

# DataCite au service des données scientifiques - identifier pour valoriser

Herbert Gruttemeier<sup>1</sup>

<sup>1</sup>Institut de l'information scientifique et technique (Inist–CNRS) - 2, Allée du Parc de Brabois, CS 10310, F-54519 Vandoeuvre-lès-Nancy, France

## Résumé

Les données de la recherche, sous forme d'objets numériques très divers, sont en train de trouver leur place dans les services d'information scientifique et technique (IST), principalement - mais pas uniquement - comme compléments des publications qui s'appuient sur ces données. L'intégration de différents types de ressources numériques avance, et doit être accompagnée par des standards d'interopérabilité, des formats communs de métadonnées et des possibilités de lier ces contenus entre eux et de les citer de manière normalisée.

Le consortium international DataCite, dans lequel l'Inist-CNRS représente la France, s'est mis comme objectif de soutenir et accélérer cette évolution. Il opère en particulier comme une agence d'enregistrement de DOI (*Digital Object Identifier*), considérant ces DOI, déjà bien établis dans le monde de l'édition, comme un outil efficace pour identifier les données de manière pérenne, pour ainsi faciliter leur découverte et pour y accéder, et puis pour les citer. DataCite a développé son propre schéma de métadonnées et a mis en place des fonctionnalités spécifiques qui favorisent le partage et la réutilisation des données. Une telle valorisation s'inscrit en particulier dans une approche de pleinement bénéficier du potentiel des open data.

Elle est aussi une contribution essentielle à une meilleure reconnaissance du travail scientifique de production, gestion et mise à disposition de données, et notamment sa prise en compte dans les critères d'évaluation. Il est d'ailleurs encourageant de voir que ces critères se s'ouvrent à des métriques alternatives, y compris celles concernant les données.

Le sujet particulier de la citation des données a récemment été l'objet de plusieurs initiatives internationales visant à harmoniser les pratiques et émettre des recommandations. Elles ont convergé, à travers le *Data Citation Synthesis Group*, vers quelques principes en train d'être largement reconnus et acceptés. Dans ce contexte, les éditeurs doivent s'adapter et clairement définir leurs politiques en termes de liens entre données et publications. On observe d'ailleurs une tendance forte vers des accords entre éditeurs et réservoirs de données.

Les actions et services de DataCite s'intègrent dans d'autres structures et initiatives internationales mises en place autour des données de la recherche et des identifiants pérennes: Research Data Alliance, WDS-ICSU, CODATA, EPIC, Data Citation Index, etc. Un exemple particulier présente le projet européen ODIN, où DataCite et l'initiative ORCID pour la création d'identifiants d'auteurs tentent de connecter les différents types d'identifiants.

## Mots-Clés:

Données de la recherche, Identification pérenne, DOI, Edition scientifique, Service information, Citation, Initiatives internationales, DataCite

## Introduction – 4<sup>e</sup> paradigme et open data

Les pratiques de recherche et les usages des chercheurs ont fortement évolué ces dernières décennies, dans un contexte d'évolutions technologiques rapides, grandes e-infrastructures partagées et production de quantités gigantesques de données dans certaines disciplines. La Science est probablement le domaine dans lequel le *big data* aura entraîné le plus de bouleversements. Il a amené le 4<sup>e</sup> paradigme scientifique [1], celui d'une recherche guidée par l'exploration de données massives (*data-driven science*).

En parallèle, ces nouvelles pratiques et opportunités ont entraîné une plus grande ouverture des données, favorisée par ailleurs par les tendances *open data* dans le domaine des données publiques, le mouvement *open access* dans celui des publications scientifiques et les possibilités du web sémantique en termes de structuration et interconnexion des données (*linked data*).

La recherche scientifique est en train d'accorder, comme conséquence de ces évolutions, une plus grande place à la gestion et valorisation des données qu'elle produit, de favoriser le partage et la réutilisation de ces données. Il en est de même pour les agences de financement de la recherche (ex. : la gestion des données dans le programme européen Horizon 2020 et son action pilote sur le libre accès à ces données) [2].

