How do Brazilian Software Engineering Researchers Perceive and Practice Open Science?*

Edson OliveiraJr[1], Christina von Flach G. Chavez[2], André F. R. Cordeiro[1], Daniela Feitosa[2]

[1]Informatics Department – State University of Maringá (UEM)
Maringá – PR – Brazil

[2]Computing Department – Federal University of Bahia (UFBA)
Salvador – BA – Brazil

edson@din.uem.br, flach@ufba.br

cordeiroandrefelipe@gmail.com, danielafeitosa@gmail.com

Abstract. *With the wide popularization and increasing adoption of Open Science, most scientific research areas have discussed its benefits to the overall society represented by any citizen. The openness process aims at promoting free availability of such researches, thus directly impacting scientific evolution. Researchers are encouraged to make scientific research artifacts open for every citizen. In the Software Engineering area we are currently experiencing international Open Science initiatives, such as the ICSE Rose Festival, the ESEM Open Science policies, and the Empirical Software Engineering journal Open Science initiative. However, a little is known about Open Science in the Brazilian Software Engineering community. Therefore, in this paper, we present and discuss the results of a survey on how do our software engineering community perceive and practice Open Science.*

1. Introduction

Main scientific areas as Medicine, Physics, and Social Sciences have benefit from Open Science [NASEM et al. 2018]. UNESCO and internationally recognized coalitions are leaders at promoting Open Science. As this is a trending movement worldwide, Software Engineering (SE), as a crosscutting and crucial scientific area, is starting dealing with it [Mendez et al. 2020]. Specific initiatives as the ICSE, ESEM and EASE conferences policies are currently guiding researcher on adopting Open Science main concepts.

As this is a brand new research topic to SE, we seek to understand **how do Brazilian software engineering researchers perceive and practice Open Science practices**.

2. Research Questions and Methodology

This study aims at surveying SE researchers, **with the purpose of** characterizing their perception and practice of open science, **from the point of view of** open science researchers, **in the context of** Brazilian SE researchers.

*Open peer review artifacts for this paper are available at https://zenodo.org/communities/opensciense2021

We defined 4 research questions: **RQ1:** *Are researchers aware of open science principles?*; **RQ2:** *What is the overall view of open science and its adoption perspective in the researchers' opinion?*; **RQ3:** *How is the open science practices experience of the researchers?*; and **RQ4:** *How are the open science practices performed by the researchers?*.

With **RQ1** we seek what researchers know about Open Science and its main practices, personal experiences, and research openness. **RQ2** discusses the general view on Open Science and its potentially adoption for prospective researches. **RQ3** is the more dense research question as we are interested in specific experiences of Open Science practice, such as, obstacles and tool support. In **RQ4** we seek to characterize how the researchers carry out common Open Science practices.

We carried out a web-based survey[1] according to the guidelines by [Linåker et al. 2015]. We target our audience in Brazilian SE researchers. Our sampling strategy was non-probabilistic. Most of our 31 respondents are male (23 - 74.19%), followed by female (8 - 25.81%). They are distributed in the following Brazilian states: BA and PR (7 - 22.58%); MS, RS, and SP (3 - 9.68%); AL and CE (2 - 6.45%); MA, PI, RJ, RN (1 - 3.23%). The majority of the respondents have more than 10 years (19 - 61.29%) of experience with software engineering followed by between six and 10 years (6 - 19.35%) and between two and five years (6 - 19.35%). None of them has less than two years.

3. Results and Discussion

3.1. RQ1: Researchers Awareness on Open Science

Our first question to the respondents was whether they have ever heard about Open Science. Almost a quarter of them (7 - 22.58%) have never, whereas the remaining (24 - 77.42%) already have.

With relation to what Open Science practices, we wanted to understand which of them the researchers know or already have used. Therefore, we obtained the following: Open Access (21 - 32.81%), Open Data (21 - 32.81%), Open Reproducible Research (15 - 23.44%), Open Science Evaluation (4 - 6.25%), Open Science Policies (3 - 4.69%), Open Science Tools (11 - 17.19%), and None of them (4 - 6.25%).

