FOSTERING CO-RESPONSIBILITY FOR OPEN METADATA QUALITY TO EVALUATE AND MONITOR OPEN SCIENCE

DOI: 10.5281/zenodo.12604738

Alicia Fátima Gómez *IE University*. Madrid (Spain)

Cristina Huidiu. *Wageningen University & Research*. Wageningen (The Netherlands)





OPFN MFTADATA WITHOUT TRANSPARENT MFTADATA OUALITY IS AN INCOMPLETE SOLUTION



______ = modifier_ob mirror object to mirror mirror_mod.mirror_object peration == "MIRROR_X": irror_mod.use_x = True mirror_mod.use_y = False irror_mod.use_z = False Operation == "MIRROR_Y" irror_mod.use_x = False lrror_mod.use_y = True irror_mod.use_z = False operation == "MIRROR_Z" irror_mod.use_x = False lrror_mod.use_y = False rror_mod.use_z = True election at the end -add ob.select= 1 er ob.select=1 ntext.scene.objects.action

Selected" + str(modific irror_ob.select = 0 bpy.context.selected_ob ita.objects[one.name].sel

mint("please select exactly

OPERATOR CLASSES -----

x mirror to the selecter pect.mirror_mirror_x" ror X"

ontext):
 context.active_object is not

WHO DO WE INFLUENCE

5

• Researchers' careers



• Funding strategies



Research & Innovation







(SOME) OPEN SCIENCE INITIATIVES

I40C 140A Funders opening calls to support open initiatives Open grants data OpenAIRE Datacite **Open Research Europe (ORE)**



OPEN SCIENCE IS A JOURNEY





- Research reproducibility
- Shorten innovation time
- Contributing to economic growth

AND

Fundamentally changed how we think about quality, bringing in new challenges:

- Predatory journals, conferences, books...
- Fake metrics
- Need for more complex research assessment
- Increased focus on metadata quality



For 470,000 records with errors:



COST OF BAD DATA

HOW CAN WE HELP IMPROVE OPEN DATA?

| | WoS | Scopus | OpenAlex |
|------------------------------|---------------|---------------|---------------|
| Whole Corpus | | | |
| Number of Records | 71,280,830 | 65,642,377 | 243,053,925 |
| Number of References | 1,765,281,799 | 2,033,522,623 | 1,845,379,285 |
| Whole Corpus - Articles Only | | | |
| Number of Records | 42,678,632 | 43,579,595 | 200,665,940 |
| Number of References | 1,400,958,343 | 1,422,650,789 | 1,636,497,394 |
| Published 2015-2022 | | | |
| Number of Records | 22,609,069 | 27,620,472 | 76,836,191 |
| Number of References | 786,437,547 | 1,035,750,923 | 840,730,834 |
| Shared Corpus (2015-2022) | | | |
| Number of Records | 16,788,282 | 16,788,282 | 16,788,282 |
| Number of References | 725,008,043 | 727,056,725 | 585,616,069 |

Table 1: Sizes of databases and of the Shared Corpus dataset, with the number of references in each

Reference Coverage Analysis of OpenAlex compared to Web of Science and Scopus



EXAMPLE 1:

- In a dataset of 405 publications in IEEE (Journals and Conference Proceedings) with TU Wien affiliation from 2021-2022 (Stand Aug.) the affiliation was missing for 118 of them (29 %), unconsistently across individual publications. For the items published in 2022, data looked better, including affiliations for 75 %.
- It was a similar issue for articles published in main publisher journals (SpringerNature, Elsevier, etc.).



