

CETAF Information Science & Technology Commission (**ISTC**) and Digitisation Working Group (**DWG**) Meeting

National Museum New building, Prague

29-30 April 2024

Please enter the NM New building (not the Historical building) via the public entrance and tell the staff that you would like to go to the CETAF meeting. They will then take you to the meeting room on the top floor.

Attendance

- Anton Güntsch (FUB-BGBM)
- David Fichtmüller (FUB-BGBM)
- Dominik Röpert (FUB-BGBM)
- Jiří Frank (NM CZ)
- Patrik Mráz (CU Prague)
- Zdeněk Vaněček (CU Prague)
- Petr Novotný (CU Prague, only Monday)
- Milan Janíček (CU Prague)
- Wouter Addink (Naturalis)
- Mathias Dillen (MeiseBG)
- Quentin Groom (MeiseBG)
- Heimo Rainer (NHMW)
- Karol Marhold (PSBC, Bratislava)
- Stefan Seifert (SNSB)
- Dagmar Triebel (SNSB)

- Barbora Rolečková (Institute of Vertebrate Biology, Czech Academy of Sciences)
- Ana Casino (CETAF)
- Laura Tilley (CETAF)
- Sabine von Mering (MfN Berlin)
- Nora Lentge-Maaß (MfN Berlin)
- Majid Vafadar (MfN Berlin)
- Jiří Danihelka (Masaryk University, Brno, CZ)
- Ota Šída (NM, CZ)
- Íñigo Granzow (MNCN+RJB-CSIC, ES)
- Martin Souček (NM, CZ)
- Adam Cironis (NM, CZ)

Draft Agenda

29 April 2024

		Notes
8:30	Arrival	
9:00	Type Specimens Working Group session - Presentation of current plans to link type specimens to names in the Taxonomic Resolution Engine Wikibase (TETTRIs projects) - Discussion on the data model	Starting notes for the session
12:00	Lunch	
	ISTC Meeting	
13:00	Welcome , adoption of agenda and introduction of participants	
13:15	Report from type specimen working group and discussion of next steps	See notes from the session linked above
13:45	Report from Citizen Science Initiative and discussion of next steps	Focus from the last year: Joint missions and assignments of specimens according to expertises, Joint development of strategies for upscaling. Work in TETTRIs Task 6.4 (NHMW, BGBM, MeiseBG, MNHM, Naturalis), work on widening audiences, extending for new mission type (handwriting), very popular missions: old handwriting. Moving specimen data based on geographic location, e.g. French specimens in Berlin can be described by French volunteers. Technical

solution difficult due to the different systems, idea: make special new missions locally and advertise them at the other system. Problem: Country or collector field needs to be present, which it often isn't after digitization, only after transcription. ORC can help but not for handwritten text.

OCR results for language detection worked quite well for 1248 specimens in Berlin.

Paris has done similar approach and used Les Herbonautes mission to confirm by users if the country or label text is German. Was done quite fast by the community.

Discussion: language detection is the limiting factor. There are different APIs available but the usually cost if there are more than a few queries for the free tiers.

Mathias: if missions can be served with IIIF manifest files, information like language could be attached to it.

Anton: With a specification for a combined strategy, it is easier to get funding from national funding sources.

Further try to use different AI/LLM tools to do most of the work automatically. In a step before that machines could decide in which (parts of the) images can be detected automatically and which might require human support.

