The EDIT Cyberplatform for Taxonomy and the Taxonomic Workflow: Selected Components

P. Ciardelli, P. Kelbert, A. Kohlbecker, N.Hoffmann, A. Güntsch, W.G. Berendsohn

Dept. of Biodiversity Informatics and Laboratories
Botanic Garden and Botanical Museum Berlin-Dahlem
Freie Universität Berlin
Königin-Luise-Str. 6-8
D-14195 Berlin-Dahlem
{ p.kelbert, a.kohlbecker, n.hoffmann, p.ciardelli, a.guentsch, w.berendsohn}@bgbm.org

Abstract: The EDIT Cyberplatform for Taxonomy is an EU-funded set of loosely coupled tools for the editing, management and presentation of taxonomic data in biology. This paper looks at the fundamental workflow issues the Cyberplatform is intended to address, then examines three of its main components from this workflow perspective. Using these components as an example, we will demonstrate concrete ways the Cyberplatform can improve and accelerate this workflow.

The paper starts by describing the Cyberplatform and its goals of loose coupling and interoperability, then looks at how these are built into the Common Data Model (CDM) Java library which forms the foundation for most Platform components. The first Platform component we examine in depth is the EDIT Desktop Taxonomic Editor, which presents a modern solution to the challenges of capturing the taxonomic workflow in software, by using techniques such as drag-and-drop, on-the-fly parsing, and unobtrusive feedback. The EDIT Specimen Explorer, the second component examined, helps find taxonomically relevant specimen and observation data by searching the GBIF (Global Biodiversity Information Facility) index using checklist-based thesauri to deliver more complete and targeted results, thereby improving and accelerating the workflow for exploring the taxonomic data available in the community as a whole. Finally, we look at a pilot project for print publishing software, which aims to remove the final bottleneck in the taxonomic workflow, the back-and-forth between taxonomist and publishing house.

1 Taxonomic Workflow

The European Distributed Institute of Taxonomy (EDIT) is an EU-funded project designed to help integrate the traditionally disparate field of scientific taxonomy as practiced in Europe. The EDIT Cyberplatform for Taxonomy brings the taxonomic workflow to the Internet, providing an open architecture to connect and integrate existing applications and developing new tools where bottlenecks and areas for improvement have been identified. Areas for improvement which will be addressed in this paper include: increasing the community's ability to exchange data by means of a common data model and improved tools for import, export and querying; improving upon the last generation of data input and editing tools with more of an emphasis on user-friendly and intuitive frontend software; and putting as much of the preparation of pre-publication drafts as possible into the hands of the taxonomist with the aim of accelerating and improving the quality of publication.

The workflow the Cyberplatform seeks to optimize is in essence the "revisionary" process by which an existing classification of a group of organisms is revised, and by which previously unclassified organisms are assigned to a "taxon", i.e. a class of organisms. As commonly modelled when developing taxonomic software, a taxon has a scientific name with a taxonomic rank (e.g. kingdom, species, etc.) and information is assigned to it such as physical and media specimens; descriptive data including geographical distribution and visual observations; and citations from existing literature.

The workflow begins either in the field with the collection and transport of specimens, or with a review of existing literature and specimens. Both of these are then followed by further analysis of specimens, the resolution of the name according to community nomenclatural codes, and peer review and publication of revisionary treatment including new taxa (if any). An example of a workflow bottleneck addressed by the Cyberplatform is the frequent difficulty in examining relevant specimen material, either because of delays inherent in obtaining physical specimens from cooperating institutions, or because of naming discrepancies in the way specimens are labelled; the component described in Chapter 4 tackles the second half of this problem.

2 A platform for Cybertaxonomy

The EDIT Platform for Cybertaxonomy, henceforth called "the Platform", covers the breadth of the taxonomic workflow, from fieldwork to publication. It provides taxonomists but also life science in general with a set of loosely coupled tools for: full, customized access to taxonomic data; editing and management of data; and collaborative work in teams. The Platform provides various tools to facilitate fieldwork, analyze data, assemble treatments, and publish efficiently. Reliability and reusability of data is a key requirement for each of these tools and thus for the Platform as a whole.

Development of the Platform is coordinated by the Dept. of Biodiversity Informatics at the Botanic Garden and Botanical Museum Berlin-Dahlem, and its various components are implemented by a team of 15 software developers and architects from multiple institutions all over Europe.

