The case for an open science in technology enhanced learning

Peter Kraker*

Know-Center,
Graz University of Technology,
Inffeldgasse 21a, 8010 Graz, Austria
E-mail: pkraker@know-center.at
*Corresponding author

Derick Leony

Department of Telematic Engineering,
Carlos III University of Madrid,
Av. Universidad, 30, Edif. Torres Quevedo,
E-28911 Leganés, Madrid, Spain
E-mail: dleony@it.uc3m.es

Wolfgang Reinhardt

Department of Computer Science,
University of Paderborn,
Fürstenallee 11, 3102 Paderborn, Germany
E-mail: wolle@upb.de

Günter Beham

Graz University of Technology,
Inffeldgasse 21a, 8010 Graz, Austria
E-mail: guenter.beham@student.tugraz.at

Abstract: In this paper, we make the case for an open science in technology enhanced learning (TEL). Open science means opening up the research process by making all of its outcomes, and the way in which these outcomes were achieved, publicly available on the World Wide Web. In our vision, the adoption of open science instruments provides a set of solid and sustainable ways to connect the disjoint communities in TEL. Furthermore, we envision that researchers in TEL would be able to reproduce the results from any paper using the instruments of open science. Therefore, we introduce the concept of open methodology, which stands for sharing the methodological details of the evaluation provided, and the tools used for data collection and analysis. We discuss the potential benefits, but also the issues of an open science, and conclude with a set of recommendations for implementing open science in TEL.

Keywords: reproducible research; comparability of research; open methodology; open access; open data; open source; vision paper; barriers to open science; implementation of open science; open science; technology enhanced learning; TEL.

Reference to this paper should be made as follows: Kraker, P., Leony, D., Reinhardt, W. and Beham, G. (2011) 'The case for an open science in technology enhanced learning', *Int. J. Technology Enhanced Learning*, Vol. 3, No. 6, pp.643–654.

Biographical notes: Peter Kraker received his Masters degree (Mag. rer. soc. oec.) in Business Administration from University of Graz in 2007 and his Bachelor degree (BSc) in Computer Science from Graz University of Technology in 2008. Currently, he is a Research Assistant at Know-Center, Graz University of Technology. His main research interest and the topic of his PhD thesis is the way in which TEL researchers use the web and in particular Web 2.0 for their work, and how artefacts generated in that process can be used to structure the field.

Derick Leony received his BSc in 2006 in Computer Science and MSc in 2007 in Telecommunications from Galileo University, Guatemala. He received his MSc in Telematic Engineering from Carlos III University of Madrid in 2010. Currently, he is a PhD student and a Teaching Assistant in the Department of Telematic Engineering at Carlos III University of Madrid. His research topic of interest is computer supported collaborative learning, with a special focus on the analysis and comparison of activity patterns in collaborative learning environments.

Wolfgang Reinhardt received his BSc in 2005 and Diploma in Computer Science in 2006 from the University of Paderborn, Germany. Currently, he works as a Research Assistant at the Computer Science Education Group at the University of Paderborn and is external PhD candidate at the Open University of The Netherlands. His PhD deals with the enhancement of awareness in research networks and the recommendation of information objects in ill-structured learning networks.

Günter Beham worked as a Researcher for Graz University of Technology and Know-Center Graz in the area of technology enhanced learning. His research interest is in ubiquitous and Open Source Software solutions for work-integrated learning. Currently, he is affiliated with Graz University of Technology as a student.

1 Introduction

Overall, modern science is a great success story, exposing exponential growth rates from its inception up until the present day (Price, 1963; Larsen and von Ins, 2010; National Science Board, 2010). Nevertheless, several weaknesses of the current science system have been pointed out in the literature. Due to the economics of publication, results tend to be exaggerated and inflated (Young et al., 2008). Furthermore, the reproducibility of results is severely limited by the fact that datasets are not distributed with the publication, and that papers do not contain all the methodological information needed to reproduce a certain research result (Knorr-Cetina, 1981). These factors, in combination with exponential growth and the ensuing information overload have contributed to simultaneous and repeated discoveries and the formation of ‘invisible colleges’ (Price, 1963). In the area of technology enhanced learning (TEL), we can observe numerous disjoint scientific communities (Gillet et al., 2009; Reinhardt et al., 2011), a phenomenon also found in other scientific areas (e.g., in human-computer interaction, Henry et al., 2007). Recent scientometric studies in the area of TEL (Kirby et al., 2005; Mauer and Khan, 2010) show that there is a low cross-citation rate and cross-authorship rate between the individual communities. As a consequence, the TEL landscape suffers from a significant degree of fragmentation, preventing the field from leveraging the effects of interdisciplinary work.

