

### WP3 meeting with AIT in Graz, 30 March 2011, 9:30-18:30

Participants: Henning Scholz (MfN), Adrian Smales (NHM), Walter Koch (AIT), Bernd Sproger (AIT), Gerda Koch (AIT), Johannes Edelsbrunner (AIT)

Introducing the AIT team:

- Bernd is working almost full time for BHLE, but sometimes has different project work
- Johannes study Mathematics, working part time for AIT and BHLE, he is mainly working on Drupal
- Odo Benda is working for OpenUp!
- Bernd Gohlisch is reporting to Walter, not working for BHLE

Taskbriefs:

- taskbriefs are written from the AIT perspective
- taskbriefs are working documents, several iterations necessary before approval; usually task briefs are used a means of project steering; created at the beginning of the project by project management, distributed to partners for comments and amendments to agree upon distribution of work; AIT has only created task briefs for themselves; other partners are welcomed to use AIT task briefs as templates for their purposes;
- **AI: Bernd send word doc of taskbriefs to Lola - done**
- **AI: Lola send taskbriefs around for corrections, amendments, own versions; to TMB & PMG first; in the second round also to people mentioned in the taskbriefs (check list of people before sending)**
- we discussed some specific aspects of taskbrief 3.1.5
- Tesseract to be used for OCR, developed at HP Labs, free, train it, already tested by AIT, integrating in the Pre-Ingest workflow which does the OCR
- Fulltext search should not be a big problem, but needs to be coordinated with the US, do they have an index so we may reuse the index; for our content we can index ourselves
- "optional" on page 3 will be removed, as OCR will be done
- Bernd successfully got the AnimalBase data, but this was not easy – information will go to the metadata (metadata enrichment)
- Index enrichment is a nice idea but to complicate now, metadata enrichment will happen at Pre-Ingest
- Index enrichment is offline enrichment, Wolfgang argue for life data to be used
- These changes have implications on the architecture – is it worth the effort going back again, unpick, evaluate and change the architecture?
- Metadata enrichment is in the current architecture
- Serious stuff is in the metadata
- Query expansion, search term amplifications – VIAF
- Taxon finder = static – metadata; Taxons (CoL) = dynamic – query expansion
- Need a glossary kind of thing or a public Wiki, what information will be where, what is static, what is dynamic, what is in metadata, what is in the index → next step is to have a dataflow diagram in D3.7 in addition to the key components
- **AI: build a list of webservices, vocabularies, enrichments over the next weeks**
- **AI: dataflow diagram for D3.7 (Pre-Ingest) with regards to enrichment**
- Taxonomic Intelligence: We work with the Taxon Finder to identify names and use the CoL database (webservice) to map these names to common names and taxa (including synonyms) if present in the database. This will enable the user to search for common names in BHL-Europe and get the relevant scientific publication with the latin species name in the text. Fulltext search (if implemented) will also allow for the search of common names directly. The integration of the AnimalBase data is further

improving the “taxonomic functionality” of BHL-Europe working in both directions (to the books but also linking back to AnimalBase to get the species page).

- CoL is multilingual, but what about the quality – based on our work we may present a proposal to Sp2000 how to improve the database
- ***AI: Henning to talk to Heimo/Wolfgang if they do any work with Sp2000; otherwise Walter will follow up with Yuri to check for webservice and data quality improvements***
- Taskbrief 3.2.2 – section 4.2: AIT will facilitate translations but not translate – it is optional, if people come and say we need to improve quality of data or translate, AIT will facilitate
- Taskbrief 3.2.2, methodology: the technical interface is REST/SOAP, ontology proxy is the component mentioned –
- ***AI: section will be revised to better explain after everybody has provided comments***
- Most webservices need to be built, not just implemented (Handle, virus scan etc)
- Have we considered working with BHL to allow them to get our metadata? We don't have enough data yet, we can kick-off the process once we have, we have standard means to help BHL
- Europeana API: it is planned to use the Europeana API (click button, extend the same search across Europeana, iFrame in BHLE portal)
- ***AI: Henning to inform Lizzy***