2.1 General architecture – loosely coupled components

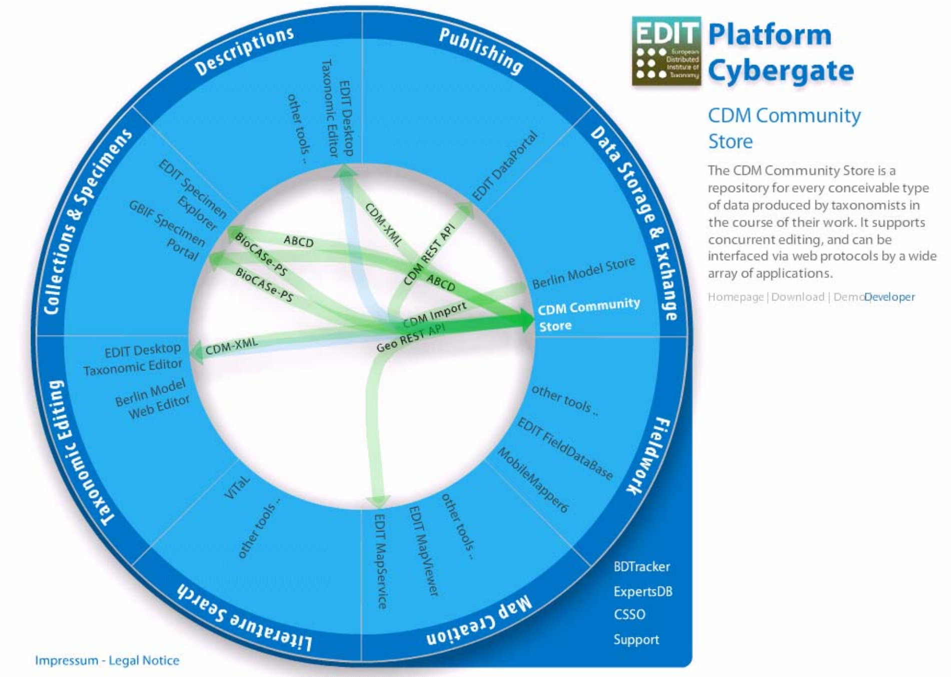
Several existing applications support various facets of the taxonomic workflow, but until now, there has been little interoperability between these applications. A main goal of the Platform is to provide an open architecture to allow connection and integration of existing applications and to provide new developments where necessary. Thus, the Platform is not a monolithic application, but rather consists of independent, interoperable components.

Platform components are generally either web services or applications, both desktop and web-based. Several components are designed to be shared by collaborating users within a community, as delimited for instance by taxonomic group or geographic focus. Each community shares a community web store based on the CDM (Common Data Model – see 2.3 below) to store and publish their data and to communicate with the public and other members. The EDIT data portal is the Platform's solution for publishing CDM data via the web, and can be easily customized by the non-technical user. Other community components are modern communication tools like blogs, forums, mailing lists and other collaborative tools based on the content management system Drupal [Dr09].

Other components include central services such as EDIT Map Services – a map generation service – and personal components including special hardware such as a water resistant GPS/GIS handhelds with integrated camera for efficient data acquisition in the field.

2.2 Integrating existing applications and data standards

Establishing interoperability between various existing applications and data standards is a major challenge. Even applications from the same domain of expertise often do not share the same data formats. Rather than attempting to implement data exchange functionality into all Platform-related applications, we decided to create instead central transformation services and data store, whereby applications exchange data either via import-export functionality in all important data formats or via web services.



The CDM Community Store is a repository for every conceivable type of data produced by taxonomists in the course of their work. It supports concurrent editing, and can be interfaced via web protocols by a wide array of applications.

Fig. 1: The interactive EDIT Platform Cybergate gives an overview on the platform and on the all-important question of how dataflow between components can be achieved. (Screenshot from http://www.cybertaxonomy.eu/)

2.3 The Common Data Model (CDM)

At the core of the platform is the Common Data Model (CDM). It covers most important domains of information including taxonomy, descriptive data, media, geographic information, literature, specimens, and person. Wherever possible, it has been made compatible with existing community data standards and thus is strongly influenced by the TDWG Ontology. TDWG, the Biodiversity Information Standards organisation, is a non-profit scientific and educational association to develop, adopt and promote standards and guidelines for the recording and exchange of data about organisms. The CDM was also strongly influenced by the experience gained in 10 years of application development for the IOPI model [Be97] and the Berlin Model [Be03], successive database models dealing with taxonomic core information such as names and taxonomic concepts; and the BioCISE model for natural history collections [Be99].

