Journal of eScience Librarianship

Volume 10 | Issue 2 Article 4

2021-03-24

Digital Object Identifier (DOI) Under the Context of Research Data Librarianship

Jia Liu
University of Cologne, Germany

ISSN 2161-3974

JeSLIB 2021; 10(2): e1180
https://doi.org/10.7191/jeslib.2021.1180

Journal of eScience Librarianship
putting the pieces together: theory and practice

Full-Length Paper

Digital Object Identifier (DOI) Under the Context of Research Data Librarianship

Jia Liu

University of Cologne, Cologne, NRW, Germany

Abstract

A digital object identifier (DOI) is an increasingly prominent persistent identifier in finding and accessing scholarly information. This paper intends to present an overview of global development and approaches in the field of DOI and DOI services with a slight geographical focus on Germany. At first, the initiation and components of the DOI system and the structure of a DOI name are explored. Next, the fundamental and specific characteristics of DOIs are described and DOIs for three (3) kinds of typical intellectual entities in the scholar communication are dealt with; then, a general DOI service pyramid is sketched with brief descriptions of functions of institutions at different levels. After that, approaches of the research data librarianship community in the field of RDM, especially DOI services, are elaborated. As examples, the DOI services provided in German research libraries as well as best practices of DOI services in a German library are introduced; and finally, the current practices and some issues dealing with DOIs are summarized. It is foreseeable that DOI, which is crucial to FAIR research data, will gain extensive recognition in the scientific world.

Correspondence: Jia Liu: liu@ub.uni-koeln.de

Received: November 13, 2019 **Accepted**: June 4, 2020 **Published**: March 24, 2021

Disclosures: The author reports no conflict of interest.

JeSLIB 2021; 10(2): e1180
https://doi.org/10.7191/jeslib.2021.1180

Introduction

A Digital Object Identifier (DOI) is a persistent identifier (PID or PI) used to identify an object uniquely in digital format. As a long-term globally unique identifier, it has been broadly adopted and employed all over the world, especially in the scientific domain. For years, it has been a standard element for identifying research data.

So far, little more than 20 years have passed after the International DOI Foundation (IDF) was launched as an organization for operating and governing DOIs in 1998. One year later, Crossref was established, which is currently the biggest DOI Registration Agency (RA) all over the world for scholarly information. DOI names have been assigned by multiple RAs from different fields worldwide and "over 27 million DOI short links to DOI names are in use" (International DOI Foundation 2019a).

With the tradition and priority in collecting, organizing, archiving, and providing access to information resources to support scientific activities, the research library community has been actively involved in research data management (RDM), which is described as "the approach taken to managing research data and other outputs of research across the research data lifecycle, and is a component of the much wider Open Science framework" (University College London 2020). Providing DOI services is one fundamental way for the research data librarian community to be involved in RDM.

On the basis of a comprehensive overview of DOI and DOI services, this study intends to explore them under the context of research data librarianship. First, the origin and structure of the DOI system and a DOI name are explored. The characteristics of a DOI and DOIs for three typical intellectual entities in the scholarly communication are described in this chapter; then, a DOI service pyramid is sketched in a general way with brief descriptions of functions of institutions locating at the different levels of the pyramid. Approaches of the research data librarianship community in the field of RDM, especially DOI services, are elaborated in the following part. As an example, the DOI services in German research libraries are discussed. At the end, the current practices and some issues dealing with DOIs are summarized.

DOI

DOI means a digital identifier of an object. "In computing, a DOI is a persistent identifier or handle used to uniquely identify objects, standardized by the International Organization for Standardization" (ISO 2012). The DOI system provides a system for persistent and actionable identifications and interoperable exchange of managed information in computer networks.

JeSLIB 2021; 10(2): e1180
https://doi.org/10.7191/jeslib.2021.1180

The DOI system and the DOI name

With the popularity of digital publishing, the problem of broken links to the digital content over the Internet becomes a continuing criticism of the web. DOI was designed to address the problem through assigning a unique identifier for an object separating from its location. The registry system of DOIs contains structured metadata and a resolution link to the actual content. If the location of the content changes, the resolution link can be easily updated and all existing links to it will still work. In October 1997, the DOI system was announced at the Frankfurt Book Fair and as mentioned above, IDF was set up to develop and manage the DOI System in the next year. 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). Although DOI originated in text publishing, developers of the DOI system have recognized the trend towards digital convergence and multimedia availability and the system was conceived as a generic framework for managing content identifications over digital networks.

Implementing the Handle System, a proprietary registry developed by the Corporation for National Research Initiatives (CNRI) for assigning PIDs, or handles, to information resources, and for resolving "those handles into the information necessary to locate, access, and otherwise make use of the resources" (Network Working Group 2003), the DOI system "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).

In 2010, the DOI system was approved as an ISO standard.

As regards to the syntax, a DOI name consists of two parts separated with a forward slash that are a DOI prefix and a DOI suffix. "The combination of a unique DOI prefix (assigned to a particular DOI registrant) and a unique DOI suffix (provided by that registrant for a specific object) is unique, and so allows the decentralized location of DOI names" (International DOI Foundation 2017). The DOI prefix is composed of a directory indicator that is designated as 10 so that a DOI could be distinguished from the other implementations of the Handle System, and a registrant code following the directory indicator and the two components are separated with a full stop (period). The DOI suffix consists of a character string of any length chosen by the registrant and customer. Each suffix must not be double to the prefix element that precedes it. The following graph (Figure 1) shows the structure of a DOI name with a hypothetical example.

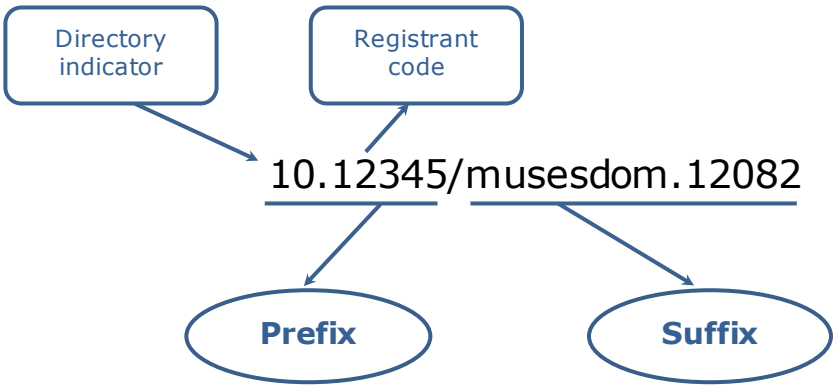

Figure 1: Structure of a DOI name

Meanwhile, DOI names could also be expressed as URLs (URIs) with the text string "https://doi.org/" prior to the DOI names. For example, the DOI name 10.116/S0098-7913(99)80002-1 could also be expressed as https://doi.org/10.116/S0098-7913(99)80002-1.

