

FORCE11 1k Challenge (#btpdf2 #1k)

What would you do with 1k today to make research communication better that doesn't involve building another tool?

Put your proposal below and we will try to select after the conference. Please put links and divide up you sections.

Open Scholar Foundation (Tobias Kuhn)

I would set up a simple "Open Scholar Foundation" with a website, where researchers can submit proofs that they are "open scholars" by showing that they make their papers, data, metadata, protocols, source code, lab notes, etc. openly available. These requests are briefly reviewed, and if approved, the applicant officially becomes an "Open Scholar" and is entitled to show a banner "Certified Open Scholar 2013" on his/her website, presentation slides, etc. Additionally, there could be annual competitions to elect the "Open Scholar of the Year".

Starting at ground zero (Melissa Haendel)

What would I do with 1K?

I would reward in 50\$ disbursements 15 graduate students/post docs for attending two sessions in the library. The first session would include a 1 hour summary of the beyond the pdf2, highlighting aspects of the data-research cycle in which there are issues surrounding research reproducibility and scholarly communication of findings. Extra money here would go towards distribution of key materials and food/coffee and any extra would be spent on more participants. The second session would be a 1hr hands on session with Library staff, where the participants would be asked to bring a yet-to-be published data set and/or publication. During this session, we would do the following: (1) Determine which aspects of the data require standardized metadata for sufficient reporting and reproducibility, e.g. perform data review; (2) Determine if there is a public repository would the data be relevant and develop workflow to deposit it there; (3) Teach information management strategies - referencing uniquely key aspects of the data or research context, version control, and linking methods and conclusions to these uniquely referenced and versioned aspects of the data; (4) Determine requirements for libraries to develop training materials and services to aid researchers in the production of reproducible science and quality data, especially as it pertains to communication of research findings. Overall

goal is to learn how to promote interaction between information scientists and research scientists, and to improve research reproducibility.

Tool for viewing and browsing ALL open access journals complete with summaries(Andrew Varnell)

I would like to make a webtool/app that brings together all open access Journals using simple RSS technology. Additionally, I would like users to be able to search/sort/view on a number of conditions. With each article, we would pull key sentences such as 'we found...' to automatically create a 5 sentence summary of the article. This would also act as a commenting hub. I see this as a valuable project because as it stands, finding and using open access is a lot of work. If we could consolidate the 100's of open access journals in a single, user-friendly site, OA articles would have a much greater chance of gaining visibility and feedback. The nice thing about this project is that it would not require users to learn any new technology in order to start accessing and using OA. If a large following adapts, we could start incorporating some of the exciting citation tools we have learned about, full search capabilities based on our own users data, etc. The spirit behind this project would be collaboration and pooling resources for a 1 stop OA shop, rather than have a number of good tools scattered around the web.

Initializing reproducibility (Melissa Haendel and Theo Bloom)

We would recruit a panel of graduate students and post-docs to review materials and methods sections in their domain specialty for a select number of PLOS journal submissions. The goal is to develop a review standard for material resources to support resource reproducibility. They would receive credit for serving on the panel by creation of a review record that they could cite (if they wished). While there are numerous "data standards", there are a few that actually represent the material resources and none that are used as part of the peer review process. The goal here would be to initiate good practice in citing materials, protocols, and resources as well as data. This will ideally lead to a better reporting cycle with respect to research resources, research reproducibility, data linking, resource attribution, and inference of new knowledge using semantic techniques.

Utilizing the ORCID Id to establish an Open-Review system (Andrew Varnell)

I would like to propose a project that would work the the ORCID system API to create a

open-review network that could be utilized for any open-access medium. We would allow users to extend their ORCID id to fields of specialty based on current publications they have listed. Using keywords from the titles of their own publications, they would be allowed to review similar works. This open-review system would be able to be placed in either a webpage or any digital science medium.

Analysing Materials and Methods from the Open Access subset of PubMed Central (Alexander Garcia)

Biotea (<http://biotea.idiginfo.org/>) is a semantically processed, fully annotated, version of the open-access subset of PubMed Central. I would like to extract from the Biotea dataset all the articles with a section “Materials and Methods”. I would then process these documents as follows. Firstly, I would semantically analyse “Materials and Methods”; from this analysis I would identify, most common words used and most common ontologies involved in the description of “Materials and Methods”. I would annotate this dataset using the Annotation Ontology and expose it over a SPARQL endpoint. Secondly, I would run a linguistic analysis over the dataset, trying to identify actions, adverbs and adjectives; I would like to know how are these linguistic elements related to those words that come from ontologies and what linguistic structures are most common in Materials and Methods. All of the above will give us information about the reality of Materials and Methods, are measures mentioned? How? What actions are indicated? How? subsequent analysis could focus on replicating experiments just by using the information provided by Materials and Methods, is this possible?

