Lightning Talk: "Kirkos," A Scientific Collaboration Tool which Incentivizes Sustainable Practices

Jonah M. Miller[1,2]

[1] *Department of Physics, University of Guelph, Guelph, ON, Canada*
[2] *Perimeter Institute for Theoretical Physics, Waterloo, ON, Canada*[*]

(Dated: 2016-05-10)

Two common challenges in scientific research, both theoretical and experimental, are remote collaboration and ensuring the reproducibility of scientific data. The former is a technical difficulty—scientists seek tools that allow them to easily collaborate over long distances. However, the latter is a cultural difficulty. Scientist are not incentivised to share their data, methods, or results with the broader scientific community. Here we propose a new project, Kirkos, that seeks to resolve both difficulties by tying good reproducibility behavior to ease of collaboration in an user-friendly interface that mimics some of the features of social media.

I. BACKGROUND

A. The problem of reproducibility

An increasingly public difficulty in scientific research is the inability of researchers to reproduce old results, either experimental or numerical. Scientific procedures are often very complicated, and researchers rarely make their procedures and their data completely public; they prefer to publish only the results and interpretation. The first reason for this tendency is logistical. Carefully documenting procedures and the interpretation of data can be a lot of work, and researchers' time is better spent working on new projects. The second reason is competitive. In the time it takes for a given line of inquiry to come to fruition, another researcher might poach one's idea.

However, this behavior is problematic. Scientific progress is built on trust and the free sharing of information. One cannot build on prior results if one is ignorant or distrustful of them. Moreover, science is fundamentally a collaborative exercise. Researchers benefit from sharing and discussing their ideas while those ideas are still in development.

Therefore, scientists need a tool that incentivises good record-keeping as researchers develop their ideas, protects them theft, and allows them to share with collaborators and publicize their work as they choose.

B. The problem of collaboration

Scientific research often involves remote collaborations. To handle the difficulties of collaborating across vast distances and different time zones, researchers need an easy way to asynchronously communicate with each other. Smaller teams might need only a file-sharing service like Dropbox, but larger teams must cobble together a collection of standard tools: wikis, email lists, code repositories, file storage, etc. One or more of these tools will probably not be provided by a given scientist's home institution. Therefore the entire tool suite is usually maintained and supported by the researchers themselves. The result is a drain on their time and a poorly maintained system.

Scientists need a single, externally maintained service that combines all of these standard tools under an easy-to-use interface.

II. KIRKOS: SCIENTIFIC DATA AS SOCIAL MEDIA

In 2011, Google launched a social media service called Google+ [2]. A core feature of this service is the ability to share content with a "circle" of people, allowing for fine-grained control over privacy. One could share information publicly, only with one's friends, or with a customizable subset of the people one had connected with online.

We propose to solve both the problem of reproducibility and the problem of collaboration by building a cloud-based toolbox based on this idea. We envision a service that allows a user to start a new "project," which automatically provides a set of tools (file servers, code repositories, mailing lists, websites, etc.), which are automatically configured in an intuitive way.

Once a user has created a project, she needs very precise control over who can view and who can contribute to the project. We therefore propose a system similar to Google+ "circles," in which a user can assign her contacts to one or more groups that can be given different access privileges to different projects. In this way, she can share specific pieces of her projects and invite specific people to contribute while otherwise maintaining privacy.

By providing a collection of collaborative tools, we will encourage researchers to use the service from the beginning of a scientific project. By designing with the intention that a project's information will be shared with an ever-widening audience until it becomes completely public, we allow scientists to both protect their intellectual property and make their research reproducible.

[*] jmiller@perimeterinstitute.ca

III. PREVIOUS EFFORTS

This is far from first proposal that has attempted to solve reproducibility problems with techniques borrowed social media. The open source community has made heavy use of websites such as Github [5] and Bitbucket [4] partly because of the social media aspect and public API, which allow developers to seamlessly include discussions, documentation, issue tracking, and with the aid of additional services, even integration testing [1].

A very promising data tool is yt-hub [3], which provides a repository for both data and the data analysis code used on it. It provides "executable arXiv" functionality where users can evaluate the data analysis code online as they explore published data. Finally, it provides a means to upload data in a users' standard workflow via integration with the yt-project analysis and visualization tool [6].

IV. PUBLIC AVAILABILITY

Since this tool is designed to enable collaboration and encourage the (eventual) free flow of information, it is absolutely essential that it be provided to the entire scientific community. This means that the institution that owns and maintains Kirkos must be willing to step back from the project and allow it to exist as an independent entity.

Another important part of making a tool publicly available is making that tool easy to use. Social media websites gain popularity partly because of their intuitive interfaces. Similarly, Kirkos must offer a variety of options for users of all backgrounds. For example, people who are familiar with version control can use a source code repository for file storage, but people accustomed to more "traditional" file storage would find this solution unusable.

V. PROJECTS GROW ORGANICALLY

Often a research project starts small. It might be the pet idea of one or two people. Sometimes, they might publish a paper based on their original idea and move on to other things. But often the original collaborators will talk with other researchers and share ideas, and over time, a research project will evolve and grow and gain new contributors. In these cases a single research project can span many papers, many computer programs, and many data sets. It is important that the people who have access to a project can change organically over time and that all of the pieces of information associated with a project are preserved in an intuitive way.

Even for relatively small projects, backing up and storing research notes and data over long periods of time is a serious challenge. Hard drives fail and cloud services usually only offer limited storage space. Kirkos must offer safe, large-volume data storage for time scales of several years.

Finally, because scientific projects last for decades, any service that caters to scientists must plan on its own obsolescence. Eventually scientists may migrate to another service, either because it is better or because circumstances force them. Therefore Kirkos must offer an easy way to export, access, and download all data stored on the service, including meta-data.

VI. SUMMARY

To conclude, we propose a collaboration platform for scientists that incentivizes good reproducibility and archival practices and that gently encourages scientists to develop their projects in such a way that they can eventually become public.

VII. ACKNOWLEDGEMENTS

The author would like to thank the many people from whom he has learned about sustainable software and scientific culture. These people include Erik Schnetter, Thomas Schibli, Matthew Turk, Gabrielle Allen, Ian Hawke, Lorena Barba, Ian Hinder, Roland Haas, Michael Schmidt, and many others.

VIII. LICENSE

[1] T. C. GmbH. Travis cs, 2011. https://travis-ci.org/.
[2] G. INC. Google plus, 2011. https://plus.google.com/.
[3] K. Kowalik, M. J. Turk, and N. J. Goldbaum. The yt hub: Data, compute, and tool sharing in the publication process. 2016. In prep. Website at https://girder.hub.yt/.
[4] A. LLC. Bitbucket, 2008. https://bitbucket.org.
[5] G. LLC. Github, 2008. https://github.com/.
[6] M. J. Turk, B. D. Smith, J. S. Oishi, S. Skory, S. W. Skillman, T. Abel, and M. L. Norman. yt: A Multi-code Analysis Toolkit for Astrophysical Simulation Data. *The Astrophysical Journal Supplement Series*, 192:9–+, Jan. 2011, 1011.3514.