ASEE 2022 ANNUAL CONFERENCE
Excellence Through Diversity
MINNEAPOLIS, MINNESOTA, JUNE 26TH-29TH, 2022 ASEE
Paper ID #36836

Using Esploro to Increase Visibility of Engineering Faculty Research Work

Qianjin Zhang (Engineering & Informatics Librarian)

Qianjin (Marina) Zhang is the Engineering & Informatics Librarian at the Lichtenberger Engineering Library, the University of Iowa. Her work focuses on providing data management support and developing library research data services. As a subject librarian, she also manages collections, and provides instruction, reference and consultation services for engineering faculty and students. She holds a MA in Information Resources & Library Science from the University of Arizona (Tucson, AZ), and a BS in Biotechnology from Jiangsu University of Science and Technology (Zhenjiang, China).

Wendy Robertson

Wendy (she/her) manages the University of Iowa Institutional Repository (Iowa Research Online) and works with the Libraries' journal publishing. She also works closely with the Graduate College regarding theses and dissertations. Her research/interest areas are metadata, preservation, accessibility, and scholarly communication.

Kari Kozak (Director, Lichtenberger Engineering Library)

Kari Kozak has been the Director of the Lichtenberger Engineering Library at The University of Iowa since July 2011. She provides instruction, reference, and consultation services to student, faculty, and staff within the departments and research centers in the College of Engineering as well as the Department of Computer Science. Kari's area of interests include outreach, instruction, innovation and design. She is available to teach classes on library resources and consult on issues including standards, patents, evaluating information and other information literacy concepts, copyright, and the Creative Space and Tool Library.

Sara Scheib

Sara is the Director of the Scholarly Impact unit at the University Libraries and works to support researchers and scholars in all stages of their careers and across disciplines in navigating the scholarly publishing landscape and maximizing the impact of their work.

James M Cox

James M. Cox is the Public Services and Emerging Technologies Librarian for the Lichtenberger Engineering Library at the University of Iowa. His interests focus on technology and tools, both physical and digital, that are incorporated into the library to enhance the user experience and expand the available services.

Amanda Bartel

Using Esploro to Increase Visibility of Engineering Faculty Research Work

Abstract

As Digital Measures is no longer used by the University of Iowa, the College of Engineering approached the Engineering Library to identify a new system to showcase their faculty work. The Engineering Library identified Esploro, a research information management system powered by Ex Libris, because there are several advantages: (1) we have full control of metadata; (2) publication records are indexed and searchable in our library catalog system; (3) publication records are discoverable by search engines; (4) Smart Harvest, an automated content harvesting feature, can load publication records from a variety of sources; (5) a complete list of publications can be easily generated for grant applications and reviews. Since the College of Engineering is the first college to adopt Esploro, there is no standard workflow for us to follow. We developed a workflow by trial and error: (1) making an initial estimate of the volume of the publication records for 110 faculty; (2) turning on the Smart Harvest to load records; (3) reviewing whether the potential records are matched to the correct people, noting especially document types and DOIs, which can indicate higher quality records from the Ex Libris database, and manually approving or rejecting the records in Esploro; (4) identifying and collecting records in Scopus to expand the Ex Libris database; (5) activating Smart Harvest again to add additional records that were missed; (6) running various reports about quality of records and improving records as needed. Facing the major challenges including the large volume of records, DOIs not existing in journal articles published in early years, and DOIs not existing in some conference proceedings and book chapters, we shared some best practices and lessons learned. Since Esploro continues to be enhanced, we will refine our workflow with new functionality as it is added.

Introduction

Digital Measures was used by University of Iowa until July 2021 to track faculty teaching, publications, research and service, and to form the basis of faculty profiles. After its cancellation, the College of Engineering started investigating a new system to showcase their faculty work. The associate dean at the College of Engineering approached the Engineering Library for assistance. The Engineering Library helped identify a feasible solution: Esploro. Esploro, a research information management system powered by Ex Libris, not only serves as our institutional repository but also can support researcher profiles. So far, 37 institutions have adopted Esploro [1]. Although Digital Measures has some similar features in managing faculty research activities, Esploro is the product which the University Libraries has been actively involved in product development. The following two subsections would be a narrative of the University Libraries partnership with Esploro's vender, Ex Libris, and the project establishment of College of Engineering faculty profiles.