In this paper, we will show how an open science would contribute to overcome these weaknesses in TEL. There is no common definition for open science yet (c.f., e.g., Gezelter, 2009; Nielsen, 2008). To us, open science means opening up the research process by making all of its outcomes, and the way in which these outcomes were achieved, publicly available on the World Wide Web. Current work on open educational resources (OER) has shown the feasibility and advantages of sharing content in educational environments (Geser, 2007). Our proposal is to bring this concept of openness to the research context of TEL. We, however, go way beyond merely providing open content.

The remainder of this paper is structured as follows: at first, we survey the instruments of open science: open access, open data, open source, and open methodology. We describe their take up in TEL and present several examples of existing implementations. Afterwards, we elaborate on our vision, detailing how an open science would benefit individual researchers, public and private stakeholders, and ultimately the whole scientific community. This is followed by a discussion of the biggest issues of open science. Concluding, we present recommendations on how to overcome these issues for the implementation of an open science in TEL.

2 Instruments of open science

In our vision of an open science, we foresee four instruments. Three of them are already well-known:

1 open access as a way to make research results available

2 open data as a way to publish the raw data

3 open source as a way to give access to research prototypes.

Additionally, we introduce the concept of open methodology: it stands for sharing the methodological details of the study provided, and the tools used for data collection and analysis.

2.1 Open access

The term open access was coined by the Budapest open access initiative (BOAI, 2002): “By ‘open access’ to this literature, we mean its free availability on the public internet, permitting any users to read, download, copy, distribute, print, search, or link to the full texts of these articles, crawl them for indexing, pass them as data to software, or use them for any other lawful purpose, without financial, legal, or technical barriers other than those inseparable from gaining access to the internet itself".

Open access is central to open science; the publication is the product of the research process, presenting the results and a basic description of the method(s) used in the research process. Therefore, the publication is the starting point for most researchers looking to access, reproduce, and build upon the results of others. The directory of open access journals (DOAJ)[1] provides structured access to more than 7,200 open access journals with more than 670,000 freely available publications. Journals listed in the DOAJ have to fulfil the access criteria defined by the BOAI and all publications must be of a scientific nature, reviewed by an editor, editorial board or peer-review process. The TeLearn Open Archive is specific to TEL, containing 1,500 open access publications and 123 open access videos to date[2]. The STELLAR Open Archive[3] contains around 4,700 references from the field, some with links to the full text.

2.2 Open data

Open data means publishing the datasets collected in the research process on the World Wide Web, without restricting their use (Murray-Rust, 2008). Open data is the next logical step after open access: since the published results are in an aggregated form, raw data is needed to make these results reproducible. Besides enabling the task of reproducing research results, open data allows for the evaluation of different hypotheses on the same dataset. In that sense, open data can be understood as reusing or recycling research data without having to expend resources in another round of data collection. Open data also allows for the aggregation of data from multiple studies to gain new insights (Murray-Rust, 2008).

In the context of TEL, there are already well-known initiatives pushing the concept of open data. One of these initiatives is DataShop (Stamper et al., 2010), which aims at providing a central repository of research data and a set of tools for its analysis. Another example of the relevance of open data in TEL is the EATEL Special Interest Group dataTEL (Drachsler et al., 2010). The main objective of dataTEL is to define benchmarking criteria in order to standardise recommendation systems in TEL.

2.3 Open source

Open Source means that software is made available under a licence that permits anyone to use, change, improve, or derive from existing source code, and sometimes even to distribute the software (Feller and Fitzgerald, 2002). In traditional software development, those rights are reserved to the copyright holder of the software. Open Source Software is often developed in public, using open platforms like SourceForge[4] or GitHub[5] to support the collaborative development.

A general example of open source in TEL is the large amount of open source virtual learning environments (VLEs) used in academic environments. In this area, the most prominent example is Moodle (Dougiamas and Taylor, 2003), due to its broad adoption in educational institutions. Some specific cases are APOSDLE[6], a European TEL project in the area of professional learning that went open source, and the development of modules to support educational specifications in the VLE.LRN[7] (del Cid et al., 2007).

