A Study on the Academic and Research Impact of Shared Contents in Institutional Repositories in Related to Performance Indicators of University Rankings

Hiroyuki Tsunoda
Tsurumi University
2-1-3 Tsurumi, Tsurumi-ku,
Yokohama 230-8501, Japan
tsunoda-h@tsurumi-u.ac.jp

Masaki Nishizawa
National Institute of Informatics
2-1-2 Hitotsubashi, Chiyoda-ku,
Tokyo 101-8430, Japan
nisizawa@nii.ac.jp

Yuan Sun
National Institute of Informatics
2-1-2 Hitotsubashi, Chiyoda-ku,
Tokyo 101-8430, Japan
yuan@nii.ac.jp

Xiaomin Liu
National Science Library,
Chinese Academy of Sciences
33 Beisihuan Xilu, Zhongguancun,
Beijing 100190, China
liuxm@mail.las.ac.cn

ABSTRACT
The purpose of this study is to investigate (a) the current established institutional repositories of the most excellent universities worldwide, (b) the total amount of contents in the individual institutional repositories, and (c) the correlation between the total amount of contents in institutional repositories and the scores of university ranking performance indicators.

Keywords
institutional repository, open access, document type, new indicator, university performance.

INTRODUCTION
Most research have been spending a lot of public funds. Researchers have responsibilities to ensure that the research output at their institutions are both findable and accessible to the public. The possibilities for publishing research outputs have considerably improved since the development of the internet. In recent years there has been an explosion of interest in Open Access (OA) publishing. OA is defined as free, online, immediate permanent access to the full-text of research articles. Two basic principles have been established in the worldwide that provide to the public free OA to publishing research outputs, frequently called gold path and green path of OA (Laakso, 2014). OA repositories are one of the recommended ways to achieve the open access vision described in the Budapest Open Access Initiative definition of open access. There are two kinds of OA repository function, to preserve research results of the authors belonging the institution (e.g., institutional repository) and to archive the digital contents (e.g., digital archive, digital library). This study focuses on the institutional repository, because the institutional repository contains a variety of research outputs, such as the intellectual properties. These intellectual properties means academic and research impact of institution on web (Antelman, 2004). Institutional repository is emerging these intellectual properties sharing in academic and research institutions (Lagzian, Abrizah, & Wee, 2015).

The purpose of this study is to investigate (a) the current established institutional repositories of the most excellent universities worldwide, (b) the total amount of contents in the individual institutional repositories, and (c) the correlation between the total amount of contents in institutional repositories and the scores of university ranking performance indicators.

METHODS AND RESULTS
Institutional Repositories of Top 100 Universities
First, this study investigates the current established institutional repositories of top 100 universities of the Times Higher Education World University Rankings 2015-2016. A directory such as the University of Nottingham’s the Directory of Open Access Repositories (OpenDOAR) and Open Access Repositories makes it easier to identify
and mine the individual repositories (Oliver & Swain, 2006; Pinfield, Salter, Bath, Hubbard, Millington, Anders, & Hussain, 2014).

Figure 1. Number of content in institutional repositories of top 100 universities.
Harvard University (450,000 or more of contents in IRs) ranked first, followed by Peking University (410,000 or more), Monash University (390,000 or more), Uppsala Universitet (370,000 or more), University College London (360,000 or more), University of Queensland (250,000 or more), Wageningen UR Corporate headquarters (250,000 or more), Rijksuniversiteit Groningen (210,000 or more), University of Cambridge (200,000 or more), and University of Oxford (190,000 or more) (Figure 1).

Document Type
Secondly, this study investigates the total amount of contents in the individual institutional repositories. The Institutional Repository contains a variety of contents and each of them has different categories for classifications.

To make statistical analysis, we followed OpenDOAR classification scheme to categorized contents of top 100 universities institutional repositories into 12 main document types, which are journal articles (JA), theses and dissertations (TD), conference and workshop papers (CP), books, chapter and section (BC), datasets (DS), multimedia and audio-visual materials (MM), unpublished reports and working papers (RP), learning objects (LO), patents (PT), software (SW), bibliographic references (BR), other special item types (OI), and 1 type of not clear document types (NC).

“Monash University Research Repository” even had used 89 kinds of document type. These were divided into 12 document types (Table 1).

<table>
<thead>
<tr>
<th>document type</th>
<th>Monash University Research Repository</th>
</tr>
</thead>
<tbody>
<tr>
<td>Journal articles</td>
<td>editorial, journal, journal article, journal issue, journal item, major review, minor review, periodical, review, survey</td>
</tr>
<tr>
<td>Books, chapters and sections</td>
<td>book, book chapter, booklet, bookplate, handbook</td>
</tr>
<tr>
<td>Unpublished reports and working papers</td>
<td>letter, report, research report, working paper</td>
</tr>
<tr>
<td>Conference and workshop papers</td>
<td>conference item, conference paper, conference proceedings, interview, poster</td>
</tr>
<tr>
<td>Multimedia and audio-visual materials</td>
<td>cartographic, chart, graphic, map, moving image, pamphlet, photograph, slide, sound, “sound recording”, speech, still image, videorecording</td>
</tr>
<tr>
<td>Learning objects</td>
<td>lecture</td>
</tr>
<tr>
<td>Bibliographic references</td>
<td>bibliography</td>
</tr>
<tr>
<td>Patents</td>
<td>patent</td>
</tr>
<tr>
<td>Software</td>
<td>programme, software, “software, multimedia”</td>
</tr>
<tr>
<td>Datasets</td>
<td>crystal structure data holding, data holding, dataset</td>
</tr>
<tr>
<td>Other special item types</td>
<td>advertisement, album cover, award, badge, brochure, calendar, card, catalogue, certificate, collection, creative work, document, ephemera, essay, expert commentary, flag, flyer, game, greeting card, guidance document, index, instructions, inventory, invitation, leaflet, manuscript/typescript, menu, mixed material, needlepoint, newspaper issue, other, poetry, postcard, score, sticker, text, “textual apparatus”, ticket, tract</td>
</tr>
</tbody>
</table>

