IFLA WLIC 2017 WROCŁAW

Submitted on: 01.09.2017

Satellite Meeting: Serials and Other Continuing Resources Section &
Acquisition and Collection Development
Open Access: Action Required
Date: 16-17 August 2017
Location: European Solidarity Center (ESC), Gdańsk (Poland)

Grey Literature in Institutional Repositories: A Case Study on the World Top 100 Universities

Hiroyuki Tsunoda
Tsurumi University, Yokohama, Japan.
E-mail address: tsunoda-h@tsurumi-u.ac.jp

Yuan Sun
National Institute of Informatics, Tokyo, Japan.
E-mail address: yuan@nii.ac.jp

Masaki Nishizawa
National Institute of Informatics, Tokyo, Japan.
E-mail address: nisizawa@nii.ac.jp

Xiaomin Liu
National Science Library, Chinese Academy of Sciences, Beijing, China.
E-mail address: liuxm@mail.las.ac.cn

Abstract:

An Institutional Repository (IR), according to Foster and Gibbons (2005), is an electronic system that captures, preserves, and provides access to the digital work products of a community. IRs with different types of digital content have appeared after 2000s. They aim to provide open access to institutional research output, to create global visibility for institutions' research, and to store and preserve other institutional digital assets, including unpublished or otherwise easily lost grey literature such as theses, working papers or technical reports. In this paper we take the world top 100 universities ranked in Times Higher Education World University Rankings 2015-2016 as example to investigate the status of contents provided in their IRs, focusing on grey literatures self-archiving. The data was collected from the Directory of Open Access Repositories (OpenDOAR) (www.opendoar.org), which is an authoritative directory of academic open access repositories run by the University of Nottingham. Up to January 2017, there are over 3,000 repositories in the OpenDOAR that are providing their access worldwide. We found that most of the top 100 universities have established the IRs. California Institute of Technology, University of Oxford, University of

Edinburgh, New York University, University of North Carolina at Chapel Hill were having the largest number of IRs, each of them operating six IRs. By searching IRs of the 100 universities individually, 192 repositories for the top 100 universities were identified, indicating each university has an average of 1.92 IRs. The number of contents is about 7 million, and 700 unique document types were identified in the total institutional repositories. We classified these document types into 12 types. As expected, journal articles (43%) have highest proportion, following by theses & dissertations(13%), conference & workshop papers(8%), book chapters & sections(6%), dataset(6%), multimedia & audio-visual materials(4%) and unpublished reports & working papers(4%). It is revealed that a wide variety of grey literature have been stored in institutional repositories, making them searchable and accessible for the public and research communities. This paper will especially focus on the availability of grey literature in IRs and discuss about new roles and possible futures for librarians.

Keywords: Grey Literature, Institutional Repository, Open Access, World University Rankings, White Literature

---

Introduction

A substantial part of scientific literature and documents ("documents" hereinafter) are produced in universities and research institutes which the researchers belong to. Among them, academic papers, preprints, books and patents are published, but many are not disclosed. Open access has recently established itself as a new circulation route for academic documents, either by the "gold route" in which publishers supply documents, or by the "green route" in which authors supply their own publications. An important part of the green route is institutional repositories (IRs) through which universities and colleges disclose documents produced by their researchers. IR, according to Foster and Gibbons (2005), is an electronic system that captures, preserves, and provides access to the digital work products of a community. IRs with different types of digital content have appeared after 2000s. IRs maintain documents for a prolonged period for free perusal and downloading by the public. The content provided by IRs has been more and more diversified in recent years. They aim to provide open access to institutional research output, to create global visibility for institutions' research, and to store and preserve other institutional digital assets, including unpublished or otherwise easily lost grey literature such as theses, working papers or technical reports.

Purposes

The purposes of the present work are twofold:

- Clarifying the present status of grey literature in IRs
- Establishing the correlation between university evaluation and the number of grey literature reposited per faculty member in IRs

Definitions

Grey literature is hard to define, and is often qualified by imprecise terms such as 'semi-public' or 'non-conventional' (Wood and Smith, 1993). A definition of grey literature was given by the third International Conference on Grey literature held in Luxembourg in 1997 (Aina, 2000), while the British Library categorizes "documents which are often difficult to identify and acquire through normal bookselling channels" as grey literature. In this paper, 'grey literature' refers to theses, data sets, multimedia and undisclosed documents.

