Jia Liu*

Digital Object Identifier (DOI) and DOI Services: An Overview

https://doi.org/10.1515/libri-2020-0018
Published online May 19, 2021

Abstract: In the establishing anniversary of the two biggest Digital Object Identifier (DOI) registration agencies all over the world, Crossref and DataCite, the paper intends to provide an overview of the development and approaches and of DOI and DOI services, from which scholarly communication has benefited greatly. At first, the author explores the initiation of DOI and differences of DOI from other persistent identifiers. After that, DOIs for different kinds of objects and DOIs' value in enhancing scholarly communication is discussed; then, in the second part, DOI services at different levels in a pyramid and those particularly in Germany are described. The active involvement of the library world are also introduced here; finally, the current situation and prospects as well as some issues dealing with DOIs and DOI services are investigated in the last part of the paper.

Keywords: Digital Object Identifier (DOI), Persistent Identifier (PI), scholarly communication

1 Introduction

2019 is the 20th anniversary of the establishment of Crossref and the 10th anniversary of that of DataCite. Both of them are global registration agencies (RAs) of DOI (Digital Object Identifier) for scholarly information. Therefore, it would be meaningful to have a review upon the approaches and development of DOI and the services related to DOI during these years.

DOI is one kind of Persistent Identifier (PI or PID) of an object. As a long-term globally unique identifier, it has been broadly adopted and employed all over the world, especially in the scientific world. For years, it has been a standard element for identifying scientific data.

It has been a little more than 20 years since the International DOI Foundation (IDF), an organization responsible for operating and governing DOIs, was launched to develop systems in 1998. According to IDF's latest statistics, "currently DOIs have been used by well over 5000 assigners, e.g., publishers, science data centres, movie studios, etc" (International DOI Foundation 2018). IDF also declares that "DOI names are assigned by multiple Registration Agencies (RA) worldwide and over 27 million short DOI links to DOI names are in use" (International DOI Foundation 2019). Furthermore, 11 years after IDF's birth, i.e. in 2009, DataCite aimed at providing DOI registration service to research data universally, came into being. The latest available statistics show that until now there are in total 19,532,085 DataCite Registrations and 17,307,178 metadata registered by DataCite are searchable (DataCite n.d. (a)). Doubtlessly, DOI has become an important kind of PI and been commonly adopted in scholarly communication. In any case, now is a good opportunity to have a review on the short but successful history of DOI and DOI services.

In order to provide a comprehensive historical review of the development and approaches of DOI and DOI services, the author has applied research methods such as literature review and case study while writing the paper. As many as possible online information resources have been collected to support the text. And in addition to elaborating a pyramid of the DOI services in details, the author takes the status quo of DOI services in Germany, where DataCite originated and hosts and DOI services have been systematically and widely provided, to provide the reader with a concrete example of such services in a country.

In the following text, at first, the origin of DOI will be explored and compared with other traditional identifiers. Some examples of DOIs assigned to variety of entities will be displayed and the value of DOI in scholarly communication will be discussed; then, the author will sketch DOI services in a general way and then introduce the DOI services particularly in Germany; after that, the current situation and prospects as well some issues dealing with DOI will be put forward.

2 DOI

The IDF's handbook describes that "DOI is an acronym for "digital object identifier," meaning a "digital identifier of

***Corresponding author: Jia Liu**, Department of Reseach and Development, University and City Library of Cologne, Universitaetsstr. 33, 50931 Cologne, North Rhine-Westphalia, Germany,
E-mail: liu@ub.uni-koeln.de. https://orcid.org/0000-0001-9525-2478

an object." A DOI name is an identifier (not a location) of an entity on digital networks. It provides a system for persistent and actionable identification and interoperable exchange of managed information on digital networks" (International DOI Foundation 2015, n.p.). As regards to the structure, a DOI is "a set of numbers, letters and other characters, constructed of two components which are a prefix [a directory indicator followed by a Registration Agency (RA) code] and a suffix (unique to a given prefix to identify the entity)" (DOI Foundation 1997). The two parts are separated with the "/" character.

Run by IDF, the DOI system implements the Handle System and provides a ready-to-use packaged system of several components:

- a specified standard numbering syntax;
- a resolution service;
- a data model incorporating a data dictionary; and
- an implementation mechanism through a social infrastructure of organizations, policies and procedures for the governance and registration of DOI names (International DOI Foundation 2015, n.p.).

2.1 The Initiation of DOI

A PI could be defined as "a long-lasting reference to a document, file, web page, or other object" (Wikipedia 2019). So far, there have in fact been many kinds of PIs employed under different contexts. For example, ISBN and ISSN are traditional bibliographic PIs that have been in use for printed publication since the 1970s while URN and URI are PIs for the digital objects in the computer network, while DOI is a comparatively newer PI concept emerging with the popularization of digital publishing.

As a matter of fact, the DOI's initiator is the publishing world, rather than any others. As early as 1998, Carol A. Risher, who was the Vice President of Copyright and New Technology at the Association of American Publisher, and William R. Rosenblatt reviewed in detail the DOI's early history starting from 1994 (Risher and Rosenblatt 1998, 18). According to their introduction, after the prototype and demonstration phase, the DOI system was announced at the Frankfurt Book Fair in October 1997 and, as mentioned above, IDF was set up to develop and manage the DOI System in the next year. They found that it was agreed that "the DOI system meets the needs of a universal identification system for digital content" and it "is just the foundation of the enabling technologies that will facilitate information retrieval and electronic commerce in a digital environment" (Risher and Rosenblatt 1998, 19). Five years later, in her book on the subject of scientific publishing, likewise, Andrea Öchsner elaborates that "the DOI system goes back to a joint initiative of three grade associations in the publishing industry (i.e., International Publisher Association; International Association of Scientific, Technical and Medical Publishers; Association of American Publishers). Although originating in text publishing, the DOI was conceived as a generic framework for managing identification of content over digital networks, recognizing the trend towards digital convergence and multimedia availability" (Öchsner 2013).

