



Guide sur les identificateurs de ressource uniques

Groupe de travail « IRU »
Version 0.9 – 2015

Remerciements

Nom	Organisme
CARTOIXIA Mathieu	Isogéo
CHARTIER Benjamin	GeoPicardie
CHATAIGNER Julie	MNHN
CHAUVEAU Chloé	GIPATGERI
COUDERCY Laurent	Onema
DAVID Benoît	CGDD/DRI/MIG
DESBOEUFS Jerome	SGMAP / Etalab
HOUBIE Frédéric	Geomatys
JACQUEMET Clément	CGDD/DRI/MIG
LEOBET Marc	CGDD/DRI/MIG
LORY Pascal	IGN
MEILLON Julien	Ifremer
PECRIAUX Gilles	IUT de Troyes (AUBE)
PHUNG Fabrice	DREAL Bretagne/COPREV
QUIMBERT Erwan	Ifremer
REBOUX Maël	Ville de Rennes / Rennes Métropole
RYCKELYNCK Guillaume	CIGAL
SATRA LE BRIS Catherine	Ifremer
TAFFOURAU Etienne	BRGM
VERGEZ Pierre	IGN
WATIER Loïc	GIPATGERI

Table des Matières

Contenu

1 Introduction	5
2 Définitions	5
2.1 Généralités	5
2.2 Identificateur ou identifiant ?	5
2.3 Quel identifiant est concerné ?	5
2.4 Comment est modélisé l'identifiant de ressource unique dans INSPIRE ?	6
2.5 Quelle organisation nationale ?	7
2.6 Identifiant résolvable	7
2.6.1 Comment "Résoudre" l'identifiant	8
2.6.2 Qui met en place le "resolver" ?	8
2.6.3 Quel est l'intérêt d'un identifiant résolvable ?	8
3 Quelles sont les bonnes pratiques communes à suivre ?	9
3.1 unicité	9
3.2 identifiants multiples	9
3.3 résolvable vers quoi ?	10
4 Résolveur national	11
5 Quelle technologie choisir ?	12
5.1 Sur quels critères ?	12
5.2 Quelles solutions existent ?	13
5.2.1 identifiant résolubles nativement	13
5.2.2 identifiants à rendre résolubles	13
5.2.3 Technologies de résolveur	16
5.3 Laquelle choisir ?	1
6 Comment migrer / mettre en place une nouvelle politique d'identifiant ?	1
6.1 Migration	1
6.2 Définir une politique d'identification des ressources	1
8 Annexes	4
8.1 FAQ INSPIRE	4
8.1.1 Pourquoi INSPIRE recommande-t-elle l'utilisation d'URIs dans le schéma http ?	4

8.1.2 Quel est le bénéfice des URIs?	4
8.1.3 Comment les différents identifiants sont-ils encodés dans les données ?	4
8.1.4 Quel est le rôle des URIs dans les services de téléchargement?	5
8.1.5 Au-delà des objets spatiaux, quels autres types sont identifiés en utilisant les URIs ? 5	
8.1.6 Y a-t-il des exemples d'URIs INSPIRE que je puisse étudier?.....	5
8.1.7 Qui alloue les URIs ?	5
8.1.8 INSPIRE va-t-il organiser un dispositif de distribution d'URI ?	5
8.1.9 Dans mon état membre, nous n'avons pas de mécanisme d'URI défini au niveau national - comment démarrer ?	5
8.1.10 Puis-je utiliser des identifiants thématiques pour les objets spatiaux ?	6
8.1.11 Les URIs doivent-elles comporter des métadonnées ?.....	6
8.1.12 Puis-je versionner les URIs?	6
8.1.13 Puis-je changer la structure des URI toutes les x années si j'en ai le besoin ??	6
8.1.14 L'utilisation d'URIs signifie-t-elle que je ne peux pas protéger mes données des téléchargements non autorisés?	6
8.1.15 Qu'obtiens-je en suivant une URI?	6
8.1.16 Comment les utilisateurs utilisent-ils mes données avec les URI?.....	6
8.1.17 Le service de découverte INSPIRE est conçu pour localiser les séries de données - comment un utilisateur trouve-t-il un objet particulier en utilisant les URI?	7
8.2 Checkliste ARE3NA.....	7
8.2.1 Gouvernance	7
8.2.2 Opérations	8
8.3.3 Architecture.....	8

1 Introduction

La gestion des identificateurs de ressources uniques (IRU) est un aspect fondamental de l'architecture de services INSPIRE car ces identificateurs font le lien entre les métadonnées de données et les métadonnées de services, et également entre les capacités des services et les ressources servies par le service.