DOI's characteristics

DOIs are open identifiers for discovery and use of, and access to research entities. They have a couple of basic characteristics and several special features that make DOI superior to other PIDs and become an established identifier for research data.

Persistence and uniqueness are fundamental characteristics of a DOI. As mentioned above, a DOI is a kind of PID of an object. In a document released by the Research Data Alliance, a PID is defined as "a long-lasting ID represented by a string that uniquely identifies a DO [digital object] and that is intended to be persistently resolved to meaningful state information about the identified DO" (Wittenburg, Hellström and Zwölf 2017). A DOI keeps the nature of a PID. Every DOI is connected with a set of richly structured metadata that describe the item for which it is assigned. Once a DOI is registered, although all elements of the accompanying metadata, except the DOI name, could be updated, the DOI name itself could neither be cancelled nor altered any longer. The DOI name stays persistently; and a DOI identifies an object globally uniquely. One DOI is solely assigned to one entity and the DOI for one object differs from that for one another object. One DOI could definitely not be the same as the other one, even though it happens that the other one has been abandoned.

Furthermore, DOIs have following significant features:

- Flexibility

  A specified standard numbering syntax belongs to the core of the DOI system. The basic structure of a DOI name is determinate. As mentioned at the beginning of this article, a DOI name consists of a prefix and a suffix. Identifying the naming authority who assigned a DOI, the prefix is relatively stable. It begins with the directory indicator that is always 10

and an RA code with 5 numbers assigned by the registration authority follows. Nonetheless, the suffix could be flexible as long as it is unique to a given prefix and does not include invalid characters, letters, or symbols. In general, the suffix could be an alphanumeric string consisting of any numbers or characters with a small exception. It could be either randomly created with the computer or reflect some features of the identified object. In the suffix of a DOI name, the amount of the revealed information about the object and the depth of the description about the object could be freely defined by the registrant or the client or both of them.

- Actionability

  The DOI system is one for identifying content entities mostly in the digital environment and it assigns an actionable identifier for entities used in computer networks. The simplest action is to locate the entity that a DOI identifies with the automatically created URL after the DOI is registered. When a DOI expressed in a form of a URL is clicked, the web page of the resource the DOI identifies appears in the next browsing window. IDF declares that "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 n.d.).

- Resolvability

  A resolution service is also one of the components of the packaged DOI system so that DOIs can be resolvable. The resolution is provided through the Handle System developed by the Corporation for National Research Initiatives. Within the DOI system, a DOI name can be resolved to values of one or more types of data related to the object identified by that DOI name, such as a URL, an e-mail address, other identifiers and descriptive metadata. By clicking on a DOI link (i.e., the URL created on the basis of the DOI), one would be guided to the resource the DOI identifies. The DOI will continue to be resolved to the related object at its new location, in case its URL changes. Additionally, "the Handle multiple resolution allows one entity to be resolved to multiple other entities; it can therefore be used to embody e.g. a parent-children relationship, or any other relationship, and is therefore suitable for describing relationships of objects (data sets)" (International DOI Foundation n.d.).

- Interoperability

  The DOI systems enable an interoperable exchange of managed information in computer networks. It is to use and work with the existing identifiers and metadata schemes. Being designed to assist identifier interoperability, the DOI system offers a persistent, semantically-interoperable resolution to related current data and is best suited to materials that will be used in services outside the direct control of the issuing assigner (e.g., public citation or managing content of value).

JeSLIB 2021; 10(2): e1180
https://doi.org/10.7191/jeslib.2021.1180

The creation and use of persistent identifiers like DOIs are important parts comprising RDM practice and RDM provider services. In 2016, the "FAIR Guiding Principles for scientific data management and stewardship" (shortly as FAIR Principles) was published, in which "the authors intended to provide guidelines to improve the findability, accessibility, interoperability, and reuse of digital assets" (GO FAIR n.d.). Although the concept of DOI came into being in the middle of 1990s, DOIs meet the requirements of the FAIR Principles. Concretely, DOIs are fully consistent with the following aspects of the FAIR Principles:

- Findable

  F1. (Meta)data are assigned a globally unique and persistent identifier (PID)

  F2. Data are described with rich metadata

  F3. Metadata clearly and explicitly include the identifier of the data it describes

  F4. (Meta)data are registered or indexed in a searchable resource

- Accessible

  A1. (Meta)data are retrievable by their identifier using a standardized communications protocol

- Interoperable

  I3. (Meta)data include qualified references to other (meta)data

- Reusable

  R1.2 (Meta)data are associated with detailed provenance

In the first formal publication of the FAIR Principles written by Mark D. Wilkinson and his colleagues, the FAIR Principles' original authors, some exemplary implementations of the FAIR Principles in the scientific community are introduced (Wilkinson et al. 2016). Dataverse, an open source web application to share, preserve, cite, explore and analyze research data, is one among the reported examples of FAIRness in the publication. In Dataverse, researchers, data authors, publishers, data distributors, and affiliated institutions all receive appropriate credit via a data citation with a PID (DOI, or handle). Dataverse makes the PID public when the dataset is published. Referring to the image from Mercè Crosas related to a dataset landing page (Crosas 2019), the author draws Figure 2 to explain how FAIR data principles are implemented in the Harvard Dataverse in which DOIs are dominating PIDs utilized.

