



## EUROPEAN DISTRIBUTED INSTITUTE OF TAXONOMY (EDIT)

### WP5: Internet Platform for Cybertaxonomy

#### INFORMATION SCIENCE & TECHNOLOGY COMMISSION (ISTC)

#### 4th Meeting, September 15-16, 2009

#### Minutes

**Present:** CBS, Utrecht: Gerrit Stegehuis; CUB, Bratislava: Eduard Stloukal; FUB-BGBM, Berlin: Anton Güntsch; HNHN, Budapest: Miklós Rajczy; MfN Berlin: Andreas Hammer; MNHN, Paris: Simon Chagnoux (for Henri Michiels); NBGB, Meise: Jean van Onacker; NHM, London: Chris Sleep (for Adrian Smales); NNM, Leiden: Marian van der Meij; RBG Edinburgh: Martin Pullan; RBG Kew: Mark Jackson; RBINS, Brussels: Cathy Emery (for Eric Danon); RMCA, Tervuren: André de Mûelenaere; SMNS, Stuttgart: Angela Jandl; UPMC/MNHN, Paris: Régine Vignes-Lebbe; UVA, Amsterdam: Yde de Jong.

**Apologies:** FUB-Informatics, Berlin: Robert Tolksdorf; IBSAS, Bratislava: Karol Marhold; MNHN, Paris: Henri Michiels (represented); Missouri Botanical Garden, St. Louis: Chuck Miller; NHM, London: Adrian Smales (represented); RBINS, Brussels: Eric Danon (represented);

**Observers:** Abigail Barker (RBGK)

**Reports:** Lisa Banfield (RBG Edinburgh); Pepe Ciardelli (BGBM, Berlin); Andreas Kohlbecker (BGBM, Berlin); Alexander Kroupa (MfN, Berlin); Patricia Mergen (RMCA, Tervuren); Andreas Müller (BGBM, Berlin); Lutz Suhrbier (FUB, Berlin).

**Chair:** Walter Berendsohn; **Minutes:** Agnes Kirchhoff.

## Day 1

### 1. Welcome, introduction of participants

### 2. Agenda

No additions to the agenda

### 3. Version 2 of the Internet Platform for Cybertaxonomy

**Introduction:** Walter Berendsohn gave an overview over the current situation of work package 5. Good progress was made over the last year, Platform version 2 will be released in time by the end of September. Problems incurred were mostly related to staff changes.

**The Platform on the Internet (WP5.2):** Andreas Müller presented developments in the EDIT Cybergate, the Common Data Model (CDM) and the Java library. The EDIT Cybergate provides access to the entire EDIT Internet Platform for Cybertaxonomy, as well as its underlying technologies, with a single graphical user access interface. The EDIT Common Data Model (CDM) and the CDM Library provide programmers with a comprehensive set of functionalities that can be used in diverse applications related to taxonomic data.

**ATBI and Monitoring field tools (WP5.2):** Alexander Kroupa presented developments such as a new data entry form (excel format), mobile devices (ArcPad, ESRI), a new compression format for map data (jps) and custom forms for mobile devices. Customised forms will be offered to users, open access maps (e.g. of Russian origin) can be provided to researchers; the captured information is stored in external dbase files.

Discussion: Digital photography of lower quality can be integrated into the portable field tool. It was asked if the field tool is compliant with EDIT GIS tools and standards such as INSPIRE: The data are independent from ArcPad or other native formats and can be exported to any format. The data will be served in the GBIF system. Species lists are stored in dbase files for retrieval in the field, new species can be added directly only via proxy (e.g. 'species1'). A CDM based export of species lists from the taxonomic editor is planned.

The ISTC members have no knowledge of parallel developments.

**Bibliographic tools (WP5.3):** Walter Berendsohn gave status report about the Virtual Taxonomic Library (ViTaL) component. Personnel problems at the NHML require rearranging the work programme. This has been discussed with the NHM library staff. Metalib and SFX (commercial ExLibris library software), is running and works. Metalib facilitates joint access to libraries and digital resources, but at present only the libraries from Kew and NHML are linked to the catalogue. Parallel developments in Biodiversity Heritage Library (BHL) and ViTaL should be synchronised. Apparently, two harvesters are developed, one with ViTaL, and another with BHL international.

**Geographic tools (WP5.4):** Patricia Mergen presented the developments of the Geoplatform

Discussion: For Platform version 3 a mirror server of the geo-services should be installed at RMCA, Tervuren. At present it is not planned to restrict the use of the services to EDIT members only. The strategy is to find more international partners willing to host a mirror. The EDIT services should be open to everybody to reach sufficient acceptance. It is important that

the GIS services are compliant with the INSPIRE directive (INfrastructure for SPatial InfoRmation in the European Community).

