Policies to support Open Source and Open science

Sayeed Choudhury

DOI: 10.4000/books.oep.16192
Publisher: OpenEdition Press, Académie des sciences, EDP Sciences
Place of publication: Paris ; Marseille
Year of publication: 2022
Published on OpenEdition Books: October 13, 2022
Series: Laboratoire d'idées
Electronic EAN: 979-10-365-4562-7

Electronic reference

The text only may be used under licence CC BY 4.0. All other elements (illustrations, imported files) are "All rights reserved", unless otherwise stated.
Software is a pillar of open science. The universities produce basically three of these pillars: articles, data and software. Those are connected and in many ways those three pillars of open science and the connections between them relate to what Roberto described as the scholarly ecosystem. In large part, those are the projects that universities do through their research. They are typically funded through grants and other kinds of external funding, and are very much the things that happen within the walls of the university. At some point, if you wish to move the impact of research and education outside the walls of the university then you have to translate them. It is useful to publish papers and to search data. It is also extremely useful to share the software.

Over the last two years we have all been experiencing COVID-19 and there has been a lot of discussion about how open science led to the vaccines that have made it possible for us to deal with it. I would submit that the most important piece of that was actually not the sharing of articles or even the data being shared, but rather the software, the code that people were using to analyse the data associated with the vaccines that we eventually developed.

The open source programmes office, or OSPO, is a university organisation that can support this translation function. It is the unit that we have created within Johns Hopkins to help our researchers translate their research and their education activities outside the walls of the university. One of the ways that we have advanced the work of the OSPO is what you heard from my previous panellists, this statement or affirmation that software is a primary research object. If you ask people within a university, either researchers or administrators or even students, ‘Do you care about open source software’? you will get different responses, but if you ask them if they care about research, every one of them will say, ‘Yes, of course I do’. Therefore, being able to make this argument that open source software supports research, is a tool, but is also a research object as we have heard, has been incredibly important in terms of advancing the work of this OSPO within Johns Hopkins University.
We have also made the case, as you heard earlier today, that software is a better way to understand what data are being produced rather than the articles that cite those data. The citation or the connection between an article and a dataset is one way, is often a direct mechanical kind of link, but the connection between data and software is much richer. If you look at the software, you can actually understand a great deal more about what data are being used and how those data are being analysed. I will talk about a specific example that describes this.

We created at Johns Hopkins this open source programmes office, or OSPO, and this is something that is used in the private sector by companies, technology companies, but also companies like American Airlines and Walmart, which are not technology companies, to manage their open source software. While there are some common elements between the OSPO in a private sector company and in a university, the OSPO at the university is also doing unique things that are important in the context of research and education and translation. We are helping with a lot of the best practices that you heard in earlier talks, which have to do with things like how to engineer your software, what licences to choose, what risks you may be introducing with those choices and how you will in fact work with publishers and even directly with society at large. The OSPO is therefore building on what has happened in the private sector, but it is expanding and extending this in the context of the university.

I am also pleased to note that while Johns Hopkins created the first OSPO in a university in the United States, there are now at least three others at the University of California, Santa Cruz, at the Rochester Institute of Technology and at the University of Vermont. It is my understanding that there will be additional universities creating these OSPOs, so we can see ourselves coming together as a network of these organisations.

Johns Hopkins is involved in a project called the Observational Health Data Sciences and Informatics, or OHDSI project. Paul Nagy is a researcher in our School of Medicine who is involved in this project. He approached me after hearing about our OSPO saying that they needed help. What he explained was that when it comes to health or medical data it is very difficult to share outside of your research team. You typically have to create a very complex legal agreement. It has to be approved by every party that signs that agreement, and this can be a time-consuming process. What the OHDSI group did instead was create open source software and share that software with each other, and by sharing the software they actually learned more about what types of data each of them were using and what types of analysis they were using on those data. This goes back to the earlier point I made where software has that richer connection with data, and that is something we have heard from previous panellists as well.
Ultimately what happened is OHDSI then ended up with a great deal of open source software and they were not managing them very well, so they came to the OSPO and asked for help, and through the OSPO we provided the kind of engineering support, management support and infrastructure support that are necessary to organise the software. I agree completely that researchers should learn about the infrastructure. They should learn about the issues. Education is very important. However, I do not believe researchers should support the infrastructure or sustain it. We heard earlier from Roberto that grants should not be used for operational expenses. I could not agree more. The OSPO is therefore the unit that helps with the education, helps with raising awareness, but that ultimately supports a team like OHDSI. What I heard clearly from our researchers was that ‘I do not want to be an expert in open source software, I just want to learn about it and then I want the experts to take over the maintenance and support’. In some sense therefore the OSPO is a centre of excellence, a centre of knowledge or a centre of information.