University Libraries' Partnership with Ex Libris

In November 2018, the University Libraries signed on with Ex Libris as an early adopter of Esploro, having previously been a development partner for the product, beginning June 2017. We formed an implementation team consisting of the Associate University Librarian, Director of Library Information Technology, Director of Scholarly Impact, Institutional Repository & Metadata Librarian, Systems Librarian, and the Program Coordinator for the Library for the Health Sciences. We also consulted with two external team members: the Director of the Research Development Office and the Senior Project Manager for Administrative Information Systems. The team met regularly with the Ex Libris Project Team to test the system, provide feedback, and report problems.

Over the course of three years, the implementation team migrated ~16,000 faculty publications, theses and dissertations, and other content from our Digital Commons instance. The team implemented various deposit workflows and oversaw the launch of three interfaces, including two versions of the administrative interface, three versions of the public interface, and the researcher profiles. The University of Iowa was the first institution to go-live with Esploro in July 2019. We officially ended our Digital Commons instance in September 2021. As of January 2022, Esploro (locally branded as Iowa Research Online) includes approximately 9,200 affiliated researchers and holds approximately 81,000 research outputs which have been downloaded over 11 million times.

College of Engineering Faculty Profiles

After the official launch of Esploro, the Director of Scholarly Impact was invited to attend a meeting with the Research Deans of the University during the Summer of 2021 to talk about how this might be beneficial to them. The Associate Dean of Engineering was interested in learning more and requested a meeting with the Director of Scholarly Impact and the Director of the Engineering Library in September 2021.

The College of Engineering was currently using a variety of different resources to collect faculty output data. They have been using Google Scholar IDs since around 2014 with the Engineering IT department running annual scrapes of the websites to create a report of publications by faculty for their annual reports. Prior studies have shown that articles and publications cited within Google Scholar can provide a large amount of the citations produced by the researchers. A study by Martin-Martin, et al, showed that 88% of the articles reviewed were found in Google Scholar [2]. There has been some concern around the quality of the data and information gathered through Google Scholar and whether it had been vetted to ensure all the articles are correctly assigned to their authors, which may be a problem for authors with common names. A study done by Ritchie, Banyas, and Sevin found that although Google Scholar provided the most unique content, more than half of it was judged to be not relevant to the author's research output

[3]. While using Google Scholar to gather faculty output is a great start, it would be helpful to include other sources. Google Scholar also has been proven to be difficult to allow for extraction of bibliographic data for analysis [4], [5]. There are features within Esploro to create a wide variety of reports and graphics currently not possible with Google Scholar.

They also had been using the Academic and Professional Record (APR) powered by Digital Measures, which had already sunset. All engineering faculty have profile pages on the Engineering website, but with static content on these pages being created by the individual, the information and data on the pages varied greatly.

During the September meeting, a more detailed demonstration of Esploro was given to the Associate Dean and his staff. The presentation featured how the profiles will look, possible reports that can be run on the data and an explanation of how the data is imported, including talking about the artificial intelligence and machine learning aspects of the system and how the librarians will be checking the data before importing. The dynamic nature of the data being constantly updated, along with the ability to run a wide variety of reports, resulted in the College giving us the green light to move ahead with the project.

Once a list of all 110 faculty within the college was obtained, it was time to start setting up the system.

Methods

Estimate Faculty Publication Data

The data in Digital Measures was not always complete and accurate, but it was better than no data. The University Libraries received an output of a subset of this data in the summer of 2019 so that we could add metadata records to Esploro. The file included anything that was published and flagged to be public that had a Digital Object Identifier (DOI) and/or a PubMed Identifier (PMID). We received these identifiers paired with the local faculty members ID. In 2020-2021, we added these records to Esploro by using the identifier to pull a record from Ex Libris' Central Discovery Index (CDI).

At the start of the project, we went to each of the faculty member's Google Scholar profile and if it a was a verified account connected to the University of Iowa, we scrolled to the bottom to get a count of the number of publications listed. We were unable to get a total of records for 20 faculty members. We compared the total records for each faculty member already in Esploro with the total found on their Google Scholar profile so that we could get an estimate of the completeness of our data.

Esploro has a function called "Smart Harvest" that will look for potential publications by a person and then assess the likelihood they are by the local person, using keywords, name variants, identifiers such as ORCID and email address, and previously connected publications [6]. These then can be loaded immediately or can be reviewed by staff, based on the quality of

the match. We opt to review all matches. Because the previously loaded publications inform the matching process, the records loaded from Digital Measures resulted in better quality matches in Smart Harvest and saved us a considerable amount of time. Common names, such as Chao Wang, can result in a very large number of matches, so in these cases in particular adding as many known items as possible is a great assistance.

