Committed to Science
– Scientific publishing in the web age –
DRAFT

The Open Science Community

Abstract

About this project:

- This file serves the collaborative drafting of a project proposal on an
Encyclopaedia of (and GitHub for) science,
as explained in this blog post.

- The .tex file was started by pasting below the leftovers from the drafting of the above blog post .

- We then turned this into LATEX format and plan to continue drafting that way.

- Unless something is clearly marked as being imported from elsewhere, all of this text is licensed CC0/Public Domain, while the LATEX code is available under the LaTeX license, with the origin being the Elsevier article bundle.

- You can **get involved**.

- **Submission of the proposal is anticipated for the end of July, 2011**

The real abstract, or at least some potential phrasing for it:

Research is the process of exploring and pushing the collaboratively defined boundaries of human knowledge through documented and contextualized observations.

Research publishing is the business of providing high-resolution snapshots of these steadily evolving boundaries. There are multiple problems with these snapshots:

- Their production takes so long that they are often already outdated when they hit the market.

- Their distribution uses sterile containers that typically bundle together numerous unrelated snapshots.

- Snapshots related to each other in space or time often end up in completely different containers.

- There is no way to zoom into (or out of) such a container in any way similar to a microscope, camera or Google map.

- There is no simple way of keeping track of new snapshots related to a given one. to tile humanity's wall of wisdom with a mosaic of the bigger picture, i.e. a snapshot of human knowledge as a whole, at any given time.

Encyclopaedias come close, and traditionally, they have been separated from the publishing containers used for original research. Technically, this is not necessary any more today, and removing this distinction may help, along with using version control, to alleviate the problems outlined above.

Key words:
Science as a wiki, GitHub for science, open access, Creative Commons, defragmentation of science, version control, digital encyclopaedia, digital collection, digital museum

---

Contents

1. Introduction

1.1. Name of the project

A bit of all of these:

- Science as a wiki

- Encyclopaedia of original research

- GitHub for Science

- Google Earth for science

- Science as it happens

1.2. Version control in scientific publishing

1.2.1. Main options

- Centralized version control

- Distributed version control

1.2.2. Precursor projects

2. Aims, goals and objectives

See also *the draft over at Species ID*

3. Timeline

4. Sustainability of the project

4.1. Sustainability of content

Key aspect: open licensing; starting with CC-BY.

4.2. Sustainability of code

Key aspect: open licensing; starting with the GNU General Public License (GPL) under which Mediawiki is available.

4.3. Sustainability of platform

4.4. Sustainability of proposal

The whole proposal has been drafted in public, so as to provide an example anyone can use to learn or teach about grant writing, to invite others to come up with similar proposals, to test the potential for public pre-submission peer review, and to stimulate the debate about doing science in the open.

5. Project team

5.1. Applicants

5.2. Partners

We have not defined any formal partnerships yet, but the following are amongst those we are considering:

- enspiral - partner for software development

- PubMed Central - content partner for seeding of the platform with content

6. Description of work

6.1. Work packages (subtasks)

6.2. Timeline

6.3. Deliverables

7. Resource requirements

8. Budget

9. Acknowledgements

People
Anyone who helped in one way or another

Tools
LaTeX, GitHub, Wikiversity, Species-ID, Google Docs, any other tools we used. Creative Commons, P2PU.

10. References

11. Figures

12. Notes from earlier stages of the drafting process

*Nice piece by David Dobbs about Jonathan Eisen's attempts to publish his father's papers online - could be cited on several points, including Knowledge Blog ("So what could be more fitting than to revamp science through a platform explicitly built to be revised, commented on, and updated?") and the four essential functions of science that are currently wrapped up in the scientific paper: registration, certification, dissemination, and preservation

Also cites Antonio Panizzi and mentions ADNI and Mendeley ("Many of the metrics and connections between papers arent accessible on the desktop, presumably because they require the servers data and processing power, and finding them on the web interface feels vaguely opaque.").

12.1. Possibly useful quotes

12.2. In focus: The encyclopaedia of original research

*Yesterday I asked one of my students if she knew what an encyclopedia is, and she said, Is it something like Wikipedia? . merging research projects, and linking them with each other as well as with the concepts and methods behind them.

