



EUROPEAN DISTRIBUTED INSTITUTE OF TAXONOMY (EDIT)

WP5: Internet Platform for Cybertaxonomy

INFORMATION SCIENCE & TECHNOLOGY COMMISSION (ISTC)

2nd Meeting, September 10-11, 2007

Botanic Garden and Botanical Museum Berlin-Dahlem

MINUTES

Participants: Agnes Kirchhoff (BGBM), Alain Empain (NBGB), Andre de Muelenaere (NBGB), Anna Weitzman, Anton Güntsch (BGBM), Christoph Häuser (SMHN), Eduard Stloukal (CUB), Eric Danon (RBINS), Franck Theeten (RMCA), Henri Michiels (MNHN), Jean Van Onacker (NBGB), Malte C. Ebach (BGBM), Marc Brugman (ETI), Mark Jackson (RBGK), Michal Lipka (MIZPAN), Miklós Rajczy (HNHM), Patricia Mergen (RMCA), Paul Richards (NHML), Martin Pullan (RBGE), Mickael Graf (SMH), Robert Tolksdorf (FUB), Simon Chagnoux (MNHN), Simon Mayo (RBGK), Walter Berendsohn (BGBM)

Proxy participants: Eduard Stloukal for IBSAS, Walter Berendsohn for MO

Apologies: Chuck Miller (MO), Karol Marhold (IBSAS), Markus Döring (BGBM), Regine Vignes (UMPC), Yde de Jong (UvA)

1. Welcome, Introduction of participants

2. Agenda

The agenda was accepted as proposed, with an additional point: discussion of “WP5 achievements by the end of the project” as drafted by the WP5 team.

3. Goals and Terms of Reference of the ISTC

The draft Terms of Reference document prepared by WP5 was acceptable to ISTC, but in the course of discussions on point 5 of the agenda it was decided to prepare a new document (Rules of Procedure) incorporating the organisational points that were included in the draft MoU template.

4. EDIT and Workpackage 5 – the first 18 months

Introduction

W. Berendsohn described how WP5 has set out during the first 18 month period of the project to:

- analyse the pre-requisites for the establishment of co-operative processes by inventorying institutional IT capabilities and processes
- model the information domain and particularly the work processes
- provide first practical applications for taxonomists in a rapid prototyping approach
- draw up a definition and design blueprint of the “Internet Platform for Cybertaxonomy”
- assess existing tools with respect to their integration into the Platform
- cooperate with the other WPs, particularly with WP6 and WP7
- liaise with other projects active in the field

Virtually all these processes were hampered to a varying extent by administrative obstacles, staffing problems, lack of resources on the institutional level, lack of communication, and pre-existing agendas. More information about EDIT and WP5 can be found at <http://wp5.e-taxonomy.eu/blog/>

Survey of Resources

Patricia Mergen gave a summary of her forthcoming report, a survey of EDIT partner IST resources and organisation. She revealed that:

- 45% of EDIT partners do not have a centralised IST service.
- Low bandwidth and core capacity of some EDIT partners may presently exclude them from accessing some of the planned EDIT WP5 platform tools.
- Larger institutions spread over different buildings or departments and are more likely to outsource IST services to third parties.
- Most institutions are in charge of their own IST infrastructure but may rely on outsourcing IST services or resources.

The general impression is that – within their limits – IST personnel is ready for and often keen on collaborating within EDIT.

Modelling the taxonomic work process

Mark Jackson summarised the process modelling effort carried out by David Taylor over the past year and also reported on the modelling of inventorying by Imre Kilian. The process model, based on interviews with taxonomists, has proven that the differences of the work processes between different subdisciplines are minor, so that it is feasible to reach out to the entire taxonomic community with a unified set of tools. The process model is not intended to serve as a direct base

for software development. However, the results document the current work practises in taxonomy and may further serve as an education tool (it may be incorporated into WP8 vision for a School of Taxonomy) and as a “background” structuring interfaces for the Platform (so that taxonomists can identify the tools and resources offered for a defined step in their work process). The revisionary and ATBI models will need to be published in a referred journal for the taxonomic community. It was noted that Anna Weitzman (USNM) and Chris Lyall (NHML) have produced a taxonomy process model similar to the unified model.

Rapid prototyping & “The Platform” - definition process

Walter Berendsohn outlined the development over the last year that led to the new definition of the Platform. As a result of discussions the approach has changed from focussing on interfacing existing applications to devising and partly implementing a new architecture based on state-of-the-art technologies. The architecture will also support the interfacing of existing applications and will help to make them interoperable, as well as providing a medium and long term migration path for existing applications.

BDTracker and Scratchpads

Malte Ebach briefly introduced these EDIT tools or services. The BDTracker is a website where software and resources can be described, revised, and annotated by the public. It offers taxonomists a way to find and compare software and communicate their experience.

Scratchpads offer a Website where taxonomists can communicate and jointly work on their data. Both tools are implemented using the Drupal content management system.