**Taxonomic core and Taxonomic Editor (WP5.5):** Pepe Ciardelli demonstrated the latest version of the Taxonomic Editor.

Discussion: Automatic identification of duplicates: at present this function is not implemented in the Editor, but a duplicate search function is available in the CDM library. The Taxonomic Editor in a multi-user environment: presently, a basic locking mechanism has been implemented, which will be further elaborated in the near future; for Internet access database connections can be used, a secure solution using a terminal server has been tested; the CDM library includes version history functions for multi user purposes (development driven by CATE). Use of different area standards: the Editor as well as the library is capable of mixing different standards for geographic data (e.g. TDWG, Euro+Med).

**Security Infrastructure, single sign on (SSO) (WP5.7):** Lutz Suhrbier explained the developments for the security infrastructure and called for more users of the SSO technology either as Service Providers or Identity Providers.

Discussion: Certificates: ways to issue certificates have to be examined by the institutions themselves; there might be different possibilities depending on the type of institution (e.g. university, or associated to a university computing centre) and on the country.

**Keys and Descriptions (WP5.6):** Régine Vignes-Lebbe demonstrated the current developments of descriptive tools and the exchange format for the EDIT Platform.

Discussion: Focus of developments on the SDD format, plans to incorporate other formats like Scratchpads: as Scratchpads are at this time not providing highly structured data, this will depend on future developments.

**Publication support (WP5.8):** Andreas Kohlbecker showed new developments in the Data Portal V2.0 and explained how the planned publication tool will work to turn CDM data into print publications. The data will be exported from CDM to XML format and transformed via XSL into a structured preformatted OpenDocument Text (ODT) file.

Discussion: Who to create the publication design and layout: the CDM will supply several layout themes, which can be selected, but for individual output the ODT file needs to be edited; an administrative interface of the CDM portal will provide layout support; HTML/CSS programming of the layout is also possible for people familiar with these web-languages.

The CDM Data Portal makes use of Drupal version 5 whereas Scratchpads use Drupal version 6 for a closer integration with Encyclopedia of Life (EoL). At the moment upgrading to version 6 is not scheduled for the CDM DataPortal.

**User testing (WP5.10):** Lisa Banfield explained testing methods and the approach to software and usability testing in EDIT. Some funding is available to several partners to support testing, but this will start in 2010.

#### 4. Use of standardised evaluation procedures for platform components

Patricia Mergen presented CASCADOSS (<http://www.cascadoss.eu/en/>) and STERNA (<http://www.sterna-net.eu/index.php/en/>). STERNA emphasises the implementation of an extensive evaluation plan. CASCADOSS is a framework for software evaluation.

Discussion: CASCADOSS evaluation could be used for marketing as an argument for further funding.

It was suggested to carry out an evaluation of Platform version 3 for GIS tools in Sept. 2010. No developer time should be used for this. The evaluation should be coordinated in close cooperation with activity 5.10 (software testing). It was agreed that we should now focus on testing. An evaluation could take place later in JPA5.

#### 5. Integration and sustainability

- Progress report

MoU on use of ViTaL: Walter Berendsohn noted that more libraries of EDIT institutions should be connected to ViTaL. Presently, it is not clear, why the Vital/Library component is not more active.

WP5 MoU are accessible via: <http://wp5.e-taxonomy.eu/blog/mou>

- Further ideas for inter-institutional agreements

Discussion: Walter Berendsohn explained the digitisation strategy with priority on Metadata capture for collections. Collections are sorted according to taxonomists needs but the biggest user groups are outside taxonomy. Indexing would open new applications and funding opportunities. Specific technical help is needed for different collection types. An MoU would be on technical collaboration for achieving this goal. Action should be taken by natural history collections worldwide. EDIT institutions could act as starting points. The IST Commission takes a leadership role on technical side.

Summary: There is no new ISTC memorandum arising from the suggested topics.

#### Day 2

##### 5. Integration and sustainability (cont.)

- Platform software licensing schemes and overview of open source programming libraries used

The software licences for the EDIT software were agreed. Up to now, no licensing scheme has been chosen for XPer<sup>2</sup>. MPL (Mozilla Public Licence) is recommended, which allows commercial use of the software.

- Sustainability and maintenance of Platform components

Walter Berendsohn reported on directors meeting (20<sup>th</sup> July 2009. Paris): the directors draw special attention to the sustainability of EDIT, to muster institutional support, Platform tools must be in use by taxonomists by the end of JPA4. The governance of products should be considered.

Walter Berendsohn pointed out that WP5 will run out of budget by August 2010. Some continuation is enabled through other projects such as PESI and BHL-Europe. The BGBM will take the responsibility for the Editor and Portal, but the commitment of other institutions is necessary to secure continued development and maintenance.

Open-source software development in a community as small as ours should be supported by coordinating staff at EDIT institutions.

The participants discussed the benefits of moving the code from BGBM SVN to a more visible system such as SourceForge. It would indicate the appearance of EDIT and additional functionalities such as forums could be useful. However, the transition from SVN to another system will be complex. The possible efforts for this transition should be evaluated. Experience of CATE could be taken as an example.

The setup of a backup site (mirror) was suggested: Edinburgh volunteers to host a mirror. It was agreed to have all source code in the central repository.

It was agreed that read access to SVN should be open for everybody, but write access should be restricted and coordinated.