Néanmoins, la situation des données en tant que contribution scientifique est infiniment plus complexe que celui des publications, varie énormément en fonction des domaines, de sorte que le but, affiché par DataCite, d'élever les données au rang de « *first class citizens in the scholarly record* » est encore loin d'être atteint. DataCite veut accompagner et accélérer ce processus. Un intérêt particulier est porté sur les *small data*, ou encore 'longue traîne' (*long tail*) de données, celles qui ne bénéficient pas nécessairement d'expertises et infrastructures au sein de leur institution et ne trouvent pas facilement une prise en charge et hébergement par les grandes archives thématiques.

### Le consortium DataCite<sup>1</sup>

En 2009, six organismes européens (grandes bibliothèques, centres d'information), dont l'Inist-CNRS pour la France, ont signé un protocole d'accord en vue de la création d'une agence dédiée à la valorisation et identification à long terme des données issues de la recherche scientifique, possédant en particulier le statut d'une 'agence d'enregistrement de DOI' (*DOI registration agency*). Les signataires partageaient en effet la conviction que les DOI (*Digital Object Identifier*)<sup>2</sup> constituent un système reconnu et efficace pour une telle identification pérenne [3].

Extrait de ce protocole : « Reconnaissant l'importance des collections de données brutes issues de la recherche comme fondement de la connaissance, et partageant une volonté commune de promouvoir et de garantir un accès pérenne à ces collections de données, nous, organismes signataires, exprimons notre intérêt à travailler de concert pour promouvoir un accès global à ces données issues de la recherche. Notre vision à long terme est d'intervenir en appui des chercheurs en mettant à leur disposition des méthodes leur permettant de localiser, d'identifier et de citer des collections de données avec un maximum de confiance. »

---

<sup>1</sup> [www.datacite.org](http://www.datacite.org)

<sup>2</sup> [www.doi.org](http://www.doi.org) [http://fr.wikipedia.org/wiki/Digital\\_Object\\_Identifier](http://fr.wikipedia.org/wiki/Digital_Object_Identifier)

DataCite avait notamment comme objectif une meilleure intégration des données de la recherche dans des services d'IST (Information scientifique et technique), ce qui implique avant tout de promouvoir la pratique de liens entre publications et données et, par là même, de la citation des données.

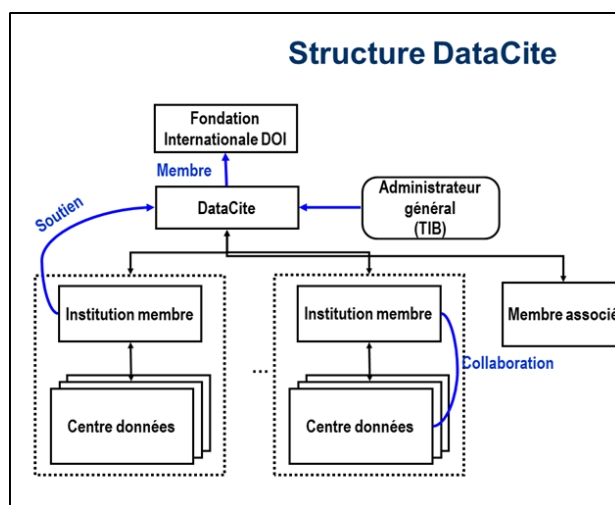
DataCite et ses membres étaient donc destinés à coopérer avec des organismes de recherche, entrepôts et centres de données, bibliothèques, mais également avec des éditeurs. Ceux-ci avaient créé, dans les années 2000, leur propre agence de DOI, CrossRef<sup>3</sup>, pour des besoins (d'identification pérenne et, à partir de là, création de services divers) dans le domaine des revues scientifiques.

Les relations entre les deux acteurs, DataCite et CrossRef, sont bonnes, il y a une sorte de *gentlemen's agreement* sur le partage des rôles pour servir les communautés respectives. CrossRef s'est associé à une déclaration commune, en 2012, entre DataCite et l'association internationale STM des éditeurs scientifiques sur les bonnes pratiques recommandées en matière de dépôt de données par les auteurs, de liens réciproques entre texte et données et de normalisation dans l'établissement de ces liens.

Il semble important de souligner que DataCite est constitué d'organismes publics, faisant un contrepoids à l'agence CrossRef mise en place par des éditeurs commerciaux. DataCite contribue ainsi au maintien de la maîtrise des données de la recherche par le secteur public, et également au mouvement *open data*, car les bénéfices en termes de visibilité et citabilité que le consortium et ses membres offrent à travers leurs services, incitent, comme de nombreux exemples le montrent, à l'ouverture des données.

