OPEN ACCESS   CC BY

THE ROLE OF PERSISTENT IDENTIFIERS IN E-SCIENCE

Rodica Cujba[1, 2], ORCID ID: 0000-0002-7982-6184

[1]*Technical University of Moldova, 168, Stefan cel Mare Blvd., MD-2004, Chisinau, Republic of Moldova*
[2]*Information Society Development Institute, 5A, Academiei Str., MD-2028, Chisinau, Republic of Moldova*
rodica.cujba@adm.utm.md, rodica.cujba@idsi.md

Abstract. The transformations in science produced by the development of ICTs are conceptualized in the model of e-Science. Information on the researchers' activities in digital form is collected in several systems outside the affiliated institutions. Manuscript submission systems, grant funding applications, data centers, citation indexes, other institutional or disciplinary repositories and personal web pages are all important sources of information. Persistent identifiers allow to discover and collect this information and provide the ability to compare, analyze and combine data with greater efficiency and accuracy. The article provides an overview of the most popular persistent identifiers used in e-Science and describes their role.

Keywords: arXiv ID, CODEN, DOI, Google Scholar Citation Profile, ISBN, ISSN, ORCID ID, PMID and PMCID, ResearcherID, Scopus Author ID.

Introduction

Information and communication technologies (ICTs) have influenced and continue to influence significantly all human spheres of activity. Since the 90's of the XXth century the Internet has become an indispensable tool not only for the scientific world where it came from, but also for business, recreation, media.

The information age or the knowledge era in which we live today is characterized by the reproduction and exchange of information and knowledge. Science becomes global. Collaboration and availability in time of information and knowledge become the main factor in the success of scientific research. The transformations in science produced by the development of ICTs are conceptualized in the notion or model of e-Science [1, 2] or Cyberscience [3, 4].

Jankowski notes that e-Science offers "added value" to science, and includes: international collaboration between researchers; increased use of interconnected computers using Grid architecture; data visualization; development of Internet-based tools and procedures; creating virtual organizational structures for conducting research; electronic distribution and publication of results [1, p. 551].

According to Liberati, e-Science is the activity of cooperation of scientists with different competencies, located in different places, but interconnected using ICT tools, in order to reach a common scientific objective [2, p. 616].

Nentwich's opinion is that Cyberscience encompasses the scientific research activities in the virtual space that are realized on the basis of two fundamental components: the digital network and high performance computers [3, p. 4].

Hine develops Nentwich's idea by saying that, unlike cyber space that cannot replace real life, Cyberscience compliments and transforms the traditional way of conducting scientific research [4, p. 37].

Information about researchers' activities in digital form is gathered in multiple systems outside researchers' home institutions. Manuscript submissions systems, grant funding applications, datacenters, citation indices, other institutional or disciplinary repositories and personal webpages are all vital sources of information. Persistent Identifiers (PIDs) enable the discovery and collection of this information, and ensure that data can be compared, matched and combined with greater efficiency and accuracy. This is vital now, as institutions work to understand their entire research portfolio. PIDs can help make research reporting more efficient and can help to demonstrate compliance with policies for research data management and open access.

PIDs uniquely identify entities within the research ecosystem and help define relations between contributors, research artefacts, and organizations. They can be used to link across disciplines and infrastructures to build clearer pictures of how research is generated. They not only provide connections between authors and articles, but also between datasets, funders and institutions, enabling better attribution and information on where research originated. PIDs are a vital part of e-infrastructure that enable open and reproducible science and enhance the exchange of research information through interoperable systems [5].

There are two conditions for assignment and usage of PIDs:

- persistent identifiers should only be assigned to resources that will be preserved for long term, that is, over several hardware and software generations;
- a persistent identifier and the services it provides should be at least as persistent as the resource identified. The resource may undergo several migrations and the outdated version / versions may no longer be accessible and / or usable. A user who has a PI of an old manifestation of a resource should be redirected to the latest version available, or to work level metadata, which may enable acquisition of the work in some other form, such as print [6].

Persistent Identifiers assigned to researchers
ORCID ID

The most popular researcher PID is Open Researcher and Contributor IDentifier (ORCID ID). It is provided by the ORCID organization. ORCID is an international, interdisciplinary, open, non-proprietary, and not-for-profit organization created by the research community for the benefit of all stakeholders.

