

Toward Easy Deposit: Lowering the Barriers of Green Open Access with Data Integration and Automation

Hui Zhang

Oregon State University Libraries, Oregon State University, 121 The Valley Library, Corvallis, OR 97330, USA; hui.zhang@oregonstate.edu; Tel.: +01-541-737-8233

Received: 13 March 2020; Accepted: 18 May 2020; Published: 20 May 2020

Abstract: This article describes the design and development of an interoperable application that supports green open access with long-term sustainability and improved user experience of article deposit. The lack of library resources and the unfriendly repository user interface are two significant barriers that hinder green open access. Tasked to implement the open access mandate, librarians at an American research university developed a comprehensive system called Easy Deposit 2 to automate the support workflow of green open access. Easy Deposit 2 is a web application that is able to harvest new publications, to source manuscripts on behalf of the library, and to facilitate self-archiving to a university's institutional repository. The article deposit rate increased from 7.40% to 25.60% with the launch of Easy Deposit 2. The results show that a computer system can implement routine tasks to support green open access with success. Recent developments in digital repository provide new opportunities for innovation, such as Easy Deposit 2, in supporting open access. Academic librarians are vital in promoting "openness" in scholarly communication, such as transparency and diversity in the sharing of publication data.

Keywords: open access; article deposit; institutional repository

1. Introduction

A combination of several major developments has caused a surge of interest in open access (OA) in recent years. First, more funding agencies, such as the US National Institutes of Health and the US National Science Foundation, have made OA publishing mandatory for grantees [1,2]. Moreover, academic libraries are considering OA as a potential solution to maintaining their access to scientific and research literature [3].

OA is a newer form of scholarly publishing, which refers to scholarly literature that is free to read online [4]. Scholarly publishing nowadays is complex, and researchers have made attempts to classify OA literature into subtypes based on factors such as the source and the license of re-use [5–7]. The focus of this article is green OA, the practice of authors depositing their post-print (peer-reviewed) manuscripts into a university's institutional repository for free public access. Green OA, or self-archiving, is usually the last option for authors if they cannot afford the article processing charges to publish in a gold or hybrid OA journal [8]. However, despite the benefits, recent studies find that only 10% of the scholarly articles were self-archived by faculty authors, even including those at institutions with robust institutional repositories and OA policies [9,10]. Several barriers can be blamed for the small percentage of green OA. The early study attributes the lack of author self-archiving to inadequate marketing and the effort required for manuscript deposit [11], while more recent studies suggest that enforcement of the OA mandate from the University and funders is required in order to see more voluntarily manuscript depositing from the faculty [12,13].

2. Support for Open Access at Oregon State University Libraries

The faculty members at Oregon State University (OSU), a public research university in the US, passed an OA policy in 2013 that mandates every faculty member to grant OSU permission to make available his or her scholarly articles and to reproduce and distribute those articles for open access. OSU Libraries (OSUL) are in charge of implementing the OA policy by promoting the policy to faculty members and depositing accepted manuscripts (post-peer review, pre-typeset) of their articles to OSU's institutional repository (ScholarsArchive@OSU), also managed by the library [14]. An early workflow to implement the university's OA mandate required human intervention at every step. First, a librarian created a search alert in the Web of Science that found all of the articles authored or co-authored by the OSU faculty, and the results were sent to a dedicated OSU email address. A library staff member was assigned as the owner of that email account and coordinated the OA workflow. For each article listed in the Web of Science search alert, library staff took the following actions:

1. Searched the SHERPA/RoMEO¹, an online database that collects publisher open access policies, to obtain the corresponding publisher's policies on copyright.
2. Found the email address of the contact author from the search alert.
3. Contacted the author to ask them to archive the article manuscript into ScholarsArchive@OSU.
4. If the author replied with the manuscript in an attachment, the library staff deposited it into ScholarsArchive@OSU on behalf of the author.

