Digital identifiers in scientific publishing and e-health

Hamid Reza Khedmatgozar
Iranian Research Institute for Information Science and Technology (IRANDOC)
Tehran, Iran

Correspondence to:
Hamid Reza Khedmatgozar
Assistant Professor in Information Technology Management
Iranian Research Institute for Information Science and Technology (IRANDOC)
No. 1090, Enghelab Avenue
P.O. Box 13185-1371
Tehran, Iran
khedmatgozar@irandoc.ac.ir
h.khedmatgozar@gmail.com

Abstract
With the growth of information technology over the past decade, digital identifiers have been introduced for the unique and stable identification of digital objects in cyberspace. Digital identifiers have applications in many contexts, including scientific publications. In addition to describing their use for scientific publications, this article presents an additional potential application in the field of electronic health and invites experts and researchers to investigate it further.

I recently read an article published in this journal in 2013 entitled "AuthorAID: An international service and chance to serve" by Barbara Gastel.[1] Barbara states that too little attention has been paid to researchers in low- and middle-income countries, and that the scientific community needs to consider the results of their research. She mentions the growing gap in scientific publications between low-income countries and the rest of the world in terms of number and content of research articles, and the process for getting them published, as well as ongoing efforts to make scientific publications more accessible to readers everywhere in the world. Since the publication of her paper, developments have helped to reduce this gap and increase the availability of scientific publications. In this article, I will try to illustrate one of these developments and research areas relevant to it.

What is a digital identifier?

Today, many of our professional and leisure activities have moved to cyberspace. Many objects and entities in cyberspace (including persons, corporate bodies, animals, books, articles, journals, and songs) have a distinct identity and are called digital objects. How to deal with them is a big issue. In general, dealing with digital objects is based on two basic approaches: (1) archiving digital objects and related information (metadata) to organise and preserve them over time ("memorable web" approach), and (2) increasing the digital object's find-ability through a deep search on the web ("navigable web" approach). More recently, a third approach called the "identifiable web" or "digital identifier" approach has emerged.[2] This third approach is close to the memorable web approach with its focus on uniqueness and persistent identification of digital objects, but it also offers the main feature of the navigable web approach by focusing on a specific mechanism for instant online access to digital objects. In this article, I introduce research and applied aspects relating to this approach.

Currently, uniform resource locators (URLs) are used to identify and access digital objects. However, using URLs as identifiers carries risks, including lack of persistency (inability to access the object due to server migration or digital object transfer – "error 404") and violations of uniqueness (multiple URLs for a specific digital object). To overcome these risks, researchers have tried to create a unique identifier with long-term persistence. The main solution proposed is the use of digital identifiers instead of URLs.[3]

The digital identifier is a code assigned to a digital object in cyberspace that is completely unique, much like a fingerprint. There is a one-to-one relationship between the digital identifier and the identified digital object. Digital identifiers are issued by independent authorities, which enhances their reliability and accuracy. Each digital identifier acts as a link so that anyone clicking on it will either be redirected to a valid URL for the object or to metadata about the object, such as its creator, publisher, and format. This redirection is achieved by a mechanism called resolution (Figure 1).

The main functions and benefits of digital identifiers are unique and persistent identification on the web, permanent maintenance of location information (even if the digital object's location changes), standardisation and enrichment of metadata, facilitation of content searching, securing of copyright, increased traffic to digital objects, reduced costs, time savings for stakeholders, exploitation of new business and research areas, and increased revenues from digital objects. This digital identifier approach has important applications in scientific publications and research data. Other potential applications include interactive television, digital museums, virtual tourism, and e-learning.[3] Based on my own investigations, another potential application is electronic health (e-health).

