Open Research in Business and Information Systems Engineering

Wil van der Aalst, Martin Bichler & Armin Heinzl

Springer

Springer

CrossMark

Open Research in Business and Information Systems Engineering

Wil van der Aalst · Martin Bichler · Armin Heinzl

1 Towards Open Research

We can all witness rapid changes in the way people conduct research, publish results, and share artifacts. This is affecting the way journals and conferences operate. It seems that we are gradually moving towards truly "open research", sometimes also referred to as "open science". Open science is the movement towards making scientific research and related artifacts (data, software, etc.) accessible to all levels of inquisition. Although digitization and the Internet have dramatically changed, globalized, and accelerated communication in general, the way that research results are communicated through journals remains fairly traditional (Groen 2007). BISE is no exception. The reviewing process is double-blind, but reviews are not publically available. A small fraction of papers are available through Springer's Open Access (Springer 2016), but the majority of BISE papers still require a subscription. Accepted papers may use or present data and software, but these are not required to be public

Prof. Dr. ir. W. van der Aalst (✉)
Department of Mathematics and Computer Science (MF 7.103),
Eindhoven University of Technology, PO Box 513,
5600 MB Eindhoven, The Netherlands
e-mail: w.m.p.v.d.aalst@tue.nl

Prof. Dr. M. Bichler
Department of Informatics, Decision Sciences and Systems,
Technical University of Munich (TUM), Boltzmannstr 3,
85748 Munich, Germany
e-mail: bichler@in.tum.de

Prof. Dr. A. Heinzl
General Management and Information Systems, University of
Mannheim, 68161 Mannheim, Germany
e-mail: heinzl@uni-mannheim.de

(thus making it difficult to verify results and compare approaches). Of course there are all kinds of practical reasons why journals like BISE do not enforce "open research" (yet). However, it is good to deliberate on the topic and seek feedback from the BISE community (Fig. 1).

Why are we discussing the topic of "open research" now? It seems that the way we publish and disseminate results is about to change: Governments are discussing the topic, and elements of open research are becoming mandatory for government-funded research. For example, the European Commission is actively pushing "open science" (European Commission 2016). The Amsterdam Call for Action on Open Science (Netherlands EU Presidency 2016) was written based on an open science meeting in April 2016 that was organized by the Dutch Presidency of the European Union. NWO, the Dutch science foundation, recently stated that "research results paid for by public funds should be freely accessible worldwide" (NWO 2016). Open research concerns both scientific publications and other forms of scientific output. There is a lot more to open research than just open access (i.e., publications that are freely accessible to all). For example, open access journals do not necessarily enforce the sharing of artifacts such as data sets and software. In fact, the last part of this editorial focuses on the sharing of these artifacts as they are vital for many of the results reported in BISE.

2 Open Publication

In 1665, Henry Oldenburg became the editor of Philosophical Transactions of the Royal Society. This was the first academic journal devoted to science, and this development coincided with the formation of scientific

🕿 Springer

Fig. 1 Open research and some of its ingredients

academies (David 2004). For example, in 1660 the Royal Society was established in England and in 1666 the French Academy of Sciences was founded (Nielsen 2011). Before the establishment of prestigious journals, there were often long discussions of the ownership of new inventions. Scientists would hide results, afraid that competitors would claim priority. Results were even encrypted to control distribution. Obviously, concealing results because competitors can claim priority does not help to advance science. Journals helped to resolve such conflicts and aided the dissemination of results. Since the creation of the first journals in the 17th century, the number of journals has been steadily growing. By now there are tens of thousands of journals, BISE being one of them.

As long as journals existed only in paper form, it was obvious that readers would need to pay for the print and distribution of published papers. However, today, researchers mostly access the electronic versions of journals. This triggers the question why publishers should still receive substantial amounts of money for the distribution of work done by the scientific community (writing and reviewing). In a way, citizens are paying twice for the same research. Taxes are used to fund scientific research, but still subscription fees need to be paid (by academic institutions and citizens alike) to access the results of this government-funded research. Moreover, researchers from developing countries do not have access to the publications in expensive journals. Therefore, discussions to enforce the free availability of scientific publications ("open access") take place at different levels. For example, NWO is enforcing an open publication policy: projects funded by this organization need to be publically available (NWO 2016). At the EU level one can observe similar developments. Pressure by governments and the academic community has led to the creation of new business models where research organizations pay a publication fee to enable open access. Consider for example Springer's Open Access policy (Springer 2016). It is good to see that funding organizations have started to realize that managing a journal (editing papers, handling review processes, etc.) and making millions of papers accessible electronically is something that requires substantial resources and a professional organization. Most attempts to create fully open journals without involving publishers have failed. There are a few notable exceptions, e.g., the PLOS (Public Library of

Science) initiative aiming at a library of open access journals and other scientific literature under an open content license (Public Library of Science 2016). Despite these exceptions, most open journals have problems in terms of reputation and sustainability. When a journal "fails", there are no guarantees that the corresponding publications remain available indefinitely. Aspects such as stability, reputation, infrastructure, accessibility, etc. need to be considered when talking about open access.

