Phase II Competition - Finalist Application Form

Created: 11/18/2016
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]

The Open Neuroimaging Laboratory is a project to facilitate finding, improving, and reusing the massive amount of brain MRI data available online. This data represents an enormous funding effort and the work and goodwill of thousands of participants. BrainBox, our first application, transforms these static MRIs into "living" matter for collaborative curation and analysis using only a Web browser. Users can work, discuss, edit and annotate MRI images simultaneously. No data are downloaded, no software installed, allowing users to incrementally improve each other's work. This increases scientific efficiency, improves public data quality, and reduces redundant effort. We

already index more than 6000 MRIs, which are ready for collaborative projects. To navigate this resource we developed MetaSearch, a meta-data filtering tool which makes it easy to find relevant data and projects to participate. Our virtual neuroimaging laboratory lowers the barriers for researchers, students, and citizen scientists to help scientific discovery.

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):

http://openneu.ro/start

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]

Significant investment has been made into collecting and sharing brain imaging data for thousands of individuals. One key challenge, however, limits the usability of this data: In order to work with it, researchers need to download the data locally; and curation, editing and analysis are then done redundantly by each research group. This painstaking process results in a large proportion of the shared data from being analysed, wasting time and funding. During phase 1 we built BrainBox – a Web app to work with shared brain imaging data directly online. Progress becomes incremental: each researcher's work improves that of the whole community. The community can then tackle projects that would be impossible for individual research groups. BrainBox enables a growing catalogue of all MRI data available online, currently indexing more than 6000 links. Our tool MetaSearch provides an interface for filtering metadata associated with these brain images and finding relevant data across all different projects. Our virtual neuroimaging laboratory enables using and enhancing open resources and social collaboration; and by lowering the requirements for working on shared data to just a Web browser, it allows not only researchers, but also citizen scientists to contribute to scientific discovery.

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]

Our initial prototype had limited visualisation and collaboration functionalities, which we proposed to improve. Our work during the last months has accomplished and exceeded this objective:

1. Visualisation. We introduced multi-colour segmentations: Users can now use different palettes to annotate MRI data. We introduced annotation layers: MRIs can now be annotated for multiple structures in parallel. We introduced text annotations: these can provide additional information about an MRI, such as data quality or imaging parameters. As volume annotations, text annotations use WebSockets and are updated in real-time. We added 9 editing tools: full-screen mode, 3D rendering, image adjustment, ruler, distant demonstration, share link, eyedropper, upload/download and a precision cursor to simplify segmentation with the finger. The layout of BrainBox is now responsive, and works on desktop computers, tablets and smartphones.

2. Collaboration. We introduced Projects, which allow users to build collections of MRIs, configure lists of collaborators and annotations. Using BrainBox projects we have indexed > 6000 MRIs, including data for patients with autism, schizophrenia, ADHD, as well as brains of many species. We introduced User pages, where users can keep track of their work, the projects they have started or those in which they collaborate. User pages are public, and become a way to discover new data and projects.

3. Access management. We use OAuth2 and a GitHub login to identify users. Project owners can now determine who can view, edit, add or remove annotations. Thanks to access management, projects can now vary from public to private, and user's contributions are traceable.

4. Server code. We rebuilt the server code which now uses Node and Express (instead of PHP), and the database which now uses Mongo (instead of MySQL). This allowed us to developed an API which now makes it easy for other applications to communicate with BrainBox.

5. New developments. We started the development of MetaSearch, a graphic filtering tool that allows to find relevant data based on all available MRI meta-data and connecting to BrainBox's API. We started the graphic design of a new interface to make our tools easier to use for a wider range of users.

6. Pre-populated data. We have bootstrapped BrainBox and MetaSearch with existing open data and created a pathway for other users to upload data in bulk. For just MRI data this can be done through BrainBox, for other metadata through a common curation route on the MetaSearch Github page.

ii. Team Contributions

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

Katja Heuer and Roberto Toro were the main developers of BrainBox's code and design. Katja Heuer designed the user interface and coded a large part of the client-side interface. Roberto Toro developed most of the server-side code as well as a large part of the client-side interface. Together they wrote the documentation of BrainBox and its API, and presented it at different scientific meetings and hackathons, engaging new participants and helping them set up their own segmentation and annotation projects.

Satrajit Ghosh provided key ideas for the design of BrainBox and is the main developer of the code and design of MetaSearch. Together with Nolan Nichols, a collaborator on the project, he developed the code to index open neuroimaging data from Amazon and other data sources, and to query and organise its metadata. These data were used to bootstrap MetaSearch and BrainBox. Amy Sterling coordinated the team of designers developing the new user experience and user interface for the Open Neuroimaging Laboratory tools. She provided key insight for the current user interface of BrainBox.

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 significant achievement of our work thus far is to have developed our tool to the point where it is able to handle very large data sets. Data can be added to BrainBox through an interactive web interface, or programmatically through BrainBox's API. This latter possibility has allowed us to easily index more than 6000 MRIs from different large scale data sharing initiatives. The API has allowed us not only to include the links to the raw data but also a series of automatic segmentations that users can now quality control and correct.