This manual workflow produced a high deposit rate, defined as the number of deposited articles divided by the total articles authored by OSU faculty, around 44% between 2013 and 2015 [15]. However, OSUL had to abandon this workflow because it could no longer commit a full-time staff to OA alone due to limited resources. The faculty feedback suggested that the main barrier for self-archiving was the time and effort required to deposit articles into the institutional repository. For example, the faculty complained of the unfriendly repository user interface and the number of required metadata fields.

To provide sustainable and long-term support for green OA, several librarians and staff members at OSUL initiated a project that would not only automate many steps of the OA support workflow but also make article self-archiving as easy as clicking a button. The result of the aforementioned project was Easy Deposit 2², a web application that harvests journal articles, automates OA promotion and outreach, and supports easy deposit into ScholarsArchive@OSU. This study of Easy Deposit 2 tried to answer the research question of whether a computer system could substitute for library staffing to support green OA. The rest of the article is outlined as follows:

- Methodology: A description of the technical details of automation at each main step of the OA support workflow and how to integrate all of the steps under Easy Deposit 2.
- Results and Discussion: A comparison of the numbers of article deposits in different periods (early workflow, with Easy Deposit 2 support, and faculty self-archiving) and a review of the lessons learned by supporting green OA with this new approach.
- Conclusion: An introduction to the future plan and an explanation as to why green OA is still valuable with the growth of other OA publishing models.

This paper demonstrates how the library can effectively utilize the recent digital library developments to achieve sustainable support for the green OA of scholarly publications.

3. Methodology

3.1. System Design of Easy Deposit 2

Easy Deposit 2 (ED2) was developed as a reboot of OSUL's support for green OA with automation. We named the proposed system Easy Deposit 2 because EasyDeposit had already been

¹ SHERPA/RoMEO: <http://sherpa.ac.uk/romeo/index.php>

² Easy Deposit 2: <https://ed2.library.oregonstate.edu/>

used as the name of another open-source toolkit³. ED2 is a comprehensive system developed using the web application framework called Ruby on Rails⁴. It contains a MySQL⁵ database as the backend storage, a deposit interface, and a dozen scripts created for a variety of functions. The system searches the publisher's database for faculty publications and harvests their metadata records. The harvested metadata records are saved in ED2's database and processed for creating the manuscripts recruiting emails. ED2 first looks for and uses the email address of the corresponding author in the metadata records. If the contact information is not available from public records, ED2 then searches OSU's electronic faculty directory for an email address with the author's name. For every author with a detected email address, ED2 sends out a message asking for self-archiving of the article manuscript into ScholarsArchive@OSU. The authors can use the deposit link embedded in the email to upload and submit their manuscripts to OSU's institutional repository. The functions of ED2 automate most of the equivalent activities in the OA workflow that were previously carried out by library staff. The three modules of ED2 and their corresponding functions are summarized in Table 1 with key technical details such as the conditions that trigger workflow activities and data collection and integration from a variety of resources.

Table 1. Three Easy Deposit 2 modules with functions provided and key technical details such as data collection and integration from different resources.

Functions Supported		Key Technical Details
Fetch Module	Harvesting faculty authored articles	The module queries the Web of Science database via API daily. Any newly published articles in the past four weeks by the university's faculty will trigger the OA workflow to recruit manuscripts.
	Parsing metadata	The module parses article metadata from harvested records and saves parsed metadata records in the local database.
Parsing and Email Module	Recruiting article manuscript	The module looks for author email addresses using either publisher's application programming interface (API) or university's directory service; sends out manuscript recruiting email to authors with all found email addresses.
Deposit Module	Deposit article manuscript into OSU's institutional repository	The author clicks the deposit link embedded in the article recruiting email to initiate the self-archiving, then uploads and ingests the manuscripts to ScholarsArchive@OSU using ED2's deposit page.

The diagram in Figure 1 illustrates the functions and modules of ED2 and how ED2 interacts with external entities such as Web of Science, the authors, and the institutional repository.

