. What seems clear is that the product cannot develop long term under the existing arrangements.

- Incorporating existing and developing standards

Biodiversity Information Standards (TDWG) – formally known as the Taxonomic Databases Working Group – is an international not-for-profit group that develops standards and protocols for sharing biodiversity data. The Group has established an Interest Group to explore avenues that incorporate observation-based monitoring of biological organisms data into existing federation mechanisms of the bioinformatics community. A potential outcome of this Interest Group is to develop a rich description standard for observational and monitoring data, to be integrated with existing collections level standards. To succeed the standard must be extensible to allow the incorporation of a diversity of data types. Furthermore, the standard should incorporate variables from more general resource discovery standards (e.g. from Dublin Core, Darwin Core 2, Access to Biological Collections Data (ABCD)). The observational data standard should cover several types of observational data held within a variety of data repositories, which include: biotic surveys, inventories, time series observational data, and checklists.

The Interest Group initially proposes that the observational data standard encompass the following:

- Content standards, which provide guidance on the types of data to be managed, including the values and structure of core database fields. These include data collection and management methods.
- Data Transfer / Sharing standards, which facilitate the sharing of data among institutions, regardless of their native database.
- Metadata / Documentation / Reference standards, which provide guidance on how to describe the content, quality, condition, and other characteristics of the data.
- Process standards, which provide guidance for tracking changes over time in definitions of taxon names as well as places.
- Representation standards, which provide guidance on spatial representation of observation data (scope to be defined).

Next Steps

The Conference agreed that some sort of "community" group is required with real aims, a proper structure and resources to help manage and develop the product. Conference proposed a number of key points for action:

1. The Recorder product and the Luxembourg Collections Module extension should be publicised
2. A web site and improved forum should be established
3. An initial Working Group should be established to oversee:
   a. Creation of the web site
   b. A survey of existing and known prospective users
   c. Formation of a consortium of interested parties who could take the product forward

In due course a more formal consortium or other body may be established to consider such topics as:

1. Scope of development
2. Canvassing community interest
3. Identifying Sponsors
4. Defining specific Goals
5. Setting out a development Roadmap
6. Establishing a long-term support environment
7. Signing up new Partners
8. Submitting Funding Proposals

Guy Colling, Tania Walisch, Steve Wilkinson and John van Breda agreed to co-ordinate action in the first instance.

References

TDWG n.d. - Biodiversity Information Standards (TDWG). Website available at http://www.tdwg.org. Accessed December 2006.

TDWGwiki Observational Data 2006 - Biodiversity Information Standards (TDWG). Observational

Data Group webhome available at http://wiki.tdwg.org/twiki/bin/view/Observational/WebHome. Accessed December 2006.

Links

Recorder:

http://www.recordersoftware.org/

Recorder Forum:

http://forums.nbn.org.uk/ (at the UK National Biodiversity Network site)

Recorder Web – call for Partners:

http://www.mnhn.lu/recherche/recorder_web.asp

Recorder Web Wiki:

http://eim.metapath.org/wiki/index.php?title=Recorder_Web

Biodiversity Information Standards (TDWG):

http://www.tdwg.org

TDWG Observational Interest Group:

http://avianknowledge.net/tdwg

http://wiki.tdwg.org/twiki/bin/view/Observational/WebHome

http://lists.tdwg.org/mailman/listinfo/tdwg-obs_lists.tdwg.org

Recorder 6 as a database for archaeological purposes

Frank Broniewski
Softwaredienstleistungen
Cloefstraße 76a
D-66693 Mettlach
frank@geo-dienstleistung.de.

Keywords: Archaeology, Luxembourg, Recorder, Database

Short Abstract

For its scientific research the EPC Project is looking for a database to store archaeological and cultural heritage information. As biological and archaeological records share a certain number of elements, Recorder seems to be a ready to use database. The article discusses the knowledge gained from testing Recorder with archaeological data.

Recorder 6 for archaeological data collection

Within the project „Space and Cultural Heritage" (EPC, „Espace et Patrimoine Culturel"), financed by the Luxembourg Research Fund (FNR „Fond National de la Recherche"), the Museum of History and Art , Luxembourg (MHNA, „*Musée National d' Histoire et d' Art*") is looking for a database to store archaeological and cultural heritage information encountered in Luxembourg. The Prehistory Department of the MNHA, who launched the EPC – project in 2003, heard through its partner Museum of Natural History (MNHN) about Recorder *6* in 2005. Several meetings with the MNHN Recorder team and Charles Copp provided the necessary background for initial tests of the Recorder *6* database.

When considering if a database for biological records is also capable of handling archaeological data, it is important to compare the nature or the structure of the data gathered. There are several sources of archeological information. The main sources are field surveys and excavations, other sources are literature references, museum collections and private collections. The database should also allow the storage of information on the different actions related to an archaeological site, e.g. different interactions on the site, owner information as well as lots of other information arising in daily administrative work. The database system should certainly handle multimedia content, names and addresses of people or literature references. Although these types of information are important, the weight of this article lies on the comparison of the structure of archaeological and biological data.

To see how well Recorder *6* performs with archaeological data, the author presents a comparison between the data structure of archaeological and biological data. Thus in the first part of the paper different archaeological methods and terms will be introduced. Those terms and methods will be compared with their biological partners. Eventually the suitability of classic Recorder's database and interface and of the thesaurus editor for archaeological information will be discussed.

A paper from Charles Copp also discusses the applicability of Recorder *6* to archaeological data. Although this article here has a similar approach, it is more biased towards the practical experience the MNHA made with Recorder *6*.

Nature of archaeological data

In archaeology, different methods exist regarding the gathering of data which determine how the data

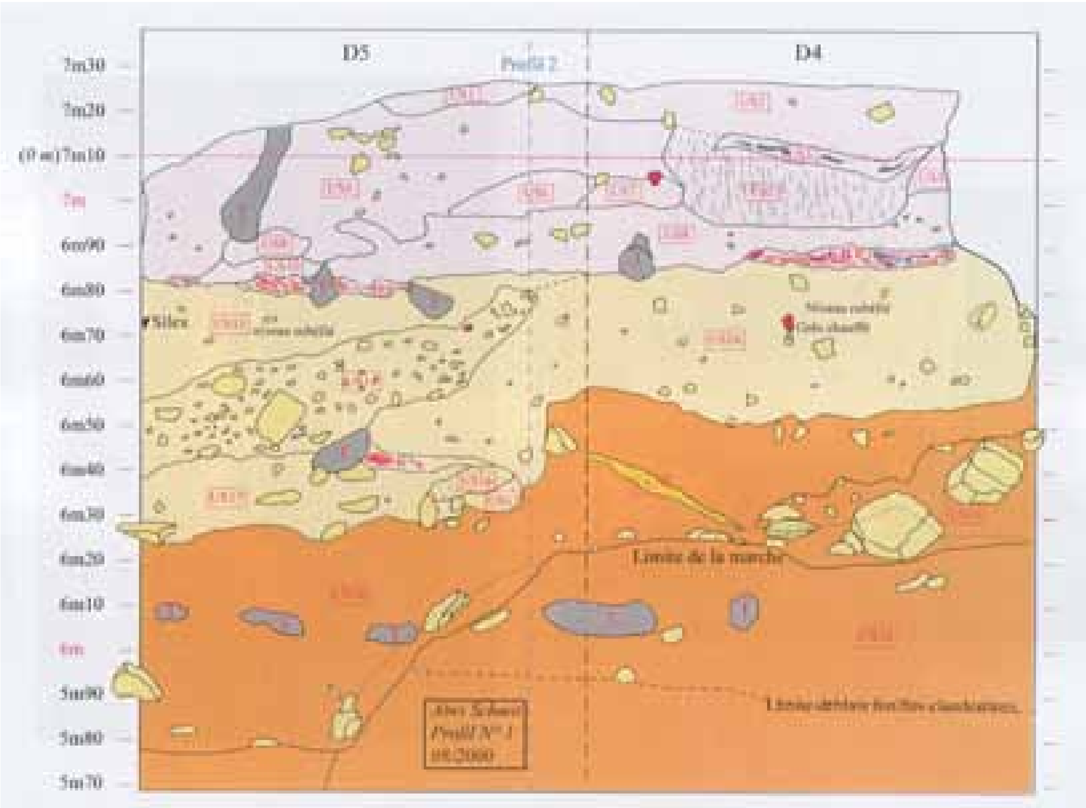


Fig. 1: Profile of a real world situation with structures and an object. (Profil Schmitt)

is collected and described. There are for example research or rescue excavations, archaeological field prospections or surveys, private collections and information in literature. Depending on the very different techniques used within each method, the amount and the type of the data gathered are very different. One cannot expect to have the same information and the same information detail from a planned excavation compared to the collecting methods of a private collector.

