**ESG virtual meeting: Progress on GeoCASe**

**Monday 27th January at 15:00hr (CET)**

*Participants*

Johanna Eder (SMNS Stuttgart) ESG Chair

Laura Tilley (CETAF) Assistant Coordinator

Babis Fassoulas (NHM Crete)

Jirka Frank (NHM Prague)

Olle Hints (Tallinn University of Technology)

Steffen Kiel (NRM Stockholm)

Jiri Kvacek (NHM Prague)

Andreas Kroh (NHM Vienna) arrived later due to technical difficulties.

Patricia Mergen (MBG Meise/RMCA)

Giles Miller (NHM London)

Michael Rasser (SMNS Stuttgart)

Rivka Rabinovich (HUJI Jerusalem)

Wiebke Walbaum (SMNS Stuttgart)

Racheal Walcott (National Museums Scotland Edinburgh)

*Agenda*

* Update from Stuttgart and Tallin on their analyses of the GeoCASe source code.
* Discuss what the next steps are.

*Minutes*

*L. Tilley*: Opened the meeting by welcoming Rachael Walcott, from the National Museums Scotland Edinburgh. Racheal is a curator of Earth Science Collections - particularly Mineralogy.

*Updates from O.Hints:*Tallinn have made good progress with analysing the functionality of the GeoCASe infrastructure (<https://github.com/MfN-Berlin/geocase-infrastructure>), and have managed to build a prototype indexing service (based on Solr) and a simple user interface on their local server, plus harvested and indexed data from providers in Estonia and Stuttgart. Link to prototype portal: <http://geocase.geocollections.info> The core elements of GeoCASe work well and the Indexing provides fast queries. In order to further test the speed of the system it would be good to have more data.

*A. Kroh*: NHM Vienna reported technical issues with providing data: Causes were identified to be an outdated version of the BioCASE provider software. After the update the data could be harvested. Some technical problems remain, but these are internal and do not affect the connection between the data source and the test portal. **Update from Olle Hints (28.01.2020):** This works as of today and some Vienna data are available in prototype.

*L. Tilley****:***  Asked what would be the next steps to improving GeoCASe and have it fully functioning?

*O. Hints* : In order to make further improvements we need to encourage more institutes to contribute - getting GeoCASe fully functioning could take years. We need data from big collections to really test the speed - a few million. In addition GeoCASe needs people to manage the data, servers, and have a functioning helpdesk. There is a lot of work to be done. Estonia (Tallinn) cannot do all of this - for now they can keep the data on the servers. The resources are currently limited and Estonia need to wait till Feb to see if they will receive national funding to cover the DiSSCo prepare roadmap (preparing the technical infrastructure - including Earth Science collections).

*G. Miller:*A lot of the Data from NHM London goes directly to GBIF. He asked whether certain database software could be linked to GeoCASe… May need to do a data mapping exercise for NHM collections to see what information would fit into GeoCASe.

There was a general discussion about adding mineral and rock data to test the indexing - i.e. to see if the data fits. There is a general problem with rock and mineral being able to fit the structure of the ABCD standard. Currently the TDWG Palaeo group are trying to extend the standard to implement Stratigraphy etc, but it is an ongoing task. *R. Walcot* - asked whether GeoCASe also stores geochemistry data - the answer is that it currently does not. GeoCASe could be used as a test bed for finding out what Rock and Mineral data is asked for and what additional fields are needed. CETAF institutes are encouraged to provide Mineral and Rock specimen metadate for testing the capabilities of GeoCASe. ESG partners volunteered to provide more data for testing the prototype:

* R. Walcott (National Museums Scotland Edinburgh
* J. Kvacek (NHM Prague)
* Steffen Kiel (NRM Sweden)

**Action:** These institutes will get in touch with O. Hints.

R. Walcott also mentioned a mineral/rock database Mindat, this may be some help...did not fully get notes.

Someone asked about if GeoCASe displays images - *O.Hints*: explained that only the links to the main images can be mapped with ABCD standard and stored in the index, and there is currently no easy way to additionally map and index thumbnail urls. Currently the test system is not able to store images / thumbnails locally.

*J. Frank:*Initiated a talk about the sustainability of the GeoCASe and the need for Hardware and distributed servers and who would host it. Maybe the DiSSCo infrastructure could provide support for this. And general questions of how GeoCASe would fit into the DiSSCo infrastructure.

There was a explanations that it will provide Earth Science data for the ELViS platform (at specimen level). There is a general agreement that GeoCASe is needed for the Earth Science community because GBIF is only catered to biologists and does not fit our requirements.