We also asked them on how they would evaluate their general experience in Open Science practices, which are: "I have some experience in Open Science practices" (15 - 48.39%), "So far, I did not know the Open Science practices" (8 - 25.81%), "I am aware of Open Science, but I have never practiced it in my researches" (6 - 19.35%), and "I have a vast experience in Open Science practices" (2 - 6.45%).

As Open Science practices are getting widely well-know, we asked researchers to whom/what they understand Open Science must be open to: all citizens (24 - 30.77%), scientists of the same research area (15 - 19.23%), scientists of other research areas (15 - 19.23%), groups specifically interested (13 - 16.67%), funding and policy makers (13 - 16.67%), companies and industry (13 - 16.67%), civil and social organizations (11 - 14.10%), no opinion (5 - 6.41%).

[1]Data available at `https://doi.org/10.5281/zenodo.5132192`

3.2. RQ2: Researchers' View of Open Science

The researchers provided us their summarized view of Open Science (38 mentions), as follows: "Open Science is an excellent opportunity for Science, especially with benefits" (13 - 34.21%); "Open Science is an opportunity for Science, with the benefits outweighing the disadvantages" (12 - 31.58%); and "Open Science is mainly positive for science, it has benefits, but also important disadvantages" (11 - 28.95%).

In addition, none of the respondents agree with: "Open Science is an unimportant bureaucratic burden for Science", "Open Science is a troubling new prospect for science", and "Open Science is a real threat to science".

We also asked the researchers whether they adopt Open Science for their researches and for what type of research. Only one person (1 - 3.22%) indicated he/she would not adopt Open Science, and three (9.67%) still not decide about it.

The remaining researchers (27 - 87.09%) state they would adopt Open Science for the following kinds of production: productions without major concerns about data confidentiality; for information surveys; for data, methods and tools, especially open source software; for all production with public funding; for empirical studies, which involve data collection; for planning (protocols), implementation and execution of studies; for sharing experimental data: interviews, questionnaires and responses; and for the development of tools, sharing of procedures and results of empirical studies.

They also mentioned some concerns on adopting Open Science such as: in Startups, the disclosure of results in an open way can make the business unfeasible; the opening of productions may not bring any return to the authors; some researches are funded by companies and therefore should be restricted; and for technological productions as there may be competition in the market.

3.3. RQ3: Practices Experience

The researchers were asked whether they faced obstacles at practicing Open Science. The majority (19 - 61.29%) said yes, whereas the remaining (12 - 38.71%) said no. From those who pointed out some obstacle, we highlight their mentions as follows: availability of data/packages (10), availability of pre-print articles (5), open but unavailable software (4), access costs (3), personal data information for access (1), lack of policies to make articles available openly (1), guidelines for Open Science in SE (1), and private data restriction (1).

We seek to identify whether researchers have used any Open Science tools. Most of them did not know or have used any tool (18 - 58.06%), whereas the remaining know or have used some (13 - 41.94%). They mentioned the following tools to suport Open Science: Zenodo (9), arXiv (6), GitHub (5), ORCID (3), GitLab (2), FigShare (1), Publons (1), Sci-Hub (1), OSF (1), Zotero (1), and Kaggle (1).

With regard to the effort put at verifying the inputs and outputs reusing or sharing when publishing research results, they answered: they do not know (16 - 51.61%); 1-2 days of work (3 - 9.68%); 10-20% of the research time (2 - 6.45%); 1-2 hours (2 - 6.45%); 30 minutes (1 - 3.23%); 4 hours (1 - 3.23%); 5-10 hours (1 - 3.23%); and from one hour to a few days (1 - 3.23%). Four researchers (12.90%) stated they put no effort to do so.

We also asked them on the research openness regarding: Efficiency - sharing data, procedures and/or to optimize science; Equity - access to all scientific results, methods, software, etc., regardless of economic or institutional capacity; Ethics - Open Science is aligned with integrity research principles; Justice - science is often funded by society, so all research results must be available to society; Impact - to surpass traditional metrics for scientific impact (larger audience, greater engagement, etc); Rigor - open access, open Data and/or open reproducible makes science easier.

