Workshop @ deRSE19: Libraries for Research Software & Engineers

Ronny Gey, Katrin Leinweber, Alexander Struck,
Robert Strötgen, Christian Pietsch

The workshop was hosted by five people from library and infrastructure environments. As information managers we discussed what libraries are already doing and should do in the future to further explore software engineering and engineers. The topic attracted participants with diverse backgrounds in research software engineering and library and information science. Our goals were to generate/strengthen awareness of the topic of Research Software & Engineers [RS(E)] in libraries, to jointly identify fields of action for libraries, to collect and collaboratively develop ideas and materials for use in libraries, and to interconnect the RSE community with libraries.

Introduction

Katrin Leinweber from TIB Hannover opened and moderated the workshop. After a short round of introductions of all the participants, we started to present the three subtopics of this workshop in short impulse lectures, from which the participants then choose one to persue. For the treatment of the topics we decided to use the Speedboat procedure[1] known from agile project management. We worked with a diagram of a sailing boat and an island. The anchor of the sailboat represents the challenges of the respective topic. The wind in the sails gives the company strength and stands for aspects that help us to achieve our goal. The goal in turn is embodied by the island, which represents the desired state for the respective topic.

Software Discovery and its tools

Discussions revealed that developers utilize general purpose search engines to find code snippets, for example in order to remind themselves how a certain algorithm was implemented. Software for re-use purpose is also searched for in general purpose search engines and within the researcher's social network.

BASE[2] as a search engine for academic audiences does show software from over 7000 sources it harvests. User Interface refinements to support this use case were discussed, such as aggregating

---

[1] "Speedboat procedure": https://klaxoon.com/blog/speed-boat-an-agile-method-to-discover.
[2] BASE: https://www.base-search.net/.

multiple versions of the same software into a single entry, with the latest version displayed primarily. It was also suggested to implement full text and source code search (in addition to metadata search) in order to increase its usefulness. Zenodo.org[3] becomes more popular as a software repository due to its integration with GitHub[4] which allows publication with a DOI and thus proper citation[5] metadata. Software Heritage[6] was shortly discussed as an archival effort to prevent loss of research software as it may happen on code collaboration platforms. We finished the discussion with ideas on how to make software and the relevant repositories more visible. A registry similar to what re3data[7] does for research data could be a promising endeavour.

Open Educational Resources

In the second track of the workshop, we set out to collect existing Open Educational Resources on RSE and to show why libraries can be good partners of the RSE community in terms of education and training. We agreed in advance that mastery of the tools is particularly important for research and that software has an outstanding importance as a tool today. First, we collected existing offers for training (Carpentries[8] – Software[9], Data[10], Library[11], ProgrammingHistorian[12], Exercism[13]), for documentation (Read the Docs[14], MkDocs[15]), and collections of interesting aspects in general (awesome lists[16]). We quickly agreed that there are already many offers and that we as libraries should rather concentrate on collecting, curating, improving and above all disseminating the existing offers. We identified potential problems for this in institutional support – although the topic of research software is currently coming to the fore, the necessary structures at the institutions are still lacking. The offers collected so far are also based on voluntary work, which is often carried out in leisure time in addition to regular work. Further, the sheer quantity of available offers is challenging.

Just as good programmers try to solve a problem by writing as little code as necessary[17], reusing well-established, shared code libraries, knowledge workers should take on the challenge of contributing to existing resources, rather than creating their own. Although the latter is a common criterion in evaluations, promotions et cetera, it exacerbates the problem of curating or even just reviewing the sheer quantity of available material.

---

[3]Zenodo – Type "Software": https://zenodo.org/search?page=1&size=20&q=&type=software.
[4]GitHub "Making Your Code Citable" https://guides.github.com/activities/citable-code/.
[5]Research Software Citation: https://cite.research-software.org/.
[6]Software Heritage: https://www.softwareheritage.org/.
[7]re3data: https://re3data.org/.
[8]The Carpentries: https://carpentries.org/.
[9]Software Carpentry: https://software-carpentry.org/.
[10]Data Carpentry: https://datacarpentry.org/.
[11]Library Carpentry: https://librarycarpentry.org/.
[12]The Programming Historian: https://programminghistorian.org/.
[13]Exercism: https://exercism.io/.
[14]Read the Docs – Create, host, and browse documentation: https://readthedocs.org/.
[15]MkDocs – Project documentation with Markdown: https://www.mkdocs.org/.
[16]For example list on GitHub: "awesome – Awesome lists about all kinds of interesting topics" https://github.com/sindresorhus/awesome.
[17]David Strauss: "All Code is Debt", February 12, 2014, https://pantheon.io/blog/all-code-debt.

