



# The open science initiative of the Empirical Software Engineering journal

Daniel Méndez Fernández<sup>1</sup> · Martin Monperrus<sup>2</sup> · Robert Feldt<sup>3,4</sup> · Thomas Zimmermann<sup>5</sup>

Published online: 2 May 2019

© Springer Science+Business Media, LLC, part of Springer Nature 2019

## 1 Introduction

Open science refers to a movement to make all research artefacts available to the public, thus, increasing transparency and reproducibility of the scientific process. Ultimately, the goal is faster scientific progress since problems can be weeded out more quickly and it is easier for new studies to build on the results of older ones.

Open science has become an umbrella term covering different aspects in research, such as open access (Piwowar et al. 2018), open data, and open-source software. We rely on the Open Research Glossary<sup>1</sup> for the following terms:

- Open Access: The effort of making peer reviewed scholarly manuscripts freely available via the Internet, permitting any user to read, download, copy, distribute, print, search, or link to the full text of these articles, [...], or use them for any lawful purpose, without financial, legal or technical barriers other than those inseparable from gaining access to the internet itself. The only constraint on reproduction and distribution [...] should be to give authors control over the integrity of their work and the right to be properly acknowledged and cited.
- Open Data: Data is freely available on the public internet permitting any user to download, copy, analyse, re-process, pass them to software or use them for any other purpose

---

<sup>1</sup><http://www.righttoresearch.org/resources/OpenResearchGlossary/>

✉ Robert Feldt  
robert.feldt@chalmers.se

Daniel Méndez Fernández  
Daniel.Mendez@tum.de

<sup>1</sup> Technical University of Munich, München, Germany  
<sup>2</sup> KTH Royal Institute of Technology, Stockholm, Sweden  
<sup>3</sup> Chalmers University of Technology, Göteborg, Sweden  
<sup>4</sup> Blekinge Institute of Technology, Karlskrona, Sweden  
<sup>5</sup> Microsoft Research, Redmond, WA 98052, USA

without financial, legal, or technical barriers other than those inseparable from gaining access to the internet itself.

- **Open Source Software:** Availability of source code for a piece of software, along with an open source license permitting reuse, adaptation, and further distribution.

Each of those principles is important to move forward scientific communities. In this editorial we will focus on open data and open source because they are fundamental for making empirical studies more transparent. An example of sharing is replication packages that contain the raw data and any material necessary for their analysis and interpretation (from study protocols to analysis scripts). Sharing data sets (Méndez Fernández and Passoth 2018): 1) increases the transparency, reproducibility, and replicability of research endeavours; 2) supports building up an overall body of knowledge in the community leading to more widely accepted and well-formed software engineering theories in the long run.

The software engineering community is making great progress in open science. We have data and artefact evaluation tracks at various conferences, the festival for open science<sup>2</sup>. The 2018 edition of the International Symposium on Empirical Software Engineering (ESEM), introduced a new open science policy<sup>3</sup> encouraging authors to make their research data publicly available. We also note that there are many open-source initiatives flourish in other research areas as well (Gent and Kotthoff 2014).

Given the positive feedback from the community and the overall rationality of open science, we are now taking a step further in our research community. We are introducing an open science initiative for the Empirical Software Engineering Journal. In this editorial, we motivate and introduce this new initiative.

## 2 The EMSE Open Science Process

### 2.1 Principles

We have designed the Open Science Initiative (OSI) of the Empirical Software Engineering Journal with the following principles in mind.

**Inclusiveness** First and foremost, we want the initiative to be inclusive by rewarding researchers committed to open science and by encouraging others to join in this research ethics and its efforts. Consequently, the process is based on a voluntary submission, and, as we believe, the required open science standards are reasonable. We want to encourage taking also small steps towards more openness and then gradually help the community transition to open practices on a wider scale.

**No Disruption** The overall aim of our initiative is to encourage authors of accepted manuscripts to publish replication packages while being as minimally disruptive as possible with the current peer review process. That is to say, the review process of submitted manuscripts shall not be disrupted at all; instead, authors will be asked upon acceptance of their manuscript to disclose their data sets which will then be reviewed.

**No one-size-fits-all Process** The overall OSI process needs to be flexible for authors and reviewers alike as well as depending on the specific study, its data and methodologies. The reason is that in our field, similar as in others, we publish works that are relatively different

<sup>2</sup><https://github.com/researchart/rose2-icse19/blob/master/cfp.md>

<sup>3</sup><http://bit.ly/esem18-openscience>

in their nature and based on a multitude of research methodologies including quantitative data and qualitative data of which some data emerges from sensitive industrial, proprietary or public sector contexts. It is thus not possible to define a one-size-fits-all policy, but we need to rely on a flexible process and continuously adapt our reviewing guidelines on a case-by-case basis.