2.2 Differences of DOI from Other PIs

Juha Hakala introduced five PI systems (ARK <Archival Resource Keys>, DOI, PURL <Persistent Uniform Resource Locator>, URN <Uniform Resource Name>, and URI <Uniform Resource Identifier>) and comprehensively and compared "their functionality against the cool URIs" in her paper. She observed that "a major difference between URN and DOI is that with the former there will be just one URN for each instance of a traditional identifier such as ISSN (International Standard Serial Number) or ISBN (International Standard Book Number)" (Hakala 2010). The URN might refer to the name of a person and the URL to their home address, while the DOI system integrates the name and their home address. DOI means more than the combination of them. Behind a DOI, a set of metadata including the name and his home address as well as other metadata crucial for identifying a unique person could be found in the framework of the DOI system.

In addition, the examples during the demonstration of the DOI System discussed by Carol A. Risher and William R. Rosenblatt graphically show the distinction between DOI and URL with a sample URL provided in a bibliographic citation resulting in the frequently seen "File Not Found" error message. A DOI found in another bibliographic citation in the same set led to an online article even after ownership of the publication changed. It has been proved that "the DOI assigned to the original article still resolves two years later to the same article despite the change in ownership, and change in physical location of the server hosting the electronic article" (Risher and Rosenblatt 1998, 14). Even though the digital content could be very often dynamic in respect of its text, location, ownership, and so on, the DOI is precisely defined by means of structured metadata and it itself remains persistent through ownership and locations changes, unaltered once assigned.

Meanwhile, there is a relationship among DOI and other PIs. According to the handbook of IDF, "DOI is a registered URI within the info-URI namespace (IETF RFC

4452, the "info" URI Scheme for Information Assets with Identifiers in Public Namespaces). DOI names may also be expressed as URLs (URIs) through a http: proxy server" (International DOI Foundation 2015). On the platform of the DataCite Fabrica system, as soon as the sate of a DOI is updated from "Draft" to "Findable," i.e. the DOI registration process terminates and a DOI is successfully registered, a URL [the lower one in Figure 1] with the DOI name linking to the online resource is automatically created, which stays the same all along with the life of the DOI, even though the original URL [the upper one in Figure 1] of the resource could be altered. Compared to other identifiers, Juha Hakala pointed out in her overview of PIs that "traditional identifiers such as ISBNs are not and will not be actionable in the Internet as such. This means, among other things, that the character string 'ISBN 951-45-9942-X' is not and will not be interpreted as a hyperlink by Web browsers, whereas a persistent identifier incorporating this ISBN is a hyperlink when expressed as HTTP (Hypertext Transfer Protocol) URI" (Hakala 2010).

The DOI system differs from standard identifier registries such as ISBN, ISRC (International Standard Recording Code), etc.: "The purpose of an identifier registry is to manage a given collection of identifiers whereas the primary purpose of the DOI system is to make a collection of identifiers actionable and interoperable, where that collection can include identifiers from many other controlled collections" (International DOI Foundation 2019). In the project proposal of the so-called ISO Project 26324, it was explained that "the Digital Object Identifier (DOI) is a system for identifying content entities in the digital environment. It provides an actionable (resolvable) persistent interoperable identifier for entities used within digital networks. The simplest action is to locate the entity that the DOI identifies. A DOI can also specify the entity in greater detail by resolving to metadata about the identified entity" (ISO 2006). A DOI name is resolvable, of which the resolution is provided through the Handle System developed by Corporation for National Research Initiatives. It means that "a DOI name can be resolved within the DOI system to values of one or more types of data relating to the object identified by that DOI name, such as a URL, an e-mail address, other identifiers and descriptive metadata" (ISO 2012).

Nevertheless, in the ISO standard for DOI, *Information and Documentation: Digital Object Identifier System* (numbered as ISO 26324:2012), it is stated clearly that "the DOI name does not replace, nor is an alternative for, an identifier used in another scheme" (ISO 2012). Similarly, it's emphasized in the Factsheet on the website of doi.org that "the DOI System explicitly recognizes other schemes. The ISO DOI specification (ISO 26324) sets out the specifications for recognizing existing schemes. At minimum, the DOI Kernel metadata must record the fact that another registry identifier exists" (International DOI Foundation 2013).

2.3 DOIs for Different Kinds of Objects

Since DOI is a kind of digital identifier of an object but not an identifier of a digital object, the DOI system provides an infrastructure for persistently and uniquely identifying objects of any type. Being designed to assist identifier interoperability, the DOI system offers a persistent, semantically-interoperable resolution to related current data and is best suited to material that will be used in services outside the direct control of the issuing assigner (e.g., public citation or managing content of value).

Figure 1: The summary view of a DOI on the DataCite Fabrica platform.

Undoubtedly, ISBN and ISSN are the most common traditional standard identifiers of publications. And it's recognised that one of the most important tasks of PI is to "render the traditional identifiers actionable in the Web and provide persistent links to the resources" (Hakala 2010). In this case, the DOI could provide a "container" for the ISBN or ISSN, i.e., an ISBN could be integrated into a DOI. For example, the ISBN of the electronic version of the monograph under the title *Introduction to Scientific Publishing: Backgrounds, Concepts, Strategies* written by Andreas Öchsner is 978-3-642-38646-6, while its DOI is **10.1007/978-3-642-38646-6**. Moreover, the new concept, "the actionable ISBN" (ISBN-A), combines ISBN with DOI. The ISBN-A is "a service powered by DOI®, in which an existing ISBN is expressed in the DOI System" and "is constructed by incorporating an ISBN into the allowed DOI syntax" (International DOI Foundation 2019).

The next is a more explicit example of a DOI, in which the abbreviation ISSN is directly embedded. The ISSN of the online version of the "world's leading multidisciplinary science journal," *Nature: International Weekly Journal of Science*, is 1476–4687 and the hypothetical DOI **10.1038/issn.1476-4687** could be used for identifying the online version of the journal.

Also for an online journal, *D-Lib Magazine*, a world famous online journal focusing on digital library research and development, the DOI name **10.1045/dlib.magazine** was assigned. The directory indicator followed by the registrant code is in the prefix while the title of the journal has been taken in the suffix of the DOI for the resource with just a tiny alternation. And each article published within the journal has a unique DOI, such as 10.1045/january2017-stocker, with the issue month and year as well the family name of the author in the suffix. Even though *D-Lib Magazine* suspended publication of new issues in July 2017, the DOIs of the article function persistently.