³ EasyDeposit toolkit: <http://easydeposit.swordapp.org/>

⁴ Ruby on Rails: <https://rubyonrails.org/>

⁵ MySQL: <https://www.mysql.com/>

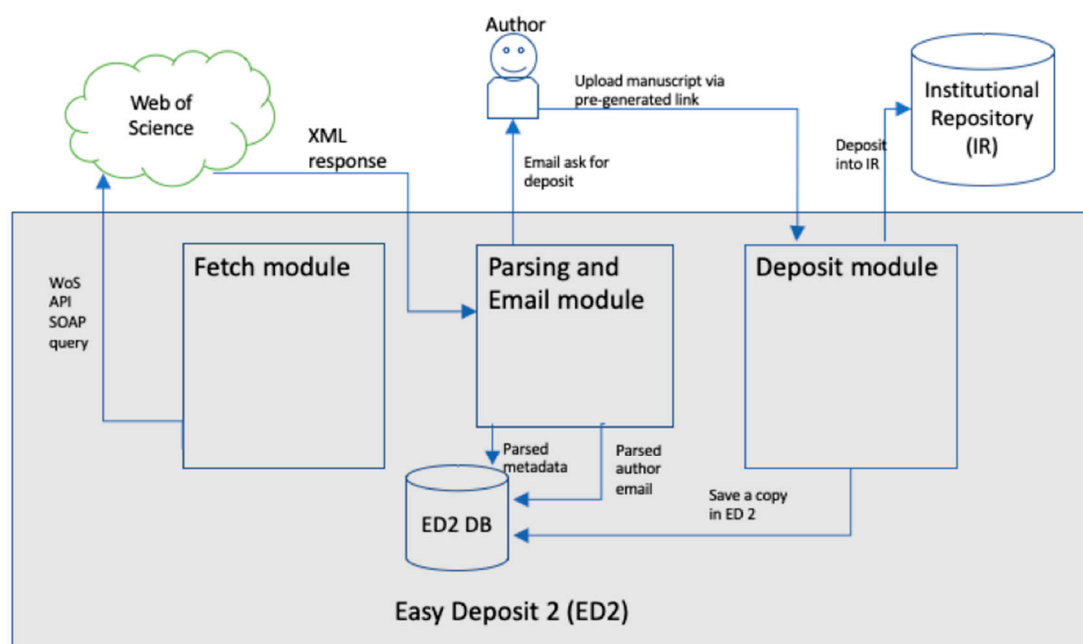


Figure 1. Easy Deposit 2 (ED2) system diagram.

The only step in the previous workflow no longer covered by ED2 is verifying the publisher's self-archiving policies. We consulted the General Counsel of the University about whether that verification is necessary in light of the University's Open Access policy. The policy grants the University (and faculty) a nonexclusive license to distribute, at the very least, the accepted manuscript version of their articles. The counsel informed us that allowing faculty to deposit their articles without library mediation would protect the library from liability.

3.2. Recruiting and Depositing Manuscripts with Easy Deposit 2

ED2 uses the Web of Science API⁶ to retrieve from the Web of Science core collections journal articles published by faculty members in the last four weeks. The extended API returns the found articles in XML with metadata fields such as the title, author, abstract, and contact email of the corresponding author. ED2 parses and saves the metadata for each article in the local database, together with the "author publication pair," which is a data structure that stores the article identifier and the author's email. The primary purpose for creating the author publication pair is to retrieve the correct metadata record during the article depositing process. If an article has several corresponding authors, ED2 will generate an author publication pair for every author. However, since April 2019, the email of the corresponding author is no longer available from the Web of Science API. The impact of relying on the commercial APIs for the future of ED2 is examined in the Discussion section.