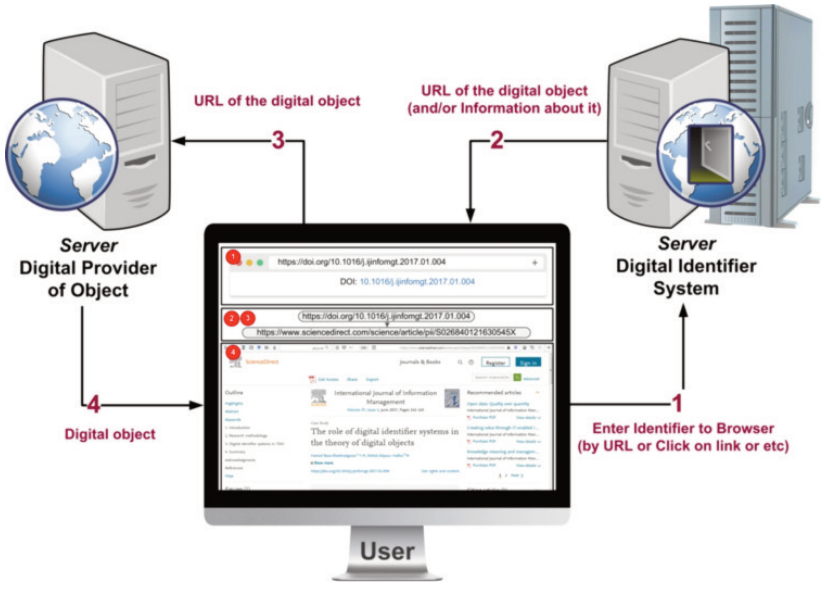

Figure 1. Example of a resolution mechanism for accessing a digital object through its digital identifier

A digital identifier is a piece of actionable code in cyberspace that works based on resolution. This mechanism has four steps. The first two steps involve redirecting the digital identifier to the location (URL) of the identified digital object and the last two steps involve providing access to this object via its URL. All of these steps are done in less than a second by your web browser.

Digital identifiers and scientific publications

In the context of scientific publications, digital identifiers and related systems can help stakeholders in the process of writing, editing, submitting, reviewing, and publishing papers in scientific journals. Some specific applications of digital identifiers are listed in Table 1.

Digital identifiers mean connecting to millions of other scientific articles via hyperlinks, with the digital identifier acting as the fingerprint of an article on the web. One of the most important applications of digital identifiers in scientific publications is the Digital Object Identifier (DOI) system and related registration agencies such as Crossref and DataCite.

DOI is the most well-known digital identification system. It is managed and controlled by the International DOI Foundation (IDF), a non-profit organisation. A DOI is permanently allocated to an object to provide a stable hyperlink that redirects to current information about that object, including its location. While information about an object can change over time, its DOI will not change. DOIs are allocated on behalf of the IDF by its registration agencies. The IDF currently has 10 registration agencies. As shown in Figure 2, DOI syntax consists of an indefinite character string including a prefix and a suffix. The prefix is a directory code that identifies the IDF as the directory and that is always 10 in DOI, followed by a registrant code, which is the unique code assigned by IDF registration agencies to the owner or publisher of the object being identified. The directory code and the registration code are separated by a full stop. After the prefix is a forward slash and then the suffix. This suffix is a unique code containing alphanumeric strings assigned by the registrant to identify the object. This suffix must be unique to the registrant and can include other identifiers such as the ISBN, ISSN, or serial number. There are no operational restrictions on the length of a DOI or any of its components, and identifiers up

Table 1. Specific applications of digital identifiers

Application	Explanation
Enabling unique and persistent identification	Digital identifiers enable the creation of unique, persistent, automated links between digital objects such as journals, publishers, editors, authors, and even components of an article such as a table or a dataset and remove human error in maintaining the links between them. They also enable automatic and reliable retrieval of information about articles.
Facilitating acceptance by indexers	In their initial evaluations of journals, many indexers, such as International Scientific Indexing, PubMed, Scopus, and Directory of Open Access Journals, ask whether the journal uses digital identifiers. Although not stated as a requirement for indexing, journals that use digital identifiers generally find it much easier to get accepted by indexers. Moreover, it has been claimed that if a journal is indexed by one of the major indexers and starts using digital identifiers, then its impact factor can increase greatly within a year.[9]
Automating the citation of articles, journals, and authors	Digital identifiers can be used to automatically monitor the level of influence of an article, author, or journal.
Increasing visibility	A big challenge for every journal and author is to bring the content of their article to the largest possible number of readers. Information about articles with digital identifiers is recorded in multiple reputable databases, which increases visibility of the articles in search engines. This increase in visibility is beneficial to journals and authors and also helps readers to find journal articles.
Preventing plagiarism	Using digital identifiers helps to maintain copyright by providing direct access to the official links for articles. iThenticate, the most reputable plagiarism detection service, only includes in its database published articles with digital identifiers. If a journal article has a digital identifier, it will automatically be entered into the database. From the time a journal subscribes to iThenticate, iThenticate will notify the journal and its publisher if another article with even the slightest resemblance to one of their articles is published.

to 4 GB can be assigned. To use the DOI system and access the DOI directory, https://doi.org/ must be placed before the DOI syntax.[4]