In contrast, academic journals, books, preprints and patents are called 'white literature' in this work.

Method

The disclosure of IR collection was investigated for the top 100 universities according to the Times Higher Education World Universities (THEWU) 2015 – 2016. The addresses of the IRs were obtained from the Directory of Open Access Repositories (OpenDOAR) and the Registry of Open Access Repository Mandates and Policies (ROARMAP) along with Google search results. The websites of these IRs provided information on the size of the collections by document type, as well as the five indices used for the university evaluation in THEWU 2015-2016. The number of faculty members (including teaching staff and students) of the 100 universities were taken from the Quacquarelli Symonds (QS) and the websites of individual organizations. These studies were conducted between January and August 2016.

Result and Discussion

Continent of Top 100 Universities

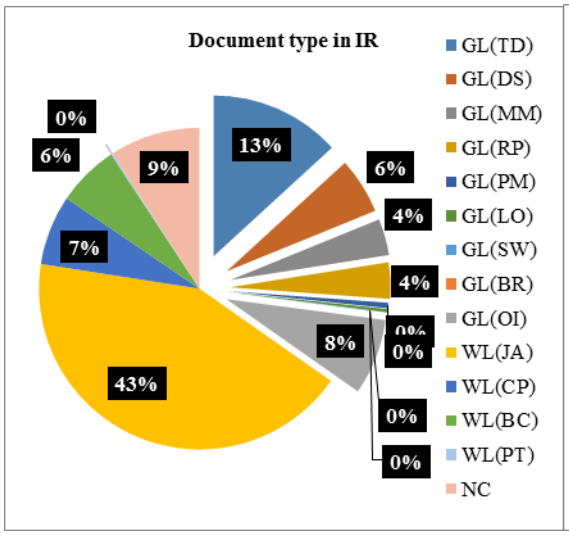
Fig. 1 shows proportion of top 100 universities by continent. Top 100 universities are unevenly distributed. Most universities are located in North America (42) and Europe (43), few in Asia (9) and Australasia (6). African universities, Caribbean universities, Central American universities, and South American universities fall out of the top 100.

Fig. 1: Proportion of Top 100 Universities by Continent - Worldwide

Current situation of IRs

Open access repositories (OAR) include: (a) IRs primarily for disclosure of research results, (b) digital archives (DA) providing digital photographs and scanned documents, and (c) digital libraries (DL) collecting electronic books. The 100 universities investigated had 1 to 6 OARs each, totaling 189 repositories. California Institute of Technology, New York University, University of Edinburgh, University of North Carolina at Chapel Hill, and University of Oxford provided six OARs each corresponding to at least one of (a), (b) and (c) defined above. The type (a) repositories were offered by 93 universities. Since the website of the IR of one of those universities was inaccessible, 92 of them were investigated. The 92 universities stored a total of 7,190,000 documents, among which 6,540,000 were accompanied by description of document types and 650,000 were not. The total of 700 document types recorded were classified into 13 categories: academic papers (JA), preprints (CP), books (BC) and patents (PT) as white literature, and theses (TD), data sets (DS), multimedia (MM), reports (RP), presentation materials (PM), teaching materials (LO), software (SW), bibliographies (BR) and others (OI) as grey literature. The total numbers of the white and grey literature items were 4,040,000 and 2,500,000, respectively.

Fig. 2 shows proportion document types in IRs and breakdown of grey literature. Fig. 3 shows the frequency of university, which stores theses and dissertations, and datasets.