Pre-Ingest and related aspects:

- More content will be delivered to Europeana before the review; delivered through the old prototype
- is shown by/at – generated out of the pre-ingest – put in IA book reader, stable URL on London server
- ***AI: AIT to finish the deployment work on the Pre-Ingest within the next 10 days, then start testing with data on NHM servers, work on the demo for the review and get more content in Europeana for the review***
- Reasons for delay with Pre-Ingest development: not a standard modul but a very complex one (see BHL wiki, London meeting, Aug 2010 – [https://bhl.wikispaces.com/file/view/BHL-E\\_PT\\_sprint02\\_203\\_Workflow\\_v01.png](https://bhl.wikispaces.com/file/view/BHL-E_PT_sprint02_203_Workflow_v01.png))
- After evaluation of available options it was decided not to take work with existing Pre-Ingest components but make one that fits our needs, this took longer then expected, it was more complicate then expected, due to ongoing discussions and changing requirements
- Pre-Ingest component has microservices: checksum, NOID, etc.
- It was discussed and agreed that we better need to communicate the progress and delays, when they occur; also provide reasons for problems and refer to documents and information even if these has been provided before
- Workshop with the CP in Tervuren on all the Pre-Ingest steps, to better understand and appreciate the process to know how to interact with the system and prepare the data properly
- It is possible to use NOID as minting process but use Handles – needs further discussions also with Atos how to proceed

Islandora and interdependencies of OAIS modules:

- Islandora provides access to the archive; arrow from Data Management to Access in the architecture diagram is Islandora, Access is the portal
- We discussed the implications of the Islandora decisions and in this context the importance of the various OAIS modules and the implications for our work; e.g.

without the Pre-Ingest we can not do anything, but without the Data Management module we are still able to serve the content to the user via the portal

- We will develop all modules to be OAIS compliant and - more important - because we need them for what we want to do, however, the development work on the different modules can run parallel and to some extent independantly
- Data Management provides access to the content providers and is very important for the effective maintenance of the data
- We can give Atos some time to work on Islandora and still work parallel on the access (side note: it was further discussed in the tech call and further evaluation is necessary if we need Islandora and which parts to avoid overlap with the portal development)

#### Portal development:

- we have seen a demo of the current Drupal work of AIT
- some Drupal modules are working
- For the review (submission of deliverables to the EC) AIT will realise simple and advanced search (see Google advanced search as example), skinned for BHL-Europe, simple browse interface, facets hopefully
- This portal demo will be still running on AIT servers with existing BHL sample data; the access to the “real” data on NHM servers will be demonstrated with the Pre-Ingest
- We have with Dennis, Wolfgang and Andreas more resources to help with the Drupal development, we have to investigate what is necessary and possible, need to have in mind that the time for getting started with the work is in some cases to much to do the work efficiently
- ***AI: Henning to send BHLE Drupal theme to AIT – Done***
- ***AI: AIT to speak to Jiri Frank and invite him to Graz if necessary to help with the application of the BHLE theme for the portal***
- ***AI: Bernd to think about and investigate modules or tasks for additional developers to work on***
- ***AI: AIT to deliver to Henning an URL for the progress report to provide evidence for the information provided in the report (achievements, performance indicators)***

#### General management and planning

- Two processes – concept and implementation, run parallel and are connected at some points, but not all the time
- Tech calls: articulate what is on the table now and will be done, issues will be discussed and worked on immediately, everything else we take notes and come back at a later stage, this will help to increase the focus of development and reduce the background noise
- Currently we have a vapourware system; real system is different – we do dynamic development, take our partners/customers/users on a journey and show them intermediate steps, ask for feedback and comments – bug fixing will be done; suggestions like “can the system do this instead of that” will be put back several months to first get tangible results
- We can not please everybody, but we have to get we working system
- Right or wrong, make it strong
- To move forward: make a list of the ambiguities within the project – what is not nailed down tightly – everybody contribute to this list – discussing every item one by one for a few minutes and make a decision following a pragmatic approach (consensus, vote - dependant on the item)
- This also needs to be done for D3.7