As for the DOI for the research data, the DOI of a dataset published in a web database, which is maintained under a project of the Faculty of Chemistry, University of Cologne, Germany could be a good though a slightly complicated example. In the DOI **10.18716/nmrshiftdb2/60000061/Chemikalienlager_Uni_Mainz_CDCl3**, the second part of the prefix, i.e. 18716, is the registrant code standing for the RA – German Central Library of Medicine (shortly "ZB MED" in German) and its partner the University Library of Cologne. And the suffix starts with "nmrshiftdb2" that indicates the involved database, the NMR (Nuclear Magnetic Resonance) database. Following that is the character string assigned by the database administrator which includes the sequence number and the abridged form of the resource title. Registered by the University Library of Cologne, the DOI name 10.18716/map/00003 was assigned to the online image collection published via MAP (Modern Academic Publishing) with the host in the same university.

2.4 DOI's Value in Scholarly Communication

With the continuous development of computing and networking technologies, data capture and mining techniques, and the support of other new technologies, scholarly communication has been much improved than before so that research practices could be more collaborative and data-intensive. As the most important medium of scholar communication, a scholarly work goes through the whole procedure of research practice and scholar communication. DOI is revolutionarily designed to identify variety of objects. Due to its characteristics such as uniqueness, actionability, interoperability, persistence, and granularity, scholarly works with DOIs could be identified, fetched, cited, transferred, and preserved in a more precise and effective way. DOIs enable the accurate identifying and collocating of one's scholarly work. They are valuable in enhancing scholarly communication.

As mentioned above, DOI was initiated by the publishers to respond to the trend of massive growth in born-digital or digitalized media. The scientific world is full of digital records. Herbert Van de Sompel et al. argue that "scholars deserve an innately digital scholarly communication system that is able to capture the digital scholarly record, make it accessible, and preserve it over time" (Sompel et al. 2014). DOIs can be used in identifying the digital record uniquely and persistently regardless of whether it is a dataset, a text file, a multimedia (audio or video) work, a graph, a 3D-object, a website or any other kind of research outcome with their rich and structured metadata mechanism behind them. Furthermore, a DOI could also identify a collection of digital records.

Once a research outcome is published, a DOI could be registered for the publication either in digital or non-digital format and then "represent" it to enter the scholarly communication circle. Nowadays, DOI links have been fairly standard for scholarly works. They help make scholarly works more "findable" as the DOI does not change from database to database for that work since it is interoperable. PANGAEA (Data Publisher for Earth & Environmental Sciences) is an information system that offers publication services for archiving, publishing, and re-usage of research data. It is declared that "all data sets in PANGAEA® are citable, fully documented and can be referenced via persistent identifier (Digital Object Identifier – DOI) – a premise for data publication" (Schindler,

Diepenbroek, and Grobe 2012). Together with other ICSU (International Council for Science, after its former name, International Council of Scientific Unions) World Data Centers, PANGAEA once even had a share in the implementation of DataCite.

In the scientific world, DOI has been an important field that should be used if possible. In particular, when a new scientific publication appears, it may have a DOI assigned but without volume or pages assigned. In this case, the DOI is the only way to uniquely reference such publication. When it is possible, it has been made a priority to always use the DOI. In the biometric database, DOI has become a standard element of an item and a search option.

Recognizing that "DOIs provide an acceptable way for competitors to provide persistent links between different systems via a fair intermediary" and "a new identifier space (separate from URLs) with which scholar works may be identified," Google Scholar, the scientific search engine from Google, "uses DOI-based links to link and to index articles from the American Physical Society (APS) and other CrossRef members" (Warner 2005, 179). With similar recognition, while conceiving the concept of a scholarly communication metadata knowledge graph (SCM-KG), A. Sadeghi and his colleagues chose DOI as the identifier of articles and books (Sadeghi et al. 2017).

3 DOI Services

The services centered with DOI are provided at different levels by different institutions.

3.1 The DOI Service Pyramid

Generally, a hierarchical system exists within the DOI service community which looks like a pyramid shown as below (Figure 2).

At the top of the hierarchy is IDF, which is "the DOI system **registration authority** and maintenance agency and the central body which governs the DOI system. It is the common management and co-ordination body for DOI Registration Agencies" (International DOI Foundation 2018).

Governed and managed by IDF, the federation of **Registration Agencies** (RA) provides DOI services and registration: "The primary role of the Registration Agencies (RAs) is to provide services to Registrants — allocating DOI name prefixes, registering DOI names and providing the necessary infrastructure to allow Registrants to declare and maintain metadata and state data" (International DOI Foundation 2019).

Based on the latest chart from IDF, nowadays there are 10 RAs all over the world covering different areas, among which Crossref and DataCite are world famous examples of RA. Grounded in 1999, Crossref (formerly styled as Cross-Ref) has been the largest DOI RA until now. It "offers a wide array of services" such as assigning DOIs to scholarly content, enabling DOI resolution and search and so on "to ensure that scholarly research metadata is registered, linked, and distributed" (Crossref 2018). Ten years later, DataCite was founded on December 1, 2009 in London, which has been a leading global non-profit organization that provides DOIs for research data and other research outputs and its members (Registrants) are distributed across four different continents. DataCite enables the registrants to assign DOIs to all their research outputs so that "their outputs become discoverable and associated metadata is made available to the community. DataCite then develops additional services to improve the DOI management experience, making it easier for our [its] members to connect and share their DOIs with the broader research ecosystem and to assess the use of their DOIs within that ecosystem" (DataCite n.d. (b)). Initially, the MDS (Metadata Store) API (Programming Application Interface) was developed to allow users to register and manage DataCite DOIs and associated metadata, which seems more "tolerant" since some special characters or even symbols were also accepted in the fields of metadata. In May 2018, the DataCite Fabrica was launched as a new DOI creation and management platform to replace the MDS web interface (Dasler 2018). It "includes all the functionalities needed to manage repositories, prefixes, DOIs and their metadata" (DataCite, n.d. (c)) through duplicating the old functions of MDS and adding some new features, e.g., the addition of a metadata form, although Fabrica has more restrictions on the characters included in DOI. MDS and Fabrica had worked together for some time; then, at the

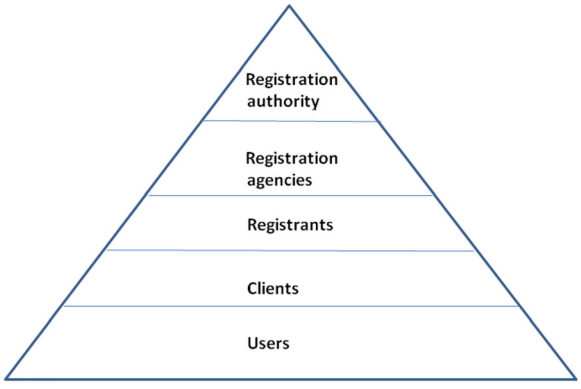

Figure 2: Pyramid of the DOI service community.

beginning of 2019, MDS exited the stage, with Fabrica working alone since.

