

Compte-rendu de l'Atelier sur les Identifiants Persistants (PIDs) organisé par le GBIF France à Paris (MNHN) le 27 juin 2012.

Formateurs :

Julien Cigare (BeBIF)
Mickael Graf (GBIF Suède)
Pere Roca Ristol (GBIF France)

Résumé :

Cet atelier avait pour but de présenter les PIDs utilisés en sciences de la vie, en particulier en biologie dans le domaine de la biodiversité.

Un **identifiant unique** est un code d'identification unique assigné à « quelque chose » de telle manière que ce « quelque chose » puisse être référencé sans ambiguïté.

Un **identifiant persistant** est un identifiant qui est assigné à un objet de façon permanente. Par exemple un numéro ISBN est assigné à un livre particulier, ce numéro est associé pour toujours à ce livre et aucun autre livre ne recevra le même numéro ISBN.

En biologie on traite d'information concernant différentes entités (spécimens, noms taxonomiques, publications, personnes, localités, caractères morphologiques, séquences ADN, ...). Il y a de nombreuses façons de référencer ces entités et, dès qu'ils sortent de leur « milieu d'origine », les identifiants utilisés peuvent devenir ambigus : il peut exister deux spécimens dont le numéro de catalogue est « BPBM37615 » (un poisson et un mollusque), il peut y avoir plusieurs articles écrits en 1998 par deux auteurs appelés « Baldwin » et « Smith ». Alors qu'un homme pourra assez facilement discerner le bon spécimen ou la bonne publication en fonction du contexte, cela sera plus difficile pour un ordinateur (dans le cadre des bases de données informatiques).

Les caractéristiques idéales d'un PID sont les suivantes :

- 1) universellement unique
- 2) immuable
- 3) création indépendante
- 4) opaque
- 5) avec des services associés.

Il existe différents types de PID couramment utilisés dans le domaine de la biodiversité et chacun a ses avantages et désavantages :

- 1) UUID : Universally Unique Identifier
- 2) DOI : Digital Object Identifier
- 3) URL : Uniform Resource Locator
- 4) PURL : Persistent URL
- 5) LSID : Life Science Identifier

Les LSID ont été présentés de façon plus approfondie, ce sont ceux qui sont recommandés par le GBIF.

Des conseils pour l'implémentation des PIDs en base de données ont été présentés ainsi que les mécanismes de résolution des PIDs et les services web pour la publication des données (partie plus technique).

Documentation :

Les informations transmises au cours de la formation se retrouvent dans différents documents mis à disposition sur :

N:\documentation\formation\identifiants_persistants_062012

GBIF (2011). A Beginner's Guide to Persistent Identifiers, version 1.0. Released on 9 February 2011. Authors : Kevin Richards, Richard White, Nicola Nicolson, Richard Pyle.

GBIF (2009). Adoption of Persistent Identifiers for Biodiversity Informatics, Recommendations of the GBIF LSID GUID Task Group. Released on 6 November 2009. Authors : Phil Cryer, Roger Hyam, Chuck Miller, Nicola Nicolson, Eamonn O Tuama, Rod Page, Jonathan Rees, Greg Riccardi, Kevin Richards, Richard White.

L'agenda et les présentations des intervenants sont également disponibles.