Usually when archaeologists do an excavation, the archaeological evidence encountered is organized in a hierarchical manner. But there are exceptions. At first, there is the area of interest which has to be examined. Within this area of interest there are test pits and soundings in which the archaeological evidence is systematically searched. Even if nothing is found within a test pit, this information needs to be stored because it helps to understand the habits and landuse in past times. Thus, even the absence of archaeological evidence is an important fact that needs to be recorded. If archaeological evidence is found within a test pit, the test pit can be extended. The archaeological evidence is then partitioned into structures and objects. Usually objects are found in a structure. In an old pit for example, that is filled with earth and sediments, there might be some charcoal within the earth filling. The structure here is the pit and the object is the charcoal. There is always a differentiation between a movable and a non-movable object. Usually structures are not movable, while objects are. One has also to keep in mind that structures might have substructures. For example, the basement of a house consists of several walls, which are built of brick stones. The house is the big structure while the walls are the substructures and the brick stones are the objects. This can even be broken down to the mortar connecting the brick stones. While this is quite theoretical and maybe obvious, the relationship of the structures helps

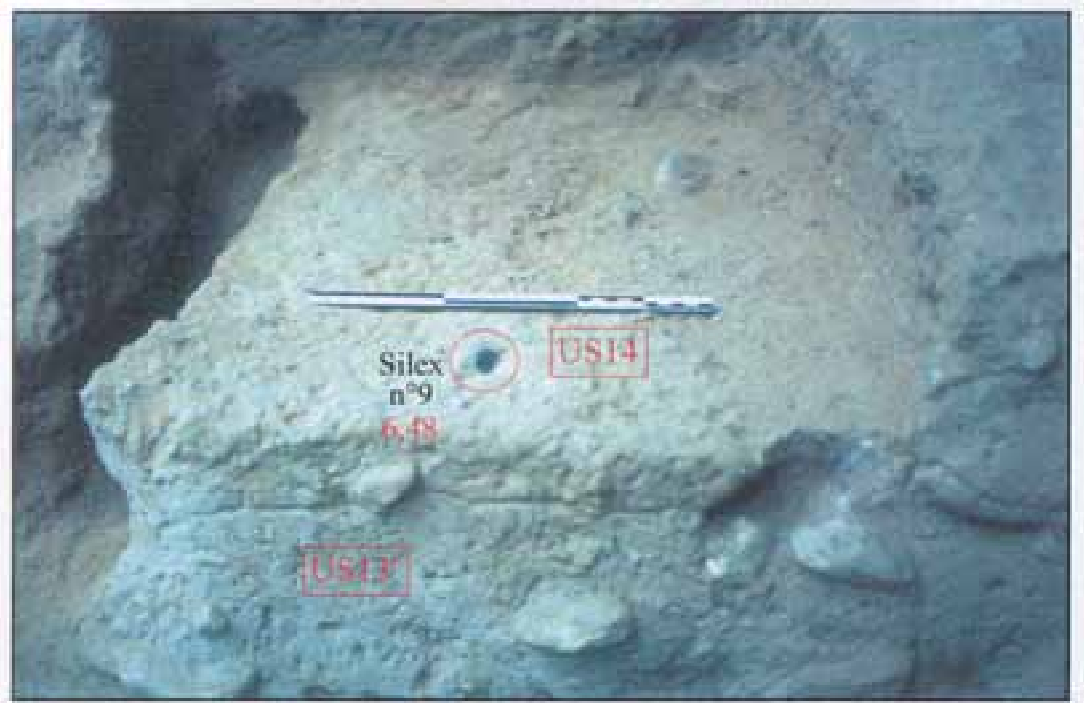


Fig. 2: Foto of a real world situation with structures and an object.

to understand the meaning of the structures one might find in archaeological studies. While a pit structure may be just a natural hole in the ground, several pits organized in a rectangular way, might be postholes of a house construction. The relationships between structures are not necessarily part of the first data acquisition in the field, but one should note and record them later on during the analysis of the excavation to allow further interpretation. Figure 1 shows a typical profile made during an excavation. One can notice the structures labelled with the red font (US1, etc.) and on the left side of the profile a silex object. Those kind of schemas help with the interpretation of a site.

While this scenario describes quite well the best case of an archaeological data acquisition, the archives are full of very vague and incomplete data. When looking through the notes of private collectors from the beginning of the last century, one often only finds a short note on the locality where an artefact was found, like for example „found at the third tree leaving the village XY in direction AB". While the information itself is interesting and valuable, it is impossible to handle it in the same schemas and criteria as standard excavation data from today. Unfortunately, this case is quite common. The descriptions of old acquisitions from a museum often just mention the city or village where the object was found and from whom it was aquired. Nevertheless this information needs to be stored in the database. Some collectors even use cryptic messages in their notes to hide the information where the object was found. And there is always the possibility of duplicate location names and the resulting inability to relate the object to the correct location.

As if the large variety of archaeological data was not enough to add to the confusion there is also a lack of international standardization of archaeological terms. Each country has its own thesaurus lists to describe archaeological data. In one country, the early middle age might start at another date than in the neighbouring country and might be called differently. Thus one does not only face the problem of dealing with different names for the same thing, but also with names that mean different things in different countries. In a country like Luxembourg, this can lead to serious problems since as much as three or four archaeological systems might have been used, the French, German, Belgian and Luxembourg system.

Suitability of Recorder

To see how well Recorder performs with archaeological data, a current report of an excavation was used for testing purpose. The test was mainly performed on the survey and the location module, the collections module was mostly left out, as managing collections is not the focus of the EPC project.

Excavation reports usually consist of a summary report and an extended description of the excavation methods and information recovered. The summary report holds ownership information, coordinates, and other information needed for a quick classification and localisation.

While most of the information in the excavation report can be entered quite easily, there are certain items which do not fit well in the current database structure of Recorder 6. Currently, structures (see „Nature of archaeological data") are not easy to implement into recorder. Structures, or all immovable data, can only be recorded as a location feature (Copp unpubl. 2006). This is a little bit unsatisfying, because usually there is a direct link between structure and object. For archaeologists it is not obvious that structures need to be recorded in the location browser while all other excavation data can be recorded in the survey browser. And of course, structures are not persistent to a location in the way a location feature is. Recorder might need to be extended to allow the linking of samples, in the archaeological case objects, to features, which are referred to as structures in the archaeological language. The feature module needs to be enhanced for allowing hierarchical links and spatial relationships between features. Just remember the case, where some postholes are grouped together to a house structure.

The survey module in Recorder 6 works already quite fine for archaeological purposes, as the essential characteristics and methodologies for both biological and archaeological information are quite similar. Survey, survey events and sampling methods are common terms in archaeology. While in the biological sense, occurrences or observations are linked to a specific sample, this is not the case in archaeology. While there exist also some prototypes, the interpretation of an archaeological object is much more flexible and may even vary regionally. Thus the link between the sample and the occurrence is not as strong.

When dealing with sites themselves and the classification into a schema, there are also some problems. When having an archaeological excavation the data gathered also serves to classify the site e.g. as Roman settlement, Palaeolithic stray find or Bronze Age burial. However, when dealing with sites found in the archives of the museum or mentioned in literature, the classification of the site is mostly not possible through the objects associated with the site. Most of the information found in the archives is very vague, e.g. where former archaeologists described a site only by mentioning it in their reports. Therefore, we can be sure to a certain degree that an archaeologist found a site, but we do not have the proof in form of data gathered. But this site should be recorded as for example a roman settlement, even if we don't have the proofing objects. Unfortunately the attachment of this kind of attribute is not possible in Recorder 6 as there is no possibility to assign this information to a survey.

Charles Copp's description of the suitability of Recorder for the EPC project should be used as a reference for a more detailed analysis (Copp unpubl. 2006).

Improvements

The previous chapter provided an overview over the suitability of Recorder from the data structure point of view. While this is definitely a very important point, the interface to the data in the database is as much important as the data itself. During the test, some issues arose concerning the interface of Recorder.

While testing the thesaurus editor and entering a few archaeological lists, it became clear that the interface of classic Recorder needs a better integration of the created thesaurus lists. Thesaurus lists can be used in archaeology to describe nearly everything one can find in the course of a survey, starting from the survey description itself and following on to the objects, for which the thesaurus can hold term lists on material, dating etc. The integrated term list editor from classic

Recorder unfortunately lacks the complexity of the thesaurus editor which permits the creation of multilevel lists. Thus crossing the gap between the simple term list tool within Recorder and the Thesaurus editor is necessary.

Dating is another problem, which is not well implemented within Recorder. While it is possible to handle the terminology and hierarchies for absolute, relative and cultural meanings, attribution of dates to contexts and specimens is problematic. Samples can be dated through linking it to a stratigraphy occurrence or similar, but this is not sufficient for archaeological purposes. Occurrences, or objects, can have a different age than the environment they were found in. Dating parts of objects is not possible either. Dates need to be attached to objects, structures, and even collections or groups of structures and objects. See Copp (2006) for more information about archaeological dating.

Conclusions

Recorder as it stands, seems already quite well suited for managing archaeological data and their associated information. Some of the differences between the nature of archaeological and biological data can only be accounted for by an extension to Recorder, similar to the collections or thesaurus module.