DataCite ne manque d'ailleurs jamais de rappeler aux éditeurs leur '*Brussels Declaration on STM Publishing*' de 2007, dans laquelle on peut lire (article 7) : « *Raw research data should be made freely available to all researchers.* » [4] Dans la réalité, il y a un nombre croissant d'accords formels entre grands éditeurs et grands réservoirs de données, la coopération entre Elsevier et le centre de données PANGAEA<sup>4</sup>, dans les sciences de la terre et de l'environnement, semble avoir joué un rôle de pionnier dans ce domaine. On peut également mentionner Dryad<sup>5</sup>, entrepôt de données partenaire des éditeurs pour les données associées aux articles publiés (et munies de DOI par Dryad).

Aujourd'hui, DataCite a largement dépassé le cadre européen initial pour devenir une organisation mondiale reconnue, comportant une trentaine de membres et ayant attribué plus de 5 millions de DOI, qui propose un ensemble de services, parmi lesquels certains sont présentés ci-dessous.



En tant qu'agence d'enregistrement de DOI, DataCite adhère à l'IDF (Fondation Internationale DOI). Le consortium DataCite est constitué de membres réguliers (qui attribuent des identifiants) et membres associés (qui soutiennent le consortium d'autres manières). La TIB Hanovre (Bibliothèque nationale allemande pour les sciences et technologies) remplit des fonctions d'administration et support technique du consortium.

<sup>3</sup> [www.crossref.org](http://www.crossref.org)

<sup>4</sup> <http://www.pangaea.de>

<sup>5</sup> <http://www.datadryad.org>

## Les services DataCite

Au-delà de la simple attribution d'identifiants DOI, DataCite a développé des services associés qui facilitent la gestion des données de leurs utilisateurs et augmentent leur visibilité.

- Grâce aux métadonnées associées aux DOI les objets enregistrés sont consultables sur la plateforme MDS (*Metadata Store*) de DataCite. Les options de recherche dans MDS sont d'autant plus puissantes que ces métadonnées sont riches et exploitent les possibilités offertes par le schéma de métadonnées DataCite. Bien entendu, les DOI sont activables à partir de cette plateforme.
- Les métadonnées sont librement accessibles et moissonables (licence *Creative Commons CC0* ou assimilée). Ce principe laisse donc la possibilité d'une exploitation commerciale (éventualité longuement discutée au sein du consortium), la société Thomson Reuters en tire d'ailleurs profit pour le développement de son *Data Citation Index*<sup>6</sup> qui vise in fine la création de services et métriques basés sur les citations (à l'image des services qu'elle propose déjà pour les publications).
- Des statistiques d'enregistrements (par agence d'attribution, par centre de données, par préfixe) et de résolutions de DOI sont mises à disposition en ligne, ces dernières permettant un début de traçabilité de consultation des données.
- Une fonctionnalité importante, encore sous-exploitée aujourd'hui, est offerte par le service de 'négociation de contenus' (*content negotiation*), développée en partenariat avec CrossRef. De façon générale, elle permet à l'utilisateur de demander une représentation particulière d'une ressource en ligne. Ici, elle se base sur un type de résolution spécifique du DOI (comparée avec la résolution 'classique' vers une *landing page*), permettant un lien direct aux métadonnées dans un format souhaité (ou plutôt : préférée, la préférence pouvant faire l'objet d'une 'négociation', d'où le nom).  
*Content negotiation* est typiquement un mécanisme de communication entre machines (*M2M*), de requêtes entre serveurs, même si DataCite propose une méthode 'manuelle' pour des navigateurs web habituels.
- A travers du site web DataCite on peut accéder aux deux services principaux actuels ayant comme objectif de répertorier les archives (*repositories*) de données existantes : *Databib*<sup>7</sup>, une initiative d'origine américaine, et *re3data.org*<sup>8</sup>, projet financé par l'Allemagne. Ces deux services sont en train de fusionner sous l'égide de DataCite.
- Bien entendu, DataCite propose une aide exhaustive en ligne destinée à guider les utilisateurs, notamment dans l'établissement de leurs fichiers XML de métadonnées, même si généralement cette aide ne remplace pas le contact direct à travers les institutions-membres comme l'Inist. Ayant fait le choix d'accepter l'enregistrement d'objets numériques très divers (choix qui ne fait pas forcément l'unanimité au sein du consortium), DataCite doit en effet fortement s'appuyer sur ses membres pour l'assistance au cas par cas dans l'utilisation de ses services.