Certaines préconisations ont déjà été faites au sein des guides INSPIRE du CNIG sur la structuration des IRU, mais les avancées du groupe de travail du CNIG sur les services ont montré qu'il était nécessaire de préciser la définition et la gestion des IRU afin d'obtenir une infrastructure pleinement opérationnelle. De plus, ces travaux ont permis de construire progressivement un consensus sur l'utilisation à terme d'IRU résolubles permettant d'accéder directement à la fiche de métadonnées de la ressource. Or la possibilité d'IRU de type http, qui est une des recommandations du guide CNIG sur les métadonnées, n'est actuellement quasiment pas suivie.

Ce guide se propose donc d'étudier les différentes possibilités de mise en place d'IRU pour aboutir à une organisation nationale cohérente.

2 Définitions

2.1 Généralités

Un identifiant permet d'identifier, de manière unique, un élément. Dans le cadre de ce document, il permet d'identifier une ressource (un jeu de données ou un service). Le partage des données et l'interaction qui se fait de plus en plus entre infrastructure de données rend nécessaire d'avoir une plus grande rigueur et une meilleure coordination à l'attribution de ces identifiants.

2.2 Identificateur ou identifiant ?

La terminologie Inspire utilise le plus souvent le terme "Identificateur" alors que le langage courant privilégie le terme "Identifiant". Les deux termes sont équivalents au niveau du sens. Dans le reste du document nous utiliserons le terme "identifiant" afin de simplifier la lecture et la compréhension du document.

2.3 Quel identifiant est concerné ?

Ce guide traite uniquement des identifiants de ressource. Une ressource dans ce cadre peut désigner un ensemble de données ou un service. Les autres types d'identifiant, les identifiants d'objet par exemple, ne sont pas concernés par ces recommandations.



Cet identifiant est à distinguer de l'élément *fileidentifieur*, qui identifie la fiche de métadonnée.

2.4 Comment est modélisé l'identifiant de ressource unique dans INSPIRE ?

EXIGENCE INSPIRE :

L'identifiant identifie la ressource de manière unique.

- Cet élément répétable est obligatoire.
- Il est constitué d'un {code} obligatoire (chaîne de caractères libre), éventuellement associé à un {espace de noms} (chaîne de caractères libre).

Au niveau modélisation, la structuration pour l'IRU peut être soit sous forme d'un *MD_Identifier*, soit d'un *RS_Identifier*, qui est une sous-classe de *MD_Identifier* qu'elle précise en rajoutant deux champs (codespace et version).

MD_Identifier

1	authority	0..1
2	code	1

RS_Identifier

1	authority	0..1
2	code	1
3	codeSpace	0..1
4	version	0..1

Mais, pour l'ISO, le *RS_Identifier* est conçu à l'origine pour la classe *MD_ReferenceSystem*, sachant que la classe *MD_ReferenceSystem* décrit le système de référence (spatial et temporel) utilisé pour le jeu de données.

Néanmoins, c'est un type qui peut également être utilisé à la place de *MD_Identifier* pour décrire l'identifiant de la ressource dans *Data_Identification*. INSPIRE indique que l'on peut utiliser *RS_Identifier* à la place de *MD_Identifier*. Il est donc nécessaire que nos outils puissent gérer les deux types de structures pour l'identifiant de nos ressources (*gmd:identifieur*).

Mais, comme le rappelle le [guide de saisie des métadonnées du CNIG](#) de décembre 2013, **la seule obligation finale sur laquelle on peut s'appuyer de manière sûre pour notre démarche est celle d'avoir un champ "code"**, qui est du type "Chaîne de caractères".

