



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”

## **D11 – Version 1 of the EDIT Platform for Cybertaxonomy**

Due date of deliverable: Month 32  
Actual submission date: Month 32

Start date of project: 01/03/2006

Duration: 5 years

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

Revision: Final

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

## 1 Introduction

This deliverable presents a summary of the developments and results achieved with regard to the development of the „Internet Platform for Cybertaxonomy“. As stated before, this is envisioned as a distributed computing platform that will: directly assist taxonomists in their daily work; increase the efficiency of producing taxonomic revisions, checklists, inventories, etc.; and make the results of the work accessible to others without delay. The basic design was described in Deliverable 7 (D7, August 2007), with details provided in a number of „project components“ cited in the D7 deliverable document.

Although the deliverable document presented here will defer technical details to the appended project component documents (underlined text), some explanations are given, adding to the length of the document. This was thought appropriate because of the complexity of the endeavour (29 components, 8 related milestones), the amount of resources directly related to D11, and the need for a comprehensive overview expressed by project participants and other interested parties. The audience for D11 includes software developers from other projects, as it is our explicit aim to attract more projects to join the effort.

## 2 What is “Platform Version 1”?

The first year of EDIT’s IT activity was largely spent setting up the organisation and defining the aims of the Platform project. The evaluation of existing taxonomic software made clear that no existing systems could be widely acceptable as the EDIT standard. As a consequence, a new technical specification was developed, presented to and accepted by EDIT’s Information Science and Technology Committee (ISTC). By the time of delivery of this specification (D7: “Draft technical design of the Internet Platform for Cybertaxonomy”, August 2007), full scale software development was set in motion. An international team of 14 developers employed by the project and actively supported by institutional staff presently works on realising the vision.

“Platform Version 1” comes shortly before the halfway point of the software development phase of the EDIT project. The deliverable represents a first attempt to unite the software developed into a recognisable, coherent structure that demonstrates the entire scope of the Platform.

The main purpose of Platform Version 1 is to demonstrate capabilities, not full usage. EDIT software development aims at a layered, multi-platform, reusable, solid, and multi-purpose system of applications. The temptation to implement quick and dirty shortcuts for demonstration purposes was avoided. As a consequence, much of the time up to now has been spent fabricating the building blocks, and in particular the Common Data Model (CDM) framework, building on the knowledge that was accumulated in more than 20 years of modelling and (more or less) experimental software implementation in taxonomy.

D11 unites and refers to 29 project “components” delivered by 11 partners. Only 3 components were contributed by other workpackages, but many of those delivered by WP5 include contributions from WP6, WP7, WP3, and WP2. We recently added the EDIT Cybergate, an experimental graphical user interface offering a single entry point to the entire Platform.

## 3 Software Components

The software products are at the heart of the informatics developments in EDIT. They are presented here under headings representing the major components developed in Workpackage 5. Sections are related, but not restricted, to actual activities as defined in the WP5 workplan.

Within the sections, a brief introduction is followed by: a description of developments and decision making processes since D7 (August 2007); a description of the current state, highlighting existing and/or imminent products; and a concluding outlook with respect to the planned development until D20 (“Platform version 2”). Special attention is given to existing or

prospective liaison with other initiatives, institutional developments and wider developments in the biodiversity informatics arena. Risks and problems incurred in the development process are treated separately under heading 5 below.

### **3.1 The Cybergate**

The EDIT Cybergate is an experimental attempt to provide access to the entire EDIT Internet Platform for Cybertaxonomy, as well as its underlying technologies, with a single graphical user access interface. This is an additional component not represented in the workplan, resulting from a desire expressed by the ISTC and users to have an overview of Platform developments and products. The Cybergate mainly addresses taxonomists, but also serves as an entry point for more technically interested audiences like developers.