Le développement de nouveaux services figure comme objectif dans le plan stratégique 2013-2016 de DataCite. Néanmoins cet objectif reste à préciser, notamment dans le contexte de la multiplication des initiatives internationales autour des données scientifiques (la principale étant aujourd'hui la *Research*

---

<sup>6</sup> <http://thomsonreuters.com/data-citation-index>

<sup>7</sup> [www.databib.org](http://www.databib.org)

<sup>8</sup> [www.re3data.org](http://www.re3data.org)

Data Alliance<sup>9</sup>, avec laquelle DataCite a signé un accord) et la concurrence croissante avec des acteurs privés comme Thomson Reuters. Il nécessite sans doute de travailler davantage en partenariat et d'intensifier les collaborations existantes.

### DataCite resourceTypeGeneral

- Dataset
- Text
- Collection
- Event
- Audiovisual
- Image
- InteractiveResource
- Model
- PhysicalObject
- Service
- Software
- Sound
- Workflow
- Other

Comme traduit par la liste de 'types généraux de ressources' de son schéma de métadonnées, DataCite se considère comme étant au service de données dans un sens large, conformément à une définition des U.S. National Institutes of Health (NIH): "*Recorded information, regardless of the form or medium on which it may be recorded including writings, films, sound recordings, pictorial reproductions, drawings, designs, or other graphic representations, procedural manuals, forms, diagrams, workflow, charts, equipment descriptions, data files, data processing or computer programs (software), statistical records, and other research data.*"

### Identifiants pérennes

L'objectif de valoriser les données scientifiques en les rendant accessibles et citables implique l'utilisation d'un système d'identification pérenne de ces données sur les réseaux. De façon générale, le choix d'identifiant dépend du type de ressources et des besoins des producteurs de ces ressources. Ciblant principalement les données de la recherche (néanmoins définies dans un sens large) et sachant que la pérennité est davantage une problématique de gouvernance d'un tel système qu'une question technique, DataCite a fait le choix des DOI (*Digital Object Identifier*), devenant ainsi une agence d'enregistrement de DOI (ou bien 'agence DOI', par abus de langage), parmi la dizaine d'agences existantes à ce jour, dont CrossRef pour les revues scientifiques. Le choix s'est basé sur plusieurs critères :

- La connaissance et adoption du système des DOI, norme ISO depuis 2012, par les auteurs et communautés scientifiques à cause de son utilisation dans le domaine de l'édition, également la claire préférence des éditeurs pour les DOI dans l'optique de lier publications et données sous-jacentes. On peut ajouter ici que des plateformes populaires telles que Figshare<sup>10</sup>, Zenodo<sup>11</sup> ou ResearchGate<sup>12</sup>, fournissent automatiquement des DOI aux documents déposés.
- Le fait que l'attribution de DOI aux données, et autres objets numériques issus de la recherche, place ces objets à un niveau équivalent aux articles publiés et contribue ainsi à leur acceptation comme contribution scientifique à part entière. Ceci va évidemment dans le sens d'une meilleure reconnaissance des chercheurs qui produisent, gèrent, publient des données.
- La demande des utilisateurs en termes d'identifiants, qui porte majoritairement sur les DOI.

<sup>9</sup> [www.rd-alliance.org](http://www.rd-alliance.org)

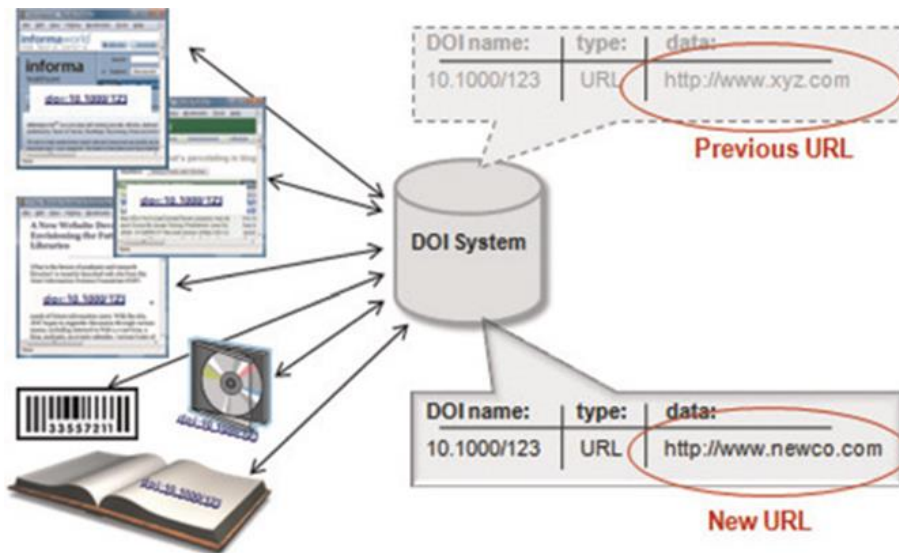
<sup>10</sup> [www.figshare.com](http://www.figshare.com)

