

Lesson development for Open Source Software best practices adoption

Mateusz Kuzak

Dutch Techcentre for Life Sciences
ELIXIR Netherlands
the Netherlands
mateusz.kuzak@dtls.nl

Jen Harrow

ELIXIR Hub Wellcome Genome Campus
Hinxton, UK

Rafael C. Jimenez

ELIXIR Hub Wellcome Genome Campus
Hinxton, UK

Paula Andrea Martinez
ELIXIR Belgium

Fotis E. Psomopoulos
Institute of Applied Biosciences, Centre for Research and Technology Hellas
Thessaloniki, Greece

Radka Svobodová Vařeková

Central European Institute of Technology (CEITEC)
National Centre for Biomolecular Research, Masaryk University
Brno, Czech Republic

Allegra Via

ELIXIR Italy, National Research Council of Italy (CNR)
Institute of Molecular Biology and Pathology (IBPM)
Italy
allegra.via@uniroma1.it

Index Terms—training, Open Source, software guidelines, best practices, recommendations, Open Science, Reproducible Research, sustainability, FAIR

Abstract—The “ELIXIR Training Platform” is partnering with The Carpentries (Software and Data Carpentry) to train life science researchers in computing and data management skills. The “ELIXIR Software development best practices” group, which is part of the ELIXIR Tools Platform, has proposed “Four simple recommendations to encourage best practices in research software” aiming to help researchers and developers to adopt Open Source Software (OSS) practices and thus improve the quality and sustainability of research software. In order to encourage researchers and developers to adopt the four recommendations (4OSS) and build FAIR software, we are developing specific and practical training materials, taking advantage of the Carpentries approach and experience in training material development and maintenance.

I. INTRODUCTION

ELIXIR [1] is an intergovernmental organization that brings together life science resources across Europe. These resources include databases, software tools, training materials, cloud storage, and supercomputers. One of the goals of ELIXIR is to coordinate these resources so that they form a single infrastructure. This infrastructure makes it easier for scientists to find and share data, exchange expertise, and agree on best practices. ELIXIR’s activities are divided into five areas called ‘Platforms’. These are Data, Tools, Interoperability, Compute and Training. The ELIXIR Training Platform coordinates training activities, trains life science researchers, and helps scientists and developers to find the training they need. The ELIXIR Tools Platform works to improve the discovery, quality and sustainability of software resources.

The Software development best practices group is part of the ELIXIR Tools platform. It coordinates activities to rise the quality and sustainability of software developed for research. The Software development best practices group in partnership with the ELIXIR Training platform, The Carpentries [2], [3], and other communities is creating a collection of training materials to help researchers and developers implement the four Open Source Software (4OSS) recommendations [4]. 4OSS are set of 4 simple recommendations aiming to help researchers and software developers to adopt Open Source Software (OSS) practices.

II. FOUR SIMPLE RECOMMENDATIONS

The 4OSS simple recommendations are as follows:

- 1) Develop publicly accessible open source code from day one. Start a project as open source from the very first day, in a publicly accessible, version controlled repository (e.g. github.com, gitlab.com and bitbucket.org). The longer a project is run in a closed manner, the harder it is to open source it later.
- 2) Make software easy to discover by providing software metadata via a popular community registry. Facilitate the discoverability of the open source software projects by registering metadata related to the software in a popular community registry (e.g. bio.tools [5]), making your source code more discoverable. Metadata might include information such as source code location, contributors, license, references and how to cite the software.
- 3) Adopt a license and comply with the licence of third-party dependencies. Provide instructions and guidelines for other projects and software to use, modify and redistribute the software and the source code. Adopt

a suitable Open Source license, include it in a publicly accessible source code repository, and ensure the software complies with the licenses of all third party dependencies.

- 4) Have a clear and transparent contribution, governance and communication processes. Open sourcing your software does not mean the software has to be developed in a publicly collaborative manner. Although it is desirable, the OSS recommendations do not mandate a strategy for collaborating with the community. However projects should be clear and transparent about how to contribute to them as well as, their governance model, and their communication channels.

