

Phase II Competition - Finalist Application Form

Created: 10/24/2016

Last updated: 11/22/2016

In this form, the six finalist teams for the Open Science Prize are asked to describe the work they have done to develop their prototypes, and make the case for why they should be considered for the phase II Prize.

The information you submit on this form will be considered alongside the prototype you have developed in deciding which team will receive the Phase II Prize.

Please note that, unless otherwise stated, the information you submit on this form will be made available publicly via the Open Science Prize website (under a [CC BY 4.0 license](#)), so that it can be assessed as part of the public voting process. All fields, except the final box for additional information are mandatory.

Your application must be completed by 11:59pm Pacific Standard Time on 21 November 2016 . You may edit this form as many times as you like before the deadline.

If you have any questions about the Prize or the review process or if you would like to provide any further information that you would not wish to be made public, please contact David Carr (d.carr@wellcome.ac.uk) or Elizabeth Kittrie (elizabeth.kittrie@nih.gov). Any technical questions regarding this form or the web platform should be directed to (openscience@wellcome.ac.uk).

Page 1

Executive Summary

Please provide a brief Executive Summary of no more than 150 words for the public voting page on the Open Science Prize website. This should be suitable for an informed lay audience, and should briefly describe your prototype and why it should be considered for the phase II Prize. [150 words]

OpenTrialsFDA increases access, discoverability and opportunities for re-use of a large volume of high quality information hidden in user-unfriendly Federal Drug Administration (FDA) drug approval packages (DAPs). These documents contain detailed information about the methods and results of clinical trials. These review packages also often contain information on clinical trials that have never been published in academic journals.

Currently, FDA documents are notoriously difficult to access, aggregate, and search. Our prototype

enables researchers and clinicians to search and access the clinical trial FDA information via a user-friendly web interface. Application Programming Interfaces (APIs) also allow third party platforms to access, search, and present the information, thus maximising discoverability, impact, and interoperability. The innovation presented here is a new application of existing open data techniques and code to resolve an important problem for evidence-based medicine, which will have a positive impact on clinical decision making globally, and so improve patient care.

Weblink for prototype

Please provide the public URL for your prototype tool or service (this will be viewed by the public for purposes of public voting):

Links include exposing the data as an API for 3rd party programmatic use, as a search UI, and example URLs matching data from FDA. Web App to Search FDA DAPs <https://fda.opentrials.net>
Web API to search FDA DAPs <https://api.opentrials.net/v1/docs/#!/search/searchFDADocuments>
<https://api.opentrials.net/v1/swagger.yaml> Integration of FDA DAPs into OpenTrials platform
(matching data from FDA with data we already have on clinical trials)
<https://explorer.opentrials.net/trials/46570f29-927f-49a8-8b1c-a35c0053b0aa>
<https://explorer.opentrials.net/trials/69e7f092-7e78-48dd-8e5b-6661b527be00>
<https://explorer.opentrials.net/trials/f8f28289-ffa4-43f7-88ec-b8d95e03bcff> Crowdsourcing task to extract indications information <http://crowdcrafting.org/project/opentrials-fda-indications/>

Your Prototype

Purpose and need

Please provide a brief summary description of the purpose of the prototype you have developed and the key challenges or needs it is seeking to address [200 words]

A trove of unbiased clinical trial data exists in the form of FDA DAPs. These contain medical and statistical reviews of clinical study reports submitted by drug companies seeking to have their drugs approved for the US market.

An FDA review reflects an awareness of the trial's existence, (so sponsor decision not to publish the trial is immaterial) and the protocol-prespecified primary outcome and statistical analytic plan (which precludes post hoc outcome switching/HARKing).

While technically 'publicly available', the information in DAPs is close to unusable for clinical trial research:

Most of the data is scanned from physical documents into image format, so it is not accessible and searchable in a machine-readable form

The navigation experience to discover data is challenging

The data is not indexed or searchable via clinical trial identifiers

OpenTrialsFDA has not only made FDA DAPs more accessible for reuse, it has added value to this data by matching it with data that the OpenTrials team is already scraping, hosting, and indexing on the database of the OpenTrials project. As all code and data created as part of OpenTrials is openly licensed, this increases the potential for reuse of the FDA DAP data in new, innovative ways.

Please summarise the work you have taken forward to develop your prototype since you were awarded the Phase I Prize in April 2016.

i. Progress

The key milestones achieved and the extent to which the goals and challenges you set out to address in your original application were delivered [400 words]

1. Wrote a code base for scraping data and files from Drugs@FDA. This continuously runs and updates as new data is published. We acquired all metadata on drugs, and downloaded all documents available on Approval History and Related Documents pages, including Letters, Labels, DAPs, and other files.
2. Performed automated text extraction on files using advanced OCR. The we built a search index over the metadata and all the text that could be extracted using this method, across all documents.
3. Utilised our large search index to run matching algorithms that search for mentions of clinical trial identifiers that are already recorded in the OpenTrials database (<https://opentrials.net/>). This set of identifiers has been acquired from a range of sources, such as ClinicalTrials.gov, EU CTR, WHO's ICTRP.

