

GROUPE INTERPÔLE

GUIDE

Bonnes pratiques autour de la citation

v1.0

SOMMAIRE

1. Périmètre du document.....	5
1.1. Objectif	5
1.2. Quels objets digitaux ?.....	5
1.3. A qui s'adresse ce document ?	5
2. Citation des jeux de données	5
2.1. Objectif et contraintes	5
2.1.1. Objectifs.....	5
2.1.2. Contraintes	6
2.2. DOI	6
2.2.1. Structure	6
2.2.2. Fonctionnement	6
2.2.1. Notion d'identificateurs de fragments (ou fragments)	7
2.2.2. Datacite.....	7
2.2.3. Alternatives.....	7
2.2.4. Limites	8
2.3. UUID.....	8
3. Liste des bonnes pratiques.....	8
3.1. Bonnes pratiques liées au suffixe.....	8
3.1.1. Neutralité du suffixe [Recommandé].....	8
3.1.2. Concision du suffixe	8
3.1.3. Utilisation des fragments pour indiquer les variantes d'un jeu de données	9
a) Avantages 9	
b) Inconvénients 9	
3.1.1. Neutralité et concision du fragment	10
3.2. Bonnes pratiques liées à la Landing page	10
3.2.1. Affichage du texte de citation	10
3.2.2. Interprétation du fragment.....	11
3.2.3. Une fiche de métadonnées comme landing page	11
3.2.4. Lien MLP / DOI	12
3.3. Bonnes pratiques liées au diffuseur de la donnée.....	12
3.3.1. Présence des informations de citation dans les fichiers de données	12
3.4. Bonnes pratiques liées au producteur du DOI	12
3.4.1. La production d'un DOI doit être soumise à une décision concertée.....	12
3.4.2. Le producteur des DOI doit être le responsable de la conservation des données	12
3.4.3. Validité des landing pages	13
3.5. Bonnes pratiques liées aux aspects bibliométriques	13
3.5.1. Informations bibliométriques d'un DOI	13
3.5.2. Attribution précise des auteurs	13
4. FAQ.....	15
4.1. Quelles données citer ?	15
4.1. Quelles métadonnées d'un DOI peuvent être modifiées ?	15
4.2. Peut-on supprimer un DOI?	15
4.1. Peut-on transférer des DOI ?	15
4.1. Que faire des DOI déjà émis et qui ne respectent pas les préconisations de ce document?	15
4.2. Deux DOI peuvent-ils pointer sur un même jeux de données ?	15

4.3.	Comment appliquer ces pratiques pour citer plusieurs jeux de données ?.....	16
4.4.	Doit-on avoir un DOI par niveau de traitement ?	16
4.5.	Où citer un DOI de données dans une publication ?	16
4.6.	Où trouver les DOI attribués par DataCite ?	16

REDACTEURS

Initiales	Nom	Organisme / Pôle
EDO	Emilie Deschamps-Ostanciaux	Form@ter
SG	Sylvie Galle	Théia
FA	François André	Aeris
SH	Steven Hosford	Cnes
FM	Frédéric Merceur	Odatis

HISTORIQUE DES REVISIONS DU DOCUMENT

Date	Modifications	Auteur
08/06/16	Création de document	FA
10/06/16	Complément du document initial / Intégration éléments discussion 08/06/16	FA, EDO, SG, SH, FM
29/05/17	Prise en compte de retours sur le précédent document. Notamment: - Ajout de la FAQ - Ajout paragraphe sur les avantages/inconvénients des fragments significatifs	FA
31/05/17	Finalisation lors du workshop Interpoles	FA,SG,EDO
01/06/17	Passage 0.2.5 -> 1.0	FA

DOCUMENTS DE REFERENCE

Nom du document	Référence
CEOS Persistent Identifier Best Practices_v1.0.pdf http://www.aeris-data.fr/rest/redirect/doi/CEOS Persistent Identifier Best Practices_v1.0.pdf	[CEOS]
OGC OpenSearch Geo and Time Extensions http:// portal.opengeospatial.org/files/?artifact_id=56866	[OSGEOT]
ANDS DOI FAQ http://www.ands.org.au/working-with-data/citation-and-identifiers/data-citation/doi-faqs	[ANDSFAQ]
ODATIS - DOI de données - FAQ http://...	[ODATISFAQ]

1. PERIMETRE DU DOCUMENT

1.1. OBJECTIF

Ce document a pour objectif:

1. de rappeler certaines définitions autour des DOI
2. de définir un certain nombre de bonnes pratiques autour de la citation d'objets digitaux. Une fois ces pratiques définies et validées,
 - une analyse d'écart pourrait être dressé au sein des différents pôles ou centre de données
 - la mise en œuvre de chacune d'elle au sein des pôles pourrait faire l'objet d'un suivi.
3. de répondre à certains questions usuelles par le biais d'une FAQ.

