



DOI de données de la recherche

QUESTIONS FREQUENTES

Auteurs : Fred Merceur
Version : 1.0
Date : 20/02/2017
DOI : <https://doi.org/10.13155/60247>

Table des matières

1. Historique des versions.....	2
2. Introduction	2
3. Un DOI, c'est quoi ?	2
4. Pourquoi attribuer des DOIs aux données de la recherche ?	3
5. Qui peut attribuer des DOI ?	3
6. Quel moyen d'accès aux données offrir depuis une Landing Page ?.....	4
7. L'organisme qui attribue les DOI doit-il être l'organisme qui assure la conservation des données ?	4
8. Comment citer des données dans une publication?.....	4
9. Ou citer un DOI de données dans une publication ?	5
10. Comment définir la liste d'auteurs d'un jeu de données ?.....	7
11. Quelle granularité choisir ?	7
12. Est-il possible de modifier les métadonnées d'un DOI ?.....	8
13. Comment gérer différentes versions d'un même jeu de données ?	8
14. Un fragment (#) de DOI, c'est quoi ?.....	9
15. Peut-on supprimer un DOI ?.....	10
16. Ou trouver les DOI attribués par DataCite ?.....	10
17. Références	10

1. Historique des versions

Version	Date	Etat	Auteurs(s)	Description
1.0	2017/02/20	V1	FM	Première version

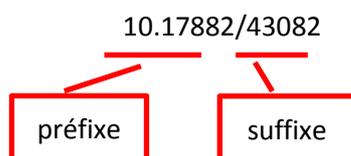
2. Introduction

Ce document propose des éléments de réponse à un ensemble de questions fréquemment posées par les scientifiques sur les DOI de données de la recherche.

3. Un DOI, c'est quoi ?

Un DOI (Digital Object Identifier) est un système de numéro d'identification unique. S'il est principalement connu pour identifier un article scientifique, un DOI peut être attribué à n'importe quelle ressource physique, numérique ou abstraite.

Un DOI se présente sous la forme d'une chaîne de caractère composée d'un préfixe propre à l'éditeur et d'un suffixe :



Un DOI peut être mentionné précédé de « DOI: » (ex : DOI:10.17882/43082). Il peut également être associé à l'URL du résolveur de DOI « <https://dx.doi.org/> » ou « <https://doi.org/> » (Exemple : <https://doi.org/10.17882/43082>). L'URL ainsi composée devient l'URL officielle de la ressource sur le WEB.

Un clic sur un DOI lance un appel vers la fondation DOI qui redirige l'internaute de manière transparente vers une page WEB : la Landing Page (LP) ou « page d'atterrissage ». Cette LP décrit la ressource associée au DOI à l'aide d'un ensemble de métadonnées et propose un moyen d'y accéder.

4. Pourquoi attribuer des DOIs aux données de la recherche ?

Un DOI attribué à un jeu de données peut répondre à plusieurs objectifs :

Fiabiliser la citation de données : Si la donnée est déplacée, il est possible de mettre à jour l'URL de la Landing Page associée au DOI. C'est en ce sens que l'utilisation des DOI dans une bibliographie permet de fiabiliser une citation. Dans une bibliographie, il faut donc citer le DOI et non pas l'URL de la Landing Page du DOI (Figure 1).

Simplifier l'accès aux données : La Landing Page d'un DOI propose un moyen d'accéder à la donnée, idéalement via une adresse HTTP/FTP directe. L'accès aux données exploitées dans une publication permet ainsi aux lecteurs de poursuivre les recherches. Les lecteurs peuvent également vérifier un résultat : la citation et l'accessibilité des données peuvent augmenter la crédibilité de l'étude.

Permettre la traçabilité de l'utilisation des données : Les publications qui citent un jeu de données à l'aide d'un DOI sont simples à repérer. Il devient ainsi possible de rendre compte ou de valoriser l'utilisation d'un jeu de données ou d'une infrastructure par l'intermédiaire du nombre et du profil des publications qui l'exploitent.