Recorder has a solid basement for storing archaeological information and already a good amount of data can be entered into Recorder without any modification. The management of collections was not part of this analysis. Nevertheless, as it seems from a first look, the collections module is already capable of handling our objects in the storage and showrooms.

Excavation data cannot be entered fully into the current version of Recorder. But creating an add-in specifically for this purpose seems possible and worthwhile.

Using the same database as the partner museum, the National Museum of Natural History, would enlarge the opportunity to join scientific forces and add a deeper level of collaboration. As both museums do access the same sources in their daily work, a database with the same structure would definitely enhance the quality of the knowledge gained. The exchange of thesaurus lists is also an interesting point for cooperation between the museums. The classification of human and animal bones, seeds from plants and other common thesaurus lists might be shared and the effort for the double work might be put into more fruitful work.

Reference

Copp C.J.T. 2006. - Notes arising form a meeting to discuss the data management and software requirements of the Prehistory Department of the Luxembourg National Museum of History and Art, 21.02.2006. Unpublished report. Musée national d'histoire et d'art, Luxembourg.

The Royal Museum for Central Africa in the era of biodiversity informatics

Patricia Mergen, Michel Louette, Jos Snoeks, Marc De Meyer, Danny Meirte
African Zoology Department, Royal Museum for Central Africa
Leuvensesteenweg 13
B-3080 Tervuren
Patricia.Mergen@africamuseum.be

Keywords: Biodiversity Informatics, Zoological collections, Royal Museum for Central Africa, Geographic Information System

Abstract

The Royal Museum for Central Africa (RMCA) is holding collections of about 10 million animal, wood and paleontological specimens, originating from the whole of the Afrotropical region (mainly from the central part) and has in recent years actively collaborated to biodiversity information projects. The collections "Xylarium" (wood samples) and "Prelude" (medicinal plants) of its Metafro Infosys project were in 2003 among the first Belgian collections presented to the Global Biodiversity Information Facility (GBIF) network with the help of the Belgian GBIF node. RMCA is the leading partner of two feasibility studies funded by the European Network of Biodiversity Information (ENBI, work package 13), namely "True Fruit Flies of the Afrotropical Region" and "Albertine Rift databases". It is a member of the FishBase Consortium which assists researchers and fisheries agents gaining access *via* the Internet to old and new literature on African fish, fisheries and aquaculture. The RMCA has broadened its activities in the field of access providing and training by setting up the Africa Biodiversity Information Centre (ABIC), which covers all the zoological and botanical groups represented in its collection. RMCA is an institutional member of the Consortium of European Taxonomical Facilities (CETAF) and of the Taxonomical Database Working group (TDWG). Through these networks it is involved in recent FP6 EU projects like SYNTHESYS (Synthesis of systematic resources) and EDIT (European Distributed Institute of Taxonomy). The RMCA is also an active collaborator of the Belgian GNOSIS project (Generalized Natural Sciences Online and Spatial Information System) based among other things on a decentralized infrastructure and on Open GIS Standards. This paper presents the activities of the RMCA in the field of biodiversity informatics and shows how the active participation of the institution in these various initiatives was, and still is beneficial both for RMCA and for its various stakeholders.

Mots clefs: Information sur la Biodiversité, collections zoologiques, Musée Royal de l'Afrique Centrale, Système d'information géographique

Le Musée Royal de l'Afrique Centrale (MRAC) héberge des collections d'environ 10 millions de spécimens d'animaux, de bois et de fossiles, originaires de l'ensemble de la région afrotropicale (principalement sa partie centrale) et a collaboré activement ces dernières années à des projets concernant l'information sur la biodiversité. Les collections «Xylarium» (échantillons de bois) et «Prélude» (plantes médicinales) des projets Metafro Infosys furent en 2003 parmi les premières collections belges présentées au réseau du Système Mondial d'Information sur la Biodiversité (SMIB) avec l'aide du nœud national belge du SMIB. Le MRAC est le coordinateur de deux études de faisabilité financé par le Réseau Européen d'Information sur la Biodiversité (ENBI, groupe de travail 13), à savoir «Les mouches de fruits de la Région afrotropicale» et «Les bases de données du Rift albertien». Il est membre du Consortium FishBase, qui a pour but d'assister les chercheurs et les pêcheurs à accéder *via* Internet à de la littérature ancienne et récente sur les poissons africains, les pêcheries et l'aquaculture. Le MRAC a élargi ses activités d'accès à l'information et de formation en créant le Centre d'Information sur la Biodiversité Africaine (CIBA), qui recouvre tous les groupes zoologiques et botaniques représentés dans ses collections. Le MRAC est membre institutionnel du Consortium Européen des institutions taxonomiques (CETAF) et du Groupe de Travail sur les bases de données taxonomiques (TDWG). De part ces réseaux, il est impliqué dans des projets européens récents au 6ième PC tels que SYNTHESYS (Synthèse des Ressources systématiques) et EDIT (Institut Distribué Européen en Taxonomie). Le MRAC est également collaborateur du projet belge GNOSIS (Système général d'information spatial en ligne sur les Sciences Naturelles), basé entre autres sur une infrastructure décentralisée et utilisant des standards SIG libre. Cet article présente les activités du MRAC dans le domaine de l'information sur la biodiversité et montre de quelle manière la participation de l'institution à ces diverses initiatives fut et est toujours bénéfique au MRAC et à ses utilisateurs.

Introduction

The Royal Museum for Central Africa (RMCA), Tervuren, Belgium is the leading research institute and knowledge centre on, *inter alia*, the biodiversity of species in the context of their natural environments in Africa, particularly Central Africa, and aims to develop an interest and understanding for the African fauna in the scientific communities.

It holds the largest biodiversity collection of Central Africa, offering a complete cross-section of reference material from many of the region's terrestrial and freshwater taxa. Furthermore, the majority of the specimens originate from the relatively poorly studied megadiversity belt in the equatorial region of Africa, and from West Africa.

African zoology: The Zoological Collections hold specimens from nearly 125,000 species: >7 million invertebrates of 117,000 species and 1,500,000 vertebrates of 6,115 species and hold the holotype material for 26,615 insect, 543 fishes, 240 birds, 104 reptiles, 81 amphibians and 36 mammals. Including paratypes, the collections hold a total of nearly 250,000 types. RMCA has e.g. nearly half of the 3,000 type specimens in the world for African freshwater fishes

Geological collections: The mineral collection consists of 10,255 ordinary minerals (in 988 species), 1,077 radioactive minerals and 170 asbestos samples (accessible under European safety regulations). There are 20,000 well-documented (micro)fossils including both fauna and flora species of Phanerozoic deposits of the Central African region.

Xylological collections: The RMCA keeps one of the largest collections of wood samples in the world (57,000 samples from nearly 13,000 different species, from 3,009 genera and 265 families). Focuses are on the commercial timbers and the family of Meliaceae.

The Library: The RMCA maintains an extensive library on African biodiversity, including the top scientific journals, but also a unique collection of rare and old publications and grey literature. Scientific staff master ten languages; consequently, the library and reprint documentation is unusually multilingual.

Biodiversity Information at RMCA

Like many other large scientific institutions throughout the world, RMCA participates in international initiatives such as the Global Biodiversity Information Facility (GBIF), the Taxonomical Database Working Group (TDWG) or the Consortium of European Taxonomic Facilities (CETAF), whose purposes are to improve access for all to Biodiversity information. The information is databased by RMCA's scientific and technical staff, thanks to ongoing efforts since several decades, and this in all the disciplines mentioned above. RMCA is also very active in the field of cooperation. Many partnerships with international scientific experts exist and trainings are organized *via* its African Biodiversity Information Center (ABIC) program, financed by the Belgian Co-operation and Development Agency. Thus, each year dozens of scientists, mainly originating from sub-Saharan Africa can undertake research visits to RMCA.

RMCA is member of the FishBase Consortium (www.fishbase.org) and is responsible for the correct introduction of the data relating to African fresh- and brackish water fish species. FishBase is currently the most important online encyclopedia on fish. It contains many tools intended for ichthyologists, fisheries and conservation agents, and the general public. Due to its membership to FishBase and thanks to an agreement with the Belgian Co-operation and Development Agency, RMCA organizes each year a training course focused on the taxonomy of African fish and on the usage of FishBase (Boden et al 2004).