Table 1. Document type of Monash University Research Repository
The most frequently document type was JA (43%), followed by TD (13%), CP (8%), BC (6%), DS (6%), RP (4%), MM (3%), OI (8%), NC (9%) and LO, PT, SW, BR which were extremely small (Figure 2). As known, so far there are lot of indicators to evaluate universities’ research.

Correlation

Finally, we this study investigates the correlation between the amount of contents in institutional repositories and the scores of university ranking indicators (Table 2). Note that the number indicating TD (theses and dissertations) was normalized by the number of students, while others for the rest 11 indicators were normalized by the number of faculties.

<table>
<thead>
<tr>
<th>Times indicators</th>
<th>nJA</th>
<th>nTD</th>
<th>nCP</th>
<th>nBC</th>
<th>nDS</th>
<th>nRP</th>
<th>nMM</th>
<th>nBR</th>
<th>nLO</th>
<th>nPT</th>
<th>nSW</th>
<th>nOI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Teaching</td>
<td>-.038</td>
<td>.186</td>
<td>-.068</td>
<td>-.191</td>
<td>.151</td>
<td>-.063</td>
<td>-.046</td>
<td>-.038</td>
<td>-.080</td>
<td>-.045</td>
<td>.060</td>
<td>.007</td>
</tr>
<tr>
<td>International outlook</td>
<td>.203$^{*}$</td>
<td>.129</td>
<td>.314$^{**}$</td>
<td>.269$^{**}$</td>
<td>.042</td>
<td>.066</td>
<td>-.113</td>
<td>.121</td>
<td>.026</td>
<td>-.038</td>
<td>-.016</td>
<td>.098</td>
</tr>
<tr>
<td>Research</td>
<td>.018</td>
<td>.202$^{*}$</td>
<td>-.048</td>
<td>-.127</td>
<td>.192</td>
<td>.050</td>
<td>-.095</td>
<td>.058</td>
<td>-.010</td>
<td>-.029</td>
<td>.058</td>
<td>.024</td>
</tr>
<tr>
<td>Citations</td>
<td>-.127</td>
<td>.028</td>
<td>-.122</td>
<td>-.032</td>
<td>.114</td>
<td>.007</td>
<td>.044</td>
<td>-.129</td>
<td>-.028</td>
<td>-.160</td>
<td>-.008</td>
<td>-.040</td>
</tr>
<tr>
<td>Industry income</td>
<td>.173</td>
<td>.083</td>
<td>.066</td>
<td>-.031</td>
<td>-.056</td>
<td>.289$^{**}$</td>
<td>-.064</td>
<td>-.038</td>
<td>.200$^{*}$</td>
<td>.022</td>
<td>.166</td>
<td>.023</td>
</tr>
</tbody>
</table>

*p < .05, **p < .01

Table 2. Correlation between contents and ranking indicators.

DISCUSSION

Control of Document Type

IRs need to use many document types, because IRs stored diverse contents. The number of document types in IRs of Top 100 Universities were around 800 kinds. However, the document type is too much. All of IRs have to control the document type.

Normalization of number of Contents

The results indicate significant positive correlation between normalized Journal Articles (nJA) and International outlook, between normalized theses and dissertations (nTD) and Research, between normalized conference and workshop papers (nCP) and International outlook, between normalized books, chapter and section (nBC) and International outlook, and between normalized unpublished reports and working papers (nRP) and Industry income.

Number of citations

Generally, the faculty in the fields of science, technology and medicine has produced more articles than the faculty in the fields of arts and humanities. Normalization is divided by the number of faculty in both fields, thereby the results indicate the absent of correlation between normalized Journal Articles (nJA) and Research, Citations.

For example, in the Times Higher Education World University Ranking, one of the most well-known world university rankings, the ranking consists of five indicators of Teaching (the learning environment), International outlook (staff, students, research), Research (volume, income and reputation), Citations (research influence), and Industry income (knowledge transfer). In this study has we regarded that total amount of contents in institutional repositories as a kind of new aspects indicating the impact of institutes on web, because these contents are available for anyone to use via the internet.
Correlation

**International outlook of Times indicators**

International outlook has been built from International-to-domestic-student ratio, International-to-domestic-staff ratio, and International collaboration. International outlook correlated with nJA, nCP, and nBC (Figure 3). Internationalism might increase the paper, conference, and book of productivity.

**Research of Times indicators**

Research has been built from Reputation survey, Research income, and Research productivity. Research correlated with nTD (Figure 3). Improvement of the research might increase the degree.

**Industry income of Times indicators**

Industry income has been built form knowledge transfer. Industry income correlated with nRP and nLO (Figure 3). Rise of the knowledge transfer from industry might increase the reports and teaching materials of productivity.

CONCLUSION

The amount of contents in institutional repositories had been independent from the other indicators. This indicator could be a candidate for a new indicator to measure the university performance.

ACKNOWLEDGMENTS

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

REFERENCES


self-archiving is allowed. Scientometrics, 99 (2), 475-494.