Ten faculty members were already active in Smart Harvest. For the six faculty members with at least 85% of their titles already in Esploro, we activated them in Smart Harvest. This percentage was not as accurate as we had hoped because we neglected to omit the theses for which they were an advisor or committee member. For every faculty member with a lower percentage than that, we knew it would save considerable time to add as many more publications as possible before activating them in Smart Harvest.

Manually Identify Publication DOIs in Scopus

To get a list of publications for each faculty member, we had two options: either ask each faculty member for an output of their Google Scholar profile or search another database for their publications. We opted for the latter as a more efficient first step, using Scopus to retrieve a list of publication DOIs for each engineering faculty member. In Scopus, we searched for each engineering faculty member's name shown as in Figure 1 and retrieved several potential author profiles for the faculty member. We did not add affiliation names in the initial search because some faculty members recently joined our university, and the affiliations of their Scopus profiles might not be updated. But for faculty members who have been affiliated with our university for a while with ambiguous names, we would add affiliation names in the search to narrow down the search results.

Figure 1. Scopus author search.

After identifying several potential Scopus author profiles for the same faculty member, we would compare the publication list of each Scopus profile shown as in Figure 2 against the faculty member's Google Scholar profile to make sure the publications were authored by our faculty member. We also encountered the situation that several Scopus author profiles belonged to the same faculty member. In that case, we requested on behalf of the faculty members to merge their Scopus author profiles.

When we confirmed the Scopus author profile for a faculty member, we exported all publication records of the faculty member in the csv format shown as in Figure 3. In the exported file, we

extracted columns for DOI and PubMed ID and added them to a separate file with the faculty ALMA ID, last name and first name. We repeated the process for over one hundred faculty members.

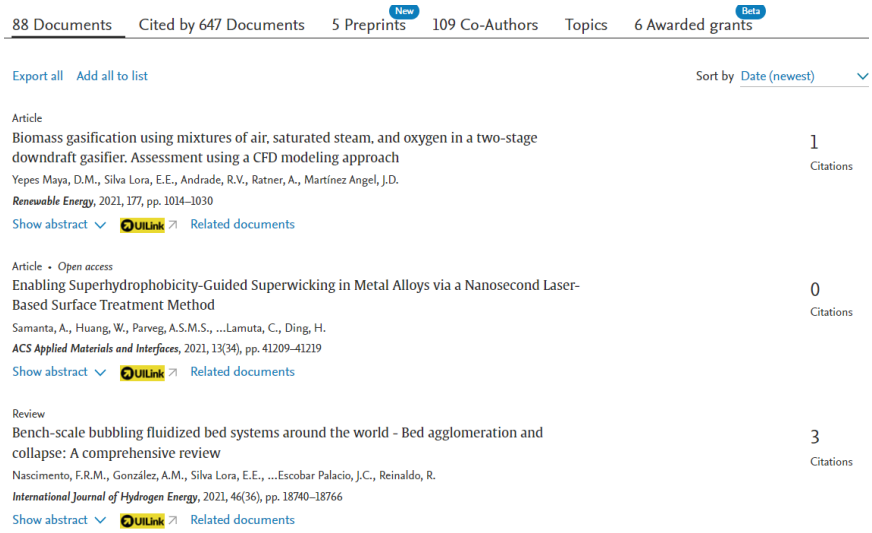

Figure 2. A list of publication records for a faculty member.

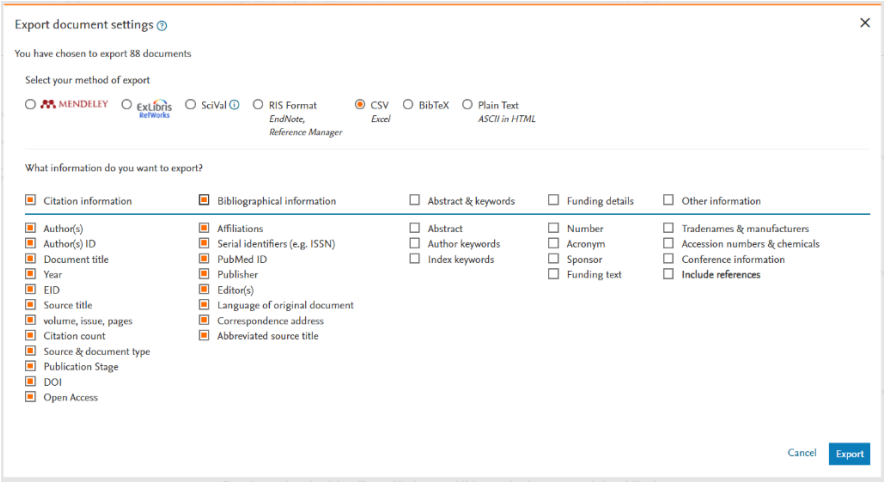

Figure 3. Exporting publication records in the csv format.

