

Minutes of the joint CETAF Digitisation Group and ISTC Meeting, *Bratislava* 12 April 2016

Executive Summary

The joint CETAF ISTC (Information Science and Technology Commission) and DWG (Digitisation Working Group) meeting took place in Bratislava, 12th of April 2016. The agenda is publicly available at

http://cetafdigitization.biowikifarm.net/cdig/ISTC_Meeting_Spring_2016_Bratislava.

Presentations as well as this minutes will be made available on the same website.

The meeting was attended by 18 participants from 11 CETAF member institutions. Purpose of the meeting was to report on important activities of the ISTC and DWG, to identify fields for a closer cooperation, and to plan concrete activities for the next year.

Action Items

- To organise two hackathons/workshops: 1) tutorial for smaller CETAF institutions implementing stable identifiers; 2) exploring links to external LOD resources (Anton Güntsch with support from CETAF secretariat).
- To try to find a better name for the “Stable Identifiers Implementers Group” (All).
- To test the draft best practices for semi-automated georeferencing developed by Naturalis (voluntary CETAF/ISTC members).

Participants

Anton Güntsch, Dominik Röpert (Berlin, BGBM); Gregor Hagedorn (Berlin, MfN); Matúš Kempa, Karol Marhold (Bratislava); Jakub Bělka, Lukáš Bělka, Katarína Gatialová (Prague); Karsten Gödderz (CETAF Secretariat); Elspeth Haston, David Harris (Edinburgh); Simon Chagnoux, Stuart McLellan (Paris); Heimo Rainer, Christian Steinwender (Vienna); Wilfred Gerritsen (Leiden); Cathy Emery (Brussels); Ari Taponen (Helsinki)

Adoption of agenda

Report from CIO meeting

Results of the CIO meeting held on the 11th of April were briefly summarised. The outcomes will be communicated in separate minutes. It was agreed that regular CIO meetings should be continued.

Stable identifier initiative, report and discussion of next steps

Anton Güntsch gave a presentation on the status of implementations of stable HTTP URI identifiers in CETAF. 11 CETAF members have now implemented the identifiers, which are increasingly used in publications and data portals. Roger Hyam (RBGE) implemented a highly useful URI tester which can be used to test existing identifier implementations (<http://herbal.rbge.info/>). Further activities include the specification of a “CETAF Specimen Preview Profile (CSPP)”, which aims to harmonise RDF elements provided by different institutions (<http://cetafidentifiers.biowikifarm.net/wiki/CSPP>) and the “Wallich Catalogue” pilot integrating distributed digitised herbarium specimens using stable identifiers and the CSPP (<http://wallich.rbge.info/>).

Activities for the coming year were discussed. It was agreed that priority should be given to

1. Broadening the group of implementers within CETAF specifically addressing smaller institutions.
2. Exploring the possibilities of linking out to external resources with a focus on scientific names, place names and person names.

Dynamic use of CETAF information

Karsten Gödderz presented developments related to the CETAF website and plans for improved exploitation of passport information. This was seen by the CETAF secretariat as the first step on the path to establish the CETAF website as a main hub for information concerning natural heritage. The next steps could include:

- a job portal
- a database of experts and expertise
- a joint query tool

In future, there would be some support available for workshops, and the development of a cost action could be undertaken as a network activity.

The website has now a prominent link to an information page for the identifier initiative. Part of this page is a simple registry of implementers which aims to encourage institutions not yet implementing the system to join (<http://cetaf.org/cetaf-stable-identifiers>). It was agreed that the name of this group (Stable Identifiers Implementers Group) should reflect a special status. Ideas for a new name should be sent to the secretariat. Karsten will add Bratislava to the current list and suggested advertising the event. It was also suggested that the website could integrate the CSPP information and the use case (Wallich Catalogue) to give them greater prominence among members.

The participants discussed the idea to set up an information system with information about scientists working in CETAF institutions and their expertise. Based on experiences made during the EDIT NoE, there were serious concerns that maintaining such a system would require substantial resources, which CETAF does not have at present. KG will report back to the CETAF secretariat.

ISTC initiatives - reports and discussion of potential for collaboration

The following presentations about important ISTC initiatives were held:

- 1) Die Herbonauten (Dominik Röpert): a collaboration between MNHN, BGBM and RBGE, which aims to set up a citizen-science platform for transcribing herbarium sheets based on software developed by MNHN.
- 2) Herbadrop (Simon Chagnoux): an EUDAT-funded collaboration of Naturalis, BGBM, MNHN, Digitarium, and RBGE together with CINES which aims to set up a joint pilot system for archiving and OCR herbarium sheets.
- 3) Visions for harmonised digital collections - prerequisites, integration, migration, challenges (Heimo Rainer): a presentation about the results of a “harmonised digital collections” workshop and the potential for improved interoperability by using appropriate vocabularies.
- 4) Results of the geo-referencing initiative (Wilfred Gerritsen): A draft best practices for semi-automated georeferencing developed by Naturalis based on a survey of existing technologies used by CETAF members and beyond.

The presentation slides will be linked to the meeting website at [http://cetafdigitization.biowikifarm.net/cdig/ISTC Meeting Spring 2016 Bratislava](http://cetafdigitization.biowikifarm.net/cdig/ISTC_Meeting_Spring_2016_Bratislava).

Naturalis will send their draft best practices to all ISTC members. In a next step, the recommendations should be tested by CETAF member organisations. Naturalis will set up the technical platform for the tests.

Access to digital collections and Digitisation on Demand

As part of the SYNTHESYS3 project, access to digital collections is being reviewed to ensure that content will be available in open access portals. Part of this work will assess different policies across partner institutes.

Access discussion included:

- whether institutes had existing policies on access to digital collection data
- issues included ignorance about copyright and the suppression of Red List and other sensitive data
- guidelines may be helpful with a view to developing a CETAF standard which institutes could follow if desired

With regard to digitisation on demand, most institutes operate some level of service at present, although the structure and content varies. Some institutes carry out 3D digitisation including CT scans on demand and may charge for it, but it was felt that it was not necessarily a good business model, and it could result in a delay in making the image more generally available.

Elspeth Haston informed the group that a short questionnaire on access and digitisation on demand was being developed to send round SYNTHESYS3 and CETAF partners.

The collaboration of the ISTC and Digitisation Working Group in mass digitisation

The following areas of collaboration were discussed:

1. Identifying the requirements in mass digitisation in terms of equipment, software and workflows, as a form of gap analysis
2. Review and prioritise the requirements and collaborate to determine the best way to get tools developed and integrated into workflows
3. Building small communities with shared interested, eg Herbadrop, DINA, Symbiota, etc
4. Expand collaboration on existing developments
5. Collaborate on solutions to storage issues

The role of the ISTC and Digitisation Working Group in developing the European Roadmap and SYNTHESYS4

The gaps identified above could potentially feed into the next SYNTHESYS project proposal and would help feed into the European Roadmap. A workshop was suggested to enable presentations and feedback was suggested.

Potential relationships with non CETAF institutes/organisations

It was agreed that any requests should be confirmed with the CETAF General Secretary before any decisions are taken.

Next meeting

It was agreed that a joint meeting was useful and so the next meeting for both the ISTC and the Digitisation Working Group will take place in Spring 2017, at least 4 weeks before the CETAF General Meeting.