1.2. QUELS OBJETS DIGITAUX ?

Les DOI sont, littéralement, des identifiants d'objets digitaux. Que sont ces *objets digitaux* ? Ils peuvent être de natures diverses. Les cas les plus fréquents sont les suivants:

- Un jeu de données (ou une collection de données)
- Un algorithme
- Un site web (celui d'un projet, d'un programme ou d'une mission)
- Un article en ligne
- ...

Dans ce document, **seuls les jeux de données et les collections seront abordés.**

1.3. A QUI S'ADRESSE CE DOCUMENT ?

La production de DOI est liée au catalogage et à la conservation à long terme des données. Ainsi, une grande partie des préconisations s'adresse aux gestionnaires de données (dans des services d'observation, des observatoires ou des pôles de données).

Toutefois, certains éléments contenus dans ce document, qui apportent un éclairage neutre sur le DOI, peuvent intéresser à la fois le producteur de la données et les représentants des organismes de tutelle

2. CITATION DES JEUX DE DONNEES

2.1. OBJECTIF ET CONTRAINTES

2.1.1. Objectifs

L'objectif premier de la citation de données est de permettre le **référencement concis, précis et fiable à travers le temps de données**. Cette citation peut s'effectuer par exemple dans un article scientifique mais aussi sur un site Internet ou dans le corps d'un message électronique.

Un deuxième objectif est de permettre également d'établir, via le mécanisme de citation, une estimation de l'importance des jeux de données au sein de la communauté via l'établissement **de mesures bibliométriques**. Un tel besoin peut également être étendu au projet/programme/mission.

2.1.2. Contraintes

Plusieurs contraintes sont à prendre en compte pour atteindre ces objectifs :

- **Non pérennité des URL**
Contrairement aux articles scientifiques qui ont une assez forte pérennité, les URL ne sont en général valides que quelques années. Par exemple, une URL peut devenir caduque suite à un changement de nom de domaine. Ainsi utiliser des URL n'est pas une solution adaptée pour la citation et la mise en place d'un mécanisme de redirection dans lequel le nom de la redirection est fixe et l'adresse de redirection est variable est nécessaire.
- **Cycle de vie du jeu de données et des métadonnées associées**
Un jeu de données évolue après qu'il a fait l'objet d'une citation : les données peuvent être corrigées, ajoutées, le format modifié, ... De même la liste des auteurs du jeu de données peut également évoluer dans le temps dans le cas d'une série de données évolutive. Ainsi, le mécanisme de citation doit pouvoir apporter de manière souple des éléments permettant de s'adapter à ces différents cas de figure.
- **Concision**
 La citation doit pouvoir figurer dans un document comme un article imprimé auquel sont attachées de fortes contraintes de place. Ainsi la citation doit donc rester courte et ne pas comporter d'éléments qui pourraient entacher son utilisation comme par exemple une impression sur plusieurs lignes du fait de la longueur, la présence d'espaces ou de caractères non alphabétiques.

2.2. DOI

2.2.1. Structure

Les DOI (Digital Object Identifier) sont une technologie largement utilisée afin de permettre le référencement d'objets digitaux. Ils sont composés des éléments suivants:

Élément	Rôle
Préfixe	Chaîne de caractère identifiant l'entité ayant produit le DOI
Suffixe	Chaîne de caractère identifiant un jeu de données au sein de l'entité pointée par le préfixe.
Métadonnées	Les métadonnées permettant de comprendre le jeu de données doivent être fournies au moment de la production du DOI. Parmi ces métadonnées certaines sont obligatoires : <ul style="list-style-type: none"> ▪ Creator (Créateur) ▪ Title (Titre) ▪ Publisher (Diffuseur) ▪ Publication year (Année de publication)
Url de landing page	La landing page (LP) est une page web qui va permettre d'avoir plus d'information sur le jeu de données et de le récupérer (sous réserve d'éventuelles restrictions).

