

Groupe de travail IRU

Guide sur les identificateurs de ressource uniques 2015

MEDDE/CGDD/DRI/MIG

Pourquoi ce guide sur les IRU ?

La gestion des identificateurs de ressource uniques (IRU) est un aspect fondamental de l'architecture de services Inspire

- Pour décrire les liens entre services et données au sein d'un catalogue de métadonnées
 - Leur référencement dans les métadonnées d'un service permet de connaître les données servies ou les services appelés
- Dans les capacités d'un service
 - Leur référencement permet d'opérer le service sur telle donnée demandée parmi l'ensemble des données servies par le service.
- Mais, dans les capacités des services un lien vers la métadonnée est aussi obligatoire.
- Or,... il est délicat lorsqu'on ajoute une nouvelle donnée à un service de connaître l'URL permettant d'accéder à la métadonnée, par exemple si la métadonnée est moissonnée.
- Même si des identifiants “non résolvable” sont possibles par rapport aux obligations INSPIRE, dans le Guide de saisie des métadonnées, la « Recommandation A » propose déjà que :
 1. l'identificateur de la ressource soit fourni sous la forme d'une URI (Uniforme Resource Identifier) au format http ;
 2. cet URI soit également une URL (adresse internet) permettant d'accéder à la ressource (ou à ses métadonnées).

Intérêt de rendre l'IRU résolvable ?

- Le service de recherche (d'un catalogue qui détient la métadonnée d'origine ou par moissonnage) permet de retrouver la métadonnée à partir de l'IRU de la ressource.
- Un lien vers la métadonnée peut être basé sur l'adresse de ce service, mais cela pose la question de la pérennité de ce lien (évolution du catalogue source, organisation des moissonnages différents, etc.).
- Une nouvelle Recommandation : *L'identifiant de la ressource (données ou service) devra pointer a minima vers la fiche de métadonnée en XML.*
- Mais, il est également possible de pointer, en plus, vers différentes ressources (version html, etc.)

Comment “Résoudre” l'IRU ?

- identifiants résolubles nativement : HTTP URIs ou URL
- identifiants à rendre résolubles : le résolveur peut être mis en place à différents niveaux :
 - Dans une infrastructure de données géographiques “locale”, dans ce cas l'identifiant pointe vers la ressource en local
 - A un niveau plus haut (par exemple national), dans ce cas l'identifiant pointe sur la ressource à ce niveau-là (par exemple sur la fiche de métadonnée moissonnée du géocatalogue).

Quelles solutions techniques pour les IRU HTTP?

Construire des IRU de type HTTP peut se faire en s'appuyant sur des organisations et des services résolveurs variés:

- les DOI (Digital Object Identifier) géré par l' IDF (International DOI Foundation)
- les URN (Uniform Resource Name) géré par l' IANA (Internet Assigned Numbers Authority),
- ...
- via son propre outil de catalogage
 - Géosource, etc.
- via le Géocatalogue national.

Quelles critères de choix d'une solution ?

- **Unicité** : Inspire impose un identifiant unique. L'unicité doit donc pouvoir être assurée quelle que soit la technologie choisie.
- **Persistance** : L'identifiant doit être persistant c'est-à-dire être valable dans le temps. Sa durée de vie doit pouvoir égaler la durée de vie du jeu de données.
- **Résolvabilité** : Un identifiant doit permettre, directement ou indirectement d'accéder à une ou plusieurs adresses du réseau.
- **Fiabilité** : L'organisme qui gère l'identifiant doit pouvoir garantir que l'identifiant est toujours valide.
- **Opacité** : L'identifiant peut être porteur de sens ou non. La préférence pour l'un ou l'autre dépend de la politique de l'organisme.
- **Evolutivité** : Possibilité de proposer de nouvelles versions de l'IRU pour correspondre aux nouvelles versions de la ressource.
- **Indépendance** : La génération de l'identifiant peut-elle être faite par différents acteurs ou est-on dépendant d'un organisme particulier ?
- **Facilité de génération** : Comment peut être généré l'identifiant, en local ou en faisant appel à un service distant ?
- **Gestion interne/externe** : Est-il possible de gérer plusieurs natures d'identifiant : des identificateurs qui auraient un sens à l'intérieur de la plateforme et des identifiants qui auraient un sens en externe ?
- **Coût** : Quel est le coût de la génération et la gestion des identifiants ?

Un tableau de synthèse dans le guide

	Unicité	Persistance	Resolvabilité	Fiabilité	Opacité	Evolutivité	Indépendance	Gestion interne/externe	Coût
3.2.1 HTTP URIs ou URL	Unique par construction	Peut nécessiter un système de redirection pour assurer une pérennité en cas de changement de nom de domaine.	URL résolvable vers une adresse. HTTP URI peut également être résolvable moyennant redirection.	nécessite de vérifier le catalogue du producteur régulièrement	Généralement peu opaque mais une opacité est possible.	Forte : on peut la changer à tout moment mais il faut prévoir un système de versionnement.	forte sauf si utilisation d'un résolveur	Gestion interne pour la création du code et externe pour utilisation du résolveur	Cout de maintenance du nom de domaine à considérer
3.2.3 URN	Unique par construction	Nécessite une définition stricte des règles de nommage pour assurer la persistance			Permettent d'avoir des identificateurs qui ont du sens, décrivent une structure de données par exemple.				
3.2.4 DOI	Unique par construction puis attribution	100% - DOI apposé à un jeu de données public et pérenne	Résolvable. Un DOI pointe vers la "Landing Page" du dataset	Toujours valide et pérenne. L'URL de la Landing page peut être modifiée si la localisation de la ressource est	Identifiant de l'organisme pour prefixer chaque code	Une fois attribué, pas d'évolution possible. Il est par contre possible de modifier les métadonnées et l'URL de la Landing	Un déclarant construit ses DOI à l'aide d'un préfixe qui lui est attribué et d'un suffixe qu'il choisit et les active en les déclarant	Gestion interne pour la création du code et externe pour déclarer le DOI auprès de Datacite	"For members of the French Higher Education and Research community, this costs a 180€ annual

Annonce d'un résolveur national

- Le résolveur national s'appuiera sur le Géocatalogue (www.geocatalogue.fr)
- Ce service permettra de fournir aux producteurs un identifiant de la ressource (donnée ou service) et de résoudre cet identifiant vers la fiche de métadonnées correspondante dans le Géocatalogue.
 - Exemple :
<http://id.geocatalogue.fr/fr-120066022-jdd-8e34893b-9991-45be-8b27-e3e403b5dd4d>
- Pour que l'identifiant soit résolvable, la fiche de métadonnées devra être moissonnée dans le Géocatalogue.