In the ISTC there was some concern over the role of scratchpads and their apparent “competition” with the platform, and how easily content built up in a scratchpads can be ported or integrated with the platform.

Cooperation and communication – SAC and Evaluation

W. Berendsohn reports on the critical assessment of EDIT activities brought forward by the SAC (Scientific Advisory Committee) and the Commission’s evaluation panel. One of the principal difficulties of the first project period was (and partly is) the lack of understanding of the function of a Network of Excellence project. As we have been told in very clear words by the EU evaluation, a NoE is about integration, not only collaboration, and it is not a research project. The ISTC needs to consider the consequences of the evaluation, and ways how to address institutional integration in an effective way with the resources available.

5. Moving towards the NoE (1): Memoranda of Understanding

The outstanding performance indicator for a Network of Excellence is the degree of “Integration” reached by the members. Evaluations have strongly criticised the project for not achieving enough in this area. WP5 proposes to establish MoUs which individual partners can join where they agree on sharing resources. A draft MoU template and two examples for topics were discussed.

MoU Template

The ISTC discussed the draft template and accepted the approach proposed by WP5, i.e. to establish an agreed MoU template document, in which a paragraph (currently paragraph 4) can be inserted with text about a specific subject matter. Template and subject matter text together form the individual MoU and are then signed by the institution. I.e. there will be a series of MoUs each identical except for para 4, and institutions will be invited to sign up to whichever they can. It was also agreed that the draft template itself should be revised, so that

- the role of the ISTC as the governing body of the MoUs during the project period needs to be clarified
- the template points to a separate document defining the Rules of Procedure (RoP) of the ISTC, the corresponding parts in the MoU are removed
- in paragraph 2 the line that states the number of EDIT partners should be left out, since this may change over time
- all acronyms should be written out in full.

The MoU will be revised and a RoP will be written based on the Terms of Reference document. RBGK has offered to show the revised MoU and the RoP documents to their legal department for informal scrutiny before a draft is circulated to all EDIT participants.

MoU content examples

Some institutions objected to the example concerning use of Drupal as a common CMS, because they already use another system. Berendsohn pointed out that the agreement referred only to taxonomic information systems as defined by paragraph 1 in the template, moreover, no institution will be obliged to sign while they cannot be convinced of the advantages of maintaining a common code base for content management. Nobody objected to having scientists using a Drupal based website at another partner's institution.

6. Moving towards the NoE (2): Common Data Model and Platform Architecture

Anton Güntsch presented the current state of architecture design and development plans. Important issues discussed by the ISTC include: **Sharing a Common Data Model.** EDIT WP5 has drafted a common data model for the Platform. We are working with the UK (NERC) funded project CATE (Creating a taxonomic e-science) in order to share such as model. **Platform Security:** Central Authorisation / Single Sign On. A central EDIT authorisation service would further integrate EDIT institutions through the use of a Shibboleth federation to reach a distributed, trusted user data management and access. The criteria for such a service will be discussed. **Hosting Services:** Bibliography, Images and Community Sites. The Platform will require several shared hosting services especially for documents such as large images and publications (i.e. pdfs). The EDIT platform, and web taxonomy in general, promote taxonomic community sites that need to be hosted. Hosting such sites is difficult for many smaller institutions given the lack of available resources. Some progress has already been made as to institutions taking on a responsibility. For example, the NHML has expressed willingness to provide a proportion of the recurrent funding for the ViTAL system in the longer term (i.e. licence fees for the Ex Libris Group applications).

The ISTC raised several points:

- Institutions will have to provide an IST service (i.e. installation etc.) to provide access or implement Platform components. Smaller institutions that don't have IST support will require an EDIT help desk (comparable to the SYNTHESYS help desk). – The intention of the project is that institutions will take on responsibility for defined components, so that long-term sustainability is ensured.
- Will current efforts be compromised by being incompatible to the EDIT platform? Should not the platform be compliant to existing systems rather than these systems integrating into the platform? – Implementation of platform components can be seen as a process depending on available institutional resources and the needs of the taxonomists in the institution. Uncoupled platform components can individually be integrated into existing systems (e.g. data conversion services). In any case, some platform compatibility can be reached by standardised import and export routines. The platform will not replace existing specimen management systems.

- EDIT and CATE? – The projects are working together to address the issue of overlapping development and interoperability.
- EDIT and TDWG? – The platform development provides many direct contributions to the TDWG architecture. The CDM is directly based on the TDWG ontology.
- Does agreeing to the platform mean acceptance of the CDM? – In the long term, yes, otherwise full interoperability cannot be achieved. However, at this point CDM structure and definitions are not fixed. A frozen version of the CDM will be released by end of 2007 to allow for software development. The next version will then not be released before end of 2008 and is open for changes.
- The platform's success depends on whether institutions and taxonomists adopt it. – Yes.
- The EDIT platform needs a single entry point (i.e. Homepage or other URL) to be presentable as a product. – This will be addressed.
- The platform should be the major deliverable of EDIT and should not sell itself short. The interoperability of EDIT is highly dependent on resources, which will be critical after the project funding ceases (2011). Therefore EDIT should be more forthcoming in PR (i.e. be more ambitious in outreach activities). – WP5 will try to contribute to this. WP5 could present the platform architecture at larger institutions in order to get feed back from institutional IST teams (i.e. a return visit to RBGK), however WP5 developer time is limited, although developer training workshops are being considered.