Certainly, Crossref and DataCite have much in common but they serve the needs of different interest groups. According to the statement from Crossref, "Crossref makes research outputs easy to find, cite, link, assess, and reuse. We're a not-for-profit membership organization that exists to make scholarly communications better" while "DataCite's mission is to be the world's leading provider of persistent identifiers for research. Through our portfolio of services, we provide the means to create, find, cite, connect, and use research." It's suggested that an institution registering datasets should join DataCite, since "DataCite develop and support tools and methods that make data more accessible and more useful"; and an institution registering journal articles, conference papers, books, ebooks, preprints, peer reviews etc., should join Crossref by which "schema and reference linking infrastructure are set up specifically to support and provide services around these content types" (Crossref 2019).

The registrant locates at the next level in the hierarchic system, while sometimes the provider or allocator is taken as the equivalent. As a member of a RA, for example, DataCite, the DOI registration provider uses DOI Fabrica to manage its clients while the user of the DataCite could access its DOI Registration service through DOI Fabrica.

As regulated by DataCite, the client within the DataCite service system has the following responsibilities:

- Providing an infrastructure where researchers can preserve and share their research outputs;
- Implementing quality control to ensure academic publication standards;
- Ensuring that publications are original enough that they do not violate any existing copyright agreement (DataCite n.d. (d)).

Presently, lots of libraries have been involved in DOI services and they may play the role of either a registrant or a client (a data center) or a user in the DOI community or simultaneously. For instance, the University Library of Cologne provides the DOI registration service to the members of the University of Cologne for their online resources as a registrant/client cooperating with ZB MED; meanwhile, as a user, this library benefits a lot from the DOI services provided by IDF and DataCite.

And here is one another concrete example of the DOI services from the research data librarianship: via DataCite, "Stanford Libraries now offers the ability to assign a DOI (digital object identifier) to an information resource hosted online by Stanford" (Stanford Libraries n.d.). Via The Stanford Libraries, DOI services are offered at different levels either without any payment or with different prices.

3.2 DOI Services in Germany

DataCite originated from a project hosted at the German National Library of Science and Technology (TIB). In February 2019, Jan Brase, who was DataCite's executive officer for its first five years, posted a guest blog under the title "10 years of DataCite: How it all began" on the website of DataCite to tell "the story of how DataCite was founded 10 years ago" (Brase 2019). In this section, DOI services in Germany are taken as a case to show such services in the landscape of a country.

At the General Assembly of DataCite in December 2012, it was decided that the DOI service would be provided free to academic institutions from 2013 with four German members of DataCite (ZBW 2013) that take care of the DOI registration service for customers from different subject areas such as those below:

- Goettingen State und University Library (SUB GOE) – humanities
- German Central Library of Medicine (ZB MED) – life sciences
- Leibniz Institute for Social Sciences (GESIS) – social sciences
- Technical Information Library (TIB) – technology and natural sciences
- German Central Library of Economic Sciences (ZBW) and GESIS – economic sciences (Lindstädt and Pletsch 2016)

Usually, the DataCite member has clients related to it. However, the University Library of Munich (LMU) is registered as an independent DataCite member.

As Jan Brase in his review of the early history of DataCite pointed out, "an important part of the approach was the inclusion of libraries as registration agencies, thereby establishing a service open to all disciplines" (Brase 2019). Many German academic libraries, most of them university libraries, have actively participated in DOI services. The current DataCite's statistics shows that five of the six (i.e. ca. 83.3%) German members are libraries. Although the proportions of the libraries among these members' clients are not so high, the numbers of the registered DOIs from the libraries are remarkable and some

of them increased sharply in a great rate from 2018 to 2019. The growing tendency of the DOIs assigned by the University Library of Cologne graphically displayed in Figure 3 is one of the proofs of this description.

Until September 27, 2020 2331 DOIs were registered with University Library of Cologne. There was a great jump as regards to the number of the registered DOIs from 2018 to 2019, and the number of DOIs registered until September 2020 has been almost double that in 2019. More and more researchers and open journal publishers from the university have applied for registration of DOIs for their publications. In fact, although DOI assignment started in 2016 with the University Library of Cologne, so far the service is not completely officially publicized. It is presumed when the service is officially open to all members of the University of Cologne that the amount of registered DOIs with the Library will reach a new height.

4 The Status Quo

Jan Brase once mentioned in his blog that (research) data has been formerly questioned or laughed at. By contrast, research data management has been a hot topic for a couple of years. As he described, "the usage of DOIs of data is globally accepted and a new generation of scientists grows up for whom it is normal to have citable datasets" (Brase 2019). Scientists are "encouraged to review their publication strategies and to favour publication channels with established DOI assignments" (Gorraiza et al. 2016).

Another encouraging fact is that although the concept of DOI came into being at the middle of 1990s, the DOI metadata is completely in line with the FAIR Guiding Principles for scientific data management and stewardship published in 2016, which advocate [metadata] should be findable, accessible, interoperable, and reusable (GO FAIR n.d.).