The Cybergate is a graphical browser and widget which facilitates exploration of the EDIT Platform for Cybertaxonomy. Due to the complexity of the EDIT Platform, a sophisticated, graphically rich user interface was needed. Adobe Flash has therefore been chosen as the underlying technology; future versions will be underpinned by Drupal CMS, which will then serve as the back-end for all - and the front-end for some - of the content of the Cybergate.

The Cybergate gives an overview of the platform with its various components, their interactions, and the underlying mechanisms, such as standards and protocols. At the same time, it provides access to products which are already available. Each of the components and interfaces is explained by a short description and a set of links to homepages and to sites for demos, downloads and development. To provide information about non-EDIT tools that could be useful for taxonomists, the Biodiversity Service & Application Tracker website (see [M5.20a BD Tracker launched](#); March 2008) is used. Where appropriate content is available, a link to the BD Tracker is provided directly in the Cybergate.

The Cybergate (<http://dev.e-taxonomy.eu/platform/>) in its current state represents a prototype, and still requires user testing and subsequent adaptation to user needs.

### **3.2 CDM and Java Library**

This is the technical base for Platform-related software development. The *EDIT Common Datamodel (CDM)* provides the data definition for all data covered by applications developed for the Platform for Cybertaxonomy. The CDM is based on community data standards and established models for taxa and names, bibliography, specimens and observations, and geographic occurrence data. The *CDM Library* is a programming library that can be shared by applications using biodiversity data. The applications can be of any kind, e.g. web services, desktop applications, or command line tools. The library corresponds to CDM information structures and manages the data in the *CDM store*, i.e. the underlying database. Business logic shared between applications is part of the library, while application-specific logic is kept separate.

In practical terms, this means that programmers using the library do not have to worry about data integrity rules defined by nomenclatural codes or the like, or how and where and in which database system the data are stored. They can use library functions to, for example, parse a taxon name received as a text string, store the resulting name elements in the database, receive notification about possible conflicts or syntax errors, and retrieve and concatenate the name from the database in order to display it in their application.

By August 2007, a draft version of CDM had been prepared, which has since undergone further refinement. With Version 1 of the Platform, CDM modelling has been finalised for the taxonomic core components and is in (final) draft state for those areas still under development (e.g. descriptive data).

While basic decisions about the technical environment to be used for software development had been made by August 2007, some further specifications were made by the end of the second project year and documented in [C5.55 Platform environment circumscribed based on existing](#)

standards (April 2008). The taxonomic core functionality was finalised (M5.15 Taxonomic core components available for platform (software and CDM specification, Aug. 2008). Web services to access the CDM were implemented and are in use by the data portal (see below); the technical background is referenced in C5.56 Interface to CDM community store implemented and usable within the platform (software, Aug. 2008). For a full technical description of the CDM Framework and the capabilities and limitations of the CDM Library as of Platform Version 1, please refer to C5.57 First CDM Framework Implementation available (Sept. 2008, delayed 2 months due to implementation pressures). This also presents an outlook on the next steps in library development.

The issue of interfacing the EDIT Scratchpads developed by WP 6 with the CDM is discussed in C6.07 Develop a migration path for appropriate content from scratchpads to the CDM (Aug. 2008, 2 months late). The definition of a single fixed interface is not possible due to the highly adaptive and user configurable character of the Scratchpad software, but a number of standards are identified to which both Scratchpads and CDM currently adhere or soon will.

### **3.3 Data Portal**

The CDM Data Portal demonstrates the data output capabilities of the EDIT Platform, as well as the integration of CDM functionality into the Drupal Content Management System (CMS - Drupal was defined as the consensus CMS for EDIT in D7). Following development with the WP6 exemplar groups of mockup sites, a prototype was developed (C5.86 Data portal prototype; May 2008) and content provided by the exemplar groups was moved into CDM stores (C5.87 WP6 exemplar group data migrated to CDM v1; May 2008). In the process, import specifications for the various formats used by the exemplar groups were generated. The prototype also demonstrated the EDIT taxonomic tree browser, which had to be newly developed because no user interface component able to handle vast taxonomic trees could be found. The taxonomic tree can be examined, browsed, filtered, and switched to different conceptual views. The Portal prototype was limited in the content covered, but already demonstrated the “themeable” nature of the interface, i.e. its appearance can easily be changed to present another look and feel. Individual themes were developed for the three WP6 exemplar groups.