RSE and the Management of Research Data

The third group discussed the close connection of research software and research data. In many cases publication and archival of research data without the used research software is useless. This causes several challenges for example for software of measure devices and simulation. Libraries have to offer new services in close cooperation with researchers and research software developers. Libraries have useful knowledge on aspects like metadata, protocols, vocabularies/ontologies, persistent identifiers. This knowledge may lead to new services about reproducibility, discovery, free or restricted access to research data and software.

Research is changing and becoming more complex. Diversity of researchers' needs and use cases require flexible cooperation between librarians, researchers and software developers. Small and concrete pilot projects should help to develop new ways of cooperation.

Future Work

The organizers consider pursuing the topic in the future. Making software repositories more accessible may be one activity. In our experience, libraries will have the most impact by connecting scientists and users to existing tools, curation forums (GitHub topics[18]) and initiatives like Carpentries[19] or Open Source Guide[20]. Stay tuned.

deRSE Conference Review

After three successful UK conferences for Research Software Engineering the first RSE conference in Germany[21] took place at the Albert Einstein Science Park[22] in Potsdam. The organizers created a wonderful atmosphere and a sophisticated and balanced schedule. Many talks were recorded and have been published in a dedicated video repository[23]. There has been praise[24] and some criticism due to some sponsors involved. There are more national conferences of this kind scheduled[25] but rumor has it that an international conference is in the making.

---

[18]GitHub topics: https://github.com/topics/.
[19]The Carpentries: https://carpentries.org/.
[20]Open Source Guides: https://opensource.guide/.
[21]deRSE19 Conference: https://www.de-rse.org/en/conf2019/.
[22]Albert Einstein Science Park: https://en.wikipedia.org/wiki/Albert_Einstein_Science_Park.
[23]deRSE19 video recordings: https://av.tib.eu/series/644.
[24]Helmholtz Open Science Newsletter vom 24.07.2019
https://os.helmholtz.de/bewusstsein-schaerfen/newsletter/archiv/newsletter-75-vom-24072019/#c19002
[25]For example in the Netherlands, see Call for contributions – NL-RSE19:
https://nl-rse.org/2019/07/09/NL-RSE-2019.html.

Ronny Gey studied business informatics at the TU-Chemnitz and then worked as a research assistant at the University of Leipzig and the FSU Jena. Since 2018 he has been working as a trainee at the University Library Leipzig and is studying library and information science at the IBI at the HU Berlin. https://orcid.org/0000-0003-1028-1670

Katrin Leinweber studied life sciences, ecology and geology, then shifted towards process automation, data analysis and software development. She supports and trains researchers and librarians in these topics since 2017. https://orcid.org/0000-0001-5135-5758

Alexander Struck has an academic background in LIS & CS. He worked for the content industry and does research on citation networks, research evaluation and research software discovery. Alexander is CIO of the Cluster of Excellence – Matters of Activity. https://orcid.org/0000-0002-1173-9228

Robert Strötgen is a historian and information scientist and has developed scientific software at various institutions such as GESIS and the Georg Eckert Institute. Since 2016 he has been head of the IT and research support services department at the University Library of the TU Braunschweig and is now deputy director. https://orcid.org/0000-0003-3320-5187

Christian Pietsch is a computational linguist in Bielefeld University Library's library technology and knowledge management department. Recent DFG projects he contributed to include CONQUAIRE (Continuous quality control for research data to ensure reproducibility) and ORCID DE. Occasionally, he is involved in running BASE (Bielefeld Academic Search Engine). https://orcid.org/0000-0001-8778-1273