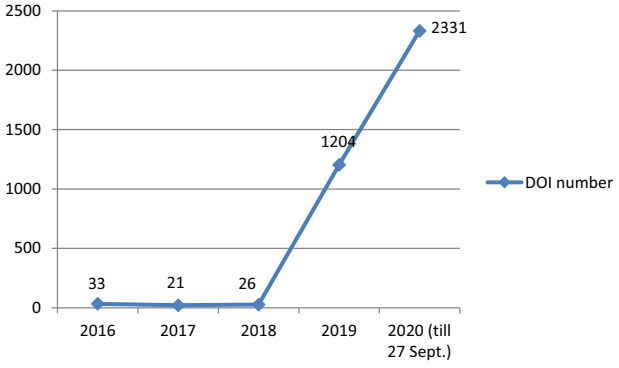

Figure 3: Statistics of DOIs registered by the University and City Library of Cologne.

4.1 DOI in Citation

Early in 2006, CrossRef has suggested including DOI in citations. In the document *CrossRef Guidelines: Using DOIs in Citations*, it listed various kinds of bibliographic formats for using DOIs to cite journal articles, books, and proceedings (Crossref 2006).

Later, in 2014, the Joint Declaration of Data Citation Principles (JDDCP) was published which requests that "a data citation must include a persistent method for identification that is machine actionable, globally unique, and widely used by a community (JAACP, principle #4)" (Fenner et al. 2019). DOI is one type of this PI and therefore itself or its expression in HTTP URL format has become an obligatory part of a data citation.

It is foreseeable that sooner or later a DOI would be a mandatory element in a standard citation style. Notably, it has been the reality to an extent. Many presses have set up regulations on citing a DOI in the citation. For example, the Atlantis Press which "is a professional publisher of scientific, technical and medical (STM) proceedings, journals and books," has proposed the following example on its website, since "a DOI is guaranteed to never change, so can be used as a persistent identifier to permanently link to an electronic article no matter where it is stored" (Atlantis Press n.d.):

> A. Fring, N. Manojlovic. G2-Calogero-Moser Lax Operators from Reduction. Journal of Nonlinear Mathematical Physics, 13 (2006), 467–478. doi:10.2991/jnmp.2006.13.4.1.

In the citation above, the position where a URL normally stands is occupied with a DOI and the accessing date is not necessary as before due to the persistence of the DOI.

Another example comes from a citation created on the basis of the American Psychological Association (APA) Style. The following is an extraction of the metadata interface of a journal article within Zotero (Figure 4), which "is a free and open-source reference management software to manage bibliographic data and related research materials" (Zotero n.d.).

Based on the metadata above, the bibliographic record in line with the APA Style is as below created, in which a URL automatically generated with DOI appears instead of the original URL that might change someday:

> Neumann, J., & Brase, J. (2014). DataCite and DOI names for research data. *Journal of Computer-Aided Molecular Design*, *28 (October 2014)*(10), 1035–1041. https://doi.org/10.1007/s10822-014-9776-5.

One similar application in the other direction is that the researcher with an ORCID (Open Researcher and Contributor Identifier) account could choose the route "Add DOI" when

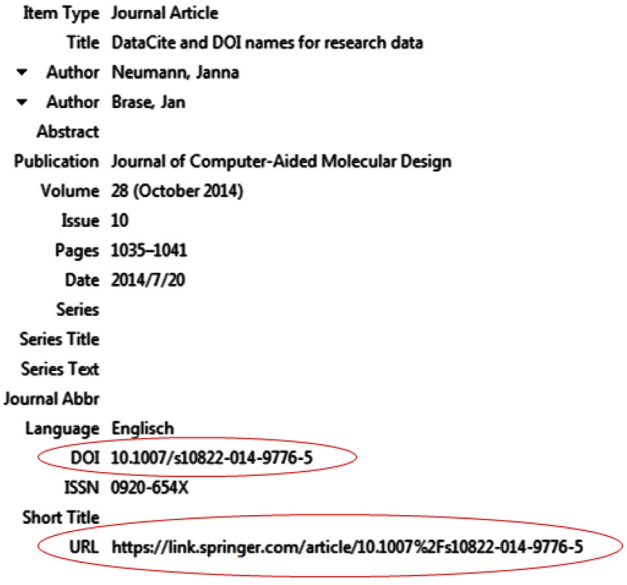

Figure 4: The interface of a bibliographic record in Zotero.

they intend to create a bibliographic record for his publication under his ORCID account (Figure 5). With this method, somehow some repeated work could be avoided in case the bibliographic information connected with the DOI could be found. The benefits of ORCID throughout one's academic career have been widely acknowledged in the scientific world. Therefore, ORCID has been a more and more common identifier for researchers: "ORCID works closely with Crossref, DataCite, and many other PID organizations to build trusted connections between IDs and other identifiers" (ORCID n.d.). ORCID and DOI seem to be both typical symbols of academic archives of modern researchers.

4.2 DOI in Bibliometric Databases

DOI has been adopted in some main streaming bibliometric databases, such as Web of Sciences, Scopus, and so on. For instance, in the database Web of Science Core Collection, DOI is one of the search options (Figure 6).

Juan Gorraiz and his colleagues implemented a study on the availability of DOI in Web of Science and Scopus and reached the conclusion that "the number of documents with a DOI is constantly increasing" in the two most important multidisciplinary databases (Gorraiza et al. 2016).

Meanwhile, more and more frequently, DOIs appear in OPACs of libraries (example refers to Figure 7).

In MAB2, a widely employed MARC format in German academic libraries, the field 552 for PI with the first indicator "a" is set up to contain the persistent identifier such as DOI shown as above.

Likewise, DOI is a part of the following bibliographic record from the repository PUBLISSO from ZB MED (Figure 8), which is "an open access publishing platform for life sciences" (Information Center for Life Sciences n.d.).

It is expected that DOI could be taken as a mandatory element in more and more bibliometric databases in the future.

4.3 Issues to be Noticed

The development of DOI and DOI service is a process of evolution during which services are improved, however, at the same time, errors or problems related to them are inevitable.

One study discloses some errors in DOI indexing by bibliometric databases, in particular, the incorrect assignment of a single DOI to multiple papers in Scopus (Franceschini, Maisano, and Mastrogiacomo 2015). Different errors of duplicate DOI names in Web of Science are also discussed in a recent publication (Xu et al. 2019; Zhu, Hu, and Liu 2019). The awareness of the bibliometric databases regarding such mistakes should have been raised.

