



Project no. 018340

Project acronym: EDIT

Project title: Toward the European Distributed Institute of Taxonomy

Instrument: Network of Excellence

Thematic Priority: Sub-Priority 1.1.6.3: "Global Change and Ecosystems"

C5.119 Report on Sustainability of the Software in the EDIT Platform for Cybertaxonomy

Due date of component: Month 46

Actual submission date: Month 46

Start date of project: 01/03/2006

Duration: 5 years

Organisation name of lead contractor for this component: 9 FUB-BGBM

Revision: First Report

Project co-funded by the European Commission within the Sixth Framework Programme (2002-2006)		
Dissemination Level		
PU	Public	X
PP	Restricted to other programme participants (including the Commission Services)	
RE	Restricted to a group specified by the consortium (including the Commission Services)	
CO	Confidential, only for members of the consortium (including the Commission Services)	

C5.119 Report on Sustainability of the Software in the EDIT Platform for Cybertaxonomy

Executive Summary

The EDIT contract and consortium agreement specify that software produced in the project will be put under a public license and be distributed as open source. Sustaining the EDIT Platform thus means supporting the open source environment that allows maintenance and further development of the Platform.

All partners involved in Platform software development are committed to attracting further project funding to improve the Platform, and some such projects are already in place (PESI, SYNTHESYS II, BHL-Europe) and an increasing number of applications are being submitted. Some partners have already declared their intention to make core staff time available to support Platform services or tools.

In JPA5, with the core Platform software reaching maturity, a campaign is being conducted to both popularise and test the usability of Platform software. First results of this can already be seen, for example, some institutions have identified the Platform as the most suitable software for e-Flora development, and the partner hosting the Fauna Europaea Database has declared the intention to move that major checklist into the Platform environment.

However, as has been stated before: it is necessary to distinguish the results of Platform development from other project based (and not reliably sustained) software suites. In taxonomic data management, long-term dependability is a key for acceptance. The modern technological basis of the Platform architecture, confirmed by the ISTC forms a base for long term joint development and support, as well as for versatility and fancy software features. Acceptance of new software is a slow process¹, and confidence in its sustainability is a key issue for decision makers to further invest in it. It is up to the EDIT partners to provide evidence that the EDIT platform is a long term solution that people and institutions can commit to. To be able to hold together the developments under the EDIT label, the future coordination of core elements of the Platform needs to be reliably secured. We had therefore proposed that EDIT announces the hiring of a full time EDIT Platform Software Coordinator² with funding committed for the period 2011-2016. We hope that this proposal can be tabled again with the BoD towards the end of JPA5, when information about the Platform has been spread more widely.

The BGBM would offer in this context to provide the working environment for the Coordinator, integrate and support the Coordinator by in-kind contributions from its Biodiversity Informatics Department, and to host the CDM development environment (hard- and software). This would be flanked by a commitment from the BGBM to maintain and improve the CDM Editor software (with 3 PM p.a.) and to maintain the technical base for the EDIT Cybergate, the CDM Data Portal and the EDIT Specimen Explorer (jointly with RMCA) until 2016. RMCA has also committed to uphold the core components of the GeoPlatform.

No formal commitments have been obtained by now, but will be negotiated, with:

- LIS in Paris, who intend to continue their development of the XPer descriptive tool and may contribute to Platform development from its own resources

¹ For example, it took 8 years of proven functionality to have the BioCASE provider software accepted outside Europe, where it had been (and is) supported and established by a series of projects.

² Responsibilities: (i) Ensuring compatibility of the CDM library with new versions of operating systems and database management systems, and with new versions of used open source software libraries; (ii) ensuring reliability and compatibility of extensions contributed by new projects and partners; (iii) ensuring continuing usability through bug fixing, helpdesk, and minor extensions; and (iv) extension of Platform usage through dissemination activities, cooperation with non-EDIT partners, and assistance in funding applications.

- MNHN in Paris who may use the Platform for their taxonomic checklists
- The MfN in Berlin, who execute the ViTaL development in JPA5, and lead the BHL Europe project for which a close connection to ViTaL has been agreed.
- NHN in Leiden and NBGB in Meise, who consider using the Platform for e-Floras
- RBGK who have previously stated their commitment to the CATE development (which directly uses and contributes to the Platform CDM library development), but who are still evaluating this in the context of their wider ranging institutional IT policies.
- The MfN in Berlin who currently lead the software tasks related to ATBI/M activities.