Exemple de DOI :

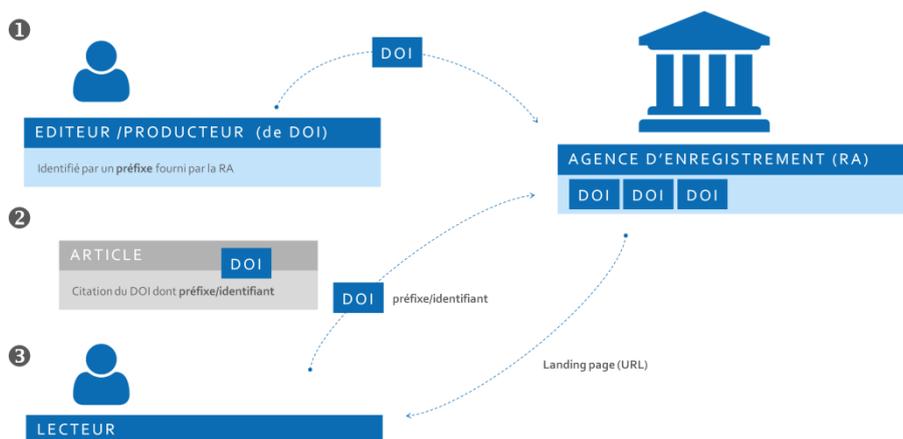


2.2.2. Fonctionnement

Les DOI sont enregistrés au sein d'agence spécialisée (Registration Agency ou R.A.). Ces agences proposent des mécanismes de résolution qui permettent de rediriger un utilisateur sur la landing page d'un DOI.

Ainsi les 3 étapes de l'utilisation d'un DOI sont :

1. Sa création et son enregistrement dans une R.A.
2. Sa citation dans un article (papier ou autre)
3. Sa résolution - via la R.A. - afin de trouver la landing page.



Une fois créé, les informations du DOI - à l'exception du suffixe - peuvent être modifiées (cf. supra) notamment l'URL de sa landing page.

2.2.1. Notion d'identificateurs de fragments (ou fragments)

La notion d'identificateur de fragments est une norme portée entre autre par le W3C (<https://www.w3.org/TR/media-frag/>) visant à pointer vers une sous-partie d'une ressource. Pour ce faire, à l'url principale est concaténée une chaîne de caractères complémentaire commençant par le signe #. Dans le reste du document on utilisera de manière équivalente les termes *fragments* et *indicateur de fragments*.

Cette norme est supportée par les préconisations DOI (https://www.doi.org/doi_handbook/5_Applications.html#5.8) lors de la résolution du DOI dans la mesure où le fragment est transmis à la LP.

Ainsi si on a le DOI `10.17882/43082` ayant comme landing page `http://monsie/maLP` la résolution du DOI `10.17882/43082#monFragment` sera `http://monsie/maLP#monFragment`.

Il est important de noter que l'utilisation des fragments fait qu'un seul DOI est publié et mis à jour.

2.2.2. Datacite

DataCite est l'organisation spécialisée dans l'enregistrement de DOI dont la cible sont des jeux de données. Son fonctionnement repose sur des correspondants à l'échelle nationale. En France, ce rôle d'intermédiaire est assuré par l'INIST.

Chaque institution scientifique peut souscrire auprès de l'INIST pour 180 €/an un contrat qui permet

- D'obtenir un ou plusieurs préfixe(s).
- D'émettre un nombre infini de DOI.

2.2.3. Alternatives

Un certain nombre d'alternatives au DOI existent. On peut citer par exemple les mécanismes ARKs, Handle ou PURLS qui passent également par la médiation d'une agence d'enregistrement.

2.2.4. Limites

L'un des engagements forts lors de la production d'un DOI est sa pérennité de la pertinence de sa landing page. Ainsi, un DOI ne peut pas être supprimé (cf. FAQ).

Dans cette optique, l'utilisation d'un DOI pour fournir un lien fiable et modifiable vers des documents non pérennes est un détournement de l'objectif principal qui mènera à l'obtention d'un grand nombre de LP vides. Dans ce cas, il convient plutôt d'utiliser des services de redirection d'url que proposent certains sites web ou qui peuvent être mis en place localement.

2.3. UUID

La création d'un identifiant universellement unique (UUID) est une opération assez simple et repose sur la production de chaînes aléatoires suffisamment longues (en général 16 à 32 caractères hexadécimaux). Pour des raisons de lecture ces chaînes sont souvent segmentées au moyen de tirets.

Un certain nombre de standards ont été spécifiés comme par exemple ISO/IEC 9834-8:2008.

L'intérêt de ces mécanismes est qu'ils permettent la génération de cet identifiant facilement, sans avoir à se préoccuper d'un contexte local. De plus l'unicité de cet identifiant sera conservée dans un autre contexte. On peut par exemple penser à l'identifiant d'un auteur qui doit pouvoir être utilisé dans plusieurs publications.