3 Open Reviewing

Profound reviewing is essential for ensuring the quality of scientific research. New ideas are often generated based on critical feedback. Incorrect or unclear results should be scrutinized by experts before they are widely distributed. The "publish or perish" culture has unfortunately created a situation where young researchers are stimulated to "write rather than read". Part of the problem is the wide range of scientific outlets; the uptake of the Internet has triggered a tsunami of journals not bounded by the physical limitations of classical paper journals. Everyone can start a new journal at any time, and for researchers it is time-consuming to manage the information overload. There is also a mismatch between the people that review and the people that submit, e.g., experienced researchers from some countries are expected to review the work of inexperienced researchers from other countries where researchers are forced to submit to journals early in their career. There should be a healthy balance between reviewing and submitting papers at the level of individuals as well as at the level of groups or even countries. The reviewing system breaks down if one group is massively submitting papers, whereas another group needs to take care of the quality control.

On the one hand, as mentioned, the number of journals is growing. On the other hand, in most established disciplines there is a fairly stable set of top-tier journals or conferences, and the competition of authors for papers in these outlets has increased significantly (Attema et al. 2014).

Regrettably, review work is hardly visible and not rewarded sufficiently in the current academic climate. An author's curriculum vitae will never reveal that the person

avoids peer review work or delivers superficial reviews. BISE uses a double-blind reviewing process. This is good in the sense that reviewers can give unbiased and independent feedback, but also renders the review process closed and invisible for the outside world. A fully open review process can set incentives for various types of strategies such as retaliation or publication cartels similar to what has been observed in online reputation systems (Ye et al. 2014). Hence, there is not a simple solution. However, it is important to think of new ways of reviewing, acknowledging the importance of true scientific interaction and improving transparency at the same time. Outstanding reviewer awards, that were introduced recently, can only be a starting point. Becoming an editorial board member can be an incentive, but more needs to be done, considering the time spent to evaluate each paper.

4 Open Data

We live in a world flooded by data (big and small). Data are collected about anything, at any time, and at any place. Consider for example the "Internet of Events" (IoE) composed of the Internet of Content (IoC), the Internet of People (IoP), the Internet of Things (IoT), and the Internet of Locations (IoL) (Van der Aalst 2016). People, devices, organizations, software systems, phones, etc. all record "events", i.e., things that happen in the real world. This is changing the way people conduct research. There is a shift from purely model-driven research and mostly conceptual research to research based on real-life data (Van der Aalst 2016). For example, we are able to monitor how people interact with software and the intelligent devices around them. As researchers we have an obligation to use this.

BISE papers increasingly depend on data. As research data used in publications become more detailed and their volume increases, it becomes more difficult to judge a paper without having access to the corresponding data. The progress of science depends on the ability to reproduce scientific results. Unfortunately, as a recent study in Nature shows (Baker 2016), most of the results described in literature *cannot* be reproduced. Based on a survey involving 1576 researchers, the Nature article (Baker 2016) reveals that 70 % of researchers have tried and failed to reproduce another scientist's experiments, and more than half have failed to reproduce their own experiments. Factors explaining this include the pressure to publish and selective reporting.

In information systems, there is no established tradition to share data and reproduce existing results. Many papers aim at originality rather than at reproducing and analyzing already published results. There are a few exceptions in more data-driven branches of information systems research

and beyond (Vlaeminck and Herrmann 2015). Consider for example the field of process mining. Most process mining papers use or provide public data in XES format. There are competitions like the Business Process Intelligence Challenge (BPIC) (Van Dongen 2016) which provide real life data, and it has become impossible to publish papers on a new process discovery or conformance checking technique without showing results for publically available data. For most other branches of BISE research this is not (yet) the case and perhaps also more difficult. There may be a range of practical limitations when sharing data. For example, data may be confidential or in a format that cannot be interpreted easily by others. However, the BISE community should pose itself questions like:

- Should all data used in BISE papers be publically available?
- How can we ensure the reproducibility of results?
- How to create a culture of sharing data and reproducing scientific results?
- How to ensure the availability of data over a longer period?

Note that it is far from trivial to make data accessible for a longer period. Published papers typically remain available "forever" (assuming a reputable publisher). However, the data used in such papers may only exist on the laptop of a PhD student or on the website of the research group. When projects end or researchers retire, the corresponding data sets often disappear. Initiatives like the 4TU Center for Research Data (2016) aim to ensure the long-term availability of data. Data sets hosted by this center have a Digital Object Identifier (DOI) and are guaranteed to be available indefinitely. Researchers can click on such a DOI link in a paper and immediately obtain access to the corresponding data. The editors of BISE are aiming for a data availability policy for the journal in the near future.

5 Open Software

Software is vital for most of the research published in BISE. In many cases novel software is developed in order to carry out the research. Consider for example a process mining paper presenting a new algorithm that is evaluated by using several data sets. The paper could not exist without the software and the data (but both can exist without the paper!). However, the paper may be accepted without providing access to the data and/or software. The authors may have made a programming error or consciously (or unconsciously) manipulated the results. The only way to verify this is by using the software and repeating the experiments. We must keep in mind that the "science is wrong" if the software is wrong.

