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"

C5.097 First report on integration of software developments in partner institutions

Due date of component: Month 41
Actual submission date: Month 41

Start date of project: 01/03/2006 Duration: 5 years

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

Revision: final

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)	

The EDIT platform for Cybertaxonomy is a suite of loosely coupled tools that already exist or are newly developed. The development of new tools and the development of adapters for the integration of existing tools take place in several EDIT partner institutions. The main partners involved here are RBG Kew, NHM London, MNHN and UPMC in Paris, CSIC Madrid, RMCA Tervuren, MfN, BGBM and FUB-Inf in Berlin, MIZPAN Warsaw, RBG Edinburgh (testing), UKBH Copenhagen. The coordination of the EDIT software development is carried out by the BGBM (task 5.2) and reported and discussed in the ISTC (Information Science and Technology Committee).

Coordination

A distributed development process requires substantial resources for coordination in order to ensure continuous joint efforts in the execution of the common objectives of the EDIT project in general and the specific objectives of Platform development. Coordination within the developer group must ensure that all partners are informed about the overall developments in the Platform at least as far as relevant for them. This includes information about new technical developments at partner institutions, frequent reminders about existing plans, information about moving targets, ascertaining the interoperability of developed software components, and planning of future developments. Existing knowledge is thus exchanged throughout the partnership with the effect of reduced research work and improved product quality.

The coordination has also to make sure that developers engage in an intensive exchange with potential users of the developed products, in order to feed back their experience and needs into the development process.

Apart from the necessary investment in time, a fundament for the success in the above mentioned tasks is a well working communication infrastructure as well as tools for collaborative development work such as code sharing and bug tracking.

Communication structures

Several communication structures have been set up during the first 3 years of the project to meet the requirements of the different partners. They include

- **Mailing-Lists:** Four mailing lists have been set up to discuss issues related to general development (dev-edit), modelling (cdm-edit), geographic tools (wp5-geo-edit), and CDM Library (dev-cdmlib). The mailing lists are intensively used. Some of the lists are strictly limited to persons directly involved, because there is a certain reluctance to intensively discuss implementation details and other internals with too wide an audience.
- **Meetings:** Over the past two years, 22 meetings (some back-to-back) took place with participation from more than 2 member institutions (meetings held during the first 18 months of the project essentially served to set up the Platform architecture). Categories:

  A general EDIT developers meeting was held in September 2007 in Berlin (http://dev.e-taxonomy.eu/trac/wiki/DeveloperMeeting2007) to get to know each other and to discuss and inform about general project objectives

  Subgroup Meetings: Tasks with a distributed developer team like 5.4 (Geotools) met regularly to discuss open issues and future development. Alternatively e-Conferences were held to discuss open issues online. The development coordinator took part in some but not all of these meetings depending on the topics discussed.

  Coordination meetings: The development coordinator met regularly with developers of tasks centred not at his home institution (WP 5.3, WP 5.6, WP 5.7, WP 5.10 and WP 7 – ATBI

sites) to discuss the current state of development in the work package as well as in the partner institutions, in order to synchronise developments.

Several meetings to were held to assess developments at partner institutions and to coordinate Platform developments with them.

In addition, conferences like e-Biosphere and TDWG were used for intensive discussions with other partners present.

Developers from Paris MNHN, Kew, MfN and Paris UPMC spend periods ranging from 2 days to a week in Berlin to work with the Platform team.

- **Emails & Skype:** Emails and chats are still the most commonly used way for daily information exchange
- **Ticket system:** The Trac system (see below) is used to inform about new or completed tasks or action points. It also serves as an archive for the most important information items exchanged via email or any other tool.

Development Environment

The most important tools supporting collaborative development are:

- **Developers Wiki** (http://dev.e-taxonomy.eu/trac/wiki). This is used to document information shared among developers. This includes for example software specifications (e.g. http://dev.e-taxonomy.eu/trac/wiki/EditorRequirements) as well as useful hints (e.g. http://dev.e-taxonomy.eu/trac/wiki/BestPractices).
- **Ticket system:** The bug-tracking tool Trac is used to manage open tasks and documenting completed work. By creating and updating tickets all relevant partners are informed (via email) about new changes and newly arising bugs and feature requests as well as comments on existing tickets. The ticket system is now intensively used by most developers. It proved to be highly valuable for the development coordinator, to stay up-to-date and to prioritize tasks according to the overall needs of the project. The ticket system is not available to the public.
- **Subversion:** All software code and technical documentation generated by EDIT developers is kept in a common version control system that has been setup at BGBM. The system allows to share code and to resolve problems when code developed at different places has to be merged. The subversion system is also used as a backup tool.

Coordination Tasks

Besides setting up the development environment and communication structures multiple tasks had to be performed by the coordination staff.

- **Defining specifications and interfaces:** In a software development project as complex and diverse as the one here undertaken, individual developers often lack a deeper understanding of the integration of their component in the entire Platform. Defining specifications and the interfaces between components thus became an important coordination task (in close cooperation with the involved partners).
- **Developer support:** During the development process documentation sometimes lags behind. For jointly used components like the CDM Library API developers thus need direct support from library developers, that has to be mediated by the development coordinator.
- **Testing and feedback**: Newly developed products and functionality in general need intensive testing. Besides capability testing this includes especially completeness testing to

ensure that the product does not only work properly but also fulfils the requirements of the Platform and the specification. This often leads to extensive discussions about the necessity of missing functionality and bug fixing. Technology support is needed when the developer's the skills, knowledge or resources cannot adequately address the requirement at hand.
Structured testing is now taken over by task 5.10 and intensive communication with 5.10 staff will be necessary. Task 5.10 will only be able to test human userinterfaces. Testing of machine readable interfaces still needs to be coordinated by task 5.2.

- **Prioritizing:** Requirements have to be prioritized to make sure that core functionality including functionality needed by other partners is addressed first within the development process. Priorities have been intensively discussed with the partners and are often organized by using the Trac ticket system.
- **Contact to users:** Crucial for the success of a product is appropriate user feedback during the development process. Regular meetings and intensive email exchange has taken place with users like the WP6 exemplar groups and others. Requirements have been discussed. Where requirements could not be incorporated into the developed tools in the foreseeable future (due to missing resources or due to other reasons) this has been explained to the user.

Integration of pre-existing software at partner institutions

- **RBGK, NHML** (and University of Oxford): There are some projects that do use platform components or are planning to do so in the future. The most important among them is CATE (Creating a taxonomic e-science, http://www.cate-project.org/). After a process of intense discussion and adaptations the CATE software is now fully based on the EDIT Common Data Model (CDM). Joint development of the CDM Java programming library is ongoing.
- **RMCA:** Pre-existing GIS developments are being adapted to the EDIT Geoplatform.
- **CSIC:** Geoplatform tools are being integrated into the museum's biogeographic research focus.
- **UvA:** The Fauna Europaea Database housed in Amsterdam has been moved to the CDM for output to the PESI (Pan European Species-directories Infrastructure) project's web interface.
- **MNHN:** Several taxonomic databases at the Museum in Paris have been mapped to the CDM to facilitate later transfer of the data to the Platform.
- **UPMC:** The development of the functionality for structured descriptive data (task 5.4) is closely linked to the ongoing development of the Xper2 Software.
- **BGBM:** The Euro+Med database and the El Salvador checklist hosted at the BGBM have been imported into the CDM, the latter is already using platform tools for editing.