JeSLIB 2021; 10(2): e1180
https://doi.org/10.7191/jeslib.2021.1180

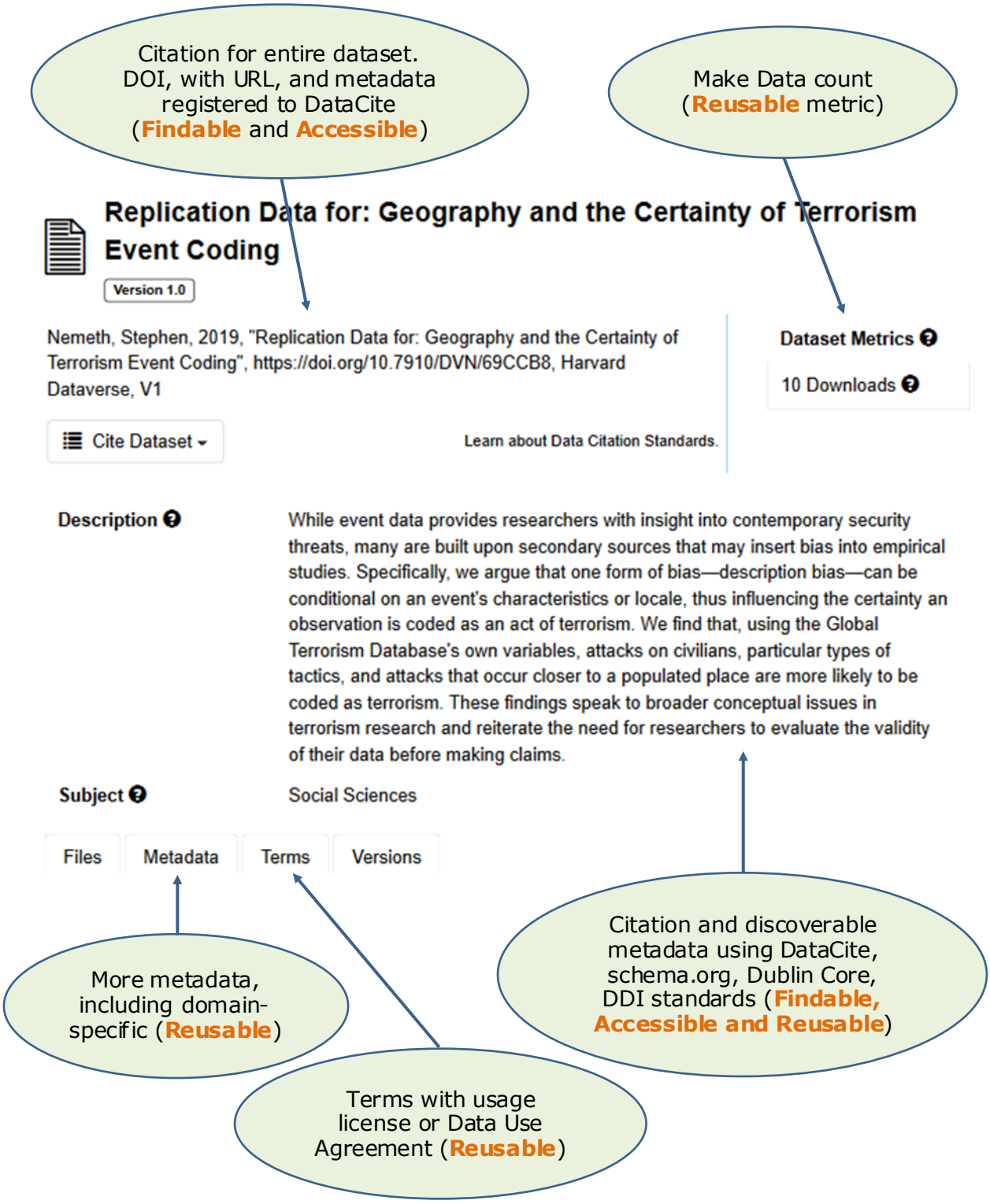

Figure 2: Implementation of FAIR data principles in Harvard Dataverse

JeSLIB 2021; 10(2): e1180
https://doi.org/10.7191/jeslib.2021.1180

Central to the realization of FAIR Principles are FAIR Digital Objects and their need to be accompanied with PIDs. In the final report and action plan from the European Commission Expert Group on FAIR Data, it was presented that "Digital Objects should be assigned unique and persistent identifiers such as a DOI or URN. This enables stable links to the object and supports citation and reuse to be tracked." (European Commission Expert Group on FAIR Data 2018) Experiences from the FAIR DO (Digital Object) Framework, which is a large research infrastructure built using FAIR DO, reflect that PIDs feature prominently in it. It was also introduced that "a PID, standing for a globally unique, persistent and resolvable identifier, is assumed to be the basis of the Internet of FAIR Data and Services" (Wittenburg 2019). A DOI is one of such PIDs and it is fundamental to FAIRness. DOIs play an outstanding role in the FAIR Metrics.

The features and characteristics of DOI make it one of the most widely known types of identifiers. Nonetheless, in addition to DOI, there are also other identifier types. In a primer publication of the National Information Standards Organization (NISO) of the United States, Carly Strasser (2015, 15) states that "researchers do not necessarily need to understand the nuances of identifiers, however, since the data repository often chooses the identifiers to be used". It is also mentioned in the ISO-document about the DOI system that "the DOI name does not replace, nor is it an alternative for, an identifier used in another scheme" (ISO 2012).

DOIs for a variety of objects

A DOI was designed to be a digital identifier of an intellectual entity. Nonetheless, the object a DOI identifies is not necessary to be a digital object. The DOI system provides an infrastructure for persistently and uniquely identifying objects of any type.

Ginny Hendricks and Rachael Lammey introduced the following list of the register contents with Crossref in a presentation at the Joint Global Infrastructure Conference:

- Journals
- Books
- Conference proceedings
- Standards
- Technical reports
- Working Papers
- Theses and dissertations
- Components (figures, tables)
- Datasets (supplementary data)
- Preprints (Hendricks & Lammey 2017)

In this section, due to the limited space, only examples of DOIs for the three most common object types in scholarly communication will be briefly discussed. The scope of objects that DOIs could identify is certainly far beyond that mentioned as below.

A book is one of the most important kinds of media in the scientific world. DOIs could be applied to identify books (mostly eBooks at this moment) persistently. On one side, undoubtedly, an International Standard Book Number (ISBN) and an International Standard Serial Number (ISSN) are the most common traditional standard identifiers of publications; and on the other hand, it is recognized that one of the most important tasks of PID is to "render the traditional identifiers actionable in the Web and provide persistent links to the resources" (Hakala 2010). Therefore, a DOI name could be designed as a "container" for an ISBN or ISSN, into which an ISBN/ISSN could be integrated. The new concept, "the actionable ISBN" (ISBN-A), has been put forward to connect an ISBN with a DOI. The ISBN-A is "a service powered by DOI®, in which an existing ISBN is expressed in the DOI System'' and it "is constructed by incorporating an ISBN into the allowed DOI syntax" (International DOI Foundation 2019b). Nowadays, it's more common that at least one traditional ISBN and a DOI containing the ISBN are provided in the copyright page of a newly published book. For example, in the copyright page of a book published by the publisher Springer, three numbers identifying the book could be found: the ISBN for the printed version (i.e., 978-3-642-39662-5), the ISBN for the eBook (i.e., 978-3-642-39663-2) and the DOI containing the ISBN of the eBook (i.e., 10.1007/978-3-642-39663-2). In this way, the DOI system, enables the ISBN to be actionable and resolvable.