¡del¿and with any info about the giants, on whose shoulders they have been built¡/del¿

¡del¿*Copying, forking (e.g. of this draft), ¡/del¿ ¡del¿*Mention "Science as a wiki" (including blog repository) and Wikis in scholarly publishing and "Towards threaded publications" ¡/del¿ ¡del¿;possibly embed Larry Lessig's talk at CERN, 18 April 2011 ¡/del¿ ¡del¿:Lessig's talk is licenced under CC-BY; could be used to highlight issues of license stacking and reuse, also with respect to the default license of the EOR¡/del¿

*EOR: earlier version

:comment on "Encyclopaedia" (incl. etymology/ kids; could go to the OLPC part), on it being open and on it being a federation of wikis

::::::¡blockquote¿"How cool would it be to fork articles, a la Github." - Jason Priem¡/blockquote¿
::::::¡blockquote¿"We need a GitHub of Science." - Marcio van Muhlen¡/blockquote¿
¡/del¿

12.2.1. Motivation

Including definition of goal. Follow SMART scheme.

12.2.2. Aims

12.3. Zooming in

Discuss what could become of the project ideas that won't end up in the final proposal, and how we plan to go about this decision.

12.4. Zooming out: Testing open vs. traditional science

*Do we need a ? (see also Panton Principles and Altmetrics manifesto)

*What online tools do scientists wish existed to facilitate their work?

:ORCID-coupled cross-platform reputation system

:see also this Nature News piece

12.5. Outlook

*Environments under consideration for drafting the full proposal: **A wiki (which)? **A Google doc (for export to which format?) **Prezi (if funder accepts that - we would like to avoid a rejection on formal grounds) (not suitabel for much more than just a presentation - really. Could be an accompanying feature, but not suitable to draft a formal proposal. Fabiana Kubke **(); for export to which format? **LaTeX Lab or ScribTeX ¡– How would you 'collaborate' on these? ARe there online versions?Fabiana Kubke 14:24, 3 May 2011 (CEST) **?

*Define a place where the actual drafting begins; ignore formatting for the moment; decision about writing environment due May 10

12.6. Notes

12.6.1. Quotes

"See also Collection of "science as a wiki" quotes."

* Sandra Bajjalieh: "All of these issues, including the trend towards judging scientists on where they publish instead of what they publish, would be solved if NSF/NIH provided and serviced a highly searchable website onto which people posted results as they obtained them. Search engine capabilities make this entirely feasible. The following features would make the system far superior to the current one of publishing in journals. 1. The comments of interested readers would be added to the posting. Thus there would be peer review. 2. Additional data and revisions that respond to comments could be added. 3. Entry time stamps would solve any issues of priority. 4. The number of "hits" and downloads a link got (similar to the information PLoS One provides for each paper) could serve as a measure of it's interest. This solution is so obvious and the benefits so numerous (NIH program directors would have current updates of research progress, no more publication costs, the ability to imbed movies and animations....)that it's really difficult to understand why there hasn't been more of a move to implement it. Do we, as a community, really want a few people regulating the flow of scientific information?" (Sandra Bajjalieh)

* Paulo Freire: "At the point of encounter there are neither utter ignorance nor perfect sages; there are only people who are attempting, together, to learn more than they now know."

* Larry Lessig's talk at CERN, 18 April 2011: http://vimeo.com/22633948 (Lessig's talk is licenced under CC-BY) :"copyright is a regulation by the state intended to change a regulation by the market; it's an exclusive right, it's a monopoly right, a property right granted by the state which is necessary to solve an inevitable market failure." ... in a more colloquial nutshell by Alex Pasternack: Copyright isn't just hurting creativity: it's killing science :: Notes: If we get above the din of this battle is that both sides agree that copyright is necessary for creative works - There is a place for sensible copyright policy but, however, not only artists rely upon copyright. Publishers do too rely upon copyright - the economic problem for publishers is different from that for the artists. We've been fighting a battle where copyright is essential but not on science where copyright is not essential. There is a trouble that few see - How accessible is information for the public? What does it mean for info to be available on the internet? It is only freely accessible if you are part of the 'elite'. Here copyright is placed to benefit the publishers - not the authors - no author has a business model that is built around profiting from this copyright. Does this limitation serve any of the purposes of copyright? What is the publishers objective? To disseminate knowledge or to profit from it? ::JSTOR archive: has