4. We identified a significant area for further investigation. It was only possible to match 58 DAPs onto known clinical trial identifiers. This is an ideal focus area for a further phase, in which we can improve our OCR methods, and the type of matching algorithms we use. It may also point to an interesting area for study where matching quality can be improved if public resources like OpenTrials have access to internal company identifiers for trials.
5. Attempted to use our search index to match indications data onto the body of clinical trial data we have in OpenTrials. This proved difficult, as there is no structured information on the indications that a document sourced from the FDA relates to. In response to this difficulty, we implemented a prototype crowdsourcing task to start to extract this information with an interested public. Again, this is an ideal area for further work with more funding.
6. Exposed our FDA DAPs, documents and metadata search index as an API for programmatic use. For the first time, it is possible to build 3rd party applications on top of this rich resource from the FDA.
7. Created a user-friendly search interface over our entire search index, which allows non-technical users to browse and discover this data with ease.
8. Made changes to the main OpenTrials platform to make this FDA data easily usable in the context of the wider body of information that OpenTrials holds on a given clinical trial. This can be seen in the examples provided, which are indicative of the matches that we could successfully make within the scope of this prototype.

ii. Team Contributions

The contributions of the team members to the development of the prototype [200 words]

The OpenTrialsFDA project is led by two experts: Dr Erick Turner (OHSU), a psychiatrist-researcher, former FDA reviewer, and transparency advocate, participating as an individual from Oregon, USA, our US-based expert; and Dr Ben Goldacre, a Senior Clinical Research Fellow in the Centre for Evidence Based Medicine at the University of Oxford, our UK expert.

The OpenTrialsFDA prototype was built by the team at Open Knowledge International, consisting of tech lead, Vitor Baptista, product owner, Paul Walsh and two developers, Victor Nitu and

Georgiana Bere. Emma Beer was the project manager, while Lieke Ploeger has worked on communications alongside Ben Meghreblian, our community manager.

iii. Significant Achievements

Any significant achievements or key success metrics you wish to highlight - this might include, for example, numbers of users, key endorsements or engagements with users, new partnerships, external funding, and so forth [200 words]

A rapid turnaround prototype has enabled us to demonstrate the value of collaboration across countries, disciplines and data to provide a high impact solution.

We ran a series of blogs and commentary on Twitter (#opentrialsFDA) tracking the development of the prototype as well as covering key FAQs.

The OpenTrialsFDA project was also presented on as part of an OpenTrials presentation at the Cochrane Colloquium in Seoul, South Korea on October 25, 2016. Two articles were published on the project.

Oregon Health and Science University (OHSU) will be hosting an OpenTrialsFDA Data Jamboree on 16 December, 2016.

Blogs:

Nov 8, 2016: OpenTrialsFDA - Frequently Asked Questions [will be republished on prototype site when it's ready]

Oct 25, 2016: OpenTrialsFDA: Jeppe Schroll on the value of regulatory data

Oct 5, 2016: OpenTrialsFDA: an interview with Erick Turner

Aug 10, 2016: OpenTrialsFDA: Unlocking the trove of clinical trial data in Drugs@FDA

May 9, 2016: OpenTrialsFDA selected as finalist in Open Science Prize

Articles

May 9, 2016 'OHSU doc, frustrated by clunky FDA database, wins grant to make it more accessible' in Portland Business Journal, <http://www.bizjournals.com/portland/blog/health-care-inc/2016/05/ohsu-doc-frustrated-by-clunky-fda-database-wins.html>

August 4, 2016: 'OpenTrialsFDA' Could Allow Research on Product Approval Packages.

Learning Points

Please briefly highlight any key learning points you took from the work that you undertook to develop your prototype [200 words max]

We knew that the accessibility of these documents was an obvious pain point for researchers in general, but we did not anticipate how difficult the Drugs@FDA website would be to scrape (our technical team considers it the most difficult site they have ever scraped data from).

While the data in scanned documents is of reasonable quality, it is still a challenging source for what is very important data, and extracting what should be structured data from tables in scanned documents is a significant challenge.

Making linkages from our search index right back to specific pages in scanned documents that are the source of the data is a technical challenge, yet one we can improve on with additional work.