And as described in the text above, the developer of the ORCID system has made efforts in making DOI more valuable in its system. Conversely, it seems that the developer of DOI, DataCite Fabrica, has tried to include the

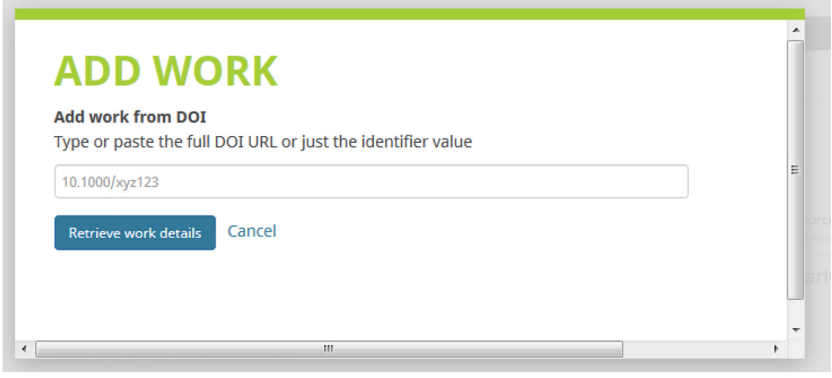

Figure 5: The interface of "Add work from DOI" on the platform of the ORCID system.

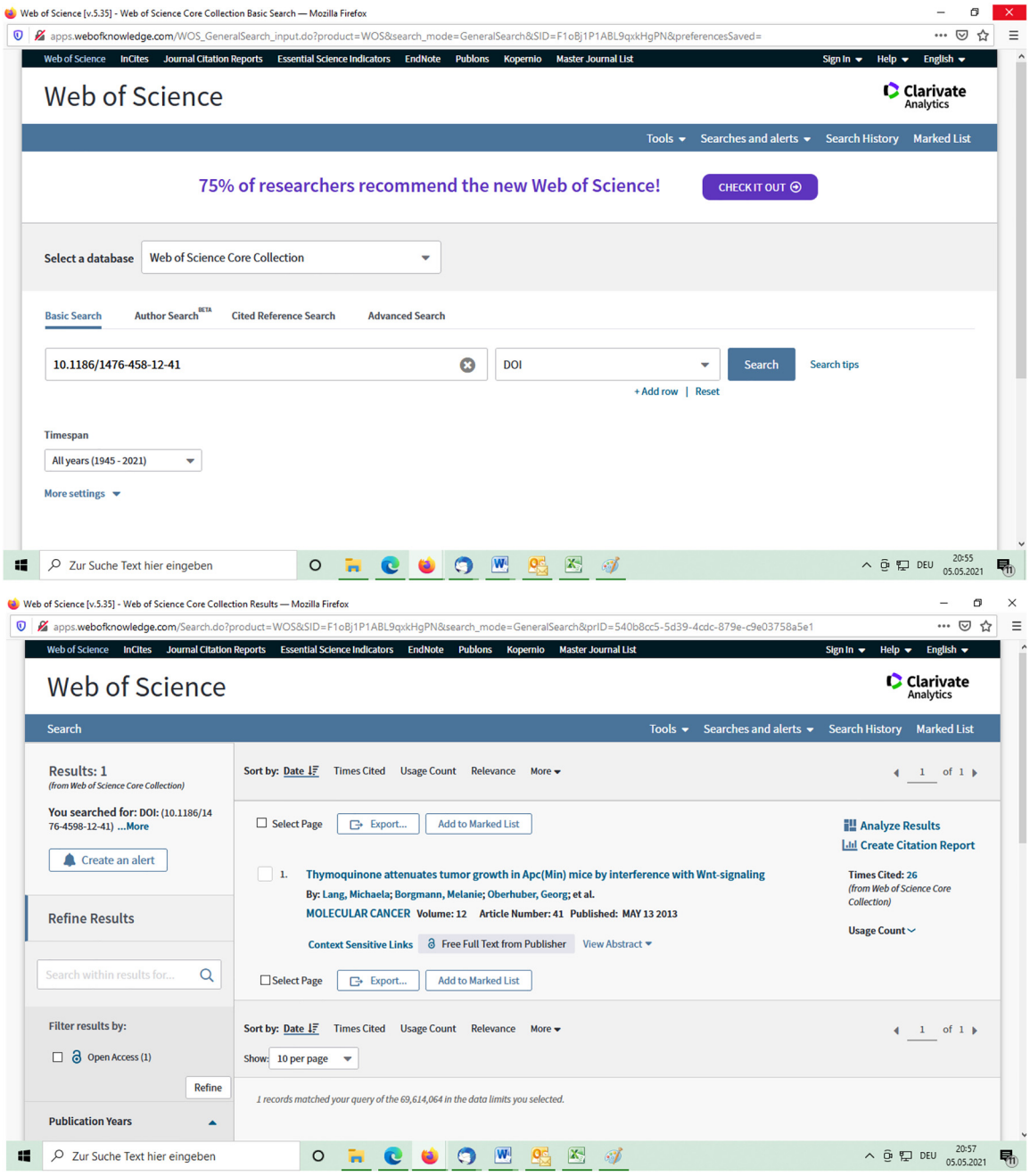

Figure 6: Search with DOI in Web of Science.

ORCID identifier in its system. In the practice, within the DOI Metadata Generator of DataCite, ORCID as the "name ID scheme" and the corresponding value for the creator could be adopted only until the first phase, when a DOI stays in "draft" state. However, the DOI with an ORCID ID could not be updated to its final status, i.e., "findable" status. The situation means that the DOI hasn't been successfully registered till the end so that it's searchable to the user out of the DataCite Fabrica system. Interestingly, the OCRID ID for the contributor is acceptable. Hopefully, in the near future, more namespaces such as ORCID could be accepted with the DataCite Fabrica.