RECOMMANDATION AUX EDITEURS :

Il est recommandé aux éditeurs de logiciels de gestion de métadonnées de savoir supporter les deux implémentations possibles.

2.5 Quelle organisation nationale ?

La gouvernance des identifiants peut se faire de différentes façons :

- Scénario 0 : Chacun choisit sa propre politique d'identifiant de ressource.

Avantages : Aucune contrainte ne s'applique et les solutions existantes peuvent être conservées.

Inconvénient : Des incohérences entre les identifiants des différents organismes risquent de survenir

- Scénario 1 : Une seule et unique politique de gestion des identifiants est proposée au niveau national.

Avantages : Les identifiants sont tous harmonisés au niveau national. Aucune incohérence ne peut survenir.

Inconvénient : Certains organismes seront contraints d'adopter un nouveau système de gestion des identifiants même s'ils ont déjà un système existant satisfaisant.

- Scénario 2 : Chacun choisit sa propre politique d'identifiant de ressources tout en respectant les bonnes pratiques définies au niveau national.

Avantages : Les organismes ayant déjà fait le choix d'un système de gestion d'identifiants peuvent le conserver moyennant le respect de bonnes pratiques.

Inconvénient : Aucune solution clé en main n'est offerte aux organismes n'ayant pas fait le choix de politique d'identifiants.

Le scénario 2 est retenu. Les bonnes pratiques à respecter dans ce cadre sont décrites dans la partie 3 *Quelles sont les bonnes pratiques communes à suivre ?*.

2.6 Identifiant résolvable

Il est possible, grâce à l'identifiant, d'accéder à la ressource ou un descriptif de la ressource. On appelle cette action "résoudre" ou "déréférencer" l'identifiant.

2.6.1 Comment “Résoudre” l’identifiant

Le lien vers la ressource peut se faire de deux manières :

- De manière directe : Si l’identifiant se présente sous la forme d’HTTP URI il suffit de ce lien pour accéder à la ressource
- De manière indirecte : Un outil appelé “Resolver” permet de transformer l’identifiant en un lien pour accéder à la ressource ou une description de la ressource.

Dans les deux cas plusieurs conditions doivent être satisfaites :

- La ressource ou le descriptif de la ressource doit être accessible.
- L’URL doit être connue du resolveur ou stable dans le temps.

2.6.2 Qui met en place le “resolver” ?

Le resolveur 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 (par exemple la fiche de métadonnée qui est présente sur le catalogue local)
- A un niveau plus haut (par exemple régional ou national), dans ce cas l’identifiant pointe sur la ressource à ce niveau-là (par exemple sur la fiche de métadonnée moissonnée du géocatalogue).

2.6.3 Quel est l’intérêt d’un identifiant résolvable ?

INSPIRE propose une architecture basée sur un ensemble de services. Ces services sont fortement liés entre eux notamment par leur cas d’utilisation : un utilisateur recherche une donnée sur le service de recherche, puis y accède via le service de visualisation ou de téléchargement. Il est donc nécessaire de faire un lien entre tous ces éléments.

Ce lien peut être basé sur les adresses des services mais cela pose la question de la pérennité de ce lien : les adresses des services peuvent changer et parfois de façon non maîtrisée par le producteur de données. Par exemple, dans les capacités du service de visualisation figure un lien vers la métadonnée de données. Ce lien peut se faire via une requête sur le catalogue de données mais cette requête n’est pas forcément stable dans le temps. Une adresse qui redirige vers le catalogue peut donc être proposée. Un changement dans l’adresse du catalogue ne nécessitera alors pas de changer toutes les requêtes renseignées dans les capacités de service mais uniquement le mécanisme de redirection.

La gestion des identifiants uniques de ressource (IRU) est un aspect fondamental dans la démarche INSPIRE, car c’est sur cette gestion des identifiants que sont bâties les associations des métadonnées et services.

