

Measurement of Open Access as an Infrastructural Challenge: The Case of Finland

Pekka OLSBO¹

Open Science Centre, University of Jyväskylä

Abstract. Finland has set numeric goals for the development of open access. However, at the moment, no system is available by which this development could be monitored. Poor quality in the metadata records in universities' research information databases prevents metadata-based analysis of open access publishing progress. This paper shows how the quality problems of Finnish publication data can be resolved through centralizing the services and processes of metadata creation and by improving the interoperability of systems involved in the processes. As a result, this study describes an environment where reliable measurement of open access is possible and presents suggested actions for improving the Finnish publication data collection.

Keywords. open access, institutional repositories, current research information systems, metadata, measurement of open access

1. Introduction

Finland has set goals for achieving open science, following the requirements and recommendations defined by the EU commission [1]. The Finnish Ministry of Education and Culture has established key performance indicators for open science, including targets for open access publishing. According to these indicators, 65% of Finnish scientific publications should be open access by 2017. The percentage should be 75% in 2018, and by the year 2020, 90% of publications should be open access [2].

To achieve these goals, measurable performance, often in the form of numeric data, is required. However, such an outcome means that those overseeing the process would need to base their evaluation on reliable metadata records. Finland, however, does not currently have a national Current Research Information System (CRIS). There are, however, research information systems in each university in Finland. The Finnish research publications portal JUULI collects the metadata created and maintained in the research information systems of the universities [3]. This means that the metadata in the Finnish portal JUULI can be trusted only insofar as the quality of metadata in the research databases of the universities can be trusted.

While the poor quality of the metadata created in Finnish research institutes regarding all publications presents a serious problem [4], establishing a reliable measurement of open access is even more challenging. One element of this challenge

¹ Corresponding Author, P.O.BOX 35, 40014 University of Jyväskylä, Finland; E-mail: pekka.olsbo@juyu.fi.

results from the current Finnish publication data collection instructions that do not require the open access status of a publication as mandatory information

The metadata collected from the year 2015's publications demonstrate the problems in quality. That year, Finnish universities reported 33,720 publications in the Finnish JUULI portal. Table 1 shows that the metadata regarding the open availability of the publications are unreported in almost 50% of cases. On the basis of this information, the only conclusion that can be made with this metadata is that the share of open access publications in Finnish universities is not measurable.

Table 1. The open access status of Finnish research publications in JUULI portal (www.juuli.fi) 2015. Data collected in January 2017.

Category	Amount	%
Not OA	10639	31 %
Gold OA	5321	16 %
Other OA (Green)	1639	5 %
No answer	16121	48 %
Total	33720	

From the year 2016 onward, the division of open access was changed from the previous collection guidelines. Now data is also collected in the category of hybrid open access publications, which makes the definition of the open access status even more demanding.

2. The Challenge of the Quality of the Metadata

The key element in making the measurement of open access publishing reliable is a clear definition of open access. Archambault et al. [5] defined open access as divided into the rational definitions of open access and the operational definitions of open access. More specifically, the rational definitions used in Archambault et al. [5] are based on the propositions made by Peter Suber [6]. The operational definitions of open access refer to definitions established by the organization gathering and evaluating the measurement data of open access. (pp 2-5).

In the case of Finland, the operational definitions of open access are described in the document, Publication Data Collection Instructions for Researchers 2016 [7]. These definitions divide open access to three categories:

1. publications published via an open access channel,
2. open access publications published within a hybrid channel, and
3. publications self-archived in the publication archives of an organization or field of science, whether immediately available or after a reasonable embargo specified by the publisher

This division places a great demand on the quality of the metadata created about the publications of Finnish scholarly authors. In other words, each research publication created in Finland would need to be categorized reliably within the designated categories. Such a requirement for reliably measuring open access causes challenge number one: the quality of the metadata.

The challenge of the quality of the metadata also includes the problem of coverage. In order to calculate the share of open access publications, the total number of publications and total number of open access publications per year in Finland must be known. However, because the exact number of open access articles each year is not accurately collected, only estimates can be made about the open availability of research [8], [9].

The second challenge in measuring open access is quantifying the progress of open access. Archambault et al. [5] described why it is quite difficult to measure the growth of open access (OA):

The reason is that growth in OA appears as the result of four main forces: (1) historical growth in the interest in OA which translates into new papers being increasingly available for free; (2) the growing interest in OA also translates into actors increasingly making available old papers for free; (3) OA policies that allow for delaying OA to scientific papers with embargo periods produce a concomitant disembargoing of scientific articles that creates additional growth in old papers being made available for free; and (4) the fact that the number of published scientific papers is growing, so even for a stable proportion of OA, the number of OA papers would keep growing (p. ii).