III. LESSON DEVELOPMENT

In order to encourage researchers and developers to adopt the 4OSS recommendations and build FAIR (Findable, Accessible, Interoperable and Reusable) software, we decided to develop specific training materials, taking advantage of the Carpentries approach and experience in training material development and maintenance [6], [7]. Here, we present a collaboration between ELIXIR and The Carpentries aimed at creating a collection of training materials to teach researchers and developers how to implement the recommendations in their research software. This involves an open and transparent content development consisting of community brainstorming, content collection and reduction and a great effort in making the material interactive with challenges and discussions.

The project was kick started with a workshop at CarpentryCon 2018 in Dublin (Ireland) [8] aimed at scaffolding the lesson. This was then followed up in August, by a 2-day lesson hackathon taking place in Utrecht (The Netherlands), where participants focused on content creation and produced a first draft of the training materials. Twenty-one participants from across the world contributed with their expertise in pedagogy, community building, Open Source software, licensing and ontologies. Results have been published on GitHub [9] under Creative Commons License (CC BY 4.0). The lesson is build around the following questions:

- 1) Make it public : What are the benefits of making my software project public from the beginning? How do I make my project publicly accessible? What resources are available to help me document my software? What are the best practices in open software development?
- 2) Use registry : Why are metadata important in research software? What are good metadata? Which are the most commonly used platforms for registering research software data.
- 3) Use licence : What is a copyright and what a licence does? Why is important that a product/code has a licence? What is the importance of third-party dependencies on your product/code? How do you choose a license for your code?
- 4) Contribution, governance and communication: How does someone start contributing to my project? What do I

need to consider about project design and governance?
How do people communicate within the project?

The content has been further reviewed and improved in an online process involving wider community of contributors and will eventually be finalised during the sprint in October at the NETTAB 2018 [10] event (Genoa, Italy). The expected release of the training materials is October 2018.

REFERENCES

- [1] "ELIXIR | A distributed infrastructure for life-science information" Internet: www.elixir-europe.org, [Sep. 16, 2018]
- [2] "carpentries.org" Internet: carpentries.org, Sep. 15, 2018 [Sep. 16, 2018]
- [3] "ELIXIR teams up with The Carpentries to boost its training programme | ELIXIR", Internet: <https://www.elixir-europe.org/news/elixir-carpentries-agreement>, Aug. 17, 2018 [Sep. 16, 2018]
- [4] Jiménez RC, Kuzak M, Alhamdoosh M et al. (2017) "Four simple recommendations to encourage best practices in research software [version 1; referees: 3 approved]" *F1000Research* [Online]. 6:876. Available: <http://dx.doi.org/10.12688/f1000research.11407.1> [Sep. 16, 2018]
- [5] "bio.tools Bioinformatics Tools and Services Discovery Portal" Internet: bio.tools, [Sep. 16, 2018]
- [6] Brown, N. C. C., & Wilson, G. (2018). "Ten quick tips for teaching programming." *PLoS Computational Biology*, [Online]. 14(4), e1006023. Available: <http://doi.org/10.1371/journal.pcbi.1006023> [Sep. 16, 2018]
- [7] Devenyi GA, Emonet R, Harris RM, Hertweck KL, Irving D, Milligan I, et al. (2018) "Ten simple rules for collaborative lesson development." *PLoS Comput Biol*, [Online] 14(3): e1005963. Available: <https://doi.org/10.1371/journal.pcbi.1005963> [Sep. 16, 2018]
- [8] "CarpentryCon 2018" Internet: www.carpentrycon.org, Sep. 11, 2018 [Sep. 16, 2018]
- [9] "4OSS lesson" Internet: softdev4research.github.io/4OSS-lesson, Sep. 7, 2018 [Sep. 16, 2018]
- [10] "Building a FAIR Bioinformatics environment Genoa, October 22-24, 2018" Internet: www.igst.it/nettab/2018, [Sep. 16, 2018]