

First Light for DOIs at ESO

Dominic Bordelon^{1,*}, Uta Grothkopf¹, and Silvia Meakins¹

¹European Southern Observatory, Karl-Schwarzschild-Str. 2, 85748 Garching near Munich, Germany,
library@eso.org

Abstract. Digital Object Identifiers (DOIs) are a helpful tool for academic research, and their use continues to grow. DOIs assist in citing and identifying research data, as well as in combating link rot. As a producer of large amounts of astronomical data, ESO has become interested in creating DOIs for its datasets. However, there are significant considerations to be undertaken in reaching this goal, and some local infrastructure is required. The ESO Library has planned, designed, and implemented a local solution for creating DOIs, consisting of an application, the ESO DOI Service, which other departments can use. This paper discusses the requirements, development process, and features of the ESO DOI Service.

1 Introduction

Digital Object Identifiers (DOIs) are an increasingly useful tool in research and scholarly communications. As a producer and provider of large amounts of astronomical data, ESO has become interested in DOIs for enhancement of the usability and visibility of its data. Within ESO, beginning in mid-2015, the Library has led the effort in planning, designing, and implementing a local solution for creating DOIs. In early 2017, the first ESO DOIs were created, and work continues towards the creation of more DOIs and better software features.

This paper describes how DOIs have come to be used at ESO, the software employed to do so, and future plans for DOIs at ESO.

2 Digital Object Identifiers

Digital Object Identifiers (DOIs) are persistent and globally unique identifiers designed to facilitate long-term access to resources. Although DOIs have many potential applications, they have been used particularly in a scholarly context since their inception in 2000. As of March 2017, roughly 143 million DOI names have been assigned.¹ A DOI is created (or “minted”) by a client via a Registration Agency (RA), which in turn is a member of the International DOI Foundation (IDF). The IDF is a nonprofit organization which acts as the central authority and maintenance agency for the DOI system. DOIs and the DOI system are described by ISO Standard 26324:2012[1], and the DOI Handbook published by the IDF is the authoritative source for information about DOIs.[2]

*e-mail: dominic.bordelon@eso.org ORCID: 0000-0002-8382-1300

¹<http://www.doi.org/factsheets/DOIKeyFacts.html> (accessed July 2017)

A DOI is conceptualized as a *digital identifier of an object* rather than an *identifier of a digital object*. It consists of the DOI name—the identifier string itself—as well as accompanying metadata which describes the resource. DOI names have a syntax (see Figure 1) which follows the pattern $10.1234/abc-678/90.1$, where: $10.$ is common to all DOIs; 1234 represents the registrant of the DOI in question; $/$ (forward slash) is a separator between the prefix and suffix; and $abc-678/90.1$, the suffix, is an arbitrary (but unique) string determined by the registrant. The accompanying metadata helps users—human or machine—make use of the described resource, and is implemented through a minimal metadata schema, the DOI Kernel, which Registration Agencies are free to supplement according to specific needs or practices.

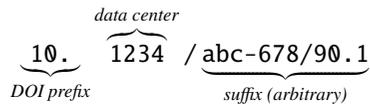


Figure 1. DOI name structure

Besides serving as a unique, unambiguous identifier for a resource, DOIs are also “actionable” or “resolvable” in the sense that a resource can be retrieved via the internet by prepending <https://doi.org/> to its DOI name. When the user requests this URL, the doi.org proxy server responds with the resource’s actual URL, and the user is seamlessly redirected. This resolution therefore promises a persistent link to the resource, which is ensured through the system’s technical infrastructure as well as the obligations undertaken by registrants. In practice, the resolution via doi.org does not return the resource (i.e., data file) to the user, but instead a “landing page.” This landing page displays useful metadata, along with a link to the resource itself (e.g., a PDF file for a journal article) if available, perhaps in multiple file formats. The address of the landing page and the resource file(s) themselves may change, but the registrant is obligated to maintain up-to-date information with the DOI system so that resolution continues to function properly. Persistence is a key goal not only of the DOI name, but also of the resolution service, in an effort to combat the problem of link rot (see e.g., Klein et al. 2014[3]).