Already in the early stages of GBIF's implementation, RMCA was very enthusiastic in becoming an active information provider in this international initiative. Information concerning the Xylarium collection (wood samples), and Prelude (medicinal plants) was among the first Belgian data a

b

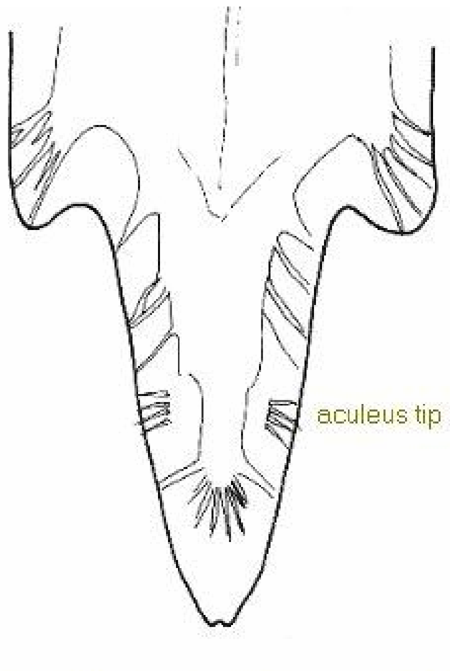


Fig. 1: Details of the species *Ceratitis hamata* De Meyer, a) oblique view of the head and thorax (automontage), b) dorsal view of the aculeus apex (drawings).

to appear on the GBIF portal. Currently, RMCA contributes to GBIF with nearly 130,000 entries from specimens of six of its biological collections and it is thus, in term of a number of entries, the first Belgian contributor to GBIF.

ENBI feasibility studies at RMCA

Work package 13 of the European Network for Biodiversity Information (ENBI) deals with making non-European biodiversity data in European repositories globally available. Many European countries have a colonial past and therefore a large share of global biodiversity information, especially from developing countries in the tropics, resides in databases of European institutions like RMCA (De Prins et al 2005).

In this context ENBI has financed four pilot studies. RMCA was the leading partner for two of these performed in close collaboration with several other Belgian and international Institutions, namely the "True Fruit Flies of the Afrotropical Region" and "The Albertine Rift Database" (Mergen et al 2005). The web portals have been implemented by the Belgian GBIF node as a sub-contracting activity. The portals, currently accessible at Bebif (n.d.) were fully implemented using OpenSource and free software and with Java and Python as programming languages. Particular attention was given to user-friendliness, end-user needs, platform and operating system independence and performance. Several search possibilities are offered from basic to advanced search interfaces as well as a very powerful full-text search option, using Apache Lucene (n.d.).

An extensive quality check has been performed on these data by the curators and other experts. Whenever available access is given to high quality images, drawings, descriptions, genetic resources and distributions maps of the specimens.

The fruit fly portal (Fig. 1) shows information from about 25,000 specimens of 169 African fruitfly species (De Meyer 1996, 1998, 2000; De Meyer & Copeland 2001, 2005; White et al 2003) and 900 host plants (De Meyer et al 2002) (Curators: Marc de Meyer, RMCA and Ian White, Natural History Museum London, NHML). The database is in the process to be extended by 200 species (genera Daucs and Bactrocera from Africa), which will be subsequently added to the web portal and updated at GBIF.

Four pilot taxa have been chosen for the Albertine Rift portal (Fig. 2):

The butterfly database (Dall'Asta & De Prins 2006) contains 24,498 records of 469 taxa and geo-referenced data for all localities incorporated. Habitus photographs are available for all species for the genera Papilio and Graphium, in total 168 taxa (Curators: Ugo Dall'Asta,, Jurate De Prins, RMCA).

Fig. 2: Birds, butterflies, flowering plants and lacustrine fish specimens from the Albertine Rift. © NBGB and RMCA

The bird database contains information on taxon and geo-referenced collection localities for approximately 35,000 specimens from the Albertine Rift (Louette et al 2002; Louette 2003; Louette et al 2006). For the 36 endemic species, photographs are included (Curators: Michel Louette, Danny Meirte, Marc Herremans, RMCA).

The cichlid fish database contains information of the specimen and frozen tissue collections of the RMCA and RBINS for the species from lakes Tanganyika and Kivu. These lakes contain a large number of endemic cichlid fishes (Snoeks et al 1994; Snoeks 2000, 2001; Verheyen et al 2003). The data represent 3,747 block records, with more than 8,000 specimens of 56 genera and about 225 species. Pictures of specimens taken in the field or drawings are shown (Curators Erik Verheyen, Royal Belgian Institute of Natural Sciences, RBINS and Jos Snoeks, RMCA).

The Rubiaceae database holds records of taxa from this plant family. The data presented focus on the Rubiaceae of the Albertine Rift region 120 species with minimum 5 specimens databased per species (Bamps 1982, Robbrecht 1988, 1993, 1996, Stoffelen et al 1990). Geo-referenced data for collection localities are available for all these specimens (Curators: Elmar Robbrecht and Piet Stoffelen National Botanical Garden of Belgium, NBGB).

For each species the user can directly check for additional information on the network of GBIF, (GBIF n. d.). A substantial part of the data has been provided to the GBIF network through the Belgian Node, using both DiGIR and BioCASE dataproviding tools. Additionally, a recent agreement between GBIF and Google Earth enables now to visualise all georeferenced records presented to GBIF (Fig. 3).

Fig. 3: Snapshot created with the GBIF occurrence data and Google Earth Services (http://ge.gbif.net), showing the sampling or observation sites of specimens from the Albertian Rift project served to GBIF (Cichlid fish: Aulonocranus dewindti (Boulenger, 1899), Bird: African Little Sparrowhawk Accipiter minullus (Daudin, 1800) limited in distribution in this region because it is parapatric to the Red-thighed Sparrowhawk Accipiter erythropus (Hartlaub, 1855) - not shown here, (see Louette 2002), Butterfly: Graphium ridleyanus (White, 1843), Rubiaceae: Fadogia ancylantha Heirn).

Conclusions

Due to its expertise gained in the fulfillment of the above mentioned project, RMCA is currently participating in several ongoing activities in the field of Biodiversity Information, both at a national and international level.

In the framework of the Network activities of the EU project SYNTHESYS for Synthesis of Systematic Resources, RMCA substantially contributes to the implementation of a demonstrator enabling a data quality check between itinerary data of biological specimen collectors and georeferenced data available through biodiversity networks, like GBIF or the Biological Collection Access Service for Europe (BioCASE). RMCA contributes to the EU Network of Excellence EDIT standing for European Distributed Institute of Taxonomy (which started in March 2006) mainly by contributing to the setting up a Cybertaxonomy Platform as well as by training and public awareness programs. About 200,000 specimens of African Amphibian specimens in RMCA are in the process of being provided to GBIF *via* the Herpnet Network managed by the University of California, Berkeley.

At a national level RMCA is a partner of the Belgian Federal Science Policy "Generalized Natural Sciences Online and Spatial Information System" project (GNOSIS). RMCA also has obtained a grant from the Belgian Biodiversity Platform for the digitalization and providing to GBIF of its African Muridae (Rodents) collection.

Finally, the Belgian Federal Science Policy has initiated a major initiative concerning the "Digitalization of the scientific and cultural inheritance of the 10 Federal Scientific Institutions (FSI)". RMCA as one of the FSI, will thus, actively continue to digitize its collections and enhance their accessibility in the upcoming years.

Acknowledgements

We would like to thank all contributing partners to the above mentioned projects, especially the Belgian GBIF Node for their technical support in the realization of the ENBI portal. The projects cited in the text are financed through several sources: EU projects ENBI, BioCASE, SYNTHESYS and EDIT, Grants from GBIF Seed Money Awards, Belgian Biodiversity Platform and the Belgian Science Policy Office.

References

Bamps P. 1982. - Flore d'Afrique Centrale (Zaïre, Rwanda, Burundi): Répertoire des lieux de récolte. Jardin Botanique National de Belgique, 224 p.

BeBIF 2006 – Belgium biodiversity information facility. ENBI feasibility studies workpackage 13. Webpage available at http://projects.bebif.be/enbi/. Accessed October 2006.

Boden, G. Musschoot, T. & Snoeks, J. 2004. - African fish biodiversity, fishbase and fish culture. Tropicultura 22 (special edition): 37-43.

Dall'Asta U. & De Prins J. 2006. - Collecties van het KMMA: vlinders / Collections du MRAC: Papillons / Collections from RMCA: Butterflies. Koninklijk Museum voor Midden-Afrika, Tervuren, 112 p.

De Meyer M. 1996. - Revision of the subgenus Ceratitis (Pardalaspis) Bezzi, 1918 (Diptera, Tephritidae, Ceratitini). Systematic Entomology 21: 15-26.

De Meyer M. 1998. - Revision of the subgenus Ceratitis (Ceratalaspis) Hancock (Diptera, Tephritidae, Ceratitini). Bulletin of Entomological Research 88: 257-290.

De Meyer, M. 2000. - Systematic revision of the subgenus Ceratitis MacLeay s.s. (Diptera, Tephritidae). Zoological Journal of the Linnean Society 128: 439-467.

De Meyer M. & Copeland R.S. 2001. - Taxonomic notes on the afrotropical subgenera Ceratitis (Acropteromma) Bezzi and C. (Hoplolophomyia) Bezzi (Diptera: Tephritidae). Cimbebasia 17: 77-84.