EXAMPLE 2:

- In January 2023 we found that in a dataset of 652 publications with TU Wien affiliation from 2018-2022 (as of Aug.) the affiliation information in Crossref was missing for all of them.
- Metadata included in MDPI was:
 - Journal metadata (title, ISSN and CODEN (where applicable)).
 - Article metadata
 - Publication type. Publication date. DOI data (including URL to article page)
 - Volume & Issue number. Page or article numbers. Title, Abstract.
 - Author details (names, and ORCIDS (where applicable))
 - Funding details (funder name/s and grant number/s)
 - References (general structure: author name, article title, year of publication, volume # and page number)



LIBRAR

This is to let you know that we have now included affiliation information in our Crossref deposits starting this week. We are still considering if we are able to update back issues with this information (in terms of capacity) and for which time-frame.

We'd like to thank you for your valuable to suggestion to improve the metadata we provide to Crossref.



https://api.crossref.org/works/ 10.3390/publications11010004

MDPI was not depositing any affiliation information at Crossref 🛞

Bianca Kramer Catriona J. Maccallum Paolo Manghi Comeron Nevin Silvio Paroni David Shottor Aaron Tay Ludo Waltman

October 27th, 2021

8

0

0

0

in

44

Shares

Now is the time to work together toward open infrastructures for scholarly metadata

As part of Open Access Week 2021, Ginny Hendricks, Bianca Kran

Paolo Manghi, Cameron Nevlon, Silvio Peroni, David Shotton, Aar the case for community action toward open infrastructures for sc the impending loss of Microsoft Academic, the need for more su contributions these can make to research equity, they outline how scholarly communications ecosystem can contribute to making a

Recent events highlight the importance and the fragility of infrast communication. The \$4.5 million grant awarded to OurResearch, to Invest in Open Infrastructure (IOI), alongside the development welcome. The news in May 2021 that Microsoft Academic was in

Microsoft Academic has been one of the key players providing m publications. Metadata on authors, affiliations, abstracts, citatio importance for scholarly literature search and research assessm making this data openly available (even though the license condit kinds of re-use allowed). It's heartening that OurResearch (of Ung through OpenAlex, to continue producing open metadata about so building on the work done by Microsoft, However, the forthcomin demonstrates the fragility of infrastructures that do not meet app governance, whether they are provided by large financially secure small grant-funded academic initiatives

66

The very reason that IOI and related as SCOSS exist, is an acknowledgeme vulnerability of open infrastructures

00

Mapping bibliographic metadata collections: the case of **OpenCitations Meta and OpenAlex**

Elia Rizzetto 1, Silvio Peroni

Research Centre for Open Scholarly Metadata, Department of Classical Philology and Italian Studies, University of Bologna, Bologna, Italy

Abstract

This study describes the methodology and analyses the results of the process of mapping entities between two large open bibliographic metadata collections, OpenCitations Meta and OpenAlex. The primary objective of this mapping is to integrate OpenAlex internal identifiers into the existing metadata of bibliographic resources in OpenCitations Meta, thereby interlinking and aligning these collections. Furthermore, analysing the output of the mapping provides a unique perspective on the consistency and accuracy of bibliographic metadata, offering a valuable tool for identifying potential inconsistencies in the processed data.

Keywords

Bibliographic collection, entity mapping, OpenCitations, OpenAlex

1. Introduction

Open bibliographic metadata collections play a pivotal role in enabling reproducible studies in the fields of bibliometrics, scientometrics and science of science and permit transparent procedures in the context of research assessment exercises, thus enabling the implementation of norms and guidelines that intend to reform the research assessment around the world, such as the Coalition for Advancing Research Assessment (CoARA2). As the volume and diversity of scholarly publications continue to expand, the need for comprehensive and interoperable bibliographic databases becomes increasingly pronounced.

This study delves into the process of mapping entities between two important open bibliographic metadata collections, OpenCitations Meta [1] and OpenAlex [2]. These mapping processes are a critical step towards enabling researchers, institutions, and platforms to access and utilise information seamlessly across diverse collections. In our work, the primary objective of this mapping is to integrate OpenAlex internal identifiers into the existing metadata of bibliographic resources (BRs) in OpenCitations Meta, thereby interlinking and aligning these collections. This paper presents the results of the mapping and provides details on the methodology adopted to accomplish this task. By shedding light on the complexities inherent in aligning bibliographic metadata collections, we aim to contribute valuable insights into the challenges and opportunities associated with such endeavours.

