

Minutes of the CETAF Information Science & Technology Commission (ISTC) Meeting

22 April 2022 (10:00 - 16:00 CEST)

Executive Summary

The meeting took place in the form of a teleconference. The agenda is publicly available at https://cetafdigitization.biowikifarm.net/c/index.php?title=ISTC_Meeting_Spring_2022. Presentations as well as these minutes will be made available on the same website. The meeting was attended by 24 participants from 12 CETAF member institutions and the DiSSCo co-ordination. The main scope of the meeting was the presentation and discussion of the DiSSCo architecture diagram, the discussion of IIIF as a CETAF standard and the presentation of EU Funding Opportunities.

Action items

- To start collaboration on semantic annotation of collection data with BiCIKL (Quentin Groom, Anton Güntsch).
- To remind directors during the next CETAF GA that technical staff (for example in the context of GeoCAsE) should link up with the ISTC (Anton Güntsch).
- To develop a CETAF recommendation on the use of IIIF (Roger Hyam).
- To organise the spring 2023 meeting as a physical joint DWG/ISTC meeting in Vienna.

Attendance

Anton Güntsch (FUB-BGBM); Karol Marhold (PSBC-IBOT SAS); Stefan Seifert (SNSB); Falko Glöckler (MfN); David Fichtmueller (BGBM, only until 11:00); Dominik Röpert (FUB-BGBM); Ni Yan (Naturalis); Sharif Islam (Naturalis/DiSSCo); Mathias Dillen (MeiseBG); Roger Hyam (RBGE); Sabine von Mering (MfN); Sam Leeflang (Naturalis); Quentin Groom (Meise); Maarten Trekels (MeiseBG); Cristina Garilao (LIB); Heimo Rainer (NHMW); Julia Pim Reis (MfN); Henry Engledow (MeiseBG); Wesley Tack (MeiseBG); Jonas Grieb (Senckenberg); Patricia Mergen (MeiseBG, RMCA); Wiebke Walbaum (SMNS); Matt Woodburn (NHM London); Paul Braun (MNHNL)

Agenda

10:00	Welcome , adoption of agenda and introduction of participants (Anton Güntsch)
10:15	BioWikiFarm - Update and discussion (Falko Glöckler)
10:50	CETAF IDs - Update and discussion of further annotation efforts (Anton Güntsch)
11:10	GeoCASE - Update and discussion on the involvement of technical staff in the CETAF process. (Falko Glöckler)
11:30	TETTRIS EU Proposal - Overview of the work plan and discussion (Quentin Groom)
11:50	Funding opportunities: HORIZON-CL6- 2023-2024-BIODIV: Digital for nature and in the HORIZON -Infrastructures 2023-2024 program (Patricia Mergen & Quentin Groom)
12:05	Lunch break
13:00	DiSSCo - General architecture and status of implementation, questions and discussion (Sam Leeflang, Sharif Islam)
14:30	Break
14:45	Towards a CETAF IIIF Recommendation - Proposal and discussion (Roger Hyam) <ul style="list-style-type: none">• Background reading “Implementation of the IIIF for Natural History Collections” summarizes what our recommendations may be https://know.dissco.eu/handle/item/294
15:30	Plans for the Digitisation Working Group (Heimo Rainer)
15:45	AOM , next meeting
16:00	End of meeting

Minutes

The agenda was adopted without changes

BioWikiFarm - Update and discussion (**Falko Glöckler**)

- It works again, but backups to be solved, BGBM can assist with offsite backups. Need update security issues
- According to German, law MfN needs to have an explicit order by the people running the individual wikis to deal with people’s accounts and other information that fall under the

GDPR.

- They are not so fast, performance issues to monitor to optimize the system. In the restore process there were no other options. Will improve. There is only one central database, which seems to slow down all Wikis.
- Financial aspects to clarify : CETAF or beyond ? Most are related to CETAF members, there are a few outside. Will ask all contact persons, umbrella would be the Biowikifarm, but may need a more official umbrella in future. Not yet for that meeting
- Is it possible to import a wiki into the farm when it is ready again ? → not impossible just decided to freeze it to resolve the issues. To check if easier to set up a new wiki or migrate it.
- Is there archiving needed ? dormant or no more contacts ?