The email sent to authors for recruiting manuscripts is generated using a template where contents such as article title and deposit link are pulled from the ED2 database. It also includes links to the OSU's OA policy and contact information for ScholarsArchive@OSU. Interested readers can find the template of the manuscript's recruiting email attached as a supplementary file of this article. The deposit link embedded in the article-recruiting email includes a "key" generated by the email of

⁶ Web of Science API: <https://clarivate.com/webofsciencegroup/solutions/xml-and-apis/>

the contact author in the SHA-2⁷ (Secure Hash Algorithm 2) cryptographic hashing standard. When the faculty member clicks that link, ED2 compares the SHA2 key with the author's email stored in its database for authentication.

Oregon State University Logged in as zhanghu All Publications Admin Logout

Easy Deposit @ Oregon State University Libraries and Press

Textural property and cell wall metabolism of 'Golden Bosc' and 'd'Anjou' pears as influenced by oxygen regimes after long-term controlled atmosphere storage

(fingerprint) f7d10d942a8eef01eca60d79c8dbf3e1bb7d24a4425f8b70d5287adceacac703
Published to ScholarsArchive@OSU on 2019-03-25 21:50:55 UTC
<https://ir.library.oregonstate.edu/concern/articles/q524jv232>

Attribute	Value(s)
wos_uid	WOS:000459827400004

abstract Controlled atmosphere (CA) allows long-term storage of European pears (*Pyrus communis* L.) without chemical treatment to deliver a natural melting (buttery and juicy) texture for consumers. However, the relationship between textural properties and cell wall metabolism as influenced by O-2 regimes has not been comprehensively determined. In this study, 'Golden Bosc' and 'd'Anjou' pears were stored in 21 (air), 2, 1, or 0.5% O-2 with < 0.5% CO2 for up to 8 and 10 months at -1.1 degrees C plus 7 d of ripening at 20 degrees C, respectively. Melting texture development in both cultivars showed high correlations with the level of water-soluble pectinuronides (WSP) and activity of beta-galactosidase (beta-GAL). Also, activities of pectin methylesterase (PME) and alpha-arabinofuranosidase (alpha-ARF) were associated with softening in 'Golden Bosc' pears. Concentrations of PMA, soluble pectinuronides (SPU) and WSP + SPU + sodium ascorbate soluble

Attached Files

1-s2.0-S0925521418310391-main.pdf

Processed Events

Publish Work Notification Email scholarsarchive@oregonstate.edu ✓
2019-03-25 21:50:55 UTC : Email initiated by scholarsarchive@oregonstate.edu at 2019-03-25 14:50:56 -0700

Publish Work scholarsarchive@oregonstate.edu ✓
2019-03-25 21:50:47 UTC : Published to the repository at 2019-03-25 14:50:55 -0700

File(s) added scholarsarchive@oregonstate.edu ✓
2019-03-25 21:50:37 UTC : 1-s2.0-S0925521418310391-main.pdf

Send article recruit with email saved in AuthorPublication ✓
2019-03-13 07:02:03 UTC : Article author recruitment email completed by scholarsarchive@oregonstate.edu at 2019-03-13 00:02:03 -0700

Fetch publication content from Web of Science ✓

Figure 2. Demonstration of Easy Deposit 2 (ED2) deposit page with a randomly selected article.

For self-archiving manuscripts into ScholarsArchive@OSU, an author only needs to interact with ED2's article deposit page without the need to create an account for upload. Figure 2 shows the actual look-and-feel of ED2's deposit page as a screenshot taken with a randomly selected article for demonstration. Metadata fields such as title and abstract are auto-populated with values from ED2's database; the author only needs to upload the manuscript and submit it for self-archiving. ED2 packs the uploaded file and its metadata in the JSON (JavaScript Object Notation) format and ingests the package into ScholarsArchive@OSU over the web as a HTTP (Hypertext Transfer Protocol) request. ED2 has a dashboard for system administrators and librarians, providing information such as the total number of journal articles that have been harvested and how many of them have been self-archived by faculty members. Access to the administration dashboard is restricted to OSU's librarians and staff members.