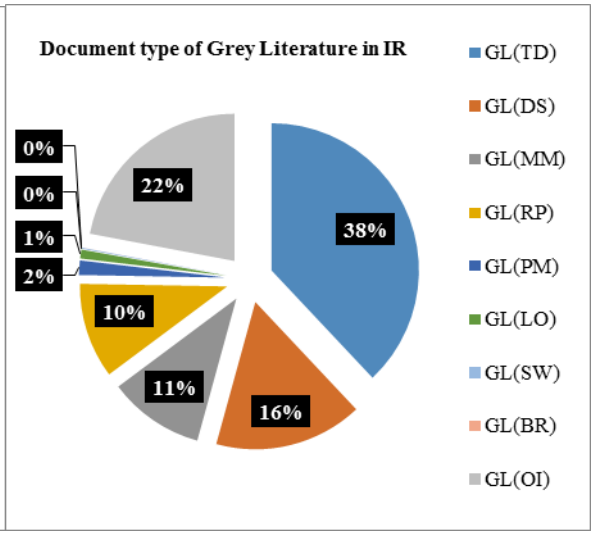
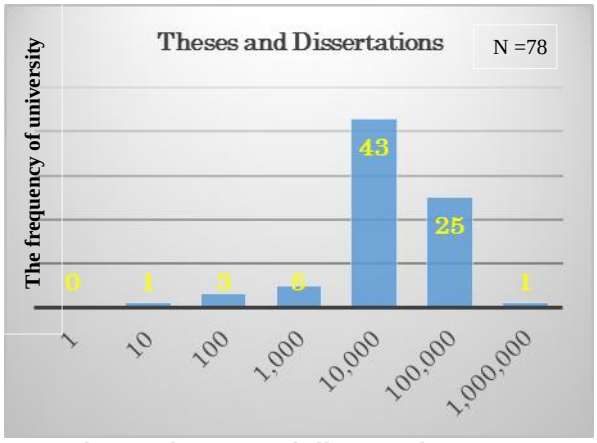


Fig. 2: Document types in IRs **Breakdown of grey literature**



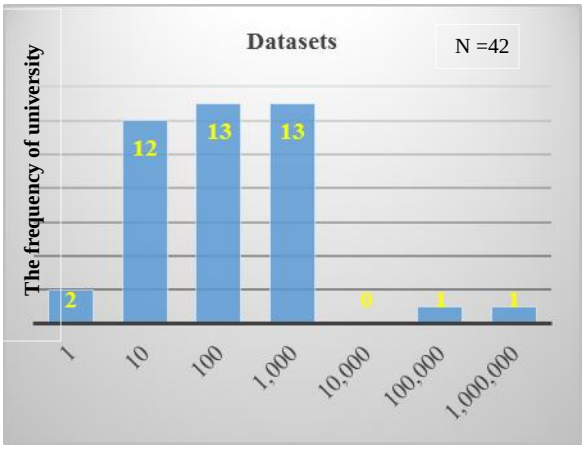


Fig. 3: Theses and dissertations **Datasets**

Grey literature in IRs

The largest share in the quantity of grey literature was represented by theses and dissertations: 950 thousand were reposited in 80 organizations. Access to theses, while having formerly been very difficult, is now far easier due to advance in disclosure through IRs chiefly with North American and European universities. Data sets were represented by 41 thousand items in 43 organizations, the majority of which was the 390 thousand of Harvard Dataverse (Harvard University); the remaining less than 15,000 were found in 42 organizations. Despite the increasing importance of open data that permits sharing basic data in open science, only three repositories were dedicated to open data: Harvard Dataverse, Edinburgh DataShare (University of Edinburgh), and Biological Magnetic Resonance Data Bank (University of Wisconsin). More open data repositories are expected to appear. Multimedia and audio-visual materials were provided by 50 organizations with 260 thousand items, including 160 thousand in Carolina Digital Repository, University of North Carolina at Chapel Hill, 30 thousand in Monash University Research Repository, 20 thousand in the KnowledgeBank at Ohio State University, and 10 thousand in the University of Toronto Research Repository. Twenty-six thousand records of unpublished reports and working papers were found in 63 organizations, including 40 thousand each in Wageningen Staff Publications (Wageningen University and Research), and in Munich RePEc Personal Archive, Ludwig-Maximilians-Universität München, and 10 thousand each in Scholarly Materials and Research@Georgia Tech (Georgia Institute of Technology) and TU Delft

Repository (Technische Universiteit Delft). Unpublished reports were distributed in a wider range than databases or multimedia. IRs had very few presentation materials, learning objects, software, and bibliographic references. Grey literature may become 'whiter' if more authors reposit more of the products in IRs. The distinction between white literature and grey literature is becoming increasingly fuzzy (Luzi, 2010, Swan, 2008 and 2011). This process has, however, not sufficiently advanced yet (Schöpfel, Prost and Le Bescond, 2012).

