e-Begi Quick Guide to DOIs

Authors: Lohitzune Solabarrieta, Ivan Manso, Andrea del Campo, Anna Rubio (AZTI)

2022-08-11

Contents

About

The present e-book is a protocol generated under e-BEGI project to guide researchers and technicians from AZTI on how to generate, process, publish and disseminate their research outputs following the FAIR principles and how to assign DOIs to convert them in digital persistent research objects.

Here to edit the book?

The code is available in AZTI's github repository repository and the book is readily available here.

Chapter 1

Introduction

The present e-book is a protocol generated under e-BEGI project to guide researchers and technicians from AZTI on how to generate, process, publish and disseminate their research outputs following the FAIR principles and how to assign DOIs to convert them in digital persistent research objects.

"Good data management is not a goal in itself, but rather is the key conduit leading to knowledge discovery and innovation, and to subsequent data and knowledge integration and reuse by the community after the data publication process." (Wilkinson et al., 2016)

Chapter 2

What is a DOI

DOI is an acronym for 'digital object identifier' (https://www.doi.org/). The DOI® system was initiated by the International DOI Foundation (IDF) in 1998 (later standardised as ISO 26324) and provides an infrastructure for persistent unique identification of objects of any type, such as data. It also provides an interoperable exchange of managed information on digital networks (that is to use, or work with, existing identifier and metadata schemes). A DOI name is assigned to data (or any other object) primarily for sharing with an interested user community or managing as intellectual property. DOI names may also be expressed as URLs (URIs).

Chapter 3

Why and how to assign a doi

WHY?

It is recommended to provide a DOI name expressed as URL to your dataset or publication and add it to their metadata in order to make them more findable and to track the original authors and science behind. This will provide proper credits to the authors and a confident tracking record.

The DOI system offers a unique set of functionalities: • Persistence, if material is moved, rearranged, or bookmarked. • Interoperability with other data from other sources. • Extensibility by adding new features and services through management of groups of DOI names. • Single management of data for multiple output formats (platform independence). • Class management of applications and services. • Dynamic updating of metadata, applications, and services.

HOW?

The DOI system is implemented through a federation of Registration Agencies (RAs) that use policies and tools developed through the IDF, which safeguards (owns or licenses on behalf of registrants) all intellectual property rights relating to the DOI system. Many millions of DOI names have been assigned to date, through a growing federation of RA world-wide.

DataCite is one of these RAs that provides DOIs specifically for research data. Although other RAs also assign DOI-s to datasets, DataCite is the most used for the DOI generation for datasets. DataCite's global community includes data centres, libraries, government agencies, research universities, etc. Membership is open to all organizations that share the data sharing mission of DataCite. The organization should have its data in repositories to become a membership and consume the DOI service at DataCite. Other way to provide a DOI to your

datasets, without becoming a member of a RA, is publishing them in established generic or community specific repositories like Zenodo, Pangaea, SEANOE or others. As mentioned before, at https://www.re3data.org/ you can find a useful and comprehensive registry to find and assess the most suitable repositories for your specific needs and communities. Examples of repositories where the user can assign a DOI without being member of any RA is shown in Table 3.

Table 3. Examples of data repositories that generate DOI.

	PANGAEA	SEANONE
URL	https://www.pangaea. de/submit/	https://www.seanoe.org /html/publish-your- data.htm
Doi assignation	Yes	Yes
Time of doi assignation	Weeks to months	24 hours
Storage limit	100 GB/dataset	10 GB/dataset & 10 days/month
Cost	Free	Free
Recommended by	Nature (https: //www.nature.com/sda ta/policies/repositories)	SeaDataNet
(https://www.seanoe.o rg/html/sdata.htm?ut m_source=hootsuite)		

AZTI recommends assigning a DOI for:

- **Campaigns:** through DataCite following the recommendation from section 5: "Assign a doi to your oceanographic campaign"
- **Datasets:** through public repositories following the recommendation from section 6: "Assign a doi to your dataset"

Chapter 4

FAIR principles

There is a need to extract the maximum benefit from our research investments/outputs and a good data management is essential to achieve this need (e.g., Roche et al., 2015). However, what constitutes 'good data management' has been largely undefined and it was generally left as a decision of the data owner. Therefore, bringing some clarity around the goals of good data management and defining simple guidelines to inform those who publish and/or preserve scientific data, is of great utility. For this purpose, Wilkinson et al (2016) described four foundational principles: Findability, Accessibility, Interoperability, and Reusability (also known as FAIR principles); that serve to guide data producers and publishers, helping to maximize the added value gained by contemporary digital publishing and, also, adhere to the expectations and requirements of the funding agencies. The FAIR principles apply not only to 'data' but also to the algorithms, processing tools, and workflows that led to those data. All digital research objects benefit from the application of these principles, since all components of the research process should be clearly available in the datasets' metadata to ensure transparency, reproducibility, and reusability (Wilkinson et al, 2016).