As noted earlier, the key performance indicators for open science in Finland are based on the share of open access publications. This makes Forces 1, 2 and 3 as described by Archambault et al. [5] relevant, thus exerting additional pressure on the need for quality metadata and the interoperability of the systems engaged in reporting processes.

The problems in metadata records generally have been categorized by Yasser [10] and further studied by Tani et al. [11]. From the five categories presented by Yasser [10], three can be applied to Finnish metadata:

(1) incorrect values, i.e., metadata records contain values that do not represent a given resource correctly even though elements are applied correctly, (2) missing information, i.e., the metadata record is not complete and (3) information loss, i.e., some details characterizing the information are lost due to the conversion of metadata from one scheme to another or due the fact that metadata is not extracted from one system to another (pp. 59-60).

All these problems can be solved by improving the processes of metadata creation and/or the interoperability of the systems involved in the collection process. As Nicholas Joint [12] remarked, the more often libraries become the mediators of green open access, the higher the chance of quality metadata being produced. This is one reason why the entire process of metadata creation and maintenance should be centralized.

3. The Progress of Open Access in Finland, 2012–2016

The publication data from Finnish universities are gathered within the national JUULI portal. However, the metadata in the JUULI portal currently does not support a high-quality analysis of the open access availability of research publications. Nevertheless, by using the metadata, I could speculate on the progress of open access in Finland and analyze the metadata problems presented by Yasser [10] in greater detail. This is

especially true when looking more closely at the development of green open access in Finland.

Finland's Ministry of Education and Culture launched the Open Science and Research Initiative [13] to both promote research information availability and to serve as the open science platform for the years 2014-2017. However, despite this goal at the national level, neither specific actions nor decisions resulting in open access publishing progress have taken place. Even at the university level, where many universities operate under a mandate for open access publishing, only a few universities are making a difference in the growth of open access. That is why, in the following analysis, I focus on describing the development in four universities in Finland who are contributing the most to this growth and for which mostly complete data is available. These include: Aalto University, the University of Helsinki, the University of Jyväskylä, and the University of Tampere.

Table 2 presents the number of open access articles in the four case universities for the years 2012–2015. These figures include both green and gold open access. The table indicates that measurable progress in open access publishing has occurred only at the Universities of Jyväskylä and Helsinki. Table 3 describes the total number of article publications from these case universities.

Table 2. The number of open access articles in four case universities. Data from the JUULI portal (www.juuli.fi).

University	2012	2013	2014	2015
Aalto	586	564	556	597
Helsinki	1020	1410	1691	1651
Jyväskylä	722	757	1114	1276
Tampere	843	714	934	767

Table 3. The total number of article publications in four case universities 2012–2015. Data from the JUULI portal (www.juuli.fi).

University	2012	2013	2014	2015
Aalto	3587	3769	3579	3360
Helsinki	10858	10867	10800	9777
Jyväskylä	2690	2847	2793	2738
Tampere	2416	2252	2430	2363

By dividing the number of open access publications by the total number of article publications per year, we can create a chart that depicts the development of open access at each case university (see Figure 1). The main question that this graph raises, however, is how reliable are these numbers? Can anything really be said about the development of open access in Finland, or even these universities, from these numbers?

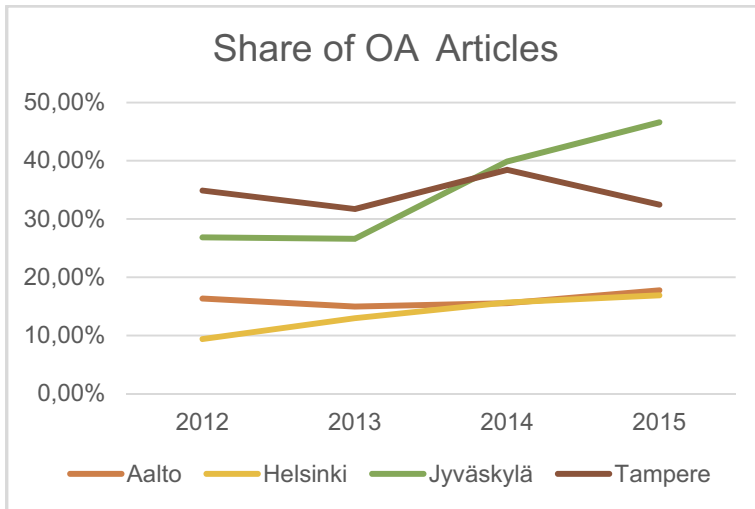


Figure 1. Share of open access articles in four case universities 2012–2015. Data from the JUULI portal (www.juuli.fi).