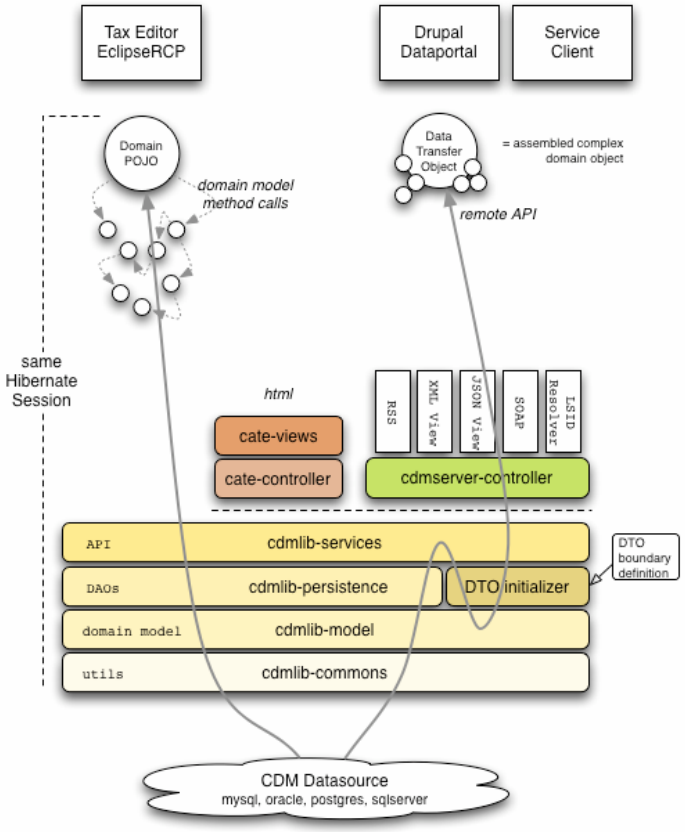


Fig. 2: Important components and libraries of the EDIT Platform.

In a model-driven design process, a Java library was built on top of the CDM. The CDM library is a multilayer architecture based on the Spring framework [Sp09], with a model layer of domain objects, a persistence layer, and a service layer exposing an API. A set of commonly used utility classes are forming an additional layer. The domain objects are mapped to and persisted in a relational database using the Hibernate object-relational mapping framework [RHM09], making it possible to use any number of different database products on the backend. An I/O Layer provides import and export functionality in various data formats, and is the glue integrating the diversity of applications serving the Platform.

2.4 GUIDs, LSIDS & PURLs

In a distributed application or platform where data is shared and available via the internet, unique identification of information is crucial. Globally unique identifiers (GUIDs) permanently identify and access data objects on the Web. TDWG's decision to use Life Science Identifiers (LSIDs) - a widely accepted URN-based identification scheme for the Life Science - is open for criticism [eSIE06], on the grounds that the introduction of LSIDs is superfluous: HTTP URIs are an adequate means of unique identification. Therefore, all identifiable pieces of information hosted in a CDM Store are accessible via a RESTful web service which supports LSIDs but also other GUID systems such as persistent uniform resource locators (PURL).

3 The Taxonomic Editor: a word processor for taxonomic data

The Taxonomic Editor is a desktop application for editing data stored in the Common Data Model. CDM data can either be stored locally for individual use or in a remote datastore for collaborative work. In a typical implementation, changes made to a CDM store are immediately reflected online in an EDIT DataPortal instance. The Editor offers import and export functionality in the standard formats of the taxonomic community in order to provide the interoperability described above. This section discusses the Editor's attempts to preserve the workflow most taxonomists are familiar with while structuring their data to make it readily available to other Platform components and members of the community.