De Meyer M., Copeland R.S., Lux S.A., Mansell M., Quilici S., Wharton R., White I.M. & Zenz J. 2002. - Annotated check list of host plants for afrotropical fruit flies (Diptera: Tephritidae) of the genus Ceratitis. Zoologische Documentatie / Documentation zoologique, Koninklijk Museum voor Midden Afrika, Tervuren, 27: 1-91.

De Meyer M. & Copeland R.S. 2005. - Description of new Ceratitis MacLeay (Diptera, Tephritidae) species from Africa. Journal of Natural History 39: 1283-1297.

De Prins J., Calabuig I. & De Meyer M. 2005 - Deliverable D13.6 from work package 13: Report of the Evaluation Workshop organized in the Royal Museum for Central Africa, Tervuren (Belgium) April 13–14, 2005. ENBI report no. WP13_D13.6_06/2005, 68 p.

GBIF n.d. - Global biodiversity information facility. Prototype data portal. Available at http://genus.species.gbif.name/. Accessed March 2007.

ge.GBIF 2005 - Prototype data portal for GBIF occurrence data and Google Earth. Available at: http://ge.gbif.net. Accessed March 2007.

Louette M. 2002. - Relationship of the Red-thighed Sparrowhawk Accipiter erythropus and the African Little Sparrowhawk A. minullus. Bulletin of the British Ornithologists' Club 122 (3): 218-222.

Louette M., Meirte D., Louage A. & Reygel A. 2002. - Type specimens of birds in the collections of the Royal Museum for Central Africa, Tervuren. Zoologische Documentatie / Documentation Zoologique, Koninklijk Museum voor Midden Afrika, Tervuren 26: 1-105.

Louette M. 2003. - Biodiversity Challenges for Central Africa and the Role of the Tervuren Museum's Programme in Data Sharing. Bonner Zoologische Beiträge 51:150.

Louette M., Meirte D. & Mergen P. 2006. - The African bird data at the Royal Museum for Central Africa in the era of biodiversity informatics. Journal of Ornithology 147(5) supplement 1: 205.

Lucene 2007 – Lucene Apache website available at http://lucene.apache.org/. Accessed March 2007.

Mergen P., Slik J.W.F, van Welzen P.C., Verheyen E., Stoffelen P., Dall'Asta U., Snoeks J., De Meyer

M., White I.M., Förther H., Wautelet F., Vander Velde A. & Robbrecht E. 2005 - Making non-European biodiversity data in European repositories globally available. Retrieved October 2006 from BeBIF portal at http://projects.bebif.be/enbi/ .

Robbrecht E. 1988. - Tropical woody Rubiaceae. Opera Botanica Belgica 1: 1-271.

Robbrecht E. 1993. - Seventy years of systematics of tropical Rubiaceae at the National Botanic Garden. Bulletin van de Nationale Plantentuin van België 62: 7-34

Robbrecht E. 1996. - Geography of African Rubiaceae with reference to glacial rain forest refuges: 564-581, in van der Maesen L.J.G, van der Burgt X.M. & van Medenbach de Rooy J.M. (eds), The biodiversity of African plants. Kluwer Academic Publishers, Dordrecht.

Snoeks J., Rüber L. & Verheyen, E. 1994. - The Tanganyika problem: comments on the taxonomy and distribution patterns of its cichlid fauna. Archiv für Hydrobiologie Beiheft Ergebnisse der. Limnologie 44: 355-372.

Snoeks J. 2000. - How well known is the ichthyodiversity of the ancient African lakes ? in Rossiter A. &

Kawanabe H. (eds), Advances in ecological research. Ancient lakes: biodiversity, ecology and evolution, Academic Press 31: 17-38.

Snoeks J. 2001. - Cichlid diversity, speciation and systematics: examples for the Great African Lakes, in Coleman R. (ed.), Cichlid research: state of the art. Journal of Aquariculture & Aquatic Sciences 9: 150-166.

Stoffelen P., Robbrecht E., Smets E. 1990 - A new species of Coffea (Rubiaceae) from central Africa, with notes on tentative other taxa. Systematics and geography of plants 69(1): 119-124.

Verheyen E., Salzburger W., Snoeks J. & Meyer A. 2003. - The origin of the superflock of cichlid fishes from Lake Victoria, East Africa. Science 300: 325-329.

White I.M., Copeland R.S. & Hancock D.L. 2003. Revision of the afrotropical genus Trirhithrum Bezzi (Diptera: Tephritidae). Cimbebasia 18, 71-137.

Grassland mapping of the commune of Niederanven

Raymond Aendekerk
Station Biologique SIAS
5, rue de Neuhaeusgen
L-2633 Senningerberg
raymond.aendekerk.biostat@pt.lu

Marc Thiel
Fondation "Hëllef fir d'Natur"
rte de Luxembourg
L-1899 Kockelscheuer
thielmarc@gmx.net

Poster abstact

Objective

The grassland mapping of the commune of Niederanven was a contract for the ministry of the environment of Luxemburg in order to collect basic information about mesophile and ecologically diverse pastures and meadows. The background was a request of the EU. It is one of several mapping projects, which aims at recordig extensively used grasslands in Luxemburg.

Method

All the grasslands of the commune of Niederanven were recorded. From every parcel, a list of species with indication of occurrence (rare, sporadic, scattered, frequent, in places dominant, aspect building) was made to have a first good quantitative overview. Due to a lack of time and because this was not part of the contract, no records according to Braun-Blanquet were made. The grasslands mapped were classified into four groups: ecologically less important (green) grasslands, grasslands with a high amount of species (>30, yellow), with species listed in the biodiversity regulation (orange) or which are managed according to a biodiversity contract (red). Further information was observed about actual managing, ecological importance and proposals for a future sustainable management. No systematic records on fauna were made.

Results

Between Mai and October 2004, a total of 472 grassland parcels with a surface of 721 ha were mapped. For every parcel, indications of place, type of use and a list of species were recorded. 149 parcels with 207 ha (29 % of the grassland) are managed intensively, 28 with 59 ha (8 % of the grassland) have a high amount of species, 273 with 421 ha (58% of the grassland) have species according the biodiversity regulation and 20 with 34 ha (5 % of the grassland) have a biodiversity contract.

The high amount of pastures and meadows with a high ecological value results from the geological substrate (keuper), different inclinations and aspect, loamy soil with lime and an extensive agriculture in this area (low density of dairy-cows and cattle). In the valleys, we found alluvium substrate with only a few wetland parcels. Around the nature reserve „Aarnescht" (Mesobrometum) we find the highest amount of different species, like for example a pasture of 7ha with 119 different herbal species. All together 354 different species were recorded,of which 75 are listed in the biodiversity-regulation.

The BioCASE network: Integrating European Collection Information Ressources

S. Klein, A. Güntsch, M. Döring, and W. G. Berendsohn
Botanic Garden and Botanical Museum Berlin-Dahlem
Dept. of Biodiversity Informatics, Freie Universität Berlin
Königin-Luise-Str. 6-8
D-14191 Berlin
biodivinf@bgbm.org, http://www.bgbm.org/biodivinf

Keywords: biological collections, national node network, meta-level data, unit-level data

Poster abstract

Common portals to existing distributed and heterogeneous biological collection information resources improve considerably the accessibility of the underlying physical collections and facilitate the analysis of combined records of species occurrences.

The BioCASE project delivered a European collection information infrastructure emphasising accurate data definitions as well as convenient provider software, with the aim of making the connection of new data nodes as easy as possible. So far, more than 150 unit level collection databases were connected to the BioCASE network. They can now be accessed with several European and national data portals (e.g. http://www.biocase.org and http://www.gbif.de/botany/) and are also served to the Global Biodiversity Information Facility GBIF (http://www.gbif.org).

In parallel, a meta level collection information network has been implemented consisting of 31 national nodes synchronized with a core meta database (CORM) hosted at the Botanical Garden Berlin-Dahlem. With this, descriptions of more than 15000 collections in 30 European countries and Israel have been made accessible via the BioCASE portal.

Species records, data flow and the Biological Records Centre

Francis Rowland
Biological Records Centre, CEH Monks Wood
Abbots Ripton, Cambs GB-PE28 2LS Huntingdon
fro@ceh.ac.uk; post

Keywords: species, biological, records, recording, data, web, online, internet, schemes

Poster abstract

Data flow within BRC focuses to a large extent on receipt of biological records from external groups (e.g. recording schemes, societies, etc), management and processing of those data, and then preparation for output.

One of the major routes for that output is via the NBN Gateway. The latter is a website which allows the user to view distribution maps and download UK wildlife data by using a variety of interactive tools. The NBN Gateway will also be employed in future as a source for the data BRC uses in its research.

Various Local Record Centres (LRCs) and Country Agencies in the UK use the Recorder software to collate their data. They are able to supply datasets in the correct format for inclusion on the NBN Gateway. The biological records which BRC deals with can come from a variety of providers, and in different formats, e.g. Excel files, output from other recording packages, recording cards, and so on.

BRC is looking to the future for the recording of biological data, and continuously seek ways to improve the flow of data, and the communication between all those involved in biological recording.