Figure 1 : A faire et à ne pas faire : dans une bibliographie, il faut citer le DOI des articles ou des jeux de données et non pas l'URL de la Landing Page de ces DOI

References

1. Ben Rais Lasram F, Guilhaumon F, Mouillot D. Fish diversity patterns in the Mediterranean Sea: deviations from a mid-domain model. *Mar Ecol Prog Ser.* 2009; 376: 253–267. doi: [10.3354/meps07786](https://doi.org/10.3354/meps07786) 
2. Rahbek C, Graves G. Multiscale assessment of patterns of avian species richness. *Proc Natl Acad Sci.* 2001; 98: 4534–4539. Available: <http://www.pnas.org/content/98/8/4534.short> PMID: [11296292](https://pubmed.ncbi.nlm.nih.gov/11296292/)
3. Macpherson E, Duarte C. Patterns in species richness, size, and latitudinal range of East Atlantic fishes. *Ecography (Cop).* 1994; 17: 242–248. Available: <http://onlinelibrary.wiley.com/doi/10.1111/j.1600-0587.1994.tb00099.x/abstract> 
4. Rex MA, Stuart CT, Hessler RR, Allen JA, Sanders HL, Wilson GFDG. Global-scale latitudinal patterns of species diversity in the deep-sea benthos. *Nature.* 1993; 636–639.
5. Macpherson E. Large-scale species-richness patterns in the Atlantic Ocean. *Proc R Soc.* 2002; 269: 1715–20. doi: [10.1098/rspb.2002.2091](https://doi.org/10.1098/rspb.2002.2091) 
6. Moutin T, Raimbault P. Primary production, carbon export and nutrients availability in western and eastern Mediterranean Sea in early summer 1996 (MINOS cruise). *J Mar Syst.* 2002; 33–34: 273–288. Available: <http://www.sciencedirect.com/science/article/pii/S0924796302000623> 

5. Qui peut attribuer des DOI ?

De nombreux éditeurs ou organismes offrent la possibilité d'attribuer des DOI. Ces éditeurs/organismes ont souscrit un contrat avec une agence de DOI. Il existe plusieurs agences dont certaines se sont spécialisées. Les éditeurs d'articles scientifiques font par exemple majoritairement appel à l'agence CrossRef. DataCite est une agence spécialisée dans l'attribution de DOI aux données de la recherche. Pour attribuer des DOI via DataCite, il faut souscrire un contrat auprès de son représentant national. En France, le représentant de DataCite est l'INIST/CNRS.

Les DOI attribués par ces différentes agences sont tous interopérables et peuvent tous être résolus à l'aide du même résolveur <https://doi.org/>.

6. Quel moyen d'accès aux données offrir depuis une Landing Page ?

Idéalement, l'accès aux données doit être proposé sous la forme d'une URL FTP/HTTP directe depuis la Landing Page du DOI. Mais l'accès à la ressource depuis une LP peut également se limiter à une simple adresse email de contact vers le producteur de la donnée, ou à un formulaire WEB de contact.

Conformément aux recommandations européennes, l'accès aux données financées par des fonds publics doit idéalement être libre et gratuit mais ce n'est pas une obligation contractuelle pour attribuer un DOI à des données.

A noter que plusieurs entrepôts de données proposent un système d'embargo sur l'accès aux données : il est possible d'attribuer un DOI à des données dont l'accès est masqué pendant une période sélectionnée, par exemple pendant la période de relecture par les pairs d'une publication scientifique.

7. L'organisme qui attribue les DOI doit-il être l'organisme qui assure la conservation des données ?

Contractuellement, un organisme peut attribuer un DOI à des données hébergées par un autre organisme. Mais cette organisation augmente le risque que la Landing Page présente des informations d'accès obsolètes au bout de plusieurs années si les relations entre ces deux organismes finissent par se distendre.