Action Points:

- Anton and Heimo will organize a meeting/workshop in the next months to further bring this topic along. Core topic should be the technical workflows to achieve and how to plan them.
- Sabine will create a contact between somebody from MfN who is in charge of the "Transkriptionswerkstatt" and Anton. (Sabine sent email to Anton on 6 May)

14:15	DiSSCo Update / Discussion (Wouter Addink)	DiSSCo is in the Transition Phase, setup of DISSCo ERIC (legal body, European Research Infrastructure Consortium) Some services and core infrastructure is already in development. Call for Hosting and founding members will come soon before the next two steps of the founding process. 17 expressions of interest to become a service provider have been submitted, 9 already in development, at various service readiness levels (SRL). Data model (part of openDS) based on GBIF Unified Model with mappings to DwC and ABCD: https://schemas.dissco.tech . Flexible Annotation Model based on Web Annotation Data Model Support for MIDS (Minimum Information about Digital Specimen), harmonized data is used to calculate MIDS score. First DOIs have been minted (via DataCite), but some limitations with will be overcome in collaboration with DataCite. Minimum Viable Product will be available this year. With 6 Core Infrastructures. Advance TRL levels from 4-5 to at least 6. Show with at a set of >2 million specimens from various collections representing different taxonomic groups and geological specimens. Regular demonstrations (every 2 months), see https://dissco.tech/blog-feed/ for upcoming and past ones. Questions/Answers: - Stable landing pages will be provided. Versions will be created when changes occur. - Subscribing to a taxonomic section and being informed when changes occur or new specimens are added, is not currently planned (certainly not for the MVP). Anton: it should be added to the list.
15:00	Coffee	
15:30	Misc topics • The new Czech biodiversity repository (Patrik Mráz) • IndExs as pilot to interlink relational	Czech biodiversity repository: Active taxonomic scene. 64 Herbaria in CZ, 10 Mio specimens, 75% in top 5 Herbaria. Of the 10 Mio: 15% database, 2,5 % freely accessible, 0,2% with images: Low digitization, interconnectedness & accessibility. National Animal Genetic Bank (NAGB), founded in 2015, member of

	databases and graph databases (Stefan Seifert, Dagmar Triebel) • Biowikifarm - How can we bring back the essential ISTC/CETAF wikis? (David Fichtmüller / Anton Güntsch) • GitHub - Does CETAF need its own repository? (Ana Casino / All)	GGBN. IndExs: Focus on Exsiccatae (set of preserved specimens with printed labels; uniform, numbered and distributed; definition under https://en.wikipedia.org/wiki/Exsiccata), Linking between library and herbarium, IndExs: Index of Exsiccatae: online db with bibliographic information. http://indexs.botanischestaatssammlung.de/ Webpage and API, Agent Identifier: Wikidata ID; Property in Wikidata: https://www.wikidata.org/wiki/Property:P12371 https://www.wikidata.org/wiki/Wikidata:Property_proposal/IndExs_Exsiccata_ID Biowikifarm: (TODO David: expend on the current situation here) Decision: wait another month to see if MfN is able to access the BWF data from the most recent state (Oct 2023) and export the pages from there. BGBM will set up a new wiki, where the pages can be imported. Two separate wikis for Identifiers and Digitization now, no merging of the content. If the import is done with the old data (Dec 2021), the new wikis will be set read-only, to be later updated with the newer information, once the current wikis become available for export again and avoid versioning issues when merging. User accounts will not be part of the export, but user names in the version history are not an issue (GDPR wise) overview of all the wikis: https://web.archive.org/web/20221220213148/http://biowikifarm.net/meta/Member_wikis CETAF GitHub: Undecided, no real use cases yet. Creating an organizational account is a good idea for now. https://github.com/IIIF/awesome-iiif
16:30	Institutional data management - potential for joint action by the ISTC • Introduction • FAIR data implementation plans and	Idea to share internal procedures and policies regarding data management.