The mission of Crossref, an independent nongovernmental company launched in 2000 and now run by research publishers, is to facilitate access, citation, and reuse of articles and books as research outputs. Through the DOI, Crossref enables researchers to easily switch from article to article at the click of a mouse, even if the articles are in different journals. The company also has other services based on the DOI, such as Similarity Check, Crossmark, and Crossref Metadata Search. Similarity Check is a service developed by iThenticate[5] that allows publishers to check the authenticity and originality of the articles that are submitted to them. Crossmark can be used to check whether or not you are using an up-to-date version of an electronic article. And Crossref Metadata Search automatically provides the information needed by some indexers without journal interference. Figure 3 shows an example of easily switching from article to article. Publishers of scientific content, especially books and journals, can access Crossref to use its services, including DOIs.[6]

> Optimal use of the specific features of digital identifiers, including persistency and uniqueness, can help to reduce the gaps between rich and poor in information and health services in the public health field.

Researchers have long wanted a better system for sharing data, as well as data archiving tools for reusing data in future studies. The need to create and manage a process that provides continuous and reliable access to data led to the establishment of DataCite in 2009 with the aim of assigning a unique identifier to scientific datasets. To this end, DataCite collaborates with a number of research libraries worldwide. According to DataCite, data should be cited in the same way that articles and books are cited. Citation allows for verification and reuse of data and allows researchers to be identified and rewarded for their contributions based on the impact of their dataset. Researchers and organisations can contact a DataCite member to obtain a DOI for each of their datasets. For example, researchers at the ResearchGate[7] scientific social network can share their datasets, receiving DOIs through DataCite.[8]

However, in developing countries, scientific publishers use DOIs significantly less than in developed countries, often due to lack of awareness and low financial capacity. As a result, the scientific outputs of researchers in developing countries are read and cited less than those in developed countries. Also, the limited citations these scientific outputs receive cannot be tracked and displayed to researchers because DOIs have not been used. This is the most important gap that the wider use of DOIs could help reduce and is consistent with the research results of Boudry and Chartron.[10] They systematically studied the use of the DOI in articles indexed in PubMed and concluded that the DOI is more frequently used by journals in developed countries than in developing countries and in international journals than in regional journals. As well as demonstrating the use of the DOI in increasing global knowledge sharing, they propose that, in order to further increase and improve sharing,

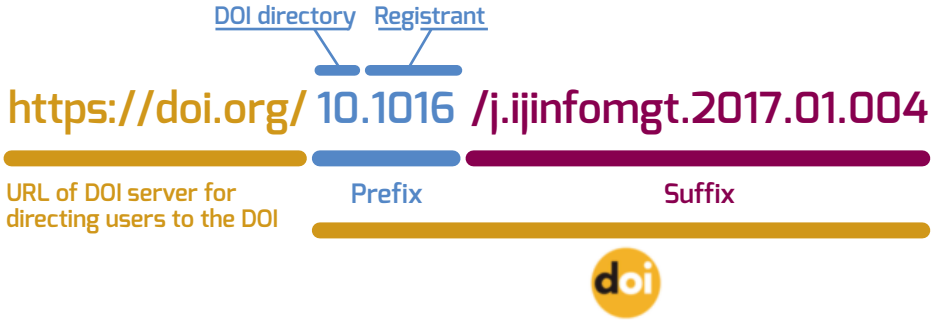

Figure 2. Syntax of a DOI
DOI syntax has two parts: prefix and suffix. The prefix consists of two parts, the directory and registrant, and is assigned by a DOI registration agency. The suffix is assigned by the owner or publisher of the identified object. In order to use this syntax to implement resolution, the web address (https://doi.org/) must be placed before the prefix.