become increasingly criticized because of the cost involved in accessing the articles in the archive. ::Lessig asks: Can we do better? ::Open access self archiving movement ::Open access publishing movement: ::Some open is free (as in free speech) some open access is free as in you dont pay for it but other copyright rules apply. ::Science Commons: "broader strategy for producing the information architecture that science needs" as per the 4 principles of Open Science (check on site). *Read write creativity / read-write communities *"Sharing is at the core of the architecture of the net" ::Note by Claudia Koltzenburg: [by whom?] Barbara van Schewick points out: On the Internet architecture level, due to corporatism, "enclosures" are rampant. The principle of network neutrality that characterized the Internet in its beginning, thirty years ago, has been put at risk particularly by profit-making interests of network providers, observes Barabara van Schewick. The effects of this amount to what economists who think in terms of traditional market economy would call a "market failure". Van Schewick holds that we (and the regulators) need to protect the factors that allowed widespread application innovation in the past (modularity, layering and the end-to-end arguments). These factors made for the openness at the core of the Internet until the early 1990s. Van Schewick recommends let users choose, and practice as much 'application agnosticism' as possible. Internet users today are mostly controlled by flatrate offers and application bundles that leave no alternatives to choose from openly. Van Schewick's argument says that users should indeed be allowed to get a sense of how much they need for what they want to do on the internet ?? and yet maintain a predictability of one's bills. – see Barbara van Schewick. Introduction. In: Internet Architecture and Innovation. Cambridge, Massachusetts/ London, England: MIT Press, 2010, 1-15. (see also Internet Architecture and Innovation) – ""'in this vein, what is the "flatrate" in academic/scholarly/scientific publishing today that lures into control?""' – Claudia Koltzenburg 12:12, 1 May 2011 (CEST)

*"In the academy [..] we need to recognise an ethical obligation [...] which is at the core of our mission which is universal access to knowledge." Entails: work needs to be free (this should be an ethical point) - We do not need (and should not practice) exclusivity about our work. *models of access that block access except to a paying elite and discourages innovation.

* Dorothea Salo: "At the risk of sounding all commie and stuff: we work toward a collective openness, or we die off one by one as the business model sustaining us as well as publishers crumbles to bits."

* Douglas Rushkoff: As soon as a network is in the hands of policy makers and their funders, this network loses its power to effect change. His conclusion is: "Create new forms that exist beyond any authority's ability to grant them protection", The Next Net. 1 March 2011. Shareable - Sharing by Design.

Vannevar Bush. As we may think. The Atlantic Monthly, 1945 *"There is a growing mountain of research. But there is increased evidence that we are being bogged down today as specialization extends. The investigator is staggered by the findings and conclusions of thousands of other workers - conclusions which he cannot find time to grasp, much less to remember, asthey appear. Yet specialization becomes increasingly necessary for progress, and the effort to bridge between disciplines is correspondingly superficial" *" Professionally our methods of transmitting and reviewing the results of research are generations old and by now are totally inadequate for their purpose.:

*". The summation of human experience us being expanded at a prodigious rate, and the means we use for threading through the consequent maze to the momentarily important item is the same as was used in the days of square-rigged ships." *" A record, if it is to be useful to science, must be continuously extended, it must be stored, and above all it must be consulted" *"Thus far we seem to be worse off than before - for we can enormously extend the record; yet even in its present bulk we can hardly consult it. This is a much larger matter than merely the extraction of data for the purposes of scientific research; it involves the entire process by which man profits by his inheritance of acquired knowledge."

12.7. Draft of actual proposal

"Use SMART(ER) approach: Specific/ Measurable/ Agreed/ Realistic/ Time constrained."
*Mention subprojects