4. Results

The effect of ED2 on green OA was evaluated by the article deposit rate, which is a widely used metric by the OA community. In this study, the article deposit rate is defined as the number of articles deposited into ScholarsArchive@OSU divided by the total number of articles published by OSU faculty members. The deposit rates shown in Table 2 were calculated over four phases:

- Pre-WoS: The period before OSUL implemented the manual workflow using the Web of Science (WoS) index.
- WoS manual: The period when OSUL committed one full-time member of staff to recruit and deposit manuscripts on behalf of the faculty.
- Period in-between: The period in between cession of the manual OA workflow and the launch of ED2.
- ED2 OA: The period when OSUL used ED2 to support OA.

⁷ SHA-2 key: <https://ruby-doc.org/stdlib-2.4.0/libdoc/digest/rdoc/Digest/SHA2.html>

Table 2. The trend of article deposit rates along with four phases of open access support between the period of 2011 and 2018: the period before OSUL implemented the manual workflow using the Web of Science index (Pre-WoS), OSUL committed one full-time member of staff to recruit and deposit manuscripts on behalf of the faculty (WoS manual), the period in between cession of the manual OA workflow and the launch of ED2 (Period in-between), and the period when OSUL used ED2 to support OA (ED2 OA).

Pre-WoS	Total Articles	Articles Deposited	Deposit Rate
2011	1615	197	12.20%
Pre-WoS Total	1615	197	12.20%
WoS Manual	Total Articles	Articles Deposited	Deposit Rate
2012	1779	799	44.91%
2013	1682	747	44.41%
2014	1731	742	42.87%
WoS Manual Total	5192	2288	44.07%
Period in-between	Total Articles	Articles Deposited	Deposit Rate
2015	2113	325	15.38%
2016	2334	152	6.51%
2017	2450	105	4.29%
2018-01 to 2018-09	1885	68	3.60%
Period in-between Total	8782	650	7.40%
ED2 OA	Total Articles	Articles Deposited	Deposit Rate
2018-10 to 2019-03	1301	333	25.60%
ED2 OA Total	1301	333	25.60%

The numbers in Table 2 are obtained by searching the underlying database with appropriate metadata fields. For the number of “Articles Deposited,” the author searched ScholarsArchive@OSU with the resource type as “Article,” and then separated the results by year with values in the “Date Created” field. For the number of “Total Articles,” the author first submitted an address search to the WoS with the query of “AD = ‘Oregon State University’,” and then used the “Year” facet to estimate the total number of articles published by the faculty. The article deposit rate tripled, from 7.40% to 25.60%, compared to the rate of the previous three years (between 2015 and 2018) since the launch of ED2. However, the article deposit rate with ED2 is about 19 percentage points lower (25.60% vs. 44.07%) than it was between 2012 and 2014 when green OA was supported by a fulltime library staff.

The reported green OA deposit rate for published journal articles is about 12%, with the considerable difference among research fields [16]. For instance, the green OA deposit rate for journal articles in the field of Library and Information Science ranged from 20% to 31% between 2012 and 2016 [17]. Overall, the green OA article deposit rate for most research fields appeared to be between 10% to 30%.

5. Discussion

Previous studies show that without an outreach and matching repository service, few faculties would voluntarily deposit manuscripts into an institutional repository, even if this was a university mandate [11,12,15]. At the author’s university, the article deposit rate for the most successful green OA was about 44% when a full-time staff was responsible for recruiting manuscripts and depositing them on the faculty’s behalf. Most academic libraries in the US cannot afford to have a full-time staff member dedicated to supporting self-archiving. The research question of this study is whether a computer system could replace library staffing for the support of green OA. The results show that ED2 can significantly increase the number of authors who would voluntarily deposit their

manuscripts into the OSU's institutional repository. The article deposit rate with ED2 was 25.60%, which is toward the top of reported green OA article deposit rates for most research fields. However, the success of green OA requires more than library support. Parties with influence and power, such as tenure and promotion committees and funding agencies, should enhance the OA mandate in their policies.