The metadata component of the DOI system enables interoperability and the creation of additional services. For example, CrossRef, a well established RA for scholarly publishers, offers services such as reference linking, “Cited-by” (tracking resources that have cited the resource in question), and a registry of research funders. DataCite, an RA specialized for research data, offers similar services, and collaborates with CrossRef to link metadata between their respective repositories.

Organizations wishing to mint their own DOIs must fulfill the following requirements: become members/clients of an RA (which is a member of the IDF); upload formatted metadata for each resource to the RA; serve a landing page for every resource; and maintain both the resource’s record in the DOI system and the target landing page URL, to which the resolver redirects. It is important to note that, by minting a DOI, the organization is making a promise of persistence of the resource, its associated metadata, and the URL association. Organizations should also expect pay for the RA’s service, e.g., a per-DOI fee.

3 ESO’s interest in DOIs

In 2015, the ESO Library began to take interest in adopting DOI creation at ESO. Such adoption would facilitate data usage and sharing, assist the organization in tracking usage of its data, and

further advance a commitment to long-term data availability. Two areas that would seem to benefit were astronomical data and publications.

As a highly productive research data center, ESO produces and shares a large amount of astronomical data, amounting to a current total of over 1.5 million images and spectra, occupying approximately 65 terabytes of disk space.² The ESO Science Archive Facility offers both raw data and “data products,” packages of pre-processed data, to the scientific community.

ESO’s Education and Outreach Department (ePOD) produces a quarterly publication, *The Messenger*, which presents ESO’s activities to the public.³ Each issue contains roughly a dozen articles, covering the topics of instrumentation, science, and organizational news.

The Library initiated conversations with these departments, and they agreed that minting DOIs would be beneficial. The next step was to make an arrangement with an RA. Since ESO anticipated using DOIs primarily for datasets (as opposed to publications), DataCite was chosen due to its prominence in this area. This choice was further facilitated by the arrangement between the Technische Informationsbibliothek (TIB; a German national library for sciences, engineering, and math based in Hannover) and DataCite: TIB allows research centers in Germany proxy membership to DataCite. The Library was eager for ESO’s signatory to be a member of the upper administration, to represent an institution-wide commitment to the responsibilities engendered by DOI minting. In March 2016, a contract was signed between ESO’s administration and TIB for the creation of DOIs.

4 Planning and points of consideration

In planning to mint DOIs, several aspects had to be considered. In addition to typical concerns of project management (e.g., time allocation, schedule, and quality assurance) and software development (e.g., choices of programming language and data storage), the local situation at ESO also raised further questions. In particular, more than one department was interested in minting DOIs, and others could arise in the future. Therefore the first question was who would have ultimate responsibility for DOIs organization-wide; since the Library initiated use of DOIs, it adopted this role. But other questions that had to be answered were:

- Who is responsible for DOIs in each “client” department?
- How would the Library deal with diverse technologies—e.g., different databases—and needs—e.g., different landing page appearance—among departments?
- How would the Library translate each department’s metadata, describing different kinds of resources, into a common format?
- How could the Library make a system that future, unknown clients would be able to use?

Each of these questions was of key importance in creating a solution.

5 Implementation: architecture and process

In August 2016, the Library began development of a software solution for minting DOIs for ESO. In February 2017, a minimum viable product was achieved, and the following month the service launched with the first batch of DOIs (articles for vol. 167 of *The Messenger*).

The Library created an application, called the ESO DOI Service, which is responsible for (1) communicating with DataCite for create–update–delete operations, and (2) serving landing pages.