Furthermore, the study investigates the mapping process's implications to assess the quality of the involved datasets. Analysing the output of the mapping provides a unique perspective on the consistency and accuracy of bibliographic metadata, offering a valuable tool for identifying potential inconsistencies in the processed data. The importance of such considerations lies in their capacity to enhance data quality, fortify interoperability, and foster a more cohesive scholarly metadata landscape.

The rest of the paper is structured as follows. In Section "Material and methods", we introduce the processed data and the mapping methodology. Then, in Section "Results", we present the result of the

120° IRCDL (The conference on Information and Research science Connecting to Digital and Library Science, February 22-23, 2024, Bressanone, Italy EMAIL: elia.rizzetto@studio.unibo.it (E. Rizzetto); silvio.peroni@unibo.it (S. Peroni)

ORCID: 0009-0003-7161-9310 (E. Rizzetto); 0000-0003-0530-4305 (S. Peroni)

2https://coara.eu



Reference Coverage Analysis of OpenAlex compared to Web of Science and Scopus

Jack H. Culbert¹, Anne Hobert², Najko Jahn², Nick Haupka², Marion Schmidt³, Paul Donner³, Philipp Mayr⁹¹

¹GESIS – Leibniz Institute for the Social Sciences, Germany. ²Niedersächsische Staats- und Universitätsbibliothek Göttingen, Georg-August-Universität, Göttingen, Germany. ³Deutsches Zentrum für Hochschul- und Wissenschaftsforschung, Berlin, Germany.

Contributing auth

OpenAlex is a promising open sou the established proprietary sources provides its data freely and openly ric studies that can be reproduced However, as OpenAlex is a rapidly is expanding and also quickly char trustworthiness of its data. In this and metadata coverage within each to help address this open question demonstrate that, when restricted lications shared by all three datab numbers and internal coverage, res and Scopus. We also demonstrate th ered by OpenAlex shows mixed re identifiers, fewer abstracts and a size article when compared to both We

2024

Feb

26

[cs.DL]

arXiv:2401.16359v2

Keywords: Bibliometrics, Open Scho Abstracts OBCID

> In this article, we focus on the importance of open research information as the foundation for transparent and responsible research assessment and discovery of research outputs. We introduce work in which we support the open research information commons by enabling, in particular, independent and small Open Access journals to provide metadata to several open data hubs (Open Citations, Wikidata, Open Research Knowledge Graph). In this context, we present The OPTIMETA Way, a means to integrate metadata collection, enrichment, and distribution in an effective and quality-ensured way that enables uptake even amongst small scholar-led publication venues. We have designed an implementation strategy for this approach in the form of two plugins for the most widely used journal publishing software, Open Journal Systems (OJS). These plugins collect, enrich, and automatically deliver citation metadata and spatio-temporal metadata for articles. Our contribution to research assessment and discovery with linked open bibliographic data is threefold. First, we enlarge the open research information data pool by advocating for the collection of enriched, user-validated metadata at the time of publication through open APIs. Second, we integrate data platforms and journals currently not included in the standard scientometric practices because of their language or lack of support from big publishing houses. Third, we allow new use cases based on location and temporal metadata that go beyond commonly used discovery features, specifically, the assessment of research activities using spatial coverage and new transdisciplinary connections between research outputs.

