Contents lists available at ScienceDirect

Journal of Informetrics

Availability of digital object identifiers (DOIs) in Web of Science and Scopus

Juan Gorraiz [a], David Melero-Fuentes [b,*], Christian Gumpenberger [a], Juan-Carlos Valderrama-Zurián [c]

[a] University of Vienna, Library and Archive Services, Bibliometrics and Publication Strategies, Vienna, Austria
[b] Catholic University of Valencia, Institute of Documentation and Information Technologies & UISYS (University of Valencia—Spanish National Research Council), Valencia, Spain
[c] University of Valencia, Department of History of Science and Documentation & UISYS (University of Valencia—Spanish National Research Council), Valencia, Spain

ARTICLE INFO

ABSTRACT

This study aims to shed light on the implementation of the digital object identifier (DOI) in the two most important multidisciplinary databases, namely Web of Science Core Collection and Scopus, within the last decade (2005–2014). The results show a generally increased percentage of items with DOI in all the disciplines in both databases, which provide very similar numbers and trends. While the percentage of citable items with a DOI has already reached 90% in the Sciences and the Social Sciences in 2014, it has remained much lower in the Arts & Humanities, exceeding 50% only since 2013. The observed values for Books and Proceedings are even lower despite the importance of these document types, particularly for the Social Sciences and the Arts & Humanities. The fact that there are still journals with a large number of items still lacking DOIs in 2014 should be alarming for the corresponding editors and should give them reason to enhance the formal quality and visibility of their journals. Finally, scientists are also encouraged to review their publication strategies and to favour publication channels with established DOI assignments.

© 2015 Elsevier Ltd. All rights reserved.

1. Introduction & background

A digital object identifier (DOI) is standard for unique and permanent online content identification and linking on the Internet. It is a unique alphanumeric string, assigned and governed by a registration agency (International DOI Foundation), which has been implemented since 2000 with on-going development (Morris, 1998; Surprenant, Blake, & Warwick, 1998; Schroeder, 1998; Simmonds, 1999; Mooney, 2001).

All DOI strings begin with a 10 and comprise of a prefix and a suffix separated by a slash. The prefix is a combination of the "10"-part, which identifies the DOI registry, followed by several (usually four) alphanumeric characters, which identify the applicant. In case the registrant is the International DOI Foundation itself, the second part of the prefix is 1000. The suffix identifies the specific object or content related to this identifier and is chosen by the registrant itself. Most legal Unicode characters are allowed in a case-insensitive manner, which means that it makes no difference whether uppercase or lowercase letters are used.

---

* Corresponding author.
  E-mail address: david.melero@uv.es (D. Melero-Fuentes).

The DOI is a permanent identifier for any kind of object or content (such as texts, images, tables, audio or video records, software, research data, etc.) in both electronic and physical forms and can refer to different hierarchical levels (e.g. journal title, article, table or image included in the article). However, the structure must be reflected in the metadata according to the indecs Content Model. Indecs is an acronym for "interoperability of data in e-commerce systems" and was a project funded by the European Community in order to provide an analysis of the metadata requirements for e-commerce content[1].

Much metadata information, including the digital object location on the Internet (such as a URL—Uniform Resource Locator), is stored within the DOI. The DOI remains unaltered for the lifetime of the object, whereas the metadata may change. Therefore, a DOI is a more robust linking option than simply referring to a uniform resource locator (URL) that may change due to negligent metadata maintenance by the publisher.

The primary purpose of the DOI system is not only the management of a collection of identifiers but also their operability and interoperability.

Organizations that meet the contractual requirements and are willing to pay a membership fee are authorized to assign DOIs. The federation of registration agencies is coordinated by the International DOI Foundation in order to implement and control the system, being a combination of the Handle System and the indecs Content Model susceptible to manipulation via a user interface. The Handle System ensures the invariability (the attributes of the object are encoded in the metadata rather than in its DOI, which remains invariable) and the uniqueness (the same DOI cannot be assigned to two different objects). The indecs Content Model is responsible for the metadata association. In case the location or describing information changes, registrants can update the metadata at any time via the user interface.

The main benefits of the DOI are to guarantee permanent and unambiguous identification of objects, to protect the copyright of material published on the Internet, to compensate content creators for their work, to keep track of content and to enable persistent citability.

Quoting Sidman and Davidson (2001), the DOI acts as a "persistent, actionable hyperlink .from the content back to the copyright owner that can travel with the file" and offers a sophisticated, transparent, and user-friendly DRM (Digital Rights Management) solution. "Even though online content does not have any sort of physical inventory, transportation, or logistics, there exists a complex chain of transactions including the sale, distribution, syndication, copyright protection, and re-use of this content".

The importance of the DOI is even gaining momentum for other document types like research data. As part of a project initiated by the German Research Foundation (DFG), the German National Library of Science and Technology (TIB) assigned its first DOI to scientific data in 2004 (Brase, Sens, & Lautenschlage, 2015). This initiative resulted in the foundation of DataCite[2] in 2009 (Neumann & Brase, 2014). During the past six years DataCite has grown into a global consortium and has so far assigned over four million DOIs to scientific datasets and other research artefacts (Brase et al., 2015). Recent studies have analysed the DOI implementation in products like the Data Citation Index (Peters, Kraker, Lex, Gumpenberger, & Gorraiz, 2015) and in the discipline Geosciences (Klump, Huber, & Diepenbroek, 2015).

Furthermore, the DOI plays a very important role for databases collecting their information from different sources in order to avoid unwanted duplicate entries and to enhance the accuracy and quality of the databases. Recent studies point us to possibly erroneous DOIs in Scopus (Franceschini, Maisano, & Mastrogiacomo, 2015; Valderrama-Zurián, Aguilar-Moya, Melero-Fuentes, & Aleixandre-Benavent, 2015).

Last but not least, the DOI is crucial for the application of new metrics, like altmetrics, an emergent field in Scientometrics (Costas, Meijer, Zahedi, & Wouters, 2012; Galligan & Dyas-Correia, 2013; Haustein et al., 2014; Haustein, Costas, & Lariviere, 2015). Although most of the tools, like Altmetric or Impactstory, also rely on other identifiers such as PubMed ids, arXiv ids or URLs, they require a DOI to achieve data accuracy (Jobmann et al., 2014). Other tools, like e.g. PlumX, also use the URL for data input, but such collected data always needs to be checked in order to avoid errors generated by potential URL instability (Peters et al., 2015).

Ducut, Liu, and Fontelo (2008) examined the availability of the URLs published in MEDLINE abstracts and estimated URL decay in these records from 1994 to 2006. The results revealed that about 81% of the URL pool was available for 90% to 100% of the time, but only 78% of these URLs contained the actual information mentioned in the MEDLINE record, and "dead" URLs constituted 16% in total. Moreover the authors demonstrated that only 519 (5.08%, $n = 10,208$) of all analysed abstracts had incorporated DOI addresses in their MEDLINE[3] abstract since the DOI was introduced in 2000.

The aim of this study is to shed light on the following issues:

1. The temporal course of DOI implementation in the main and most popular multidisciplinary citation databases Web of Science Core Collection (WoSCC) and Scopus in the last decade 2005–2014.
2. Differences according to publication types and subject areas.
3. The identification of subject categories and journals that still have not used DOIs or that have been indexed in these databases without any DOI between 2005 and 2014.

---

[1] http://www.doi.org/topics/indecs/indecs_framework_2000.pdf (accessed on August 2015).
[2] https://www.datacite.org (accessed on August 2015).
[3] MEDLINE is also included in Scopus.