Our tool has reached the point where it becomes interesting for neuroimagers, and several groups have manifested their interest in adopting it; among many others, researchers from the ENIGMA consortium interested in stroke recovery, researchers from the ABIDE consortium interested in autism, and NITRC, a large repository of neuroimaging tools and data which could start indexing their data using BrainBox.

Learning Points

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

As we have combed through the publicly available data that we know of, we realize how disorganized and inaccessible the data are, and how much additional time and effort is required to harmonize. There is also a serious impediment to scientific progress when the current sources for protected data (e.g., the NIMH archive, ADNI) generate significant redundancies by not allowing a pathway for people to collaboratively work on it as a community. We have once again experienced the benefits of open source development. Thanks to the immediate feedback from our users we were able to adapt our design and extend our tools in directions that make it more useful. We have also reached a point where we see the limits of our prototyping approach: As our code becomes more complex, we have learned the importance of documentation and coding style. A major necessity for the future is the implementation of a testing methodology that would allow us to add new functionalities more robustly. We have also learned the contrast that exists between the impact that our current work is having on our community, and the little that time-consuming code development is rewarded: for academic careers, coding counts for peanuts.

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

Our project will allow researchers to work directly on online MRI data, creating of a collaborative layer of human annotation. This will make possible a collective enterprise to incrementally assess and improve the quality of open data, enhancing rigor and reproducibility. Our will allow anyone with a Web browser to work with a growing database of containing already thousands of MRIs; and anyone with a brain scan to view it, work on it and share it. Our tool can have an important impact for outreach and education, and engage a large community students and citizen scientists in neuroimaging research.

ii. Innovation

The degree of innovation associated with the tool or service

The Open Neuroimaging Laboratory is the first platform that allows direct access to openly available data and provides tools to interact with such data. The key innovations are: 1) creation of a scalable infrastructure for individuals, whether researchers or general public, to add brain imaging data to the Web, 2) to allow curating the data through distributed collaboration, and 3) to increase the utility of data by sharing the curation results openly. In general, this prototype extends the power of the Web to brain science by creating a read-write infrastructure that scales across individuals, laboratories, and institutions.

iii. Utility

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

Most brain imaging laboratories perform a significant amount of manual assessment and editing on brain images without any collaborative tool. This is even more significant in clinical settings such as brain tumors (meningiomas and glioblastomas) and infant brains, where such manual checks and edits are essential. We have already initiated several collaborations on this front with researchers and clinicians. Some of their de-identified data are already available in BrainBox. Thus BrainBox can provide clinical utility, research utility, and educational utility.

iv. Feasibility & Technical Merit

The feasibility and technical merit of the prototype

The prototype is already operational and demonstrates feasibility of the project as well as the adoption of various modern Web technologies. BrainBox and MetaSearch handle authentication, authorization, and data harmonization. By using GitHub, we provide a common platform for other developers to contribute, and at the same time track the history of the project. The prototype has started linking with common data elements and ontologies to minimize the amount of data wrangling that will be necessary in the future. By providing simple templates and useful applications, it also crowdsources the submission of data in curated form to leverage these applications.

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]

Thus far, we have done all work openly on GitHub so that any developer can jump in. BrainBox has

been presented at different scientific meetings and hackathons, engaging new participants and researchers. By relying on distributed open data and open application, we minimize cost of ownership and development.

Technologically, we have initiated conversations with the NIMH Data Archive to understand how the data in the archive and other NIH data repositories that require institutional authorization can be enriched and collaborated on with tools such as BrainBox. We leverage public data stored in well maintained repositories (e.g., Amazon S3, Dataverse, Open Science Framework, Zenodo, Figshare). We have also opened conversation with the Inter Planetary File System (IPFS) developers to understand how a distributed file system can help sustain and preserve public data. At present, the Institut Pasteur is supporting the maintenance of the backend server. Moving forward we intend to minimize backend services and increase peer-to-peer client side applications, which are more sustainable.

In the short term, we intend to collaborate with NIH projects such as ReproNim, BD2K Enigma, and NIF, foundation projects such as the CRN at Stanford, as well as EU projects like the Human Brain Project and the UK BioBank. We also would like to connect with other protected databases such as COINS, SchizConnect, and GENUS. Connecting with these other databases will increase the availability of data.

In the longer term, we expect that distributed databases and services will be developed and maintained by interested projects, institutions, and individuals. The distributed nature of the project increases future sustainability. BrainBox has already inspired new applications. MindControl is an application being built at UCSF to support multiple sclerosis investigators. We also envision a world of personal data stores coupled with search engines and applications that changes the dynamics of data exchange and reuse. For now we want to focus on increasing the benefits of open data.

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]

Beyond the visible elements of the project, resources have also to be allocated to infrastructure development and maintenance. We have already invested heavily in these areas, but there is still much to be done. We are extremely appreciative of the work of many researchers who have put brain imaging data on the Web. We hope this platform will encourage others to do so as well. Finally, it is imperative that as a global scientific community we can openly discuss ethics, policies, and protections to individuals around data publishing and its use to maximize impact on human health.