Enriching the scholarly metadata commons with citation metadata and

spatio-temporal metadata to support responsible research assessment

and research discovery

Keywords: open research information, citation metadata, spatio-temporal metadata, scholarly

Daniel Nüst1*, Gazi Yücel2, Anette Cordts2, Christian Hauschke2

2TIB - Leibniz Information Centre for Science and Technology, Hannover, Germany

1Institute for Geoinformatics, University of Münster, Germany

* Correspondence: Daniel Nüst, daniel.nuest@uni-muenster.de

publishing, research assessment, discovery

Abstract



METADATA BEST PRACTICES

| Documentation | | | | |
|--|----------------------|---|---|------------------|
| Home > Documentation > Metadata principles and | practices > Me | tadata best practices | Q Search this section | |
| Documentation | Met | adata best practices 🖉 | | |
| Setting up as a member | < Met | adata principles and practices | Abstracts > | |
| The research nexus | Best p | ractices, like principles, are aspirational for our membe | ers, but we'll do our best to help you meet them. | |
| Metadata principles and practices | Our sy | stems, schema, and practices have evolved over time | and, as with many organizations, we need to | |
| Metadata best practices | assess | s that honestly with a goal of meeting the best practice | e in the future. | |
| Abstracts | Cross | ef metadata requirements for content registration are | minimal but meeting minimum requirements only | |
| Bibliographic metadata | means collect | that you have succeeded in registering your content a is recommended to improve discoverability and connection | and DOIs. Most of the optional metadata we ect content persistently to the scholarly record. | |
| Funding metadata | There | are nuances and best practices for both different type | es of content and different types of metadata. | |
| License metadata | Best p | ractices are available for the following: | | |
| Multi-language material and | Best p | ractices for key metadata elements Best r | practices for key record types | |
| uansiations | • <u>Ab</u> | Contributors | | |
| References | • <u>Bib</u> | Contributor metadata is expressed consiste | ently across record types (excluding Grants), an | d includes |
| Relationship metadata | | contributor names, roles, identifiers, alterna | te names, and affiliation information. A contribu | utor is a single |
| Version control, corrections, and | : | person or a group of people/organization th | at has contributed in some way to the content l | being registered |
| retractions | • <u>Fu</u> r | Do includo: | | |
| Books and chapters | • <u>Lic</u> • Mu | bo include. | | |
| Conference proceedings and papers | • Rei | correct names, so authors and other con | ntributors can be matched to citations | |
| Datasets | • Vei | a complete contributor list so that contri | butors can receive credit for their work, and to | help make your |
| | | Content more discoverable Contributor rolo(s) at least one for each | h contributor, but cupply as many as apply | |
| | | OPCID iDc, so that authors can be disam | biguated and connected to the recearch they w | write and cupper |
| | | Affiliations and ROR IDs so that contribut | tor institutions can be identified and research o | utputs can be |
| | | traced by institution | | |

https://www.crossref.org/documentation/principles-practices/best-practices/

UNIVERSITY

LIBRARY

WAGENINGEN

DataCite Metadata Schema Documentation Version 4.5 DataCite Metadata Working Group Mar 12, 2024 Contents 1 Contents 2 1.1 Introduction 1.2 DataCite Metadata Properties Note: DataCite Metadata Working Group. (2024). DataCite Metadata Schema for the Publication and Citation of Research Data and Other Research Outputs. Version 4.5. DataCite e.V. https:// org/10.14454/g8e5-6293 2. Creator Obligation: Mandatory Contributors from the DataCite Metadata Working Group: Occurrences: 1-n Jan Ashton, British Library (co-chair of working group) Definition: The main researchers involved in producing the data, or the authors of the publication, in priority order. For instruments this is the manufacturer or developer of the instrument. To supply Isabel Bernal, Spanish National Research Council (CSIC) (cd multiple creators, repeat this property. Allowed values, examples, other constraints: Felix Burger, TIB May be a corporate/institutional or personal name. Note: DataCite infrastructure supports up to · Madeleine de Smaele, TU Delft Library 10,000 names. For name lists above that size, consider attribution via linking to the related metadata. Samantha Foulger, ETH Zurich Sub-properties. · Vanessa Gabriel, University Library of the LMU Munich 2.1 creatorName Ted Habermann, Metadata Game Changers - 2.1.a nameType Joseph Padfield, The National Gallery 2.2 givenName Sarah Ramdeen, Columbia University 2.3 familyName 2.4 nameldentifier 2.4.a nameldentifierScheme 1 2.4.b schemeURI • 2.5 affiliation