The ISTC approves the further development of the platform provided that EDIT will not expect existing systems to change and integrate into the platform immediately.

7. Revision of the 2nd JPA

W. Berendsohn reports from the Network Steering Committee meeting (12 Sept. 2007) on progress towards a second work plan acceptable to the European Commission. This will include the deprecation of the existing extensive list of deliverables to internal deliverables (“components”) and the formulation of few project deliverables. Of these, one corresponding to WP5, the publication of the draft architecture, has already been delivered. No comments were made by the ISTC.

8. EDIT Certification

Malte Ebach introduced different models for certification that may be applied in EDIT. The ISTC has the following recommendations

- The Platform is presently too under-developed for a certification process.
- On a practical level, EDIT should start with Self Certification on BDTracker (i.e. compatibility of existing programs and services to platform components).
- Certified products tend to be older and are not favoured over newer products. Therefore it is necessary to encourage users to embrace certification.

9. Preparing for the 3rd JPA

Development was impeded to a certain extent by the need to react to evaluation and redesign of the 2nd JPA. The following outline is the current state of planning in existing

WP5.1 (Integration)

- ISTC will meet once a year. The next meeting is tentatively scheduled for October 2008 (depending on when TDWG is scheduled).
- Promote WP5 and services such as the BDTracker to bioinformaticians and taxonomists.

- Contribute to further develop TDWG standards and collaborate in TDWG interest and task groups. WP5 does not intend to develop standards in parallel to TDWG.

WP5.2 (Platform development)

The CDM warrants more activities dedicated to developing it.

WP5.3 (ViTaL)

- NHM (BHL) needs to communicate with INOTAXA and TDWG in order to come to some agreement as to interoperability as well as to standards. Currently the BHL is producing unstructured data that needs to be standardised in some way.
- NHM (BHL) interaction with EDIT is unknown to INOTAXA. Julius Welby needs to discuss EDIT tasks with Chris Lyall (INOTAXA) in order to prevent parallel development and create synergies.

WP5.4 (Geo-services)

Patricia Mergen, who is now coordinating this activity, presented an outline of plans for the 3rd JPA. Outreach to other partner's GIS departments or activities is a central component. No comments were made by the ISTC.

WP5.5 (Taxonomic core)

The tasks of 5.5 should now concentrate on the conversion of data complying to different information standards (i.e. TDWG-RDF to the CDM).

WP5.6 (Descriptive components)

ISTC recommendations:

- Separate key generation tools from descriptive tools.
- Interactive keys are more in demand than traditional keys.
- Many descriptive packages are not integrated with many taxonomic packages. An integrated package therefore would be a big step forward and will be appealing to taxonomists. -- At least in the long term, EDIT will develop an integrated descriptive/taxonomic tool.

WP5.7 (Security components)

One or two institutions start to generate a pre-configured environment and become part of the Shibboleth federation and to have more extended applications for service providers.

WP5.8 (Publication components)

No funds from the EDIT WP5 core are assigned to this task presently, and scratchpad developments provide at least some of the necessary steps. However, web publication is progressing in close collaboration with the exemplar groups of WP6.

WP5.9 (Specimen component)

This activity is also not funded by EDIT, but clearly the necessity for a GBIF/BioCASE interface exists that is specifically optimised for the taxonomist as the user. A linkage to GBIF is included in this activity as well as further linkages with the planned SYSNTHSYS 2 project.

WP5 achievements by the end of the project (February 2011)

WP5 has been asked by the SAC and reviewers to present a vision of achievements to be expected by the end of the project. The following points were discussed and approved by the ISTC:

- Memoranda of Understanding (MoUs) have been formulated covering all major areas of taxonomic information and software implementation and use. All EDIT partners have signed a number of such MoUs.
- A common architecture (i.e. Common Data Model, shared services, interfaces and implementation technologies) for taxonomic IT applications has been agreed and implemented or prototyped, and serves as a reference standard for the participating institutions.
- Taxonomists internationally use components of the EDIT platform for cybertaxonomy (i.e. either adhere directly to the platform, are interoperable, or at least semantically compatible to it).
- Individual EDIT partners or third parties have committed to provide future support for defined components of the Platform.
- All platform applications and services are accessible through an EDIT portal with an intuitive user interface that follows the taxonomic workprocess in guiding the taxonomist towards the needed resources.