

Portal Development Workshop

Lola – Intro and welcome:

- Running over 2 sessions, today and morning
- Main objectives today – eval main portal features implemented to date
- Walk through features 1 by 1 and capture feedback
- does it meet functionality
- feedback +

Tomorrow's session

- Potentially today – review to ensure we have a system by month 28 which will meet reqs
- Review what needs to be implemented and consider resources and timeline – mainly for developers
- Ready to produce updated portal roadmap ahead of august
- LN set up issue tracker, We will be using issue tracker directly to incorporate feedback
- Review based on extract from scrum backlog deliverable entries for core features
- Development team have met and translated the priority 1 features into backlog items
- Kickoff Walkthrough – BSproger
- LN raising issues directly to github

BS demo adv search 1:

- JH – not certain on ordering of which search being tested – simple or advanced
- Fransisco – confusion need order to session
- Fransisco – need structure to define how we can drive feedback session and what to test/how to test
- JH – this session for initial feedback; full testing to follow later
- This Session is not for testing, validation of implemented features.

--

Feature Run through (order from spreadsheet)

- 1 – ok
- 2 – [ok – tbc autocomplete]
- 3 – ok [need content and testing to fully confirm diacritic searching]

- 4 – Categories for simple search following bibliographic search: title, author, [plus taxon name – searching in content with found names/ubio tools]
- Discussion of taxon names, OCR and how found names treated, OCR scan of documents to extract names and store position within docs; approach intent to store found taxa in metadata – name finding

- 5 – GRIB integration – no non digitized content in BHL; but how to integrate with the GRIB which holds such. Solution suggestion – link to grib search in output from normal search to enable searcher to follow along to GRIB; need further follow up with GRIB side.
- ** 6 – ok (standard date forms) – query by century format to move to a new feature (enhancement with for example specific cultural date ranges)

- 7 – taxon not yet there, but query across different fields – ok
- 8 – browse index/category [ties to 4], add the other 2 categories – taxon, subject/discipline – ok as for 4
- 9 – things missing, status (digi or non digi), resource type – xsl available to modify style – overall ok
- ** 10 – sort by category/ (defined by field), type – we have primary sort - ok; need to add a feature for secondary/tertiary sort
- ** 11 – we rely on xsl for display format, add button for direct download of bibliodata in [various/selectable – mets/olef/xml/pdf] format - easy to do - ok
- 12 – ok
- 13 – ok
- 14 – ok++
- ** 15 – ok add button for re-run query/delete search rather than link – ok; rename save search results in user profile to save search queries and set off by default
- 16 – can bookmark a query URL ok (q: for a long URL can we makeashortlink)
- 17 – ok – bounding (search) box to go for now – unreliable in operation also if OCR is bad. [main concern, page level metadata and speed]
- 18 – ok
 - 19 – visits needs to be added to main page also; exports and stats ok
 - 20 – portal drupal customisation, news aggregation needs to be configured in – reassign this config over to AWD
 - 21 – wilddcarding implemented ok – word correction not in, add as new feature (can be done via Solr – Andreas has ideas) – reprioritised – still P1 but end of list
- 22 – front truncation is a big potential hit; this is embedded string search (eg for len des nat – an in string search - to find annalen des natural; will not be a default function for simple search – option to allow user to turn on (this feature will be useful for scientists, not such for other users) – end of P1 list
- 23 – ok

- 24 – drupal personalisation – ok
 - 25 – should show first page, not blanks for docs/content viewer – not visible so back to 50%
- 26 – ok
- 27 – transfer to Jiri – will be done end of June
- 28 – transfer to Jiri – will be done end of June

Round table review –

- Heimo – search function is good, keeping taxon separate from biblio is good. Quite satisfied, agreement that substring searching is important as a scientist
- Kimmo would like to keep bookreader with search function/highlighting
- Francisco – pleased with the solution, would like scientific option for searching available, happy as a personal setting
- LN: self declared not a good representative user ☺
- Andreas: portal functional and useable, some very important and essential in there, a few features still to come
- JH: ‘Yes’
- Dennis: simple search fine,
- Larissa: very important to have a google like search auto-complete from a non science perspective; Image search by plate/page important. Simple search result would like year of publication visibility
- Jiri F: would like search option in bookviewer to find pages/details quickly; overall positive, google search autocomplete ideal for default users.

Overall voiced consensus is favourable

LN – testing plan will be looking for bugs on live deployed environment

Lola: next steps we need to review priority 1 features still outstanding for month 28 release, and decide priority/relevance

Session 2

Lola: review of work so far, consideration of priorities for other items and timelines/resources for that work.