2.5.a affiliationIdentifie

2.5.b affiliationIdentifierSchem

https://schema.datacite.org/meta/kernel-4.5/

SOME IMPLICATIONS

Our sources (OCI, https://identifiers.org/oci) DOI DOI NIH tional Institute of Health NIH-OCC OpenAIRE JaLC INDEX PMID DOI. PMID. Crossr

OpenCitations Index is a bibliographic index recording citations between publications, using the data available in particular bibliographic databases, where such citations are treated as first-class data entities, with accompanying properties, and are identified by an Open Citation Identifier

OpenCitations Meta is a database that stores and delivers bibliographic metadata (title, type, venue, volume, issue, pages, publication date, identifiers, authors, editors, publishers) for all publications involved in the OpenCitations Index – every entity in OpenCitations Meta is assigned persistent internal identifier called OpenCitations Meta Identifier (OMID, https://identifiers.org/omid)

| | INDEX | Crossref | DataCite | PubMed | OpenAIRE |
|----------|---------------|---------------|-------------|-------------|------------|
| INDEX | 1,975,157,320 | 1,563,218,160 | 169,814,412 | 695,988,810 | 14,645,838 |
| Crossref | | 1,100,964,483 | 27,051 | 458,309,297 | 3,917,329 |
| DataCite | | | 169,663,255 | 9,623 | 114,483 |
| PubMed | | | | 227,958,101 | 9,711,789 |
| OpenAIRE | | | | | 902,237 |

Peroni, S. (2023, noviembre 29). Assessing research needs data transparency - or the OpenCitations' vision for an open citation and bibliographic data ecosystem. https://doi.org/10.5281/zenodo.10222549



PMC, ARXIV

Peroni, S. (2021, noviembre 19). Current and future sources of open citations. Austrian DataCite Consortium 2021. https://doi.org/10.5281/zenodo.5713191

How to be indexed in DOCI and Meta

- High-quality citation data necessary
- Citation links via relatedIdentifiers, having relationType equal to either "References" or "Cites" and relatedIdentifierTvpe equal to "doi'
- Metadata gathered for OpenCitations Meta are: identifiers, title, authors, publication date, venue, volume, issue, page, type, publisher, editor

| 14: | "18.4230/lipics.stacs.2020.42" |
|-------------------------------|---|
| type: | "dois" |
| attributes: | |
| doi: | "18.4230/lipics.stacs.2020.42" |
| identifiers: | 0 |
| ► creators: | L1 |
| titles: | EI . |
| publisher: | "Schloss Dagstuhl - Leibniz-Zentrum für Informatik" |
| container: | 0 |
| publicationYear: | 2028 |
| subjects: | L1 |
| <pre>> contributors:</pre> | L-1 |
| dates: | D |
| Language: | "en" |
| types: | ω |
| relatedIdentifiers: | |
| P 8: | (_) |
| 1: | (_) |
| ▶ Z: | () |
| * 3: | |
| relationType: | "References" |
| relatedIdentifier: | "https://doi.org/10.4230/LIPIcs.ICALP.2017.31" |
| relatedIdentifierType: | "doi" |

Other paths of collaboration

Help in developing our open source software and services

Curatorial involvement to improve OpenCitations data - we will work on a graphical tool to facilitate this (medium-term plan)

Collaborations with other like-minded infrastructures to develop federated access to open scholarly information of all types, thereby returning control over such information to the academic community that generated it in the first place



Learn more

crossref.org/members/prep/

Richer metadata makes content useful. Make sure your work can be found.

Crossref The content of this site is licensed under a Creative Commons Attribution 4.0 International License

Privacy



REPORT

https://www.crossref.org/documentation/reports/participation-reports/







GOOD EXAMPLES



"RESEARCHERS WHO DO NOT HAVE ORCID IDS INCLUDED IN THEIR CROSSREF METADATA RISK NOT BEING COUNTED IN THESE AUDITS AND REPORTS."

