

Guide PID d'ODATIS

Guide ANR COPILOtE sur les identifiants pérennes (PID) et leurs recommandations dans le contexte du pôle ODATIS

Informations du document

Titre court	
Guide PID d'ODATIS	
Titre long	
Guide ANR COPiLOtE sur les identifiants pérennes (PID) et leurs recommandations dans le contexte du pôle ODATIS	
Auteur	
Cécile NYS (OceanScope/IFREMER) ¹ , Frédéric MERCEUR (IFREMER) ² , Erwann QUIMBERT (IFREMER) ³	
Dissémination	Copyright
Publique	© ODATIS, 2022
DOI	
10.13155/87217	
Citation	
Nys, C., Merceur, F., Quimbert, E. (2022). Guide ANR COPiLOtE sur les identifiants pérennes (PID) et leurs recommandations dans le contexte du pôle ODATIS. ODATIS. https://doi.org/10.13155/87217	

¹ Cécile NYS – ORCID : <https://orcid.org/0000-0002-6938-5004>

² Frédéric MERCEUR – ORCID : <https://orcid.org/0000-0003-2154-7622>

³ Erwann QUIMBERT – ORCID : <https://orcid.org/0000-0003-3452-1333>

Historique

Version	Auteurs	Date	Commentaires
0.0	Cécile NYS (OceanScope/IFREMER), Frédéric MERCEUR (IFREMER)	22 juin 2021	Reprise du chapitre « PID guidelines » du guide Copilote en version draft (version 0.9)
0.1	Cécile NYS (OceanScope/IFREMER)	22 juin 2021	Relecture et corrections
0.2	Cécile NYS (OceanScope/IFREMER)	24 juin 2021	Ajout première liste de références utilisées
0.3	Cécile NYS (OceanScope/IFREMER), Frédéric MERCEUR (IFREMER)	2 juillet 2021	Relecture du document et commentaires de Fred Merceur (réunion du 2 juillet)
0.4	Cécile NYS (OceanScope/Ifremer)	6 octobre 2021	Révision et corrections suite à échanges avec Fred Merceur
0.5	Cécile NYS	4 novembre 2021 ; 4, 5 et 6 janvier 2022.	Relecture, ajouts et corrections ; Ajout d'une section « Choix de l'entrepôt » ; Ajout des renvois de références et la liste complète des références.
0.6	Erwann QUIMBERT (IFREMER)	7 janvier 2022	Relecture, ajouts et corrections
0.7	Cécile NYS	11 janvier 2022	Révisions
0.8	Sabine SCHMIDT	13 janvier 2022	Corrections et révisions
0.9	Cécile NYS	14 janvier 2022 ; 20 janvier 2022	Reprise des corrections et révisions
0.10	Sabine SCHMIDT, Erwann QUIMBERT	20 janvier 2022 ; 21 janvier 2022	Corrections et révisions
0.11	Cécile NYS	24 mars 2022	Relecture finale et mise en page
1.0	Cécile NYS, Frédéric MERCEUR & Erwann QUIMBERT	24 mars 2022	Version diffusable en ligne
1.1	Cécile NYS, Frédéric MERCEUR & Erwann QUIMBERT	24 mars 2022	Ajout du DOI du guide

Table des matières

TABLE DES MATIERES	4
TABLES DES ILLUSTRATIONS	4
ABREVIATIONS	5
1. INTRODUCTION	6
2. PRESENTATION GENERALE DES PID.....	6
3. SYNTAXE ET STRUCTURE DES PID.....	8
4. COMMENT CHOISIR UN PID ?	10
4.1. ORCID.....	10
4.2. DOI.....	11
5. COMMENT CHOISIR LA GRANULARITE D’UN DOI POUR UN JEU DE DONNEES ?.....	12
6. CHOIX DE L’ENTREPOT	13
7. PID GRAPHS	14
8. CONCLUSION	15
9. REFERENCES.....	16

Tables des illustrations

FIGURE 1. FORMAT ET RECOMMANDATIONS POUR LE DOI DU PÔLE ODATIS.....	9
FIGURE 2. REPRESENTATION SCHEMATIQUE DU PID GRAPH AVEC DES OBJETS NUMERIQUES CONNECTES PAR DIFFERENTS TYPES DE PID.....	14