2.4 Open methodology

Knorr-Cetina (1981) already showed that papers do not contain all the methodological information needed to reproduce a certain research result. An elimination process is taking place during the production of the paper in which information is decontextualised and typified. Scientific work is usually done in a different way than it is reported on later. This is also backed by Latour and Woolgar (1979) who found that science is not a structured process but rather an array of incoherent observations, which need to be ordered subsequently. Furthermore, there are certain procedural remarks that are too detailed to be included in a publication. The way that these procedural remarks look like differs greatly depending on the method used. Currently, these procedural details are mostly exchanged through personal communication or joint work.

Open methodology aims at bridging this gap between raw data and final results. The term open methodology might seem unnecessary because sharing the methodology that was followed in a research work is part of the purpose of a scientific publication. But open methodology goes beyond of what is described in the paper. It encompasses such things as experimental setups, scripts written for computer simulations, and aggregation rules in qualitative data analysis. Open methodology improves the way in which the methodology is shared so that the paper is complemented with an explicit and detailed procedure on how to analyse the data collected and to generate the obtained results.

Certainly, there is the problem of tacit knowledge (Polanyi, 1983) that cannot be communicated in a codified way. In the context of computer-supported evaluations though, every step of the evaluation processed by a computer can potentially be recorded. These steps can then be packed into self-executing files, similar to application setup files. Another researcher would be able to run the self-executing file later on, and could follow the evaluation as the steps are reproduced on his computer. In some cases, this would only include the data analysis, e.g. when interviews or experiments were carried out. In other cases, such as simulations, this could involve data collection as well.

As a basic example, we will analyse the paper 'Interactive teaching tools for spatial sampling' by Bowman et al. (2010) from the *Journal of Statistical Software* (*JSTAT*). The paper describes software that is designed to convey the concepts involved in the statistical analysis of data that is measured over a spatial region. It then presents two scenarios in which this software can be used in teaching. All articles in JSTAT are published open access on their website[8]. The source code and the data that was used for producing the results are available for download alongside with the paper. Readers and reviewers are able to download the software package 'rpanel' and use it with the Open Source Software programme R[10] to follow the scenarios described in the paper. Furthermore, due to the open source license associated with the source code, readers are able to build upon this software to incorporate further examples (as suggested in the paper). However, there are certain obstacles to the reproduction of the results. If one is not familiar with R, there is no information on how to install and load the 'rpanel' package in the R software. Furthermore, the exact function calls that were used to produce the examples are not provided in the paper.

While these bits of information are too detailed for an article they are vital to the reproduction of results. Lack of this information makes it almost impossible for someone unfamiliar with R to reproduce the teaching exercise described in the paper, and leaves everyone else guessing what the actual steps were. This simple example can be applied to many other types of empirical research. How can one reproduce results from a large dataset if detailed instructions for data pre-processing are missing? How does one evaluate qualitative research, if the exact aggregation rules are unknown? How can computer simulations be compared if the exact parameters are not given?

There are already certain approaches, which contribute to an open methodology. The Stanford Exploration project (SEP) is pioneer work in the field of reproducible research and has its focus on reproducing numerical results in publications (Schwab et al., 2000). Authors are required to document the creation of numerical results from their input data. Alongside with this documentation, authors have to provide make files (input files for the UNIX utility make that serve as database for reproducing the figures containing numerical results). Fomel and Hennenfent (2007) propose to use SCons (Dubois et al., 2003) in conjunction with Madagascar[11] as a framework for reproducible research in scientific computing. In their view, the make utility is not universal enough for a broad support of reproducibility. myExperiment (Goble and De Roure, 2007) allows researchers to openly document their workflows which can be used to let others know about their setup. SAHARA Labs[12] is Open Source Software that enables remote access to lab equipment. Since all experiments and the respective setups can be recorded, this would allow sharing the experiment itself rather than only the details. Currently, remote labs are more geared towards students, but there are efforts to adapt them for researchers as well. Osentoski et al. (2010) report on a shared lab package that can be used for reproducible, shared experimentation with robots.