CALLING FOR MORE AWARENESS

| Participation Reports | | | Crossref |
|---|-------------------------------|----------------|---|
| Find a member | Learn more | | |
| | | | |
| Richer metadata makes content useful. Make sure your work can be found. | Colegio Oficial de Enfermeria | de Madrid | 55 Total registered content items |
| | | | |
| Content type: Journal articles | | | 0 |
| | | | Journal articles 55 |
| Journal articles 🗸 🔍 Search by title | | | Current content |
| References | ORCID IDs | Funder Regist | ry IDs |
| 0% | 0% | | 0 % |
| Funding award numbers | Crossmark enabled | Text mining Uf | () RLs |
| 0 % | 0 % | | 100% |
| | Similarity Check LIRLs | Abstracts | 0 |
| 0% | 100 % | | 100 % |

https://www.crossref.org/members/prep/39727

https://www.crossref.org/members/prep/7372

| Find a member | Learn more | | |
|--|-------------------------|--------------|--|
| | | | |
| Richer metadata makes content usef Make sure your work can be found. | ul. University of Dubro | ovnik | 179 Total registered content items |
| Content type: Journal articles | | | |
| | | | Journal articles 17 |
| Journal articles 🗸 🔍 Search | by title | | Current content V |
| References | ORCID IDs | Funder | () Registry IDs |
| 0% | | 0% | 0 % |
| | Crossmark enabled | 🕖 Text mi | ning URLs |
| Funding award numbers | | | 0% |
| Funding award numbers | | 0 % | |
| Funding award numbers | | | |



https://www.crossref.org/members/prep/22598



MONITOR OPEN SCIENCE



https://coara.eu/agreement/the-agreement-full-text/

JNIVERSITY

LIBRARY

WAGENINGEN



OPEN SCIENCE MONITOR

UPDATED METHODOLOGICAL NOTE

https://www.makingspeechestalk.com/ch/Open_Science_Monitor/

Propose new trends

BARCELONA DECLARATION ON OPEN RESEARCH INFORMATION

- 1. We will make **openness the default** for the research information we use and produce
- 2. We will work with **services and systems that support and enable** open research information
- 3. We will support the **sustainability of infrastructures** for open research information
- 4. We will support **collective action** to accelerate the transition to openness of research information





stranded at airports and shocked the scientific community. When Biden took office in 2021, he quickly overturned the ban. And he has taken other steps to reform immigration for professionguidance for workers in science, technology seeking visas to come to the United States. This has increased the number of STEM visas being issued, according to the agency,

But one area in which the candidates' By Sarah Wild policies have more closely aligned - and are unlikely to change - is relations with China. In 2018, under Trump, the US Department of Justice launched the China Initiative, a programme meant to safeguard US laboratories and businesses against espionage. The inititive and ended it, arguing that the programme Sch. Commun. 12, eP16288; 2024). had been perceived as using racial profiling to achieve its aims. Biden nonetheless con-research relies on the chain of footnotes," tinued with reforms introduced by Trump explains author Martin Eve, a researcher that required US universities and research in literature, technology and publishing at namely, that it tracked only articles with DOIs, organizations that were awarded more than Birkbeck, University of London, "If you can't and that it did not search every digital repos-\$50 million ner year in federal research verify what someone else has said at some itory for articles (he did not check whether funding to prove that they have instituted other point, you're just trusting to blind faith items with a DOI were stored in institutional a research-security programme, including for artefacts that you can no longer read repositories, for example). tougher scrutiny of foreign travel.

Such policies have made US institutions decrease under Biden.

Last year. Republican lawmakers in the US - are unique fingerprints used to identify and House of Representatives wrote that it had link to publications, such as scholarly articles that if you have a DOI, it's there forever," says been "unwise" of the Biden administration to and official reports. Crossrefis the largest DOI Mikael Laakso, who studies scholarly pubend the China Initiative, sparking fear among registration agency, allocating the identifiers civil-liberties advocates that they would try to about 20,000 members, including publishto reinstate the programme. They hope that ers, museums and other institutions. a renewed Biden administration would stave off such efforts, but aren't sure what would happen under a second Trump term.