Abréviations

Abréviations	Signification
CDS	Centre(s) de Données et Services
CTS	CoreTrustSeal (certification)
DOI	Digital Object Identifier (Identifiant unique d'un objet digital)
FAIR	Findable (Trouvable) Accessible (Accessible) Interoperable (Interopérable/Interopérabilité) Reusable (Réutilisable)
GRID	Global Research Identifier Database
IGSN	International Geo Sample Number
ISBN	International Standard Book Number
ISNI	International Standard Name Identifier
ISSN	International Standard Serial Number
LEI	Legal Entity Identifier
NBN	National Bibliography Number
ORCID	Open Researcher and Contributor ID
PID	Persistent IDentifier (Identifiant permanent)
PMC ID	PubMed Central Identification
PSI	Public Sector Information
RDA	Research Data Alliance
ROR	Research Organization Registry
SWH-ID	SoftWare Heritage persistent IDentifiers
URI	Uniform Resource Identifier (Identifiant unique et uniforme d'une ressource en ligne)
URL	Uniform Resource Locator
URN	Uniform Resource Name (a type of URI)
(URN-)NBN	(Uniform Resource Name-) National Bibliography Number
UUID	Universally Unique IDentifier (Identifiant unique universellement – ex : DOI, ISBN, URI, etc.)
WoS ResearchID	Web of Science ResearchID

1. Introduction

Il est actuellement fréquent d’entendre parler et d’assigner un DOI à un jeu de données. Mais qu’est-ce qu’un DOI (Digital Object Identifier) et quel est son intérêt ? Quelle forme prend-t-il ? Nous essayerons de répondre à ces questions dans ce guide de recommandations des PID.

Dans le cadre du projet Copilote, nous parlerons plutôt des « *Persistent IDentifiers* (PID) » (Identifiants pérennes *en français*) dont le DOI est un exemple. Les PID sont déclinés en fonction du type d’objet numérique que nous désirons identifier de façon pérenne.

Ce guide est destiné aux responsables des Centres de Données et Services (CDS) en charge de la création et de l’attribution de PID/DOI dans leur structure.

Ce guide est constitué de six sections principales : une présentation générale des PID, la syntaxe et la structure des PID, le choix du PID en fonction de l’objet numérique, la granularité du DOI pour un jeu de données, le choix de l’entrepôt, et une brève introduction sur le concept des « *PID graphs* » (toiles d’information).

2. Présentation générale des PID

Un PID (*Persistent IDentifier*) est un identifiant pérenne d’une ressource, d’un objet numérique.

Le PID prend la forme d’une chaîne de caractères. Celle-ci est unique au niveau mondial et est associée à un service qui localise la ressource dans le temps, même si son emplacement change (un résolveur). Cette chaîne de caractère et ce service forment une URL qui est redirigée de façon transparente pour un usager, soit vers la ressource elle-même, soit vers une *landing page* (page d’atterrissage) où est décrite la ressource (métadonnées de la ressource) et qui fournit un lien vers cette ressource. Ce lien (URL) est pérenne car il est possible de mettre à jour le lien de redirection auprès de la base de données du PID. Un PID peut être attribué à quasiment tout ! : des documents, des données, des personnes, des échantillons, des expériences, des logiciels, des modèles, des définitions de variables, etc. Les PID sont, bien entendu, indispensables si on veut se conformer aux principes FAIR.

A ce jour, les catégories les plus fréquentes d’objets numériques ayant un PID (Madden et al., 2020f, 2020a, 2020e, 2020c, 2020b, 2020d) sont :

- Jeux de données ;
- Personnes/Chercheurs/Individus ;
- Institutions et organisations ;
- Logiciels ;
- Publications.

Il existe plusieurs systèmes de PID comme :

- DOI ;
- ORCID ;
- ROR ;
- Handle ;
- IGSN ;
- Etc.

Chaque catégorie d'objet numérique peut avoir un ou plusieurs systèmes de PID possibles. Voici une liste non exhaustive de différents systèmes de PID déjà existants par catégorie d'objets (Madden et al., 2020f, 2020a, 2020e, 2020c, 2020b, 2020d & PID forum⁴) :

- Jeux de données : DOI, Handle, URN-NBN.
- Personnes/Chercheurs/Individus : ORCID, ISNI, WoS ResearchID, Scopus Author ID, Research ID.
- Institutions et organisations : ROR, CrossRef FunderID, ISNI, GRID, Ringgold, Wikidata, LEI, PSI.
- Logiciel : DOI, SWH-ID, Commit hashes.
- Publication : DOI, ISBN, Handle, arXiv ID, Wikidata item, ISSN, PMC ID.

Les PID ne sont pas un concept récent. Par exemple, cela fait longtemps que les livres publiés par des éditeurs à travers le monde se voient attribuer un numéro ISBN, ce qui en facilite le catalogage. Dans le monde académique, cela fait des années également que les PID sont utilisés pour les publications. La plupart des publications ont un PID de type DOI, ce qui permet d'en suivre l'impact. Cela est pris compte, notamment, dans l'évaluation de la notoriété et du rayonnement d'un chercheur, d'un laboratoire, d'une université.

