

The use of DOI within catalogues for scientific resources

di Piero Attanasio

[1.12.2003]

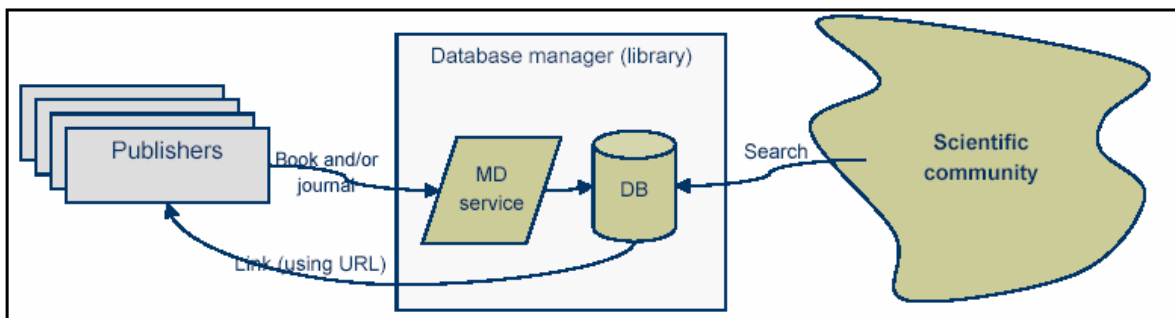
doi: 10.1392/BC2.0

The DOI (Digital Object Identifier) is a standard identifier for any Intellectual property entity in the Internet. Many applications have been implemented using the DOI (see www.doi.org) and thus people may tend to forget that the first value of the DOI is precisely to be a unique identifier, which facilitates the communication between the IT systems of parties in the publishing value chain. This includes also the interoperability between publishers and library world.

The show-case that is here described concerns the possible collaboration between the mEDRA system and scientific resources databases and may apply to different situations, such as the bio-medical databases provided by the "Istituto Zooprofilattico Sperimentale della Lombardia e dell'Emilia-Romagna" in Brescia; the Italian mathematical journals database provided by the SIBA of the University of Lecce; and the Italian legal journals database provided by the "Istituto di Documentazione Giuridica of the CNR" (Italian national research centre).

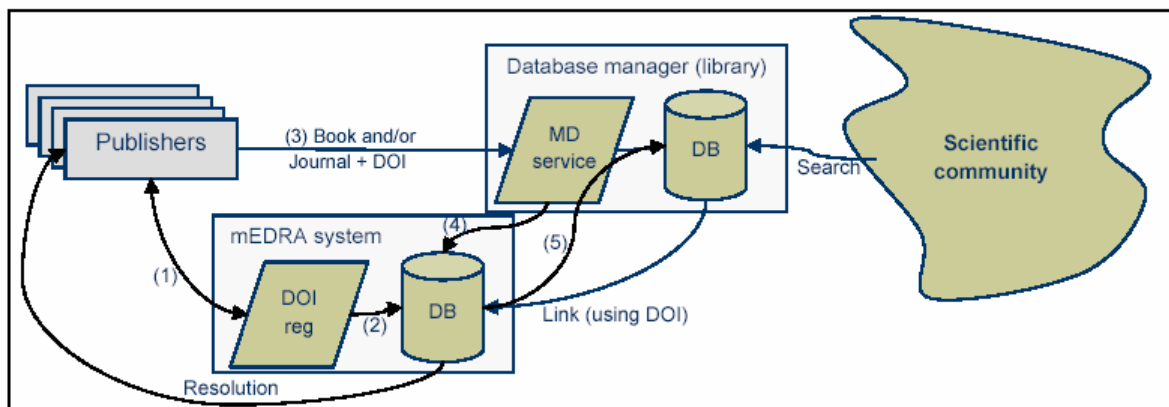
Such databases collect bibliographic data on journal articles and are a key access point to bibliographic information for Italian scientific communities. Figure 1 shows the current production process for the creation of the databases. Publishers send the journal or the book to the library, which creates metadata on individual item "books-in-hands". If the resource is available on line, the library includes in the database the URL to create a link directly from the database to the resource.

Fig. 1 – Current production process for creation of bibliographic database



When a publisher registers a DOI on any digital content (either journal article, or e-book, or individual book chapter) he also provides some metadata to mEDRA. Such metadata are structured according to mEDRA metadata schemas¹. It is possible to set up a new process to fuel the bibliographic database (Figure 2).

