Towards Open Science in Acoustics: Foundations and Best Practices

Sascha Spors[1], Matthias Geier[1] and Hagen Wierstorf[2]
[1] *Institute of Communications Engineering, University of Rostock, Germany, Email: Sascha.Spors@uni-rostock.de*
[2] *Filmuniversität Babelsberg KONRAD WOLF*

The Scientific Method

Before introducing the open science approach in detail it is worthwhile to review the foundations of the scientific method. It may be referred to as *'A method of procedure that has characterized natural science since the 17th century, consisting in systematic observation, measurement, and experiment, and the formulation, testing, and modification of hypotheses.'* [1]. Science builds upon a set of well accepted methods which are commonly believed to ensure above formulated principles. Besides these, the reproducibility of results is one of the cornerstones of the scientific method. It may be defined as *'Reproducibility is the ability of an entire analysis of an experiment or study to be duplicated, either by the same researcher or by someone else working independently, whereas reproducing an experiment is called replicating it.'* [2]. The irreproducibility of a wide range of scientific results has drawn significant attention in the last decade [3, 4, 5, 6, 7, 8, 9]. In order to track down the problem of irreproducibility, the application areas of the scientific method have been classified into three branches by [10, 11]:

1. deductive
2. empirical
3. computational

The deductive branch covers results derived e.g. by formal logic and mathematics, the empirical branch e.g. statistical analysis of controlled experiments. The first two are traditional branches, while the last one is considered as a potentially novel branch. It covers results derived by large-scale simulations and data-driven computational science. The measures that have to be taken to ensure reproducibilitly in the traditional branches are quite well known. This does not hold for the third, computational, branch. Here the main intellectual contributions may be encoded in software only.

Besides problems in the research methods themselves, results are often not reproducible since necessary supplementary material as protocols, data and implementations are not available. In many cases only the published results are accessible to other researchers. Open science focuses on the ease of access to scientific data underlying research and therefore supports the ease of reproducibility.

An alternative view on the problem of reproducibility is given by asking **who should benefit from my research**? Potential answers to this question ordered by the degree of transparency are:

- ☐ myself
- ☐ my future self
- ☐ my colleagues
- ☐ other researchers
- ☐ all people in the world
- ☐ science itself

While research conducted for myself does not require any transparency in terms of underlying data, research performed for the sake of science itself should be as open as possible to support scientific advancement.

Open Science

Open science follows the general demand for socialization of knowledge. More specifically, it aims at making the research process transparent on all levels. The advent of scientific journals can be seen as a first step towards openness. Open science has been discussed on a broad level in the last decade. However the exact meaning of the term is still not fully settled. Many measures are discussed to achieve transparency in the research process [12, 13, 14, 15, 16]. We review the most common elements of open science in the following, followed by application examples in the context of a listening experiment.

Elements of Open Science

The various measures used to gain open and transparent science may be classified as follows:

- **Open Source**
  Availability of source code of the implementations used in the research process, e.g. for numerical simulations, statistical analyses and graphical user interfaces.
- **Open (Science) Data**
  Availability of the raw/processed data underlying the research, whereas data does not only refer to electronic resources but also to protocols, material samples, etc. For instance impulse responses, anechoic stimuli and anthropomorphic data.
- **Open Access**
  Free access to published research output, e.g. articles.
- **Open Methodology**
  Detailed documentation of the methodology underlying the research. For instance hypothesis and experimental procedure.

- **Open Notebook Science**
  Availability of the primary record of research results, e.g. laboratory notebooks or detailed mathematical derivations.

- **Open Educational Resources**
  Freely accessible resources for teaching, learning and research. For instance textbooks, slides and examples.

- **Open Peer Review** [17]
  Various measures to increase the transparency of the peer review process, e.g. availability of review comments and crowdsourced review.

Example – Listening Experiment

The workflow of a typical perceptual experiment serves as an example for the usage of the different elements of open science. A perceptual study may be decomposed into the following stages:

1. **Idea**
   The results of a preceding experiment or a discussion among colleagues is often the advent for new research projects. These ideas reveal for instance the motivation behind the research. Publishing the ideas in the spirit of Open Notebook Science makes them accessible to the public.