Most commonly, DOIs are assigned to identify online journals and online journal articles in a format which fits to describe lots of details about them. Publishers are strongly recommended to use ISSN numbers as a part of the title-level DOIs so as to offer persistent and resolvable links to their journals. According to the webpage of the International Center of the ISSN, "the ISSN Standard, ISO 3297:2007, provides internationally accepted recommendations to use ISSN as suffix for title-level DOIs: 'To construct a DOI suffix using an ISSN, precede the ISSN (including the hyphen) with the lowercase letters "issn" and a period'" (International Standard Serial Number International Center n.d.). Crossref provides best practices in drafting DOIs including ISSNs (Wilkinson 2020a). For example, 1994-4683 is the ISSN of an online journal and the title-level DOI of the journal is 10.5930/issn.1994-4683. One other example of a DOI for an e-journal is 10.1045/dlib.magazine, which is a journal-level DOI assigned for D-Lib Magazine, an internationally respected online journal focusing on digital library research and development. Each article published within the journal has a unique article-level DOI, such as 10.1045/January2017-nuest providing more details about the article. In this DOI name, the issuing month (January) and publishing year (2017), and the family name of the author (Nuest) could be found in the suffix. In fact, D-Lib Magazine suspended publication of new issues in July 2017. This case shows how important and valuable DOIs are with persistence of intellectual entities. No matter if during online publishing, open access to, or long-term preservation of an online journal, a DOI is a crucial element.

JeSLIB 2021; 10(2): e1180
https://doi.org/10.7191/jeslib.2021.1180

As for DOIs for research data, a DOI of a dataset published in the database nmrshiftdb2 could be taken as a good example. In the DOI name 10.18716/nmrshiftdb2/77196, the second part of the prefix, i.e. 18716, is the registrant code standing for the RA—German Central Library of Medicine (shortly as ZB MED in German) and University and City Library of Cologne (USB Koeln, the German abbreviation), Germany; and the suffix starts with "nmrshiftdb2" indicating the involved database and after that is the sequence number of the dataset designated by the administrator of the database.

DOI services

DOI services discussed here refer to services that center on DOIs, such as registering and managing as well as resolving DOIs.

The DOI service pyramid

DOI services are provided at different levels by various institutions. A hierarchical system that crosses disciplines and regions exists within the DOI service community. The following pyramid (Figure 3) shows the typical hierarchy mechanism briefly.

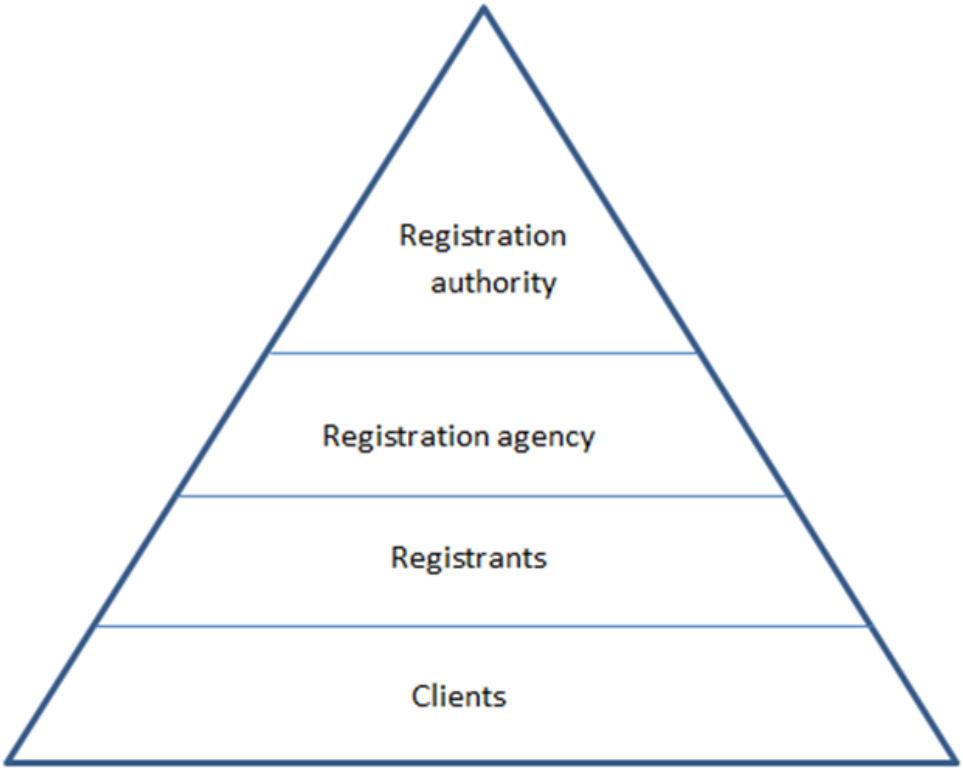

Figure 3: The DOI service pyramid

- Registration authority – at the top of the pyramid

  As the unique "DOI system registration authority and maintenance agency and the central body which governs the DOI system" (International DOI Foundation 2018), IDF stands at the top of the DOI service pyramid. "It is the common management and coordination body for DOI Registration Agencies" (International DOI Foundation 2018).

- Registration Agencies (RAs) – at the second level of the pyramid

  RAs are governed and managed directly by IDF and their primary role is to provide services to Registrants, such as allocating DOI name prefixes, registering DOI names and providing the necessary infrastructure to enable Registrants to declare and maintain metadata and status data.

  Currently, there are 10 RAs all over the world covering different areas and Crossref and DataCite are the two most well-known successful examples among them. Crossref (formerly styled as CrossRef) was grounded in 1999 and it's the largest DOI RA, while DataCite was established in 2009 and registered as "DataCite—International Data Citation Initiative" in Hannover, Germany (DataCite n.d.b).

- Registrant – at the third level of the pyramid

  Sometimes, provider or allocator is taken as the equivalent of registrant. A registrant can be any individual or organization that wishes to uniquely identify entities using the DOI system. It's not necessary for a registrant to be a member of IDF. However, the registrant must have an agreed relationship as a customer or client of a RA with which it registers DOI names. If a registrant has multiple types of content or application requirements, it may choose to contract with several RAs to provide services. It's a registrant's duty to ensure appropriate content management of their own materials (maintenance of URLs and data), either directly or by contract (e.g. with RA). The second part of the prefix of a DOI name is the numerical code standing for a registrant. Registrants may update the metadata for their DOIs in their repositories at any time, for example, when publication information changes or when an object moves to another URL.