L'utilisation de PID permet aussi une citation plus aisée d'un objet numérique et sa réutilisation. Par le biais des PID, du fait qu'ils soient plus facilement identifiables que du simple texte (ils sont aussi uniformes et harmonisés), il est ainsi possible d'effectuer la bibliométrie de l'objet, à savoir d'effectuer une analyse quantitative de l'objet. Pour cela, la bibliométrie fait appel à l'utilisation d'indicateurs bibliométriques tels que : facteur d'impact, *h-index*, *alternative metrics*, etc. (Anonymous, 2020).

Une autre utilité des PID est, notamment, l'interopérabilité : le dialogue machine-to-machine est rendu possible par de tels identifiants. Ces PID sont également intéressants car ils permettent une citation et une réutilisation plus aisée de jeux de données et de logiciels. Il est ainsi plus facile d'attribuer les droits d'auteurs de façon 'correcte' (ne permettant pas la confusion). Un PID, une fois attribué, ne peut pas disparaître ou plutôt la *landing page* qui y est associée ne devrait pas disparaître. Cette *Landing Page* reprend les métadonnées de l'objet numérique considéré. Voici, ci-dessous, une énumération, non-exhaustive, des métadonnées reprises pour les objets numériques suivants :

⁴ PID forum: <https://www.pidforum.org>

- Un jeu de donnée : le(s) producteur(s), le titre, un résumé et/ou la méthode d'acquisition, l'emprise géographique, les paramètres mesurées, l'année de création, les modifications opérées, la licence d'utilisation, etc.
- Un identifiant de personne (ex. ORCID) : l'identifiant attribué, le(s) nom(s), le prénom(s), affiliation (ou les affiliations), etc.
- Institution : le nom en langue d'origine, le nom traduit en anglais, l'abréviation officielle, l'adresse (pays inclut), etc.
- Logiciel : le(s) créateur(s), le nom, le langage, la version, etc.
- Publication : le(s) auteur(s), l'affiliation des auteurs, le titre, l'année de publication, le journal/la revue, l'éditeur, un résumé, mot(s) clés), etc.

Les PID sont un des éléments indispensables aux principes FAIR (voir le « [Guide principes FAIR - Principes FAIR dans le contexte du pôle ODATIS⁵](#) ») et ils sont très importants dans le contexte du pôle ODATIS (Schmidt et al., 2020). L'objectif est que tous les centres de données et services (CDS) du pôle ODATIS, et même l'IR Data Terra dans son ensemble, adoptent des principes homogènes, ce qui facilitera l'accès aux données.

Ce besoin d'uniformisation des CDS ODATIS et ceux des autres pôles de l'IR Data Terra concerne, par exemple, la publication de jeux de données qui devra se faire avec la mise en place d'un DOI afin d'en améliorer la traçabilité. Les *landing pages* associées au DOI devront être les plus homogènes possibles pour trouver à minima les mêmes informations (métadonnées riches, niveau de granularité du jeu de données, etc.).

L'objectif est donc d'harmoniser les pratiques des CDS d'ODATIS au sujet des PID et de la mise en place de DOI sur les jeux de données.

3. Syntaxe et Structure des PID

Comme indiqué précédemment, un PID est une chaîne de caractère unique au niveau mondial et est associé à un résolveur. Cette chaîne de caractère et ce résolveur forment une URL qui est redirigée de façon transparente pour un usager soit vers la ressource elle-même, soit vers une « *Landing Page* » où elle est décrite avec un lien vers cette ressource.

Les PID ne sont pas construits et créés de façon totalement arbitraire. En effet, chaque PID doit être unique et répondre à une syntaxe contrôlée.

⁵ Guide principes FAIR - Principes FAIR dans le contexte du pôle ODATIS : <https://doi.org/10.13155/87107>

La syntaxe contrôlée d'un PID est composée, dans le cadre du DOI, de deux parties : (DOI)/ préfixe / suffixe (Deschamps-Ostanciaux et al., 2017).

Le préfixe est spécifique à l'entité qui l'a créé et qui est en charge de la gestion de la *landing page* du PID. Ce préfixe est attribué à l'établissement en charge du DOI par DataCite (ou autre fondation en charge des DOI, tel que CrossRef). Le suffixe, quant à lui est décidé de façon arbitraire par l'établissement en charge de l'attribution du DOI. Exemple, le préfixe, pour tout jeu de données déposé (uniquement) sur SEANOE est « 10.17882 ». Le préfixe attribué à tout jeu de données déposé dans Sextant est quant à lui « 10.12770 ».

