

Réunion sur le DOI avec l'INIST - 3 avril 2014

Personnes présentes

Personnes de l'INIST :

- Herbert Grüttemeier (herbert.gruttemeier@inist.fr)
- Marie-Christine Jacquemot-Perbal (marie-christine.jacquemot@inist.fr)
- Sylvie Leblanc (sylvie.leblanc@inist.fr)
- Mohamed Yahia (mohamed.yahia@inist.fr)

Les membres de la liste gestion de données datamanag@ujf-grenoble.fr dont / et :

Brice Boudevillain
Guillaume Nord
Jerôme Le Tanou
Guillaume Mella
Patrick Bernaud
Nadège Meunier
Catherine Pequegnat
Andrea Walpersdorf
Grégory Teste
Véronique Chaffard
.....

DataCite

DataCite (<https://www.datacite.org>) est une agence internationale d'enregistrement de DOI. L'INIST est un membre de DataCite et peut donc attribuer des DOI (<http://www.inist.fr/?Attribution-de-DOI&lang=fr>). Les DOI sont basés sur le système Handle. DataCite garantit la pérennité des identifiants. Les DOI sont surtout connus dans le domaine des publications. DataCite se concentre sur les données de la recherche, au sens large (peut inclure la littérature grise, du multimédia...). Les objets qui nous intéressent ici sont les bases de données. La propriété intellectuelle n'est pas affectée par l'attribution d'un DOI. Les bases de données sont protégées par le droit *sui generis* (droit spécifique au producteur de base de données). La principale règle de l'utilisateur est de toujours maintenir un URL stable qui pointe un identifiant. Le coût est de 180 € par an via l'INIST, quel que soit le nombre de DOI attribués.

Obligations du contractant

- Mettre en libre accès une page avec les métadonnées et les conditions d'accès à la base de données : libre accès, embargo, restriction d'accès... Dans le cas d'un embargo préciser la date à laquelle l'embargo prend fin. Encouragement de DataCite à rendre les données publique (à terme).
- S'assurer que l'URL derrière chaque DOI est accessible et veiller à ce que l'URL reste stable.
- Si l'objet n'est plus accessible (problème de conversion de format...), cela doit être affiché sur la landing page (page présentant de l'information sur la base de données).
- L'URL doit être associée de façon unique à l'objet pour lequel on attribue le DOI, pas à une liste d'objets.
- S'assurer que chaque DOI a un suffixe unique.

DataCite a introduit un contrôle automatique, s'il y a un problème le contractant sera contacté.

Métadonnées

Voir http://schema.datacite.org/meta/kernel-3/doc/DataCite-MetadataKernel_v3.0.pdf, paragraphe 2.1.

Les métadonnées doivent être librement accessibles, dans le cadre d'une licence type Creative Commons CC0 (<http://creativecommons.org>, [http://fr.wikipedia.org/wiki/Licence CC0](http://fr.wikipedia.org/wiki/Licence_CC0)) (mais pas nécessaire de l'indiquer explicitement, cela peut même poser des problèmes). DataCite offre toutefois la possibilité pour le contractant de ne pas afficher les métadonnées, mais cela doit rester un cas exceptionnel, et va dans le sens contraire de l'objectif recherché. Les bases de données peuvent quant à elles avoir une restriction d'accès ou un embargo (mais il y a un encouragement de DataCite à les rendre publiques à terme).

5 métadonnées obligatoires :

- le DOI
- le créateur / entité
- la date de création de la base ou de mise à disposition de celle-ci
- le titre
- le publisher.

Une fois le DOI attribué, elles ne peuvent pas être modifiées.

Ces 5 métadonnées obligatoires sont celles qui vont apparaître dans la citation de la base de données.

La date peut-être la date de création de la base ou de sa mise à disposition, c'est la date qu'on veut voir apparaître dans la citation (pour plus de précision d'autres champs date sont présents dans les métadonnées optionnelles).