3.1 User acceptance

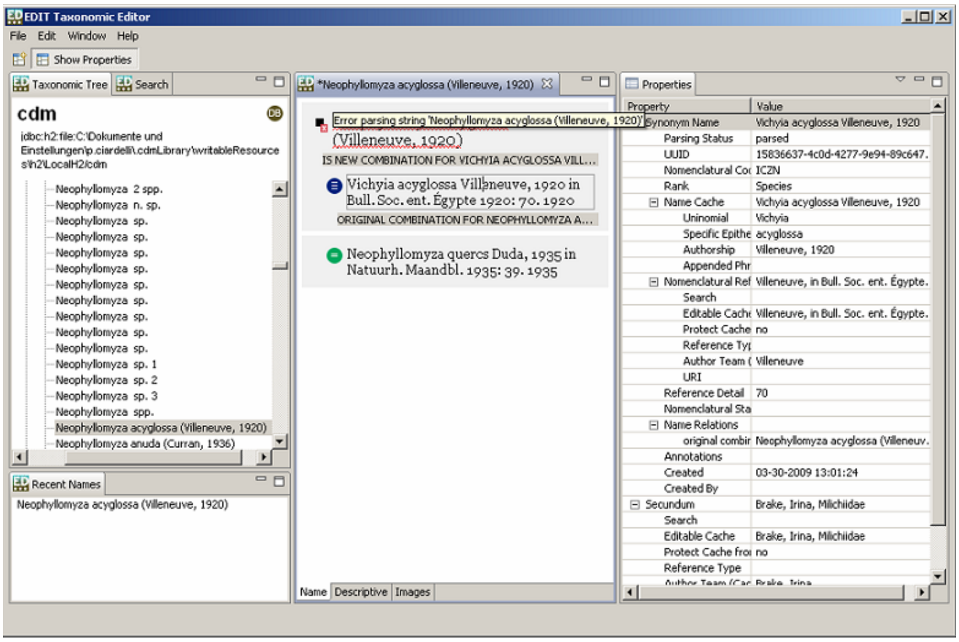
The interface gestures used in the Taxonomic Editor arose largely from experiences with previous generations of taxonomic editing software, where user acceptance posed a significant problem: while the models behind this software generally accommodated the range of taxonomic data edited by the community at large, the methods of input were onerous. Editing programs relying on web-based forms tended to slow the taxonomic workflow when compared to taxonomy done with an unformatted word processing program such as Microsoft Word, and taxonomists by and large reacted with a "thanks-but-no-thanks".

The challenge therefore was to structure the Taxonomic Editor such that input retains gestures from traditional word processors while ensuring data is saved in a structured format. The solution was a three-paned application window, with navigation to the left, a free text area in the center, and an atomized view of the data in the right-hand panel. The user enters data in the free text area, while an on-the-fly parser atomizes the data for storage in the CDM. The goal of the parser is a fairly humble 95% accuracy rate; the great variety of taxonomic communities expected to use the Platform – for instance, the substantial differences between the biological and zoological nomenclatural codes – necessitate a fairly pragmatic implementation. The user can check whether the parser has done its work correctly, and make fine-grained corrections, in the right-hand panel.

3.2 Eclipse Rich Client Platform

A key part of the Editor philosophy is that this interplay between free form input and highly structured storage is only successful when the user is given sufficient feedback as to how his or her input is interpreted by the parser. This is where the choice of technology – the Eclipse Rich Client Platform - comes in.

The Eclipse Java SDK has an extremely rich set of editing functionality that has grown over the course of Eclipse's existence as an open-source project. The Eclipse Rich Client Platform makes this functionality generic, and allows applications that use the RCP libraries to extend this functionality to their own ends. Much of this functionality is extremely attractive from a taxonomic workflow point of view. For instance, the Eclipse SDK keeps track of all Java syntax errors within the current project using a system of markers and annotations; this can be extended to the taxonomic realm to show all instances in the taxonomist's current project that violate community naming conventions. This pre-built functionality make RCP development a far more attractive proposal than DIY web applications; for instance, autocomplete in and of itself is a fairly mature feature in most Javascript libraries, but in the RCP sphere, it ties into annotations, parsing, project resources, and so forth. Autocomplete is an especially important feature for taxonomic work, where scientific names and references are frequently re-used in different contexts, and duplicate detection and removal has traditionally consumed an inordinate amount of IT resources.



cdm

jdbc:h2:file:C:\Dokumente und Einstellungen\p.ciardelli\.cdmLibrary\writableResources\h2\LocalH2\cdm