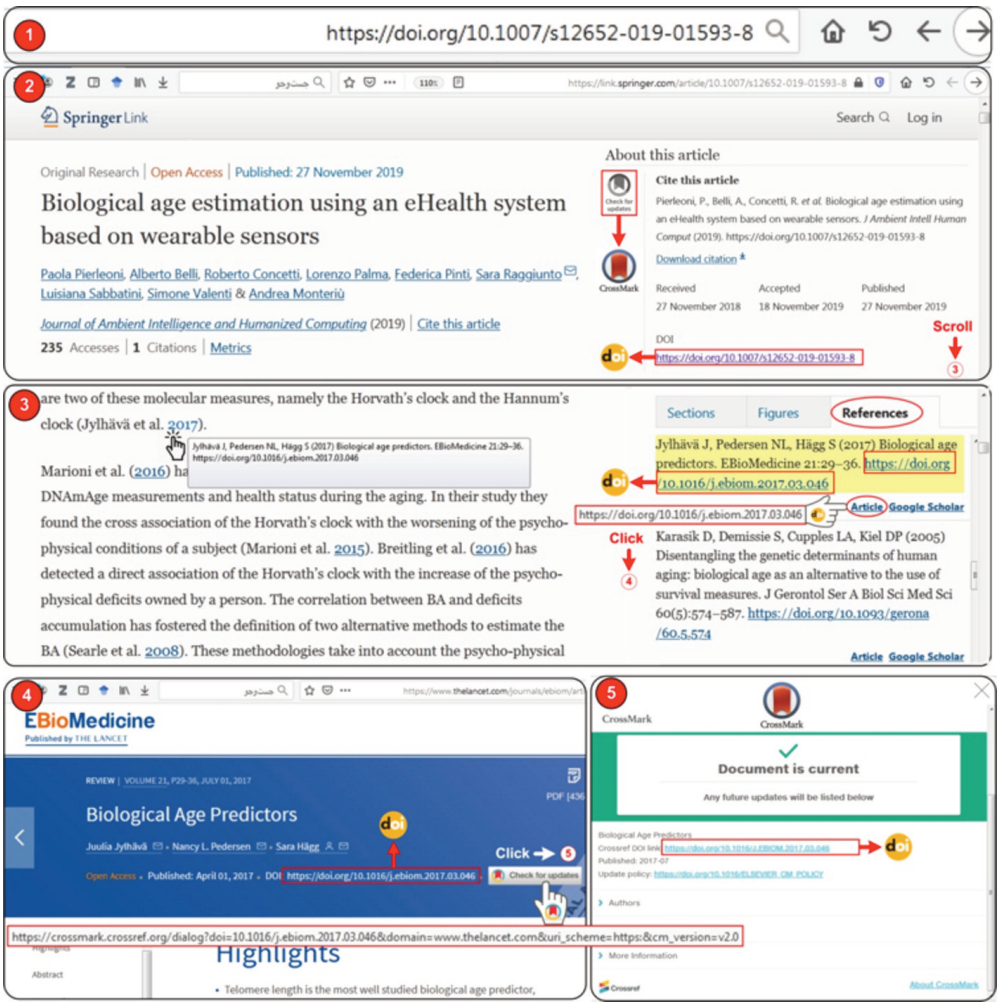

Figure 3. Example of easily switching from article to article using DOIs
When you read an article, you see that another article has been referenced. You can easily access that reference by clicking on its DOI. You can also click on the CrossMark icon to check whether you are viewing the latest version of the article.

necessary measures must be undertaken to allocate this identifier to journals in developing counties.

Digital identifiers and e-health

In today's world, health, technology, and business have found a strong link in the concept of e-health. The dominant term in the technology aspect of this concept is the internet. [11] Systems such as electronic health records and health management information systems (HMISs) also fall under this concept. An electronic health record is a longitudinal electronic record of a patient's health information generated as a result of one or more encounters in a care delivery setting.[12] An HMIS is a system whereby health data are recorded, stored, retrieved, and processed to improve decision-making.[13]

At the macro level, the literature in this area indicates that in order to achieve universal health coverage, a unique health identifier is essential to identify individuals in a country, especially patients.[14] Efforts have been made in different countries to assign identifiers to patients. For example, Mills and colleagues[14] have compared efforts in the United Kingdom, Slovenia, South Korea, and Thailand. They point out that, in some countries, there are problems with the lack of access to identifiers at all times, as well as the existence of local identifiers and the lack of links between local and national identifiers. Studies in other countries such as Cambodia, Laos, and Myanmar,[15] as well as Nigeria,[16] point to other similar problems. It seems that the benefits of a digital identifier solution, including guaranteed stability and metadata, can be used to address some of these problems. A joint study of unique health identifiers and digital identifiers could be of interest to researchers.

One of the basic processes in the field of health in general and e-health in particular is the unique identification of all kinds of non-patient stakeholders and objects, such as nurses, hospitals, medicines, therapists, and pharmacies. If these stakeholders and objects are uniquely identified, it is possible to link them and their data, connect different HMISs, and thereby monitor the stakeholders and objects. For example, Sensmeier and colleagues[17]

emphasise the need for the unique digital identification of nurses in both the health/treatment process and digital information systems in order to monitor their actions and evaluate their performance. Given the advantages of digital identifiers, it seems they would be suitable for uniquely and persistently identifying and monitoring non-patient stakeholders and objects.