2. **Design of Experiment**
   The experiment is designed on basis of the initial idea. This involves for instance the formulation of a hypothesis, the definition of an experimental procedure, stimuli and subjects. Open Methodology makes the design available.

3. **Computation**
   The generation of stimuli may involve mathematical deviations for numerical simulations, the implementation of signal processing, as well as data like impulse responses and anechoic recordings. For the experiment itself often a control logic and graphical user interface is implemented. These can be made available following Open Notebook Science, Open Data and Open Source.

4. **Experiment**
   The experiment is conducted in the next step. Here the responses are collected from the test subjects potentially together with other data like head orientation. This constitutes raw data which is processed later in the analysis stage. The data can be published as Open Data as long as no privacy issues are raised.

5. **Analysis**
   The statistical analysis of the raw data is the basis of the generalization of the individual results. This may involve the anonymization and outlier removal as first steps. The analysis procedure and the processed data can be made transparent using the elements Open Methodology, Open Source and Open Data.

6. **Manuscript**
   The design and the results of the perceptual study are compiled into a manuscript. It is composed from text, references and visualizations of the analyzed results. The manuscript may be published as Open Access publication.

7. **Pre-Publication Peer Review**
   The publication process typically involves some kind of quality assessment. The manuscript is evaluated by independent reviewers who provide ratings and suggestions for improvement. If accepted the manuscript is then revised. The peer review process can be made transparent using Open Peer Review.

8. **Publication**
   After successful incorporation of comments from the pre-publication peer review, the manuscript is ready for publication. It may be accompanied by supplementary material in order to improve the ease of reproducibility. In the case of conference papers a presentation may be given which contains other or additional material. The elements involved for publication are Open Access, Open Source and Open Data.

9. **Aftermath**
   After publication, the presented research may be replicated by others. Post-publication review will consider the correctness and relevance of the research. The feedback may require to publish errata or revise the underlying data and code. Finally open questions may bring up ideas for the next research project.

Copyright and Licenses

Licenses are an important aspect of open science. Scientific code and data that is made available to the public should be accompanied by a clear license statement. Otherwise, due to copyright laws, re-distribution and re-use is most likely forbidden. The exact legal implications are rather complex, hard to oversee and depend on the country whose copyright laws apply.

In order to explicitly allow re-distribution and re-use – under well-defined conditions – there are a variety of licensing frameworks which can be used for open science. For text, images, artwork and similar works, the family of creative commons (CC) licenses[1] is a good choice [18], especially the rather permissive variant CC BY and the more restrictive variant CC BY SA. When publishing source code, one of the many available open source licenses should be used. Examples are the permissive BSD or MIT licenses and the more restrictive GNU General Public License (GPL), just to name a few.

To promote wide-spread legal re-use of research results, a permissive license should be preferred. There are even tools which allow authors to waive all their rights, as far as legally possible, most notably CC0. In order to aid in the selection of an appropriate licensing strategy,

[1] `https://creativecommons.org`

recommendations for open science have been given for instance in [19].

Management of Research Data

The systematic management of research data is an important prerequisite for open and reproducible science. This covers the internal as well as public handling of data, whereas the term research data is not restricted to electronic resources. The following principles for research data management have been compiled from the recommendations given in [20, 21, 22, 23]:

- develop a comprehensive data management plan
- use workflow tracking in the research process
- make data findable, accessible, interoperable and reusable (FAIR) [23]
- apply open licensing models
- offer training and qualification

The data management plan should document the data raised and processed, its handling and archival. The use of workflow tracking (e.g. Apache Subversion[2], Git[3]) makes the research process transparent and allows to discover for instance the source of problems. The FAIR policy in conjunction with open licensing models promotes the re-use of published data. Training and qualification is an essential cornerstone of a decent data management strategy. It should highlight the importance of data management and should introduce the data management plan as well as the tools used for its realization.

Note, the instantiation of appropriate data management measures is mandatory in a number of funding schemes.

Empirical Analysis of Open Science

Open science has been evaluated in a number of empirical studies. This covers the actual state with respect to data publications up to the impact of open access publications. The current policy of journals has for instance been evaluated in [24, 25]. The studies reveal that so far there is no common call for sharing code and data, and that the requirements are rather homogeneous. The offer to deposit underlying data and code is taken only by a minority of authors. However, the situation has improved considerably in the past years.