They answered as follows (IR = an Important Reason, NR = Not a reason, NA = I don't know/I don't have enough information, RR = a relatively important reason, MR = the most important reason): **Efficiency** - IR (14 - 45.16%) and MR (17 - 54.84%); **Equity** - IR (14 - 45.16%), MR (12 - 38.71%), NA (3 - 9.68%), and RR (2 - 6.45%); **Ethics** - MR (16 - 51.61%), IR (13 - 41.94%), NR (1 - 3.23%), and RR (1 - 3.23%); **Justice** - IR (17 - 54.84%), MR (12 - 38.71%), and RR (2 - 6.45%); **Impact** - IR (16 - 51.61%), MR (7 - 22.58%), RR (5 - 16.13%), NR (2 - 6.45%), and NA (1 - 3.23%); and **Rigor** - MR (15 - 48.39%), IR (10 - 32.26%), RR (4 - 12.90%), and NA (2 - 6.45%).

We asked researchers what barriers they should face in case of their institutions adopt Open Science practices. They state: the lack of clear policies (20 - 27.03%); the lack of adequate infrastructure (16 - 21.62%), financing restrictions (14 - 18.92%), time limitations (14 - 18.92%), and fears and uncertainties (10 - 13.51%).

None researches has ever received any kind of Open Science training in their institutions. We also asked them whether they received any kind of support related to Open Science practices in terms of the following criteria: guidelines - (web page/brochure/videos), policies, recommendations; technical infrastructure - models, software, storage, databases, publishing and/or data repositories etc.; specialized support - experts on different aspects of Open Science, research data committees, courses, workshops etc.; financial support and rewards; and career prospects and recognition.

They answered as follows (AS = adequate support, EC = more encouragement or sufficient support, NR = not relevant, and NA = I don't know/I don't have enough information): **guidelines** - EC (16 - 51.61%) and NA (15 - 48.39%); **technical infrastructure** - EC (16 - 51.61%), NA (14 - 45.16%), and AS (1 - 3.23%); **specialized support** - EC (16 - 51.61%), NA (14 - 45.16%), and NR (1 - 3.23%); **financial support and rewards** - EC (16 - 51.61%), NA (13 - 41.94%), and NR (2 - 6.45%); and **career prospects and recognition** - EC (17 - 54.84%), NA (13 - 41.94%), and NR (1 - 3.23%).

With regard to research public licensing, we asked them whether they have ever used any Creative Commons licenses in their researches and artifacts. Most of them have used as follows: CC0 (no restrictions) (8 - 25.81%), BY (attribution) (7 - 22.58%), BY-NC (attribution + non-commercial) (5 - 16.13%), BY-SA (attribution + equal share) (3 - 9.68%), and BY-NC-ND (attribution + non-commercial + no derivatives) (1 - 3.23%). None of the researchers have used BY-ND (attribution + no derivations) or BY-NC-SA (attribution + non-commercial + equal share). One research has used other type of licensing, and six have not used any.

3.4. RQ4: Carrying Out Practices

With regard to the availability of published papers in journals with no restrictions, the researchers affirm that: "Some articles I published are available free and open, with licens-

ing that allows reuse" (16 - 51.61%); "All articles I've published are freely and openly available, with licensing that allows reuse" (7 - 22.58%); "Some articles I published are available with restrictions on reuse" (6 - 19.35%); and "All articles I published are available with restrictions on reuse" (2 - 6.45%). None of the researchers states that "No articles I have published are freely and openly available".

We also asked them whether they believe their experimental elements (resources, algorithms, methods, protocols, etc) are freely accessible and allow reproduction by other groups. They state that: "Some experimental elements are freely accessible and allow the reproduction of the research by other researchers" (19 - 61.29%); and "All experimental elements are freely accessible and allow the reproduction of the research by other researchers" (12 - 38.71%). None of them state that "The experimental elements are not accessible and do not allow the reproduction of the research by other researchers".