Load Publication DOIs to Esploro

The files for each person were then manually merged into a csv file for each department or for several departments. The columns we used are last name, first name, local identifier, DOI, and PMID. Cataloging-Metadata department staff uploaded these and did standard review and cleanup of the results. We should also note that there is now a process where we can add contents from CDI using title, date, and other standard bibliographic information, and not only using identifier. However, this process did not exist until we were almost done with this project.

Run Smart Harvest and Initial Review

We ran Smart Harvest in Esploro for each department separately, to capture all relevant publications based on the faculty names. In a few cases for faculty with very common names, we needed to add a last run date corresponding with their degree to reduce the number of potential matches. We processed each department separately so that we could fine tune our procedures and training as the project continued.
Esploro classifies matches with local researchers into three levels of status: uncertain match, strong match, and very strong match. A system standard author matching report was then created for each smart harvest run. This file was initially reviewed by Engineering library staff members who noted X if it was a clear match, and NO if it was clearly not a match and left the cell blank if they could not easily tell if it was a match when skimming the spreadsheet. A comment column was also added to indicate when something was by the correct person, but the asset type was incorrect (such as a conference proceeding coming in as a journal article).

Since many of the publications contain DOIs, it would be easy to identify the author information, especially the affiliation information, for the faculty's publications on the literature database, publisher's website, or the institutional repository. For the publications that do not contain DOIs, we would search the titles in Google Scholar or check the publications against the faculty's resume which can be found on the institutional website. The number of publication records reviewed can be found in Table 1. The publications from the Department of Industrial and Systems Engineering are not included because they were not processed in this way.

Table 1. The number of publication records reviewed in spreadsheets.

	Mechanical Engineering	**Electrical and Computer Engineering**	**Civil and Environmental Engineering**	**Biomedical Engineering**	**Chemical and Biochemical Engineering**
Approved, marked as X	154	191	44	76	49
Disapproved, marked as "NO"	85	77	1	460	81
Unsure	28	85	73	47	25
Total	267	353	118	583	155

After this review was completed, the rows that were marked as an X were extracted from the original sheet and saved to a new csv file. The newly added comment column on the new csv file was also removed. This csv file was used for the "update approvals" process. We then added conditional formatting in the spreadsheet to the identifier column to mark duplicate lines (the same publication will appear once for each potential author match). The original sheet was filtered to those marked "NO" and we double checked those, especially duplicated records, that should or should not be entirely deleted because there is only one record for the publication but multiple matches for co-authors. For example, if a publication turned out not to be our faculty member's work due to name ambiguity but happened to be a work of the other faculty member,

we would only disconnect the publication record with the faculty member profile instead of deleting the entire record. The system identifiers of all these rejected records were put in a text file, which was used to create a set of records. This set was processed with the “Delete Set of Research Assets” job. These steps for batch approval and batch deletion took only a few minutes total.

Further Manual Review in Author Matching Approval Task List

The remaining unsure items required more careful checks so were reviewed in the author matching approval task list as shown in Figure 4. This list can be filtered by author, strength of match, asset type, and date. The system presents information in an easier to read manner and the DOI is a clickable link. Twenty assets are displayed on a screen, and they can be approved or deleted (all or selected) as a group. Books needed to be checked carefully as they were often dissertations by the faculty members’ students. Other resources were correct, such as preprints, or abstracts, but came in with an incorrect asset type; after approval these needed to be corrected.

The affiliation of the author on the publications displays. This is particularly helpful for Engineering Library staff members who can check where the faculty members previously worked or received degrees, to confirm the match was correct.