CETAF IDs - Update and discussion of further annotation efforts (Anton Güntsch)

- Project which could be extension of COST Mobilise, competitive, but would provide funds for STMS, workshops, on collections use ORCIDs, disambiguate in collections hope that we will get it. BiCiKL has also a task on it to linking people .. to linking people Anton: it is all about people, can be extended to other elements, BiCiKL is more (geographies and people), Cost Mobilise it is people only. ...To extend the services even more, need additional projects and activities.
- Henry : Botany pilot : uses data from Bionomia . Botany pilot is on linking people additionally.
- Last stages the submission : to submit to Biodiversity Journal. Round Tripping of data, GBIF, specimens to people on bionomia but difficult to get it back to the CMS. How to deal with data coming back already linked.
- BiCiKL what are the next plans : Annotation work, there will be a meeting in Seville, with BGBM, Naturalis and Meise. Would be good to include more partners. Should start at the end of this year, but would like to start it earlier → only few CETAF members are part of BiCiKL and try to open it more and have a larger workshop later on ? Would be useful it is the goal to extend the Botany Pilot. Implementation within the RDF and how to make it available. See with TDWG if there can be a session (deadline passed but could be in an existing one)
- World Flora Online, would have needs for this and have submitted a session for TDWG, this would fit in there .. or in several sessions at TDWG.
- It is a Botany Pilot, in animals should also work with them as structured differently. Botany pilot is generic and can be used with animals as well. It is because the test was done with botanists, but can be open to other groups .
- The Botany Pilot video is here <https://youtu.be/W-LFUOKlpe8>

GeoCASE - Update and discussion on the involvement of technical staff in the CETAF process. (Falko Glöckler)

- CETAF ESG and GeoCASE group, (which is independent from CETAF) and can bring in collections from external data providers too. If it works smoothly within the CETAF community. ,, , ie DiSSCo providers. Solve technical barriers.
- RBINS is using Darwin Core to see if they can use ABCD to provide their earth science data, RMCA has older BioCase version could upgrade technically but curator not ready to provide data;
- Stefan Seifert : has older version of ABCD 2.0 would need to map to EFG
- New Info (Cristina): LIB (Hamburg) has mineralogical and geo- and paleontological collections. There are plans to bring these into GeoCASE.
- Anton will mention it during next CETAF meeting, to let them know of issue, there are regular meeting of WG coordinators to work better together between ISTC and ESG/

- GeoCase colleagues . Already clarity on who is who.
- <https://geocase.eu/>
- Naturalis/DiSSCo work on sharing geological data: <https://github.com/TomDijkema/BioCAsE-GeoCAsE-docs>
-

TETTRIS EU Proposal - Overview of the work plan and discussion (**Quentin Groom**)

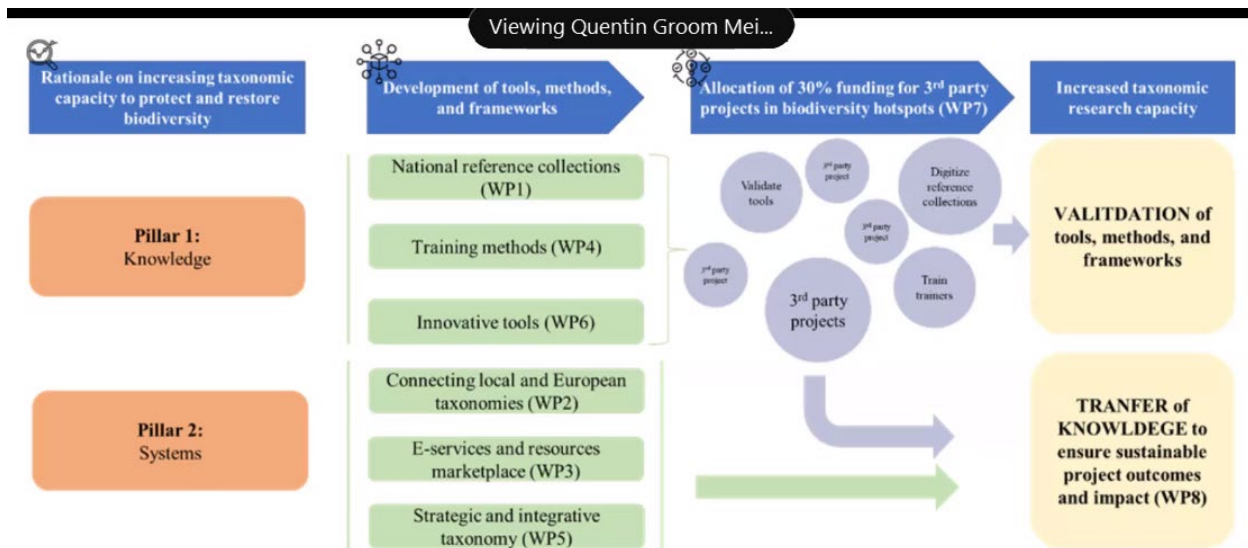
- Building capacity in taxonomy, know if accepted in May to start in fall 2022
- It is IA (things that already exist and bring them closer to market , 17 partners and 11 countries CoL is based in the UK, not NHM or Kew ..