With regard to make a preliminary version of papers accessible via an open and external repository, they affirm that: "I have never published a draft manuscript of my scientific research on open platforms outside the journal." (18 - 58.06%); "I have sometimes published a preliminary version of the manuscript of my scientific research on open platforms outside the journal." (12 - 38.71%); and "I always publish a draft manuscript of my scientific research on open platforms outside the journal." (1 - 3.23%).

We asked them whether the general public can access and reuse their data and other materials of their researches with no authors' consent. Thus, they stated that: yes, some data and other materials (16 - 51.61%); yes, all data and other materials (12 - 38.71%); and no, data and other materials cannot be accessed and reused without the author's consent (3 - 9.68%).

The researchers were asked whether they use only open software and tools in their researches. Therefore, they answered as follows: some tools and software are open source (18 - 58.06%); all of them are open source (10 - 32.26%); and none of them is open source (3 - 9.68%).

As expected, the researchers are aware of Open Science, especially for the open access, open data, and open reproducible research concepts. We expected a wider knowledge on open science tools as: (i) researchers are from the SE area in which software is the main artifact; and (ii) open repositories are popular these days, as several journals require at least an external URL with data sets and other artifacts.

We expected a higher perceived general experience in Open Science practices from the researchers as more than a half of them (51.61%) do not report experiences in such practices. We understand this might be because most of the Brazilian funding agencies, SE journals and conferences, or research career tenure processes do not demand promoting open science practices. Some initiatives on this are the Brazilian FAPESP funding agency, and ESEM and EASE conferences, and the Empirical Software Engineering journal.

As expected all researchers agree Open Science is an opportunity for science evolution. In addition, only one researcher clearly stated he/she would not adopt Open Science and three are not yet decided to do it. We understand this might happens due to: (i) this is a very recent trend in the SE community; (ii) they do not have enough information or experience on its practices; or (iii) they were not stimulated by his/her research

environment (i.e., institution, funding agency).

Although the majority of researchers would adopt Open Science, they highlight some concerns on it. We also expected this as, for instance, data confidentiality is still a major discussion topic, and the openness of results might make the business unfeasible. We understand these concerns should be further investigated to provide a common ground for researchers and the business stakeholders.

Some obstacles are tackled by the researchers at practicing Open Science. This is might be related to the effort put to make researches open. Although most researchers could not say anything on it, some reported, for instance, 1-2 hours and 1-2 days working on their researches to make them somehow open. Thus, we believe such obstacles clearly can be overcome as soon and well as we start to disseminate Open Science practices proper examples and references.

The researchers find openness is an important or is the most important reason for diversity, efficiency, equity, ethics, justice, impact, and rigor. We understand these elements are key for the dissemination process, especially for ethics, rigor, and impact.

Researchers reported on the following barriers to Open Science adoption by their institutions: lack of policies, adequate infrastructure, financial and time restrictions. In addition, they state they need mainly more encouragement and a better technical infrastructure to overcome these barriers. We understand such barriers may be associated with the recent Open Science movement in Brazil, especially in the area of Software Engineering, thus widely promoting it can be a way to be prepared for an institution movement towards Open Science, as in several European universities.

Researchers used to make available their journal paper preprints free and open to be reused and their experimental elements to be reproduced. In addition, they use open source tools/software in their researches. We already expected these practices, especially because we are dealing with the Software Engineering area.

We understand we need to increasingly disseminate openness practices as a way to evolve our research area and to provide all citizens access to our researches as premier scientific areas do.

4. Future Research

As future research we intend to increase the sample size with more Brazilian researchers and replicate this survey in a worldwide fashion for the Software Engineering area.

References

Linåker, J. et al. (2015). Guidelines for conducting surveys in software engineering. Technical Report 1.1, Lund University.

Mendez, D. et al. (2020). Open science in software engineering. In *Contemporary Empirical Methods in Software Engineering*, pages 477–501.

NASEM, N. A. o. S., Engineering, Medicine, et al. (2018). *Open science by design: Realizing a vision for 21st century research*. National Academies Press.