²<http://www.eso.org/public/about-eso/esoglance/> (accessed July 2017)

³<https://www.eso.org/sci/publications/messenger/>

The service’s clients are various departments which want to mint DOIs, such as the Data Archive or ePOD. A client sends a structured request for a new DOI, containing the desired DOI name and resource metadata, to the service via an HTTP API; the service creates its own record and attempts to register the DOI with DataCite; and finally the service gives the client a response indicating whether creation was successful or not. Editing and deaccessioning DOIs each follow a similar pattern. As soon as the DOI is registered with DataCite, the ESO DOI Service serves a landing page at <http://doi.eso.org/<DOI>> (where <DOI> is replaced by the DOI name), displaying the metadata that the service has for that resource. This architecture is depicted in Figure 2.

The ESO DOI Service was developed using Test-Driven Development (TDD) and using common application design patterns such as Model–View–Controller and Data Mapper.[4] The decision was also made early to decouple the application from those of other departments, and to communicate with clients only via the HTTP interface, to avoid any unnecessary dependency. As a result, the ESO DOI Service and a client application can be “black boxes” to one another: internal details can safely change as long as the agreed-upon HTTP interface is maintained. The ESO DOI Service does not need to “know” which database system (for example) a client department uses and vice versa.

The central data model of the ESO DOI Service is the DataCite Metadata Schema (DMS).[5] This decision was made to answer the question of how to deal with multiple types of resources. It also serves the goal of producing the XML required by DataCite for DOI minting, which is the key output of the system. Semantic mapping takes place outside of the scope of the application; each client determines, in consultation with the Library, how its data needs to be translated to fit the metadata scheme. A simple example is that *Messenger* authors and principal investigators of Archive datasets both become DMS creators. Generally these mappings are not difficult, but they are needed to mint DOIs with DataCite.

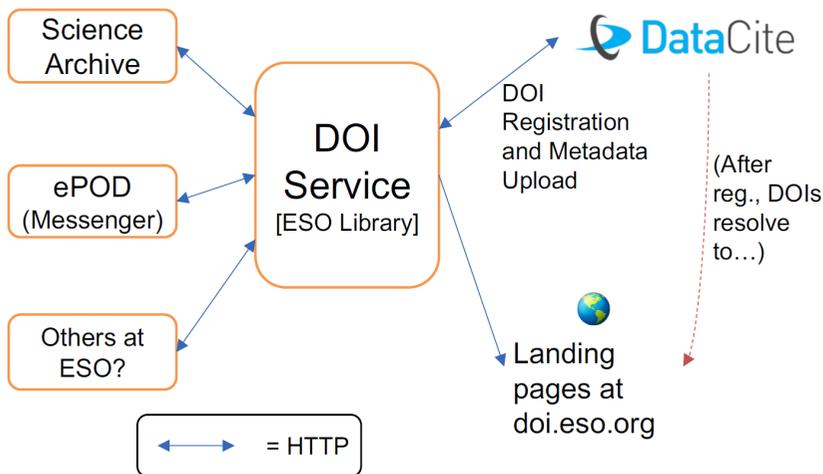


Figure 2. DOI Service architecture within ESO.

6 Features

In addition to the ESO DOI Service’s core functionality described above, a few more features have already been implemented.

6.1 Namespacing & extensibility

ESO DOIs begin their suffix with a namespace, such as /0722-6691/ in the DOI 10.18727/0722-6691/5010. Although this namespace is, by design, meaningless to the global DOI system, the ESO DOI Service uses it for next-consecutive-integer DOI naming. In the example above, the next DOI minted in the same namespace would end with 5011. This allows a client to request a DOI without having to specify the full name, instead specifying only the namespace.

The ESO DOI Service also uses the namespace to “know” more about what kind of resource is being described. The basic data model can be extended for the sake of richer landing page content, and the visual design of the landing page itself can change based on the namespace. This has already been implemented for *Messenger* articles, whose landing pages display volume/section/page information (which is outside of the core data model), and which display labels like “author” instead of “creator.” Such behavior is possible for as many namespaces as desired.