8. Comment citer des données dans une publication ?

DataCite suggère de citer un jeu de données dans un format proche des formats de citation courants des publications (DataCite Metadata Working Group, 2016) :

Creator (PublicationYear): Title. Publisher. DOI

Exemple :

Delcroix Thierry; Alory Gaël; Téchiné Philippe; Diverrès Denis; Varillon David; Cravatte Sophie; Gouriou Yves; Grelet Jacques; Jacquin Stéphane; Kestenare Elodie; Maes Christophe; Morrow Rosemary; Perrier Julien; Reverdin Gilles; Roubaud Fabrice; (2002): Sea Surface Salinity data from Voluntary Observing Ships Network; The French Sea Surface Salinity Observation Service (SSS OS). <https://doi.org/10.6096/SSS-LEGOS>

9. Ou citer un DOI de données dans une publication ?

Certaines revues commencent à imposer des consignes spécifiques. La revue Science demande que les données soient citées dans la section *Acknowledgements*. La revue PLOS ONE a mis en place une section *Data Availability* dans l'entête de ses articles (Figure 2).

Figure 2 : Section Data Availability dans un article publié dans la revue PLOS ONE

<p>Data Availability: The raw data underlying this analysis (date and time of each detection at each site, plus signal-to-noise ratio) are all accessible through http://doi.org/10.17882/46124.</p>
<p>Funding: Deployment and recovery cruises were funded by the French Polar Institute (IPEV; MD174/175, MD185, MD189, MD193, MD197, MD200, MD201 cruises) and by NOAA/PMEL and INCOIS (SK 307). Additional support for the acoustic experiments was provided by the Institut National des Sciences de l'Univers of CNRS. E. C.L. was supported by a Ph.D. fellowship from the University of Brest and from the Regional Council of Brittany (Conseil Régional de Bretagne). Moorings and hydrophones were funded by the Regional Council of Brittany (CPER). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.</p>
<p>Competing interests: The authors have declared that no competing interests exist.</p>

Introduction

As the preferred target of commercial whalers, the Antarctic blue whales (*Balaenoptera*

Dans les revues qui ne donnent pas de consignes précises, les auteurs citent majoritairement les données dans la section *Acknowledgements* ou dans la section *References* (Figure 3).

Figure 3 : Citation d'un jeu de données dans la section References d'une publication

The pacific oyster *C. gigas* reproduction has been well studied on the Atlantic coast, fitting hatchery, oyster culture and ecology needs. Since 2008, a national monitoring network that aims to observe the reproductive status of the Pacific oyster over a wide geographic range (see <http://www.ifremer.fr/velyger>, **Pouvreau et al., 2016**) helps to understand the spatio-temporal variability in the reproductive performance of this

References

Pouvreau et al., 2000b Stephane Pouvreau, Aline Gangnery, Jerome Tiapari, Franck Lagarde, Matthieu Gamier, Alain Bodoy
Gametogenic cycle and reproductive effort of the tropical blacklip pearl oyster, *Pinctada margaritifera* (Bivalvia: Pteriidae) cultivated in Takapoto atoll (French Polynesia)
 Aquat. Living Resour., 13 (1) (2000), pp. 37–48 Publisher's official version: [http://doi.org/10.1016/S0990-7440\(00\)00135-2](http://doi.org/10.1016/S0990-7440(00)00135-2) Open Access version: <http://archimer.ifremer.fr/doc/00000/1391/>
 Article |  PDF (406 K) | View Record in Scopus | Citing articles (67)

Pouvreau et al., 2016 Stephane Pouvreau, Daniele Maurer, Isabelle Auby, Franck Lagarde, Patrik Le Gall, H  l  ne Cochet, Anne-Lise Bouquet, Am  lie Geay, Dominique Mille
VELYGER Database: The Oyster Larvae Monitoring French Project
 SEANOE (2016) <http://doi.org/10.17882/41888>

Rayssac et al., 2012 N. Rayssac, A. P  rignon, E. Gervasoni, F. Pernet, Gall P. Le, F. Lagarde
 Projet PRONAMFED 2010–2011   valuation du potentiel d'approvisionnement naturel en