En effet, l’identifiant unique de ressource va devoir être utilisé dans les métadonnées de service (MDS), via "*operatesOn*", et dans les Capacités (pour répondre au besoin d’identification des couches)

Or, il est délicat lorsqu'on ajoute une nouvelle donnée à un service de connaître l'URL permettant d'accéder à la métadonnée. En fonction du processus qui a mené à obtenir la donnée, la métadonnée a pu être obtenue par transfert sans que l'adresse soit connue.

Il est donc primordial dans ce cas que le localisateur de la ressource de la métadonnée de donnée renseigne bien une URL vers la métadonnée. Pour garantir cette possibilité, il est nécessaire de lever toute ambiguïté sur la gestion des IRU pour permettre d'obtenir ces URL.



Même si des identifiant “non résolvable” sont possibles par rapport aux obligations INSPIRE, le [guide de saisie des métadonnées du CNIG](#) préconise la solution d’avoir un champ “code” sous forme d’une IRU au format http, mais laisse le choix pour que cette URL retourne soit la ressource elle-même, soit ses métadonnées.

Exemple : http://www.ign.fr/bloc_identifiant_la_donnee

3 Quelles sont les bonnes pratiques communes à suivre ?

3.1 unicité

Le producteur de la ressource devra pouvoir garantir l’unicité de l’identifiant de la ressource qu’il publie.

Un identifiant ne peut être associé qu’à une et une seule ressource. En conséquence, les identifiants ne doivent pas être réutilisés même si la ressource a cessé d’exister.

En fonction de la technologie retenue, cette unicité peut se faire de façon statistique ou par construction.

Pour “pérenniser” cette unicité, il est important lors du moissonnage et de la duplication des fiches de métadonnées de garantir la non-modification de ces identifiants.

Dans le cas de la duplication de la donnée, se référer au [guide de gestion des catalogues](#) du CNIG.

3.2 identifiants multiples

La métadonnée INSPIRE de la ressource offre la possibilité de renseigner plusieurs identifiants de ressources uniques.

Cette possibilité peut permettre de gérer des identifiants différents issus de contextes différents. Par exemple, un acteur français ayant une politique d’identifiant de ressource incompatible avec les recommandations de ce guide pourra fournir dans sa fiche de métadonnées deux identifiants de ressource : l’un conforme à sa politique actuelle et l’autre conforme à ce guide.

3.3 resolvable vers quoi ?

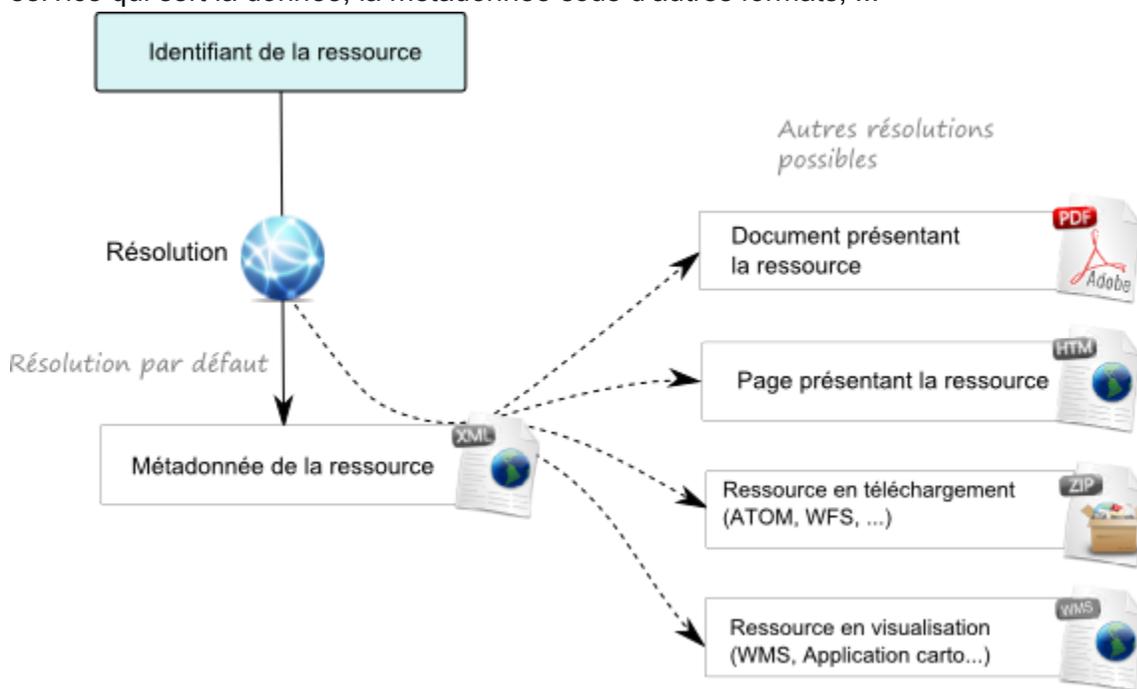
Une fois qu'il est décidé que l'identifiant pointe vers quelque chose, la question est de savoir vers quelle ressource et dans quel format.

