

Editors' Note

Peter Wittenburg^{1†} & George Strawn^{2†}

¹Max Planck Computing and Data Facility, Gießenbachstraße 2, 85748 Garching, Germany

²US National Academy of Sciences, Washington DC 20418, USA

Citation: Wittenburg, P., George S.: Editors' note. *Data Intelligence* 3(1), 1-4 (2021). doi: 10.1162/dint_e_00068

In 2019 the German Leibniz research organization sponsored a conference on Open Science (OS) with the idea to publish some of the presented papers in the *Data Intelligence* journal. Becoming engaged as editors, we recognized that the term “Open Science” was coined about 10 years ago with the intention as pointed out by Michael Nielson: “OS is the idea that scientific knowledge of all kinds should be openly shared as early as is practical in the discovery process”. Crow and Tanenbaum[®] stated in 2020 that with OS a great return of investment could be achieved: for each invested dollar about 140 dollars were returned. However, after having participated in many meetings where the ideal of OS was presented repeatedly, after having read many policy papers from many different research organizations and funders, and after having realized that the practices in the data labs have not changed substantially yet, we decided that it is time to review the state of OS in a broader manner.

The conference presentations, especially those four and an additional one from the Library of the Peking University, China, selected for publication showed that librarians were pushing activities to foster OS without much support from the hierarchies in the research organizations to influence practices. We must thank the librarians for their energy, but the effect of this activity was that the concept of “Open by Design” shifted to the concept of “Open by Publication” and that researchers tend to believe that OS is something some librarians will do for them at the end of projects. It is not only the experience of COVID-19 which demonstrated that this concept change is not appropriate to foster data-driven research. Not only in the medical sector it is a must to exchange digital objects, be it data, metadata, software or other research artifacts, as quickly as possible. This is true for other research areas as well, just think of data about earthquakes, climatic influences, etc. And indeed, researchers exchange data very quickly amongst their peers, i.e., in limited personal circuits. OS, however, is meant to replace this accidental practice of sharing

[†] Corresponding authors: Peter Wittenburg (Email: peter.wittenburg@mpcdf.mpg.de; ORCID: 0000-0003-3538-0106); George Strawn (Email: gostrawn@gmail.com; ORCID: 0000-0003-4098-0464).

[®] M. Crow, G. Tanenbaum. We must tear down the barriers that impede scientific progress, December 2020. Available at: <https://www.scientificamerican.com/article/we-must-tear-down-the-barriers-that-impede-scientific-progress/>.

by a systematic approach which can be compared to the change to systematically publish research results with the help of journal papers centuries ago.

It should be noted that “Open by Design” implies (1) to carry out systematic exchange from the beginning and not have to wait for years until publications have been created, (2) to apply suitable mechanisms to the required documentation immediately and not to engage curators to do the hard and expensive documentation and curation work after years, (3) to exchange the whole richness of data as being generated and not just the few data sets that are connected to publications. “OS by Design”, however, requires changing practices in the labs, which is much harder to achieve and will not be liked by researchers as long as efficient support tools are missing.

Being aware of differences between policy level statements and data lab practices, we thought that it would be important for a special issue on OS to not just include papers from the conference, but to ask a few distinguished colleagues with different backgrounds to write a paper on their view on OS. With the exception of one colleague who was under an enormous time pressure due to COVID-19 research all accepted our invitation. The results are eight invited papers about OS and one paper describing data lab practices based on deep insights into about 70 research infrastructure projects.

The statements indeed show a broad spectrum of opinions. Paolo Budroni, a philosopher by education, puts our discussions on OS into the historical context indicating that openness was always an important issue influenced by the technological possibilities. Heather Joseph, a librarian by training, makes an excellent statement pro OS very much aligned with official policy reports on OS. Jonathan Clark, with his strong publishing background, puts the importance of trust in data and the value of links between digital objects into the centre of an OS domain. John Wood, based on his many years of experience with research projects, argues that we need to lower the expectations to make OS feasible. Klaus Tochtermann, a computer scientist by background and active in developing large research infrastructures in Germany, demonstrates that already in the area of integrating metadata across disciplines there are many roadblocks to overcome. George Strawn, based on his experience with getting the Internet started and other IT projects, argues that the usual hype cycle curve will become true again and that it will take time until realistic OS scenarios will become daily practices. Peter Wittenburg, based on his involvement in setting up large research infrastructures and his close relation with many data labs, also argues that implementing a fair “OS by Design” scenario will take time due to several non-technical roadblocks.