In the education realm, in this past semester of fall 2021, we conducted a course at Johns Hopkins called the Semesters of Code. This was taught by Stephen Walli, a Microsoft executive with over 40 years of experience with software engineering. He mentioned to me that while graduates from universities like Johns Hopkins have very good skills, companies have discovered that it takes anywhere from three to six months to retrain them to understand how to work in an open source project, how to work as part of a team. The idea therefore is that if we can start the educational process earlier in their undergraduate career then they will be better prepared to work at places like Microsoft. However, we also recognise the tremendous impact and value to them learning these skills regardless of whether they work for a company, become a researcher or work in a non-profit.

The open source software was the framework through which we talked about software engineering practices. In addition to the lectures, where Stephen covered topics that were not necessarily related to engineering, like the legal issues, the economic issues, the history of open source, each student was assigned to a project, a real world open source project, where individual contributors and developers from those projects were mentors. The students worked on the projects, and they were given direct access to a mentor who guided them through the project, who helped them learn about the project and who then helped them work their way through their particular assignments. We received the evaluation and feedback a few weeks ago and I am pleased to note that the students very much enjoyed this. We actually had one student who had to drop the course because he is graduating, and he could not fit the course into his work, but he continued to work on the project. We had another student who did not finish the work he had decided to do, but then he said, ‘Can I do
this independently’? and continued working with the mentor. The students therefore really enjoyed this opportunity to focus on real world projects.

One of those real world projects is called Lutece\(^3\), an open source digital services platform in the city of Paris. It has been used for many years to provide services to the citizens of Paris, and students at Johns Hopkins have worked on the code with Lutece, made contributions to that. Our OSPO took those contributions, engineered them and updated them, and contributed back to Lutece. We are using this at a community centre in West Baltimore called the St Francis Neighbourhood Center. You therefore have a software platform developed in Paris, worked on by students and software engineers at Johns Hopkins and being used in a community centre in West Baltimore. I mention this to highlight that if we had to sign a legal agreement to have Paris, Johns Hopkins and this community centre work together it would have taken a very long time. Instead, the open source licence of Lutece is the legal agreement under which we did our work.

One other point I therefore wish to make about open source, particularly as it relates to impact outside of universities, is it is a much easier way to partner. You do not have to have a grant. You do not have to have a legal agreement. You have the open source licence itself, which is the way in which you can conduct work with other partners.

I had mentioned that there are multiple universities within the United States now that are forming these kinds of OSPOs. There is a network called OSPO++ that has been created as a community and as a network of these OSPOs\(^4\). It is not only within the United States. We have participants from Europe and other parts of the world who are coming together to talk about how OSPOs can continue to support the kinds of activities within universities and to increase the amount of impact beyond the university itself.

Recently, OSPO++ has worked on a best practices guide through the United States National Academies of Science. The NAS, as it is called, recently had a committee look at open science and as part of its final report produced a primer on open source software. The OSPO++ brought together representatives from universities, companies and governmental and nongovernmental organisations, including the Open Forum Europe, which is a think tank that has worked closely with the European Commission, to update and create this primer on open source software\(^5\). An important policy recommendation that has been made by this NAS report matches a recommendation made within the European Union to create 20 of these so-called OSPOs, both within universities and government, and have those OSPOs work together in a coordinated way. The hope is that the United States will create a network of OSPOs,
Europe and other parts of the world will create OSPOs, and we can start to coordinate that activity, start to work towards more common policies, best practices and help our researchers manage their open source software.

**To go a step further**

“There is a very important growing movement around ethics in data. The same is true of software. We sometimes talk about biases that are embedded within data, within the tools or algorithms used to process those data through software. There is absolutely an ethical component to working on software, even if it is open source, but one of the good things about open source software is it is transparent, so we have a better chance of seeing what is actually being done in terms of analysing the data.”

**References**