"Relations with China won't improve in the foreseeable future, but they could get worse," says Jenny Lee, a higher-education researcher and vice-president for international affairs at the University of Arizona in Tucson.

The elections in November will undoubt edly affect government policies on many scientific issues. But for Barron, as for many others, science is just one of many concerns that he has about a potential second term for Trump, who has questioned the legitimacy of the 2020 election, promoted misinformation on a number of fronts, and signalled that he would institute new rules that critics argue would make it easier to fire career government employees opposing his politics, "I would put myself in the camp that is most worried about Two million articles are not properly archived. democracy," Barron says.



engineering and mathematics (STEM) who are Analysis of DOIs suggests digital preservation isn't keeping up with burgeoning scholarly knowledge.

The sample of DOIs included in the study was made up of a random selection of up to 1.000 ore than one-quarter of scholarly registered to each member organization. Twenarticles are not being properly ty-eight per cent of these works - more than archived and preserved, a study two million articles - did not appear in a major of more than seven million digdigital archive, despite having an active DOI ital publications suggests. The Only 58% of the DOIs referenced works that had ative led to a number of arrests of scientists findings indicate that systems to preserve been stored in at least one archive. The other with Chinese heritage, but when Biden took papers online have failed to keep pace with 14% were excluded from the study because they office, his administration reviewed the initia the growth of research output (M. P. Eve J. Libr. were published too recently, were not journal articles or did not have an identifiable source. "Our entire epistemology of science and

Preservation challenge

Eve notes that the study has limitations

Nevertheless, preservation specialists have Eve, who is also involved in research and welcomed the analysis. "It's been hard to know wary of collaborating with scientists in China. development at digital infrastructure organ the real extent of the digital preservation experts say. And in fact, studies have shown ization Crossref, checked whether 7,438,037 challenge," says William Kilbride, managing that scientific collaborations between the works labelled with digital object identifiers director of the Digital Preservation Coalition. United States and China have continued to (DOIs) are held in archives. DOIs - which con-beadquartered in York, UK, which publishes a sist of a string of numbers, letters and symbols handbook of good preservation practice. "Many people have the blind assumption

lishing at the Hanken School of Economics in Helsinki, "But that doesn't mean that the link will always work." Kate Wittenberg, managing director of the

digital archiving service Portico in New York City, warns that small publishers are at higher risk of failing to preserve articles than are large ones. "It costs money to preserve content," she says, adding that archiving involves infrastructure, technology and expertise that many smaller organizations do not have access to. Eve's study suggests some measures that could improve digital preservation, including stronger requirements at DOI registration agencies and better education and awareness of the issue among publishers and researchers

"Everybody thinks of the immediate gains they might get from having a paper out somewhere, but we really should be thinking about the long-term sustainability of the research ecosystem," Eve says, "After you've been dead for 100 years, are people going to be able to

UNIVERSITY WAGENINGENUR LIBRARY

https://asistdl.onlinelibrary.wiley.com/doi/10.1002/asi.24460

Received: 3 September 2020 Revised: 28 December 2020 Accepted: 25 January 2021 DOI: 10.1002/asi.24460

RESEARCH ARTICLE



Open is not forever: A study of vanished open access journals

Mikael Laakso¹ | Lisa Matthias² | Najko Jahn³

¹Information Systems Science, Hanken School of Economics, Helsinki, Finland ²Department of Political Science, John F. Kennedy Institute, Freie Universität Berlin, Berlin, Germany ³Göttingen State and University Library, University of Göttingen, Göttingen, Germany

Correspondence

Mikael Laakso, Information Systems Science, Hanken School of Economics Arkadiankatu 22, 00100 Helsinki, Finland Email: mikael.laakso@hanken.fi