Par son triplet protocole(DOI)/préfixe/suffixe, le DOI constitue également un UUID dont l'utilisation est vouée à la diffusion et à la publication.

3. LISTE DES BONNES PRATIQUES

Cette section propose un certain nombre de bonnes pratiques autour de la mise en place. Pour chaque pratique un rappel de la problématique est apporté ainsi que certains éléments d'explication des choix proposés. De plus un niveau de recommandation est également indiqué.

3.1. BONNES PRATIQUES LIÉES AU SUFFIXE

3.1.1. Neutralité du suffixe [Recommandé]

Les recommandations officielles n'indiquent aucune contrainte sur le formalisme du suffixe. Certains DOI produits comportent ainsi des parties significatives comme le nom d'un projet, d'un site d'observation... Toutefois, ces éléments n'ont pas forcément une pérennité importante construire un suffixe contenant de telles parties peut s'avérer contradictoire avec celle que l'on souhaite apporter au DOI.

DOI-SUFF-01 [Recommandé] : Le suffixe ne doit pas être porteur de sens.

Cette recommandation est très fréquemment reprise dans les différents documents de préconisation comme par exemple [CEOS] ou [ANDSFAQ].

3.1.2. Concision du suffixe

Un certain nombre de mécanismes d'identifiants proposent des identifiants basés sur des UUID de 16 à 32 caractères. Ainsi, on peut, par exemple, citer l'identifiant OrCID (orcid.org).

Dans le cas du mécanisme de type DOI, l'utilisation de tel mécanisme peut être, à la fois, contre-productif en produisant des DOI trop long (cf. supra) tout en n'étant pas nécessaire, le préfixe permettant de garantir une universalité.

Ainsi le suffixe doit être aussi court que possible dans le contexte de son préfixe. On pourrait ainsi envisager des suffixes simplement croissants (1,2,3...).

DOI-SUFF-02 [Recommandé]: L'universalité du DOI doit être portée alors fois par le préfixe et le suffixe (et non par le suffixe seul)

Cette recommandation reprend REC_03 dans [CEOS]. Cette recommandation peut également être rapprochée du nommage des classes en java dans lequel une partie du nom du package correspond à l'organisme réalisant le développement afin de garantir l'unicité universelle du nom de la classe.

3.1.3. Utilisation des fragments pour indiquer les variantes d'un jeu de données

Un jeu de données peut exister sous la forme de plusieurs variantes:

- Sous forme de sous-ensemble (subset) du jeu de données plus large (restriction temporelle, restriction géographique,...)¹.
- Sous forme de version d'un jeu de données (en cas de retraitement)
- Sous forme d'extraction (snapshot) d'un jeu de données à un moment précis (pour un article par exemple)

Un grand nombre de questions concernant les DOI concerne la manière de mettre en place un mécanisme de DOI permettant de répondre à ces différentes questions. Une réponse est apportée par la norme DOI qui permet de mettre en place des liens entre des DOI (ex: *isPartOf*, *isPreviousVersionOf*, ...). Ces solutions semblent complexes à mettre en place de manière homogène et à maintenir. Ainsi nous déconseillons leur utilisation.

A la place nous préconisons l'utilisation de fragments (cf. supra) pour désigner toutes les variantes d'un unique DOI associé au . En effet, avec l'utilisation des fragments, la LP peut tirer profit de cette information complémentaire pour afficher des services plus pertinents comme par exemple un lien de téléchargement plus précis. Ainsi, la fiche de métadonnées perd son rôle uniquement statique. **En effet le fragment lui ajoute des informations relatives à l'extraction des données.**

a) Avantages

- Avec les fragments, un seul DOI doit être mis en place pour décrire l'ensemble des citation d'un jeu de données.
- Les fragments permettent d'apporter une **première réponse à la question de la citation fine des données** (cf. recommandations RDA : https://rd-alliance.org/system/files/documents/RDA-DC-Recommendations_151020.pdf).

b) Inconvénients

L'utilisation des fragments apporte un nouvel élément porteur de sens. **Il faudra donc veillez à maintenir le sens apporté par un fragment** du moment où on a autorisé la production de DOI les utilisant:

- Si la landing page est hébergée sur un autre système il faudra que celui-ci supporte l'interprétation des fragments déjà émis.
- En cas d'évolution du mécanisme des fragments (par exemple du nom des paramètres), une compatibilité descendante avec les fragments déjà émis.
- Enfin, la LP doit signaler l'incompréhension d'un fragment transmis.