**Community Support is Essential** The success of the Open Science Initiative depends on support from the research community. Therefore, we strive to implement the Open Science Initiative as a community effort and to evolve and refine it based on input from the community. We hope to contribute to promoting a data-sharing culture in the community, where authors publicly archive their data and related material required to understand and reproduce their results. In the long term, we foresee good open science becoming the norm and we need to get there together and based on an active and shared discussion.

## 2.2 Details of the Process

The Open Science Initiative will be implemented in a minimally intrusive manner without changing existing reviewing processes. All submissions to EMSE will, also in the future, undergo the same, rigorous review process regardless of whether authors decide to disclose their data and materials or not.

When a paper reaches the final stages of its review process, typically when it has been given a minor revision, the authors will be proposed and encouraged to submit a replication package. The replication package then undergoes a specific review by one member of a newly created open science board. Two members of our editorial board acts as Open Science Chairs and coordinates the process between the authors and the open science board reviewer.

The reviews are made according to a review guideline dedicated to replication packages. The process is, for now, single-blinded, same as the review process of manuscripts.

The detailed reference documentation of the open science peer review process, which we will continuously update also based on the community feedback, can be found at <https://github.com/emsejournal/openscience/>.

## 2.3 Badges

Accepted articles for which the authors disclosed their data are promoted via a specific badge that makes explicit the availability of, for instance, open data sets. Badges have shown to be an effective incentive that increases the participation in open science initiatives (Rowhani-Farid et al. 2017; Baranski et al. 2016). Badges further acknowledge open science practices and make explicit to the community that the code or data has been disclosed according to a certain quality standard.

After extensive analysis of the existing badge systems as well as what would constitute reasonable open science practices to our field, we have decided to begin with a single badge. We intend to adapt this choice based on feedback from the community. Having a single badge makes it easier for open science reviewers to focus their assessment on replication packages and easier for authors to understand the decisions. We plan to use the well-accepted OSI “Open-data” badge<sup>4</sup>.

<sup>4</sup><https://osf.io/tvyxz/wiki/1.%20View%20the%20Badges/>

Upon approval, a manuscript then receives this open science badge to make the readers aware of the existence of a peer-reviewed replication package.

### 3 Outlook

To the best of our knowledge and at the time of writing, this is the first Open Science initiative among the journals in the software engineering research community. We aim at having a positive impact on our community as a whole, and to build a strong community support. We hope that over time other journals in our community will follow.

Yet, the initiative is intended only as a first step. The scope of open science is larger than the availability of reviewed replication packages. Among the directions that we now consider are:

1. Preregistered reports, which are studies accepted only based on previously reviewed study methodology / protocol rather than on fully described results in order to reduce, *inter alia*, potential publication bias.
2. Open reviews of replication packages in order to increase transparency and support the community of authors.

We cordially invite the community to join us in this endeavour and to actively shape it with contributions and active participation in the corresponding online community under <http://bit.ly/emse-osi-group>.

Come join us in this exciting journey to make software engineering more open! We all stand to gain.

### References

- Baranski E, Hardwicke TE, Piechowski S, Falkenberg L-S, Kennett C, Slowik A, Sonnleitner C, Hess-Holden C, Errington T, Fiedler S, Nosek B, Kidwell MC, Lazarević LB (2016) Badges to acknowledge open practices: a simple, low-cost, effective method for increasing transparency. *PLoS Biol* 14:5
- Gent IP, Kotthoff L (2014) Recomputation. org: Experiences of its first year and lessons learned. In: *Proceedings of the 2014 IEEE/ACM 7th international conference on utility and cloud computing*. IEEE Computer Society, pp 968–973
- Méndez Fernández D, Passoth J-H (2018) Empirical software Engineering: From discipline to interdiscipline. *J Syst Softw* 148:170–179
- Piwowar H, Priem J, Larivière V, Alperin JP, Matthias L, Norlander B, Farley A, West J, Haustein S (2018) The State of OA: a Large-scale Analysis of the Prevalence and Impact of Open Access Articles. *PeerJ*
- Rowhani-Farid A, Allen M, Barnett AG (2017) What incentives increase data sharing in health and medical research? A systematic review. *Res Integr Peer Rev* 2(1):4

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.