- Client – at the bottom of the pyramid

  A client (or a data center) in the DOI service hierarchy system has direct contacts with researchers and different clients might have various responsibilities. For example, DataCite provides a list of general responsibilities of a client as below:

  - 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.a).

JeSLIB 2021; 10(2): e1180
https://doi.org/10.7191/jeslib.2021.1180

DOI services under the context of research data librarianship

The increasingly computational nature of research across all disciplines has resulted in both new challenges to and opportunities for libraries. The computer center is dominating in RDM and digital publishing, and the universal open access to the networked resources also challenges the library's traditional role as the center of the information resources. Librarians must rapidly evolve and expand their roles in the data-driven world. In order to keep standing in an important position of the research data ecosystem, the research library community has taken advantage of their knowledge and expertise in describing and indexing resources, building and providing access to collections, and providing support for the long-term stewardship of digital resources to support scientific reproducibility by ensuring that research data remains findable, accessible, interoperable, and reusable (FAIR). As Andrew M. Cox and Stephen Pinfield comment (2014), "RDM is a fascinating area of academic library activity" and it is "the center of many complex and changing expectations." The research data librarians need to identify which role they could play in RDM and figure out what skills, policies, and services they should develop to fit the role.

The Association of European Research Libraries (LIBER—abbreviation of its name in French) presents that "the FAIR Data Principles are essential for libraries who want to foster and extend research data services" (LIBER 2017). In reality, libraries play an important role in implementing FAIR principles. Recognizing that making research data FAIR is the basis of making science more open, transparent, and profitable, the research data librarianship community offers support in planning, organizing, and sharing FAIR research data to cover the entire data life cycle; helping researchers to implement the FAIR metrics in assessing their research data so as to make their research outputs FAIR. Mark D. Wilkinson and his colleagues once established an actionable design framework and exemplar metrics for FAIRness to assist scientists in improving the FAIRness of their research data (Wilkinson et al. 2018). In an RDM and an infrastructure project launched by the Fraunhofer Society, which is the largest organization for applied research and development services in Europe, DOI is implemented as the PID and the FAIR metrics is employed to create a FAIR research data infrastructure and culture (Beyan 2018). In addition, some libraries, for example, the Leibniz Information Centre for Science and Technology and University Library (shortly as TIB in German), have planned to "expand the offer of standardized metrics (citation and usage statistics) to research data, so that it could be shown how research data is reused" (Niemeyer 2019). DOIs feature prominently in these metrics.

Differing from other PIDs, DOIs "require a minimal amount of metadata to be provided at the point of obtaining each DOI" and also "a commitment from the provider to maintain the URL associated with the DOI" (Simons 2012). Research libraries could perfectly fulfill the requirements because of their tradition in providing or controlling qualified metadata, maintaining access to and preservation of research data. In his review of the early history of the development of DataCite,

Jan Brase once pointed out, "an important part of the approach was the inclusion of libraries, thereby establishing a service open to all disciplines" (Brase 2019). Although none of the 10 RAs in the latest list from IDF is a library, it has been very common for a library to provide services centered on DOIs. On one hand, more and more scientists recognize the necessity and importance of a universally unique identifier such as DOI and ORCID (Open Researcher and Contributor Identifier) for their research outputs; on the other hand, the research library community finds that it's a good chance for participating more in RDM through providing DOI services.

In the DOI service pyramid discussed at the beginning of this chapter, with the support from RAs that locate at the higher level, an academic library mostly functions as a registrant, and provides services to a researcher or a client which stays at the bottom of the pyramid to promote the FAIRness of research data. It's important to note that some research libraries play roles as both registrants and clients in that these libraries also self-publish entities online, such as journal articles, theses and dissertations, as well as digitized objects.

The most typical DOI service provided by a research library is assigning DOI names for research outputs. For instance, celebrating its DOI services for 10th anniversary in September 2019, the DOI Desk of the Library of the Swiss Federal Institute of Technology Zurich (ETH) is the official DOI registration office for Switzerland's universities and research sectors. As a DataCite registrant, it offers DOI registration services to organizational units (not individuals) from the field of higher education and research in Switzerland. It registers DOIs for digital objects that can be either primary data (research data) or secondary data (e.g. dissertations, working papers, articles, digitized resources) (Hirschmann 2019).

And no matter whether they register DOIs or not, a few academic libraries provide a DOI resolver, through which the user could be guided to the entity identified with a DOI. It also happens that a library provides other services related to DOIs. For example, with the service so-called DOI-Search (shown as DOI-Suche in its web page), the University Library of Heidelberg (UB Heidelberg, the German abbreviation) enables its user to check if UB Heidelberg holds the online journal identified with a DOI, and if so whether the library has a license for access to the full text of an article published in the journal (Universitaetsbibliothek Heidelberg n.d.). In addition to the DOI registration service, the DOI Desk of the ETH Library also provides such a service of DOI resolution. A DOI name could be resolved if it is typed or copied in the blank frame for resolving DOI names in the web page of the DOI Desk. The resolution leads to the corresponding document to be opened in the web browser (Hirschmann 2019).

With respect to the data center registered with a library, there are other solutions. For instance, data centers registered with TIB have two ways to have DOIs for their objects registered and metadata for DOI assignment uploaded. Data centers that only want to register a few DOIs per year can use the web-based user interface of DataCite Fabrica, while for larger registration quantities, automatic

DOI registration is a better solution (TIB n.d.). Before 2019, the automatic DOI registration could be realized via DataCite Metadata Store (MDS) Application Programming Interface (API). Nowadays, since DataCite MDS does not exist any longer, another API has been designed by Data Fabrica to achieve the function.

DOI services in the German research data librarianship community

Among the global efforts of RDM, the German research libraries have done much in promoting and improving RDM services. Annette Strauch advocates that the university library is a good partner in RDM in the practice. "A successful implementation of RDM and the creation of data management plans in the university libraries and their research-related services is grounded on specific support measures. University libraries provide assistance and answers to technical queries" (Strauch 2020). Cooperating with computer centers of the German universities, the German university libraries provide first-level consultant and support to technical inquiries related to metadata creation and metadata standards, data repositories, online publishing, and Open Access during the research process. At the same time, a couple of German university libraries, for example, the University Library Hildesheim, provides many training courses to researchers and frequently host webinars on the subject of RDM (Strauch 2020).