Distribution of grey literature in the top 100 universities

Fig. 4 shows the distribution of grey and white literature as well as documents of unknown type in individual repositories.

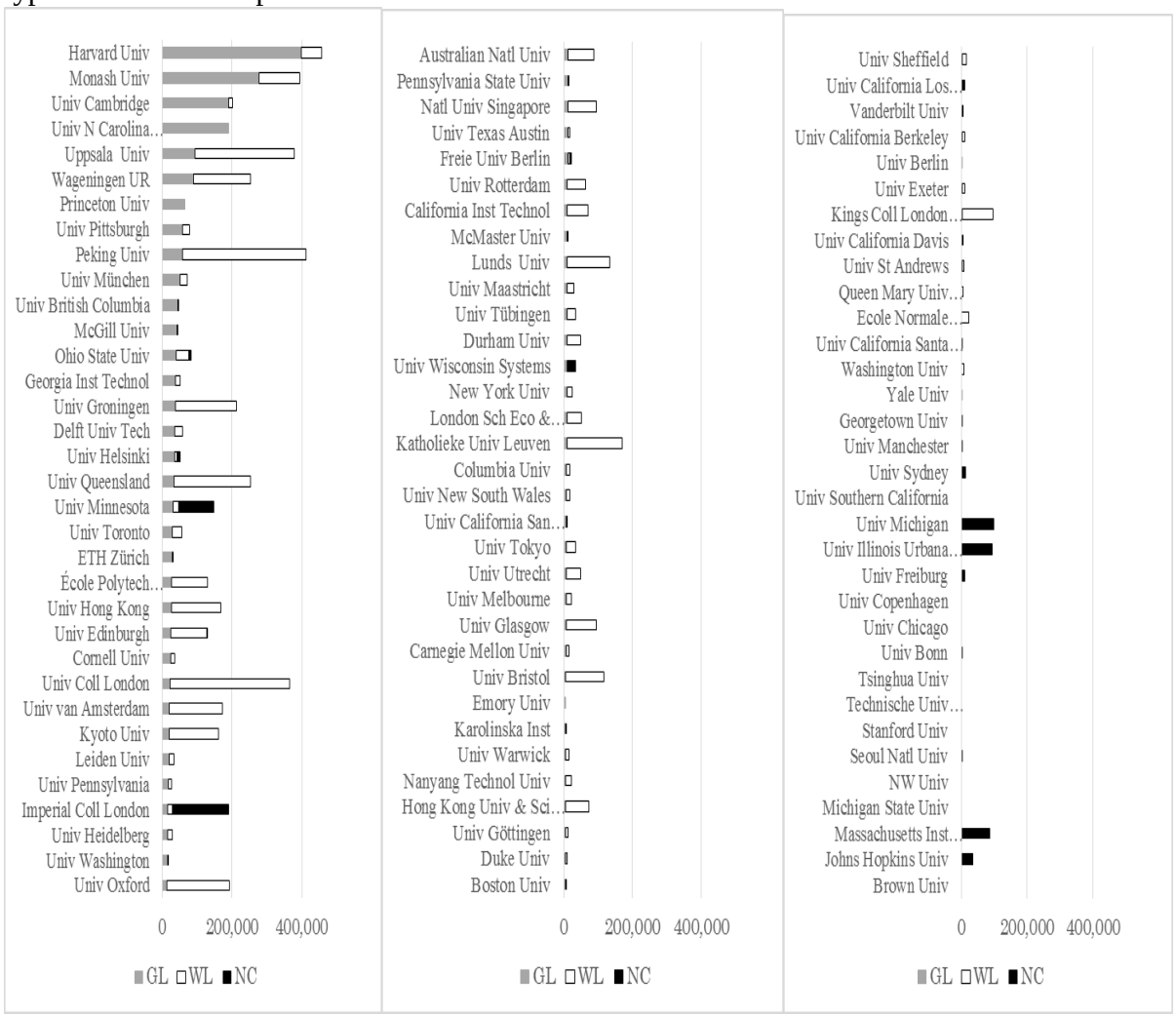


Fig. 4: Distribution of grey and white literature and documents of unknown type

The largest amount of grey literature is provided by Harvard University via three OARs: Digital Access to Scholarship at Harvard, Harvard Dataverse, and Harvard Smithsonian Digital Video Library. Harvard Dataverse houses more than 390 thousand data sets. Monash University follows with more than 120 thousand theses in Monash University Research Repository. The third-largest is University of Cambridge with five OARs: Digital Himalaya, Apollo, ESC Publication, Teaching and Learning Research Programme Publications and Computer Laboratory Technical Reports - Cambridge Univ. Apollo (formerly DSpace @ Cambridge) provides more than 170 thousand chemical structures. In contrast, a number of universities, including Peking University, University College London, University of Glasgow, University of Bristol, and King's College London, has more white literature than

grey in the repositories, possibly because (1) public grant providers often require that the outcome of the works funded be disclosed in an open access, (2) open-access documents tend to be cited more frequently than those published in book form, and (3) university evaluation prompts faculty members to place their research outputs in IRs.

Correlation with university evaluation indicators

The number of grey literature reposited in IRs per faculty member was calculated by the following formula (1).

$$\text{GL-Index} = \frac{\text{Grey literature - Thesis and Dissertation}}{\text{Faculty}} \quad (1)$$

Table 1 shows the top 8 universities by GL-Index, which is high enough to compare with other measurements.