- Seventeen deliverables

- Total Budget: €5,997,182
- Person months: 473
- Three years, six months

- 30% for third party funding grant €1,800,000

- Third party grants, even if not a partner directly



The structure of TETTRIS

- Third party calls, beneficiaries cannot apply.
- Was the only proposal submitted

Funding opportunities: HORIZON-CL6-2024-BIODIV: Digital for nature (**Patricia Mergen & Quentin Groom**)

Horizon BioDIV :

Interdisciplinary assessment of changes affecting terrestrial and freshwater ecosystems, building on observation programmes

One project of €3,000,000

HORIZON-CL6-2023-BIODIV

Addressing biodiversity decline and promoting Nature-based Solutions in higher education

Coordination and Support Action

One project of €3,000,000

Mentions Museums specifically

HORIZON-CL6-2024-BIODIV

Invasive alien species

Innovation Action

Two projects of €6,000,000

One for terrestrial ecosystems and one for aquatic ecosystems

HORIZON-CL6-2024-BIODIV

Digital for nature

Innovation Action

Two project of €8,000,000

Long list of objectives and tools, who can lead it , how to contribute ?

Robotics, AI , data modeling, almost all in there

Naturalis is interested in playing a strong role in this topic, though this still needs to be discussed internally.

Naturalis is also interested in the topic Invasive alien species.

List of additional opportunities , ask draft programs for more info and target calls

DiSSCo - General architecture and status of implementation, questions and discussion (**Sam Leeflang, Sharif Islam**)

- ...DiSSCo Services: <https://www.dissco.eu/services/>
- MIDS <https://www.tdwg.org/community/cd/mids/>
- <https://kubernetes.demo.nsidr.org/>

Henry: Identifying plant, organs ? is it only machine annotated ? What to do if as an user you notice something is wrong ?

Demo that can use machine learning can be used to annotate, but as part of the ECAS system the idea is that humans can also annotate the object information

Roger : object ? How to handle object quality ? how to handle cascading changes and interlinked changes. → have still to work on the data model and progress on these parts.

Activities in DiSSCo Prepare WP6 are looking into these aspects. If there are images, if name changes, collector name changes, what it triggers on other linked data, is not just the technical parts, but also policies and right managements who can change and approve changes.

Anton: What is the relation to GBIF ? Will this be provided to GBIF ? Institutions provide directly to GBIF ? Do they have to pass via DiSSCo in future ? → not official, aim to curate and annotate the data, with experts, citizen and machine learning. They are not aggregator, they can be a service between GBIF and the institutions that have been curated checked by DiSSCo. But to further discuss

Quentin : Image infrastructures, here this could be very useful to provide additional capacity for Machine learning on images, without that it does thing GBIF is doing. Do not want to compete with GBIF. Need to fill in a specific niche.

Sam: if have end points to run tests, contact him to make further trials. For example with the Botany pilot .

Anton : Specimen explorer : is this an API ? Should it use the API or implement something in parallel → would use the API. Just separate back and front end APIs. Can also build on top of those APIs.

SPARCL interface : is it ok to expose this to the public as it is massive ? Is this done elsewhere .? Comes from a BiCiKL WP and it indeed asks a lot. To be discussed. Not sure if it is doable yet.

Stefan : BioCase take the Archives or Live data ? For the moment live data but may move to archives. They can handle 2.06, Also 2.1 should work, but GBIF may not be able to handle 2.1
Anton Tools for data transformation can be used and further explored

ELVIS use case for TA .

Anton: is this specifically about the instruments → general but use case in ELVIS

ELVIS has a good interface, to update, instruments, shared vocabularies, why have multiple entry points.

Roger : Context, words, it is not easy to fill in. can mean different things, can change over time. Need high level vocabularies and need something human readable. Good to have an image, who to contact,

Pay attention to no ask too many efforts compared to what we actually need;

Towards a CETAF IIF Recommendation - Proposal and discussion (Roger Hyam)

- Why have a standard approach to publishing high resolution images and collections of images of a single specimen?
 - Enables building common tools
 - Enables sharing of skills and workers
 - Enables integration of data
- Why use IIF ([International Image Interoperability Framework](#))?
 - It is bigger than us - widely used in the humanities as well as sciences.
 - There are tools available and a community developing them.
 - We don't want to reinvent the wheel.
 - Because of Synthesys+ WP 4.3 we have already implemented it at a number of institutions.
- Why a recommendation?
 - We don't have a recommendation for JPEG or UTF-8 so why this?
 - There are ways to use IIF that need to be established as best practice within our domain.
 - We would recommend not just to use it but make suggestions as to how to use it.
- What do we need to do?
 - Decide here if we think it might be a good idea.
 - Take the [existing Synthesys+ output](#) and turn it into a Wiki that we can jointly edit under the CETAF.
 - Understand any formal process?
 - Discuss how we reach consensus.
 - Discuss timescales.