10. Comment d  finir la liste d'auteurs d'un jeu de donn  es ?

Les diff  rents syst  mes qui permettent l'attribution de DOI n'offrent pas tous la m  me liste de m  tadonn  es pour d  crire un jeu de donn  es. Parmi la liste de m  tadonn  es communes    l'ensemble des syst  mes, il y a par contre les 5 m  tadonn  es obligatoires impos  es par DataCite pour attribuer un DOI : Le Titre, l'ann  e, l'  diteur, la liste des auteurs et le DOI lui-m  me. Ces 5 m  tadonn  es serviront    composer la citation du jeu de donn  es.

La m  tadonn  e Auteur peut   tre compos  e d'une liste de personnes identifi  s par leurs noms et pr  noms. Il est   galement possible d'attribuer un jeu de donn  es    une ou plusieurs collectivit  s d'auteurs (ex : un nom de projet, un nom d'organisme, ...).

En option, DataCite accepte une liste de contributeurs. Cette m  tadonn  e peut   tre notamment utilis  e pour lister les personnes ayant contribu  s aux jeux de donn  es en compl  ment d'une collectivit   d'auteurs.

Le choix des auteurs peut aussi d  pendre de la strat  gie de gestion de versions (voir par paragraphe 12).

11. Quelle granularit   choisir ?

Il n'y a pas de r  ponse automatique    cette question, c'est un choix qui d  pend des donn  es et de l'usage de ces donn  es par les scientifiques.

Les DOI ont pour principal objectif de faciliter et de fiabiliser la citation des données dans les publications scientifiques : il est donc possible de définir une stratégie sur la base du niveau d'agrégation des données qui sont généralement exploitées dans une publication.

Par exemple, si les DOI sont positionnés au niveau d'une mesure et les publications scientifiques exploitent généralement des milliers de ces mesures, il ne sera pas possible de citer ces milliers de DOI de données : dans ce cas, les DOI ne sont pas bien positionnés.

Si des publications scientifiques exploitent différents niveaux d'une même masse de données, il est possible d'attribuer des DOI spécifiques à ces différents niveaux. Dans la Landing Page de ces DOI, il est alors possible d'offrir des liens pour permettre au lecteur de naviguer d'un DOI à l'autre.

12. Est-il possible de modifier les métadonnées d'un DOI ?

Oui, il est possible de modifier à tout moment l'ensemble des métadonnées d'un DOI à l'exception du DOI lui-même. Il est donc possible, par exemple, d'ajouter un auteur, de modifier l'ordre des auteurs ou de faire une correction dans le titre. Ce n'est pas forcément idéal et notamment si le jeu de données a déjà été cité, mais ça reste possible.

13. Comment gérer différentes versions d'un même jeu de données ?

Comme pour la granularité, il n'y a pas d'obligation par rapport à la gestion des versions d'un jeu de données évolutif. Idéalement, le DOI doit permettre l'accès aux anciennes versions pour permettre la reproductibilité des expériences passées mais ce n'est pas une obligation contractuelle. Pour les données volumineuses, le coût de stockage peut, par exemple, être incompatible avec la conservation de l'ensemble des versions.

Si différentes versions d'un jeu de données doivent continuer à être accessibles de manière indépendante et citées de manière spécifique (par exemple avec une liste d'auteurs différents), il faut alors attribuer un DOI spécifique à chaque version.

Il est également possible de proposer un accès à différentes versions d'un même jeu de données au sein de la Landing Page d'un même DOI (Figure 4). Ce choix peut être complété par une gestion de fragments (voir paragraphe 13).