The barriers and incentives of open source and data publications accompanying scientific publications have been raised in a questionnaire within the machine learning community [11]. The main barrier named was the effort for documentation and clean-up, the main incentive the encouragement of scientific advancement.

Various studies have investigated the presumed advantage in terms of citations of open access publications, e.g. [26, 27]. An clear advantage could be shown in most cases and less to no advantage in rare cases.

[2]`https://subversion.apache.org`
[3]`https://git-scm.com`

Personal Experience

Besides using open toolboxes and datasets, our own experience with open science began with the open source release of the SoundScape Renderer (SSR) [28] in 2010[4]. Since then various software toolboxes, datasets, open access papers and open educational resources have been published[5,6]. We use Redmine[7], SVN and Git for internal data management. Public releases are mainly available on GitHub[8] and referenced by Zenodo[9], if required. Challenges were the initial effort to get familiar with the workflow and license models. We furthermore did not find a convenient solution for version tracking (of a high number) of large files, like for instance room acoustic measurements. A clear benefit is the need for documentation, clean-up and internal discussion before releasing data to the public. While this effort seems to be undesirable at first glance, it turns out to save a lot of work later on. The feedback from the community is very positive and a number of bugs have been reported. We expect that our publicly released scientific code and data contain less bugs than if we would have used it only internally.

A common concern often raised is the misuse of openly published data by others, for instance for commercial activities. So far we have not seen any misuse of our materials and hence cannot confirm this concern.

Conclusions

We have introduced the concept of open science and discussed its application in the field of acoustics. The reproducibility of results is an essential cornerstone of the scientific method. Open science by itself does not ensure the ease of reproducibility of published results. For instance the accessibility of code for a numeric simulation without proper documentation can be very limited and may lead to irreproducibility of the simulation results. The FAIR policy [23] – *make data findable, accessible, interoperable and reusable* – of the European Commission is a first step towards quality standards for data publications. Other initiatives define minimum quality standards for data publications [29] or have instantiated a peer review for the reproducibility of published findings.

Training and qualification has to be understood as an integral part of research data management and open science. It should convey its foundations as well as common tools and workflows.

The current movement towards open science is still in a phase of development and consolidation. The scientific ecosystem, for instance journals and funding organizations, has taken up the topic in order to improve the management of research data and the reproducibility of scientific findings. However, at the current state most

[4]`https://github.com/SoundScapeRenderer`
[5]`https://github.com/spatialaudio`
[6]`https://github.com/sfstoolbox`
[7]`http://www.redmine.org`
[8]`https://github.com`
[9]`https://zenodo.org`

institutional evaluation measures do not account for an engagement in open science. This is strongly linked to the problem of personal attribution when contributing to open source and open datasets. The scientific track record nowadays builds heavily on such measures as citation index and third party funding which potentially hinder the deep involvement into open science.

References

[1] "Scientific method," in *Oxford Dictionaries: British and World English*, 2016, retrieved 28 May 2016. [Online]. Available: https://en.oxforddictionaries.com/definition/scientific_method

[2] J. T. Leek and R. D. Peng, "Opinion: Reproducible research can still be wrong: Adopting a prevention approach," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 112, no. 6, pp. 1645–1646, 2015.

[3] A. Borgwardt, *Wissenschaft auf Abwegen? Zum drohenden Qualitätsverlust in der Wissenschaft.* Friedrich-Ebert-Stiftung, 2014.

[4] "Retraction watch," http://retractionwatch.com, accessed 24.11.2016.

[5] J. P. Ioannidis, "Why most published research findings are false," *PLoS Medicine*, vol. 2, no. 8, 2005.

[6] O. S. Collaboration *et al.*, "Estimating the reproducibility of psychological science," *Science*, vol. 349, no. 6251, 2015.

[7] I. Chalmers and P. Glasziou, "Avoidable waste in the production and reporting of research evidence," *Obstetrics & Gynecology*, vol. 114, no. 6, pp. 1341–1345, 2009.

[8] L. P. Freedman, I. M. Cockburn, and T. S. Simcoe, "The economics of reproducibility in preclinical research," *PLoS Biology*, vol. 13, no. 6, 2015.