6.2 Machine-readable embedded metadata

FORCE11, a multidisciplinary organization centered upon advances in scholarly communication, released in 2014 a “Joint Declaration of Data Citation Principles,” whose purpose is to help provide “a foundation of robust, accessible data” for “sound, reproducible scholarship” through the expression of eight principles.[6] Although a full discussion of the Declaration is outside the scope of this article, the principle of “Access” through metadata is particularly relevant to the ESO DOI Service. In turn, Fenner et al. have produced an extremely helpful “Road Map” to fulfilling these principles.[7]

One recommendation by Fenner et al. is the use of <meta> tags in landing pages, with Dublin Core attributes. The purpose of these tags is for landing pages to interoperate with reference manager applications, since this is how they extract metadata. In a similar vein, ESO DOI Service landing pages also contain JSON-LD using Schema.org, which improves discoverability (for example by search engines).

6.3 “Cite this article”

ESO DOI Service landing pages also have a simple “Cite this article” (or “Cite this dataset”) feature. A user can copy-paste the pregenerated reference, so citing the resource is quite easy—and guaranteed to include the DOI, if the user does it this way.

7 Outlook and conclusion

Returning to the questions posed in Section 4, the chosen approach seems to have addressed these challenges. The Library deals with diverse technologies by offering its own service, accessible via a common, language-agnostic interface (HTTP), but whose internals are isolated from the applications of other departments. The adoption of the DataCite Metadata Schema (DMS) as the core data model provides a common target for the metadata of various departments and resource types, and also allows for easy translation to the XML required for registering the DOI with DataCite. Finally, the use of namespaces and a data model extensible beyond the DMS enables the application to serve diverse wishes.

However, this local solution—like the DOI system as a whole—is not merely technological, but must also rely upon critical social components in order to succeed. The Library must continue to communicate with client departments to begin minting DOIs for Archive datasets, and the ESO DOI

Service's metadata schema and API must be well conveyed for clients to use them effectively. Any technical problems that arise will also require interdepartmental coordination.

Although the ESO DOI Service is an operational application, there are several "nice-to-have" features which have yet to be implemented. These include: search of ESO DOIs; automated handling of inverse relations between resources; downloadable bibliographic formats for landing-page users; regular checks for resource persistence; and perhaps others. The Library hopes to accomplish these in the short-to-medium term.

DOIs are an important component in contemporary research and scholarly communications, but they require a global infrastructure, and each participating organization must somehow integrate its data, using its own infrastructure. For ESO, the ESO DOI Service provides that integration. In crafting such a solution, the Library has found another way to continue its mission of supporting researchers and data providers for the benefit of the astronomy community as a whole.

References

- [1] International Standards Organization, *ISO 26324:2012 Information and documentation — Digital object identifier system* (ISO, Geneva, 2012)
- [2] International DOI Foundation, *DOI® Handbook*, <https://doi.org/10.1000/182> (accessed July 2017)
- [3] M. Klein, H. Van de Sompel, R. Sanderson et al., *PLoS ONE* **9(12)**, e115253 (2014), <https://doi.org/10.1371/journal.pone.0115253>
- [4] M. Fowler, D. Rice, M. Foemmel et al., *Patterns of Enterprise Application Architecture*, Addison-Wesley, Boston, (2002), ISBN 0-321-12742-0
- [5] DataCite Metadata Working Group, *DataCite Metadata Schema Documentation for the Publication and Citation of Research Data* Version 4.0 (DataCite e.V., 2016) <https://doi.org/10.5438/0012>
- [6] Data Citation Synthesis Group, "Joint Declaration of Data Citation Principles" (San Diego, FORCE11, 2014) <https://www.force11.org/group/joint-declaration-data-citation-principles-final> (accessed July 2017)
- [7] M. Fenner, M. Crosa, J. Grethe et al., *A Data Citation Roadmap for Scholarly Data Repositories*, bioRxiv **97196**, (2016) <https://doi.org/10.1101/097196>
- [8] N. Paskin, "Digital Object Identifier (DOI®) System", *Encyclopedia of Library and Information Sciences, Third Edition* (CRC Press, 2009) 1586-1592