	progress at Naturalis (see: Naturalis FAIR Data Playbook) • Potential for joint documentation and collaboration	
17:00	AOB • Upcoming funding opportunities • Next meeting (where?)	OSCARS Open Science Cluster Action for Research & Society Cascading funds : https://oscars-project.eu/oscars-first-open-call, second call in September 2024 Closing the research gaps on Essential Biogeochemical Ocean Variables (EOVs) in support of global assessments HORIZON-CL6-2024-CLIMATE-02-2 https://ec.europa.eu/info/funding-tenders/opportunities/portal/screen/opportunities/topic-details/horizon-cl6-2024-climate-02-2 → collaborations with GEO BON ? Exploiting the potential of secondary bio-based products HORIZON-MISS-2024-NEB-01-01https://ec.europa.eu/info/funding-tenders/opportunities/portal/screen/opportunities/topic-details/horizon-cl5-2024-d6-01-04 European Partnership of Agriculture of Data HORIZON-CL6-2024-GOVERNANCE-02-01https://ec.europa.eu/info/funding-tenders/opportunities/portal/screen/opportunities/topic-details/horizon-cl6-2024-governance-02-01 Draft program Pillar 1 Infrastructure not yet published but should be : anytime soon with calls opening in fall 2024 and more opportunities for our community. https://ec.europa.eu/info/funding-tenders/opportunities/docs/2021-2027/horizon/wp-call/2023-2024/wp-3-research-infrastructures_horizon-2023-2024_en.pdf but only very few calls (those 02) for September 2024

17:15	End of ISTC meeting	
17:30	Guided tour	
20:00	Dinner	The restaurant - Restaurace Hajnovka, Vinohradská 25 Hajnovka.cz Location - https://maps.app.goo.gl/RgxRK32H7NGMvzwZA

30 April 2024

	DWG meeting	Notes
8:30	Arrival	
9:00	DWG revival	
9:15	Recap DWG activities in the past	
9:30	Rationale of future DWG activities	
	National / Institutional Digitisation Initiatives and alignment with DiSSCo - GBIF - EOSC etc. roundtable	
10:00	Workflows in Collections • Preparation of materials for object level digitization • levels digitization ◦ Objects ◦ Taxonomy ◦ Standards: DwC, ABCD, AV; Latimer Core • Post digitization	

11:00	Current activities • OSCAR (H. Rainer) • MfN (N. Lentge-Maaß) • SKILLS4EOSC (H. Rainer)	
11:45	WrapUp	
12:00	End of DWG meeting	

Logistics

Venue

NATIONAL MUSEUM New Building, Vinohradská 1, Prague, 110 00, Czech Republic

5th (top) floor, meeting room TOP

https://www.google.com/maps/place/Nov%C3%A1+budova+N%C3%A1rodn%C3%ADho+muzea/@50.0799324,14.4290962,17z/data=!3m1!4b1!4m6!3m5!1s0x470b948d824eafb9:0xf2af9ac4bebc611a!8m2!3d50.079929!4d14.4316711!16s%2Fg%2F11r_pkwbz?autouser=0&entry=tts

Hotels

Hotel Museum – ***not available on the Booking.com at the moment***, [Kontakt | Hotel Museum](#)

Hotel Clarion – Tylovo náměstí, [Clarion Hotel Prague City. Praha – ceny aktualizovány 2024 \(booking.com\)](#)

Novotel Praha – Kateřinská (I.P.Pavlova), [Novotel Praha Wenceslas Square, Praha – ceny aktualizovány 2024 \(booking.com\)](#)

Hotel Arkada - Balbínova, [Arkada Hotel Praha. Praha – ceny aktualizovány 2024 \(booking.com\)](#)

Hotel Alton - Legerova, [Alton, Praha – ceny aktualizovány 2024 \(booking.com\)](#)

Hotel Ankora – Kateřinská (I.P.Pavlova), [Hotel Ankora, Praha – ceny aktualizovány 2024 \(booking.com\)](#)

Hotel Rimska – ***mandatory accommodation at least for 2 nights***, [Aparthotel Rimska 14, Praha – ceny aktualizovány 2024 \(booking.com\)](#)

Hotel Amarilis – Štěpánská, [Amarilis, Praha – ceny aktualizovány 2024 \(booking.com\)](#)

Hotel Mamaison Residence – Na Rybníčku (I.P.Pavlova), [Mamaison Residence Downtown Prague, Praha – ceny aktualizovány 2024 \(booking.com\)](#)