

Gestion des Données Ouvertes dans L'Agriculture et la Nutrition

Ce cours en ligne est le fruit d'une collaboration entre les partenaires de GODAN Action, y compris Wageningen Environmental Research (WUR), AgroKnow, AidData, l'Organisation des Nations Unies pour l'Alimentation et l'Agriculture (FAO), le Forum Mondial sur la Recherche Agricole (GFAR), l'Institut des Etudes du Développement (IDS), le Land Portal, l'Open Data Institute (ODI) et le Centre Technique de Coopération Agricole et Rurale (CTA).



GODAN Action est un projet de trois ans du Département pour le Développement International du Royaume-Uni pour permettre aux utilisateurs, producteurs et intermédiaires de données de s'engager efficacement avec les données ouvertes et maximiser leur potentiel d'impact dans les secteurs de l'agriculture et de l'alimentation. Nous travaillons en particulier à renforcer les capacités, à promouvoir des normes communes et les meilleures pratiques et à améliorer la manière dont nous mesurons l'impact. [www.godan.info]

MODULE 2: UTILISATION DES DONNÉES OUVERTES

LEÇON 2.5 : Référencement des données



Photo by [Neil Palmer \(CIAT\)](#) licensed under CC BY-SA 2.0

Objectifs et résultats d'apprentissage

Le partage de la recherche scientifique est un processus établi de longue date, depuis la création de la Royal Society en Angleterre en 1660. La publication d'articles dans des revues scientifiques a permis aux scientifiques de revendiquer leurs découvertes tout en les partageant avec d'autres. Des siècles plus tard, notre capacité de partager les résultats va maintenant bien au-delà de la simple publication et s'étend aux données recueillies dans le cadre du processus de recherche. Cette leçon examine comment des principes similaires sont maintenant utilisés pour découvrir et référencer les données scientifiques.



À la fin de cette leçon, vous devrez être en mesure :

- D'expliquer l'importance de la citation
- En énumérer les principales caractéristiques et expliquer chacune d'elles.
- De comprendre la valeur des identificateurs persistants dans les données
- D'identifier les bonnes pratiques existantes pour l'utilisation des identificateurs persistants.

Sommaire

Module 2 : Utilisation des données ouvertes.....	2
Leçon 2.5: Données de référencement.....	2
Objectifs et résultats d'apprentissage.....	2
Liste des illustrations.....	4
1. Introduction	5
2. Arguments en faveur du partage des données scientifiques.....	5
3 Identificateurs de données persistants sur le Web.....	7
3.1 Fournisseurs de DOI.....	9
4. Discussion	10
Références	11

Liste des illustrations

Illustration 1 Le schéma 5 étoiles pour les données ouvertes liées entre elles.....	7
Illustration 2 Exemple de DOI pour les données.....	8
Illustration 3 Utilisation d'un DOI pour récupérer une notice bibliographique.....	9

1. Introduction

Le partage de la recherche scientifique est un processus établi de longue date, depuis la création de la Royal Society en Angleterre en 1660. La publication d'articles dans des revues scientifiques a permis aux scientifiques de revendiquer leurs découvertes tout en les partageant avec d'autres. Bien des siècles plus tard, la publication d'articles scientifiques ne consiste pas seulement à revendiquer des résultats, mais aussi à les utiliser comme mesure de performance.

Le mode classique de diffusion des résultats scientifiques est la publication dans des revues professionnelles. Les revues professionnelles font l'objet d'un examen par les pairs et d'un contrôle de qualité rigoureux, ce qui signifie que la quantité de travail nécessaire pour être publiée n'est pas négligeable.

La récompense de ce travail est l'inclusion dans l'"index des citations". L'index sert à évaluer le rendement des scientifiques. Cela signifie que l'effort de publication est récompensé tant au sein de la communauté qu'au niveau de la carrière et des promotions.

Les citations répondent donc à un certain nombre d'objectifs clés :

- revendiquer la propriété
- avoir un impact
- référencement cohérent
- facilité de découverte.