DOI-SUFF-03 [Recommandé] : les variantes d'un jeu de données doivent être indiquées via des fragments et ne doivent pas donner lieu à la production de nouveaux DOI

¹ De manière plus générale, ce cas la correspond aux données obtenues par application, **à un moment donné**, d'une requête (*query*) composée de différents critères de recherche (temps, zone géographique, paramètre physique, nom de programme,...) sur un ou plusieurs ensembles de données.

3.1.1. Neutralité et concision du fragment

Il est important que la présence du fragment n'aille pas à l'encontre des exigences susmentionnées, notamment celle de concision.

Ainsi, le contenu du fragment doit être un identifiant ayant la taille la plus petite possible. Ceci implique que la signification réelle du fragment doit être stockée dans le système d'information le produisant et accessible par la LP. Hors de cas triviaux, ceci correspond à la notion de magasin de requêtes (query-store) définie dans les recommandations RDA. La mise en place de ce mécanisme impliquant, comme indiqué précédemment, une maintenance du contenu des requêtes.

DOI-SUFF-04 [Recommandé] : les fragments doivent être concis et non significatifs. Leur réelle signification doit être conservée et maintenue au sein de query stores.

Remarque:

La non significativité des fragments entraîne l'obligation de passer par l'intermédiaire du portail de données pour obtenir le fragment. Ceci peut sembler antagoniste avec le développement actuel des interface REST d'extraction des données. En effet dans de tels mécanismes les paramètres d'extraction sont passés à la requête HTTP via une succession de clés/valeurs accolée à l'url. Ainsi on peut avoir une requête du type:

`http://mon.service.extraction/identifiantJeuDeDonnées?param1=valeur1¶m2&valeur2...`

Avec de tels mécanismes l'utilisateur peut librement agir sur les différents paramètres

Adopter la même syntaxe pour les fragments des DOI - `prefixe:suffixe#param1=valeur1¶m2&valeur2...` - apporte la même liberté à l'utilisateur qui peut modifier/adapter lui même la citation.

Si le nombre de paramètre reste faible on peut considérer que la citation reste suffisamment concise. Dans ce cas l'utilisation de fragments explicites apporte les avantages suivants:

- Pas besoin de mettre en place un query-store.
- Homogénéité entre l'interface REST et la citation
- Liberté pour l'utilisateur.

Toutefois, il faut bien se rappeler que l'extraction a un aspect ponctuel alors que le DOI à une vocation pérenne. Les inconvénients sont alors les suivants:

- La compatibilité ascendante est plus complexe à mettre en œuvre car il est difficile de savoir quels sont les paramètres et les valeurs qui sont utilisés par les utilisateurs, l'analyse d'un query-store n'étant pas possible. Ainsi, elle doit être effectuée systématiquement sur tous les paramètres/valeurs.
- Pour éviter que cette tâche de maintenance ne devienne trop lourde, il faut veiller à choisir de manière pertinente les noms des clés et des valeurs (en utilisant par exemple des dénominations issues de dictionnaires adoptés par la communauté)
- Il ne faut pas oublier qu'un des paramètres doit être la date de la requête.

3.2. BONNES PRATIQUES LIEES A LA LANDING PAGE

3.2.1. Affichage du texte de citation

La landing page doit rappeler le texte de citation afin de permettre son incorporation rapide dans un article. La structuration de la citation proposé par DataCite est le suivant :

Creator PublicationYear: Title. Publisher. Identifier

Remarques:

- Plusieurs style peuvent être appliqués à cette structure (APA par exemple). Datacite propose une un service de formatage : <http://crosscite.org/citeproc/>
- Le texte de citation doit prendre en compte d'éventuels fragments passés dans l'URL de la LP.

DOI-LP-01 [Recommandé]: Affichage du texte de citation sur la LP

3.2.2. Interprétation du fragment

La LP doit être capable de retranscrire un éventuel fragment sous forme compréhensible par l'utilisateur. Elle doit pouvoir signaler des cas comme :

- un fragment non compréhensible
- une donnée n'étant plus disponible dans la version identique à celle présente lorsque le fragment a été généré.
- ...

DOI-LP-02 [Recommandé]: Interprétation d'un éventuel fragment par la LP

3.2.3. Une fiche de métadonnées comme landing page

Dans le cas d'un DOI portant sur un jeu de données, il est pertinent de faire en sorte que la landing page soit une fiche d'un catalogue de métadonnées celle ci affichant des informations plus complètes que les métadonnées du DOI. Ceci permet la mise en œuvre de DOI-SUFF-01, les informations autrefois portées par le suffixe apparaissant dans le fiche de métadonnées.