ORCID ID is composed of 16-digit numbers that is compatible with the ISO Standard (ISO 27729), e.g. 0000-0001-2345-6789. Initially ORCID iD is randomly assigned by the ORCID Registry. When stored, the ORCID iD should be expressed as a full https URI: https://ORCID.org/xxxx-xxxx-xxxx-xxxx, complete with the protocol (https://), and with hyphens in the identifier (xxxx-xxxx-xxxx-xxxx).

ORCID system is based on collaboration amongst publishers, universities, funding bodies, researchers, and other stakeholders in scholarly communications. ORCID is

committed to allowing individual researchers to create, claim, manage and control the visibility of their data or to optionally delegate the management of their data to their university or another third party [7].

How to use ORCID iD? Include your ORCID ID when you:

- Submit manuscripts – publishers are adding ORCID IDs of contributors to a publication's metadata, correctly attributing your work to you;
- Apply for funding – many funding bodies systems has ORCID integrated into the grand submission workflow;
- Accept peer review assignments – you can be acknowledged for your contribution to the research community;
- Renew professional memberships - organisations are increasingly using ORCID ID to help link and disambiguate members' identity;
- Create your poster or power point presentations – get your QR Code for your ORCID ID and include it on your poster or presentation;
- Communicate with colleagues - add your ORCID ID to your email signature or business card [8].

ResearcherId (Web of Science ResearcherId)

ResearcherID is an identifying system for scientific authors. The system was introduced in January 2008 by Thomson Reuters.

ResearcherID consists of alphanumeric characters and the last 4 numbers contains the year you registered.

Researchers can use ResearcherID to claim their published works and link their unique and persistent ResearcherID number to these works for correct attribution. In this way, they can also keep their publication list up to date and online.

The combined use of the Digital Object Identifier with the ResearcherID allows a unique association of authors and research articles. It can be used to link researchers with registered trials or identify colleagues and collaborators in the same field of research

In April 2019, ResearcherID was integrated with Publons, a Clarivate Analytics owned platform, where researchers can track their publications, peer reviewing activity, and journal editing work. With ResearcherID now hosted on Publons researchers can keep a more comprehensive view of their research output and contributions in one place. This is particularly important for researchers in fields that predominantly use peer-reviewed conference articles (computer science) or in fields that focus on publishing books and chapters in books (humanities and disciplines in the social sciences).

ResearcherID and Publons are also integrated with Web of Science, and ORCID, enabling data to be exchanged between these databases [9].

Scopus Author ID

Scopus Author ID allows to track own publications indexed in the Scopus citation database and build metric reports. One can use Scopus to follow the outputs of other researchers and institutions and identify potential collaborators [10].

A Scopus Author ID is a unique 11-digit identifier automatically assigned by Elsevier in its ScienceDirect and Scopus products the first time one of your publications is indexed in Scopus.

Users can search the lookup tool to locate an author's profile, which includes the identifier, references, citations of work, h-index, and subject areas.

Due to variant publishing names Scopus may generate multiple IDs for the same author. Scopus offers the possibility to merge all your IDs to optimize the metrics.

Like in case of ResearcherID, at Scopus it is easy for researchers to freely import their research papers to ORCID through a direct link on the author detail page [11].

Google Scholar Citation Profile

Google Scholar Citations provide a simple way for authors to keep track of citations to their articles. You can check who is citing your publications, graph citations over time, and compute several citation metrics.

In order to create a Google Scholar Citation profile, you need a Google Account. Once the profile is set up, it will automatically update.

You can add groups of related articles, not just one article at a time; and your citation metrics are computed and updated automatically as Google Scholar finds new citations to your work on the web. You can choose to have your list of articles updated automatically or review the updates yourself, or to manually update your articles at any time [12].

A profile generally lists your name, chosen keywords of research interest, generated citation metrics, and citations (including links to citing articles).

You can also make your profile public, so that it may appear in Google Scholar results when people search for your name. In order to be included in Google Scholar search results, you will need to make your profile public and needs to have a verified email address at your university (non-institutional email addresses are not suitable for this purpose). To do this, click the 'Edit' link next to the 'No Verified email', and add your email address. Your email address with not be displayed on your public profile [13].

ORCID has created a tool that allows you to import citations from BibTeX (.bib) files into your ORCID record, including files exported from Google Scholar and other popular citation management tools.

Persistent Identifiers assigned to publications
ISBN

The International Standard Book Number (ISBN) is the persistent identifier for books. ISBNs identify printed or digital books and are used as inventory-tracking devices. The code is usually found on the back cover of a book and includes a barcode. One shortcoming of the ISBN is that it does not provide information on all versions of a particular book because each has its own unique code, which might not be an issue since it is recognizable.