Alors que les deux premiers sont des avantages pour l'auteur, les deux derniers aident à la réutilisation. Tout comme à la fin de cette leçon, les références aident à mettre les gens en contact avec des données probantes ou des documents de support. Les références contiennent au minimum le nom de l'auteur, le lieu de publication et l'année de publication. Ces facteurs aident à la fois à indiquer l'impact, mais aussi à faciliter la découverte, en permettant aux utilisateurs de trouver le bon numéro de la revue et le numéro de page de l'article qui les intéresse.

L'ensemble de ces quatre aspects signifie que les scientifiques sont incités et récompensés pour la publication de travaux de recherche de haute qualité. La citation constitue un important mécanisme de confiance qui montre aux lecteurs que vous êtes un auteur responsable et que vous avez cité des travaux dignes de confiance dans votre domaine.

2. Arguments en faveur du partage des données scientifiques

La publication d'articles de recherche est une partie importante de la méthode scientifique. Plus importante que l'impact personnel, cette méthode est conçue

pour faire progresser l'enseignement, la recherche et la pratique d'un domaine d'étude académique ou universitaire donné.

Le partage de la recherche scientifique est essentiel à l'avancement de l'enseignement, de la recherche et de la pratique. Il permet aux scientifiques d'un domaine d'étude de valider et de mettre à profit les recherches des autres chercheurs.

La validation et l'expansion des résultats de la recherche peuvent nécessiter l'accès à des ressources rares, comme des télescopes ou des composés chimiques, mais plus récemment, de nombreuses expériences ont utilisé des données comme intrants. Il y a donc conséquemment eu des appels en faveur du partage des données dans le cadre du processus de publication scientifique.

En principe, les scientifiques sont prêts à fournir des données, mais le travail supplémentaire nécessaire qu'ils doivent effectuer dans le traitement, la documentation du contexte et l'assurance qualité de ces données n'est souvent ni apprécié ni reconnu (Brase et al., 2015). Inversement, les données semblent indiquer que ceux qui rendent les données disponibles ont reçu plus de citations que les études similaires pour lesquelles les données n'étaient pas disponibles, ce qui donne à penser que les données peuvent accroître l'impact (Piwowar, 2013).

L'importance du partage des données de recherche a fait en sorte que de nombreuses parties ont maintenant mis en place des politiques et des mandats qui font respecter la publication des données de recherche.

Bon nombre des plus grandes revues et des plus grands éditeurs ont maintenant des exigences en matière de partage des données qui doivent être satisfaites. Pour en savoir plus sur certaines de ces questions, cliquez ici :

- Science's [Data Deposition Policy](#) (Politique de dépôt des données de la science)¹; faites défiler vers le bas jusqu'à la section " Disponibilité des données et du matériel après la publication ".
- Springer Nature's [Data Sharing Policies](#)²; Voir la politique de la Nature comme exemple spécifique de journal sur la Disponibilité des données, du matériel et des méthodes ([Availability of data, materials, and methods](#)³)
- Wiley's [Data Sharing Service](#) (Service de partage de données de Wiley)
- American Geophysical Union's [Publication Data Policy](#)⁴ (Politique de publication des données de l'American Geophysical Union)
- Sage's [Replication policy](#)⁵ (Politique de réplification de Sage)

¹ <http://www.sciencemag.org/authors/science-editorial-policies>

² <http://www.springernature.com/gp/authors/research-data-policy/?countryChanged=true>

³ <http://www.nature.com/authors/policies/availability.html>

⁴ <https://authorservices.wiley.com/author-resources/Journal-Authors/licensing-openaccess/open-access/data-sharing.html>

⁵ <https://us.sagepub.com/en-us/nam/journal/big-data-society#ARTICLETYPES>

L'un des problèmes que posent bon nombre de ces politiques est que les données accompagnent toujours un article de revue. Les données elles-mêmes ne sont pas un objet de recherche de première classe qui peut être référencé.

En 2004, la Bibliothèque nationale allemande des sciences et de la technologie (TIB) a attribué son premier identificateur d'objet numérique (DOI) aux données scientifiques afin de rendre les ensembles de données scientifiques accessibles aux résultats de recherche.

Le DOI est un identificateur persistant utilisé pour identifier de manière unique des objets, normalisé par l'Organisation internationale de normalisation (ISO). Le DOI vise à fournir un identifiant "résoluble" à un objet numérique, tel qu'une URL. Contrairement à l'ISBN et à l'ISRC, qui ne sont que des identificateurs, un DOI est exploitable et interopérable.

