

July 18, 2016

Erin (Program Manager), Gabi (Technical Manager), Jackie, Martin, Carolyn

Action Items

Gabi will provide Carolyn a login to the GGBN document library
Carolyn to send API information to Gabi
Carolyn will ask Bianca for a breakdown of publications in BHL by year
Martin and Carolyn will debrief BHL's Technical Developer Mike Lichtenberg
GGBN team will look at BHL and draft a list of bullet points of how we might collaborate
We'll all touch base again in a month or so

Notes

GGBN is the Global Genome Biodiversity Network:
http://www.ggbn.org/ggbn_portal/

GGBN Document library - assemble PDFs related to tissues
Interested in facilitating data mining within the library

Kinds of Documents:
combination of published, open access, grey lit, archival materials.
50 documents currently
Tags, assigned by GGBN team
Full text

Just looked at BHL, typed in genome, lots of interesting matches

Interested in linking out to content already in BHL, to avoid duplication of effort
APIs might provide a good way to do that

Document Library Technical Background (Gabi)
Using AI Fresco

open source document management system
Tag documents with self-defined key words
Full text search
PDF
Stored binary files
underlying SOLR and Java

Use SOLR for GGBN data portal
GBIF is using as well
Google and eBay and big data platforms
You need some knowledge about fine tuning
Depends on how you define fields
For example, you could spend 4 hours indexing something and then
if you change one field, reduced to 20 minutes
So does require some experience on how to optimize
GBIF has lots of experience that GGBN has been able to draw from
Would be happy to work with us if might be helpful

For users, they can search and see sample in data portal, the idea is
that here also is the reference (PDF), and here is the sample

BHL is only open access
Title and metadata
We don't tag in the same way

For molecular stuff, most interested in 1960 and younger
Do we have a sense of how much content in BHL is from last 50
years?

Get from Bianca, breakdown of publications by year. Always growing
because as part of permissions agreements with publishers, there
are moving walls based on in-copyright materials
About 400 in copyright titles right now that we have permission for

It's a plus if we can find a way technically how the platforms can
communicate to each other

Carolyn will send a link to the API documentation

If we can have access to the stable url of articles, that would be enough to tag it and deliver through GGBN portal
However, we could not do full text search on those as is, but maybe we could find a solution as well;
Using the tags and categories, perhaps there is a way these could be sent back to BHL
Some GGBN managers (Katie, Jackie and Gabi) assign those tags

Collections in BHL

If tags were exported, might be possible to use them to build a GGBN Collection in BHL

Could pull those altogether in one place
It's a manual process but using tags and the BHL API, it might be possible to streamline the identification of the item IDs to help facilitate adding items to the collection

Gabi will send Carolyn a login to the document library

Carolyn to send API documentation to Gabi to look into cross-linking
We'll give Mike a heads up and see if he would like access to document library as well

Touch base in a month or so

GGBN will explore the two libraries, think about how to collaborate and compile a draft of bullet points how we think we can collaborate