Table 1: Top 8 Universities by GL-Index (> 10.00)

University	GL-Index
Harvard University	93.93
University of North Carolina at Chapel Hill	53.77
Monash University	46.16
University of Cambridge	36.93
Wageningen UR Corporate headquarters	32.86
University of Pittsburgh	13.40
École Polytechnique Fédérale de Lausanne	12.51
Ludwig-Maximilians-Universität München	11.54

The number of theses and dissertations per student in IRs was calculated by the following formula (2).

$$\text{TD-Index} = \frac{\text{Thesis and Dissertation}}{\text{Student}} \quad (2)$$

Table 2 shows the top 11 universities by TD-Index, which is high enough to compare with other measurements.

Table 2: Top 11 Universities by TD-Index (> 1.00)

University	GL-Index
Princeton University	8.11
California Institute of Technology	3.85
Uppsala Universitet	2.79
Monash University	2.34
McGill University	1.44
Eidgenössische Technische Hochschule Zürich	1.29
Peking University	1.29
University of Hong Kong	1.25
University of Helsinki	1.25
Technische Universiteit Delft	1.07
Imperial College London	1.04

Table 3 shows correlation coefficients between university evaluation indicators and number of grey literature per faculty member in IRs. All correlation coefficients were not significant at the 5% level. However, all five indicators showed a positive correlation between GL-Index.

Table 3: Correlation coefficients between university evaluation indicators and GL-Index

	Teaching	International Outlook	Research	Citations	Industry Income
GL-Index	0.08	0.05	0.12	0.09	0.02

Fig. 5 shows correlation between GL-Index and university evaluation indicators, which are divided into four groups by continent. The very few universities in North America, Europe and Australasia, which have high index values, have determined correlation coefficients. However, Asian universities are not involved in the decision.