The reliability of metadata in the JUULI portal can be tested more accurately by looking closely at the development of green open access in these four case universities. During 2012–2015, green open access was categorized in the JUULI portal as “other OA availability.” Table 4 shows the number of green open access articles in the four case universities for the years 2012–2015.

Table 4. The number of green open access articles in four case universities 2012–2015. Data from JUULI portal (www.juuli.fi).

University	2012	2013	2014	2015
Aalto	12	8	15	24
Helsinki	157	440	448	549
Jyväskylä	276	279	568	664
Tampere	77	51	179	178

Figure 2 shows the share of green open access articles in each case university. However, the share of green open access is very modest in three of the universities, with the exception being the University of Jyväskylä, which shows rapid development in recent years.

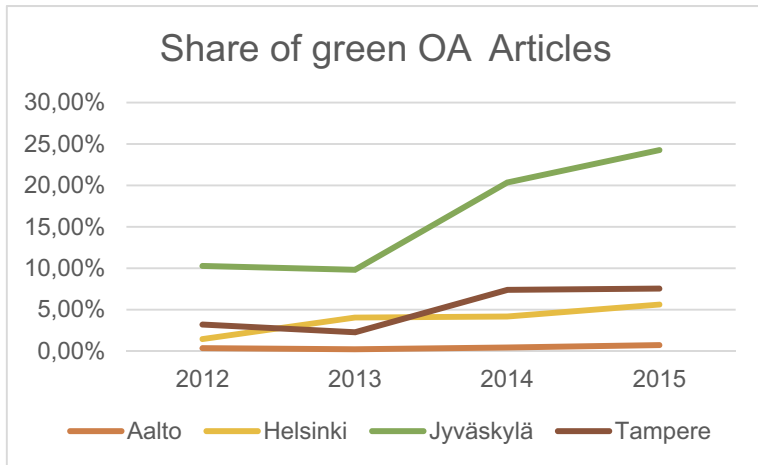


Figure 2. The share of green open access articles in four case universities 2012–2015. Data from the JUULI portal (www.juuli.fi).

In order to test the problems in metadata records categorized by Yasser [10] and Tani et al. [11], the original sources of the metadata must be used. In the case of green open access, this means the institutional repositories and the current research information systems of the case universities. Because both the repositories and the CRISs include data from the year 2016 as well, the analysis, as presented in the figure and tables, can be expanded to cover this additional data.

The data available in the institutional repositories of the universities regarding the number of deposited articles differ significantly from similar data available in the JUULI portal. Table 5 shows the number of deposited articles from the four case universities between 2012 and 2016. This comparison between Tables 4 and 5 demonstrates that no conclusions about the development of open access can be drawn using the data from the JUULI portal.

Table 5. The number of green open access articles in the repositories of four case universities 2012–2016. Data collected from the repositories Aaltodoc [15], Helda [16], JYX [17] and TamPub [18] on February 20, 2017.

University	2012	2013	2014	2015	2016
Aalto	134	162	249	277	448
Helsinki	586	806	978	1166	1598
Jyväskylä	265	371	642	889	1057
Tampere	207	154	236	266	333

I now analyze the problems with the metadata records in the JUULI portal using the three categories as developed by Yasser [10]:

1. *Incorrect values, i.e., metadata records contain values that do not represent a given resource correctly, even though elements are applied correctly.* In comparing the data in the JUULI portal to the original repositories and CRISs for the four case universities, information about the green open access is incorrect or missing in the JUULI portal in a large number of the metadata records. The inaccuracy of a value can be a result of human behavior or a lack of interoperability between the repository and the CRIS.
2. *Missing information, i.e., the metadata record is not complete.* As presented in Table 1, the most significant problem regarding the information about the open availability of publications in the JUULI portal is missing information. Once again, this is due to two reasons: The people responsible for maintaining the CRIS do not fill in the information about green open access, or this information is not extracted from the repository into the CRIS. In some cases, there is only a small qualitative difference between missing information and information loss. Missing information may be due to information loss.
3. *Information loss, i.e., some details characterizing the information are lost due to the conversion of metadata from one scheme to another or due the fact that metadata is not extracted from one system to another.* The basic infrastructure of green open access should be quite simple: Metadata are created or imported into the CRIS; the metadata are then transferred to the repository, where they are enriched with open access information; and then the updated records are imported back into the CRIS. These metadata are then transferred to the JUULI portal. But if the interoperability between the repository and CRIS, or between the CRIS and JUULI, is incomplete, then information is in danger of being lost.