A forum and steering panel for the coordination of new requirements was requested. Governance with decisive power is needed for critical decisions (e.g. for Drupal upgrades, fundamental CDM changes). A constant person from one institution should organise this process and the consortium members should support the funding of the person by the end of next year for 5 years.

It was suggested to integrate EDIT in the business model of LifeWatch. For LifeWatch depending on national commitments it is not yet clear how far the national financing scheme will work. EDIT and LifeWatch have only partly overlaps: EDIT is focussed on the development of applications at institutional level whereas LifeWatch is more focussed on a networking level.

Are there plans to install platform components in participating ISCT institutions? Yes, if the researchers at the institutions require this from the IT departments. The products should be introduced to the researchers. CBS will continue to use its own system. Kew is in a decision making process at the moment.

## **6. Preparing for the 5th Joint Programme of Activities**

**Activity 5.1** Integration of IST departments; over-all WP5 coordination.

Walter Berendsohn asked the ISTC members, who would be interested in an EDIT platform introduction session (training, demonstration) at their institution.

The following ISTC members expressed interest: RBINS, Tervuren: beginning of January; SMNS, Stuttgart: January; RMCA, Tervuren: not before the documentation is finished; MNHN, Paris: January; NNH, Leiden: not before January; MfN, Berlin: December or January; NHM, London: January.

Some sessions could be organised in joint events: Belgium (RMCA, RBINS, NBGB); Berlin (BGBM, MfN); London (RBGK, NHML).

The Demonstration sessions could be synchronised with the recruiting for usability testing.

Volunteers for usability testing are asked to contact RBGE. RBGE will recruit testers at TDWG.

The ISTC institutions are invited to consider how new projects can collaborate with EDIT to derive additional synergies.

The Biodiversity Service & Application Tracker (BDTracker), a collection of links to software, tools and resources useful to taxonomists, needs to be updated and extended now. It was discussed if the BD tracker should be handed over to TDWG or GBIF. Walter Berendsohn will ask TDWG for their requirements to move the BD tracker.

#### **Activity 5.2** Platform Implementation and Data Model

Developer tutorials: It was discussed if there is need for more developer tutorials, and for more non-English tutorials. Tutorials can be offered in other languages only if the translation is funded with additional resources. Patricia Mergen will try to follow up translations of demonstrations and documentations in a separate project. Local workshops should be organised by EDIT members.

#### **Activity 5.3.** Bibliographic tools

The ViTaL programme is under revision and to be connected with BHL and BHL-Europe.

#### **Activity 5.4** Geographic tools

A video explaining the GIS components is created.

A map services cache is prepared by Pere Roca Ristol in Madrid (CSIC-MNCN). Mirroring becomes more important for the stability of the system with growing number of users.

There will be more support of the geo-mailing list with the new IT person in Tervuren.

It was recommended to Pere Roca to meet the GIS staff from NHM London, RBG Kew and RMCA Tervuren.

The code-base of the GIS software tools is not included in the subversion repository at present. Reporting in one environment would be helpful for bug-tracking. It should be aimed for better ways of integrating the geographic tools into the CDM components (activity 5.2.)

#### **Activity 5.5** Taxonomic core:

For the Taxonomic Editor, specimen-related functions should be improved.

#### **Activity 5.6** Keys and descriptions:

The display of descriptive data will be extended. The implementation of keys has to be discussed. Natural language descriptions will be provided in cooperation with exemplar groups and a PhD student in Paris.

A joint meeting with Key2Nature is planned for September 2010.

User documentation and guidelines are planned.

#### **Activity 5.7** Communication, security infrastructure

The ISTC members are invited to let the FUB-Inf know if they want to join the single sign on system.

#### **Activity 5.8** Publication tools

There is a priority on the development of the portal installation kit and the support of print publication.

RBG Edinburgh has already developments in this direction. A user friendly tool would be helpful. Pdf format output would be useful as well as predefined formats. E-Publishing should be supported too.

A layer for pre-processing between the library and the output process is planned. It would lessen the dependency on programmers work, but the concept for this is not finished yet.

**Activity 5.9** Specimen access

This activity was finished in 2008.

**Activity 5.10** Software and usability testing

See presentation and discussion above.

**7. A consolidated ISTC for all CETAF-related projects?**

SYNTHESYS II members, who are not in EDIT will be invited to join the ISTC.

Should the ISTC be a CETAF structure? Objections from ISTC institutions should be sent to Walter Berendsohn.

The ISTC members support the continuation of the ISTC beyond the project period. The ISTC was received positively in evaluations and by the directors. ISTC members think that the committee could make sense for the following reasons: communication of the IT processes, sharing responsibilities, inform about software developments, discussing IT strategies, planning at the local level, information on running international projects, early warning about new systems being developed, see demos of existing systems, present activities within institutions, inform about duplicate developments, feedback on experience with the use of EDIT products, agreements on standards and co-operations.

**8. Next meeting**

5. and 6. October 2010, at the BGBM (dates of conflicting meetings need to be checked).