<sup>11</sup> [www.zenodo.org](http://www.zenodo.org)

<sup>12</sup> [www.researchgate.net](http://www.researchgate.net)

- La présence de cette « infrastructure sociale », constituée par l'*International DOI Foundation (IDF)*, à travers son réseau d'agences et leurs membres – dans le cas DataCite : la trentaine de bibliothèques et centres d'information à travers le monde, agréés d'attribuer des DOI. Cette attribution est soumise à la signature d'un contrat avec les organismes demandeurs, précisant les obligations de chaque partie, garant supplémentaire de la pérennité des *landing pages* et de l'accès aux données.
- L'obligation d'associer des métadonnées aux DOI, facilitant la recherche et ainsi l'accès aux données. Elles peuvent être mises à jour à tout moment. De façon générale, la valeur stratégique des métadonnées augmente (ce qui contraste avec la tendance de les rendre librement accessible, que cette « information brute elle-même ne vaut quasiment plus rien dans la mesure où elle est disponible partout sur internet » [5]), surtout avec les promesses des technologies sémantiques et des *linked open data*. Le consortium DataCite a créé son propre schéma de métadonnées, basé sur un cadre (*metadata kernel*) défini par l'IDF et suffisamment général pour couvrir un large éventail d'objets numériques [6].

Il est à noter que le système des DOI est conçu pour être interopérable avec d'autres schémas d'identifiants et métadonnées. Il constitue un cas particulier du système des Handle<sup>13</sup>, eux-mêmes des identifiants pérennes. Les DOI se basent sur les Handle pour les composantes 'nommage' et 'résolution', on peut dire que techniquement, Handle et DOI, c'est la même chose [7]. Une certaine concurrence entre les deux, longtemps entretenue par le fait que les premiers sont gratuits et les seconds payants, sans parler de la connotation « édition commerciale » pour les seconds, a cédé la place à une coexistence, voir convergence, des deux systèmes, traduit aussi par le rapprochement entre DataCite et le consortium EPIC (*European Persistent Identifier Consortium*)<sup>14</sup>, qui préconise les Handle. Certains membres de DataCite utilisent les deux, suivant les cas et en fonction des granularités des objets et sous-objets.



Un identifiant pérenne (*Persistent identifier* ou *PID*) est un identifiant qui est assigné à un objet de façon unique et permanente. Il est disponible et gérable à long terme ; il ne changera pas si l'objet est renommé ou déplacé (changement de site, d'URL, d'entrepôt de données...).

<sup>13</sup> <http://www.handle.net>

<sup>14</sup> <http://www.pidconsortium.eu>

A noter aussi que ‘bonne’ traduction de ‘digital object identifier’ est plutôt ‘identifiant numérique d’objet’ (au lieu de ‘identifiant d’objet numérique’), car un DOI peut être attribué à une entité quelconque – physique, numérique ou abstraite – ‘physical object’, par exemple, fait partie des types de ressources du schéma de métadonnées DataCite.

Les identifiants pérennes ne se limitent pas aux objets. Des initiatives diverses ont été lancées ces dernières années visant à établir des systèmes d’identification d’institutions ou de personnes/auteurs (parmi lesquels le référentiel d’autorités français IdRef<sup>15</sup>). Dans ce domaine, on peut probablement parier sur une large adoption par le monde scientifique des identifiants ORCID (*Open Researcher and Contributor ID*)<sup>16</sup>, grâce à une notoriété déjà obtenue par le grand nombre d’organisations qui soutiennent ORCID - éditeurs, institutions de recherche, sociétés savantes, fondations. Ceci malgré quelques critiques récurrentes : dominance américaine, forte implication des éditeurs commerciaux, réduction des individus à un simple numéro...