Dans la suite du document on appellera *MLP* (metadata landing page) une telle fiche de métadonnées.

La MLP doit comporter au minimum les informations suivantes:

- Titre
- Résumé
- Personne à contacter pour les données
- Lien d'accès aux données
- Liens d'informations sur les formats de données (outils,...)

Le lien permettant d'accéder aux données peut prendre plusieurs formes plus ou moins précises, par ordre de préférence:

- Lien de téléchargement : lien permettant de récupérer le jeu de données ou un surensemble de celui-ci. La personne consultant la MLP est grandement guidé.
- Lien de recherche: lien permettant d'accéder à une interface de recherche/sélection/extraction de données. La personne consultant la MLP doit effectuer l'extraction manuellement.
- Lien de commande: lien pointant vers un formulaire de commande de données. L'extraction est faite par un opérateur du centre de données et ensuite mise à disposition du demandeur

Les informations suivantes sont également fortement recommandées :

- Empreinte temporelle
- Empreinte spatiale
- Généalogie des données (liens avec d'autres jeux de données, historique des différentes versions du jeu de données)
- Moyens de conservation mis en œuvre autour de la donnée
- Formats d'extraction disponibles

Ces problématiques de contenu de la landing page sont en fait du ressort des **problématiques relatives à la mise en place d'un catalogue (granularité, contenu, lien entre les pages)**.

DOI-LP-03 : La landing page d'un DOI d'un jeu de données doit être une fiche de métadonnées structurée

Remarque: Il peut être intéressant d'uniformiser l'identifiant interne de la MLP et le DOI. En vertu de DOI-SUFF-02, les candidats pour cet identifiant pourraient, par exemple, être de la forme *prefixe_suffixe*.

3.2.4. Lien MLP / DOI

L'ensemble des **métadonnées du DOI doivent être extraites de manière automatisable à partir des informations de la MLP.**

Ceci est valable au moment de la publication du DOI mais également **lorsqu'une des métadonnées est modifiée** (exemple: ajout d'un nouvel auteur).

Remarque: Même si elle est automatisable, la production d'un DOI ne doit pas forcément être automatique.

DOI-LP-04 : L'ensemble des métadonnées du DOI doivent provenir de manière automatisable de la fiche de métadonnées utilisée comme landing page. Les métadonnées du DOI doivent être mise à jour si le contenu du MLP l'est.

3.3. BONNES PRATIQUES LIEES AU DIFFUSEUR DE LA DONNEE

3.3.1. Présence des informations de citation dans les fichiers de données

DOI-LP-01 indique la nécessité de présenter le texte de la citation dans la MLP. Ce texte, dont le DOI, doit également, si possible être attaché dans les fichiers de données diffusés.

Il peut par exemple être mis en place dans les en tête des fichiers NetCDF. Il peut être également mis dans un fichier ReadMe.txt inclus dans l'archive des données.

DOI-DIFF-01 : Le fichier de données doit contenir un rappel des éléments de citations. Au minimum son DOI.

3.4. BONNES PRATIQUES LIEES AU PRODUCTEUR DU DOI

3.4.1. La production d'un DOI doit être soumise à une décision concertée

Dès qu'un DOI est produit et donc peut être cité dans un article, il est nécessaire qu'une démarche de qualité - incluant notamment des traitements visant à sa pérennisation - soit mise œuvre autour des données concernées. De fait la décision de produire un DOI pour un jeu de données doit faire l'objet d'une décision concerté entre les différents acteurs de la donnée : équipes scientifiques, équipes techniques, ...

DOI-PROD-01 [Recommandé] : La production d'un DOI est une décision

3.4.2. Le producteur des DOI doit être le responsable de la conservation des données

Come indiqué précédemment, l'enregistrement d'un DOI est un acte d'engagement important. Il n'est donc pas conseillé d'émettre un DOI pour des données dont on aurait pas la maîtrise et donc pour lesquelles on ne pourrait pas garantir la pérennité.

DOI-PROD-02 [Recommandé]: Le producteur d'un DOI est le responsable de la conservation des données

3.4.3. Validité des landing pages

Il est nécessaire qu'un producteur de DOI mette en place un mécanisme qui va régulièrement - quotidiennement par exemple - vérifier que les différentes landing page sont toujours actives et ne correspondent pas à des liens morts.

Ce mécanisme de vérification doit faire remonter automatiquement des alertes dans le système de surveillance du SI.