Le créateur n'est pas nécessairement une personne physique, il peut être un organisme ou une entité.

Métadonnées optionnelles : mots clés, résumé, rôle des différentes personnes, financements, géolocalisation... Remplir ces champs renforce la visibilité de la base de données.

Schéma propre (pas une norme ISO). C'est au gestionnaire de données de faire la conversion.

Pour remplir les métadonnées, il faut que le centre charge lui-même le fichier contenant les métadonnées (xml). Il y a une API. On peut donc automatiser ce processus. Si on le souhaite, il existe également un formulaire pour des cas ponctuels. Une aide peut être demandée à l'INIST. L'INIST fournit un accès à un espace test permettant d'enregistrer temporairement des DOI.

Bases de données évolutives

Si on modifie de manière substantielle la base de données, il faut lui attribuer un nouveau DOI. On peut mettre une granularité dans le DOI. Exemple : le programme ARGO produit des données mises à jour continuellement. Dans ce cas l'option choisie est un enregistrement de DOI mensuel, qui reflète l'état de la base de données à une certaine date. C'est au gestionnaire de la base de données de s'organiser. On peut par exemple mettre une numérotation dans le DOI.

C'est la responsabilité du gestionnaire de la base de données de donner un suffixe unique. DataCite peut dans de rares cas donner plusieurs préfixes (pour des catégories d'objets très différents par

exemple).

Statistiques d'utilisation

En ce qui concerne les statistiques sur l'utilisation du lien, peut être fournit la statistique de résolution : nombre de fois qu'on clique sur un DOI. On peut avoir ces statistiques par centre.

Thomson Reuters met actuellement en place un Data Citation Index, s'appuyant sur DataCite. Le Data Citation Index sera intégré au Web of Science, avec des statistiques d'utilisation beaucoup plus fines. Thomson Reuters scanne les articles pour voir si un jeu de données est cité.

Incitation à la citation des DOI dans les articles

En terme de normalisation, nous sommes encore dans les débuts. Changement de culture progressif.

En cas de fin de contrat

Les métadonnées et URL doivent rester accessibles. A priori, si modifications de celles-ci, contacter l'INIST. Modalités restent encore à définir.

Niveaux des bases de données

Pour un jeu de données, on peut avoir plusieurs niveaux. Les différents niveaux peuvent par exemple être le service d'observations, ou un niveau de granularité plus fin.

Parfois dans un article on peut avoir besoin de citer tout un service d'observations, dans lequel on extrait différentes informations ou pour donner le cadre général de l'étude. Actuellement, l'usage est de citer le site web. Ce niveau n'est pas exclu pour l'attribution d'un DOI : on peut attribuer un DOI à un service.

La granularité est le choix du gestionnaire de la base de données, elle doit être cohérente, se baser sur la possibilité de citation (si par exemple trop fin pour être cité, pas d'intérêt).

Bases de données pour lesquelles on n'est pas producteur

Cas où on enrichit une base de données déjà existante dont on n'est pas propriétaire.

Il faut trouver un accord avec le propriétaire de la base de données. Si quelqu'un utilise les données initiales, il peut être intéressant pour lui d'avoir un lien vers les données enrichies. Dans les métadonnées optionnelles, il est possible de mettre un lien vers la base de données originale. Question à se poser : est-ce suffisamment enrichi pour apporter quelque chose de nouveau.

Dans les métadonnées optionnelles, on peut indiquer une relation avec le portail où sont déjà présentes les données.

Contrat entre OSUG-INIST

Sera présenté à la Direction de l'OSUG .

Resif envisage un autre contrat avec l'INIST compte tenu de ses spécificités

Pour information : Réunion le 07/04 à l'INSU avec Françoise Genova , chargée de mission bases de données INSU et Nadège Meunier

Le 04/04/2014, Emilie Chirouze, stagiaire M2P IST