In the course of subsequent development, the taxonomists concerned with the WP6 exemplar groups defined a priority list of additional content areas needed for inclusion in the portal. These have already been partly implemented, as they were deemed by the taxonomists as necessary prerequisites before presenting the portal to users outside the exemplar groups. As a consequence, C6.10 Report on discussions of the first version of the Data Portal with the user communities (Aug. 2008) was delayed for three months. The report provides guidelines for further development of the portal, but also highlights the fact that a discussion of the target audience is needed; there are doubts that a single portal will be useful for both the general public (such as that targeted by data aggregators such as the Encyclopaedia of Life) and an audience of taxonomic specialists. The inclusion in the data portal of output for structured descriptive data including identification keys is planned for JPA-4, when the interfaces of the CDM with such data are finalised (see section 3.6 below).

The specification of the content by the exemplar groups and other users made clear that taxonomists want to publish a wide range of data items in their web revisions, including considerable amounts of content already published in print. In a workshop organised by WP6, the results of the IPR task in WP5 were presented (in this context see also C5.23 EDIT Intellectual Property Licence Options (Use Agreements) Appraisal, May 2008). The workshop resulted in a series of recommendations on how to use existing copyrighted materials in web publications and how authors of new treatments can retain their rights. The workshop also specified a number of issues with regard to citation and peer review of web publications. See C6.09 Report resulting from a workshop on web publication issues on IPR (May 2008) for further details.

### **3.4 The EDIT Taxonomic Editor**

The first 18 months period of the EDIT project established the need for sophisticated editor software supporting data input, including, e.g., the nomenclatural details in a taxonomic treatment. It also became clear that no existing tools were adequate as a model for development, and thus it was decided to use innovative techniques to develop an entirely new kind of editor software for taxonomists.

The EDIT Taxonomic Editor is a data input and editing tool, the central software for the taxonomist working on data held in CDM stores. Traditional and time consuming form-based data entry is replaced with word-processor like text input. State-of-the-art parsing technology analyses the free text and disassembles it into its syntactic components without further user interaction. Version 1.0 already covers the basic functionality necessary for entering and editing taxonomic treatments consisting of names, taxa, references, synonymies, and hierarchies, as well as categorized free text information linked to individual taxa (such as statements regarding geographic distribution or type specimens). See [C5.75: V1.0 of the CDM taxonomic editor available](#) (software, Nov. 2008 - 3 months delay) for further details; technical specifications are referenced in [M5.15 Taxonomic core components available for platform](#) (software and data specification, Aug. 2008). The development was delayed by unexpected difficulties in synchronising the different editing panels with the database backend. Version 1.0 is still somewhat wanting in performance and functionality beyond the taxonomic core.

The Editor will continue to be improved by the addition of functionality and performance enhancements almost until the end of the EDIT project. Until Platform Version 2, the emphasis will be on adding functionality, performance enhancements, and full operating system independence (see C5.75 for a list of features). Improvements will be incremental, i.e. intermediate versions (1.1, 1.2, ...) will be released as new functionality is implemented and stable.

Version 2 will also include the functionality necessary to handle different taxonomic concepts and their interrelationships. A workshop was held in collaboration with WP6 to define the requirements, and a wiki site was set up for continued discussion: [C5.74 / C6.8: Specification of concept-related workflows to be supported by the taxonomic editor](#) (wiki site, Sept. 2008).