The contribution of Jean Claude Burgelman is special in this context since he was one of the key persons who solved the policy puzzle to get final agreements of all member states to make the European Open Science Cloud (EOSC) happen, which is an initiative for building an infrastructure to pave the way towards OS in Europe. This serious and fair description of complex political activities that created many frustrations but finally were successful is in our views a unique document worth elaborating on. We therefore asked a few persons who were involved in these processes arguing from highly different backgrounds and points of views to respond with comments on this paper. Again, all six experts whom we asked to participate in this exercise in addition to the editors agreed to make statements: Natalia Manola (IT researcher and

OpenAIRE infrastructure chair), Edit Herczog (RDA council member and ex-member of European Parliament), Per Öster (CSC Director and involved in e-Infrastructures), Barend Mons (Bioinformatics researcher, chair of early EOSC boards and GO FAIR leader), Hanifeh Khayyeri (member of Swedish Research Council and of EOSC Board) and Dimitris Koureas (Biodiversity researcher and leader of DISSCO research infrastructure). We see this elaboration about the EOSC process as a start for further open discussions which should take place in 2021 and which may help to shape EOSC and thus help in establishing an OS domain.

Another paper by Keith Jefferey et. al. was added which describes in broader terms the current practices in the data labs inspired by deep and recent insights into about 70 research infrastructures in Europe. It is meant to indicate how distant OS policy and data practices in the data labs still are and which hurdles need to be overcome to make progress in “OS by Design” affecting the practices.

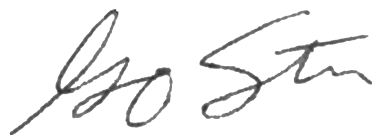
Due to the scientific and economic perspectives we believe that there is no doubt that OS will become a daily practice for all researchers. But we should be aware that setting up systematic procedures implementing “OS by Design” and thus covering the richness of Digital Objects in terms of volumes and types will still take some time. Essential roadblocks need to be overcome, unrealistic expectations need to be reduced, tools and mechanisms need to be developed that are attractive for the researchers to adapt their practices, and the gap between policy level documents and practices needs to be closed.

It is time to thank the authors of the conference papers, of the invited statements on OS and the comments to Jean Claude’s note on EOSC for their excellent contributions and to thank Fenghong Liu who put forth the idea to organize such an issue and the editorial office for all their efforts to get the special issue of the *Data Intelligence Journal* on OS published.

January 2021



Guest Editors:



AUTHOR BIOGRAPHY



Peter Wittenburg was Executive Director of Research Data Alliance (RDA) Europe, Member of RDA Technical Advisory Board, Scientific Coordinator of European Data Infrastructure (EUDAT) and Technical Director of the CLARIN and DOBES Research Infrastructures. He set up and led the Technical Group with about 35 experts at Max Planck Institute (MPI) for Psycholinguistics and then led the Language Archiving Group with about 25 experts. Since 2000 he has played leading roles in a variety of European (funded by the European Commission) and national projects (funded by MPS, DFG, BMBF, NWO) and ISO initiatives (ISO TC37/SC4). He won the Heinz Billing Award of the Max Planck Society (MPS) for the advancement of scientific computation in 2011 and received an honorary doctorate from University Tübingen in 2013.
ORCID: 0000-0003-3538-0106



George Strawn is currently the director of the Board on Research Data and Information at the National Academies of Sciences, Engineering, and Medicine where he focuses on Open Science and FAIR data. Prior to joining the Academies, Dr. Strawn was the director of the National Coordination Office (NCO) for the Networking and Information Technology Research and Development (NITRD) Program and co-chair of the NITRD interagency committee.
ORCID: 0000-0003-4098-0464