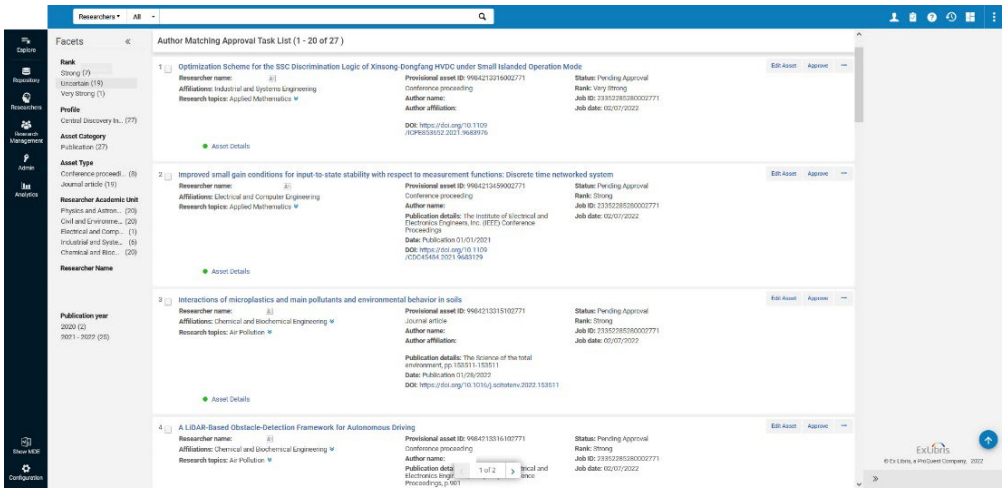

Figure 4. Screenshot of the author matching approval task list in Esploro.

Working with Faculty and Staff on Their Profiles

While the profiles were getting set up and assets were being approved, we began to have conversations with the faculty and staff to get feedback and create more complete profiles. The Associate Dean for the College of Engineering sent out a message to all the faculty near the end of November that had two goals: letting them know about this project and asking them to

connect their ORCID. Connecting their ORCIDs to the University would send an update to Esploro. The ORCID provides Smart Harvest with a good match point that is used when it appears in the data shared by publishers. The profiles were not yet live, but each faculty member could log-in to the system and see their profile page. We had several faculty members reach out after receiving this email. Luckily, there has been no negative responses to this project with the largest change requested so far is one faculty just wanting to know if we could change content on the profile/biography page that had little to do with the published works content. The faculty and staff within the College have been supportive of this project.

Outcomes

This project doubled the number of publications connected to researchers. Table 2 shows the total publications for all co-authors rather than a count of total publications. Each co-authored publication is counted for each author because the goal is to have complete publications for every author. After Smart Harvest's initial run for faculty members, the harvesting added an additional 102 author publications.

Table 2. The number of author publications by method of adding records.

Record Source	**Total**	**Percentage of Total**
Migrated from previous repository	496	4.9%
Digital Measures Import	3,944	38.9%
Scopus Import	3,891	38.3%
Smart Harvest	1,692	16.7%
Other	124	1.2%

Table 3 shows the number of records for authors, grouped by faculty rank. Because this project relied on adding materials by DOI, we missed many conference abstracts that lack a DOI. The new import by title option in Esploro would have given us better results.

Table 3. Total number of publications by author faculty rank.

	Total in Google Scholar	**Total in Esploro**	**Percentage in Esploro**
Professor+Dean	13,192	10,659	80.8%
Associate Professor	1,928	1,587	82.3%
Assistant Professor	597	460	77.1%
Lecturer	30	33	110.0%
Professor of Instruction	36	36	100.0%
Associate Professor of Instruction	25	134	536.0%

The project added formally published materials. In the terminology of Esploro, conference paper, conference poster, conference presentation, and presentation are not formal publications and are not in the CDI database. Esploro currently does not add patents. The total number of publications by asset types compared with the totals added from the commencement of discussions about the project (25 August 2021) are shown below in Table 4.

Table 4. Total publications by type for the College of Engineering.

Asset Type	Number of Publications	Total Added via Project	% of Total added via Project
Abstract	78	37	47.4%
Accepted manuscript	3	0	0.0%
Book	106	103	97.2%
Book chapter	411	298	72.5%
Conference paper	1	0	0.0%
Conference poster	6	0	0.0%
Conference presentation	7	0	0.0%
Conference proceeding	1,761	1,380	78.4%
Dataset	26	7	26.9%
Edited book	8	1	12.5%
Editorial	24	2	8.3%
Encyclopedia entry	8	5	62.5%
Journal article	6,491	2,556	39.4%
Letter/Communication	22	2	9.1%
Magazine article	3	0	0.0%
Other	27	18	66.7%
Patent	47	0	0.0%
Preprint	72	60	83.3%
Presentation	3	0	0.0%
Report	45	15	33.3%
Review	13	9	69.2%

All the records added are fully editable and there are mechanisms for researchers to contact administrators to request corrections. The records are discoverable through the Esploro search and profile interfaces, the Libraries' Primo discovery service, Google, and in the case of full-text content, Google Scholar.