BRC staff are currently beginning a project to look at online data capture for a range of smaller recording schemes. BRC hope to produce an online system that is simple and aimed at voluntary recorders in schemes and societies. It will include registration for recorders, allowing them to enter records for any online schemes BRC are running online, as well as tools for scheme or area organisers to manage those data.

Online data entry may be incidental, or else for particular grid referenced locations, specified by the registered user as part of their profile. Online recording forms will be designed in a way that means records will automatically be in NBN Exchange Format from the beginning.

Data will be held in a 'warehouse' database. These data will be made accessible to scheme or area organisers, allowing them to more rapidly verify and validate records. Many of the tools used by BRC and NBN Gateway staff to carry out validation (e.g. of grid references, species names, and so on) will be made available online to these expert organisers. Once data have been vetted, they can continue through the BRC data integration process, and can then be made available on the NBN Gateway.

Using the BRC online data entry facilities, registered recorders will also be able to view lists of species they have recorded and see maps of their distribution. Although larger organisations are likely to continue using Recorder software for collation of data, this service maybe useful to those schemes that require the technical support of BRC.

Participants list

Aendekerk, Raymond - Station Biologique SIAS | 5, rue de Neuhaeusgen | L-2633 Senningerberg | biostat@pt.lu

Barbier, Yvan (Dr) - Centre de Recherche de la Nature, des Forêts et du Bois | Avenue de la Faculté, 22 | B-B-5030 Gembloux | y.barbier@mrw.wallonie.be

Bechet, Georges (Dr.) - 25, rue Munster | L-2160 Luxembourg | georges.bechet@mnhn.lu

Betts, Hannah - Monkstone House | City Road | GB-PE1 1JY Peterborough | Hannah.Betts@JNCC.gov.uk

Broniewski, Frank (Diplom Geograf) - Cloefstraße 76a | D-66693 Mettlach | frank.broniewski@mnha.etat.lu

Broome, Andrew - Estate Office, Alderley Park, | Congleton Road | GB-SK10 4TG Macclesfield | andrew.broome@astrazeneca.com

Brown, Ross (Dr) - AstraZeneca UK Ltd. Brixham Environmental Laboratory | Freshwater Quarry, Devon | GB-TQ5 8BA Brixham | ross.brown@astrazeneca.com

Cal, Morgan - Dorset Software | New Quay Road | GB-BH15 4AF Poole | cal.morgan@dorset-software.com

Carrières, Evelyne - 27, rue des lentilles | F-F-54800 Sponville | evelyne.carrieres-kam@orange.fr

Caspari, Steffen (Dr) - Zentrum für Biodokumentation | Am Bergwerk 10 | D-66578 Landsweiler-Reden | s.caspari@biodokumentation.saarland.de

Cellina, Sandra - Sandra Cellina | 26, rue du Parc | L-3542 Dudelange | cecellina@yahoo.com

Colling, Guy (Dr.) - Musée national d'histoire naturelle | 25, rue Munster | L-2160 Luxemnbourg | gcolling@mnhn.lu

Copp, Charles - 8 The Paddock | GB-BS21 6JU Clevedon | eim@globalnet.co.uk

Downey, Liam (Dr) - The Heritage Council | Parliament Street | IE-NA Kilkenny | liam.downey@mail.com

Dufrêne, Marc (Dr.) - Centre de Recherche de la Nature, des Forêts et du Bois | AVENUE MARECHAL JUIN, 23 | B-5030 GEMBLOUX | M.DUFRENE@MRW.WALLONIE.BE

Frank, Guy - Ministère de la Culture, de l'Enseignement Supérieur et de la Recherche | montée de la Pétrusse, 20 | L-2912 Luxembourg | guy.frank@mc.etat.lu

Gordon, Barker - Heelis | Kemble Drive | GB-SN2 2NA Swindon | gordon.barker@nationaltrust.org.uk

Guinet, Jean-Michel - 25, rue Munster | L-2160 Luxembourg | jmguinet@mnhn.lu

Güntsch, Anton - Botanic Garden and Botanical Museum Berlin-Dahlem | Königin-Luise-Str. 6-8 | D-14191 Berlin | a.guentsch@bgbm.org

Hardwick, Richard (Dr) - Chaussée de Watermael, 93 | B-1160 Bruxelles | rch@skynet.be

Hastert, Marc - Am Fuurt, no 10 | L- 7418

Buschdorf | marc.hastert@ehl.lu

Helminger, Thierry (Dipl. Biol.) - Musée national d'histoire naturelle | 25, rue Munster | L-2160 Luxembourg | thelminger@mnhn.lu

Hoffmann, Lucien (Prof. Dr.) - Public Research Center-Gabriel Lippmann | rue du Brill, 41 | L- 4422 Belvaux | hoffmann@lippmann.lu

Houtgraaf, Dirk - PO Box 9517 | NL-2300 RA Leiden | houtgraaf@naturalis.nl

Hussey, Charles - Natural History Museum | Cromwell Road | GB-SW7 5BD London | c.hussey@nhm.ac.uk

Jimenez-Krause, Daniel - ZADI | Villichgasse 17 | D-53177 Bonn | jk@zadi.de

Johnson, Rebecca - Solway Heritage, Campbell House, The Crichton | Bankend Road | GB-DG1 4ZB Dumfries | rjohnson-dgerc@solwayheritage.co.uk

Kingston, Naomi (Dr) - 7 Ely Place | IE- Dublin 2 | naomi.kingston@environ.ie

Klein, Stefan - Botanic Garden & Botanical Museum Berlin-Dahlem | Königin-Luise-Str. 6-8 | D-14191 Berlin | s.klein@bgbm.org

Kremer, Gilles - 9, rue Bofferding | L- 4911 Bascharage | gilles.kremer@mnhn.lu

Krippel, Yves (Ing. agr. dipl.) - Rue de Rollingen 18A | L-7475 Schoos | yves.krippel@mnhn.lu

Kunsch, Manuel - 25, rue Munster | L-2160 Luxembourg

Lorgé, Patric - Kräizhaff | Route de Luxembourg | L-.18899 Kockelscheuer | col@luxnatur.lu

Lysaght, Liam (Dr) - Rothe House | Parliament Street | IE-NA Kilkenny | llysaght@heritage-council.com

Maunsell, Michael - Tipperary Institute | Nenagh Road | IE-Co. Tipperary Thurles | mmaunsell@tippinst.ie

May, Rudolf - Konstantinstrasse | 110 | D-53179 Bonn | rudolf.may@bfn.de McFerran, Damian (Dr) - Ulster Museum | Botanic Gardens | NIE-BT9 5AB Belfast | damian.mcferran@magni.org.uk

McLean, Alistair - Weston Park Museum | Western Bank | GB-S10 2TP Sheffield | alistair.mclean@sheffieldgalleries.org.uk

Meisch, Jim - Musée National d'Histoire Naturelle | 25, rue Munster | L-2160 Luxembourg | jmeisch@mnhn.lu

Mergen, Patricia (Dr) - Royal Museum for Central Africa, Department of Zoology | Leuvensesteenweg 13 | B-3080 Tervuren | patricia.mergen@africamuseum.be

Moes, Marc - Rue de Dippach 27 | L-8055 Bertrange | marcmoes@pt.lu

Mony, Frédéric - Chateau du Montet | Rue du Doyen Roubault | F-54500 Vandoeuvre les Nancy | frederic.mony@ciril.fr

Mozek, Ronny - rue Fernand Mertens, 39 | L-3258 Bettembourg | ronnymozek@yahoo.de

Muhammad, Asif - SELU, School of biological and molecular sciences | Oxford Brookes University,Gipsy Lane,Headington Campus, U.K. | GB-OX3 0BP Oxford | 04062274@brookes.ac.uk

Müller, Edgar - Zur Dynamitfabrik 8 | D-66793 Saarwellingen | edgar-mueller@loparit.de

Nederbragt, Hans - Oude Delft 200 | NL-2611 HH Delft | hans@trezorix.nl

Peiffer, Patrick - 20, haaptstrooss | L-9181 Tadler | Patrick.Peiffer@bi.etat.lu

Peltzer, Raymond - rue Zénon Bernard 62 | L-4031 Esch-Alzette | rpeltzer@pt.lu

Rick, Susanne - 241 rue de Luxembourg | L-8077 Bertrange | susanne.rick@mnha.etat.lu

Ries, Christian (Dr) - Musée national d'histoire naturelle | 25, rue Munster | L-2160 Luxembourg | cries@mnhn.lu

Rissone, Adrian - Natural History Museum | Cromwell Road | GB-SW7 5BD London | A.Rissone@nhm.ac.uk

Roper, Charles - Sussex Biodiversity Record Centre | Woods Mill | GB-BN5 9SD Henfield | charlesroper@sussexwt.org.uk

Roscher, Sabine - ZADI | Villichgasse 17 | D-53177 Bonn | roscher@zadi.de

Rowland, Francis - CEH Monks Wood | Abbots Ripton | GB-PE28 2LS Huntingdon | fro@ceh.ac.uk