Le suffixe d'un jeu de donnée est l'identifiant spécifique au jeu de donnée attribué au sein de l'entité gestionnaire de la *Landing Page* et du PID. Ce suffixe peut prendre la forme d'une suite de chiffre, d'une chaîne de lettres ou d'une suite de chiffres et de lettres. **Le premier conseil très important** est que **le suffixe doit être opaque**, c'est-à-dire qu'il ne doit pas avoir une signification en tant que tel. Il est ainsi conseillé de ne pas introduire, par exemple, le nom d'un laboratoire ou d'une UMR ou d'une organisation, dans le suffixe. En effet, ces noms sont susceptibles de changer dans le temps et il ne sera pas possible de changer le suffixe après-coup. Dès lors, préférez les suffixes opaques tels qu'une combinaison de chiffres et de lettres aléatoires. **Un second conseil est d'éviter des suffixes à rallonge**, en effet plus le suffixe est long, plus il risque d'être mal retranscrit et/ou cité. Un suffixe trop long rendrait également le PID illisible (imaginer devoir mettre un DOI qui ferait trois lignes) et prendrait trop de place dans les listes de références. Les utilisateurs seraient, dès lors, plus susceptibles de l'omettre, ce qui rendrait l'intérêt d'un PID caduque. **Définissez donc un suffixe court, de préférence de moins de 16 caractères**. DataCite recommande de mettre un suffixe le plus court et le plus opaque possible afin que le suffixe ne soit pas porteur de sens.

L'ensemble « préfixe + suffixe » est dès lors une combinaison unique au monde.

DOI : FORMAT & RECOMMANDATIONS

DOI = préfixe + suffixe

Format : doi.org/préfixe/suffixe

PRÉFIXE – caractéristiques	SUFFIXE - recommandations
<ul style="list-style-type: none"> • Spécifique à l'entité • Attribué par une fondation en charge des DOI (e.g., DataCite) 	<ul style="list-style-type: none"> • Opaque • Court • Moins de 16 caractères



Figure 1. Format et recommandations pour le DOI du pôle ODATIS

4. Comment choisir un PID ?

Les critères de choix doivent être pris en fonction de la pérennité de l'organisme qui porte le PID, la préférence de la communauté, la reconnaissance du PID par les éditeurs de revues, la conformité du schéma de métadonnées, etc. Dans certains cas, les choix sont assez simples comme pour la citation publique de jeux de données ou d'un document : le DOI s'impose. Ou pour l'identification publique de contributeurs, l'ORCID devient incontournable. Cependant le choix peut être plus compliqué comme pour l'identification d'organismes biologiques (DOI, IGSN, etc.).

4.1. ORCID

ORCID (<https://orcid.org/>) est un système de PID qui permet d'identifier et de distinguer les contributeurs de la recherche (scientifiques, ingénieurs, techniciens, etc.). L'identifiant ORCID est une liste de 4 blocs de « 4 numéros et lettres », ex : [0000-0002-6938-5004](https://orcid.org/0000-0002-6938-5004)⁶. Ce système est porté par un organisme à but non lucratif basé aux Etats-Unis d'Amérique.

Le gouvernement français a initié un consortium « Communauté française ORCID » en octobre 2019. Ce consortium est porté par contrat et inclut 36 établissements français (dont les établissements de recherche) afin de faciliter l'implémentation de l'identifiant dans les établissements. La *Landing Page* d'un ORCID comprend de base le nom de la personne, qu'il est souhaitable de compléter par l'organisme de rattachement, et en général la liste de ses publications.

Il est à noter que l'obtention d'un ORCID est une démarche personnelle et qu'un établissement ne peut pas contraindre un employé à le créer. Un problème est que de nombreuses *landing pages* d'ORCID sont vides. Pour éviter ce problème, l'IFREMER a développé à partir des API d'ORCID un formulaire où l'utilisateur autorise ARCHIMER (la base de données documentaire de l'IFREMER pour son personnel) à alimenter et mettre à jour automatiquement sa page ORCID. Les informations sont mises à jour quotidiennement pour chaque utilisateur.

Pour SEANOE, l'ensemble des dépôts bénéficie d'un DOI auprès de DataCite. De ce fait, au moment de l'envoi par SEANOE du remerciement du dépôt, le contributeur est informé que DataCite peut automatiquement alimenter son compte ORCID avec les dépôts effectués dans DataCite.

⁶ <https://orcid.org/0000-0002-6938-5004>

4.2. DOI

Un DOI peut être associé à une publication, ainsi qu'à un jeu de données. Un DOI est une chaîne de caractères avec un préfixe qui correspond à l'organisme qui l'attribue et un suffixe. La combinaison des deux (préfixe + suffixe) est une chaîne unique au niveau mondial. Ce DOI, comme tout PID, est une redirection vers une URL pointant vers la *Landing Page*.