GH: run through P1 features yet to be implemented; a number already done, others fall into several specific categories – taxonomic, name recognition, basket functionality, document delivery, help and feedback mechanisms, Grib links/integration – these are where we should focus today in groups

Help/Feedback

- Help for search, download – described originally as link to bhl portal wiki or tutorial
- Feedback for OCR, scanning quality or portal functionality

- BS: help can be contributed by users
- JH: need to have a minimum set of manuals and documentation from user perspective
- JH: user feedback strongly desired on US portal, need to consider how
- JH: bug reporting more related to issue tracker than user feedback
- GH: Jiri F to consider options from a WP5 perspective and come back to us in the next week – delivery as collaboration with AIT (implementation) JF for solution

Basket / document delivery (pdf generator)

Download options – single page, item or content in general – range of pages, various quality, colour/non colour, file formats, download help (intersection with above)

- JH most important thing to download a range of pages (article from a book) with one click;
- BS quality of extracted images in submissions may be an issue, delivery from archival side to create derivatives. Request made on portal, delivered from AS. (guest interface?)
- LN need to determine specific chunks to target
- JH image quality, formats: PDF OCR tif,
- LN strict OAIS dissemination from access, archiving only should hold highest quality objects; in our case fedora has tools to enable
- WK: view is while AIT listed that this in support of Atos
- GH agreed Atos to take lead on the dissemination component
- LN do we integrate into portal or hold separate – provision via web services
- GH agree where to present the tools in next TMB and what we can deliver in next 2 months. Need to determine what is deliverable quickly and practically for core release and what needs significant resourcing

GRIB

- GH we would like to integrate grib – but level of integration is the question.
- Do we fully integrate, or use as to pass on where objects aren't found – use case for average user likely to be can't find, request scan.
- Boris – how do we combine login auth, Grib not an open system for unauthorised. Don't want random users incrementing counters.
- JH for users of portal, only important option is - can indicate request for an object
- Boris can set up a default portal user via API– discussed this option before but not yet implemented, to enable count of requests for an object
- JH is this possible without the user going to the Grib
- Boris view/search options only make sense if Grib data is in the portal. A question of how we want the user to interact
- GH – do we aggregate the search results
- LN we would need to harvest Grib or use a federated search for live (much more complicated)
- BS for portal easiest if Grib data available as OAI – we harvest as a 'nondigitized' collection. Is Grib data available as OAI-PMH
- JH not yet – once Grib stable and released can plan in
- Boris formal Grib release in sept

GH Boris to follow up with AIT (AIT provide OAI harvester), format to be agreed later

GH We want to do this, will target for November, still P1, but not for core release.

Taxonomy and Name Searching

GH a range of issues – what is practical for august release?

- JH most important is integrating ubio/name finder
- GH can we implement ubio for august, do anything with general names and should/can we bring in CoL/SP2000 services
- Heimo – SP2k concerned about derivative publication, CoL setting up services based on checklists – positive about use of these. CoL have several batches of services for name finding, will be xml/json – will be available for testing by end June. Download of corpus of CoL to BHL is a concern for SP2k. BHL ideally would use a cache of CoL to avoid performance issues rather than a download. EoL using checklist 2010, major changes for 2011, web service interface will be better solution
- Walter – identified NHMW responsibility for implementation works
- GH – given current status of CoL – we agree this is target for November rather than core
- JH – ubio needs to be available at this point
- GH – what is priority for VIAF service? August or later?
- JH – same as ubio,
- Henning – for august do we aim to offer core with equivalent functionality of BHL US or more
- Heimo – would be good to have this implemented via webservice rather than download – possible could be done by august. Person names as
- Henning – europeana have already implemented this
- GH – VIAF web service integration will be delivered in core release

- GH - Which other aspects do we need to consider
- JH - content viewer – users will expect ToC, title page, functionality as in BHL US
- WK - ToC needs data from scan files – this doesn't come from metadata
- Martin - important to capture these data during scanning process; BHL US look for scan data xml, needs to be in SIP. Also review portal and enhance on case by case basis.
- BS – option to add a list to go direct to object (page may match)
- Martin – simple option where no data available is to provide a false ToC
- GH – our issue is that many items already scanned. There have been mismatches with BHL US portal where scans don't always match the metadata.
- JH – data should be available for existing BHL content
- Heimo – gathering this data is very important for inclusion in Best Practice Guide