As a result of the analysis above, the causes for the problems in the metadata records reflect two main factors: the human factor and the infrastructural factor. The elimination of problems related to these factors has been one key element for the development of self-archiving and centralized open science services at the University of Jyväskylä.

4. Conclusions: The “Jyväskylä Model” in Practice

The information gathered from research information systems in Finnish universities is not reliable enough for measuring open access development. This is true for both gold and green open access. Therefore, two essential actions must be implemented:

1. Gold, green, and hybrid open access papers must be identified and cataloged within the research information systems with high-quality reliability, and
2. The interoperability of institutional repositories and current research information systems must be improved so that the complete information about the deposited papers in the repository is exported automatically and accurately to the research information system.

As part of the project Finland: A Model Country for Green Open Access, we have studied these processes in the Open Science Centre of the University of Jyväskylä. We

have created a model that results in high-quality metadata creation and automated processes between the repository and the CRIS [14]. The basic idea of the “Jyväskylä model” is to centralize all aspects of the self-archiving and open access processes lying within the responsibility of the professionals at the university library. Even within the library, just a few professionals handle the metadata input and the repository–CRIS processes. Researchers do as little as possible and, in some cases, nothing at all. The motivation for this structure is that these processes and workflows need expertise and extensive knowledge about various aspects of publishing and project work, such as the legal aspects, publisher policies, funder mandates, and requirements set by the Finnish government ministries. When experienced professionals attend to collecting the necessary data on publications by the university’s researchers, the creation of metadata is much faster and more reliable. Additionally, this system is also cost effective in that fewer people are needed to input and transfer the information.

The infrastructure of the Jyväskylä model is described in Olsbo et al. [15] and it is based on automated movement of metadata between the repository and CRIS. This structure ensures that all changes and additions to the metadata are included in both systems. Consequently, all the metadata transferred to the national JUULI portal is up to date and complete.

This model of centralized publications data gathering, particularly related to open access services, was implemented at the University of Jyväskylä in 2014. Since then, the share of green open access publications at the university has risen from 16% to 51%. Moreover, the staff resources needed for tracking all aspects of university publications, as well as for reporting the research activities to the Ministry of Education and Culture, have been reduced considerably. Figure 3 shows the development of openness of peer reviewed articles in the University of Jyväskylä 2015–2016.

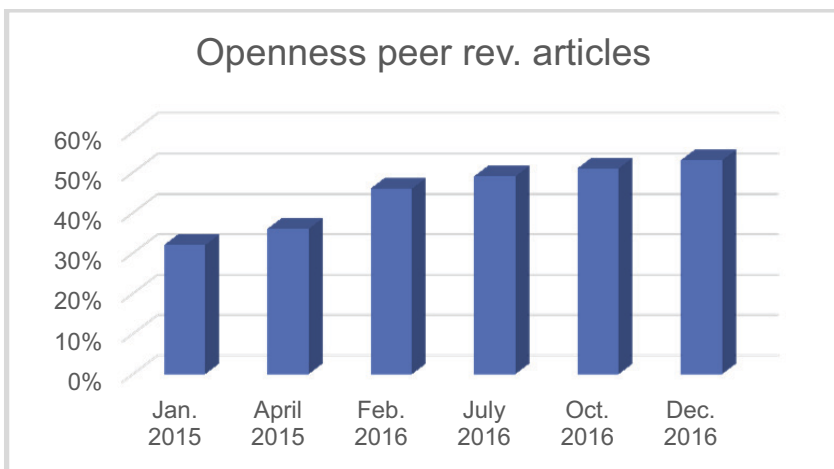


Figure 3. Development of openness of peer reviewed articles in the University of Jyväskylä 2015–2016. Data from the TUTKA research information database (<http://tutka.jyu.fi/tutka/>).

High-quality metadata and the seamless interoperability of the repository and CRIS ensure the possibility of reliably analyzing the development of open access.

Figure 3 shows the development of open access peer reviewed articles at the University of Jyväskylä for the years of 2015 and 2016.

A prototype has been developed by the University of Jyväskylä for open access monitoring that automatically analyzes the metadata in the CRIS and gathers up-to-date information about the development of open access. The model of practicing and monitoring open access created in the University of Jyväskylä can be implemented at most universities in Finland. When this model is combined with the new national VIRT A Publication Information Service and system architecture [20], it would enable the reliable analysis and assessment of the open access development in Finland.

Acknowledgements

This study was funded by the Finnish Ministry of Education and Culture's Open Science and Research initiative (see more: <http://openscience.fi/>).