Pour les centres désirant attribuer des DOI, il est nécessaire de choisir une agence de DOI (CrossRef, DataCite, etc.). Il existe une dizaine d'agences, mais chaque agence est spécialisée : par exemple CrossRef dans l'attribution de DOI aux documents (ex : les publications). Pour attribuer des DOIs aux données, la plupart des centres utilisent DataCite qui est spécialisée dans l'attribution de DOI aux jeux de données. Le représentant de DataCite en France est l'INIST qui permet de souscrire un contrat, sous forme d'un abonnement (pour tout organisme publique celui-ci était de 180€/an en 2020), afin de pouvoir attribuer un nombre illimité de DOI. Une fois le contrat souscrit, il est possible de demander un ou plusieurs préfixes. L'Ifremer a par exemple demandé 7 préfixes (pour SEANOE, le SISMER, ARCHIMER, etc.).

Pour activer un DOI, il y a deux possibilités :

- soit le demandeur passe par des formulaires web proposés par DataCite ;
- soit le demandeur passe par l'API de DataCite.

Dans les deux cas, pour activer un DOI, il est nécessaire de poster l'URL de la *Landing Page* et de fournir cinq métadonnées obligatoires : *DOI*, *Creator*⁷, *Title*⁸, *Publisher*⁹, *Publication*, *Year*¹⁰ et un ensemble de métadonnées optionnelles. L'URL et les métadonnées obligatoires doivent être maintenues à jour. La *Landing Page* doit indiquer comment accéder à la donnée (URL, formulaire d'enregistrement, page de login, email, services de sélection des données), ou si la donnée n'est plus accessible. **Il est important de noter qu'un DOI ne peut pas être supprimé !**

Il existe également un consensus, au niveau de la France, pour que les jeux de données publiques utilisent le DOI comme type de PID.

⁷ *Creator* : Auteur et/ou producteur (de données)

⁸ *Title* : Titre

⁹ *Publisher* : Editeur

¹⁰ *Year* : Année

5. Comment choisir la granularité d'un DOI pour un jeu de données ?

Pour choisir la granularité d'un DOI, il faut se souvenir de la finalité du DOI qui est la citation du jeu de données dans un article scientifique, un rapport, un catalogue de données et métadonnées, *map viewer*, etc. Un DOI peut être attribué à plusieurs sous-ensembles d'un même jeu de donnée : par paramètre(s), par région(s) d'étude, par année/période temporelle, par unité statistique, etc. Pour déterminer le meilleur niveau de granularité, il faut envisager toutes les réutilisations et citations possibles de ce jeu de données. Si des consignes ou accords pour la granularité existent au sein d'une communauté, il est conseillé de les suivre. Si des incertitudes persistent, il est recommandé d'en discuter directement avec l'entrepôt envisagé pour savoir ce qu'ils recommandent. Il est donc conseillé d'imaginer comment le jeu de données sera utilisé dans un article scientifique. Il est d'ailleurs possible d'attribuer un DOI à plusieurs niveaux d'un même jeu de donnée. Il faut donc réfléchir à la granularité du DOI à l'intérieur d'un jeu de données en fonction de la taille des fichiers, de leur utilisation, du nombre de fichiers, des paramètres, des régions d'étude, des versions, etc.

Pour la gestion des versions, le choix doit être effectué en fonction :

- Du besoin de citation spécifique des différentes versions,
- De la reproductibilité,
- Des contraintes techniques liées au volume des données,
- De l'entrepôt retenu,
- Etc.

Dans SEANOE, l'utilisateur n'a pas la possibilité de modifier lui-même le jeu de données par une version plus récente après l'attribution d'un DOI. SEANOE a donc une discussion avec l'utilisateur pour évaluer ce qu'il veut modifier. Il y a donc plusieurs possibilités comme obtenir un nouveau DOI, ou bien conserver le même DOI pour l'ensemble des versions. La deuxième solution est préconisée, car elle permet de simplifier les consignes aux auteurs et d'avoir une *Landing Page* unique pour les différentes versions, ce qui donne une meilleure visibilité dans Google. De plus, cela permet aux usagers d'accéder à la version la plus récente du jeu de données et de laisser les versions plus anciennes en accès à la demande. Il est aussi possible d'attribuer des fragments pour citer une version spécifique dans un seul DOI.

Pour le choix des auteurs, il est possible soit d'avoir une liste de personnes, soit une liste de collectivité d'auteurs (associés à une liste de contributeurs). Ceci a une incidence sur le choix des versions car un auteur arrivant en cours de projet peut avoir la paternité de jeux de données précédant son arrivée (ce qui peut être problématique).