Scheer, Anne (DI) - Naturpark Öewersauer | 15 rte de Lultzhausen | L-9650 Esch-sur-Sûre | anne.scheer@naturpark-sure.lu

Schneider, Thomas - Klinkerstr. 92 | D-66663 Merzig | ct.schneider.mzg@t-online.de

Sigwart, Julia - National Museum of Ireland (Natural History) | Merrion Street | IE-2 Dublin | julia.sigwart@ucd.ie

Sowa, Frank - rue de Nospelt, 2 | L-8394 Olm | frank.sowa@sicona.lu

Speers, Larry - GBIF | Universitetsparken 15 | DK-2100 Copemhagen | LSpeers@GBIF.ORG

Story, Sam - Dorset Software | New Quay Road | GB-BH15 4AF Poole | sam.story@dorset-software.com

Thiel, Marc (Dipl. Agr. Ing.) - om Eer, 10 | L-6850 Manternach | thielmarc@gmx.net

Tholl, Marie-Thérèse - Duerfstrooss 31 | L-9647 Doncols | mttholl@pt.lu

Tweddle, John (Dr) - The Natural History Museum | Cromwell Road | GB-SW7 5BD London | j.tweddle@nhm.ac.uk

van Breda, John - Corbiere House | New Quay Rd | GB-BH15 4AF Poole | john.vanbreda@dorset-software.com

Walisch, Tania - Musée National d'Histoire Naturelle | 25, rue Munster | L-2160 Luxembourg | twalisch@mnhn.lu

Weiler, Marie-Antoinette - Musée national d'histoire naturelle | 25, rue Munster | L-2160 Luxembourg

Wilkinson, Steve (Dr) - Joint Nature Conservation Committee | City Road | GB-PE1 1JY Peterborough | steve.wilkinson@jncc.gov.uk

Wolff, Frank - Ministère de l'environnement | Montée de la Pétrusse, 18 | L-2918 Luxembourg | frank.wolff@mev.etat.lu

Wörz, Arno (Dr.) - Naturkundemuseum | Rosenstein 1 | D-70191 Stuttgart | woerz.smns@naturkundemuseum-bw.de

Zimmer, Jean-Marie - 25, rue Munster | L-2160 Luxembourg | jmzimmer@mnhn.lu

Zupins, Ivars - Natural History Museum of Latvia (Latvijas Dabas muzejs) | K.Barona 4 | LV-1712 Riga | ivars.zupins@dabasmuzejs.gov.lv

INSTRUCTIONS TO AUTHORS

Please check our internet site http://www.mnhn.lu for the latest version of these instructions!

Scope

FERRANTIA is a series of monographic works (20-250 pages in final layout) dealing with life and earth sciences, preferably related in some way or other to the Grand-Duchy of Luxembourg.

It publishes original results of botanical, zoological, ecological, geological, mineralogical, paleontological, geophysical and astrophysical research and related fields.

A complete issue of FERRANTIA may be devoted to several papers on a single topic as the responsibility of an invited editor.

Copyright

The submission of a manuscript to FERRANTIA implies that the paper must not have been accepted for publication or be under consideration elsewhere.

Copyright of a published paper, including illustrations, becomes the property of the publisher. Requests to reproduce material from FERRANTIA should be addressed to the editor.

Reviewing

Articles submitted for publication are reviewed by the editorial board and by one or two referees. The final decision on acceptance or rejection of the manuscript is taken by the editorial board. Manuscripts not prepared according to the following instructions to authors will be returned for correction prior to review.

Nomenclature

Papers with a systematic content should strictly follow the International Codes of Nomenclature.

Specimens

We recommend that the authors should deposit at least a part of the type material in the MNHN collections.

Publication dates

FERRANTIA pays special attention to publication dates, which are always specified to the day of publication.

Manuscripts

Manuscripts, without limitation of the number of pages, must conform strictly to the instructions to authors, and should be sent to the Editor:

FERRANTIA
Travaux scientifiques du Musée national d'histoire naturelle de Luxembourg
25, rue Munster
L-2160 Luxembourg

Format

Manuscripts must be submitted as paper copy in triplicate in A4 format, preferably double-spaced, with margins of at least 3 cm and all pages numbered.

If possible, an electronic version of the text may also be sent as unformatted Word document (PC or MAC) (font Times New Roman, 10 pt). Tables (Word, Excel) and figures (300 dpi in the page size of the journal) may also be sent electronically.

Structure

Papers are to be written in simple, correct and concise French, German or English. They should be organized as follows:

- a brief title (should not include the name of new taxa);
- a suggested running head (no more than 50 characters);
- name(s) and first name(s) of author(s), followed by their full address(es) and, if possible, e-mail or fax number;
- abstracts in English, French and German, each 200-800 words long; new taxa names should be included in the abstract; the abstract should be precise and descriptive, in order to be reproduced as such in data bases; avoid vague sentences such as "three new species are described" or "species are compared to species already known"; include precise differential characters;
- text of the article, in the following order: Introduction, Abbreviations used, Material and methods, Results and/or Observations, Discussion, Acknowledgements, References. The arrangement of the parts "Results/ Observations" and "Discussion" may be modulated according to the length and subject of the article; very long papers may include a table of contents;
- for systematic descriptions, each description should follow the order: name of taxon with author and date, synonymy, type material, etymology, material examined, distribution, diagnosis and/or description, remarks.
- description of geological features should include type level, type horizon, type locality. This order may be

adapted according to the concerned groups: consult a recent issue of FERRANTIA;

- taxon names must be stated with author (and publication date, separated by a comma, where appropriate) at least once at the first mention. At subsequent mentions of the same taxon, or other taxa of the same genus, the genus name may be abbreviated (Rosa canina L. to R. canina).
- use n. sp., n. gen., n. fam., etc. for new taxa;
- use italicized words only for taxa of generic and sub-generic ranks;
- use lowercase characters for authority names
- references to illustrations and tables should be indicated as follows: (Fig. 1), (Fig. a, d), (Fig. 2a-d), (Figs 3; 6), (Figs 3-5; Tab. 2); (Tab. 1); for German texts use Abb. instead of Fig.
- footnotes should not be used.

Tables and figures

Copies of all figures and tables should be included with the manuscript. They can be either included in the text at the correct locations with their legends or referenced in the text and included as annexes.

The editorial board will pay special attention to the quality and relevance of illustrations. Colored illustrations are accepted where appropriate and necessary.

Line drawings must be in Indian ink or high quality laser printouts; high contrast Fotographs are required,

Illustrations can be grouped into composite plates the elements of which are identified by letters (a, b, c...). Plates are not placed at the end of the article: they will be considered as figures and numbered as such. Arrange figures to fit in one (70 x 200 mm) or two columns (144 x 200 mm) or one half page (144 x 100 mm). Letters, numbers, etc., for each figure, are to be indicated on an accompanying overlay, not on the original figure. They will be inserted by the printer. A scale bar is required for each figure, when appropriate. No diagram or table is to exceed one page; longer tables should be divided.

References

In main text, references to authors, in lower case, should be presented without comma before year, as follows: Smith (2001), Smith (2001, 2002), (Smith 2001), (Smith 2001; Jones 2002), (Smith & Jones 2003, 2005), (Smith, Jones & Johnson 2003), Smith (2001: 1; 2003: 5), Smith (2001: fig. 2).

References should be presented as follows, in alphabetical order. Do not abbreviate journal names:

Høeg J. T. & Lützen J. 1985. - Comparative morphology and phylogeny of the family Thompsoniidae (Cirripedia: Rhizocephala: Akentrogonida) with description of three new genera and seven new species. Zoologica Scripta 22: 363-386.

Marshall C. R. 1987. - Lungfish: phylogeny and parsimony, in Bernis W. E., Burggren W. W. & Kemp N. E. (eds), The Biology and Evolution of Lungfishes, Journal of Morphology 1: 151-152.

Röckel D., Korn W. & Kohn A. J. 1995. - Manual of the Living Conidae. Volume 1: Indo-Pacific Region. Christa Hemmen, Wiesbaden, 517 p.

Schwaner T. D. 1985. - Population structure of black tiger snakes, Notechis ater niger, on off-shore islands of South Australia: 35-46, in Grigg G., Shine R. & Ehmann H. (eds), Biology of Australasian Frogs and Reptiles. Surrey Beatty and Sons, Sydney.

Gerecke R. et al. 2005. - Die Fauna der Quellen und des hyporheischen Interstitials in Luxemburg unter besonderer Berücksichtigung der Milben (Acari), Muschelkrebse (Ostracoda) und Ruderfusskrebse (Copepoda). Ferrantia 41, Musée national d'histoire naturelle, Luxembourg, 140 p.

Proofs and reprints

Proofs will be sent to the author (or the first author) for correction and must be returned within two weeks by priority air mail. Authors will receive twenty-five reprints free of charge; further reprints can be ordered at a charge indicated on a form supplied with the proofs.