12.7.1. Potential problems

*Barriers to expert participation **Giving credit is key, and if wiki contributions (or any other science 2.0 activities) would be recognized in academic career terms (#altmetrics; requires), scientists would be

willing to reallocate their time accordingly

*"'What does "publishing" mean in a wiki context?"' The current use of the term "publishing" in itself can be taken as an illustration of how commercial codes and practices seeping into academic culture have not been counteracted successfully since the invention of the Web. In academic CVs, research output in print only (and in electronic but non-open access format) still figures as a "publication", even though the meaning of "to publish" as "make generally known" and "disseminate to the public" has seen fundamental and indeed groundbreaking changes with the Web as a publishing platform. Indeed, "the public" itself has changed fundamentally because today, a "publication" can be made accessible on the web "'without"' . Had academic institutions been more interested in the benefit offered by such opportunities, publishing openly would be much more widely accepted today. In this light, nothing should be claimed to be a "publication" any longer unless it is open, maybe even "'in the sense of Open publishing"': "Open publishing is a process of creating news or other content that is transparent to the readers. They can contribute a story and see it instantly appear in the pool of stories publicly available. Those stories are filtered as little as possible to help the readers find the stories they want. Readers can see editorial decisions being made by others. They can see how to get involved and help make editorial decisions. If they can think of a better way for the software to help shape editorial decisions, they can copy the software because it is free and change it and start their own site. If they want to redistribute the news, they can, preferably on an open publishing site."

12.7.2. Notes

- Wiki stats tools: Article-level traffic stats, Trending topics, Edit stats, Edit stats for new pages :See also MyADS in astronomy

- virtuelles Museum

- Museum fish MRI & Tierstimmenarchiv

- Micropayments for culture - similar for science

- using Google docs

- Open Science Games? Any equivalent to open vs. public peer review?

- eResearch talk by Mark Gahegan

- , via Twitter

- Filipe Cruz, in Skype chat of May 6, at 19:11 - superfabs: there are a few sites dedicated to harboring science papers and journals in digital free for download formats. would be nice to do a list of them atleast, to analyze and figure out how to better complement them? its similar work i think :Link provided a bit later: http://xdatelier.org/2010/12/11/open-access-repositories/

- More from that chat: scannopolis 19:16 @superfabs, I think that the aim of the project is too wide maybe? Fabiana Kubke 19:17 @scann can you be more specific? Jonas Öberg 19:18 Worthwhile to note in this discussion is that Paul Boshears has been thinking about a project to extend and make Open Journal System (OJS) more suitable for accepting non-text material. Might tie in with other open access projects. scann 19:18 @fabs, yes, I think maybe you need to define what are you going to consider "science", or in which areas you are going to primarily focus on to work 19:18 for example, the scientist who come from humanistic or social sciences network are more resistant to work with things like wikis Ian Sullivan 19:20 and the scientists from many of the hard sciences will need hosting for the large datasets that make up a lot of their source material Fabiana Kubke 19:20 @scann Ah, I see - I was thinking about that earlier today - how do I phrase this to say "as a first step we will do this in this area" - I thought concentrating on one specific (I was thinking Chagas would be a good candidate - I am more familiar with some implications)

- Ian Sullivan: http://grants.gov/ in the US has a comprehensive listing for all grant opportunities open at the federal level

- http://www.skollfoundation.org/

- https://creativecommons.org/weblog/entry/23831

- Jamendo - company based on distributing CC-licensed music

- FigShare as virtual museum?

- Discussion of Gantt chart tools

- Community building in ecology

- Use case (from Wilbanks talk): Reverse Causal Reasoning

- Open science article in the Guardian

- "open-access repository for all research findings, which would let scientists log their hypotheses and methodologies before an experiment, and their results afterwards, regardless of outcome"

- Bruce Alberts "Our goal as teachers and educators should be to expose our students to the discovery process and to excite them about challenges at the frontiers of knowledge." (B. Alberts, "A Wakeup Call for Science Faculty", Cell, vol. 123, 2005, pp. 739-741. DOI:10.1016/j.cell.2005.11.014) Also: "Old habits die hard, and I have been disappointed to discover that this is especially true in academia.", from the same source

- Wiss-ki

- Note to self: For basic help with GitHub, see http://help.github.com/git-cheat-sheets/ .

13. Potential funding schemes

13.1. Calls for proposals

-

13.2. Funders with good match in scope

-

13.3. Prizes and competitions

- NIH reuse app challenge

13.4. Microfinancing

Invite crowd-sourcing, with link to a description of a project already funded by that source and ideally with some overlap to the current proposal. What role do funders have to play in bringing research into the web age?

- Startl - startup support for socially responsible businesses

- Bitcoin; option for mining

- KulturWertMark

- Kickstarter - could the subprojects perhaps be submitted there, or to similar places (indiegogo, or rockethub)?