One more issue dealing with the DataCite DOI services is related to the obsolete DOIs. So far, a DOI in the "findable" status registered on the platform of either the DataCite MDS or Fabrica could not be deleted. In case it is not necessary any longer, the status of such a DOI would be degraded to "registered" and then it is out of the searchable

Sel.-kennz.	078	e	a	ZDB-23-DGG-eba
1.Person/Fam.	100		a	Bauer, Oswald
Haupttitel	331		a	Zeitungen vor der Zeitung
Titelzusatz	335		a	Die Fuggerzeitungen (1568-1605) und das frühmoderne Nachrichtensystem
Verantw.-ang.	359		a	Oswald Bauer
Veröff.ang.	419		a	Berlin
			b	Akademie Verlag
			c	[2011]
Copyright-Datum	419	d	c	©2011
Ersch.-jahr	425	a	a	2011
Umfangsangabe	433		a	1 online resource
HT 1.Reihe+Zäh	451		a	Colloquia Augustana
			v	28
Bandang. Vorl.	455		a	28
ISBN	540		a	9783050051598
Persist.Ident.	552	a	a	10.1524/9783050051598
Elektr. Adresse	655	e	u	https://doi.org/10.1524/9783050051598
			A	0
Elektr. Adresse	655	e	3	Cover
			u	https://www.degruyter.com/cover/covers/9783050051598.jpg
			A	2

Figure 7: A part of a bibliographic record in HBZ.

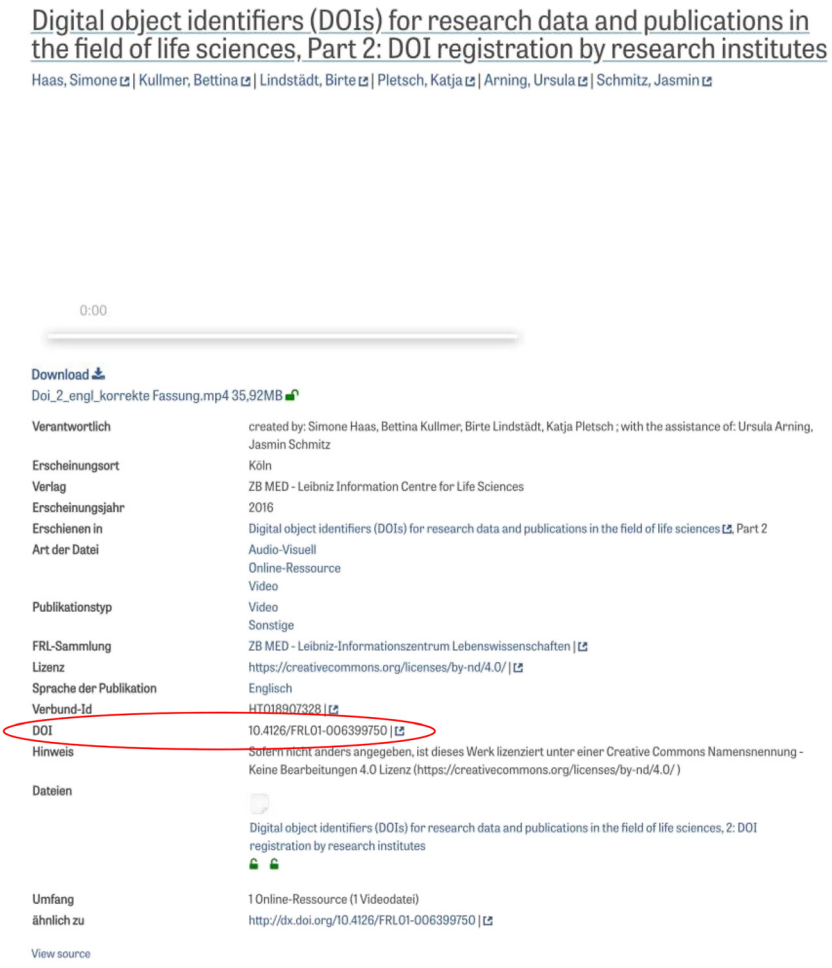

Figure 8: A bibliographic record from the repository PUBLISSO.

scope. Meanwhile, it is suggested to set up the tombstone page for this type of DOI. But the DOI still remains in the database and a great number of such invalid DOIs have accumulated with some institutions. For example, within the 483,869 DOIs registered with the British Library, 14,896 DOIs (about 3.08%) are obsolete while the proportion of such DOIs by the University Library of Ludwig-Maximilians University in Munich, Germany is astonishingly as high as 23.41% (4526–19,332).[1] The author wonders whether, in addition to the current solutions to deal with the abandoned DOIs, someday another solution could be achieved.

5 Conclusion

At the beginning of DataCite (15 years ago), DOI was called "evil" and thought of only as belonging to the publishers (Brase, Lautenschlager, and Sens 2015). However, much more attention has been paid to DOIs in recent years, and they have been widely accepted for some time. And although DOI was initially designed to be a digital identifier, it's not an identifier of only digital objects. It's foreseeable that the scope of the objects identified with DOI could be gradually expanded. DOI is being taken as a standard element in the identification system to promote the seamless scholarly communication, while more libraries would be involved in the work related to DOIs.

References

Atlantis Press. n.d. *Citing and Using a Digital Object Identifier (DOI)*. https://www.atlantis-press.com/using-dois (accessed April 9, 2019).

Brase, J. 2019. *10 Years of DataCite: How it All Began*. https://blog.datacite.org/how-it-all-began/ (accessed February 25, 2019).

Brase, J., M. Lautenschlager, and I. Sens. 2015. "The Tenth Anniversary of Assigning DOI Names to Scientific Data and a Five Year History of DataCite." *D-Lib Magazine* 21 (1/2). https://doi.org/10.1045/january2015-brase (accessed April 06, 2021).

CrossRef. 2006. *CrossRef Guidelines: Using DOIs in Citations. Version 1.0*. https://files.atlantis-press.com/file-downloads/CrossRef_Ref_Guide_1_0.pdf (accessed August 20, 2019).

Crossref. 2018. *Metadata Enables Connections*. https://www.crossref.org/services/ (accessed April 06, 2021).

Crossref. 2019. *DataCite*. https://www.crossref.org/community/datacite/ (accessed August 27, 2019).

Dasler, R. 2018. *DataCite 2018 Wrap-Up and 2019 Preview*. https://blog.datacite.org/2019-preview/ (accessed August 27, 2019).

DataCite. n.d. (a). *DataCite Statistics*. https://stats.datacite.org/ (accessed August 27, 2019).