### **3.5 Geo-Platform**

The development of geographic applications for taxonomists was one of the prominent points in JPA-1 of WP5, but was critically delayed due to administrative problems. Since D7, which did not focus on geographic applications, much of this initial vision has finally been realised. It covers not only the anticipated needs of the publishing taxonomist (especially a free map library – see <http://edit.csic.es/GISdownloads.html> – and a tool for freely configurable distribution maps for web and print publication), but also paves the way for applications that will be useful for the WP7-ATBI sites and for the analysis of distribution data in biodiversity research. The considerable interlinkage of the individual components, the relatively high number of institutions involved in the development process, and the necessity for intense communication to ensure the component's incorporation in the platform led to delays and a more or less simultaneous delivery of many of the components just in time for D11 (which makes the component reports somewhat repetitive). A short but comprehensive description of the MapViewer site and services, including examples, is provided in [C5.41 Documentation for GIS webapplication available](#) (Sept., delivered Nov. 2008), as well as in briefer form in [C5.66 Regional distribution map generation webservice for online and print quality maps released and online available](#) (June, delivered Nov. 2008). A more technical description, including an outlook on further tasks (now partly accomplished), is provided in [C5.64 Global background maps available in print quality](#) (March 2008); the Application Programming Interface allowing developers to include EDIT functionalities in their Web portals is referenced in [C5.68 API Documentation for map generation webservices available](#) (wiki site, Oct. 2008). The prerequisites for print-ready maps were summarised in [M5.28 Report](#)

on Guidelines to Produce Publication ready maps (Sept. 2008), and the state of the implementation was reported in C5.67 Map generation webservice enhanced for Publication ready maps (print quality) (Sept., delivered Nov. 2008). The state of the implementation of GIS services depicting point locations are further described in C5.65 Point (occurrence) map generation webservice for online and print quality maps released and online available (software; May 2008, implemented with delays, report delivered Nov. 2008); it includes the description of the tool developed to depict itineraries, which accepts a number of additional data input formats not yet covered by the MapViewer application.

Feedback from partners has been received and will lead to further improvements of the tools, ameliorating their accessibility for individual taxonomists as well as for software developers. Incorporation into the WP7 ATBI site portals is under way, and their use within ongoing taxonomic projects will be demonstrated in the Fauna Iberica and Euro+Med PlantBase portals.

### **3.6 Descriptive components**

Descriptive tools are software applications that allow some or all of the following functions: the input and editing of structured descriptive data; the subsequent generation of description texts; fixed and interactive identification keys; and data analysis based on descriptive data. Descriptive tools are necessarily part of the Internet Platform for Cybertaxonomy, and structured (and unstructured) descriptive data are accommodated by the Common Data Model. However, an analysis of existing tools (C5.50 Report on integration of existing descriptive tools into the Platform, Aug. 2008) resulted in the assessment that several of them have already been adopted by biologists, and have proven their value in numerous applications and taxonomic keys. The European project Key2Nature, with which EDIT collaborates, will also contribute to promote the interest in such systems and to distribute descriptive applications for multiple audiences. This adheres to the original philosophy of EDIT WP5, that existing, widely used tools should be made interoperable with the Platform. A standard exchange format for descriptive data already exists: the TDWG-SDD (Structured Descriptive Data) XML schema, which builds on the older DELTA standard (Descriptive Language for Taxonomy). SDD was developed in cooperation with the developers of several existing tools.

As a first step in EDIT development, structured descriptive data were included in the CDM and further refined in the course of the development of an SDD import and export module for the CDM library (C5.79 SDD Import/Export module for CDM v.1, software; and – with more details - M5.18 SDD import export module functional, Nov. 2008. The export part is in preparation).

Like other data stored in the CDM, descriptive data will be accessible for display in web pages, e.g. in the EDIT Data Portal and in the EDIT Taxonomic Editor. Accessibility will be achieved using an API developed for the CDM store and will make use of tools customized specifically to offer new content visualizations (e.g. by using a tool transforming structured descriptive data into readable texts, i.e. “natural language” output, and interactive keys). Accessibility for other applications, e.g. Scratchpads, will be realized using the export functionality.