The ISBN has 10 or 13 digit digits and contains five types of identifying information or "elements" such as: a prefix, registration group (e.g., individual country or territory), registrant, publication, and "checksum" (i.e., a figure that is used to detect errors).

By purchasing an ISBN for your publication, you ensure a better chance of it being found in a search. [14, 15]

ISSN

The International Standard Serial Number (ISSN) comprises of 7 digits followed by a "check digit" to identify any errors in citations. ISSNs help to identify serial publications, which are those that are published regularly in sequence, such as magazines, journals, newspapers, and databases. They don't identify content or certify its validity. Although ISSNs do not identify the journal owner, if a journal name changes, a new ISSN is necessary.

This is important when considering a title change because the change will affect cataloging and indexing.

ISSNs also have some limitations. There might be an ISSN for printed text and a different one for electronic versions. "E-journals" are becoming quite popular given the Internet capabilities. Publishers are finding it much less costly to publish their serials online, and many do so exclusively. In addition, many of these journals are open access publications, which allow researchers to more easily gather needed information.

The ISSN format is "ISSN 0000-0000," which is printed on the journal cover. For electronic versions, "e-ISSN" is posted on the home page of the publication [15].

DOI

Digital Object Identifier (DOI) is a persistent identifier or handle used to identify objects uniquely. The DOI system has been standardised through the International Standards Organisation, ISO (within the responsibility of committee ISO TC46/SC9, Identification and documentation) as ISO 26324, Digital Object Identifier System. The Standard was approved in November 2010 and published in May 2012 [16].

DOI is a unique alphanumeric string assigned by a registration agency to identify content and provide a persistent link to its location on the Internet.

All DOI numbers begin with a 10 and contain a prefix and a suffix separated by a slash. The prefix is a unique number of four or more digits assigned to organizations; the suffix is assigned by the publisher and was designed to be flexible with publisher identification standards.

Publishers and repositories often assign DOIs at the time of publication. The benefit of a DOI is that it is persistent and will not change even if the item's location online (URL) changes. They help to mitigate the challenge of broken links [17].

It is recommended that when DOIs are available, you include them for both print and electronic sources. The DOI can also be found on the database landing page for the article [18].

PubMed Indexing and PubMed Referencing Number

PubMed Indexing Number (PMID) is used for articles in the PubMed database, which contains more than 27 million citations from several sources. Like DOI, PMIDs are unique identifiers and are assigned to each record in the PubMed system. The code is found at the end of a PubMed citation and below the article's author list and title.

The PubMed Central referencing number (PMCID) is required for listing in the National Institutes of Health (NIH) grant proposals and reports. PMCIDs are assigned to an article in PubMed Central, which differs from the PubMed database. These articles comply with NIH's Public Access Policy. PMCIDs are listed below the abstract [15].

CODEN

CODEN was designed and published in 1953 by a researcher at the Chronic Disease Research Institute to help him remember his reference publications, but became popular with scientists of all disciplines and is used for bibliographic indexing. Usually, CODEN is for articles related to chemistry; however, publishers of non-chemistry articles can request a CODEN from the International CODEN Service.

A CODEN is a six-character, alphanumeric code for both serial and non-serial science publications. For serial publications the first four letters represent the journal title, and the

fifth and sixth letters refer to one of the first six letters of the alphabet and the check character, respectively. For non-serial publications, the first two characters of the CODEN are numbers followed by letters. The fifth character is taken from the entire alphabet, and the sixth is the check character [15].

ARXIV ID

ARXIV ID is an identifier of a document in arXiv pre-print archive (arxiv.org). It is a repository that offers open access to more than 1.5 million e-prints approved for posting after moderation, but not full peer review, in the fields of physics, mathematics, computer science, quantitative biology, quantitative finance, statistics, electrical engineering and systems science, and economics [19].

The canonical form of arXiv ID identifiers from January 2015 (1501) is arXiv:YYMM.NNNNN, with 5-digits for the sequence number within the month.
arXiv is owned and operated by Cornell University, a private not-for-profit educational institution and submissions to arXiv should conform to Cornell University academic standards [20].

Users can retrieve papers from arXiv via the web interface. Registered authors may use the web interface to submit their articles to arXiv. Authors can also update their submissions if they choose, though previous versions remain available [19].