DataCite. n.d. (b). *Getting Started*. https://support.datacite.org/docs/getting-started (accessed September 06, 2019).

DataCite. n.d. (c). *DataCite Fabrica Stage*. https://doi.stage.datacite.org/about (accessed April 06, 2021).

DataCite. n.d. (d). *DataCite Member and Client Responsibility*. https://support.datacite.org/v1.2.0/docs/community-responsibility (accessed August 20, 2019).

DOI Foundation. 1997. *A Guide to Using Digital Object Identifiers: For Creators, Publishers, and Information Providers*. http://www.doi.org/guidebook/1098-guidebook.html (accessed December 14, 2018).

Fenner, M., M. Crosas, J. S. Grethe, D. Kennedy, H. Hermjakob, P. Rocca-Serra, G. Durand, R. Berjon, S. Karcher, M. Martone, and T. Clark. 2019. "A Data Citation Roadmap for Scholarly Data Repositories." *Scientific Data* 6: 28.

Franceschini, F., D. Maisano, and L. Mastrogiacomo. 2015. "Errors in DOI Indexing by Bibliometric Databases." *Scientometrics* 102 (3): 2181–6.

GO FAIR. n.d. *FAIR Principles*. https://www.go-fair.org/fair-principles/ (accessed June 19, 2019).

Gorraiza, J., D. Melero-Fuentes, C. Gumpenberger, and J.-C. Valderrama-Zurián. 2016. "Availability of Digital Object Identifiers (DOIs) in Web of Science and Scopus." *Journal of Informetrics* 10 (1): 98–109.

Hakala, J. 2010. *Persistent Identifiers: An Overview*. http://www.persid.org/downloads/PI-intro-2010-09-22.pdf (accessed January 23, 2019).

Information Center for Life Sciences. n.d. *PUBLISSO: ZB MED Publication Portal for Life Sciences*. https://www.publisso.de/en/ (accessed September 10, 2019).

International DOI Foundation. 2013. *Factsheet. DOI® System and Standard Identifier Schemes*. https://www.doi.org/factsheets/DOI Identifiers.html (accessed April 10, 2019).

International DOI Foundation. 2015. *1 Introduction*. https://www.doi.org/doi_handbook/1_Introduction.html (accessed December 11, 2018).

International DOI Foundation. 2018. *Factsheet: Key Facts on Digital Object Identifier System*. https://www.doi.org/factsheets/DOIKeyFacts.html (accessed April 2, 2019).

International DOI Foundation. 2019. *DOI Registration Agencies*. https://www.doi.org/registration_agencies.html (accessed August 16, 2019).

ISO. 2006. *Terms of Reference for ISO TC 46/SC 9 Working Group 7: ISO Project 26324, Digital Object Identifier (DOI) System*. http://www.collectionscanada.gc.ca/iso/tc46sc9/wg7/wg7n1.pdf (accessed April 10, 2019).

ISO. 2012. *ISO 26324:2012(en): Information and Documentation – Digital Object Identifier System*. https://www.iso.org/obp/ui/#iso:std:iso:26324:ed-1:v1:en (accessed December 22, 2018).

Lindstädt, B., and K. Pletsch. 2016. *Neu bei der DOI-Registrierung?: ihr Weg zum DOI!* https://os.helmholtz.de/fileadmin/user_upload/os.helmholtz.de/Workshops/ws_datacite_helmholtz_ws1.pdf (accessed April 26, 2018).

Öchsner, A. 2013. *Introduction to Scientific Publishing: Backgrounds, Concepts, Strategies*. Heidelberg; New York; Dordrecht; London: Springer.

ORCID. n.d. *What are Persistent Identifiers (PIDs)?* https://support.orcid.org/hc/en-us/articles/360006971013-What-are-Persistent-identifiers-PIDs (accessed September 10, 2019).

1 Calculated on the basis of the document: DataCite n.d., "DataCite statistics [Registration by members]." Accessed September 6, 2019. https://stats.datacite.org/.

Risher, C. A., and W. A. Rosenblatt. 1998. "The Digital Object Identifier: an Electronic Publishing Tool for the Entire Information Community." *Serials Review* 24 (3–4): 12–20.

Sadeghi, A., C. Lange, M.-E. Vidal, and S. Auer. 2017. "Integration of Scholarly Communication Metadata Using Knowledge Graphs." In *Research and Advanced Technology for Digital Libraries* (*Lecture Notes in Computer Science*, Vol. 10450), edited by J. Kamps, G. Tsakonas, Y. Manolopoulos, L. Iliadis, and I. Karydis. Cham: Springer. TPDL 2017.

Schindler, U., M. Diepenbroek, and H. Grobe. 2012. "PANGAEA®: Research Data Enters Scholarly Communication." *Geophysical Research Abstracts* 14: 13378.

Sompel, H., S. Payette, J. Erickson, C. Lagoze, and S. Warner. 2014. "Rethinking Scholarly Communication: Building the System that Scholars Deserve." *D-Lib Magazine* 10 (9).

Stanford Libraries. n.d. *DOI Services*. https://library.stanford.edu/data-services/doi-services (accessed August 14, 2019).

Warner, S. 2005. "The Transformation of Scholarly Communication." *Learned Publishing* 18 (3): 177–85.

Wikipedia. 2019. *Persistent Identifier*. Last modified August 27, 2019 https://en.wikipedia.org/wiki/Persistent_identifier (accessed January 21, 2019).

Xu, S., L. Hao, X. An, D. Zai, and H. Pang. 2019. "Types of DOI Errors of Cited References in Web of Science with a Cleaning Method." *Scientometrics* 120: 1427–37.

ZBW. 2013. *DOI-Vergabe für Akademische Einrichtungen Kostenlos: Neuer Service der Deutschen DataCite-Mitglieder*. http://www.zbw.eu/de/ueber-uns/arbeitsschwerpunkte/forschungsdatenmanagement/dara/ (accessed August 24, 2018).

Zhu, J., G. Hu, and W. Liu. 2019. "DOI Errors and Possible Solutions for Web of Science." *Scientometrics* 118: 709–18.

Zotero. n.d. *Zotero*. https://www.zotero.org/ (accessed September 10, 2019).