There are about 10 institutions who have implemented IIF

GBIF new data model, where they need camera trap images → digital representation of digital objects, but can directly see it useful for camera trap images;

Services can be CETAF branded ?

Do we need an extra Wiki or can it be in the digitization wiki. Who runs the digitization wiki ? > to be investigated further.

Falko : They are separate Wiki's for IDs and digitization, Would prefer to include IIF into the digitization Wiki

Mathias : Lots of different CMS and models. It is a bit the same for the images (format, storage, metadata ?) → is this an issue with IIF implementation → Experience from Edinburgh : plants in the wild, cultivated .. it works but difficult to integrate fully, takes some time

Maarten : if change infrastructure keep IIF in mind. It is possible to build it upon an existing infrastructure but is more work and less optimal.

Mathias : set up a new server and to modify it to make it work, foresees that it will be the same for others. Most institutions will probably not change their whole infrastructures regularly. Technical part is not so trivial.

Quentin : Every institution to have a IIF to maintain is challenging to see if DiSSCo can have a central IIF service to maintain it. Could be an economy of scale.

Anton: even if there is a central facility some cannot store it centrally it internationally so to see if possible.

Roger: Economy of funding could be an argument. Ie put them in Zenodo. Images not looked at much, so online storages can increase to view it. Maybe does not need to be instantly available, people can wait to get the high level images.

Pat : Machine to Machine access, could be used more often. Need real time access with hot access or cold storage (archives) Zenodo is archiving. Would need real time or near real time

Mathias: agrees on usage needs that can increase, to see towards EOSC if they can have a real time access ? But still in construction.

Quentin : To take time to continue to work on this. Devote time to this

Plans for the **Digitisation Working Group (Heimo Rainer)**

Separate or together. ?

If Face to face better together
If online could be separated

CETAF Exec : Think tank, Recommendations, Guidelines

Open Science strategies : impact on their daily routines ? How is it implemented? Do it like Collection Group ? What is overlap with CG, ISTC ..

MIDS : should be continued, via TDWG, maybe not there. Too many open questions.

Henry: Not so much fan of MIDS, descriptions were not clear. Not relevant for most curators .. what is the point ? How to clean the data ? If publish online , data is in too much poor state even if a lot, difficult to get the values. It does not help much. Can be more specific with the databases.

Difference on what IT people need, and curators need. If set up MIDS and it is not used ... ?

MNHN Paris had interviews about MIDS and how it can be useful for them. May need more discussions with partners; if MIDS is a part of OpenDS how can be ignore it.

MIDS should not be calculated manually but produced digitally.. How many specimens at which level, if data are not so that it can be done automatically it has no use, then the estimations are ok.

Mathias : Using MIDS, with specifications not ready yet. MIDS levels to be calculated. Uncertainties to be answered.

Heimo : Getting a MIDS level on their collections, if calculated from provided data. For example using GBIF data. MIDS 2 and 3 still needs to be defined. In some institutions already try to fill in information to reach higher MIDS levels.

Anton : All mass digitization in Europe, need some KPIs where MIDS are useful . To get an impression of the issues in a whole.

Anton : ISTC and Collection Groups with the curators.

Heimo: Communication between the groups important

Anton: if physical meeting to stay together each their own agenda and topics. Back to back.

Heimo : ok and ahead of the CETAF spring meeting.. Next one could be in Vienna.

AOM, next meeting

- Next CETAF 2023 spring meeting which should be a physical meeting ... Could be in Berlin (constructions ongoing)

TDWG : Anton symposium on national data centers, Quentin with ENA, GBIF , Worl Flora Online , CETAF side meeting → indicate to Pat what you need to something extra; Cetaf side meeting could be a dinner.

Join information

Meeting link:

<https://fu-berlin.webex.com/fu-berlin-en/j.php?MTID=m0499416fd56475fb19ead7f5157cef4e>

Meeting number:

2733 140 1870

Password:

Jf3JCmnt2q3

Host key:

199830

Join by video system

Dial 27331401870@fu-berlin.webex.com

You can also dial 62.109.219.4 and enter your meeting number.

Join by phone

+49-619-6781-9736 Germany Toll

+49-89-95467578 Germany Toll 2

Access code: 2733 140 1870

[Global call-in numbers](#)