Abstract

The preservation of the scholarly record has been a point of concern since the beginning of knowledge production. With print publications, the responsibility rested primarily with librarians, but the shift toward digital publishing and, in particular, the introduction of open access (OA) have caused ambiguity and complexity. Consequently, the long-term accessibility of journals is not always guaranteed, and they can even disappear from the web completely. The focus of this exploratory study is on the phenomenon of vanished journals, something that has not been carried out before. For the analysis, we consulted several major bibliographic indexes, such as Scopus, Ulrichsweb, and the Directory of Open Access Journals, and traced the journals through the Internet Archive's Wayback Machine. We found 174 OA journals that, through lack of comprehensive and open archives, vanished from the web between 2000 and 2019, spanning all major research disciplines and geographic regions of the world. Our results raise vital concern for the integrity of the scholarly record and highlight the urgency to take collaborative action to ensure continued access and prevent the loss of more scholarly knowledge. We encourage those interested in the phenomenon of vanished journals to use the public dataset for their own research.

1 | INTRODUCTION

The preservation of the scholarly record requires sustained and direct action, which begins with the question of responsibility. Preservation refers to a set of activities to ensure the long-term accessibility and usability of original material, such as environmental control, disaster planning, storage and handling, digitization, and maintenance of digitally stored material (American Library Association, 2008; Northeast Document Conservation Center, 2015; UNESCO/UBC Vancouver Declaration, 2012). Library collections of printed academic journals and books secure long-term access through physical copies, but the shift

from analog to digital gave rise to uncertainty as to who is responsible for preserving scholarly literature in electronic formats-publishers, libraries, authors, or academic institutions (Day, 1998; Fenton, 2006; Johnson, Watkinson, & Mabe, 2018; Meddings, 2011; Moulaison & Million, 2015; Science Europe, 2018; Waters, 2005). This ambiguity can be dangerous since electronic resources are vulnerable to various threats, such as hardware or software failure, natural disasters, or economic failure. If there is no general agreement whose responsibility it is to preserve electronic resources, no one will be responsible, and we risk losing large parts of the scholarly record due to inaction. Exactly how much digital journal content has already been lost is

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited

@ 2021 The Authors, Journal of the Association for Information Science and Technology published by Wiley Periodicals LLC on behalf of Association for Information Science and Technology

J Assoc Inf Sci Technol. 2021;72:1099-1112



yourself."

get access to the things you've worked on?'

https://www.nature.com/articles/d41586-024-00616-5

CHALLENGES

- Open ≠ free €
- Support small journals to improve metadata availability in Crossref and other open platforms
- Data quality at scale
- Entity linkability
- Automated feedback loops
- Define research in open science and metrics with more practical aims (nice results are not reliable if the data are not accurate)



RESPONSIBILITIES OF DIFFERENT ACTORS ARE SHIFTING

Responsibilities of Universities and research institutes

Responsibilities of open infra providers

Responsibilities of publishers Responsibilities of commercial infra providers

Policy makers and funders



FOUR PILLARS FOR IMPROVING METADATA QUALITY

- **Collaboration**: Engage stakeholders (researchers, publishers, research organisations, etc.).
- **Standardization**: Adopt common metadata standards and common identifiers.
- Automated Curation: Use tools to validate and enhance metadata in a transparent and automated way.
- Education: Raise awareness about the impact of metadata.



The very reason that IOI and related initiatives such as SCOSS exist, is an acknowledgement of the current vulnerability of open infrastructures

66

99

Now it's time to work together toward open infrastructures for scholarly metadata. LSE Blog, 27 October 2021



THANKS FOR THE ATTENTION

For further questions:

Alicia Fátima Gómez IE University Library Madrid (Spain) <u>alicia.gomez@ie.edu</u> Cristina Huidiu Wageningen University & Research Wageningen (The Netherlands) <u>cristina.huidiu@wur.nl</u>