Health research continues to generate a large number of valuable datasets. An example is the field of genetics and in particular genomics (most recently in relation to COVID-19). In order to have reliable and permanent global access to these datasets, more attention needs to be paid to assigning digital identifiers to them. The work at DataCite can be a precursor to these efforts, which should be of great interest to health professionals and researchers.

Conclusion

In the field of scientific publications, digital identifiers simultaneously provide unique, persistent identification and persistent access to scientific outputs such as articles and datasets. Through the use of digital identifiers in low- and middle-income countries, research outputs will become more accessible to the world at large, thereby reducing the research gap between richer and poorer nations. So, if you are a reader, author, reviewer, or editor of scientific research, I suggest you become more familiar with digital identifiers and related services.

The use of digital identifiers in e-health is a new research subject of potential interest to researchers in both information technology and health. It can also be considered as a business opportunity for practitioners in these areas. In my opinion, optimal use of the specific features of digital identifiers, including persistency and uniqueness, can also help to reduce the gaps between rich and poor in information and health services in the public health field. For example, digital identification of a person living in a slum area or low-income country combined with persistent access to his or her health profile can enable centralised monitoring of all health services provided to the person by all stakeholders. Decision-makers can then exploit the persistent and unique identification of digital objects including patients, stakeholders, and datasets to identify health service gaps between rich and poor areas and work to reduce them.

Conflicts of interest

The author declares no conflicts of interest.

References

1. Gastel B. AuthorAID: An international service and chance to serve. Med Writ. 2013;22(4): 284–7. doi: 10.1179/2047480613Z.000000000151
2. Khedmatgozar HR, Alipour-Hafezi M. The role of digital identifier systems in the theory of digital objects. Int J Inform Manage. 2017;37(3):162–5. doi: 10.1016/j.ijinfomgt.2017.01.004
3. Khedmatgozar HR, Alipour-Hafezi M. A basic comparative framework for evaluation of digital identifier systems. J Digit Inf Manag. 2015;13(3):190-7.
4. International DOI Foundation (IDF). DOI Handbook. 2020 [cited 2020 Apr 9]. Available from: https://www.doi.org/10.1000/182
5. iThenticate: Plagiarism Detection Software. 2020 [cited 2020 Apr 9]. Available from: http://www.ithenticate.com/
6. Crossref. 2020 [cited 2020 Apr 9]. Available from: https://www.crossref.org/
7. ResearchGate. 2020 [cited 2020 Apr 9]. Available from: https://www.researchgate.net/
8. DataCite. 2020 [cited 2020 Apr 9]. Available from: https://datacite.org/
9. Braile DM. After the impact factor, the DOI. Rev Bras Cir Cardiovasc. 2011; 26(3):I-II. doi: 10.5935/1678-9741.20110001
10. Boudry C, Chartron G. Availability of digital object identifiers in publications archived by PubMed. Scientometrics. 2017;110(3):1453–69. doi: 10.1007/s11192-016-2225-6
11. Oh H, Rizo C, Enkin M, Jadad A. What is eHealth (3): A systematic review of published definitions. J Med Internet Res. 2005;7(1):e1. doi: 10.2196/jmir.7.1.e1
12. Health Information and Management Systems Society. EHR: Electronic health record. 2020 [cited 2010 Mar 20]. Available from: https://www.himss.org/electronic-health-records
13. Endriyas M, Alano A, Mekonnen E, et al. Understanding performance data: Health management information system data accuracy in Southern Nations Nationalities and People's Region, Ethiopia. BMC Health Serv Res. 2019;19(1):175. doi: 10.1186/s12913-019-3991-7
14. Mills S, Lee JK, Rassekh BM, et al. Unique health identifiers for universal health coverage. Journal of Health, Population and Nutrition. 2019;38 (1):22. doi: 10.1186/s41043-019-0180-6
15. Stahl M, Roth S, Thorell L, Parry J. On the road to universal health coverage: every person matters. ADB Briefs. 2016; No. 56:1–8.
16. Chukwu E. The case for a unique digital patient ID scheme in Nigeria. J Health Med Informat. 2017;8(3):267–71. doi: 10.4172/2157-7420.1000267
17. Sensmeier J, Androwich I, Baernholdt M, et al. Demonstrating the value of nursing care through use of a unique nurse identifier. Online J Nurs Inform. 2019;23(2).

Author information

Hamid Reza Khedmatgozar is an Assistant Professor of Information Technology Management at the Iranian Research Institute for Information Science and Technology (IRANDOC) in Tehran, Iran. His research interests revolve around IT adoption, digital identifier systems for information objects, e-business models, risk management, neural networks, and multiple-criteria decision-making.