DataCite et ORCID entretiennent des relations privilégiées. Dans le projet européen ODIN (*ORCID and DataCite Interoperability Network*)<sup>17</sup> ils ont développé, avec d’autres partenaires, des prototypes de services intégrant les deux types d’identifiants. Ce travail se poursuit dans le cadre d’Horizon 2020 avec le projet THOR (*Technical and Human Infrastructure for Open Research*). A noter que le «C» dans «ORCID» signifie ‘contributor’, preuve de la volonté de prise en compte de tous les contributeurs à la recherche scientifique, au-delà des auteurs de publications (par exemple les gestionnaires de données).

Les identifiants pérennes (*PID*’s) constituent un domaine de recherche sous plusieurs aspects, surtout dans une perspective de gestion automatisée de grandes masses de données, dont le volume et la complexité dépasseront les moyens de l’intervention humaine. On s’intéresse en particulier aux types d’information qu’ils doivent contenir pour être ‘actionnables’ par des machines [8].

Concernant la syntaxe des DOI, il y a une discussion sur le choix entre identifiants dits sémantiques, ou signifiants, et identifiants opaques, souvent générés par les machines. Généralement on préconise la deuxième solution, car le lien sémantique entre l’identifiant et la ressource peut être brisé et on peut rencontrer des problèmes de cohérence et de généricité [9].

## **Publication et citation de données**

Des initiatives internationales visant à promouvoir la publication et la citation des données ont vu le jour les dernières années, dans le but d’établir des lignes directrices et de créer les fondements d’une normalisation de ces pratiques. Une publication de l’OCDE de 2009 qui aborde largement cette problématique a d’ailleurs joué un rôle important dans ce contexte [10]. Il convient de souligner que ‘publication’ n’est pas identique à ‘mise en ligne’. Dans le premier cas les données ont fait l’objet d’un processus qui les a enrichies, qui a produit de la valeur ajoutée. Ceci peut être fait de plusieurs manières : curation, choix de formats standards et interopérables, archivage à long terme, attribution d’identifiant pérenne, choix de licence (de préférence *open*), dépôt dans une archive reconnue (de préférence une qui possède un label de qualité), *peer review*, voire publication dans un *data journal*.

---

<sup>15</sup> [www.idref.fr](http://www.idref.fr)

<sup>16</sup> [www.orcid.org](http://www.orcid.org)

<sup>17</sup> [www.odin-project.eu](http://www.odin-project.eu)

Le but d'une telle publication 'formelle' est de faciliter la citation, la vérification et la réutilisation des données. Avec la conséquence positive que la valorisation et reconnaissance du créateur des données se rapproche ainsi de celles de l'auteur d'une publication.

De façon générale, la publication conditionne le partage, et les enjeux du partage des données, on le sait, ne sont pas seulement scientifiques, mais aussi économiques et sociétaux.

Quant à la citation des données, les pratiques ne sont pas seulement encore peu fréquentes, mais aussi peu harmonisées. Des efforts concertés ont récemment pu être observés dans le but d'améliorer cette situation. A partir de travaux du *Digital Curation Centre*<sup>18</sup> britannique, d'un rapport [11] d'un groupe de travail de CODATA<sup>19</sup> et de principes établis par l'initiative FORCE11<sup>20</sup> sous forme d'un document nommé *Amsterdam Manifesto* [12], ainsi que de réflexions au sein de DataCite, un *Data Citation Synthesis Group* d'une quinzaine d'experts s'est établi, qui a produit en 2013 une « *Joint Declaration of Data Citation Principles* » [13], approuvée et soutenue par de nombreuses institutions à travers le monde et ainsi devenue une référence dans la matière. Ces principes tiennent compte de la nécessité qu'ils soient à la fois applicables par les humains et exploitables par les machines.

Ce qui est à souligner ici : la prise en compte de l'impact des données dans les nouvelles métriques alternatives, pour laquelle il y a déjà des exemples prometteurs [14], dépend aussi du degré de normalisation de la citation de ces données.