Full Report

1. Aims and Scope

The EDIT Contract defines a sustained Internet Platform for Cybertaxonomy as one of the outcomes of the EDIT project (see section 4 below). This will have to address both data and software, but software is the focus of this report.

2. Open source software and commercial options

The EDIT Consortium Agreement clearly states: “Software and data standard development produced in the course of the project will be put under a public license” (see below, section 5). This reinforces statements in the Contract’s Description of Work (DoW) that EDIT will produce and use open source software (section 6) as far as possible. This principle has been followed in the creation of software for the EDIT Platform for Cybertaxonomy, taking into account that underlying server operating systems and databases in use at EDIT institutions may also be commercial products.³ Software developed with EDIT funding for the Platform was and will be put under open source public licenses and makes use of software libraries created by the open source community (section 7).

This excludes the option to sell Platform components as commercial software. However, as open source communities demonstrate, there may be options to offer paid-for services such as packaging, training, and implementation of special features (which subsequently have to be released to the public). These options have to be explored over the next years, however, at present they are not expected to cover more than a fraction of the basic cost to sustain the software. The single most important source of funding for institutions taking responsibility for software components will be use and further development within the scope of soft-money projects. However, these institutions will be responsible to bridge gaps in project financing and providing the basic infrastructure needed to maintain the software.

3. What is implied in sustaining the Platform?

4 cross-cutting issues can be identified for Platform sustainability:

- 1. Ensuring continuing usability** (including bug fixing, helpdesk, imports, minor extensions).
- 2. Extension of usage**, dissemination activities and cooperation with non-EDIT partners.
- 3. Ensuring compatibility** (with new versions of operating systems, database management systems, and used open source software libraries; with extensions contributed by new projects and partners; with new hardware, etc.).
- 4. Extension of functionality** (e.g. to cover collection management, to fully support input and analysis of descriptive data, etc.).

Points 1-3 above are essential coordinating activities, aimed at maintaining functionality and momentum, while activities under point 4 need to be addressed by new projects.

³ Institutional policies vary among EDIT institutions and cannot be determined by the taxonomic workflow alone; in many cases these decisions are determined more by the necessities of the institutions administration and public relations, etc. Software developed as part of the EDIT Platform therefore had to be compatible with a range of options that are present in EDIT institutions. This was one of the main reasons to create an entirely new software architecture for the EDIT Platform. In the core Platform, the database access is isolated from the “business logic” layer in the Java-based CDM Programming Library, and that in turn is used either directly through an API (Application Programming Interface) or by means of web services. This allows institutions to freely choose their database management system and operating system, and to build new software based on the functionality in the library.

In July 2009, a survey of the EDIT partners active in WP5 software development resulted in a wish-list with about 4 full time positions distributed among partners to maintain current activities. The Board of Directors has indicated that no immediate support on that scale can be dedicated from institutional core funding while the Platform tools are not yet in widespread use. Instead, project funding was recommended for the time being.

In the following, the present situation with respect to major Platform components is described.

3.1. CDM and Java programming library, developer group coordination (WP 5.2)

Lead: BGBM Berlin

This is the central technical task which provides the base for all products in the Platform. We envision a setting in which a loose group of developers situated at different institutions (project or core staff) program or improve software using the CDM library. To avoid compatibility problems, central coordination of the open source code is necessary. This should be a full position (12 PM p.a.) for a biodiversity informatician, executing the tasks referred to under points 1-3 above.

The **BGBM** offers in this context:

- Hosting the CDM development environment, including the development group tools (such as the TRAC task management).
- Providing the work environment and workspace for the coordination, including technical know-how and support from the BGBM Biodiversity Informatics Department's staff.
- Training of software developers in using the CDM library
- Documentation of the CDM library.

Projects: The Platform and the CDM library has become a central element of the BGBM's biodiversity informatics strategy, so a number of projects or project applications are directly connected to that development, among them PESI (until April 2011, with EDIT partners **MNHN, NHM, UvA**), ViBRANT (lead: NHML, with several EDIT Partners, pending), i4Life (with several EDIT partners, pending) and several pending and forthcoming applications to the German research council (DFG). RBGK's application for the eMonocot proposal and connected development of the CATE software (based on the CDM library) is also very important in this context. The Platform software has also been targeted and budgeted as a component of the LifeWatch infrastructure.

However, the question of central coordination of open source development, including training and maintenance of the documentation presently remains unsolved.