3. Identificateurs de données persistants sur le Web

Le concept d'identificateurs exploitables a d'abord été esquissé par Tim Berners-Lee dans le cadre du schéma 5 étoiles pour les données ouvertes liées.

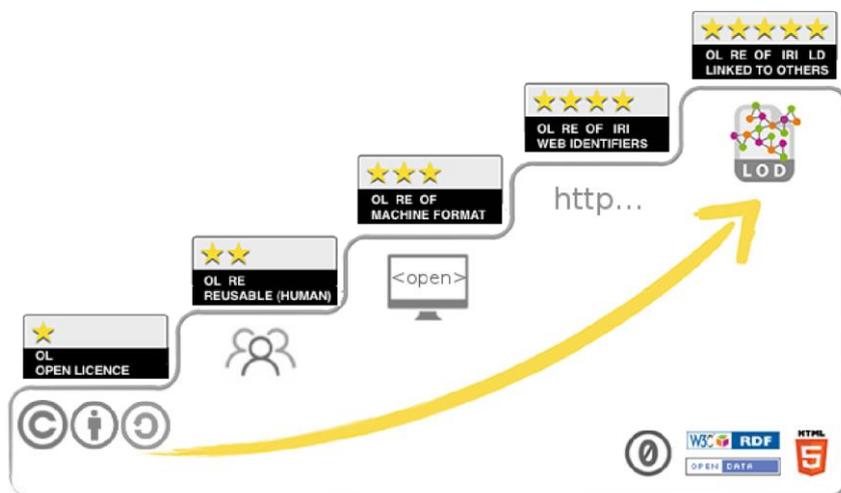


Illustration 1 Le schéma 5 étoiles pour les données ouvertes liées entre elles

Ce schéma décrit le processus d'ouverture et de mise en relation des données sur le Web. Dans le cadre du schéma, la quatrième étape concerne les identificateurs. Le principe général de la 4ème étape est d'utiliser des identificateurs basés sur le Web (URI) pour désigner les objets. De cette façon, les gens peuvent trouver plus d'informations sur vos objets et les cibler grâce à l'utilisation d'un URI résoluble.

Contrairement aux livres, qui utilisent un numéro ISBN comme identificateur, un identificateur Web n'a pas besoin d'un catalogue pour cet identificateur.

Cependant, les identificateurs basés sur le Web peuvent être créés par n'importe qui, sur n'importe quel domaine, et ne détiennent donc pas le même niveau d'autorité.

L'idée de l'identificateur d'objet numérique était de fournir un identificateur fiable et résoluble pour les objets numériques. Essentiellement, un DOI est un identificateur permanent basé sur le Web qui pointe vers l'objet numérique sur le Web, quel que soit son emplacement. Alors que n'importe qui peut mettre un objet numérique sur le Web, les DOI ne sont disponibles que pour les objets qui ont été approuvés comme étant de haute qualité. Essentiellement, le DOI tente de reproduire le même impact qu'un index de citations de revues, uniquement sur le Web (Data Citation Synthesis Group 2014).