The ED2 system was developed as an extension of OSU's repository service and was integrated into OSU's institutional repository. There are existing efforts to harvest publications using the Web of Science API [18]. The Bibliographic Management System⁸, developed by the Stanford University Library, inspired us to initiate the ED2 project. ED2 is an innovation in that it is the first comprehensive system specially developed to support green OA. It is the article deposit portal of ScholarsArchive@OSU, the software agent that automates the workflow of green OA, and the database of faculty publications with their OA status. A librarian can use ED2's dashboard to answer important questions such as how many articles faculty members have published in the last year and how many of them have been self-archived into ScholarsArchive@OSU.

Access to metadata records of scholarly publications has become the bottleneck of running ED2. The author email is vital for ED2 because it is required for the outreach, the recruiting of the manuscripts, and the authentication for deposit. ED2 was able to pull out the corresponding author email(s) from the Web of Science API until April 2019, when the company decided to exclude author email(s) from the API output. An alternative solution was developed to obtain author emails by looking for author names in OSU's staff directory. However, for all of the articles harvested by ED2, fewer than 50% are found to have at least one current faculty author. The lesson learned here is that the library cannot rely on a single and commercial source for data, because the library has little influence over vendor decisions, and the priorities of both parties are not aligned in many situations. Several popular APIs—Elsevier⁹, Crossref¹⁰, and ORCID¹¹, have been tested to determine whether they could replace the API from the Web of Science. The results show that none of the APIs include author contact information, and the availability of metadata, such as the abstract and copyright information in the API output, is inconsistent. A long-term solution for the problem of data accessibility may be that non-profit organizations, such as Crossref and ORCID, collaborate and provide critical information such as copyright and contact details through their APIs.

6. Conclusions

Recent technological developments in the digital repository and web services provide new opportunities for innovation in supporting OA publishing. Easy Deposit 2 is a comprehensive system developed to prompt OA and to facilitate the author self-archiving. ED2 was designed to substitute dedicated support staff by automating major OA tasks, such as harvesting faculty-published articles, source manuscripts, and article depositing into an institutional repository. ED2 significantly increased the number of article manuscripts voluntarily deposited by faculty members. The article deposit rate to OSU's institutional repository was raised from 7.40% to 25.60% with the support of ED2. However, a computer system, like ED2, does not replace the knowledge, commitment, and value represented by professional librarians and library research support services.

The priority for the future development of ED2 will be to diversify the data sources, with a preference for community-based or non-profit organizations. The proposed methods include harvesting article metadata from Crossref and searching for author contact information in ORCID. Breaking the monopoly of commercial indices on publication data is critical for the long-term sustainability of supporting OA and promoting “openness” in scholarly communication. Academic

⁸ Bibliographic Management System: https://github.com/sul-dlss/sul_pub

⁹ Elsevier API: <https://dev.elsevier.com/>

¹⁰ Crossref API: <https://www.crossref.org/services/metadata-delivery/rest-api/>

¹¹ ORCID API: <https://orcid.org/organizations/integrators/API>

librarians have vital roles in a more open and transparent sharing of publication data by collaboration with scientific communities and non-profit publication data aggregators.

Funding: This research received no external funding.