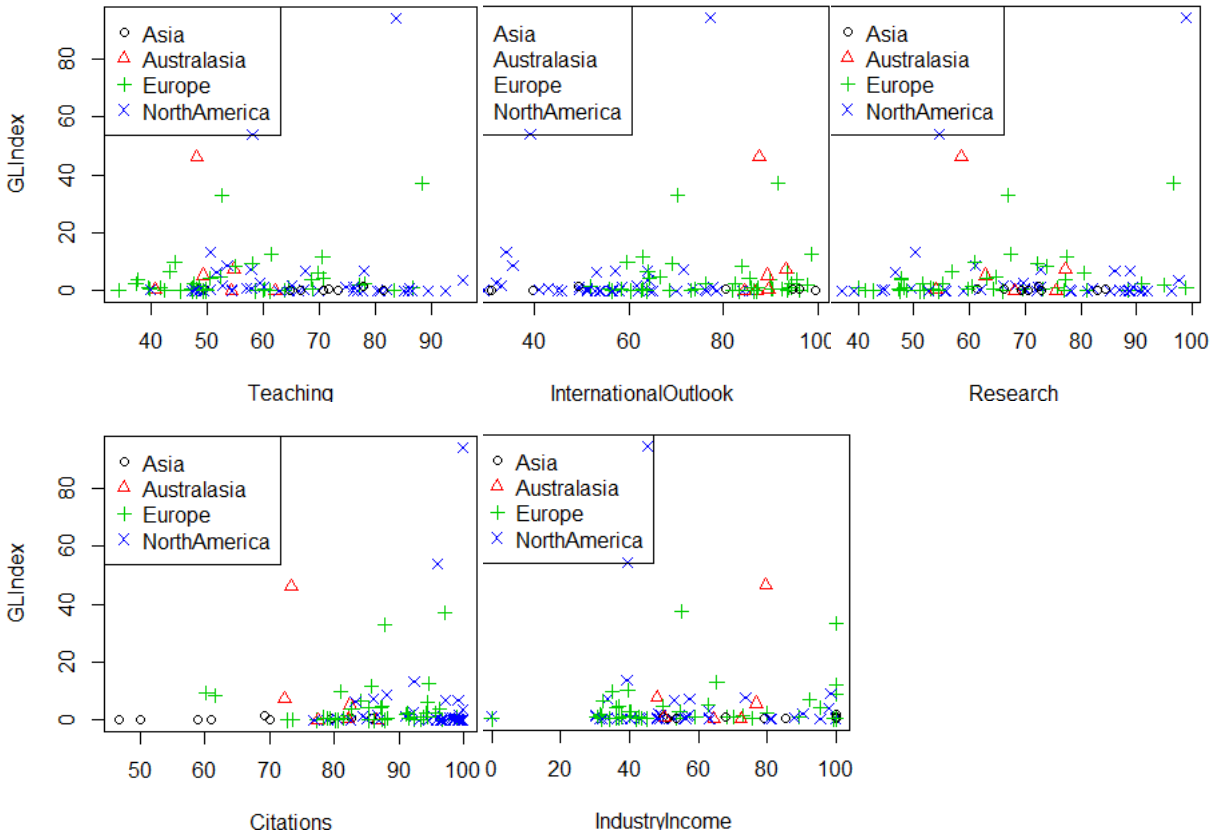


Fig. 5: GL-Index and university evaluation indicators

Table 4 shows correlation coefficients between university evaluation indicators and number of theses and dissertations per student in IRs. Research was significant at the 5% level. The other four indicators were not significant between TD-Index.

Table 4: Correlation coefficients between university evaluation indicators and TD-Index

	Teaching	International Outlook	Research	Citations	Industry Income
TD-Index	0.19	0.13	0.20*	0.03	0.08

Fig. 6 shows correlation between TD-Index and university evaluation indicators, which are divided into four groups by continent. A couple of universities in North America, which have high index values, and some universities in Asia, Australasia and Europe have determined correlation coefficients.