6. Choix de l'entrepôt

Pour choisir un entrepôt pour publier un jeu de données, il faut au préalable :

- Vérifier la disponibilité d'un entrepôt institutionnel,
- Vérifier les recommandations des éditeurs d'articles (certains éditeurs imposent une liste d'entrepôts),
- Explorer les annuaires d'entrepôts (ex : <https://fairsharing.org/databases/>).

Dans le cadre d'ODATIS, le choix se porte sur deux entrepôts de données, SEANOE et SEXTANT.

SEANOE peut être utilisé pour toutes les données orphelines qui ne sont pas déjà incluses dans un catalogue de métadonnées. SEANOE propose :

- Un entrepôt de jeux de données marines recommandés par ODATIS, SeaDataNet, Elsevier, PLOS, etc. ;
- Un service accessible gratuitement à l'ensemble de la communauté internationale ;
- Un service où les données sont obligatoirement en libre accès après un embargo éventuel de 2 ans avec une licence Creative Commons (avec la recommandation d'utiliser de préférence la licence CC-BY) ;
- Un service où tous les jeux de données disposent d'un DOI ;
- Un service où les dépôts sont validés en 24h (ouvrables) maximum, avec un maximum de 100Go par dépôt ;
- Un outil optimisé pour le référencement dans Google.
- Permet d'avoir une veille sur les citations et sur les articles qui ont cités la donnée. Ce qui permet aussi d'avoir une analyse des citations sous forme graphique. (Bibliométrie)
- Si le jeu de donnée déposé est intéressant pour être déposé dans EMODnet, celui-ci est automatiquement poussé via EMODnet ingestion qui permet au centre de données européen d'organiser son ingestion.

Sextant - Chaque Centre de Données et de Services (CDS) ODATIS expose ses produits et ses données à travers un catalogue de métadonnées qui est accessible depuis le [portail ODATIS](#)¹¹. Celui-ci est basé sur Sextant via l'utilisation d'une API, ce qui permet que :

- Tous les produits peuvent bénéficier de la fonction de DOI-fication (de Sextant) pour se voir attribuer un DOI. Pour cela, il suffit de demander la création du DOI lorsque la description du produit est finalisée. Une demande est envoyée à l'équipe d'administration de Sextant qui prend ensuite le relai et attribue le DOI dans un délai de 24h00.
- Un produit peut, dans sa description, faire référence à un DOI créé par un autre entrepôt de données.

¹¹ Catalogue ODATIS : <https://www.odatis-ocean.fr/donnees-et-services/acces-aux-donnees/catalogue-complet>

7. PID graphs

L'utilisation de PID peut être perçue comme le tissage d'une gigantesque toile d'informations qui se croisent, s'associent, se superposent, etc. Dans la communauté PID¹², cette toile d'informations est appelée *PID graph*, tel illustré dans la Figure 1.

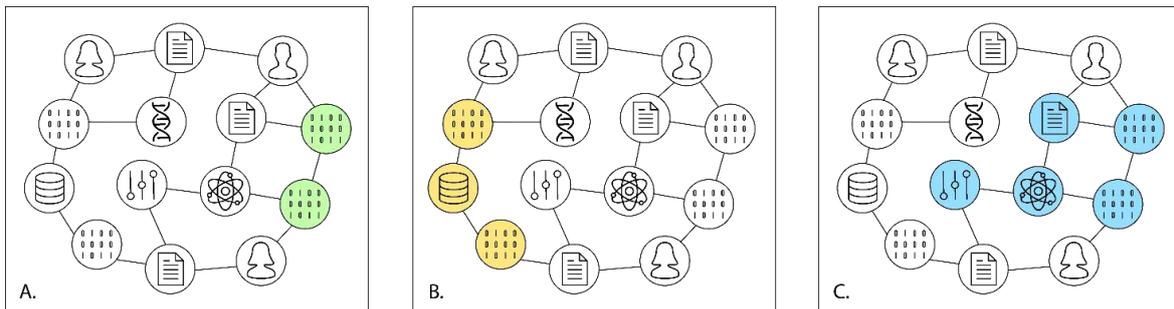


Figure 2. Représentation schématique du PID graph avec des objets numériques connectés par différents types de PID.

La figure A illustre différentes versions d'un code logiciel ; la figure B des jeux de données hébergés par un entrepôt de donnée spécifique et la figure C l'ensemble des objets numériques reliés à un Research Object (Objet de recherche) (Fenner & Aryani, 2019).