- Neophyllomyza 2 spp.
- Neophyllomyza n. sp.
- Neophyllomyza sp.
- Neophyllomyza sp.
- Neophyllomyza sp.
- Neophyllomyza sp.
- Neophyllomyza sp.
- Neophyllomyza sp.
- Neophyllomyza sp.
- Neophyllomyza sp.
- Neophyllomyza sp.
- Neophyllomyza sp. 1
- Neophyllomyza sp. 2
- Neophyllomyza sp. 3
- Neophyllomyza spp.
- Neophyllomyza acyglossa (Villeneuve, 1920)
- Neophyllomyza anuda (Curran, 1936)

Neophyllomyza acyglossa (Villeneuve, 1920)

Error parsing string 'Neophyllomyza acyglossa (Villeneuve, 1920)'

(Villeneuve, 1920)

IS NEW COMBINATION FOR VICHYIA ACYGLOSSA VILL...

Vichyia acyglossa Villeneuve, 1920 in Bull. Soc. ent. Égypte 1920: 70. 1920

ORIGINAL COMBINATION FOR NEOPHYLLOMYZA A...

Neophyllomyza quercs Duda, 1935 in Natuurh. Maandbl. 1935: 39. 1935

Property	Value
Synonym Name	Vichyia acyglossa Villeneuve, 1920
Parsing Status	parsed
UUID	15836637-4c0d-4277-9e94-89c647.
Nomenclatural Coc	ICZN
Rank	Species
⊟ Name Cache	Vichyia acyglossa Villeneuve, 1920
Uninomial	Vichyia
Specific Epithe	acyglossa
Authorship	Villeneuve, 1920
Appended Phr	
⊟ Nomenclatural Ref	Villeneuve, in Bull. Soc. ent. Égypte.
Search	
Editable Cach	Villeneuve, in Bull. Soc. ent. Égypte.
Protect Cache	no
Reference Typ	
Author Team (	Villeneuve
URI	
Reference Detail	70
Nomenclatural Sta	
⊟ Name Relations	
original combir	Neophyllomyza acyglossa (Villeneuv.
Annotations	
Created	03-30-2009 13:01:24
Created By	
⊟ Secundum	Brake, Irina, Milichiidae
Search	
Editable Cache	Brake, Irina, Milichiidae
Protect Cache fro	no
Reference Type	
Author Team (Cac	Brake, Irina

Fig. 3: The 3-paned view of the Taxonomic Editor: from left to right, navigation; free text entry; and the line of text which has focus broken down into its atomized components. The first entry could not be parsed; hovering over the annotation in the line's margin displays feedback from the parser.

3.3 User feedback

As the user enters a scientific name, the parser keeps him or her informed on its progress in separating the name into its constituent parts. This feedback is presented in two ways, first, as a red or yellow squiggly line drawn underneath the offending section of text, and second, in the left margin of the central input area, as a warning or error icon. Hovering over either the squiggly line or the icon gives a description of the error or warning. An error could be either a violation of the nomenclatural rules in use (i.e. botanical or zoological), or a taxonomic error, such as a species not sharing the same family name as its taxonomical parent. A warning could be shown when a name shared in different locations within the project is changed by the user, who is thus made aware that this change has repercussions beyond the current document.

The RCP offers endless possibilities for extending user feedback and accelerating the user's workflow. For instance, the so-called "quick fix" feature could be implemented such that when a user changes a shared name, the warning comes along with possible actions, such as "change the name in all the locations it is used", or "create a new name for this one instance". The philosophy of the Editor be summarized as follows: don't burden the user with confirmation popups during input that impede his or her workflow, but make him or her aware that a violation has occurred, and make it as simple as possible to remedy the situation. If in the above example, a user generally wants to create a new name 75% of the time, this should be done by default without forcing the user to choose, but the feedback mechanisms should make the user aware that input was interpreted in a certain way, and that there are other interpretations that can be applied retroactively.

4 The EDIT Specimen Explorer: primary biodiversity data and the Common Data Model

A crucial part of the taxonomic workflow is access to specimen and observation material. Since the international taxonomic community switched to storing data in structured formats like the CDM, observation and specimen databases worldwide have been gradually linked together, forming a huge number of records, each documenting the occurrence of a particular organism at a given location at a certain point in time. A main driver of this process has been GBIF (Global Biodiversity Information Facility), which has been networking worldwide biodiversity data and making it freely available on the Internet since 2001. International initiatives such as BioCASE (Biological Collection Access Service for Europe) and SYNTHESYS (SYNTHESIS OF SYSTEMATIC RESOURCES) share this vision of free and open access to the world's primary biodiversity data, and important precursors and partners to GBIF.