Purely analytical research can be evaluated and replicated based on the paper only. However, more and more academic work is based on an implemented system that cannot be fully described in an academic paper.

Some papers report on software systems that have only existed on the PhD student's computer. Authors may describe the architecture of a complex system that only partly existed. Functionality suggested in the paper may not have been implemented. Unfortunately, such practices seem widespread (just take a random sample of papers presenting complex IT artifacts and ask the authors to provide the software). For an external party, results are almost impossible to evaluate without access to the code. The reviewer needs to make guesses based on the reputation of the authors. This is undesirable, because ensuring the reliability and reproducibility of scientific results is one of our main contributions to society.

Fortunately, more and more research projects develop open source software as an important by-product of research. People can inspect reported software artifacts and even modify and improve them. Open research adopts ideas and the mindset originating from the open source community. Note that "open source" software is by definition "open software". However, "open software" doesn't need to be "open source". For example, people can share an executable program without sharing the source code.

It is important to create a "level playing field" in research. For example, there may be two competing research groups. Assume group A provides open source software and group B only uses/develops proprietary software. Group B can use ideas from group A and write papers comparing the performance of its software with the software of the other group. This doesn't hold in the other direction. Even when group A is sure that the results of group B are flawed, it cannot demonstrate this easily.

There are also a few questions to put to the BISE community related to software:

- Should software reported in BISE papers be publically available?
- When is the use of proprietary software acceptable?
- Should authors with an industrial background be treated differently?
- How can the availability of software be ensured over a longer period?
- How long should software be available?

Providing open software is easier said than done. Rapid technological advances make it difficult to maintain software just for the purpose of reproducibility. Some journals have introduced new policies and they publish reproducibility articles (Wolke et al. 2016). Such articles include a validation by reviewers; they also provide access to the source code in a repository and possibly a virtual machine. This makes it possible for readers to reproduce the results of a system-oriented paper with the respective experiments at relatively low cost.

6 Final Remarks

This editorial aims to trigger a discussion on "open research" in the BISE community. Open research relates to open access of publications and novel ways of reviewing. It also refers to opening up the artifacts (data and software) publications build upon. Scientific journals exist since the 17th century. However, due to the digitization of science, the "rules of the game" are changing rapidly. For example, reproducibility is a key concern and new possibilities in our digital society should be exploited to facilitate this. We should reward authors who share data and software. It is probably too early to make this mandatory for all BISE papers, but a trend towards more "openness" is inevitable and also highly desirable. Sharing artifacts and providing transparency will help us to advance science faster.

We hope that our thoughts will facilitate the further development of BISE and trigger discussions within the community regarding open research.

References

Attema AE, Brouwer WBF, van Exel J (2014) Your right arm for a publication in AER? Econ Inq 52(1):495–502

Baker M (2016) 1,500 scientists lift the lid on reproducibility. Nat 533:452–454

David P (2004) Understanding the emergence of "open science" institutions: functionalist economics in historical context. Ind Corp Chang 13(4):571–589

European Commission (2016) Open innovation, open science, open to the world: a vision for Europe. https://ec.europa.eu/digital-single-market/en/news/open-innovation-open-science-open-world-vision-europe. Accessed 15 Sept 2016

Groen FK (2007) Access to medical knowledge: libraries, digitization, and the public good. Scarecrow, Lanham

Nielsen M (2011) Reinventing discovery: the new era of networked science. Princeton University Press, Princeton

NWO (2016) Open access at NWO. http://www.nwo.nl/en/policies/open+science/open+access+publishing. Accessed 15 Sept 2016

Netherlands EU Presidency (2016) Amsterdam call for action on open science. https://english.eu2016.nl/documents/reports/2016/04/04/amsterdam-call-for-action-on-open-science. Accessed 15 Sept 2016

Public Library of Science (2016) Public library of science (PLOS): open for discovery. http://www.plos.org. Accessed 18 Sept 2016

Springer (2016) SpringerOpen. http://www.springeropen.com. Accessed 15 Sept 2016

4TU Center for Research Data (2016) Data collections. http://data.4tu.nl/. Accessed 15 Sept 2016

Van der Aalst W (2016) Chapter 1 data science in action. Process mining: data science in action. Springer, Berlin

Van Dongen B (2016) Business process intelligence challenge (BPIC). http://www.win.tue.nl/bpi/doku.php?id=2016:challenge. Accessed 15 Sept 2016

Vlaeminck S, Herrmann L-K (2015) Data policies and data archives: a new paradigm for academic publishing in economic sciences? New avenues for electronic publishing in the age of infinite collections and citizen science: scale, openness and trust. Proceedings of the 19th International Conference on Electronic Publishing, pp 145–155. http://hdl.handle.net/10419/121278. Accessed 15 Sept 2016

Wolke A, Bichler M, Chirigati F, Steeves V (2016) Reproducible experiments on dynamic resource allocation in cloud data centers. Inf Syst 59:98–101

Ye S, Gao G, Viswanathan S (2014) Strategic behavior in online reputation systems: evidence from revoking on eBay. MIS Q 38(4):1033–1056