SMART protocols for 100 “Materials and Methods” with Smart Protocols (Olga Giraldo)

I would like to have 5 graduate students processing 100 pre-selected published papers; they should focus on “Materials and Methods” for reproducibility purposes. I would like them to represent “Materials and Methods” as Smart Protocols. If the information in “Materials and Methods” is not enough then they should look into the rest of the paper. I would use the US\$ 1000 as an incentive for the students participating in these work sessions; an ipad would be raffle among them. The main aim of this workshop is to use Smart Protocols in order to represent the experimental protocols from the analysed papers. In this way I would like to accomplish: i) experimental validation from an independent party for Smart Protocols, ii) enhancing Smart Protocols by eliciting knowledge from these sources (100 pre-selected published papers), iii) generate RDF, and expose over an SPARQL endpoint the resulting “Smart Protocols” dataset, and iv) gather/provide evidence for the usability of Smart Protocols as a valid approach for representing experimental protocols and making reproducibility a reality

in scholarly communication.

The New Scholar Challenge at Sepublica2013 (Alexander Garcia, Robert Stevens, Phillip Lord, Christoph Lange)

We would use the award, US\$1000, to sponsor a challenge at the Sepublica Workshop. How are we changing scholarly communication? How are we promoting author led publishing? How are we organising and promoting change independently from third party interests and focusing on requirements that we as researchers have. We would like to award the most outrageous ideas presented in this challenge.

Understanding the scientific communication that we already have.

Phillip Lord

Scientific communication is already better, and it is already happening. It is present on the web in numerous websites, wiki's and blogs where scientists describe their own work, and their views others. This is where publication is being used to purely to communicate, rather than as a badge of honour, born by the necessity for assessment.

Unfortunately, this form of publication is not treated with seriousness. Pubmed will not index them; libraries will not archive them as scientific journals. It is grey literature. So, with 1k, we will pay for two crawls from www.archive-it.org; we will use this as an inducement for scientists to submit a link to their website, and we will archive all of these after 6 months and 1 year.

This will give two main outputs: archives that will be accessible through archive.org, and as a separate collection from archive-it, open for analysis. And a collection of links to sites where academics are already using the website to communicate.

“What would you do with 1k today that would change scholarly communication for the better? ”
Challenge entry by Kriste Krstovski

Current scholarly article retrieval systems cover many scientific disciplines ranging from the life

sciences and biomedicine (PubMed), physics and astrophysics (ADS), computer science (ACM Digital Library), electrical engineering (IEEE Xplore), and mechanical engineering (ASME Digital Library), to name a few. Each of these, and many other scholarly literature retrieval systems, employs its own user interfaces and ranking algorithms which in most cases rely on representing documents using the vector space model where most of the representation is done with term-frequency vectors and/or are augmented by eigenvalue or singular value decomposition. Since the first introduction of the vector space model for performing document similarity comparisons, variations that offer more semantic representation have been proposed. Recently, there has been a lot of attention given to statistical topic modeling such as latent Dirichlet allocations (LDA) which have proven to be highly effective at discovering hidden structure in document collections. While these and other recent state-of-the-art latent variable models have in the past proven useful in various text mining tasks, these models have not as yet been fully exploited on big, real-world data sets such as scholarly article retrieval systems. Thanks to the generous support of the Force 11 movement, I recently had the opportunity to learn first-hand the challenges and obstacles found in the current scholarly article retrieval systems. I also had the opportunity to interact with developers and researchers that deal with issues in scholarly retrieval systems on daily bases.

With the knowledge and experiences gained at the Beyond the PDF 2 conference and my research efforts in the field of computer science, I would like to use the finding provided by this challenge to utilize and explore current state-of-the-art techniques found in the fields of machine learning and natural language processing (NLP) for the purpose of improving many aspects of the current scholarly retrieval systems and to further my current research efforts on utilizing latent variable models in the SAO/NASA Astrophysics Data System (ADS). First and foremost, I would like to promote the issues and challenges in the scholarly retrieval systems in the community of computer science researchers. One way of doing this would be to pursue the effort to organize a workshop in one of the leading natural language processing conferences. The idea of organizing one such workshop was well received by Prof. Eduard Hovy and Prof. Graeme Hurst in my recent communication with them during the Beyond the PDF 2 conference.