However, searching for this information using scientific names is a fraught process in taxonomy: different literature, institutions, and collections use different scientific names for the same taxon, and some data is accessible mainly via its common vernacular name. This section outlines efforts within GBIF and EDIT projects to provide generic access to diverse data sources, and looks at a thesaurus-based approach to addressing some of the weaknesses of data retrieval in the workflow.

4.1 Drawing and linking databases: distributed network of heterogeneous databases

GBIF harvests and indexes remote datasets shared on the Internet using generic wrappers, with additional processing for quality improvement, internationalization, and so forth.. Through open communication protocols and data exchange standards, the GBIF index overcomes the problem of data heterogeneity and currently lists 174 million occurrence records.

The BioCASE Provider Software (BPS) - the so-called BioCASE wrapper used by the GBIF index - is an XML databinding middleware used as an abstraction layer on top of a database. It is agnostic to the kind of data being exchanged and the underlying conceptual schema, and can be used to set up distributed networks.

Observation and specimen data can be accessed in either of two ways:

- retrieval of a particular record (or a set of records) via the provider's web service (the wrapper);

- access to cached data in the GBIF-index via general or specialized data portals.

4.2 Accessing primary biodiversity data

The BioCASE portal at http://search.biocase.org/europe allows users to search European biodiversity data and derive thematic subsets of the GBIF index [Ho06]. The BioCASE portal uses TOQE (Thesaurus Optimized Query Expander) [Hf07], a generic XML-based web-service. It allows access to thesaurus databases of any kind, and uses a fixed set of methods to hide the complexity of the underlying thesaurus. Results are delivered as well-formed XML documents. The BioCASE portal uses the Euro+Med Plantbase checklist for European flora and the Fauna Europaea checklist for European fauna to expand user queries to include synonyms and related concepts for a given name.

Using thesaurus expansion in the BioCASE portal improves search efficiency by discovering records that may (or in some cases may not) be identified by a related name, but it is an on/off mechanism: it does not allow users to control the query expansion process.

To address this shortcoming, a prototype [Ke08a] for checklist-driven access to collection and observation data was developed within the SYNTHESYS project to improve the BioCASE portal's query mechanism by giving users full control of the query expansion process, allowing them to:

- choose which thesaurus to use;
- choose to include or exclude types of relationships (synonyms, related taxa in the taxonomic hierarchy, related taxonomic concepts such as misapplied names);
- individually mark or unmark “related” names for inclusion in the search.

This prototype became the EDIT Specimen Explorer, a new search portal for taxonomists to explore GBIF data. Available in 11 languages, it provides users with fast and easy-to-use access to worldwide biodiversity data, and offers full control over query expansion. The portal accepts one or more Latin names, suggests related query terms for both zoological and botanical data, expands the query accordingly and offers complete BioCASE portal functionality for resulting specimen and observation unit data.

4.3 Exporting primary biodiversity data

Unit data from the GBIF network can be imported into the user's Common Data Model (CDM) in two steps. First, unit details are exported in either XML or .csv format from the EDIT “Specimen and observation explorer for taxonomists” by click of a button. Second, the file is imported into the CDM via the import function of the EDIT Desktop Taxonomic Editor. Users can then edit this primary data and build their own dataset for a chosen thematic focus using the Editor.

The process then comes full circle if the user chooses to publish his or her CDM data via the BioCASE wrapper, allowing GBIF to (re)harvest the dataset and expose it to the GBIF network. The data is then searchable via web portals such as the GBIF data portal or the EDIT Specimen and Observation Explorer for Taxonomists [Ke08b].