Pour les raisons de lien entre services explicités ci-dessus, il est primordial qu'il y ait un lien vers la fiche de métadonnée en XML.

RECOMMANDATIONS NATIONALES :

L'identifiant de la ressource (jeu de données ou service) devra pointer a minima vers la fiche de métadonnée en XML.

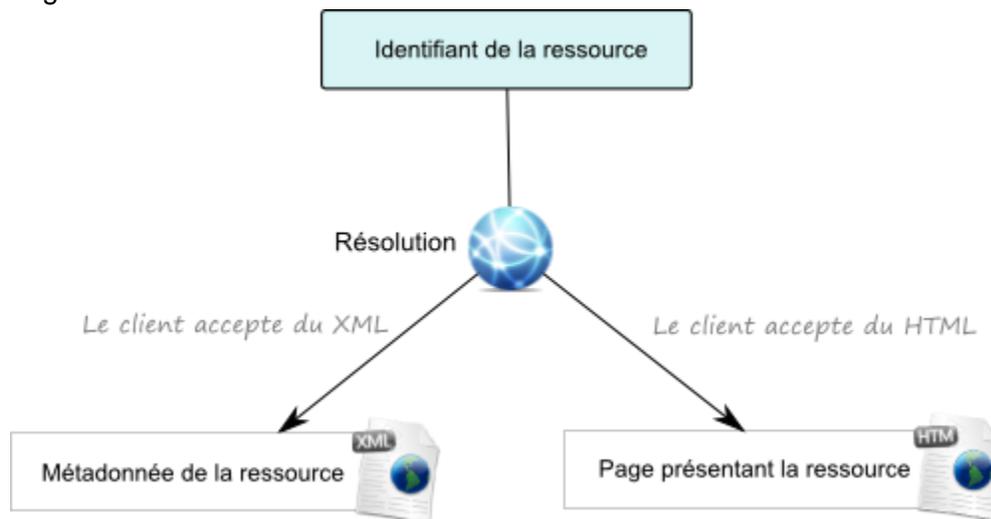
Il est également possible de pointer, en plus, vers différentes ressources, notamment les ressources renseignées sur la fiche de métadonnée : vers la données en téléchargement, un service qui sert la donnée, la métadonnée sous d'autres formats, ...



L'accès à une ressource en particulier peut être paramétré, par exemple via un postfix ajouté à l'URL.



La ressource vers laquelle pointe l'identifiant peut également être choisie via un mécanisme de négociation de contenu.



4 Résolveur national

Le résolveur national s'appuiera sur le Géocatalogue (www.geocatalogue.fr), qui est le point focal pour les métadonnées INSPIRE dans le cadre de la mise en œuvre de la directive en France.

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.

Le choix sera laissé aux producteurs de renseigner ou non l'identifiant dans la fiche de métadonnées. Les outils de catalogage (eg. Géosource) pourront permettre de faciliter la saisie de cet identifiant par des mécanismes automatiques.

Pour que l'identifiant soit résolvable, la fiche de métadonnées devra être moissonnée dans le Géocatalogue. Par défaut, l'identifiant sera résolue vers la fiche XML. Néanmoins, si cela est supporté par le client la fiche au format HTML pourra être renvoyée.