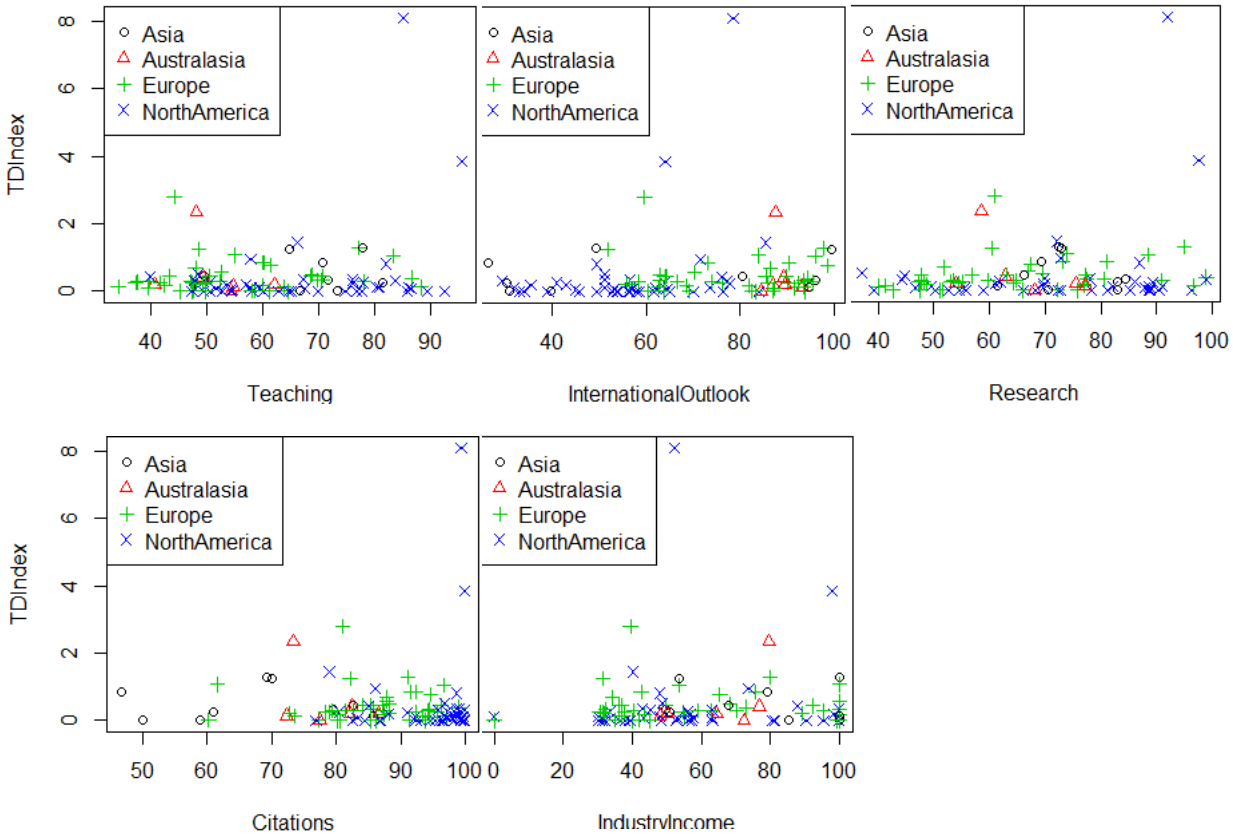


Fig. 6: TD-Index and university evaluation indicators

The Pearson's correlation coefficients between university evaluation indicators and number of grey literature per faculty member in IRs were all positive but statistically insignificant. This is presumably because a small number of universities top-ranked for individual evaluation indicators tend to reposit a large amount of grey literature in IRs. This tendency is particularly clear for universities highly ranked with respect to the indices research, teaching and citations. The grey literature reposited by the members of these universities is likely to contain many scientifically important contributions.

Conclusion

The current situation of disclosure of grey literature was reviewed by analyzing documents reposited in IRs. The situation in individual IRs proved to be diverse. This study revealed that some universities provide a great number of grey literature. However, the number of white literatures, such as journal article and book, in institutional repositories was more than grey literatures. How to promote disclosure of grey literature could be one of the most important issues for institutional repositories for universities.

Acknowledgments

This work was supported by JSPS KAKENHI Grant Numbers 25330388, 25280121.

References

Aina, L.O. (2000). “Grey literature and library and information studies: a global perspective”, International Journal on Grey Literature, Vol. 1 Issue: 4, pp.179-182

Directory of Open Access Repositories (OpenDOAR) http://www.opendoar.org/ (accessed 20 June 2017).