3 An open science for TEL

Price (1963) used the term ‘invisible colleges’ to describe communities of a few hundred scholars that work closely together in a field. These scholars always go to the same conferences, work in the same projects, and exchange staff amongst them. Thus, they form a tightly knit community that transfers knowledge through joint work rather than written artefacts (i.e., papers). This trend can also be witnessed within the research domain of TEL. TEL engages in supporting learning with adequate technologies (Dror, 2008). While the term itself is rather new, TEL-related topics have been researched in various communities for quite some time, e.g. distance education, instructional design, educational technology, and computer-supported collaborative learning. TEL aims at unifying these fields; as a result, it draws on various disciplines, such as pedagogy, computer science, psychology, cognitive science, and social science. Analysis of the field, however, shows that the disciplines do not work together in an interdisciplinary sense, but rather form disjoint scientific communities (Gillet et al., 2009; Reinhardt et al., 2011).

This is empirically backed by bibliometric and scientometric studies. Kirby et al. (2005) performed a citation analysis of subfields of instructional design and the learning sciences on the basis of three journals, one magazine, and two conferences from 1991 to 2001. They found that only 2.5% of the authors have published in both fields, and only 0.5% have published in peer-reviewed journals from either field. The average cross-field citation rate is equally low with just 0.4%–0.5%. Maurer and Khan (2010) conducted a scientometric analysis for five journals and two conferences. In line with Kirby et al. they find almost no overlap between authors in journals and in conferences.

In our vision, the adoption of open science methodologies provides a set of solid and sustainable ways to connect the disjoint communities in TEL. One of the most relevant instruments of open science for this purpose is open access; it allows for an exchange of research findings, thus providing a sense of awareness about the subjects that the community is currently researching. Open source can also act as a connector of research groups. By releasing and using Open source tools, the involved groups establish discussions on implementation issues as well as methods for improving the tool. Consequently, open source provides a level of awareness about the technical work carried out by the community, and it also provides a scaffold for future software implementations. With the adoption of open data and open methodology, we envision a more collaborative approach to TEL research. The main conversation threads that can arise from these instruments are discussing approaches to collect and analyse the available data, and the exchange of research results found on a certain set of data.

Apart from connecting researchers better, we envision that researchers in TEL would be able to reproduce the results from any paper using the instruments of open science. Standardised formats would enable them to compare the results to their own, and build upon each other's achievements. An example would be a research group working on the description of a project proposal, where one of the main points is the list of objectives that the project is aiming to achieve. In an open science environment, researchers could refer to finished projects that provide reproducible and comparable information about the results achieved during that project. Thus, researchers proposing a new project can build on and improve the results of the previous project. This in turn reduces redundant design and development, making TEL research more efficient. With the comparability of research results, it becomes more easily to determine which approach fits best to a given problem, increasing the effectiveness of TEL research.

If research prototypes are more widely available, they can potentially be used in practice much earlier, generating a greater impact on practice and more visibility for TEL research. Thus, an open science would not only benefit the researchers but also public and private stakeholders in applied science. In traditional science, teachers may have access to TEL research publications, but to get their hands on research prototypes, they must be part of a project or trial. Otherwise they have to wait until the respective research prototype becomes a product, which usually takes years to complete. With an open science, the time between research and application in practice can be reduced significantly: teachers would be able to download and reproduce the proposed examples, and take them directly to the classroom. Furthermore, they could contribute to improve the prototype by giving feedback and/or committing source code to the Open Source repository. In that way, both sides win.

Open science would also be benefit open innovation[13]. The idea of open innovation is to include external knowledge in the organisational innovation process (Chesbrough, 2003). Open Source Software created in TEL can be used to initiate open innovation in TEL stakeholder businesses (Gassmann and Enkel, 2004). One prime example is Moodle, which came out of a PhD research project (Dougiamas and Taylor, 2003). Moodle has since spawned an international commercial network, including hosting, consulting, and courseware developers[14]. Similarly, open access has the potential to contribute to open innovation processes in companies, and open data and open methodology could help to improve research in R&D departments of companies.

4 Issues of open science

There are several issues that must be addressed on the way to an open science in TEL. For one, there are legal, technical, and practical problems that need to be solved. This includes guidelines for publishing additional material, standardised formats, appropriate licenses, and proper citation methods. Nevertheless, when it comes to open science, the main issues are of a social nature. A recent survey amongst computer scientists (Stodden, 2010) shows that the largest barriers to sharing data and code are

1 the time it takes to clean up and document for release

2 the possibility that code/data may be used without citation

3 legal barriers, such as copyright

4 the potential loss of future publications using this code/data

5 that competitors may get an advantage.