Next Steps

Reviewing profiles with Faculty and Staff

Once the Smart Harvest was turned on, fifteen faculty were identified to add some additional information to their profiles to improve which materials were retrieved. All these faculty have

common names and had not linked their ORCID with the University, so the system was not as able to easily distinguish them from other researchers. We reached out to each of these researchers individually to ask them to connect their ORCID. Keywords from their research were added to their profiles to aid the system in finding affiliated publications.

A few faculty members have begun to review their publication lists. A new problem we have discovered is when a faculty member is involved in a research group that writes articles and papers (e.g., PREDICT-HD Investigators of the Huntington Study Group, their article DOI: 10.1037/a0029218). If the publisher does not include all members of the research group in their metadata, then there is no way for the indexes to identify the publication as by our faculty member. In those cases, we will need to seek out publications in a different manner and add them when a faculty member notifies us.

Next, the Associate Dean of the College and the librarians will be holding a college wide session so the faculty can learn how to update the biography information, request edits, removal, and adding publications and other works to their profiles. This will also be the time to answer any questions they may have.

College of Engineering Uses

When finished, this project will provide a database of all research products (journal articles, books, chapters, etc.) in one place. This database will hopefully one day include additional patents as well. The bibliographic information within these profiles will allow the college or researchers to create data and graphical displays of their research in ways that are not feasible using Google Scholar.

These reports can be included in annual reports, newsletters, or magazines showing faculty output over the last year or any period. They can be broken down by department or a wide range of other parameters. In addition to the external reports, the data can also be used to highlight new work or publications for the internal newsletters or other communications to increase awareness of research done throughout the college.

Esploro can import citation statistics for publications which will allow for even more facets to measure faculty output. Hopefully, one day this will be automatically added.

Conclusions

Starting new faculty profiles for the entire college can be a challenging task, and it may be up to chance and resources in terms of having an existing infrastructure, developing a team, and establishing the workflow. Issues of the number and quality of publication records, the functionality of Esploro's Smart Harvest, library staffing capacity, and administrative support

can make a difference in how the project is conducted as well as its overall success. We hope that the workflow provided above will help academic librarians who also have Esploro at their libraries consider how their libraries can approach the project of faculty publication profiles.

References

[1] ExLibris. “Esploro customer community: Discover Esploro customers around the world.” ExLibrisGroup.com. https://exlibrisgroup.com/products/esploro-research-services-platform/esploro-customer-community/ (accessed March 21, 2022).

[2] A. Martin-Martin, M. Thelwall, E. Orduna-Malea, and E. D. Lopez-Cozar, "Google Scholar, Microsoft Academic, Scopus, Dimensions, Web of Science, and OpenCitations' COCI: A multidisciplinary comparison of coverage via citations," (in English), Scientometrics, vol. 126, no. 1, pp. 871-906, Jan. 2021, doi: https://doi.org/10.1007/s11192-020-03690-4.

[3] S. Ritchie, K. Banyas, and C. Sevin, "A comparison of selected bibliographic database search retrieval for agricultural information," Issues in Science and Technology Librarianship, no. 93, 2019, doi: https://doi.org/10.29173/istl48.

[4] E. D. Lopez-Cozar, E. Orduna-Malea, and A. Martin-Martin, "Google Scholar as a data source for research assessment," (in English), Springer Hbk, pp. 95-127, 2019, doi: https://doi.org/10.1007/978-3-030-02511-3_4.

[5] G. Halevi, H. Moed, and J. Bar-Ilan, "Suitability of Google Scholar as a source of scientific information and as a source of data for scientific evaluation—Review of the Literature," Journal of Informetrics, vol. 11, no. 3, pp. 823-834, Aug. 2017, doi: https://doi.org/10.1016/j.joi.2017.06.005.

[6] ExLibris. "Esploro smart harvesting framework." ExLibrisGroup.com. https://knowledge.exlibrisgroup.com/Esploro/Product_Documentation/Esploro_Online_Help_(English)/Esploro_Smart_Harvesting_Framework (accessed January 31, 2022).