Tweets:

Juan Pablo Alperin @juancommander 19 Mar

we could pay for #crossref membership for low-income countries so they can have DOI's and take advantage of more of these tools #btpdf2 #1K

Daniel O'Donnell @DanielPaulOD 20 Mar

#btpdf2 #1k establish collaborations between OA journals in mid and low income economies w journals/pubs in high income. Learn from e.other

Martin Fenner @mfenner 20 Mar

David Shotton mentions Open Citations project at #btpdf2. Great project to open up the existing citation graph <http://opencitations.wordpress.com>

Merce Crosas @mercecrosas 19 Mar

#btpdf2 1K challenge idea: publish common standard guidelines for referencing data

#datacitation endorsed by publishers and scholars

push Lukas Koster @lukask 20 Mar

@carlystrasser @datacite @TAC_NISO And the DCC @digitalcuration dataset citation guide

<http://www.dcc.ac.uk/resources/how-guides/cite-datasets> ... #btpdf2

Eugenio Battaglia @battagliaem 20 Mar

#btpdf2 #1k for #scholrev olutionaries, addressing the open science challenge!

Force11.org @force11rescomm 20 Mar

@cameronneylon #1k challenge? have heard several #btpdf2 armchair pitches for "locked in a room" "on a lifeboat" figure it out notions...

Heather Piwowar @researchremix 20 Mar

#btpdf2 #1k A YC/techstars incubator for scholarly communication startups. Mentorship, leg up in biz, marketing, funding, recruiting, etc.

Claudiu Mihăilă @ClaudiuMihaila 20 Mar

@ontowonka @jeroenbosman give them #1K to help improve their IR system; improves discoverability & scholars will return to them #btpdf2

Stian Håklev @houshuang 20 Mar

Hackfest to create tools/workflows/documentation on using Scholarly Markdown+Git for academic authoring/collab #1k #btpdf2 @CameronNeylon

Stian Håklev @houshuang 20 Mar

Haven't talked much about authoring tools/workflows here. Anyone excited about scholarly Markdown? <http://blogs.plos.org/mfenner/2012/12/13/a-call-for-scholarly-markdown/> ... #btpdf2 @mfenner

"what would you do w/ \$1K to change #scholarlypub?" asks @pgroth at #btpdf2. A: experiments/prototypes, like #OA Resolver / @OARReR

Ian Mulvany @IanMulvany 20 Mar

#btpdf2 1k challenge idea, allow publication of single peer reviewed data sets and figures. Accelerate storytelling. Lightweight pubs.

Stian Håklev @houshuang 19 Mar

Give grad students researching in other countries grants to help translate thesis #1k #btpdf2
<http://reganmian.net/blog/2011/06/01/chinese-translation-of-ma-thesis-on-top-level-courses-available/> ... <http://reganmian.net/blog/2008/03/07/a-fair-trade-logo-for-academic-research/> ...

Shashi Mudunuri @ShashiAJE 19 Mar

#btpdf2 #1k Give two grants to generate #rubriq scores for papers. Help change peer review for the better two papers at a time.

Chris Evelo @Chris_Evelo 19 Mar

.@kaythaney I would probably do what @pgroth did, spend the 1k on a challenge, making many ppl think about science communication

Peter Murray-Rust @petermurrayrust 19 Mar

@pgroth #btpdf2 asks what would we do with 1K dollars to change #scholarlypub? I'd give it as prizes for undergraduate competition

jeroen bosman @jeroenbosman 19 Mar

#btpdf2 #1K idea for impr sch. comm: let's all switch to either APA, Chicago or Nature/numbered style and ditch the 1000s others (sorry MLA)

The Amsterdam Manifesto on Data Citation Principles

Proposed by Mercè Crosas, Todd Carpenter, Christine Borgman and David Shotton

We apply to the **Force11 1K Challenge** for funds to provide the manpower and resources to set up and manage a web site on which to post the new **Amsterdam Manifesto for Data Citation**(https://docs.google.com/document/d/1ON0yy2_jT2VxL_Cdm03HgMSNnN1A6VzvDQrTi577-ig/e/dit) developed during the Beyond the PDF 2 Conference.

The Amsterdam Manifesto web site should be linked to from the Force11 site, or could even be a new section within the Force11 site, if that is more cost effective.

The site will have the following features:

- 1) A page displaying the manifesto (based on the example provided by the Panton Principles - <http://pantonprinciples.org/>).
- 2) A page permitting individuals to sign / endorse the manifesto (see <http://pantonprinciples.org/endorse/>).
- 3) A page for comments, where visitors can add references to other relevant information or additional suggestions for data citation (see <http://pantonprinciples.org/comment/>).

We would also like to ask Force11 to help us in reaching out to publishers, scholars and relevant organizations by announcing the creation of this web site and encouraging the endorsement of the Amsterdam Manifesto.