Page layout of final publication

paper size	170 x 240 mm
page size	144 x 200 mm
nomber of columns	2
column width	70 mm
space between columns	4 mm
top margin	22 mm
bottom margin	18 mm
inside margin	15 mm
outside margin	11 mm

Fonts

Body text: Palatino linotype (serif), 9pt

Titles, legends, headers, footers: Trebuchet (sans-serif)

LISTE DES NUMÉROS PARUS À CETTE DATE:

Les volumes de la serie»FERRANTIA» paraissent à intervalles non réguliers.

Travaux scientifiques du Musée national d'histoire naturelle (1981-1999)

I Atlas provisoire des Insectes du Grand-Duché de Luxembourg. Lepidoptera. 1ère partie (Rhopalocera, Hesperiidae). Marc Meyer et Alphonse Pelles, 1981.

II Nouvelles études paléontologiques et biostratigraphiques sur les Ammonites du Grand-Duché de Luxembourg, de la Province du Luxembourg et de la région Lorraine attenante. Pierre L. Maubeuge, 1984.

III Revision of the recent Western Europe species of genus *Potamocypris* (Crustacea, Ostracoda). Part 1: Species with short swimming setae on the second antennae. Claude Meisch, 1984.

IV Hétéroptères du Grand-Duché de Luxembourg

1. *Psallus (Hylopsallus) pseudoplatani* n. sp. (Miridae, Phylinae) et espèces apparentées. Léopold Reichling, 1984.

2. Quelques espèces peu connues, rares ou inattendues. Léopold Reichling, 1985.

V La bryoflore du Grand-Duché de Luxembourg: taxons nouveaux, rares ou méconnus. Ph. De Zuttere, J. Werner et R. Schumacker, 1985.

VI Revision of the recent Western Europe species of genus *Potamocypris* (Crustacea, Ostracoda). Part 2: Species with long swimming setae on the second antennae. Claude Meisch, 1985.

VII Les Bryozoaires du Grand-Duché de Luxembourg et des régions limitrophes. Gaby Geimer et Jos. Massard, 1986.

VIII Répartition et écologie des macrolichens épiphytiques dans le Grand-Duché de Luxembourg. Elisabeth Wagner-Schaber, 1987.

IX La limite nord-orientale de l'aire de *Conopodium majus* (Gouan) Loret en Europe occidentale. Régine Fabri, 1987.

X Epifaune et endofaune de *Liogryphaea arcuata* (Lamarck). Contribution à l'écologie des populations de *Liogryphaea arcuata* (Lamarck) dans le Sinémurien au NE du Bassin de Paris. Armand Hary, 1987.

XI Liste rouge des Bryophytes du Grand-Duché de Luxembourg. Jean Werner, 1987.

XII Relic stratified scress occurences in the Oesling (Grand-Duchy of Luxembourg), approximate age and some fabric properties. Peter A. Riezebos, 1987.

XIII Die Gastropodenfauna der «angulata-Zone» des Steinbruchs «Reckingerwald» bei Brouch. Hellmut Meier et Kurt Meiers, 1988.

XIV Les lichens épiphytiques et leurs champignons lichénicoles (macrolichens exceptés) du Luxembourg. Paul Diederich, 1989.

XV Liste annotée des Ostracodes actuels non-marins trouvés en France (Crustacea, Ostracoda). Claude Meisch, Karel Wouters et Koen Martens, 1989.

XVI Atlas des lichens épiphytiques et de leurs champignons lichénicoles (macrolichens exceptés) du Luxembourg. Paul Diederich, 1990.

XVII Beitrag zur Faunistik und Ökologie der Schmetterlinge im ehemaligen Erzabbaugebiet "Haardt" bei Düdelingen. Jos. Cungs, 1991.

XVIII Moosflora und -Vegetation der Mesobrometen über Steinmergelkeuper im Luxemburger und im Bitburger Gutland. Jean Werner, 1992

19 Ostracoda. Nico W. Broodbakker, Koen Martens, Claude Meisch, Trajan K. Petkovski and Karel Wouters, 1993.

20 Les haies au Grand-Duché de Luxembourg. Konjev Desender, Didier Drugmand, Marc Moes, Claudio Walzberg, 1993.

21 Ecology and Vegetation of Mt Trikora, New Guinea (Irian Jaya / Indonesia). Jean-Marie Mangen, 1993.

22 A checklist of the recent non-marine ostracods (Crustacea, Ostracoda) from the inland waters of South America and adjacent islands. Koen Martens & Francis Behen, 1993.

23 Ostracoda. Claude Meisch, Roland Fuhrmann, Karel Wouters, Gabriele Beyer and Trajan Petrovski, 1996.

24 Die Moosflora des Luxemburger Oeslings. Jean Werner, 1996.

25 Atlas des ptéridophytes des régions lorraines et vosgiennes, avec les territoires adjacents, Georges Henri Parent, 1997.

26 Evaluation de la qualité des cours d'eau au Luxembourg en tant qu'habitat pour la loutre. Groupe Loutre Luxembourg, 1997.

27 Notes Paléontologiques et Biostratigraphiques sur le Grand Duché de Luxembourg et les régions voisines. Pierre Louis Maubeuge & Dominique Delsate, 1997.

28 Die Moosflora der Kleinen Luxemburger Schweiz (Müllertal). Florian Hans, 1998.

29 Etude sur les genres *Globorilusopsis* Maubeuge, 1994 et *Simoniceras* n. gen. du Lias Supérieur du Grand-Duché de Luxembourg (Calyptoptomatida). Pierre Louis Maubeuge, 1998.

30 L'Ichthyofaune du Toarcien luxembourgeois. Cadre général et catalogue statistique. Dominique Delsate, 1999.

31 Proceedings of the 3rd European Batdetector Workshop. 16-20 August 1996 Larochette (Lux.). Christine Harbusch & Jacques Pir (eds.), 1999.

32 Les collections paléontologiques du Musée national d'histoire naturelle de Luxembourg. Fossiles du Trias et du Jurassique. Dominique Delsate, Chris Duffin & Robi Weis, 1999.

FERRANTIA (2002-)

33 Die Fledermäuse Luxemburgs (Mammalia : Chiroptera). Christine Harbusch, Edmée Engel, Jacques Pir, 2002.

34 The Protura of Luxembourg. Andrzej Szeptycki, Norbert Stomp, Wanda M. Weiner, 2003.

35 Liste rouge des bryophytes du Luxembourg. Jean Werner, 2003.

36 Paléontologie au Luxembourg. Simone Guérin-Franiatte (éd.), 2003.

37 Verbreitungsatlas der Amphibien des Großherzogtums Luxemburg. Roland Proess (éd.), 2003.

38 Trois études sur la Zone Rouge de Verdun. I. Herpétofaune. II. La diversité floristique. III. Les sites d'intérêt botanique et zoologique. Georges H. Parent, 2004.

39 Verbreitungsatlas der Heuschrecken des Großherzogtums Luxemburg. Roland Proess, 2004.

40 Les macrolichens de Belgique, du Luxembourg et du nord de la France - Clés de détermination. E. Sérusiaux, P. Diederich & J. Lambinon, 2004.

41 Die Fauna der Quellen und des hyporheischen Interstitials in Luxemburg unter besonderer Berücksichtigung der Milben (Acari), Muschelkrebse (Ostracoda) und Ruderfusskrebse (Copepoda). Reinhard Gerecke, Fabio Stoch, Claude Meisch, Isabel Schrankel, 2005.

42 Red List of the Vascular Plants of Luxembourg. Guy Colling, 2005.

43 Contribution à la climatologie du Luxembourg. Analyses historiques, scénarios futurs. Christian Ries (éd.), 2005.

44 Sandstone Landscapes in Europe - Past, Present and Future. Proceedings of the 2nd International Conference on Sandstone Landscapes. Vianden (Luxembourg) 25-28.05.2005. Christian Ries & Yves Krippel (Editors), 2005.

45 Additions et corrections au catalogue des plantes vasculaires de l'arrondissement de Montmédy. Etude sur l'évolution séculaire de la flore. Georges H. Parent, 2006.

46 Beiträge zur Paläontologie de Unterdevons Luxemburgs (1). Christian Franke (Hrsg.), 2006.

47 Verbreitungsatlas der Libellen des Großherzogtums Luxemburg. Roland Proess, 2006.

48 Les Hêtres tortillards, *Fagus sylvatica* L. var. *tortuosa* Pépin, de Lorraine, dans leur contexte européen. Georges H. Parent, 2006.

49 Inventaire minéralogique du Luxembourg - Stolzembourg, Schimpach, Goesdorf. Simon Philippo (éd.), 2007.

50 Inventaire de la biodiversité dans la forêt «Schnellert» (Commune de Berdorf). Marc Meyer & Evelyne Carrières (éds.), 2007.

51 Proceedings of the first international Recorder conference. Luxembourg 2-3 December 2005. Tania Walisch (Editor), 2007.

Envoyez votre commande aux adresses indiquées à la page 2 de la couverture.