3.2. ATBI&M Tools and Services (5.2.7)

Current Lead: MfN Berlin

As one of the requisites for continuation of the EDIT ATBI activities, the following WP-5 related tasks need to be maintained:

- ATBI-Data management: Data entry, data cleaning (incl. enquiries with researchers), data export (website, BioCASE input, CARDOBS database): 2,5 PM p.a.⁴
- Continued development of the field tool software (incl. updating of the report/manual listing different options and setup of an information desk): 2 PM p.a.
- Further use and requirement definition for Geo-tools: minor costs only.

The entire cost of 4.5 PM would be incurred at the institution maintaining the ATBI&M activity, which is expected to cover costs related to hardware, basic technical system maintenance and workspace for the staff involved.

The ATBI&M activity has been identified as a priority area by EDIT directors. However, the future of the software tasks is directly related to the yet-to-be solved question of which EDIT institution will take the lead in this area.

⁴ The numbers given are calculated based on the experience gained over the past 2 years.

3.3. Geo-Platform (5.4)

Lead: RMCA, Tervuren

The development of the Geo-Platform components have mainly been driven by RMCA and CSIC-MNCN, with collaboration from MIZPAN and other partners. Because of continuing staffing problems at CSIC, RMCA has taken over the activities for the last project year. For the Geo-Platform, the following components should be ensured:

- Maintaining Map services
- Provision and addition of digital contents (i.e. freely available base maps)
- Maintaining the map-viewer interface
- Training activities

RMCA is willing to maintain geospatial components beyond the EDIT project, committing to host the principal site for the services (with other institutions expected to mirror to balance loads), tend to bugs, and provide documentation and training as much as possible.

Projects: RMCA's activities in the context of SYNTHESYS 2, GBIF, and collaborative projects with African partners (funded until 2012) are using the GeoPlatform and contribute to further sustain it. The EDIT Mapviewer was accepted for use in the AEGOS (African-European Georessources Observation System) and CABIN (Central African Biodiversity Information Network) projects. Some further project applications are pending. In the training and documentation context, collaboration exists with the geospatial activities of the GBIF secretariat and the French GBIF node.

We know that other EDIT partners have GIS-related staff, so there is further potential for support which may play a role in future sustenance of this part of the platform. This will be followed up during JPA5.

3.4. Descriptive tools (5.6)

Lead: MNHN/UPMC, Paris

The **LIS** (Laboratoire Informatique et Systématique) will maintain and improve their software Xper2. Xper2 2.0 is now free software of the university Paris 6 with a Creative Commons License. The work on Xper2 is independent of EDIT, but it contributes to the Platform. This software will continue to be freely accessible and downloadable from the LIS server. The import/export SDD/CDM software developed in the context of the Platform, same as the output routines for “natural language” descriptions, form part of the CDM library and are thus in the public domain.

These tools, as well as those which would be highly desirable additions to the Platform functionality (e.g. to compute printable and interactive keys) need to be maintained and adopted according to user needs in the future. 3 PM p.a. are thought to be adequate for these tasks after the end of EDIT's project period, if central library coordination is ensured.

3.5. Virtual Taxonomic Library (5.3)

Lead: MfN, Berlin

The high costs that had been suggested in the draft report in July 2009 have been one of the reasons to discontinue the use of the MetaLib commercial software for the EDIT ViTaL task. After a process of evaluation and mutual consideration, a substantial amount of possible synergies emerged with the Biodiversity Heritage Library Europe project, lead by the **MfN** in Berlin. The catalogue components of EDIT ViTaL are being merged with the respective BHL components, and the ViTaL component of direct access to (sources of) subscriptions of digitised literature incorporated into a joint ViTaL/BHLE development. The ViTaL web portal will access a references index that is planned to be build in corporation with BHL-Europe by the German

common library network GBV (Gemeinsamer Bibliotheksverbund, Göttingen), which handles the relationships with the providers of the proprietary PICA+ Software from OCLC within their existing full-access agreement (this has been one of the major obstacles with the previously used MetaLib product). The index will be part of the BHL-Europe technological infrastructure, so service will be maintained (also for the ViTaL portal) and the index will further be developed after the ending of the EDIT program until the end of the BHL-Europe program. In case of having an open source solution, which is reproducible for other BHL instances, BHL may continue maintenance of the system for the future and even further develop it to be one of the core components of the BHL architecture. Costs for sustaining the updating of library catalogues and subscription information in the ViTaL context still have to be assessed, but will be substantially lower than those previously anticipated.