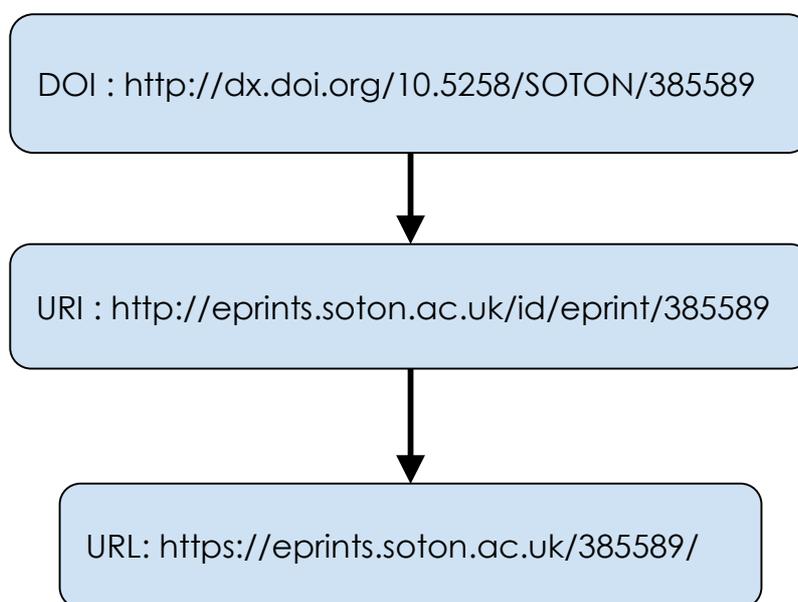


Illustration 2 Exemple de DOI pour les données

L'illustration 2 montre comment un DOI référence un ensemble de données sur le Web. Dans ce cas, l'ensemble de données est disponible dans un dépôt institutionnel. Le référentiel fournit à la fois un identificateur unique pour l'ensemble des données (URI) ainsi qu'une page Web lisible par l'homme sur l'ensemble des données (URL). Le DOI est simplement un autre identificateur pour l'ensemble des données qui pointe vers l'URI réel. L'objectif du DOI est de fournir un identifiant permanent et faisant autorité pour tout objet numérique, quel que soit l'emplacement de l'objet sur le Web.

En plus de fournir un moteur de résolution pour ces identificateurs persistants, le service Web DOI est également capable de fournir directement des métadonnées sur l'objet numérique par la négociation de contenu, comme le montre l'illustration 3.

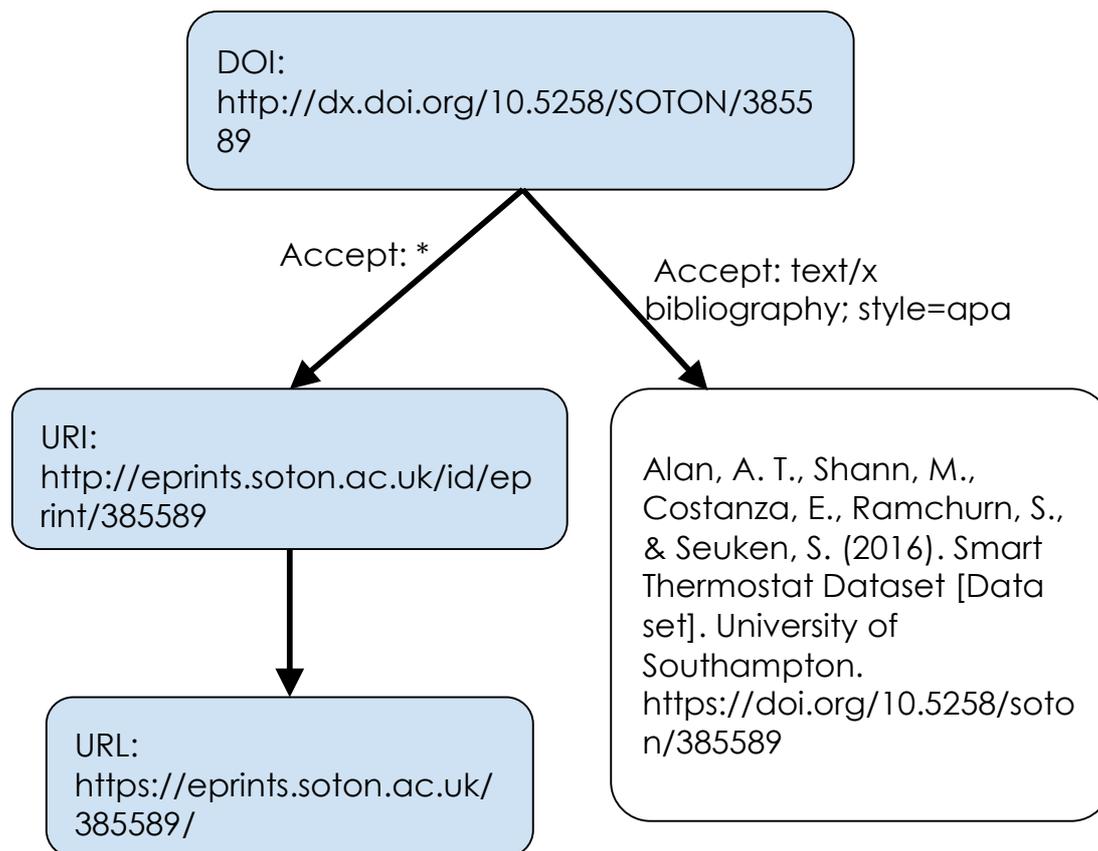


Illustration 3 Utilisation d'un DOI pour récupérer un enregistrement bibliographique