DOI services are one of the fundamental RDM services the German research libraries offer to their customers. In December 2012, it was announced at the General Assembly of DataCite that since 2013, on the platform of the DataCite Fabrica, the five (5) following German institutions would provide DOI registration services for free to customers from different subject areas:

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

Among the five institutions listed above, the first four (i.e., TIB, GESIS, ZB MED, and ZBW) had already been DataCite members before the Assembly. SUB GOE joined them later. And except GESIS, the other four institutions in the list are libraries.

In fact, DataCite sprouted from a project hosted at TIB, which acts in the capacity of the German National Library of Science and Technology, as well as architecture, chemistry, computer science, mathematics and physics. Under the title

JeSLIB 2021; 10(2): e1180
https://doi.org/10.7191/jeslib.2021.1180

"Publication and Citability of Primary Data" in English, the project was funded by the German Research Foundation (Niemeyer 2019). In a blog article written by Jan Brase, the effort from TIB in promoting the establishment of DataCite has been described in detail (Brase 2019). The growing demand on a kind of globally unique identifier for research data was recognized during the project and the first DOI for research data was registered in 2004 at TIB. In 2005, TIB became the world's first DOI RA for research data. Later, in 2009, DataCite was launched in 2009 as a distributed organization with a central business office located at TIB (DataCite n.d.c). From the beginning of its history, TIB has been one of the seven founding members of DataCite.

Standing at the leading position of the German librarianship, TIB actively promotes German libraries to be involved in DataCite's services.

With the leadership of and impact from these institutions, many German academic libraries, most of which are university libraries, have actively participated in providing DOI services. In the recent DataCite member list, four (4) of the German members are libraries. Although the library's proportion within the clients of these members is not yet prominent, the quantities of the DOIs registered with libraries are outstanding and some of them increase sharply at a great rate. The general growing tendency of the DOIs assigned by the USB Koeln, where the author works, revealed in Figure 4 is one of the proofs supporting this description.

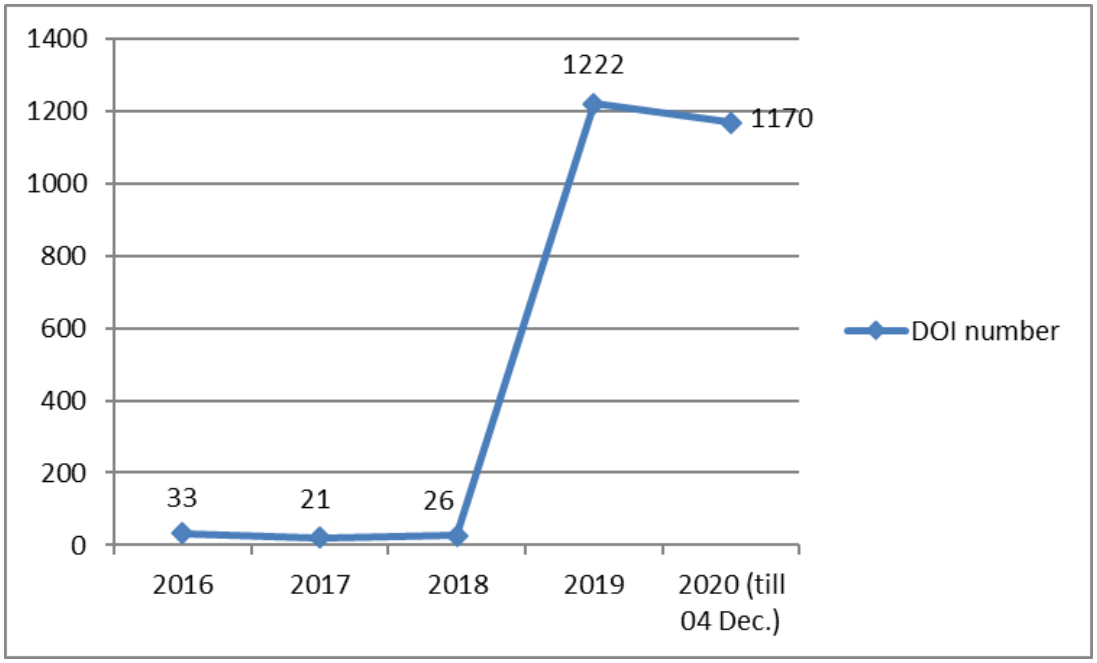

Figure 4: Statistics of DOI numbers registered with USB Koeln

Till December 4th, 2020, totally 2,472 DOIs have been registered with USB Koeln, which cooperates with ZB MED to provide DOI assignment service via DataCite Fabrica since 2016. The quantity of the DOIs registered with USB Koeln increases

dramatically from 2018 to 2019. Without the influence of the Corona crisis, there might be more requests for DOI registration in 2020. The most important reasons leading to the significant increase are that the request for DOI registration from the database nmrshiftdb2 overwhelmingly rises, and many more scientists of the USB Koeln apply for DOIs for the digital objects they created. The database nmrshiftdb2 is the continuation of a NMR (Nuclear Magnetic Resonance) database and it is maintained under a project of the Faculty of Chemistry, University of Cologne, Germany. The demand on DOIs for the datasets in nmrshiftdb2 increases quite a lot with the quantity of the datasets in the database growing quickly.

The best practices of DOI registration with USB Koeln could be briefly outlined as below:

1. A user fills all metadata describing the digital object in a form on the platform of DataCite DOI Metadata Generator and an XML file will be created at the end of this step. A DOI is one of the mandatory elements in the form and it is assigned by the user under the naming rules set up after the negotiation between the user and USB Koeln;
2. With a Handle ticket within which the metadata XML file and the URL of the object are included, the user sends a request for DOI registration for the object to the Open Technology Real Services (OTRS) system;
3. After receiving the ticket, a librarian of USB Koeln who manages DOI assignments, saves the metadata file in a local repository and registers the DOI with the method creating a DOI through uploading a metadata file.

Since April 2020, steps 1 and 2 have been implemented automatically with the datasets from database nmrshiftdb2 due to the great number of requests for the registered DOIs. An API has been programmed to enable the automatic DOI registration to accelerate processing of the massive requests for DOI registrations.

In addition to these three steps mentioned above, USB Koeln is also responsible for maintaining and updating metadata and URLs of DOIs while institutions take care of contents of the related digital objects.

Conclusion

At present, as Jan Brase finds, "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 favor publication channels with established DOI assignments" (Gorraiza et al. 2016).

A DOI has been a mandatory element in a standard citation style. In the main stream bibliometric databases, such as Web of Science and Scopus, a DOI is included in a bibliographic record as long as it's available and searching a DOI is one of the search options.