Its existence was one of the precipitating factors that led to the current movement in scientific publishing known as *open access*. Mathematicians and scientists regularly upload their papers to arXiv.org for worldwide access and sometimes for reviews before they are published in peer-reviewed journals.

Conclusions

If you are undertaking research, you will need to set up persistent digital identifiers that uniquely distinguish you as a researcher and connect you to your research activity.

Taking into consideration that ORCID and Google Scholar are easy to access and manage, the author of the paper considers that every researcher should set up at least ORCID ID and Google Scholar Profile. The author calls you to include your ORCID ID when you apply for grants, submit manuscripts, accept peer review assignments, renew professional profile or create your poster presentation. Include your ORCID ID in your email signature or business card. At the same time, Google Scholar offers many useful tools like grouping related papers, updated citation metrics for your works.

Regarding identifiers used for publications, every identifier has its own purpose, but, just to mention, referring to an online document by its DOI is supposed to provide a more stable link than simply using its URL.

Nevertheless, you can set up all of researcher identifiers overviewed in this article, as well as a publication can have one or more different identifiers.

As a researcher, you should recognize these various identifiers to make your searches less time-consuming.

References

1. Jankowski NICHOLAs W. Exploring e-Science: An Introduction [online]. In: *Journal of Computer-Mediated Communication*, 2007, 12 (2), pp 549-562 [cited 27.08.2019]. Available: https://doi.org/10.1111/j.1083-6101.2007.00337.x

2. Liberati Diego. Networked Experiments in Global E-Science. In: *Handbook of Research on Electronic Collaboration and Organizational Synergy*. IGI Global, 2009, pp. 615-625. DOI: 10.4018/978-1-60566-106-3.ch040.

3. Nentwich Michael. *Cyberscience: Research in the Age of the Internet*. Vienna: Austrian Academy of Sciences, 2003. ISBN 3-7001-3188-7.

4. Hine Christine. *Systematics as Cyberscience: Computers, Change, and Continuity in Science*. The MIT Press, 2008, 320 p. ISBN 978-0-26-208371-3.

5. Duine Maaike. The importance of Persistent Identifies and the Role of THOR [online]. Overview of the presentation at Digital Infrastructure for Research 2018 Conference [cited 28.08.2019]. Available: https://www.digitalinfrastructures.eu/content/importance-persistent-identifiers-and-role-thor

6. Hakala Juha. Persistent identifiers – an overview [online presentation], [cited 28.08.2019]. Available: https://slideplayer.com/slide/3892164/

7. What is the relationship between ISNI and ORCID? Internet resource of ORCID - https://support.orcid.org/hc/en-us/articles/360006973973-What-is-the-relationship-between-ISNI-and-ORCID-

8. ORCID ID and research identifiers. Internet resource of the Library of the University of Qeensland Australia - https://web.library.uq.edu.au/node/4566/1

9. ResearcherID. Wikipedia. Internet resource - https://en.wikipedia.org/wiki/ResearcherID

10. What is Scopus Author ID? Research Identity. Internet resource - https://utas.libguides.com/ManageID/Scopus

11. Researcher Profiles, Identifiers and Social Networks: Maximise your Impact. Scopus Author Identifier. Internet resource - http://unimelb.libguides.com/researcher_profiles/scopus

12. Google Scholar. Internet resource - https://scholar.google.com/intl/en/scholar/citations.html

13. My Citations (Google Scholar). Researcher Identifier and your Online Research Profile. Internet resource - https://libguides.usc.edu.au/researcheridentifiers/gsch

14. ISBN's and Identifiers: Standards Basics for Indpendent Publishers [online presentation], [cited: 31.08.2019]. Available: https://www.slideshare.net/bisg/isbns-and-identifiers-standards-basics-for-the-independent-publishing-community-featuring-phil-madans-executive-director-of-digital-publishing-technology-at-hachette-book-group

15. ISBN, DOI, ISSN: A Quick Guide to Publication Identifiers. Enago Academy. Internet resource - https://www.enago.com/academy/isbn-doi-issn-a-quick-guide-to-publication-identifiers/.

16. DOI.org. Internet resource - http://doi.org.

17. Whar are DOIs and PMIDs? Library Guides. Library of the University of Chicago - https://guides.lib.uchicago.edu/c.php?g=297381&p=3260251

18. What is digital object identifier, or DOI. Learning APA Style - https://apastyle.apa.org/learn/faqs/what-is-doi

19. arXiv archive - https://arxiv.org

20. arXiv. Wikipedia. Internet resource - https://en.wikipedia.org/wiki/ArXiv