3.1. Fournisseurs de DOI

Un DOI peut être lié à un ensemble de données de plusieurs façons :

- [Figshare](https://figshare.com)⁶ fournira un DOI pour toute œuvre déposée, ce qui comprend les données.
- [Zenodo](https://zenodo.org)⁷ fournit également des DOI pour tout type de résultats de recherche, y compris les ensembles de données.
- [Dryad](http://datadryad.org)⁸ fournit des DOI pour les soumissions de données liées à des articles (moyennant des frais, qui comprennent le stockage, la conservation, l'archivage et la vérification des meilleures pratiques).

Il existe également un certain nombre de référentiels spécifiques à un domaine qui fournissent un DOI pour les données. La revue [Scientific Data](https://www.nature.com/sdata/policies/repositories)⁹ tient à jour une bonne liste de référentiels que vous pouvez consulter.

Le choix d'un fournisseur de DOI dépendra de circonstances particulières. Figshare et Zenodo sont tous deux des services gratuits, par exemple, tandis

⁶ <https://figshare.com>

⁷ <https://zenodo.org>

⁸ <http://datadryad.org>

⁹ <https://www.nature.com/sdata/policies/repositories>

que Dryad facture le service qu'ils offrent pour couvrir les frais de stockage et d'archivage courants.

Dryad accepte les données relatives aux publications ; si elles ne sont pas associées à un article, elles ne seront pas acceptées. Figshare et Zenodo acceptent tout résultat de recherche, qu'il soit lié ou non à la publication. En ce sens, Figshare et Zenodo s'appliquent plus largement à tout résultat de recherche.

Les données peuvent être plus faciles à découvrir dans les dépôts Dryad ou les dépôts spécifiques à un domaine (DSR) que dans les dépôts à usage général comme Figshare et Zenodo. Les données sont susceptibles d'être formatées de manière standard et plus facilement consultables en ligne ou à l'aide d'autres outils si elles sont archivées dans Dryad ou dans des DSR. Cela est susceptible d'encourager la réutilisation.

Figshare est une entité commerciale à but lucratif, Zenodo est géré par le CERN et a été soutenu par le projet européen OpenAIRE à un moment donné, tandis que Dryad est une entité à but non lucratif soutenue par des subventions de recherche et des cotisations de membre pour les organisations.

4. Discussion

Les livres et les articles de revues bénéficient depuis longtemps d'une infrastructure qui les rend faciles à citer, un élément clé dans le processus de recherche et le discours universitaire. Datacite (le groupe responsable du DOI) estime que les objets numériques doivent être citables de la même manière.

Les DOI Datacite sont conçus pour :

- Soutenir l'attribution et le mérite appropriés
- Soutenir la collaboration et la réutilisation des données
- permettre la reproductibilité des résultats
- Favoriser des progrès plus rapides et plus efficaces de la recherche, et
- Fournir les moyens de partager les données avec les futurs chercheurs

Pour les chercheurs, la possibilité de publier et de citer des objets numériques et des objets de première classe incite à publier des données ouvertes de haute qualité et à fort impact. Les plates-formes telles que Figshare facilitent de plus en plus la publication de ces données et la création d'identificateurs de données persistants.

Références

- Brase, J., Sens, I. and Lautenschlager, M. 2015. The tenth anniversary of assigning DOI names to scientific data and a five year history of DataCite. *D-Lib magazine* 21 (1/2). <http://dx.doi.org/10.1045/january2015-brase>

- Data Citation Synthesis Group (2014). Joint Declaration of Data Citation Principles. Martone M. (ed.), FORCE11, San Diego, CA, USA. Available at <https://www.force11.org/group/joint-declaration-data-citation-principlesfinal>
- Piwowar, H.A. and Vision, T.J. 2013. Data reuse and the open data citation advantage. *PeerJ* 1 (2013): e175.