Prenons l'exemple d'un jeu de données auquel nous aurions attribué un DOI. Sa *Landing Page* fait appel à d'autres PID au travers les différentes informations :

- Les **auteurs et/ou producteurs d'un jeu de données**, qui ont souvent un identifiant de type ORCID ; les **affiliations des auteurs** peuvent elles aussi faire l'objet d'un PID tel que le ROR.
- L'**emprise géographique** du jeu de données, qui peut être piochée dans une norme préétablie de type ISO ou autre vocabulaire contrôlé, ainsi l'utilisateur est sûr de son interprétation du lieu identifié.
- Les **paramètres P01**¹³ du BODC¹⁴, dans le cadre d'un jeu de donnée océanographique physique et/ou chimiques : ces paramètres font appel à un thesaurus centralisant l'ensemble des paramètres et construits sur un modèle sémantique. De sorte qu'un même type de prélèvement soit nommé partout de la même façon, pour faciliter l'interopérabilité de jeu de données issus de laboratoires distincts. Ces paramètres P01 sont associé à une **URI** qui a elle-même sa propre *Landing Page* définissant ainsi de manière claire et non équivoque le paramètre en question.

¹² PID forum : <https://www.pidforum.org>

¹³ P01: BODC Parameter Usage Vocabulary (https://vocab.seadatanet.org/v_bodc_vocab_v2/search.asp?lib=P01)

¹⁴ BODC : British Oceanographic Data Center (<https://www.bodc.ac.uk/>)

8. Conclusion

La finalité de ce guide est d'apporter les recommandations de base pour l'attribution des DOI aux produits gérés par les Centre(s) de Données et Services (CDS) d'ODATIS.

Il est important de retenir que chaque CDS :

- A la possibilité de gérer lui-même ses DOI s'il en a la capacité (exemple du CDS AVISO+). Les DOI seront alors indiqués dans les fiches de métadonnées des produits et accessibles à travers le catalogue ODATIS.
- Peut bénéficier des fonctions de DOI-fication proposées par les entrepôts Sextant et SEANOE.

9. Références

- Anonymous. (2020). *La bibliométrie | Sciences Po Grenoble*. <http://www.sciencespo-grenoble.fr/les-ressources/appui-a-la-recherche/la-bibliometrie/>. Accessed 2020-11-12.
- Buck, J., Merceur, F., & Exter, K. (2020). *A compilation of current best practices on PIDs for the ENVRI-FAIR PID task force*.
- Cousijn, H., Kenall, A., Ganley, E., Harrison, M., Kernohan, D., Lemberger, T., Murphy, F., Polischuk, P., Taylor, S., Martone, M., & Clark, T. (2018). A data citation roadmap for scientific publishers. *Scientific Data*, 5, 1–11. <https://doi.org/10.1038/sdata.2018.259>
- DataCite Metadata Working Group, de Smaele, M., Dasler, R., Ashtin, J., Bernal Martínez, I., Burger, M., Fenner, M., Habermann, T., Ilik, V., Jacobson, M., Raugh, A., la Roi, A., Roy, S., Yahia, M., & Zolly, L. (2016). DataCite Metadata Schema Documentation for the Publication and Citation of Research Data. In *DataCite - International Data Citation DataCite*. DataCite. <https://doi.org/10.14454/7xq3-zf69>
- Deschamps-Ostanciaux, E., Galle, S., André, F., Hosford, S., & Merceur, F. (2017). Good practices for data citation. In *Groupe interpôle - Guide*. https://www7.obs-mip.fr/wp-content-aeris/uploads/sites/18/2017/05/INTERPOLE_GUIDE_CITATION_V1.0_ENG.pdf
- Fenner, M., & Aryani, A. (2019). *Introducing the PID Graph*. FREYA. <https://doi.org/10.5438/jwvf-8a66>
- Ferguson, C., Mcentrye, J., Bunakov, V., Lambert, S., Kotarski, R., Stewart, S., Macewan, A., Fenner, M., Cruse, P., van Horik, R., Dohna, T., Koop-Jacobsen, K., Schindler, U., & McCafferty, S. (2018). *D3.1 Survey of Current PID Services Landscape*. <https://doi.org/10.5281/zenodo.1324296>
- Groth, P., Cousijn, H., Clark, T., & Goble, C. (2020). FAIR Data Reuse – the Path through Data Citation. *Data Intelligence*, 2(1–2), 78–86. https://doi.org/10.1162/dint_a_00030
- Hellström, M., Heughebaert, A., Kotarski, R., Manghi, P., Matthews, B., Ritz, R., Conrad, A. S., Weigel, T., Wittenburg, P., & Valle, M. (2020). *Second draft Persistent Identifier (PID) policy for the European Open Science Cloud (EOSC)*. <https://doi.org/10.5281/zenodo.3780423>
- Hellström, M., Johnsson, M., Vermeulen, A., Lear, D., & Fiebig, M. (2019). *Report on identification and citation service case studies*. <http://www.envriplus.eu/wp-content/uploads/2015/08/D6.3.pdf>
- Hellström, M., Lassi, M., Vermeulen, A., Huber, R., Stocker, M., Atkinson, M., & Fiebig, M. (2020). *D6.1 - A system design for data identifier and citation services for environmental RIs projects to prepare an ENVRIPLUS strategy to negotiate with external organisations*. ENVRIplus.
- Leadbetter, A., Raymond, L., Chandler, C., Pikula, L., Pissierssens, P., & Urban, E. (2013). Ocean Data Publication Cookbook. In *IOC Manuals and Guides* (Vol. 64, p. 41). UNESCO. <http://www.odip.eu/documents/odip/downloads/48/mg64.pdf>
- Madden, F., van Horik, R., van de Sandt, S., Lavasa, A., & Cousijn, H. (2020a). Dataset Identifiers. In *Guides to Choosing Persistent Identifiers - Version 3*. <https://doi.org/10.5281/zenodo.4192174>
- Madden, F., van Horik, R., van de Sandt, S., Lavasa, A., & Cousijn, H. (2020b). Organization Identifiers. In *Guides to Choosing Persistent Identifiers - Version 3*. <https://doi.org/10.5281/zenodo.4192174>
- Madden, F., van Horik, R., van de Sandt, S., Lavasa, A., & Cousijn, H. (2020c). People Identifiers. In *Guides to Choosing Persistent Identifiers - Version 3*. <https://doi.org/10.5281/zenodo.4192174>
- Madden, F., van Horik, R., van de Sandt, S., Lavasa, A., & Cousijn, H. (2020d). PID tool tables. In *Guides to Choosing Persistent Identifiers - Version 3*. <https://doi.org/10.5281/zenodo.4192174>