L'identifiant sera construit à partir de l'élément *fileidentifiant*, qui identifie la fiche de métadonnée. Généralement, le type d'identifiant utilisé par les outils de catalogage (eg. Géosource) est de type UUID (voir chapitre 5.2.2.4). A noter que dans le Géocatalogue le *fileidentifiant* est forcément unique et obligatoire.

Exemple : <http://id.geocatalogue.fr/fr-120066022-jdd-8e34893b-9991-45be-8b27-e3e403b5dd4d>



Pour plus d'informations : <http://www.geocatalogue.fr/#!/ResolveurURI> (à venir)

5 Quelle technologie choisir ?

Construire des IRU de type HTTP peut se faire en s'appuyant sur des organisations et des services résolveurs variés, comme par exemple les DOI (Digital Object Identifier) géré par l' IDF (International DOI Foundation) ou les URN (Uniform Resource Name) géré par l' IANA (Internet Assigned Numbers Authority), ou encore via son propre outil de catalogage ou via le Géocatalogue national.

5.1 Sur quels critères ?

Le choix d'un identifiant de ressource unique est intimement lié à la stratégie d'un organisme. Un certain nombre de contraintes propres à chacun (partenariats, stabilité de l'organisme, organisation distribuée ou non...) peuvent influencer sur la préférence de telles ou telles technologies.

- **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-elle dépendante 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 ?

5.2 Quelles solutions existent ?

5.2.1 identifiant résolubles nativement

5.2.1.1 HTTP URIs ou URL

Les HTTP URIs (Universal Resource Identifiers) sont des types d'identifiants très répandus, notamment dans le domaine du Web sémantique. Ils permettent à la fois d'identifier une ressource de manière unique et de la localiser.

Exemple : *http://data.ign.fr/data/IGNF_BDCARTOr_3-1_TOPONYMIE*

Cette solution a l'avantage de ne pas nécessiter que l'utilisateur connaisse l'existence du résolveur.

Pour pérenniser la localisation de la ressource, il est conseillé de définir un schéma d'HTTP URI indépendamment de la localisation réelle de la ressource puis de rediriger ces URI vers la ressource.



Note : Même si certaines ressources de l'organisation ne sont pas disponibles en ligne, les HTTP URI peuvent tout de même être utilisées. En effet il est désormais communément admis que les HTTP URIs ne pointent pas vers une donnée (ne soient pas des URLs).

5.2.2 identifiants à rendre résolubles

5.2.2.1 URN (Uniform Resource Name)

Les URN (RFC 2141) permettent d'identifier une ressource dans un espace de nommage particulier.

En s'inspirant de règles de structuration d'URN existantes (celles de l'OGC), on pourrait définir une structuration de type :

urn:geodata:{code pays}:{Identifiant de l'organisme}:{bloc identifiant de la donnée}

Ce qui donnerait en France la règle :

urn:geodata:fr:{code SIREN de l'organisme}:{bloc identifiant de la donnée}

Donnant par exemple : *urn:geodata:fr:243500139:510dee31-2037-e311-a543-0050568c385a*



Si on veut une vraie URN il faut la faire reconnaître par l'IANA (Internet Assigned Numbers Authority). Aujourd'hui le registre officiel des URN ne contient pas d'espace de nommage d'urn "geodata" : <http://www.iana.org/assignments/urn-namespaces/urn-namespaces.xhtml>

5.2.2.2 DOI (Digital Object Identifier)

L'IFREMER permet d'utiliser dans Sextant des DOI (par exemple 10.12770/1282383d-9b35-4eaa-a9d6-4b0c24c0cfc9) comme identifiant unique de série de données. Cette solution est particulièrement intéressante pour citer des jeux de données dans des articles scientifiques.

Les DOI sont gérés par l>IDF (International DOI Foundation): <http://www.doi.org/index.html>

Les DOI peuvent être encodés comme des URI HTTP, (comme par exemple : <http://dx.doi.org/10.1000/182>), mais un DOI ne renvoie pas forcément la fiche de MD correspondant à la donnée.