Meanwhile, a DOI is connected with other open PIDs. ORCID provides a persistent digital identifier to distinguish one researcher from another one. Currently, ORCIDs for institutions are also available. With DataCite Profiles, "researchers can connect their ORCID profiles and automatically update their ORCID record when their ORCID ID is included in the metadata of a newly registered DOI" (DataCite 2018). In another direction, it's declared that "ORCID works closely with Crossref, DataCite, and many other PID organizations to build trusted connections between IDs and other identifiers" (ORCID n.d.b). Figure 5 shows the possibility of supplementing a citation of a publication to the scholar record of a researcher through imputing a DOI. On 15th June 2017, Crossref, DataCite and ORCID co-hosted the Joint Global Infrastructure Conference at the Korea Chamber of Commerce & Industry (ORCID n.d.). More clues about their cooperation could be easily found on the Internet. Both ORCID and DOI have become typical symbols of academic archives of modern researchers.

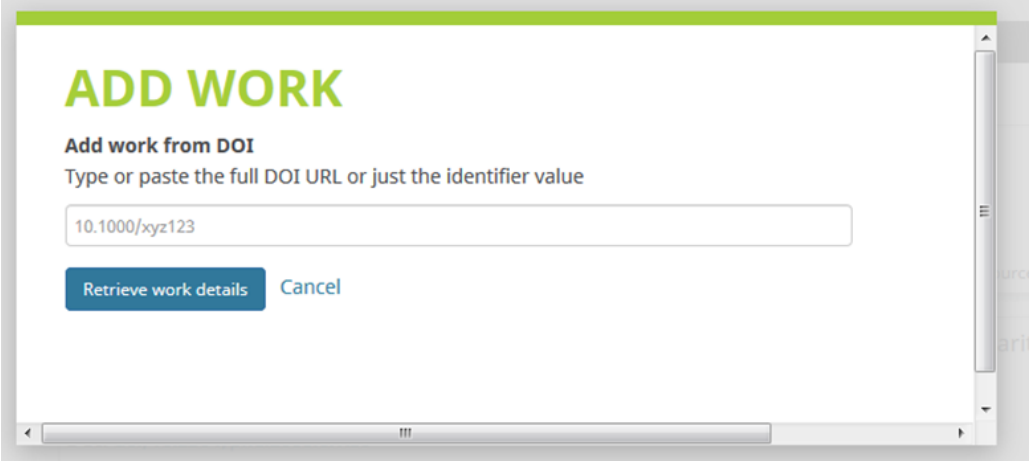

Figure 5: Interface of "Add work from DOI" on the platform of the ORCID system

And as the ISO standard specifies, DOI means a "digital identifier of an object" rather than an "identifier of a digital object" (ISO 2012). It's foreseeable that more and more types of intellectual entities would be identified with DOIs.

Some problems during the development of DOI and DOI services, such as repeated DOIs and obsolete DOIs, have been paid attention to. Most repeated DOIs assigned during the early history of DOI in the bibliometric databases have been corrected. And the current universal conundrum is the accumulation of abandoned DOIs. Various solutions including establishing tombstone pages have been put forward. Crossref addresses this issue in its own way: if the content of a Crossref member stops being maintained, Crossref is entitled to redirect the related DOIs to an archive or a Defunct DOI page hosted by Crossref (Wilkinson 2020b). Besides, persistent access to the content the DOI identifies is also a key issue with which to be concerned. DataCite preserves the metadata associated with the DOI and intends to retain its resolvability to not only the metadata in perpetuity, but also for the content files connected to the digital object and this can vary by client. In some IRs they have a public DOI request service which is

offered as a DataCite client, but their policy is that they do not mint a DOI for a researcher unless they have a copy of the associated content file, as this is the only measure they have to be able to ensure preservation of access, as well the client's ability to update the underlying URLs of DOIs that have been minted should the object URL ever change.

The importance and necessity of maintaining metadata, and especially URLs related to DOIs when the information changes, has been recognized because DOIs do not refresh themselves (Wilkinson 2020c). As Anna Tolwinska reminds the researcher, it's very important to update the URL connected with a DOI if the publisher of the digital object changes its location on the web. When the content is moved, its publisher should update its URL to keep the DOI functioning. The researcher can also make any minor corrections to his metadata or add additional metadata, such as funding and license information, at any time (Scholastica 2019).

The research data librarianship community should develop more roles in RDM (for instance, data curation, RDM training, etc.) and have close cooperation with computer centers and data centers. More research libraries should join in the group for providing DOI registration and resolution services as well as other related services, for instance, preserving the metadata and content of the digital object connected with a DOI when allowed. DOIs are proving very successful with the persistent identification of content. Research libraries should continue to explore new ways to enable the persistent access to content.

References

Beyan, Oya, Andrea Wuchner, Dirk Eisengräber-Pabst, Christoph Quix, Christian Zaschke, and Oliver Schumacher. 2018. "Research Data in the Fraunhofer Digital Project: Creating a FAIR Research Data Infrastructure and Culture." Presented at the *Workshop on Research Objects (RO2018) Workshop at IEEE eScience 2018 (RO2018)*, Amsterdam, Netherlands: Zenodo. https://doi.org/10.5281/zenodo.1441011

Brase, J. 2019. "10 years of DataCite: How It All Began." Accessed February 25, 2019. https://doi.org/10.5438/g4x3-r450

Cox, Andrew M. and Stephen Pinfield. 2014. "Research Data Management and Libraries: Current Activities and Priorities." *Journal of Librarianship and Information Science* 46(4): 299-316. https://doi.org/10.1177/0961000613492542

Crosas, Mercè. 2019. "The FAIR Guiding Principles: Implementation in Dataverse." Accessed May 19, 2020. https://scholar.harvard.edu/files/mercecrosas/files/fairdata-dataverse-mercecrosas.pdf

Crossref. 2018. "Find a Service." Accessed August 20, 2019. https://www.crossref.org/services

DataCite. 2018. "How Do I Activate ORCID Auto-update?" Accessed December 06, 2020. https://support.datacite.org/v1.1/docs/how-to-activate-orcid-auto-update

———. n.d.a. "Terms and Conditions." Accessed May 13, 2020. https://datacite.org/terms.html

———. n.d.b. "DataCite, Member and Client Responsibility." Accessed August 20, 2019. https://support.datacite.org/docs/community-responsibility

———. n.d.c. "Staff." Accessed June 4, 2020. https://datacite.org/staff.html

DOI Foundation. 1997. "A Guide to Using Digital Object Identifiers: for Creators, Publishers, and Information Providers." Accessed December 14, 2018. http://www.doi.org/guidebook/1098-guidebook.html