Foster, N.F. and Gibbons, S. (2005). “Understanding faculty to improve content recruitment for institutional repositories”, D-Lib Magazine, Vol. 11 No. 1 http://www.dlib.org/dlib/january05/foster/01foster.html (accessed 20 June 2017).

Georgia Institute of Technology. Scholarly Materials And Research @ Georgia Tech https://smartech.gatech.edu/ (accessed 15 May 2016).

Harvard University. Digital Access to Scholarship at Harvard https://dash.harvard.edu/ (accessed 15 May 2016).

Harvard University. Harvard Dataverse https://dataverse.harvard.edu/ (accessed 15 May 2016).

Harvard University. Harvard Smithsonian Digital Video Library http://www.hsdvl.org/ (accessed 15 May 2016).

Luzi, D. (2010). “Grey Documents in Open Archives”, The Grey Journal, Vol. 6, No.3, pp.137-144.

Monash University. Research Repository http://arrow.monash.edu.au/vital/access/manager/Index (accessed 22 May 2016).

Ludwig-Maximilians-Universität München. Munich RePEc Personal Archive https://mpra.ub.uni-muenchen.de/ (accessed 28 March 2016).

Ohio State University. KnowledgeBank at OSU https://kb.osu.edu/dspace/ (accessed 25 May 2016).

Quacquarelli Symonds (QS) University Rankings - Top Universities https://www.topuniversities.com/university-rankings (accessed 20 June 2017).

Registry of Open Access Repository Mandates and Policies (ROARMAP) https://roarmap.eprints.org/ (accessed 20 June 2017).

Schöpfel, J., Prost, H. and Le Bescond, I. (2012). “Open Is Not Enough: Grey Literature in Institutional Repositories”, GL 13 : Thirteenth International Conference on Grey Literature : The Grey Circuit. From Social Networking to Wealth Creation. Washington, 5-6 December 2011, Dec 2011, France. 2012. <sic 00908862>

Swan, A. (2008). “Study on the availability of UK academic "grey literature" to UK SMEs: Report to the JISC Scholarly Communications Group”, Tech. rep., JISC, London.

Swan, A. (2011). “Institutional repositories - now and next”, In P. Dale, J. Beard, & M. Holland (eds.), University Libraries and Digital Learning Environments, pp. 119-134. Ashgate Publishing, Farnham.

Technische Universiteit Delft. TU Delft Repository https://repository.tudelft.nl/ (accessed 31 May 2016).

Times Higher Education World Universities (THEWU) https://www.timeshighereducation.com/world-university-rankings (accessed 20 June 2017).

University of Cambridge. Apollo (formerly DSpace @ Cambridge) https://www.repository.cam.ac.uk/ (accessed 5 June 2016).

University of Cambridge. Computer Laboratory Technical Reports http://www.cl.cam.ac.uk/techreports/ (accessed 20 June 2016).

University of Cambridge. Digital Himalaya http://www.digitalhimalaya.com/ (accessed 20 June 2016).

University of Cambridge. ESC Publication https://eprints.esc.cam.ac.uk/ (accessed 5 June 2016).

University of Cambridge. Teaching and Learning Research Programme Publications http://www.tlrp.org/dspace/ (not accessed 5 June 2016).

University of Edinburgh. Edinburgh DataShare http://datashare.is.ed.ac.uk/ (accessed 6 June 2016).

University of North Carolina at Chapel Hill. Carolina Digital Repository https://cdr.lib.unc.edu/ (accessed 10 June 2016).

University of Toronto. Research Repository https://tspace.library.utoronto.ca/ (accessed 12 June 2016).

University of Wisconsin. Biological Magnetic Resonance Data Bank http://www.bmrb.wisc.edu/ (accessed 17 June 2016).

Wageningen University and Research. Wageningen Staff Publications (formerly Wageningen Yield) http://library.wur.nl/WebQuery/wurpubs?A240=OK (accessed 13 June 2016).

Wood, D.N. and Smith, A.W. (1993). “SIGLE: A Model for International Co‐operation”, Interlending & Document Supply, Vol. 21 Issue 1, pp.18-22