## The Noble Eight-Fold Path to Citing Data

1. Importance
2. Credit and attribution
3. Evidence
4. Unique Identification
5. Access
6. Persistence
7. Specificity and verifiability
8. Interoperability and flexibility

Source : S. Callaghan

Insistant sur la place essentielle d'une bonne pratique de la citation dans l'écosystème de la communication scientifique, la « *Joint Declaration of Data Citation Principles* » est constituée de huit principes directeurs, décrivant les objectifs, les fonctions et les caractéristiques de la citation des données.

### Perspectives

Après cinq années de développement enthousiaste, dans plusieurs directions, et la reconnaissance en tant que voix influente dans les discussions internationales autour des données de la recherche,

<sup>18</sup> <http://www.dcc.ac.uk>

<sup>19</sup> [www.codata.org](http://www.codata.org)

<sup>20</sup> [www.force11.org](http://www.force11.org)



DataCite ne pourra pas faire l'économie d'une re-précision de ses objectifs stratégiques. Si la phase initiale du consortium était caractérisée par une certaine fierté de la croissance rapide du nombre de 'clients' et de DOI attribués, aujourd'hui des questions sur les volumes et types de contenus à enregistrer se posent, ainsi que des questions sur la qualité – qualité des données et qualité des centres avec lesquels on travaille. DataCite s'est associé à des tentatives, par exemple par RDA et par ICSU-WDS<sup>21</sup>, de mise en place de certifications de réservoirs de données – comme le *Data Seal of Approval*<sup>22</sup> créé par DANS<sup>23</sup>.

Quant aux données elles-mêmes, quelles sont celles qui méritent d'avoir un 'label DOI' ? Ou accepte-t-on tout, sans restrictions ? Est-ce que tout objet est susceptible d'être cité dans une communication scientifique ?

Dans ce contexte, comment va DataCite se positionner à terme par rapport au *Data Citation Index* de Thomson Reuters (qui, lui, applique des critères de qualité assez stricts) ?

D'autres discussions concernent les types de contenus qu'on accepte. La décision de se baser sur une définition très large de 'données' entraîne souvent un jugement critique sur l'hétérogénéité des objets enregistrés. Et de nouveaux objets frappent à la porte de DataCite, pas seulement des objets numériques comme les nano-publications [15], mais également des ressources plutôt physiques, voire non matérielles : expériences, échantillons, stations de terrain, enregistreurs, campagnes d'observation...

Le besoin de mesurer l'utilité et l'impact d'objets scientifiques se généralise, au-delà de l'objet 'publication', notamment de la part des financeurs de la recherche. Pouvoir normaliser et faciliter la citation, pour évaluer l'utilisation d'une infrastructure, par exemple, est considéré comme prometteur dans ce contexte.

Un autre débat tourne autour des politiques 'open' au sein du consortium, à savoir autour du degré d'exigence de rendre les données librement accessibles. Malgré une position générale de DataCite en faveur de l'*open data*, l'accès aux données primaires à travers sa plateforme est dans beaucoup de cas (bien justifiés, en général) soumis à des périodes d'embargo, pouvant atteindre plusieurs années.

L'autonomie de ses membres est un principe du consortium, mais des réponses communes aux questions ci-dessus paraissent comme indispensables. DataCite doit donc se diriger vers une plus grande homogénéisation des politiques et pratiques existantes. Par exemple, la Commission européenne ne semble pas apprécier la diversité des modèles économiques parmi les membres européens.

Les réponses à donner demandent aussi des degrés d'expertise croissants au sein des institutions membres. Les personnes concernées sont majoritairement des bibliothécaires et autres professionnels de l'information. Dans le cadre des relations avec utilisateurs ils sont en train d'acquérir de nouvelles compétences, et DataCite contribue ainsi à la transition de leur rôle, vers celui des *data librarians* ou *data scientists*.

---

<sup>21</sup> [www.icsu-wds.org](http://www.icsu-wds.org)

<sup>22</sup> [www.datasealofapproval.org](http://www.datasealofapproval.org)

<sup>23</sup> [www.dans.knaw.nl/en](http://www.dans.knaw.nl/en)

