Untitled

colomb_julien

4/17/2019

Contents

A Figure centered knowledge exchange

I call here "figure-unit" the ensemble of the figure and its metadata, it is probably very close to what Thomas names smart figures.

The figure as a unit (figure-unit)

Minimal Info

As a minimal unit we would like to have:

- a figure
- a title
- a legend
- automatically generated tags

Tags created via sourcedata and/or other Text mining technologies.

Extensions

- link to the data
- link to table of material used (with RRID)
- link to protocols used (doi/protocols.io ?)
- link to analysis documentation/code
- Other metadata (??)

This information would be useful (necessary?) when it is time to write a manuscript. It could also be used to make more/better tagging. For example, if we get a specific antibody in the material list, the protein recognized by the antibody can be use as a tag. This may be important when people use nicknames/synonyms in the figure legend.

In particular, for mutant/transgenic mouses, via the MGI number, we could have access to a lot of information (which is directly linked to ontologies).

figure-unit creation and sharing in one lab

At the moment figures are shared via slack in the group. I think we need to find an output which is shareable via slack. Slack commenting could be used as an informal peer review of the figures.

We could use the HUbox (basically an open source version of the dropbox) to add **version control** to figures, but it does not solve the problem of attaching (meta-)data to the figure file.

Texture can add the information to the figures, but is not close to be ready for our purpose, as it does not allow any visualization of the figure in a different software (i.e. no push to slack).

At the moment, I am thinking about looking into using R and Rmarkdown to produce both a computer readable version (XML, texture-like) and a PDF version (shareable via slack) from the users entries. Creating a small app to upload the figure, add a title and a legend as well as links to any other element, and produce a PDF out of it is easy; I will need to look into how we could create an XML file from that input. NB: it could also be coded in python.

Search/delivery of figure-units

The tags linked to the figure could be searched via keywords, but I think the most interesting is to be able to create a network of figures (what figures goes with what other figures), and/or combine this information with information we could gather from people searching (figures they published, material they use, protocols they use/plan to use,...)

We could therefore calculate a score of relevance for each researcher and deliver only the most significant figures to them ("This results seem to be very relevant for your research, check it out!")

This is yet very hypothetical, though, and this is a problem we will not solve alone. Interestingly, an initiative to make better data search has just started:https://www.go-fair.org/implementation-networks/overview/discovery/, I enrolled in the list and hope we will get some nice things going that way.

As a first system for dissemination, we might use a simple static blog website (on an intranet to restrict access, we would create with Hugo?). It has several advantages:

- blog posts can be tagged and we could see how the tagging system could work.
- blog posts could be created via the same code as the PDF and XML version of the figure-unit.
- We will anyway create a blog platform for the SFB (see last section) and could use the same basic design for an blog website with restricted access.

figure-unit peer review

While informal commenting via slack and/or Hubox is possible, one will probably need a better formalized peer review for publication-ready figures. A manual peer review is possible, but it might be interesting to use the pull request functions of git-based tools (GitHub-Gitlab-GIN) as a peer review step.

One could build a specific repository for figure ready files, lab members could only make changes on their branch and would need to send a pull request if they want their change to be incorporated in the master branch. At that point, we can not only have a manual peer review (science, statistics, legend,...), but incorporate automatic checks (is the data attached, are there enough tags, ...).

NB: This is also be linked to the blog idea, since our static blog website will be developed via git based tools.

Other information

Keisuke

The Larkum lab has already a data manager for the human brain project related work. His work is presently focused on collecting metadata from the software people are using to do their electrophysiology and imaging, as well as looking into the ontologies to be used. Since we talk about huge data file, GIN might be a good solution to add some structure to the people's data management.

Server

I talked with Richard today. The SFB server we plan on buying cannot be used to store the raw data. First because it will probably not be large enough, and second because a server lifetime is 7 years, while the DFG/HU requires a plan to archive the raw data for a least 10 years.

On the other hand, 40 000 euro for a server to share figures and metadata is a bit overkill. I am still puzzled about what we can do with this... The ITB could give us a virtual server for testing if we need to play around stuff before buying a server, by the way.

Blog website

We will build a Rblogdown/Hugo static website. It is technically quite simple (see https://rdmpromotion. rbind.io for instance). Its first use will be to blog about a seminar series. I was indeed present on Tuesday and could convince the participant to take notes collaboratively and transform these notes into blog posts. The scope of the website will be developed with the postdocs and PhD students of the SFB, but it will probably work both as an outreach inside and outside the SFG.

We can easily create a clone of it with restricted access once it is done, this might be done on the server by the way.

Future publication: thoughts

Since my output should be papers, I am trying to think about possibilities... One idea is to prove that creating figure-units at the end of the experiment is an efficient way to work and gives better quality data (in comparison with the usual way of writing papers years after the experiments were done). To achieve this comparison, we could plan to produce figure-units for each figure of manuscripts which are ready to be published or were recently published. We could then compare the material and method section with the data we would produce. It would have the nice side-effect to produce open data for these papers, maybe also creating "papers of the future" (http://scientificpaperofthefuture.org/spf.html). In the future, we would see how much faster it is to produce these figure-units just after the experiment is done (or while the experiment is done).

Other contacts

- I will talk with the people behind GIN to know more about it. wachtler@biologie.uni-muenchen.de
- There is another figure based tool in development, the founder is a friend of a friend and I will talk to him soon: nate@flashpub.io
- I am in contact with Peter Kraker and the go fair initiative. pkraker@openknowledgemaps.org