Descriptive tools are currently used by the Exemplar Groups for taxon descriptions and keys, e.g. the program DeltaAccess (Diversity Descriptions). This will provide the base for further testing of the incorporation of this kind of information. One EDIT partner (MNHN/UPMC) is using and developing descriptive software based on Java, which has the potential for direct integration into the CDM framework library.

### **3.7 Bibliography (ViTaL)**

ViTaL (the Virtual Taxonomic Library) is a set of tools designed to help taxonomists locate and use taxonomic literature more easily. Preparations for the development of this component of the Platform were well underway by the time of D7 (August 2007). The results of an analysis of

requirements and of available software solutions were used to present a design document and implementation plan. It was decided to use the commercial software product Ex Libris for the implementation of a metasearching mechanism for literature (Ex Libris Metalib) and for the linking-through to available full text content (Ex Libris SFX). The main further functions of ViTaL are the aggregation (harvesting) of taxonomically relevant references from taxonomic sites and the development of a fast search mechanism for these, as well as a mechanism to notify digitisation projects for taxonomic literature about references in use or requested by taxonomists but not yet available in digital form.

The ViTaL metasearch service based on MetaLib is designed to provide ViTaL users with the ability to search the catalogues of the EDIT partner libraries, which include many of the most significant collections of taxonomic literature available. It also offers results from other library-defined resources and Google Scholar searches. All EDIT libraries were contacted (M5.23 All EDIT institution libraries contacted with invitation to participate in ViTaL, March 2008) and asked to make it possible to include their catalogues and subscriptions in the metadata service. The response to this call (see M5.24 Details of Z39.50 catalogue resources provided to the ViTaL team (NHML) by the initial set of actively collaborating institutions; Nov. 2008) was encouraging: up to now, not only the initial set agreed by in all 10 EDIT members have answered and 9 were able to provide OPAC details, which will make it possible to include their catalogues in ViTaL.

User search results include bibliographic and location information for hard copy access and loan requests, and also include links to ViTaL's OpenURL resolver (see below), giving access, where possible, to the full-text of articles for download. Difficulties encountered with the support of the commercial software resulted in a slight delay in component C5.29 Pilot of Metalib functional for human access (Sept. on-line, fully functional in Nov. 2008), but the service is now available under <http://vital.nhm.ac.uk/V>.

The OpenURL service with SFX provides ViTaL users with links to full-text articles where these are available for free download or via subscriptions belonging to their institutions. It also offers other library-defined resources, including the online public access catalogue (OPAC) holdings and Google Scholar searches. A functional prototype of the service for inclusion in web portals such as Falx (see below) is available: C5.27 Pilot website for OpenURL service (Sept. 2008).

At the same time, the portal for machine and human access to the metadatabase was also launched (see C5.28 Portal prototype functional for machine access to Metadatabase; Sept. 2008). These are the software components making up the bibliographic aggregation and search application, known collectively as Falx. For the purposes of the pilot, Falx harvests the bibliographies of EDIT scratchpad sites using their BibTeX download facility. The bibliographic data is stored in a MySQL database (the meta-database). This content is then used to build a full-text searchable index using Apache Lucene, a Java indexing application. Finally, the index is exposed to the web for both human and machine use at the URL <http://taxonlib.org/bibsearch>.

### **3.8 Specimen access**

Accelerated access to specimen information and to taxonomic literature are probably the two most pressing issues if the aim is to increase productivity of the taxonomic work process. Specimen access is covered by the GBIF system and services cooperating with GBIF, such as those elaborated by the BioCASE and SYNTHESYS projects in Europe. However, for JPA3 the need for a specimen access service specifically tailored for taxonomists' needs was defined and action taken, first with an analysis of available taxonomy-oriented institutional access portals (C5.89 List of data items and searchable data items in existing portals, May 2008). Since the new portal was to harness BioCASE technology, and the BioCASE portal software is based on the GBIF index, the next component described options to improve that database for taxonomists' use: C5.90 Inclusion of missing data items in GBIF index (June 2008). The respective desiderata

were communicated to the GBIF Secretariat in Copenhagen and will be considered for future versions of the GBIF Index.