- Madden, F., van Horik, R., van de Sandt, S., Lavasa, A., & Cousijn, H. (2020e). Publication Identifiers. In *Guides to Choosing Persistent Identifiers - Version 3*. <https://doi.org/10.5281/zenodo.4192174>
- Madden, F., van Horik, R., van de Sandt, S., Lavasa, A., & Cousijn, H. (2020f). Software Identifiers. In *Guides to Choosing Persistent Identifiers - Version 3*. <https://doi.org/10.5281/zenodo.4192174>
- Merceur, F. (2016). *DOIs for ocean data General principles and selected examples (Argo , French cruises)* (Issue May, pp. 1–13). Ifremer. <https://doi.org/10.13155/44515>
- Merceur, F. (2017). *DOI de données de la recherche: Question fréquentes*. <https://doi.org/10.13155/60247>
- ODATIS. (2021). *Ateliers Techniques*. [www.Odatis-Ocean.Fr](https://www.odatis-ocean.fr/activites/ateliers-techniques). <https://www.odatis-ocean.fr/activites/ateliers-techniques>
- Rauber, A., Asmi, A., Uytvanck, D. Van, & Pröll, S. (2015). *Data Citation of Evolving Data - Recommendations of the Working Group on Data Citation (WGDC)* (p. 2). Research Data Alliance.
- Schmidt, S., Maudire, G., Nys, C., Sudre, J., Harscoat, V., Dibarboure, G., & Huynh, F. (2020). Streamlining Data and Service Centers for Easier Access to Data and Analytical Services: The Strategy of ODATIS as the Gateway to French Marine Data. *Frontiers in Marine Science*, 7(December), 1–10. <https://doi.org/10.3389/fmars.2020.548126>
- Stall, S., Cruse, P., Cousijn, H., Cutcher-Gershenfeld, J., de Waard, A., Hanson, B., Heber, J., Lehnert, K., Parsons, M., Robinson, E., Witt, M., Wyborn, L., & Yarmey, L. (2018). Data Sharing and Citations: New Author Guidelines Promoting Open and FAIR Data in the Earth, Space, and Environmental Sciences. *Science Editor*, 41(3), 83–87.
- Sudre, J., Nys, C., Boulanger, D., Carval, T., & Mercier, C. (2021). *Compte rendu de l'atelier technique des 15 et 16 décembre 2020 - COPILOtE: FAIR, Identifiants Permanents, autoévaluation CTS*. <https://doi.org/10.13155/81251>
- Wilkinson, M. D., Sansone, S. A., Schultes, E., Doorn, P., Da Silva Santos, L. O. B., & Dumontier, M. (2018). A design framework and exemplar metrics for FAIRness. *Scientific Data*, 5, 7–10. <https://doi.org/10.1038/sdata.2018.118>