There is the possibility of getting scooped when being open about one's own research. This is not so much an issue when data and source are not published before the paper.

Nonetheless, some researchers view their source code as an investment, which they do not want to give away for free; research also involves competition. David (2004) includes a summary of the economic virtues and vices of copyright protection in the context of open science; being its main virtue to provide an incentive for creative publications. On the upside, openness can lead to new collaborations beyond the usual suspects of cooperation partners (Waldrop, 2008). Furthermore, being open can help to claim priority on new ideas or discoveries.

In the case of source code, potential questions from users and patents/IP constraints play an important role, whereas in the case of data, privacy issues are ranked highly (Stodden, 2010). Privacy is a very delicate matter that needs the outmost attention. In the case of TEL, protecting the privacy of study participants is a requirement and implies that every dataset must be anonymised before its publication. Several strategies for anonymisation have emerged (c.f., e.g., Zhou et al., 2008), but in this complex field, no general solution can be applied. Furthermore, it is possible to de-anonymise information based on background information that is usually not taken into account during an anonymisation process (Nayaranan and Shamtikov, 2008).

Lastly, researchers are concerned about their reputation. When others find errors in their source code, data, or methodology, this could have damaging effects on their standing. In analogy of the open source movement, however, this could also be an opportunity for them, fixing errors that would otherwise have been repeated in successive studies. Correcting errors in someone else's code is not seen as a revelation of incompetence but rather as a collaborative act of creating something better together. Furthermore, in cases of large-scale errors, detection happens with or without an open science (Dalton, 2002). The issue of reputation is highly related to the academic reward system; current academic systems place new journals, including open access ones, at a disadvantaged place (Björk, 2004). Nevertheless, research into citations has given indication that open access papers are cited earlier and more often than their non-open access counterparts (Eysenbach, 2006).

5 Recommendations for implementing an open science in TEL

Open science is first and foremost a community effort. In fact we are arguing that reproducibility and comparability should become two of the standard criteria that every reviewer has to judge when assessing a paper. In addition to relevance, novelty, significance, and soundness, reviewers would have to assess both

1 to which degree they were able to reproduce the results

2 to which degree the results can be compared to previous efforts.

These two criteria should be of equal importance as the established criteria, giving incentive to the authors to actually apply the instruments of open science.

In addition, journals and conferences ought to make the submission of source code, data, and methodological descriptions together with the paper mandatory for them to be published. Conferences and journals themselves should in turn commit to making the papers openly accessible. The case of the genetic sequence database GenBank[15], which stores DNA sequences and makes them available to the public, has shown that if publishers and conference organisers adopt new standards, they can be propagated quickly within the community. The huge success of GenBank is due to the fact that many journals adopted the Bermuda principles (Marshall, 2001), which state among other things that DNA sequences should be rapidly released into the public domain. The recent popularity of open access and the dataTEL initiative (Verbert et al., 2011) is an indicator that these standards will meet a community that already has a certain interest in making its research more open.

With regard to the more technical problems, we propose to review existing initiatives, e.g., DataCite (Brase, 2009) for data citations, or the proposal for reproducible research standard in the computational sciences by Stodden (2009). However, there remains another challenge that we have not touched upon: awareness about the existence of tools, methods and data. It is not sufficient to only provide free repositories; for the success of an open science initiative, it is crucial to make the research community aware about the existence of such repositories, and to support researchers in use and re-use of the available data and other artefacts.

Even though these problems represent rather complex challenges, we are confident that they can be overcome on our way to an open science in TEL. The wide application of open science will help making TEL research more understandable and usable both in our everyday research and in educational practice.

Acknowledgements

The authors thank three anonymous reviewers for their valuable comments and *IJTEL* for providing the opportunity of a Young Researchers Special Issue. This contribution is partly funded by the Know-Center, which is funded within the Austrian COMET programme – Competence Centres for Excellent Technologies – under the auspices of the Austrian Federal Ministry of Transport, Innovation and Technology, the Austrian Federal Ministry of Economy, Family and Youth, and the State of Styria. COMET is managed by the Austrian Research Promotion Agency FFG.

References

Björk, B-C. (2004) 'Open access to scientific publications – an analysis of the barriers to change?', *Information Research*, Vol. 9, No. 2, available at http://InformationR.net/ir/9-2/paper170.html (accessed on 14 December 2011).