References

- [1] European Commission. Commission recommendation of 17.7.2012 on access to and preservation of scientific information. Available at: http://ec.europa.eu/research/science-society/document_library/pdf_06/recommendation-access-and-preservation-scientific-information_en.pdf
- [2] Niinimäki, S. The Open Science and Research Initiative in Finland. Presentation, Annual JPI Conference, Brussels, 15.1.2016. Available at: https://www.era-learn.eu/events/annual-joint-programming-2015-new-date-2016/topic-3-strategies-for-fostering-open-knowledge-and-open-access-in-research/02_ERALEARN_2020_Niinimki.pdf
- [3] Finnish research publications portal JUULI. Available at: <http://www.juuli.fi/>
- [4] Ilva, J. Towards Reliable Data – Counting the Finnish Open Access Publications. *Procedia Computer Science*, Volume 106, 2017, pp 299-304. <http://dx.doi.org/10.1016/j.procs.2017.03.029>.
- [5] Archambault, E., Amyot, D., Deschamps, P., Nicol, A., & Provencher, F. Proportion of Open Access Papers Published in Peer-Reviewed Journals at the European and World Levels—1996–2013. Copyright, Fair Use, Scholarly Communication, etc. 2014; Paper 8. Available at: <http://digitalcommons.unl.edu/cgi/viewcontent.cgi?article=1007&context=scholcom>
- [6] Suber, P. *Open Access*. Cambridge, Mass.: MIT Press; 2012. Available at: https://mitpress.mit.edu/sites/default/files/9780262517638_Open_Access_PDF_Version.pdf
- [7] Ministry of Education and Culture. Publication data collection instructions for researchers 2016, pp. 14-15. Available at: <https://confluence.csc.fi/display/suorat/Julkaisutiedonkeruun+utkijaohjeistukset>
- [8] Bjork B., Roos A., & Lauri, M. Scientific journal publishing: yearly volume and open access availability. *Information Research*. 2009; 14(1). Available at: https://helda.helsinki.fi/bitstream/handle/10227/615/bjork_roos_lauri.pdf
- [9] Khabsa M. & Giles C.L. The number of scholarly documents on the public web. *PLoS One*. 2014; 9(5): e93949. pp 59-60. <http://dx.doi.org/10.1371/journal.pone.0093949>
- [10] Yasser, C.M. An Analysis of Problems in Metadata Records. *Journal of Library Metadata*, 2011; 11:2, pp. 51-62. DOI: 10.1080/19386389.2011.570654. Available at: <http://dx.doi.org/10.1080/19386389.2011.570654>
- [11] Tani, A., Candela, L., & Castelli, D. Dealing with metadata quality: The legacy of digital library efforts. *Information Processing & Management*. Vol. 49, Issue 6, November 2013, pp. 1194–1205. <http://dx.doi.org/10.1016/j.ipm.2013.05.003>
- [12] Joint, N. Institutional repositories, self-archiving and the role of the library. *Library Review*, 2006; Vol. 55 Iss: 2, pp.81–84. <http://dx.doi.org/10.1108/00242530610649576>
- [13] Ministry of Education and Culture. Open science and research leads to surprising discoveries and creative insights. Open science and research roadmap 2014–2017. Reports of the Ministry of Education and Culture, Finland 2014: 21. Available at: http://www.minedu.fi/OPM/Julkaisut/2014/Avoimen_tieteen_ja_tutkimuksen_tiekartta_2014_2017.htm?lang=en

- [14] Suomi rinnakkaistallentamisen mallimaaksi (SURIMA) –hanke [Finland – A model country for green open access project]. Project publications. Available at: <https://jyx.jyu.fi/dspace/handle/123456789/52527>
- [15] Olsbo, P., Hautala, T. & Kananen, J. Modelling centralized self-archiving process in Finland. Poster in Repository Fringe conference, Edinburgh, 1.8.2016. Available at: <http://urn.fi/URN:NBN:fi:jyu-201612295282>
- [16] Aaltodoc – Aalto University publication archive. Available at: <https://aaltodoc.aalto.fi/>
- [17] Helda – Digital Repository of the University of Helsinki. Available at: <https://helda.helsinki.fi/>
- [18] JYX – Jyväskylä University Digital Archive. Available at: <https://jyx.jyu.fi/>
- [19] Tampub – institutional repository of the University of Tampere. Available at: <http://tampub.uta.fi/>
- [20] VIRT A Publication Information Service. CSC – IT Center for Science Ltd; 2017. Available at: <https://confluence.csc.fi/display/VIR/Esittelymateriaaleja?preview=/50870665/68724326/VIRTA%20publication%20information%20service%20in%20Finland.pdf>