L'IFREMER a défini les règles suivantes (cf. la fiche de MD *Argo floats data and metadata from Global Data Assembly Centre (Argo GDAC)* <http://sextant.ifremer.fr/geonetwork/srv/fre/xml.metadata.get?uuid=1282383d-9b35-4eaa-a9d6-4b0c24c0cfc9>).

Un premier identifiant de la donnée reprend le fileidentifiant (*fileIdentifiant* : 1282383d-9b35-4eaa-a9d6-4b0c24c0cfc9) et un deuxième identifiant de la donnée reprend le DOI, lui même structuré avec le préfixe de gestion de ses DOI, suivi de la valeur du fileidentifiant, mais qui n'est pas sous forme HTTP.

```
<gmd:identifiant>
  <gmd:MD_Identifiant>
    <gmd:code>
      <gco:CharacterString>1282383d-9b35-4eaa-a9d6-
        4b0c24c0cfc9</gco:CharacterString>
    </gmd:code>
  </gmd:MD_Identifiant>
</gmd:identifiant>
<gmd:identifiant>
```

```
<gmd:MD_Identifier>
  <gmd:code>
    <gco:CharacterString>DOI:10.12770/1282383d-9b35-4eaa-a9d6-
    4b0c24c0cfc9</gco:CharacterString>
  </gmd:code>
</gmd:MD_Identifier>
</gmd:identifiant>
```

La résolution du DOI indiqué (via le site <http://www.doi.org/index.html>, ou en générant l'URL: <http://dx.doi.org/10.12770/1282383d-9b35-4eaa-a9d6-4b0c24c0cfc9>) , produit la redirection vers une page html <http://sextant.ifremer.fr/record/1282383d-9b35-4eaa-a9d6-4b0c24c0cfc9/> qui présente une partie des informations de la fiche de MD, avec des actions complémentaires.



Pour plus d'infos : <http://www.doi.org>

5.2.2.3 LSID

Les identifiants LSID (Life Science Identifier) sont similaires aux DOI mais sont davantage utilisés dans le domaine des sciences de la vie. Il prend la forme d'une URN particulière : **URN:LSID:<autorité d'identification>:<espace de nom>:<identification de l'objet>[:<version>]** et est associé à un système de résolution.

Les identifiants doivent avoir des services associés "LSID Resolution Service" (pour résoudre l'identifiant) OU un "LSID Resolution Discovery Service" (pour trouver le service qui permet de résoudre l'identifiant) OU un "LSID Assigning Service" (pour assigner des identifiants en fonction d'un certain nombre de propriétés).

Ils ne seront pas détaillés dans ce guide.

5.2.2.4 UUID

UUID (Universal Unique Identifier) est un type d'identifiant qui permet de générer des identifiants "statistiquement" uniques, leur création étant basée sur la génération d'une longue (128 bits) chaîne de caractères aléatoires. Il permet de générer des identifiants sans aucune centralisation. Beaucoup de langage ont des bibliothèque qui permettent de générer ce type d'identifiant facilement.

Pour être résolvable, cette chaîne de caractère peut être associée à un résolveur.

Les UUIDs peuvent être associés à un espace de nommage.



Pour plus d'infos : <http://tools.ietf.org/html/rfc4122.html>

5.2.3 Technologies de résolveur

Plusieurs technologies permettent de transformer une URI en URI résolvable.

5.2.3.1 PURL

Une URL est une adresse qui pointe vers une ressource du Web. Une PURL (Persistent Uniform Resource Locator) est une adresse qui pointe vers une autre adresse. Cette technologie permet, tout en s'affranchissant des modifications d'URL qui peuvent survenir sur l'infrastructure, de conserver une URL stable dans le temps. Cette PURL est souvent composée d'un identifiant unique stable et d'une partie fixe (qui permet de résoudre cet identifiant).



Pour plus d'infos : <https://purl.org/docs/help.html#overview>

5.2.3.2 ARK

ARK (Archival Resource Key) fonctionne de façon similaire, en faisant toujours précéder l'identifiant par ark: pour être à même de reconnaître avec certitude la partie liée à l'identifiant et celle liée au resolver.