BOAI (2002) *Budapest Open Access Initiative*, available at http://www.soros.org/openaccess/read.shtml (accessed on 29 April 2011).

Bowman, A.W. et al. (2010) 'Interactive teaching tools for spatial sampling', *Journal of Statistical Software*, Vol. 36, No. 13, pp.1–17.

Brase, J. (2009) 'DataCite – a global registration agency for research data', in *Fourth International Conference on Cooperation and Promotion of Information Resources in Science and Technology*, IEEE, pp.257–261.

Chesbrough, H.W. (2003) *Open Innovation: The New Imperative for Creating and Profiting from Technology*, Harvard Business School Press, Boston.

Dalton, R. (2002) 'Misconduct: the stars who fell to earth', *Nature*, Vol. 420, No. 6917, pp.728–729.

David, P.A. (2004) 'Can "open science" be protected from the evolving regime of IPR protections?', *Journal of Institutional and Theoretical Economics JITE*, Vol. 160, No. 1, pp.9–34.

del Cid, J.P.E. et al. (2007) 'Implementation of a learning design run-time environment for the. LRN learning management', *Journal of Interactive Media in Education, Special Issue: Adaptation and IMS Learning Design*, Special Issue: Adaption and IMS Learning Design, available at http://jime.open.ac.uk/jime/article/viewArticle/2007-7/323 (accessed on 4 January 2012).

Dougiamas, M. and Taylor, P. (2003) 'Moodle: using learning communities to create an open source course management system', in D. Lassner and C. McNaught (Eds.): *Proceedings of World Conference on Educational Multimedia, Hypermedia and Telecommunications 2003*, AACE, Honolulu, Hawaii, USA, pp.171–178.

Drachsler, H. et al. (2010) 'Issues and considerations regarding sharable data sets for recommender systems in technology enhanced learning', *Procedia Computer Science*, Vol. 1, No. 2, pp.2849–2858.

Dror, I.E. (2008) 'Technology enhanced learning: the good, the bad, and the ugly', *Pragmatics & Cognition*, Vol. 2, No. 16, pp.215–223.

Dubois, P.F., Epperly, T. and Kumfert, G. (2003) 'Why Johnny can't build [portable scientific software]', *Computing in Science & Engineering*, Vol. 5, No. 5, p.83–88.

Eysenbach, G. (2006) 'Citation advantage of open access articles', *PLoS Biology*, Vol. 4, No. 5, pp.692–697.

Feller, J. and Fitzgerald, B. (2002) *Understanding Open Source Software Development*, Addison-Wesley, Boston.

Fomel, S. and Hennenfent, G. (2007) 'Reproducible computational experiments using SCons', in *IEEE International Conference on Acoustics Speech and Signal Processing*, IEEE Computer Society, pp.IV-1257–IV-1260.

Gassmann, O. and Enkel, E. (2004) 'Towards a theory of open innovation: three core process archetypes', in *R&D Management Conference*.

Geser, G. (2007) *Open Educational Practices and Resources: The OlcOs Roadmap 2012*, Revista de Universidad y Sociedad del Conocimiento, Vol. 4, No. 1, pp.1–9.

Gezelter, D. (2009) 'What, exactly, is open science? The open science project', available at http://www.openscience.org/blog/?p=269 (accessed on 27 April 2011).

Gillet, D., Scott, P. and Sutherland, R. (2009) 'STELLAR European research network of excellence in technology enhanced learning', in *International Conference on Engineering Education & Research*.

Goble, C. and De Roure, D.C. (2007) 'myExperiment: social networking for workflow-using e-scientists', in *Proceedings of the 2nd Workshop on Workflows in Support of Large-Scale Science*, ACM, pp.1–2.

Henry, N., Goodell, H., Elmqvist, N. and Fekete, J-D. (2007) '20 years of four HCI conferences: a visual exploration', *Intl. Journal of Human-Computer Interaction*, Vol. 23, No. 3. pp.239–285.

Kirby, J., Hoadley, C. and Carr-Chellman, A. (2005) 'Instructional systems design and the learning sciences: a citation analysis', *Educational Technology Research and Development*, Vol. 53, No. 1, pp.37–48.

Knorr-Cetina, K. (1981) *The Manufacture of Knowledge: An Essay on the Constructivist and Contextual Nature of Science*, Pergamon Press, Oxford.