DOI-PROD-03 : Le producteur doit mettre en place un mécanisme de vérification de l'état des LP.

3.5. BONNES PRATIQUES LIEES AUX ASPECTS BIBLIOMETRIQUES

3.5.1. Informations bibliométriques d'un DOI

Le DOI est un UUID dont le rôle principal est la citation au sein de publications. Il est donc important de mettre en place un certain nombre de service bibliométriques autour d'un DOI afin de connaître le nombre de fois où il est cité. Indirectement, il peut être également utile de savoir si un DOI ne fait actuellement l'objet d'aucune citation.

Ce service a été mis en place dans certains systèmes d'information mais n'est pour le moment pas proposé par les principaux acteurs du domaine mais certains travaux, menés en collaborations avec les éditeurs sont en cours. On peut citer par exemple <https://dliservice.research-infrastructures.eu>. C'est vers ces mécanismes tiers qu'il faudrait se tourner.

DOI-BIB-01 : Les informations bibliométriques d'utilisation d'un DOI devront pouvoir être calculées nativement via les futurs mécanismes tiers.

Il est possible que ces futurs mécanismes permettent de faire le distinguo entre les différentes variantes d'un DOI (fragments). Il serait toutefois intéressant que les services bibliométriques permettent l'utilisation de caractères *joker* (notamment le "*").

3.5.2. Attribution précise des auteurs

Les DOI gèrent la notion d'auteur. Pour illustrer ce point, on peut souligner qu'en cas de présence d'un identifiant OrcId, le profil Orcid du contributeur est automatiquement mis à jour.

Ainsi, à travers le mécanisme des DOI s'ouvre la possibilité de calculer des métriques précises pour les producteurs de jeux de données eux même, comme par exemple le h-index.

Là aussi, les services ne sont pas encore disponibles. Toutefois, on peut émettre un certain nombre d'hypothèses :

- De tels services seront proposés par les principaux acteurs du domaine
- Ces services ne prendront pas en compte les fragments.

La conséquence du dernier point est que le calcul du h-index prendra en compte uniquement les informations disponibles au moment du calcul au sein de Datacite c'est à dire la dernière liste d'auteurs.

Plusieurs cas de figure se présentent :

a) Le jeu de donnée n'est pas évolutif

Dans ce cas la liste d'auteur est figée. Les solutions susmentionnées peuvent s'appliquer.

b) Le jeu de données est évolutif

b1) Pas de besoin d'attribution fine

Si le producteur du jeu de données n'est pas intéressé par une attribution fine des auteurs, la meilleure pratique est de faire que la liste des auteurs transmises à Datacite (et donc celle de la MLP) contiennent l'intégralité des différents auteurs étant intervenus sur le jeu de données sans suppression.

b2) Besoin d'attribution fine

Dans le cas d'un jeu de données évolutif, le producteur de données peut vouloir que l'attribution portent uniquement sur les auteurs concernés par les données téléchargées.

Comme susmentionné, **dans ce cas, le mécanisme des fragments sera inefficace.**

Plusieurs solutions sont possibles :

- Publier un DOI pour chaque fichier produit
- Publier un DOI lors de chaque changement d'auteur

Ces solutions ont l'avantage de leur simplicité. Toutefois, elles présentent comme inconvénient la production d'un grand nombre de DOI dont la plupart ne seront inutilisés.

Une solution, plus proche de l'utilisation des fragments, peut être la suivante :

1. **Ajouter pour les auteurs, une période d'activité au sein de la MLP.**
2. Lorsque des données sont téléchargées, au lieu de proposer une citation avec le DOI du jeu de données (monDoi#monFragment), le système doit proposer une **citation qui convertit le DOI à fragment en un DOI statique (par exemple monDoi__monFragment)**. Ce DOI aura la liste des auteurs réellement concernés par les données.
3. Publier ce DOI statique sur Datacite (soit systématiquement, soit via un mécanisme de demande de la part de la personne ayant téléchargé la donnée). Ce DOI sera compatible avec le mécanisme de calcul de l'attribution de Datacite.

DOI-BIB-02 : Les informations d'attribution des auteurs doivent pouvoir être calculées nativement via les futurs mécanismes tiers.

S'il n'est pas besoin d'attribution fine, le mécanisme des fragments doit être utilisé.

Dans le cas contraire, le mécanisme des fragments doit être adapté pour le jeu de données afin de publier des DOI statiques correspondant aux jeux réellement utilisés dans des publications.