Fig. 2 – Possible new production process and value added



The new workflow may be described as follows:

1. Publishers assign DOIs on their content.
2. At the same time, they register the metadata in the mEDRA database
3. When publishers send the item to the library they communicate also the relative list of DOIs
4. The database manager submits the list of DOIs to mEDRA...
5. ...and receives back the metadata that may fuel directly the bibliographic database. The metadata service of the library just validates such metadata and possibly add further pieces of information that are necessary for the specific services offered to final users.

The use of unique identification system and standard metadata schemas enable all the actors involved to exchange data in automatic way, through batch procedures that have to be established not for every publisher but between mEDRA and the database manager.

Furthermore, the DOI may be used by the database manager to create a link for the benefit of final users. The DOI-based link will exploit the persistence functionality of the whole system and this allows the metadata manager to provide better service to users.

The link between scientific databases and mEDRA may facilitate also the DOI assignment to the publishers backlist. This is a delicate issue for many publishers since it is possible that the metadata for backlist items are not as well structured as for the current ones. In such case, it is possible to figure out that a publisher - once he/she desires to assign DOIs to backlist items - asks the database manager to send the relative metadata to mEDRA for the registration. The use of standard information tool facilitate bi-directional communication.

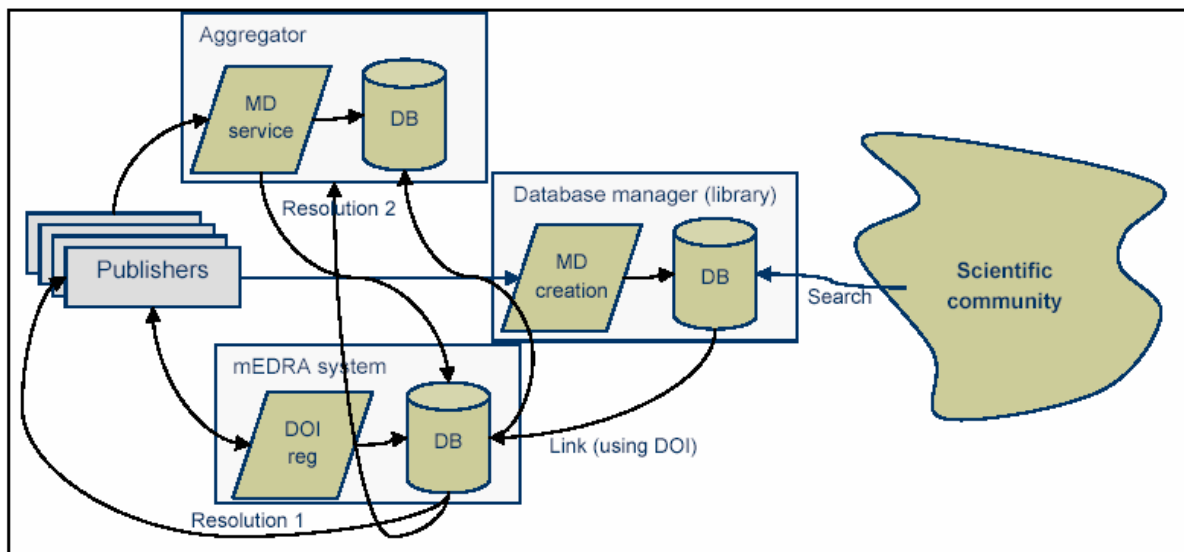
¹ see <http://www.medra.org/en/schema.htm>

It is to be noted that DOIs may be assigned to any IP entity, and thus also to articles that have only a printed version. In such case the resolution will be done to any relevant piece of information about the content (e.g. an abstract accessible on line, a service to order the article etc.). Publishers will control such resolution, being the DOI registrant.

Possible developments

The model is scalable for further development. For instance, a digital aggregator may exploit the existence of structured metadata and identification scheme to aggregate content on a particular issue acting as secondary publisher (Figure 3). Also in this case a direct communication between the mEDRA database and the aggregator one may speed up the process. Publishers will be facilitated to access the aggregator services² by using the same identification system.

Fig. 3 – Possible implementation involving secondary publishers



In such case, it is possible to see that a multiple resolution requirement and, for users (particularly libraries), an appropriate copy problem will arise. When a scholar queries the bibliographic database and find a content, the DOI system make possible to handle multiple resolutions. In our case, Resolution 1 will be to the publisher website and Resolution 2 will be to the aggregator service. What is the best source to access the content is up to the user. Let suppose that the query comes from a university library that has subscribed the access to the aggregator database, while in the publisher website the same content is available in “pay-per-download” form. Multiple resolution may facilitate the choice between the two options: the library system itself may control that the individual user will access the appropriate copy, avoiding double payment for the same content.

² A demonstration for such business is the relation between Casalini Digital Division and mEDRA: <http://dx.medra.org/10.1392/BC1.0>