*L. Tilley:* Will talk to Ana Casino (Executive director) on how CETAF can further support the needs of GeoCASe development: technically, financially, etc. Based on the outcome of this meeting. *Plus Ana is deputy coordinator communication and engagement for DiSSCo.*

There was a general agreement that we need to discuss how the workload of GeoCASe could be split between CETAF ESG institutes, but first we need to hear the opinion of Falko Glöckler about the outcome of this meeting. Two years ago, we did develop a work plan, but this would need to be revised.

We should keep in touch with the TDWG palaeo group who are also trying to extend the Darwin Core ABCD schema to accommodate Earth Science specimen data.

***P. Mergen:*** Noted the importance of constantly promoting the importance of Earth Science in DiSSCo Prepare to make sure our community needs are considered. Also how important the CETAF ESG community is for providing best practices. These best practices have been considered by Flemish research institutes and universities with small collections.

**Next steps:**

*L. Tilley* : Write the minutes and disseminate them to the ESG group. Report the outcome of these minutes to Falko Glöckler opinion about the reporting’s of this meeting and how we could potentially split the workload.

* *L. Tilley*: To talk to Ana Casino (Executive Director CETAF) on how CETAF can support the needs reported during this meeting.

Once we have an opinions/feedback from Falko Glöckler and Ana Casino, the ESG can arrange another meeting to organise how to move forward.

Further discussions can take place during the next ESG physical meeting during the 47th CETAF general meeting taking place in May (Tervuren, Brussels).

**Feedback/Comments from Falko Glöckler (09.02.2020)**

It’s great that the test prototype is up and running. It’s a good starting point to improve the search and other functionality of GeoCASe in a distributed development approach.

I have some news from our side: We are going to hire more (permanent) developers for different activities at the MfN. So I will be finally able to dedicate some of their time for further developing GeoCASe.

The job announcements are not published yet, but I would expect them to be out by end of March latest. Hopefully, the team will be complete by end of May then (unfortunately, recruitment takes some time here).

There will be lots of tasks for the new development team (GeoCASe will be only one of them), so a distributed development within the CETAF & DiSSCo institutions would still be the best approach to avoid bottle necks and sustain the progress on the portal.

Before the next (virtual or physical) ESG meeting I would like to outline a (rather technical) roadmap and align this with the experiences Olle and his team have made. I will contact them regarding the details.

I like the idea of revising the work plan we made two years ago, as I feel this might be the right time for kicking off the teamwork of both, a technical team and a community support team. In order to do so we should agree on the top priorities based on the user stories in GitHub and refine the before-mentioned technical roadmap draft. Of course it is also very important to align this with the DiSSCo roadmap.

In the next paragraphs I’m trying to address some of the topics and questions from the minutes:

GeoCASe role in the DiSSCo infrastructure (addressing the question by Jiri Frank):

From our side it is planned to use GeoCASe as a domain specific data source for the DiSSCo infrastructure, but this will be discussed and specified in the activities of WP 6 in DiSSCo Prepare. From the current state of the discussions, the collection data will remain “living” in the institutions’ databases and will be referenced (on a kind of meta-level) as a part of the Digital Specimen in DiSSCo. This means that GeoCASe should probably not being considered as an repository for “pushing” geo-collection data to DiSSCo (although it definitely should be able to do this, but not exclusively). Instead, the further development of GeoCASe should focus on the discovery mechanisms and web interfaces (both human and machine readable). So, this will enable GeoCASe to serve as a registry of specimens in geo-collections for DiSSCo as this will be one of (potentially many more) use-cases for accessing the data via web APIs.

Images in GeoCASe:

Having said the above, it might be also clear, that GeoCASe should not be hosting images of the specimen. However, it makes totally sense to create (or host) thumbnails of the images in order to avoid performance issues. But this would be only temporary depending on the schedule of regular re-indexing.

Hosting GeoCASe:

MfN will continue hosting GeoCASe and the data. In the context of the DiSSCo infrastructure there could be the opportunity to set up a mirror in other institutions as a fall back in case of temporary outages.

Data standards and their extensions:

I would like to encourage all ESG members to participate in the workshops of the TDWG paleo group. Some might know, that the ABCD-EFG and (DarwinCore) paleo group has been merged during the Biodiversity Next. The conveners of the group (including myself) are aiming at regular (virtual) meetings in order to make essential progress on certain topics of extending the data standards. Regarding the geochemical data the community could discuss in one of the meetings to which extent such data should be in covered by the standard or whether there is a stable service to only reference to (in order to be resolved in data portals like GeoCASe for better discovery by geochemical composition).

These discussions should be closely aligned with the DiSSCo Prepare task 1.2. (User stories and needs in the geo-community, led by MfN). The ESG should be encouraged to actively participate in the upcoming workshops as well. A first one will be during the SPNHC conference in Edinburgh in June.