Figure 4 : Accès aux différentes versions du jeu de données Dyfamed (<http://doi.org/10.17882/43749>)

File	Size	Format	Processing	Access	Key
1988-2015 deployments	20 MB	NetCDF, Excel	Quality controlled data	Open access	43298
2010-2014 deployments	14 MB	NC, NetCDF	Quality controlled data	Restricted access	43276
2010-2013 deployments	2 MB	XLS, XLSX	Quality controlled data	Restricted access	43283

Attribuer un DOI unique à un jeu de données évolutif présente de nombreux avantages :

- Les consignes de citation sont plus simples à donner aux utilisateurs du jeu de données. Ils n'ont qu'un seul DOI à appréhender.
- Pour améliorer la visibilité du DOI dans les moteurs de recherches (ex : Google), il est préférable de diffuser une seule Landing Page qui concentre un maximum de citations et donc de liens (*backlink*) plutôt que de multiples Landing Page de contenu quasiment identiques.
- Si un DOI spécifique est attribuée à chaque version, les moteurs de recherches ne présenteront par forcément la version la plus récentes en avant dans leurs listes de résultats. Un internaute peut donc découvrir une version obsolète du jeu de données et ne pas se rendre compte qu'il existe une version plus récente.

Un DOI unique attribué à un jeu de données évolutif peut, par contre, avoir une incidence sur la définition des auteurs (voir paragraphe 9). Si un jeu évolue sur plusieurs années, de nouveaux auteurs peuvent rejoindre le projet à l'origine des données ou, au contraire, quitter ce projet. Si des auteurs sont ajoutés au DOI, ils seront associés au jeux de données dans son ensemble. Le DOI leur attribuera donc la paternité d'une partie des données auxquelles ils n'ont pas contribué. Ce problème, s'il est considéré comme bloquant par certains auteurs, peut être contourné en proposant une collectivité d'auteurs (ex : le nom du projet) comme auteur de la donnée.

14. Un fragment (#) de DOI, c'est quoi ?

Certains éditeurs permettent la citation d'un sous-ensemble d'un DOI (ex : une version, une zone géographique, ...) à l'aide d'un fragment (#) de DOI.

C'est par exemple le cas des données des flotteurs Argo. Pour permettre la reproductibilité des études qui exploitent les données Argo, un *snapshot* de l'ensemble du jeu de données est figé tous les mois. Ce *snapshot* contient l'ensemble des données Argo disponibles au moment de la création du *snapshot*.

La citation du jeu de données global s'effectue en citant le DOI Argo sans paramètre :

Argo (2000). Argo float data and metadata from Global Data Assembly Centre (Argo GDAC). SEANOE. <http://doi.org/10.17882/42182>

La Landing Page du DOI propose alors l'ensemble des versions disponibles. La citation d'un *snapshot* spécifique se fait en ajoutant le fragment attribué au *snapshot* précédé du caractère # à la fin du DOI :

Argo (2016). Argo float data and metadata from Global Data Assembly Centre (Argo GDAC) - Snapshot of Argo GDAC of February, 8th 2016. SEANOE. <http://doi.org/10.17882/42182#42350>

La Landing Page du DOI ainsi appelée présente alors des informations relatives au snapshot.

15. Peut-on supprimer un DOI ?

Il n'est pas possible de supprimer un DOI. Contractuellement, un organisme qui attribue un DOI s'engage à maintenir en ligne la Landing Page du DOI. Si, un jeu de données devient inaccessible, la Landing Page doit alors présenter aux internautes les raisons de la suppression de l'accès à la donnée.

Certains articles scientifiques continuent d'être cités plusieurs dizaines d'années après leur publication. L'attribution de DOI aux données doit permettre une citation fiable et pérenne sur les mêmes échelles de temps.

16. Ou trouver les DOI attribués par DataCite ?

Les métadonnées des DOI de données sont accessibles par l'intermédiaire du site <https://search.datacite.org/>.

17. Références

DataCite Metadata Working Group. (2016). *DataCite Metadata Schema Documentation for the Publication and Citation of Research Data*. doi:10.5438/0012

Groupe interpôle. (2016). *Bonnes pratiques autour de la citation*.

Merceur, F. (2016). *DOIs for ocean data, general principles and selected examples (Argo, French cruises)*. doi:10.13155/44515