The FDA recently announced “a new look and a new web address”; this (minor) update means we will have to revise our data acquisition methods

Case for Phase II Prize

Please make the case for why your prototype should be considered for the Phase II Open Science Prize against the following key criteria [100 words each]:

i. Impact

The current and potential future impact of the tool or service in terms of advancing research and generating health and societal benefit

While the FDA is currently beta testing APIs to improve access to a small proportion of its data, no APIs exist to access the efficacy and safety data in FDA DAPs.

OpenTrialsFDA will enable academic researchers to:

-access and search unbiased descriptions of what happened in clinical trials of drugs used by billions of patients in the US and worldwide

- reveal discrepancies between clinical trial data in FDA DAPs and published journal articles
- access data on entire clinical trials whose existence cannot be ascertained elsewhere.

Integrating this data on the OpenTrials platform will also enable multiple new activities previously impossible.

ii. Innovation

The degree of innovation associated with the tool or service

As far as we know this is the first time the text of the FDA's Drug Approval Packages have been made available for search and matching with the support of domain experts. The innovation will provide the research world with important information on clinical trials, improving the quality of research, and allowing evidence-based treatment decisions to be properly informed by a complete and unspun evidence base.

In addition to making the FDA data available and accessible on the OpenTrials platform, all code is published under an open source license. All additional data created will be licensed for permissive, open reuse.

iii. Utility

The level of demand and utility associated with the proposed service or tool

Strong connections with key communities will ensure that the results of the project are taken up and adopted. Should we successfully pass to Phase II of the Prize we would like to continue to iterate the OpenTrialsFDA prototype to make sure that it is used and useful to researchers, clinicians and patients. We would begin intensive user testing with representatives from these groups. We will be able to implement user testing sessions as we have for OpenTrials, allowing us to build out new feature requests, iterate and improve for all users, ensuring that we optimise the utility of the prototype.

iv. Feasibility & Technical Merit

The feasibility and technical merit of the prototype

The prototype exists - we have successfully:

- Built a working, friendly UI
- Imported, indexed, and made searchable over 55,000 documents relating to DAPs

The prototype works: currently you can:

- Search the entire text of all documents
- See initial matching of documents to trials in the OpenTrials database
- Search within all documents of a certain type (e.g. Medical Review or Statistical Review)
- Search for a drug by its brand name or generic name
- Search across all documents and metadata programmatically using our API, facilitating third-party analysis and extending functionality
- Use our prototype UI for crowdsourcing indication information

Development & sustainability plan

Please briefly describe your vision, and any tangible steps you have taken, to develop your prototype into a sustainable tool or service that advances the goals of open science [400 words]

We have taken inaccessible unsearchable documents, OCR'ed them, and made the resulting data open and public. It will remain open and public, and searchable. However, the lesson of the poor quality sharing at the FDA's own website is this: there is a big difference between information being strictly publicly available at all, in any form; and information that is shared in the most accessible, searchable, usable, and high impact fashion possible.

Should we be successful in proceeding to Phase II, we would like to do the following:

- Increase the number of documents matched with the OpenTrials database by improving our search, OCR methods, and matching algorithms, so that the documents are more discoverable, and linked to other descriptions of the same trials and treatments.
- Implement the crowdsourcing of elements of the FDA documentation
- Ensure sustainability by bringing it into the larger OpenTrials project more seamlessly.
- Improve the search and annotation interfaces.

Making the data discoverable in the context of OpenTrials, will thus allow users to mine the data, uncover new associations and discoveries, and extract new value that was previously hidden. This has the potential to make significant contributions to biomedical research and its healthcare applications, increase transparency and data credibility, decrease medical and public health risks and generate societal benefits yet unknown.

Much richer information is still to be found within the documents. This includes unbiased data on

clinical efficacy and safety, which is crucial to work of physicians and other prescribers, who are charged with making fully informed prescribing decisions. Such data are also of interest to the patients who want to be fully informed about drug efficacy and safety before taking them. In addition to the more clinically oriented medical and statistical reviews, there are other review types, which will likely be used by future researchers. These reviews are authored by professionals representing other disciplines such as chemistry, pharmacology-toxicology (preclinical safety), and biopharmaceutics (which deals with drug absorption, distribution, metabolism, excretion, interactions, etc).

In order to advance the goals of open science, we would also like to write about our experiences in building this product, particularly in the value of academics and coders collaborating on tools to solve solutions, rather than in just publishing academic papers. We believe these types of collaborations allow us to build solutions faster that have a wider applicability and impact than academic papers. We believe this is the future of open science.

Final comments

Please use the box below to provide any further information you would like to add, that has not been addressed in the questions above [200 words]

Finally, Open Science as a discipline and as a movement will be furthered by a model innovation unlocking hidden data treasures and improving science-based decision-making through openness.