3.6. The EDITor - EDIT Taxonomic Editor (5.5)

Lead: BGBM

BGBM is willing to commit resources (core staff if necessary) to maintain the EDITor with respect to functionality, training and momentum (an estimated 3 PM p.a. if set in a team with the CDM and CDM Library work). Limited additional input from user institutions will be required e.g. for internationalisation. Implementing additional features (e.g. full integration of descriptive data editing) will require additional resources.

3.7 Security Infrastructure (5.7)

Lead: ISTC

The software in place by the end of the project should cover the applications existing at that time and the needs of the institutions participating. The Shibboleth proxy software is by no means restricted to EDIT applications and has been set up as an open source project. The **EDIT ISTC** will discuss coordination of this activity beyond the EDIT project period.

3.8. CDM Data Portal (5.8)

Current lead: BGBM

BGBM offers to continue hosting the Drupal-based Data Portal Websites in place by the end of the EDIT project period (decided up to now: Cichorieae exemplar group, Palmweb, Euro+Med PlantBase, Trees of El Salvador). Support for portals set up at other institutions needs to be negotiated, especially if additional functionality is required (this normally occurs in the context of new projects with associated funding and may also imply implementation of new functions in the CDM library). As experienced by the Scratchpad group, updating the underlying Content Management System Drupal may be a problem in the future and increase the costs of the implementation of new functionality in the Data Portal. The EDIT consortium should think about a way to **jointly support the maintenance of the various Drupal-based developments.**

3.9. Cybergate

Current lead: BGBM

The Flash-based Cybergate website needs no further technical development but continuous updating of the linked EDIT components and their descriptions. Texts and links needs to be provided by the **partner institutions** responsible for that component. **BGBM** commits to continue hosting the Cybergate page and including the information provided by partners.

3.10. BD Tracker

Current lead for software: BGBM; content: geo-tools: CSIC; descriptive tools: LIS Paris; collection tools: RMCA.

The **BGBM** offers to continue to host this Drupal based software and provide the technical maintenance. However, data provision (the description of non-Platform software useful for

taxonomists) is not included. For descriptive systems, we hope that the LIS in Paris will continue to feed in information. For other types of software (also, e.g., molecular bioinformatics tools and collection management software, both not yet included), **partners** should take responsibility. The organisation for Biodiversity Information Standards (TDWG) has been approached to consider taking over this task in the long term.

3.11. EDIT Specimen Explorer

(Software: BGBM, Implementations: RMCA, BGBM)

This web portal is based on the BioCASE/SYNTHESYS software, connected to the GBIF cache database, and a generic query expansion that allows to connect provider-defined checklists to provide (e.g.) synonyms. It provides an ABCD interface to the CDM to store the specimen and observation records retrieved.

BGBM will continue to support the BioCASE/SYNTHESYS portal software (and the Berlin GBIF mirror) through the SYNTHESYS-2 project (project period ends in August 2013). The system is also used for the GeoCASE Portal at the **MfN** and in the context of several pending applications to the German research council (DFG). **RMCA** plans to continue to use and adapt the software for their African partner institutions through the “Belgian Cooperation”. RMCA hosts a GBIF mirror (<http://gbif.africamuseum.be>) and already pre-installed the infrastructure to have a mirror in Kinshasa, too. The CABIN project uses the same software and BioCASE providers are also used in the context of the STERNA project. RMCA are offering these services both to the staff of their museum and the African partners, so they acknowledge that this is part of the work of the Cybertaxonomy unit and will be maintained beyond EDIT. However, third party funding will be thought after for expansions and major software upgrades.

4. Annex: Sustainability in the EDIT Contract

“The integrating activities in WP5 with respect to the Internet platform for Cybertaxonomy have to lead to durable and sustained work processes and software components within the EDIT infrastructure. This activity aims at the creation of a network of European Service Centres where individual institutions or small consortia take responsibility for long-term help desk function and/or maintenance for software or information systems (digitisation, data provision, data harvesting, and analysis), ... [DoW p. 59]

An EDIT set of software programs and standards, specified, tested and implemented within the project and maintained and enhanced beyond the project's period in the spirit of open source software development, will not only forge an institutionalised European biodiversity informatics community, but also keep the development open for use and contributions through institutions not taking part in the initial network. [DoW p.66]

During the last two years of the project, an activity within WP3 will specifically address the sharing of responsibility for long term maintenance of the software components constituting the Platform. [DoW p. 68]

5. Annex: Software in the EDIT Consortium Agreement

9.4.6 Database and Software Access Rights

The provisions regarding Access Rights granted in this Consortium Agreement shall not apply in respect of any software programs or related source/object codes and documentation produced or used by any Party in the course of the Project.

