Transform DOI system into a giant science hub

Khaled Moustafa
Editor of the Arabic Preprint Server (ArabiXiv)
Email: khaled.moustafa@arabxiv.org

In its current mode of identification of scientific publications, the digital object identifier (DOI) is not more than a web linking of published material to their publishing sources. When a given DOI is searched in the DOI website (doi.org), we are redirected to the publishing websites, if the material is available, or an error message (Not Found) will appear if the DOI-associated content is not available or has moved to a new location. To bestow a worthwhile value to DOI assignments, I suggest the establishment of a unique persistent DOI database (for e.g., as a DOI hub, DOI library, or DOI indexer) in which all the DOI assigned by publishers and journals will be listed in one and same place with basics bibliographic metadata and complete citation information, including the DOI link itself, authors' names, manuscripts' titles, publishing source, date of publication, and ideally abstracts and full text if available for free (open access). As a result, when a DOI is searched in the DOI hub, full bibliographic information should be retrievable regardless of its status in the publishing source. Basic indexation information and metadata associated with published articles will thus be always accessible and findable independently from the publishing sources. A unique, general and long-term preserved DOI hub will make it easy to search, find and cite scientific literature from the various scientific fields even if a journal or publisher ceases its publishing activity.

Keywords: digital object identifier; doi; persistent archive; bibliography database, science indexation; science archive; doi hub; doi indexer; doi library; science hub;

The digital object identifier (DOI), as its name suggests, is a numerical identifier of material or immaterial objects such as books and articles printed on paper or published online ([Paskin 2010](#)). Millions of DOI names are assigned yearly through a group of accredited DOI registration agencies such as the Crossref, which started to assign DOIs to scholarly communications in the year 2000, and it is now used by more than 4,800 members ([Digital Object Identifier, 2021](#)). In January 2021, the number of DOIs assigned by Crossref exceeded 120 million records ([Crossref, 2021a](#)).

However, the DOI is not an essential part of the publishing process – it is not a proxy for publications' quality–, but it is supposed to facilitate cross-citation of scholarly publications and to make publications easy to find, link and reuse ([Crossref, 2021b](#)). Its goal is to simplify

the management of intellectual contents on a digital environment ([Paskin 1999](#)). The DOI is often presented to be persistent, unique, resolvable and interoperable ([Paskin 2010](#)). Nonetheless, one of the requirements that DOI resolves properly (redirect users to the source of information) is that the host (i.e., the journal's or publisher's website) must be maintained alive so that the DOI links must always resolve to a response page ([Crossref, 2021b](#)). A third-party archive should therefore be used to preserve the DOI-identified content in a case that a journal or publisher ceases to host the corresponding content ([Crossref, 2021b](#)). As such, the DOI is not an indexer or aggregator, but a simple linker to a publication's website. In such a case, when a journal moves or shuts down, the assigned DOI links will not resolve properly, unless the DOI-identified content is repointed out to the right location that contains it. In other words, the current state of DOI functioning is not as persistent and useful as it is supposed to be. It is merely a form of URL (uniform resource locator) that users click to go the publishing journal/source without a real value over URL links. This mere location/linking function should raise questions about the real usefulness of the DOI system over web linkage.

Therefore, to avoid such hurdles and to add a worthwhile value to the DOI/Crossref system, I suggest that all the DOI records assigned by publishers, journals, DOI registrant of any type should be listed/archived in one and same place (for example in a framework called "DOI indexer, DOI library, DOI archive, DOI repository or DOI hub"). The indexation should be done with at least minimal bibliographic and complete citation information, including the DOI link itself to the original source, authors' names, manuscripts' titles, publishers' or journals' names, dates of publication, and ideally with abstracts/full text whenever available for free. The links to full texts of DOI-identified contents should also remain resolvable to the original publishing sources (journal, publisher, etc.).

Listing or archiving all the DOI assigned by different journals, publishers and other DOI registrars (for digital objects) in one and same place will make a global and giant science/DOI hub. Within such a universal and unique DOI indexer, it will be easy to search, find and retrieve basic bibliographic information on any DOI-assigned content or object. A further advantage is that even if a journal or publisher ceases its publishing activities for any reason, or if it moves into a new website or address, the citation metadata associated with the DOI itself will remain always available on the DOI hub, regardless of the situation of the publishing source.

The International DOI Foundation and its registrant agencies, particularly Crossref, are well positioned to develop such a DOI hub either on their current networks: doi.org and/or crossref.org or on a new dedicated framework. Users can then get the complete bibliographic information from the new DOI hub by searching publications by their titles, authors' names or DOIs, in an independent way from publishers' websites. A unique and

general DOI library will be complementary to other existing bibliography databases, such as PubMed, but it will certainly be larger than PubMed as PubMed is mainly dedicated on biomedical literature while a DOI hub will be a huge multidisciplinary and multilingual bibliography database with a proven utility than to merely linking publications to external publishers' webpages as is currently the case with DOI links.

References

CrossRef, 2021a (<https://www.crossref.org/blog/new-public-data-file-120-million-metadata-records/>), accessed July 01, 2021

CrossRef, 2021b (<https://www.crossref.org/membership/terms/>) accessed July 01, 2021

Digital Object Identifier, 2021 (https://www.doi.org/doi_handbook/1_Introduction.html) accessed July 01, 2021

Paskin, N. (1999). "The digital object identifier system: digital technology meets content management." Interlending & document supply.

Paskin, N. (2010). "Digital object identifier (DOI®) system." Encyclopedia of library and information sciences **3**: 1586-1592.