The product developed in accordance with identified taxonomists' requirements (M5.21a Operational specimen web portal; Aug. 2008) is the EDIT Specimen and Observation Explorer (<http://search.biocase.org/edit/>), a web portal for taxonomists that provides access to data in the GBIF system. Queries for names can be expanded using user-selected taxonomic thesauri (taxonomic checklists, such as Fauna Europaea and Euro+Med Plantbase) and selected name relationships contained therein (e.g. synonyms). The query results obtained can then be imported directly into a CDM data store (C5.91 CDM storage functionality implemented; Oct. 2008). Programming in the respective activity has been concluded. However, further taxonomic thesauri need to be included to fully harness the tool's potential. Talks are under way to include botanical databases with global scope (EDIT Partners Missouri Botanical Gardens, Royal Botanic Gardens Kew); additional zoological databases of that kind still need to be identified.

### **3.9 Security**

A common security infrastructure within EDIT (with the possibility of extension into the biodiversity research community at large) was already planned at the time of D7 (August 2007) and a preliminary decision was taken to use the Shibboleth standard-based application framework to implement a single-sign-on mechanism (SSO) and identity providers within the EDIT network. Simply put, the software developed will allow developers to adapt their biodiversity informatics applications to become part of a large network, to which users can sign on with a single username and password.

Development of the applications began with a software specification (C5.82 Specification of CSSO profiles for the platform covering the use cases web application, web service and desktop application; June 2008); subsequent setup of the infrastructure (C5.83 Setup the initial CSSO security infrastructure for the platform; Aug., document in Sept. 2008) including the installation and configuration of an identity provider, service provider and public key infrastructure; and the partial integration with the EDIT Experts Database (C5.84 Integrate the CSSO Identity Provider component with the Experts-DB, Aug. 2008).

A common security infrastructure would be a significant step towards the integration of institutional IT infrastructures. The results were presented to the ISTC committee, members of which are responsible for actually implementing the SSO mechanism at the institutional level. The solution was welcomed and will enter into testing at several EDIT member institutions. Acceptance was helped by the fact that the Shibboleth framework is already widely used among university and other service providers (for example, the JStore Journal Storage archive is using Shibboleth). In addition, this activity in EDIT is currently contributing to the planning of the proposed LifeWatch initiative, a large-scale biodiversity research infrastructure in Europe which will also need an SSO mechanism into which EDIT could be integrated.

## **4 Integration**

Institutional and domain integration are central aims of Network of Excellence projects. EDIT members agreed that modern data and information processing offered a golden opportunity for synergetic collaboration among institutions and taxonomic researchers and thus integration. We think that D11 demonstrates that this assumption was basically correct, and that considerable progress has been made. Ultimate success will depend on broadening the support within the EDIT partnership, but also within the taxonomic and biodiversity informatics communities at large.

D11 shows a great number of integrative cross-links, both between institutions and between work packages. While WP5 is driving the process, there are direct links with other workpackages:



- With WP7 through the development of field tools (a component of D20, i.e. Platform Version 2), by the integration of the geoplatform and the ATBI field sites, and with the construction of Drupal-based ATBI web portals.
- With WP6 through close cooperation with the Exemplar Groups, as well as through the complementarity of Scratchpads, which have a potential of opening up further communities, and which also use the Drupal CDM.
- With WP3 through the development of taxonomic infrastructure (with the PESI project as one of the most promising developments for Platform adoption) and cross-links with GBIF.
- With WP2 through software development and potential “prospecting” of users in the context of the experts database (another interface implemented in Drupal).