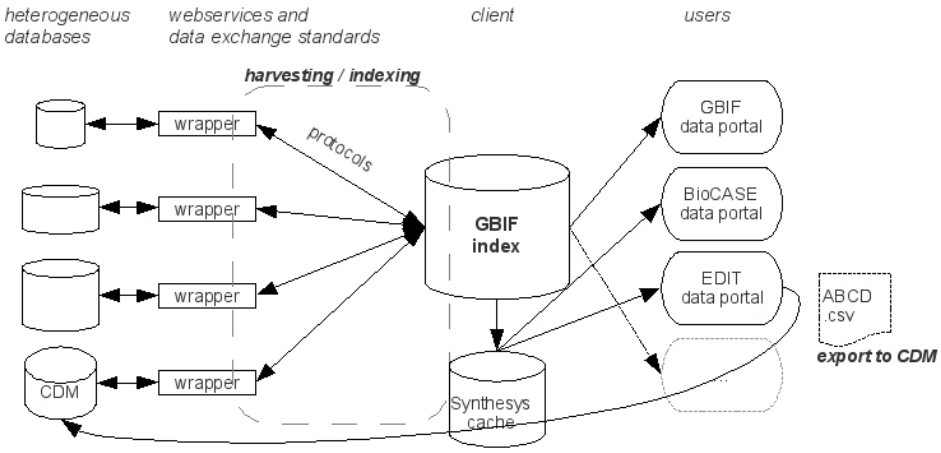


Fig. 4: the flow of data from GBIF into the CDM.

5 Print publishing – print-ready at the click of a button

The final step in the taxonomic workflow is publication, the bread-and-butter of the practicing scientist. This section deals with printed, as opposed to online, publication, which can take the form of taxonomic checklists, monographies, or journal articles.

Data errors are often first spotted when taxonomists have the publisher's proofs in their hands. Traditionally, this represented a bottleneck in the workflow; preparation of these proofs was costly and time-consuming. This section describes an EDIT prototype project to give the taxonomist the ability to prepare a printer-ready copy of his or her work directly from a taxonomic database. The solution had to be flexible, customizable, and exact: several formats require a high degree of precision in the printer's copy; geographic distribution data for instance is sometimes represented using a key that requires spacing on the printed page to be correct down to fractions of a centimeter.

5.1 Case study – creating the Med-Checklist

In 2008, in cooperation with the European "Euro+Med Plantbase" project, the second volume of the Med-Checklist - a synonymic catalogue of vascular plant taxa found growing wild in the countries surrounding the Mediterranean Sea - was ready for publication [Gr08]. The source data was stored in the "Euro+Med Plantbase", a MS SQL database implementing the "Berlin Taxonomic Information Model" (Berlin Model). A team of experts throughout Europe contributed to the database. To make the publication process considerably more flexible, a system was created to generate a printer's copy according to a precise layout, including index and bibliography, directly from the Euro+Med database.

5.2 Rule-based layout with XSL

Every application and data model using such a publishing system should be able to import and export at will. Since XML formats are the de facto standard for exchanging data, it was decided to realize the typesetting process using the "eXtensible Stylesheet Language" (XSL) family. XSL is based on a family of recommendations of the World Wide Web Consortium (W3C), describing a concept for transforming and presenting XML documents.

The XSL encompasses three technologies:

- eXtensible Stylesheet Language Transformations (XSLT) is a language to translate an XML document into another XML document with a potentially different schema or arrangement of data. The purpose of these translations is to bring the original data into a format easily understood by the importing system.

- XML Path Language (XPath) is a language with which parts of an XML document may be addressed.

- eXtensible Stylesheet Language Formatting Objects (XSL-FO) describe how text, images and other graphical elements should be arranged on a page

The typesetting process itself is divided into two steps. Step one transforms the original data into the desired form and specifies the layout for the objects to be formatted. Actual typesetting takes place in step two using formatting software called the "FO Processor" The format of the output medium may vary depending on the chosen FO Processor, but is commonly PDF or PostScript.

The Med-Checklist was published at the end of 2008 using this approach. While a successful proof of concept, this implementation was extremely specialized. The XML data output, the translation templates, and the layout specifications were all created specifically for this project. Although implementation was not especially difficult, it would need to be made much more generic and flexible to fit into everyday taxonomic workflow.

5.3 Generic approach – creating word processor templates from CDM data

EDIT and the taxonomic community would benefit greatly from a generic approach that leaves publication layout completely in the hands of the user. The Med-Checklist project proved that it is possible to transform an XML document as output from a taxonomic database into another, pre-publication XML document. The roadmap is to translate already existing XML export mechanisms into OpenDocument format (ODF) files. ODF is an ISO-certified open standard that has been adopted worldwide by numerous organizations.