Software and data standard development produced in the course of the project will be put under a public license.

However concerning Software programs and Database that are defined as Pre-Existing Know-How, the rules concerning ownership and Access Rights as defined in this Consortium Agreement shall remain applicable.

Access Rights to Software that is defined as Pre-Existing Know-How shall not include access to Source Code but only access to object Code. However Access Rights to Source Code may be granted subject to a separate agreement to be concluded between the Contractors concerned. [p. 43]

6. Annex: Software in the EDIT Contract

EDIT will require a strategy for accessing data that is to be protected and this will be developed within WP6 Unifying Revisionary Taxonomy, WP5 and WP3. The data and **software developed through the EDIT network will be freely and openly available** and IPR issues will arise chiefly in relation with proprietary the shared software tools and proprietary or restricted datasets that are brought into the network to form part of the platform [soft]ware developed (see WP5 and WP6, ...). [DoW, p. 32]

.. IPR issues related to software will be developed through interaction between the NSC and the BoD, with advice from an Intellectual Property Use and Dissemination Committee (IPUDC), if any. [DoW p. 43]

Improving production and dissemination of taxonomic knowledge is the *raison d'être* of EDIT. EDIT will address this problem, at its necessarily limited scale, by ...and by implementing the IT tools which will improve communication, collaboration and dissemination. These activities will produce Knowledge ... and it is the intention of EDIT that this Knowledge is free, open-access and disseminated freely as widely as possible. **Software will be produced or converted to an open-source licence model;** where that is impossible, the consortium will use appropriate legal mechanisms to protect them. [DoW p. 44]

An EDIT set of software programs and standards, specified, tested and implemented within the project and maintained and enhanced beyond the project's period **in the spirit of open source software development**, will not only forge an institutionalised European biodiversity informatics community, but also keep the development open for use and contributions through institutions not taking part in the initial network. [DoW p.66]

7. Annex: EDIT Software licenses

Software	Open source	License	Notes
Platform Cybergate (Flash application)	yes	MPL (Mozilla public licence)	MPL allows inclusion of the code in commercial software
Web-publishing component CDM Data Portal Module (based on Drupal)	yes	GPL 2.0 (GNU public licence 2.0)	GPL is a must because Drupal uses this scheme. This excludes commercial use of the code.
BD-Tracker Web Module documenting software of use for taxonomists (based on Drupal)	yes	GPL 2.0	
CDM programming library Java code, including all data access, import and export functions	yes	MPL	
Taxonomic Editor (using the Eclipse framework)	yes	EPL (Eclipse public license)	EPL also allows commercial use of the code
ViTaL Falx reference aggregator (Python code)	yes	Not yet stated	Source released in EDIT subversion system
ViTaL metasearch (using MetaLib)	no	Commercial software	Metalib by Ex Libris Ltd. Use will be discontinued
ViTaL/BHL Reference Index (through German common library network GBV)	no	Commercial sw with established usage agreement support	PICA+ Software from OCLC
GeoPlatform EDIT map viewer Distribution-mapservice; Point-mapservice	yes	Creative Commons BY-NC-SA	http://creativecommons.org/licenses/by-nc-sa/2.5/es/
Geo Platform Itinerary Tool	yes	GPLv2	http://www.gnu.org/licenses/old-licenses/gpl-2.0.txt
Geo Platform Coordinate converter	yes	-	Own licence disclaimer
Single Sign-on and security: Shibboleth	yes	Apache License, Version 2.0	http://www.apache.org/licenses/LICENSE-2.0.html
EDIT Specimen Explorer (using 'cherrypy' webapplication framework software)	yes	MPL BSD (cheerypy)	Cherrypy software based on BSD Licence
XPer Java code	no	Creative common licence	Next version software free for non-commercial use, but not open source
ATBI Database	design yes	uses commercial software	MS Access & Excell
ATBI Field tools for testing: TerraSync (Trimble), ArcPad (ESRI)	no	commercial software	
ATBI - BioCase Provider Software	yes	MPL	