Acknowledgments: Thanks are extended to the members of Easy Deposit 2 project (Josh Gum, Sarah Imholt, Steve VanTuyl, and Margaret Mellinger) for their contributions and help in the development and operation of the system.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Borrego, Á. Institutional Repositories versus ResearchGate: The Depositing Habits of Spanish Researchers. *Learn. Publ.* **2017**, *30*, 185–192, doi:10.1002/leap.1099.
2. Laakso, M.; Bo-Christer, B. Anatomy of Open Access Publishing: A Study of Longitudinal Development and Internal Structure. *BMC Med.* **2012**, *10*, 124, doi:10.1186/1741-7015-10-124.
3. Robert, S. UC Faculty to Elsevier: Restart Negotiations, or Else. Berkeley News. Available online: <https://news.berkeley.edu/2019/08/07/uc-faculty-to-elsevier-restart-negotiations-or-else/> (accessed on 7 August 2019).
4. Steinhart, G. LibGuides: Open Access Publishing: What Is Open Access? Available online: <https://guides.library.cornell.edu/openaccess/home> (accessed on 4 March 2020).
5. Dallmeier-Tiessen, S.; Bettina, G.; Robert, D.; Jenni, H.; Peter, I.-K.; Deborah, K.; Simon, L. Open Access Publishing—Models and Attributes. 2010. Available online: https://pure.mpg.de/pubman/faces/ViewItemOverviewPage.jsp?itemId=item_1759194 (accessed on 24 April 2020).
6. Burtle, L. GSU Library Research Guides: Open Access: Types of OA. Available online: <http://research.library.gsu.edu/c.php?g=115588&p=754380> (accessed on 24 April 2020).
7. Collister, L. LibGuides: Open Access @ Pitt: Types of OA. Available online: <https://pitt.libguides.com/openaccess/typesofoa> (accessed on 24 April 2020).
8. Heather, P.; Priem, J.; Larivière, V.; Alperin, J.P.; Matthias, L.; Norlander, B.; Farley, A.; West, J.; Haustein, S. The State of OA: A Large-Scale Analysis of the Prevalence and Impact of Open Access Articles. *PeerJ* **2018**, *6*, e4375, doi:10.7717/peerj.4375.
9. Tillman, R. Where Are We Now? Survey on Rates of Faculty Self-Deposit in Institutional Repositories. *J. Librariansh. Sch. Commun.* **2017**, *5*, eP2203, doi:10.7710/2162-3309.2203.
10. Stevan, H.; Brody, T.; Vallières, F.; Carr, L.; Hitchcock, S.; Gingras, Y.; Oppenheim, C.; Hajjem, C.; Eberhard, R.H. The Access/Impact Problem and the Green and Gold Roads to Open Access: An Update. *Ser. Rev.* **2008**, *34*, 36–40, doi:10.1016/j.serrev.2007.12.005.
11. Björk, B.-C. Open Access—Are the Barriers to Change Receding? *Publications* **2013**, *1*, 5–15, doi:10.3390/publications1010005.
12. Tomasz, N.; Murray, A. The Critical Role of Institutional Services in Open Access Advocacy. *Int. J. Digit. Curation* **2013**, *8*, 84–106, doi:10.2218/ijdc.v8i1.238.
13. Jenny, F.; Spezi, V.; Proberts, S.; Creaser, C. Towards an Understanding of the Relationship between Disciplinary Research Cultures and Open Access Repository Behaviors. *J. Assoc. Inf. Sci. Technol.* **2016**, *67*, 2710–2724, doi:10.1002/asi.23621.
14. Oregon State University Open Access Policy | Libraries | Oregon State University. Available online: <https://library.oregonstate.edu/open-access> (accessed on 11 February 2020).
15. Hui, Z.; Boock, M.; Wirth, A. It Takes More than a Mandate: Factors That Contribute to Increased Rates of Article Deposit to an Institutional Repository. *J. Librariansh. Sch. Commun.* **2015**, *3*, eP1208, doi:10.7710/2162-3309.1208.
16. Bo-Christer, B.; Laakso, M.; Welling, P.; Patrik, P. Anatomy of Green Open Access. *J. Assoc. Inf. Sci. Technol.* **2014**, *65*, 237–250. doi:10.1002/asi.22963.

17. Emery, J. How Green Is Our Valley? Five-Year Study of Selected LIS Journals from Taylor & Francis for Green Deposit of Articles. Insights: *UKSG J.* **2018**, doi:10.1629/uksg.406.
18. Belle, S.L.; Clark, J.A. Citations as Data: Harvesting the Scholarly Record of Your University to Enrich Institutional Knowledge and Support Research. *Coll. Res. Libr.* **2017**, *78*, 952, doi:10.5860/crl.78.7.952.



© 2020 by the author. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).