## Bibliographie

- [1] HEY (T.), TANSLEY (S.) et TOLLE (K.) (Eds.), The fourth paradigm: data-intensive scientific discovery, Redmond, WA (Etats Unis), Microsoft Research, 2009
- [2] Commission Européenne (2013, 11 déc.), Guidelines on Data Management in Horizon 2020. Consulté le 17 janv. 2015. [http://ec.europa.eu/research/participants/data/ref/h2020/grants\\_manual/hi/oa\\_pilot/h2020-hi-oa-data-mgt\\_en.pdf](http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-data-mgt_en.pdf)
- [3] BRASE (J.), FARQUHAR (A.), GASTL (A.), GRUTTEMEIER (H.) et al., Approach for a Joint Global Registration Agency for Research Data, dans Information Services & Use (IOS Press) vol. 29, 2009, pp. 13-27
- [4] STM – International Association of Scientific, Technical and Medical Publishers (2007, 1er nov.), Brussels Declaration, sur le site STM. Consulté le 14 janv. 2015. [www.stm-assoc.org/public-affairs/resources/brussels-declaration](http://www.stm-assoc.org/public-affairs/resources/brussels-declaration)
- [5] HUGUENY (H.), La paradoxale valeur des métadonnées du livre, dans Livres Hebdo, 9 déc. 2014
- [6] DataCite (oct. 2014), DataCite Metadata Schema for the Publication and Citation of Research Data, version 3.1. Consulté le 12 janv. 2015. <http://dx.doi.org/10.5438/0010>
- [7] PASKIN (N.), Digital Object Identifier (DOI) System, dans BATES (M.J.) et MAACK (M.N.) (Eds.), Encyclopedia of Library and Information Sciences, Third Edition, Taylor & Francis, 2009
- [8] WEIGEL (T.), KINDERMANN (S.) et LAUTENSCHLAGER (M.), Actionable Persistent Identifier Collections, dans Data Science Journal, vol. 12, 2014, pp. 191-206. [www.jstage.jst.go.jp/article/dsj/12/0/12\\_12-058/pdf](http://www.jstage.jst.go.jp/article/dsj/12/0/12_12-058/pdf)
- [9] BERMÈS (E.), Des identifiants pérennes pour les ressources numériques, Actes des Journées Réseaux (JRES) 2007. Consulté le 17 janv. 2015. <http://2007.jres.org/planning/pdf/163.pdf>
- [10] GREEN (T.), We Need Publishing Standards for Datasets and Data Tables, OECD Publishing White Papers, 2009, version révisée fév.2010. Consulté le 15 janv. 2015. <http://dx.doi.org/10.1787/787355886123>
- [11] CODATA-ICSTI Task Group on Data Citation Standards and Practices (2013), Out of Cite, Out of Mind: The Current State of Practice, Policy, and Technology for the Citation of Data, dans Data Science Journal, vol. 12, 2013, pp. 1-75. <http://dx.doi.org/10.2481/dsj.OSOM13-043>
- [12] FORCE11 (mars 2013), The Amsterdam Manifesto on Data Citation Principles, sur le site de FORCE11. Consulté le 17 janv. 2015. [www.force11.org/AmsterdamManifesto](http://www.force11.org/AmsterdamManifesto)
- [13] Data Citation Synthesis Group (mars 2014), Joint Declaration of Data Citation Principles – FINAL, sur le site de FORCE11. Consulté le 17 janv. 2015. [www.force11.org/datacitation](http://www.force11.org/datacitation)
- [14] FENNER (M.), Altmetrics and Other Novel Measures for Scientific Impact, dans BARTLING (S.) et FRIESIKE (S.) (Eds.), Opening Science -The Evolving Guide on How the Internet is Changing Research, Collaboration and Scholarly Publishing , Springer, 2014, pp. 179-189

[http://dx.doi.org/10.1007/978-3-319-00026-8\\_12](http://dx.doi.org/10.1007/978-3-319-00026-8_12)

[15] THOMPSON (M.), MONS (B.), ROOS (M.), SCHULTES (E.), Data Publishing using Nanopublications, dans Tiny Transactions on Computer Science (Tiny ToCS), vol.1, 2012.

<http://tinytocs.org/vol1/papers/tinytocs-v1-thompson.pdf>

*Préprint d'un texte publié dans BROUDOUX (E.) & CHARTRON (G.) (Eds.) "Big data – Open data. Quelles valeurs? Quels enjeux?", ISBN 978-2-80-730031-6, De Boeck Supérieur, 2015, pp. 229-240*

*Preprint of an article published in BROUDOUX (E.) & CHARTRON (G.) (Eds.) "Big data – Open data. Quelles valeurs? Quels enjeux?", ISBN 978-2-80-730031-6, De Boeck Supérieur, 2015, pp. 229-240*