[9] D. W. Howells, E. S. Sena, and M. R. Macleod, "Bringing rigour to translational medicine," *Nature Reviews Neurology*, vol. 10, no. 1, pp. 37–43, 2014.

[10] M. Shahram, V. Stodden, D. L. Donoho, A. Maleki, and I. U. Rahman, "15 years of reproducible research in computational harmonic analysis," *Computing in Science and Engineering*, vol. 11, no. 1, pp. 8–18, 2009.

[11] V. Stodden, "Resolving reproducibility in computational science: Tools, policy, and culture," 10 2014, talk given at the University of Arizona.

[12] N. Pontika, P. Knoth, M. Cancellieri, and S. Pearce, "Fostering open science to research using a taxonomy and an elearning portal," in *Proceedings of the 15th International Conference on Knowledge Technologies and Data-driven Business*, 2015, pp. 11:1–11:8. [Online]. Available: http://doi.acm.org/10.1145/2809563.2809571

[13] "Reproducibility and reliability of biomedical research: improving research practice," The Academy of Medical Sciences, Tech. Rep., 2015.

[10] `https://github.com/spatialaudio/DAGA2017_towards_open_science_in_acoustics`

[14] S. Albagli, M. L. Maciel, and A. H. Abdo, Eds., *Open Science, open issues*, 2015.

[15] R. Vrana, "Open science, open access and open educational resources: Challenges and opportunities," in *Information and Communication Technology, Electronics and Microelectronics (MIPRO)*, May 2015, pp. 886–890.

[16] P. Kraker, D. Leony, W. Reinhardt, and G. Beham, "The case for an open science in technology enhanced learning," *International Journal Technology of Enhanced Learning*, vol. 3, no. 6, pp. 643–654, 2011.

[17] E. Ford, "Defining and characterizing open peer review: A review of the literature," *Library Faculty Publications and Presentations, Portland State University*, 2013.

[18] T. Kreuzer, *Open Content - A Practical Guide to Using Creative Commons Licences.* Wikimedia Deutschland, the German Commission for UNESCO and the North Rhine-Westphalian Library Service Centre (hbz), 2014.

[19] V. Stodden, "The legal framework for reproducible scientific research - licensing and copyright," *Computing in Science and Engineering*, vol. 11, no. 1, pp. 35–40, 2009.

[20] *Sicherung guter wissenschaftlicher Praxis.* Deutsche Forschungsgemeinschaft, 2013.

[21] Hochschulrektorenkonferenz, "Management von Forschungsdaten - eine zentrale strategische Herausforderung für Hochschulleitungen."

[22] V. Stodden and S. Miguez, "Best practices for computational science: Software infrastructure and environments for reproducible and extensible research," *Journal of open research software*, vol. 2, no. 1, pp. 1–6, 2014.

[23] European Research Commission, "H2020 programme: Guidelines of FAIR data management in horizon 2020," version 3.0, 26.7.2016.

[24] V. Stodden, P. Guo, and Z. Ma, "Toward reproducible computational research: Empirical anaysis of data and code policy adoption by journals," *PLoS ONE*, vol. 8, no. 6, pp. 1–8, 2013.

[25] A. A. Alsheikh-Ali, W. Qureshi, M. H. Al-Mallah, and J. P. Ioannidis, "Public availability of published research data in high-impact journals," *PLoS ONE*, vol. 6, no. 9, pp. 1–4, 2011.

[26] A. Swan, "The open access citation advantage: Studies and results to date," http://eprints.soton.ac.uk/id/eprint/268516, 2010.

[27] M. J. McCabe and C. M. Snyder, "Identifying the effect of open access on citations using a panel of science journals," *Economic Inquiry*, vol. 52, no. 4, pp. 1284–1300, 2014.

[28] M. Geier, J. Ahrens, A. Möhl, S. Spors, J. Loh, and K. Bredies, "The soundscape renderer: A versatile framework for spatial audio reproduction," in *Proceedings of the DEGA WFS Symposium.* Ilmenau, Germany: Deutsche Gesellschaft für Akustik (DEGA), September 2007.

[29] The TOP Guidelines Committe, "Promoting an open research culture: The top guidelines for journals," version 2, 27.6.2015.