This approach has explicit advantages:

- By providing the user with the ability to import her data into a word processor, full control is returned to the user. Nevertheless, it must be made clear that this is a final stage in a publishing process, as changes to the data are not made to the database from which it was drawn.

- Modern word processors like OpenOffice.org work with so-called styles, similar to CSS (Cascading Style Sheets) used in XHTML documents, where a layout is specified for a class of elements. The styles approach gives the user full control over the layout of the data and allows him or her to change it in a consistent manner.

Integrating this approach into the Cyberplatform and putting the ability to create professional, printed publications into the hands of the practicing taxonomist would remove a substantial bottleneck from the taxonomic workflow.

6 Conclusion

This paper has examined some of the ways the Cyberplatform hopes to improve the taxonomic workflow. Foremost among these is to improve the taxonomic community's ability to access existing data by introducing a Common Data Model and making it compatible with existing community standards, offering tools for exchanging data between data models, and improving the tools with which existing data is queried. The manual input and editing of taxonomic data should become more user-friendly and less time-consuming with the introduction of the Taxonomic Editor. Finally, the prototyped print publishing software could remove what has traditionally been one of the biggest bottlenecks in the workflow, the labored back and forth between taxonomist and publisher.

The authors gratefully acknowledge the support of: the EU 6th Network of Excellence Project EDIT (European Distributed Institute of Taxonomy, contract EVR1-CT-2000-40004); the EU 6th Framework Project EU 6th Framework Project SYNTHESYS, contract Riii3-CT-2003-506117); and the EU 7th Framework Project PESI (Pan-European Species directories Infrastructure, contract 223806).

Literaturverzeichnis

[Be97] Walter G. Berendsohn: A taxonomic information model for botanical databases - The IOPI Model. - Taxon 46:283-309. 1997.

[Be99] Walter G. Berendsohn, Anastacios Anagnostopoulos, Gregor Hagedorn, Jasmin Jakupovic, Pier Luigi Nimis, Benito Valdés, Anton Güntsch, Richard J. Pankhurst & Richard White: A comprehensive reference model for biological collections and surveys. Taxon 48: 511-562. 1999.

[Be03] Walter G. Berendsohn, Markus Döring, Marc Geoffroy, Karl Glück, Anton Güntsch, Andrea Hahn, Wolf-Henning Kusber, Jinling Li, Dominik Röpert & Frank Specht: The Berlin Taxonomic Information Model. Schriftenreihe Vegetationsk. 39: 15-42. 2003.

[Dr09] Drupal.org Community Plumbing - URL http://drupal.org/, 2009

[eSIE06] e-Science Institute, Edinburgh: Second International Workshop on Globally Unique Identifiers for Biodiversity Informatics (GUID-2), URI: http://wiki.gbif.org/guidwiki/wikka.php?wakka=GUID2Report. Edinburgh. 2006

[Gr08] Greuter ..Mchl. 2.

[Hf07] Hoffmann, N.; Kelbert, P.; Ciardelli, P. & Güntsch, A.: TOQE - A Thesaurus Optimized Query Expander. Proceedings of the TDWG annual meeting, Bratislava, Slovakia, 2007.

[Ho06] Holetschek, J.; Güntsch, A.; Oancea, C.; Döring, M.; Berendsohn, W. G.: Prototyping a Generic Slice Generation System for the GBIF Index. Pp. 51-52 in Belbin, L., Rissoné, A. and Weitzman, A. (eds.). Proceedings of TDWG, St Louis, MI, 2006.

[Ke08a] Kelbert, P., Güntsch, A., Hoffmann, N., Berendsohn, W.G. 2008: Checklist-driven access to European collection and observation data: Pp. 373 in: Gradstein, S. R. et al. (ed.): Systematics 2008 Programme and Abstracts, Göttingen 7-11 April 2008. Universitätsverlag Göttingen, Göttingen.

[Ke08b] Kelbert, P. et al.: The new EDIT specimen and observation explorer for taxonomists:Pp. 10-12 in EDIT Newsletter 11, ISSN 1962-3402, October 2008

[RHM09]Red Hat Middleware: Relational Persistence for Java and .NET, URI: http://www.hibernate.org/, 2009

[RR07] Richardson, L.; Ruby, S.: RESTful Web Services - Web services for the real world. O'Reilly Media. 2007

[Sp09] SpringSource, URI: http://www.springsource.org/, 2009