

Lightning Talk: A Proposal for the Measurement and Documentation of Research Software Sustainability in Interactive Metadata Repositories

Stephan Druskat¹

¹Department of German Studies and Linguistics
Humboldt-Universität zu Berlin, Germany

4th Workshop on Sustainable Software for Science: Practice and Experiences, 12 Sep 2016, Manchester, UK



Software sustainability – Challenges

Two identified technical barriers to software sustainability are

- the identification of good software, i.e., software built to be sustainable, and
- the discoverability of software [1].

Experience from the humanities: Computational infrastructure is rarely funded, software is not made sustainable due to lack of experience.

Metadata repositories can provide a solution (and be a target for SMPs), if they

- measure and document the *technical sustainability* [2], and
- document resulting metrics and more general features of software.

Existing solutions (e.g., SciencePAD, EGI Applications Database, DiRT Directory, GitHub, Open Hub, Zenodo) offer rather *limited* and/or only *implicit* information on technical sustainability.

Software sustainability – Definition

- *Technical sustainability of software products* is under-defined!
- Approximation: A software's capacity for longevity and evolution [2].

Proposal: Define in analogy to three-dimensional general model of *sustainability* [3].

The goals of technical software sustainability are

- ① *ensuring the existence of the software,*
- ② *preserving the potential for productive operation of the software,*
- ③ *creating and retaining possibilities for further development and adaptation of the software.*

Factors that contribute to these goals can be found in software **metadata** (considering a broad definition): *technologies, documentation, publication, licensing*, but also (*potentially*) *available human resources, usage statistics*, etc. These metadata can be recorded in a repository along with a description of features, applications, etc.

Criteria-based measurement of technical sustainability

Measurement over the metadata should be **based on criteria** categorized along the lines of the defined goals, and result in **reliable and reproducible metrics**.

Challenges:

- Finding relevant criteria: SSI's assessment criteria as starting point
- Finding quantifiable criteria (cf. below)
- Weighting criteria: Empirical elicitation, user-defined weights
- Constructing the metrics

Metrics for technical sustainability of software

The criteria above can be categorized along two different scales: **objectivity** and **quantifiability**. It is therefore necessary to construct different metrics taking into account different categorizations.

Proposal: 3 metrics

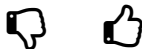
Type	obj	quant	repr	tamp	Example
"Hard" metric	✓	✓	✓	✓	License
"Semi-hard" metric	✓		✓	✓	Build system used
"Soft" metric					Intuitive UI

obj: objective criteria, *quant*: quantifiable, *repr*: reproducible, *tamp*: tamper-safe



User interaction

- Review and evaluate metadata: Can be used to detect bad data
- Vote/grade specific metadata points: Contribute to “soft” metric
- Document use of software (citations!): Contribute to harder metrics
- Gamification: Attract users



7



Data accumulation

The metadata to be held in the repository can be gathered through different means.

- Direct input by the originator of the software
 - Harvesting data from existing repositories via, e.g., GitHub and Open Hub APIs
 - Dedicated crawling of source code repositories, etc.
- The latter two methods can be used for verification and a preliminary quantification.

Conclusion and future research

An interactive metadata repository for research software that measures and documents technical sustainability can be a valuable tool for **software discovery**, the **identification of sustainable software**, and represents a natural **target point for SMP documentation**.

Future research: PhD thesis

Next steps:

- criteria elicitation and compilation
- crowd-sourced weighting of criteria

Thank you!

References

- [1] S. Hettrick,
“Research software sustainability: Report on a Knowledge Exchange workshop,”
The Software Sustainability Institute, Tech. Rep., 2016. [Online]. Available:
<http://repository.jisc.ac.uk/6332/>
- [2] C. Becker, R. Chitchyan, L. Duboc, S. Easterbrook, B. Penzenstadler, N. Seyff, and
C. C. Venters,
“Sustainability design and software: The Karlskrona Manifesto,”
in *2015 IEEE/ACM 37th IEEE International Conference on Software Engineering*,
vol. 2, May 2015, pp. 467–476.
- [3] J. Jörissen, J. Kopfmüller, V. Brandl, and M. Paetau,
Ein integratives Konzept nachhaltiger Entwicklung,
ser. Wissenschaftliche Berichte FZKA. Karlsruhe: Forschungszentrum Karlsruhe,
1999, no. 6393. [Online]. Available:
<http://www.itas.kit.edu/pub/v/1999/joua99a.pdf>