Findable: Each dataset should be identified by a unique persistent identifier and described by rich, standardized metadata that clearly include the persistent identifier. The metadata record should be indexed in a catalogue and carried with the data.

Accessible: The dataset and its metadata record should be retrievable by using the persistent identifier and a standardized communications protocol. In turn, that protocol should allow for authentication and authorization, where necessary. All metadata records should remain accessible even when the datasets they describe are not easily accessible. It should not be confused with Open Data, since FAIR's Accesible principle grants the data owner the degree to which data is available, or advertised (metadata) (Mons et al., 2017).

Interoperable: Both metadata and datasets use formal, accessible, shared, and broadly applicable vocabularies and/or ontologies to describe themselves. They should also use vocabularies that follow FAIR principles and provide qualified references to other relevant metadata and data. Importantly, the data and metadata should be machine accessible and parsable.

Reusable: To meet this principle, data must already be findable, accessible, and interoperable. Additionally, the data and metadata should be sufficiently richly described that it can be readily integrated with other data sources. Published data objects should contain enough information on their provenance to enable them to be properly cited and should meet domain-relevant community standards.

Chapter 5

Assign a Doi to an oceanographic campaign

Would you like to assign a DOI to your oceanographic campaign? You are in the right place. Here you have some recommendations to assign a DOI to you AZTI campaign:

1. Prepare a summary of your campaign details, including at least, but open to other and more detailed information:

a. Summary of the campaign
b. Geographical area
c. Date and frequency
d. Contact
e. Any other relevan information related to your campaign

2. Contact FULANITO Y MENGANITO to create a specific web section inside www.azti.es/Campaigns/XXX assigned to your campaign

3. Contact Edorta Aranguena (earanguena@azti.es), Lohitzune Solabarrieta (lsolabarrieta@azti.es) and/or Andrea del Campo (adelcampo@azti.es) to assign a DOI to the website created in the previous step

The DOI will be assigned through DataCite, which is one of the settled Registration Agencies (RI) authorized to assign a DOI and where AZTI is a member.

4. Contact back FULANITO Y MENGANITO to add the DOI generated in the previous step 3 to the website details.

5. Decide if you want to publish your dataset, following the steps in the ebook section "6.- Assign a doi to your dataset". Include that doi (if generated) or the way to obtain the dataset (if exist) in the www.azti.es/Campaigns/XXX site of your campaign

Chapter 6

Assign a doi to a dataset

Would you like to assign a DOI to your dataset? You are in the right place. AZTI recommends you publish your our datasets in public data repositories that will automatically assign a doi to your datasets. Here you have some recommendations to assign a DOI to your AZTI dataset:

1. Prepare a summary of your dataset details, including at least, but open to other and more detailed information:

   a. Summary of the dataset, including standardized variable names
   b. Geographical area
   c. Date and frequency
   d. Quality Control (QC) applied to the dataset
   e. Contact
   f. Any other relevant information related to your dataset

2. Find the most suitable repository for your specific need and communities https://www.re3data.org/

3. Connect to the repository website and follow the instructions to upload your data

4. Contact Lohitzune Solabarrieta (lsolabarrieta@azti.es) and/or Andrea del Campo (adelcampo@azti.es) if you have any questions.

Chapter 7

Assign a doi to other objects

Under construction. Will come soon

SORRY FOR THE INCONVENIENCES

References

Roche D. G., Kruuk L. E. B., Lanfear R. & Binning S. A. Public Data Archiving in Ecology and Evolution: How Well Are We Doing? PLOS Biol. 13, e1002295 (2015).

Tanhua T, Pouliquen S, Hausman J, O'Brien K, Bricher P, de Bruin T, Buck JJH, Burger EF, Carval T, Casey KS, Diggs S, Giorgetti A, Glaves H, Harscoat V, Kinkade D, Muelbert JH, Novellino A, Pfeil B, Pulsifer PL, Van de Putte A, Robinson E, Schaap D, Smirnov A, Smith N, Snowden D, Spears T, Stall S, Tacoma M, Thijsse P, Tronstad S, Vandenberghe T, Wengren M, Wyborn L and Zhao Z (2019) Ocean FAIR Data Services. Front. Mar. Sci. 6:440. 'DOI: 10.3389/fmars.2019.00440.

Wilkinson, M. D., Dumontier, M., Aalbersberg, I. J., Appleton, G., Axton, M., Baak, A., … & Mons, B. (2016). The FAIR Guiding Principles for scientific data management and stewardship. Scientific data, 3(1), 1-9.

Contacts

This GitHub ebook is generated by AZTI team, for more information contact:

Lohitzune Solabarrieta (lsolabarrieta@azti.es), Ivan Manso (imanso@azti.es),
Andrea del Campo (adelcampo@azti.es), Anna Rubio (arubio@azti.es)