4. FAQ

4.1. QUELLES DONNEES CITER ?

D'après [ANDSFAQ]

Un DOI peut s'appliquer sur des données totalement publiques ou disponibles avec un accès restreint (ex: ouverture aux membres d'une communauté). Dans ces cas la LP est publique et comporte soit des liens permettant la récupération de la données, soit des informations de contact du médiateur de la communauté.

Pour des données privées, l'utilisation du mécanisme externe du DOI semble un peu contradictoire et d'autres mécanismes doivent être utilisés.

4.1. QUELLES METADONNEES D'UN DOI PEUVENT ETRE MODIFIEES ?

D'après [ODATISFAQ]

A l'exception du suffixe, toutes les métadonnées peuvent être modifiées.

4.2. PEUT-ON SUPPRIMER UN DOI?

D'après [ODATISFAQ]

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 LP doit alors présenter aux internautes les raisons de la suppression de l'accès à la donnée.

4.1. PEUT-ON TRANSFERER DES DOI ?

Il est totalement possible de transférer la responsabilité d'un DOI d'une entité à une autre du moment où la deuxième entité est bien enregistrée auprès de DataCite (et est donc dépositaire d'un préfixe différent du premier).

Dans ce cas, aucun changement n'est effectué dans le nom du DOI: le préfixe et le suffixe restent les mêmes. Le transfert de responsabilité est purement administratif. Toutefois, en général, un tel transfert est suivi d'un changement de l'URL de la LP.

4.1. QUE FAIRE DES DOI DEJA EMIS ET QUI NE RESPECTENT PAS LES PRECONISATIONS DE CE DOCUMENT?

Certaines des préconisations de ce document peuvent être mise en œuvre progressivement après l'émission du DOI comme par exemple celles qui concernent le contenu de la LP.

Par contre, certaines ne peuvent s'appliquer a posteriori. Notamment celles qui concernent le suffixe. Par contre, dans ce cas, il peut être intéressant d'effectuer des recherches bibliographiques afin d'estimer si le DOI est réellement utilisé afin de mesurer les conséquences de la production d'un deuxième suffixe concernant le même jeu de données (c.à.d la même LP). Dans ce cas, il faudra bien maintenir actif le premier DOI et de s'assurer que les informations des deux DOI conserveront les même informations dans le temps.

4.2. DEUX DOI PEUVENT-ILS POINTER SUR UN MEME JEUX DE DONNEES ?

Deux cas de figure se présentent:

1) Les DOI ont des préfixes différents

Dans ce cas, la réponse doit être clairement *non*. En effet, seul le centre responsable de la conservation doit publier les DOI. La mise en place de DOI sur des données gérées par un autre centre constituent un risque de LP morte à moyen ou long terme.

2) Les DOI ont le même préfixe

Dans ce cas la réponse est *idéalement non*. Toutefois, ce cas est possible:

- S'il s'agit de l'émission d'un DOI plus conforme aux préconisations ainsi qu'indiqué plus haut.
- Il est également possible qu'un mécanisme de fragments puisse générer des DOI différents pour un même jeu de données. Toutefois, il convient d'essayer d'éviter que cela se produise.

4.3. COMMENT APPLIQUER CES PRATIQUES POUR CITER PLUSIEURS JEUX DE DONNEES ?

La citation précise de données issues d'un même producteur de données, peuvent porter sur plusieurs jeux de données (et donc plusieurs DOI ayant le même préfixe). Ainsi, elle peut devenir longue et donc difficilement citable.

Dans ce cas là, une bonne solution consiste à mettre en place une fiche de métadonnées décrivant l'extraction et contenant, dans la section *liens*, l'ensemble des DOI. A cette page ont peu faire correspondre un DOI qui sera facilement citable.

4.4. DOIT-ON AVOIR UN DOI PAR NIVEAU DE TRAITEMENT ?

Certains jeux de données comportent plusieurs niveaux de traitement (données brutes, paramètres géophysiques, ...). Normalement, le catalogage devrait distinguer ces niveaux de traitements dans plusieurs fiches de métadonnées distinctes, chacune possédant son DOI.

Remarque:

Si les fiches comportent une section généalogie, celle-ci pourrait utiliser les DOI des jeux de données parents.

4.5. OU CITER UN DOI DE DONNEES DANS UNE PUBLICATION ?

D'après [ODATISFAQ]

Certaines revues commencent à imposer des consignes spécifiques. Par exemple, 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.

4.6. OU TROUVER LES DOI ATTRIBUES PAR DATA CITE ?

D'après [ODATISFAQ]

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