ETH Library. 2020. "ETH Zurich DOI Desk." Accessed May 28, 2020. https://library.ethz.ch/en/publishing-and-archiving/publishing-and-registering/doi-desk.html

European Commission Expert Group on FAIR Data. 2018. "Turning FAIR into Reality: Final Report and Action Plan from the European Commission Expert Group on FAIR Data." Accessed May 24, 2020. https://ec.europa.eu/info/sites/info/files/turning_fair_into_reality_0.pdf

GO FAIR. n.d. "FAIR Principles." Accessed June 19, 2019. https://www.go-fair.org/fair-principles

Gorraiz, Juan, David Melero-Fuentes, Christian Gumpenberger, and Juan-Carlos Valderrama-Zurián. 2016. "Availability of Digital Object Identifiers (DOIs) in Web of Science and Scopus." *Journal of Informetrics* 10(1): 98–109. https://doi.org/10.1016/j.joi.2015.11.008

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

Hendricks, Ginny. 2019. "DataCite." Accessed August 27, 2019. https://www.crossref.org/community/datacite

Hendricks, Ginny and Rachael Lammey. 2017. "Introduction to Crossref." Presented at the *Joint Global Infrastructure Conference*, Korea Chamber of Commerce & Industry, June 15, 2017. https://doi.org/10.23640/07243.5131786.v1

Hirschmann, Barbara. 2019. "10 Years DOI Desk at ETH Zurich." Presented at *ETH Zurich*, Zurich, Switzerland, September 13, 2019. https://doi.org/10.3929/ethz-b-000365054

International DOI Foundation. 2015. "1. Introduction." Accessed December 11, 2018. https://www.doi.org/doi_handbook/1_Introduction.html

———. 2017. "2. Numbering." Last updated: March 21, 2017. Accessed April 09, 2019. https://www.doi.org/doi_handbook/2_Numbering.html

———. 2018. "7. International DOI Foundation." Accessed August 16, 2019. https://www.doi.org/doi_handbook/7_IDF.html

———. 2019a. "DOI Registration Agencies." Accessed August 16, 2019. https://www.doi.org/registration_agencies.html

———. 2019b. "Factsheet. DOI® System and the ISBN System." Accessed August 16, 2019. https://www.doi.org/factsheets/ISBN-A.html

———. n.d. "Overview of DOI." Accessed May 4, 2020. https://www.doi.org/overview/041217DOIOverview.pdf

International Standard Serial Number International Center. n.d. "Use of ISSN for DOI Identifiers." Accessed December 4, 2020. https://www.issn.org/understanding-the-issn/issn-uses/use-of-issn-in-doi

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

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

LIBER. 2017. "Implementing FAIR Data Principles: the Role of Libraries." Accessed May 22, 2020. https://libereurope.eu/blog/2017/12/08/implementing-fair-data-principles-role-libraries

———. 2018. "Data Librarian, Expert on Research Data Management, Description, Archiving and Dissemination." Accessed May 24, 2020. https://libereurope.eu/blog/2018/05/14/data-librarian-expert-on-research-data-management

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

Niemeyer, S. 2019. "Happy Birthday DataCite!" Accessed May 8, 2020. https://blogs.tib.eu/wp/tib/2019/12/19/happy-birthday-datacite-en

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

ORCID. n.d. "Joint Global Infrastructure Conference". Accessed May 13, 2020. https://orcid.org/content/joint-global-infrastructure-conference

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. https://doi.org/10.1016/S0098-7913(99)80002-1

Scholastica. 2019. "Getting Started with DOIs at Your Journals: Interview with Anna Tolwinska, Crossref." Published April 16, 2016. Updated November 10, 2019. Accessed December 07, 2020. https://blog.scholasticahq.com/post/getting-started-with-dois-at-your-journal-interview-with-anna-tolwinska-crossref

Simons, Natasha. 2012. "Implementing DOIs for Research Data." *D-Lib Magazine* 18(5/6). https://doi.org/10.1045/may2012-simons

Strasser, Carly. 2015. *Research Data Management: a Primer Publication of the National Information Standards Organization*. Baltimore: National Information Standards Organisation (NISO). Accessed February 21, 2018. https://wiki.lib.sun.ac.za/images/2/24/PrimerRDM-2015-0727.pdf

Strauch, Annette. 2020. "Universitätsbibliotheken heute: Partner im Forschungsdatenmanagement in der Praxis = University Libraries Today: Partners in Research Data Management in Practice." *ABI Technik* 40(2): 177-186. https://doi.org/10.1515/abitech-2020-2008

TIB. n.d. "Information for Data Centres." Accessed May 26, 2020. https://www.tib.eu/en/publishing-archiving/doi-service/information-for-data-centres

Universitaetsbibliothek Heidelberg. n.d. "DOI-Suche." Accessed May 24, 2020. https://www.ub.uni-heidelberg.de/cgi-bin/doi.cgi

University College London. 2020. "FAIR data." Accessed May 24, 2020. https://www.ucl.ac.uk/library/open-science/fair-data

Wikipedia. 2019. "Persistent Identifier." Accessed January 21, 2019. https://en.wikipedia.org/wiki/Persistent_identifier

Wilkinson, Laura J. 2020a. "ISSNs and ISBNs." Last updated April 08, 2020. https://www.crossref.org/education/content-registration/descriptive-metadata/issns-isbns

———. 2020b. "Understanding Your Member Obligations." Accessed May 12, 2020. https://www.crossref.org/education/metadata-stewardship/understanding-your-member-obligations

———. 2020c. "Maintaining Your Metadata." Accessed December 07, 2020. https://www.crossref.org/education/metadata-stewardship/maintaining-your-metadata/#00556

Wilkinson, Mark D., Michel Dumontier, IJsbrand Jan Aalbersberg, et al. 2016. "The FAIR Guiding Principles for Scientific Data Management and Stewardship." *Scientific Data* 3, 160018. https://doi.org/10.1038/sdata.2016.18

Wilkinson, Mark D., Susanna-Assunta Sansone, Erik Schultes, Peter Doorn,Luiz Olavo Bonino da Silva Santos,and Michel Dumontier. 2018. "A Design Framework and Exemplar Metrics for FAIRness." *Scientific Data* 5, 180118. https://doi.org/10.1038/sdata.2018.118

Wittenburg, Peter. 2019. "Large Research Infrastructure Building using FAIR Digital Objects." Published on November 20, 2019. Accessed May 19, 2020. https://hdl.handle.net/11366/1223

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