Institutional collaboration is on the one hand exemplified by the fact that 10 member institutions are directly involved in the programming effort. Platform developments are also closely watched by the wider community of institutions represented in the ISTC, which has been decidedly supportive of the workplan. The direct incorporation of Platform tools into the institutional IT structure is currently considered by several member institutions, ranging from adopting the CDM and Taxonomic Editor for checklist work, or connecting the institutional taxonomic backbone to the EDIT Specimen Explorer, to technical issues such as implementation of Shibboleth. All this is in spite of the fact that most EDIT software products are still in early to late beta state. It seems that the chosen technologies and implementation pathways offer an opportunity for broad acceptance in the biodiversity informatics community. Several international projects have stated their direct interest in the platform: the CATE (Creating a Taxonomic E-science, <http://www.cate-project.org/>) project in Britain will be based on the CDM and CDM library; in PESI the European flora and the fauna checklist dataset will be moved to the EDIT platform during 2009; Fauna Iberica will use the EDIT Geoplatform; and the global nomenclator projects have expressed their direct interest in using the CDM for a common names infrastructure (a recent development, based on an EDIT D11 component elaborated by WP3: [C3.2.1 Contribution to the Global Name Architecture \[GBIF/ECAT white paper\]](#); April 2008). The potential here lies in the large number of contributors and editors of taxonomic content, as well as in the substantial development experience and resources in these projects, which would give the Platform the momentum necessary for open source development.

## 5 Risks

Software developments using new technologies, or using innovative approaches in implementing known techniques in a new and complex domain such as biodiversity information, incur the risk of unforeseen technical problems. EDIT Platform development was no exception, but the resulting delays were kept to a minimum (first version of the Editor, map generation software, Data Portal). Outsourcing the development is not necessarily a solution, as demonstrated by the frustrations with MetaLib support, and resulting delays in the ViTaL activity. However, technology decisions have been based also on current knowledge about the reliability of techniques and products, so that this risk is not considered a serious challenge to the project's success.

The complexity of the domain in itself is a risk in that software developers must have highly specialised interdisciplinary knowledge – they must be biodiversity informaticians. The loss of staff presents a formidable risk of failure for a project with restricted time and staff. Such problems have caused considerable headaches within EDIT, as exemplified by the need to find a replacement for the leader of the WP5 development team, by delays due to staff shortage in portal development, and by the difficulties encountered in the recruitment for ATBI tasks. Fortunately, it was possible to minimise effects by relying on the now considerable breadth of biodiversity informatics activities in EDIT institutions.

User orientation is another factor; insufficient communication with users in the software development phase may lead to lack of acceptance of the products. Fortunately, all EDIT development centres are directly located in institutions forming part of the primary user community, and the EDIT Exemplar Groups, with their links to the community, reinforce this. However, up to this point most programming efforts have been focussed on preparing the technical base for the products; things like the CDM modelling and the Java library are only becoming interesting to a wider public now that actual products based on them are appearing. In JPA-4, user testing of Platform products will play a prominent role.

Communication and proper project planning within the EDIT developer team is also not always optimal. With the team distributed over 10 institutions, coordination of development becomes a serious challenge. Adding to this is the close cooperation with other projects mentioned above. The risks were minimised by implementing a project management tool, by focussing more time of the leader of the developers group on coordinative tasks, by attempting to minimise travel time of developers, by exchanging developers among institutions, and by the introduction of elements of modern software development strategies for communication and progress control within the developer group.

Finally, the question of sustainability needs to be addressed. Long-term sustainability of the EDIT Platform will depend on its acceptance, both within and outside the core EDIT community. A few EDIT institutions declared their interest in maintaining software or service components beyond the project period, and a central helpdesk email has been established ([editsupport@bgbm.org](mailto:editsupport@bgbm.org)). Another strategy to ensure this is the incorporation of the EDIT platform in further projects; a first step towards this was done with PESI, and for some aspects LifeWatch (of which EDIT was one of the initiators) may play a role.

In summary, we think that the risks are calculable. Our goal in 2009 is to demonstrate that the foundation achieved during the first half of the EDIT software development period can now be fully harnessed to improve the now publicly available products in their performance, scope, reliability and usability.