Larsen, P.O. and von Ins, M. (2010) 'The rate of growth in scientific publication and the decline in coverage provided by science citation index', *Scientometrics*, Vol. 84, No. 3, pp.575–603.

Latour, B. and Woolgar, S. (1979) *Laboratory Life: The Construction of Scientific Facts*, Sage Publications, Beverly Hills.

Marshall, E. (2001) 'Bermuda rules: community spirit, with teeth', *Science*, Vol. 291, No. 5507, pp.1192–1192.

Maurer, H. and Khan, M.S. (2010) 'Research trends in the field of e-learning from 2003 to 2008: a scientometric and content analysis for selected journals and conferences using visualization', *Interactive Technology and Smart Education*, Vol. 7, No. 1, pp.5–18.

Murray-Rust, P. (2008) 'Open data in science', *Serials Review*, Vol. 34, No. 1, pp.52–64.

Narayanan, A. and Shmatikov, V. (2008) 'Robust de-anonymization of large sparse datasets in security and privacy', *SP 2008. IEEE Symposium on. IEEE*, pp.111–125.

National Science Board (2010) *Science and Engineering Labor Force*, National Science Foundation, Arlington.

Nielsen, M. (2008) 'The future of science', Michael Nielsen Blog, available at http://michaelnielsen.org/blog/the-future-of-science-2/ (accessed on 27 April 2011).

Osentoski, S. et al. (2010) 'Brown ROS package: reproducibility for shared experimentation and learning from demonstration', in *AAAI-10 Robot Workshop*.

Polanyi, M. (1983) *The Tacit Dimension*, Peter Smith Gloucester, MA.

Price, D.J.D.S. (1963) *Little Science, Big Science*, Columbia University Press, New York.

Reinhardt, W., Meier, C., Drachsler, H. and Sloep, P. (2011) 'Analyzing 5 years of EC-TEL proceedings', in C.D. Kloos, D. Gillet, R.M.C. Garcìa, F. Wild and M. Wolpers (Eds.): *Towards Ubiquitous Learning. Proceedings of the 6th European Conference on Technology Enhanced Learning*, pp 531–536, Springer Berlin/Heidelberg.

Schwab, M., Karrenbach, N. and Claerbout, J. (2000) 'Making scientific computations reproducible', *Computing in Science & Engineering*, Vol. 2, No. 6, pp.61–67.

Stamper, J. et al. (2010) 'PSLC DataShop: a data analysis service for the learning science community', in V. Aleven, J. Kay and J. Mostow (Eds.): *Intelligent Tutoring Systems*, p.455, Springer Berlin/Heidelberg.

Stodden, V. (2009) 'The legal framework for reproducible scientific research: licensing and copyright', *Computing in Science & Engineering*, Vol. 11, No. 1, pp.35–40.

Stodden, V. (2010) 'The scientific method in practice: reproducibility in the computational sciences', Sloan School Working Paper, 4773-10, MIT, available at http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1550193 (accessed on 10 April 2011).

Verbert, K. et al. (2011) 'Dataset-driven research for improving recommender systems for learning', in *1st International Conference on Learning Analytics & Knowledge*, Banff, Canada.

Waldrop, M. (2008) 'Science 2.0', *Scientific American*, Vol. 298, No. 5, pp.46–51.

Young, N.S., Ioannidis, J.P. and Al-Ubaydli, O. (2008) 'Why current publication practices may distort science', *PLoS medicine*, Vol. 5, No. 10, p.e201.

Zhou, B., Pei, J. and Luk, W. (2008) 'A brief survey on anonymization techniques for privacy preserving publishing of social network data', *ACM SIGKDD Explorations Newsletter*, Vol. 10, No. 2.

Notes

1 http://doaj.org.
2 13/11/2011.
3 http://oa.stellarnet.eu.
4 http://www.sourceforge.net.
5 https://github.com/.
6 http://www.aposdle.org.
7 http://dotlrn.org/.
8 http://www.jstatsoft.org.
9 http://www.jstatsoft.org/v36/i13.
10 http://www.r-project.org/.
11 http://www.ahay.org/wiki/Main_Page.
12 http://sourceforge.net/projects/labshare-sahara/.
13 We thank an anonymous reviewer for pointing out the connection to open innovation to us.
14 See http://moodle.com for an overview.
15 http://www.ncbi.nlm.nih.gov/genbank/.