Exemple : [http://monsite.fr/ark:/NAAN/Identifiant\[Option\]](http://monsite.fr/ark:/NAAN/Identifiant[Option])

avec le NAAN (Name Assigning Authority Number) et les options qui permettent d'accéder à différentes ressources (donnée, métadonnée, ...)



Pour plus d'infos : www.cdlib.org/services/uc3/arkspec.pdf

5.3 Laquelle choisir ?

Pour choisir une solution technique par rapport à une autre il est nécessaire de la confronter avec les besoins. Voici un tableau qui présente les avantages et inconvénients des différentes technologies au regard des critères énoncés plus haut.

	Unicité	Persistance	Resolvabilité	Fiabilité	Opacité	Evolutivité	Indépendance	Gestion interne/externe	Coût
<u>3.2.1 HTTP URIs ou URL</u>	Unique par construction	Peut nécessiter un système de redirection pour assurer une pérennité en cas de changement de nom de domaine.	URL resolvable vers une adresse. HTTP URI peut également être resolvable 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	Coût de maintenance du nom de domaine à considérer
<u>3.2.3 URN</u>	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.				
<u>3.2.4 DOI</u>	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 préfixer 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

				modifiée. Le DOI reste inchangée.		page.	à un organisme comme DataCite		<p>subscription fee for an unlimited number of DOIs.</p> <p>For all other users, this costs a 180€ annual subscription fee for up to 500 DOIs assigned. A descending cost scale is applied above 500 DOIs."</p>
3.2.6 <u>UUID</u>	Unique statistiquement				Totalement opaque. Peut être préfixé par une chaîne de caractère pour diminuer l'opacité.	Tout le système de résolubilité est basé sur le resolver, un changement de resolver peut donc permettre de faire évoluer les redirections.			

6 Comment migrer / mettre en place une nouvelle politique d'identifiant ?

6.1 Migration

Une nouvelle politique d'identifiant peut tout à fait se faire en cohabitation avec la politique précédente. Les ressources ont alors plusieurs identifiants (cf. 3.2 *identifiants multiples*)

Cette méthode permet de ne pas nuire aux applications en place qui utilisaient l'ancien identifiant.

6.2 Définir une politique d'identification des ressources

La définition d'une nouvelle politique d'identification d'une politique d'identification des ressources implique différents aspects :

- La gouvernance : Comment seront assignés les identifiants ?
- Les fonctionnalités : Quelles opérations sont nécessaires pour gérer les identifiants ?
- L'architecture : Comment est structuré un identifiant ? Quels services seront mis en place autour de cet identifiant (ex: resolver) ?
- Le financement : Quels sont les coûts et bénéfices de cette mise en place ?

Gouvernance :

Quel est l'objectif de cette politique ?

Il est nécessaire en premier lieu de lister les différents cas d'utilisation que les identifiants vont permettre de résoudre.

A quel niveau définir la politique d'identifiant ?

Une politique d'identifiant peut se faire à différents niveaux de l'organisation. Par exemple, l'attribution peut se faire au niveau d'une commune, d'une plateforme régionale, d'un organisme d'état. Il s'agit donc de définir à quel niveau est généré l'identifiant afin de pouvoir recenser les acteurs concernés.

Quelles sont les ressources concernées ?

Il est important de recenser les ressources auxquelles un identifiant sera associé : métadonnées, ensemble de données, service, objets, listes de code, ...

Quels sont les acteurs concernés ?

Toutes les parties prenantes doivent être intégrées directement ou indirectement au processus. Il est primordial de définir les rôles de chacun et de préciser qui est en charge d'attribuer un identifiant aux ressources. La liste des acteurs concernés doit pouvoir évoluer pour s'adapter aux changements.

Comment intégrer de nouveaux acteurs ?

De nouveaux groupes, qui gèrent les identifiants de nouvelles ressources ou de nouveaux types de ressource, doivent pouvoir être intégrés au besoin.

Quelle sera l'organisation ?

Dans le cas où la politique d'identifiants concernerait plusieurs acteurs, de manière centralisée (par exemple une plateforme régionale et des plateformes locales) il faut définir comment se fera l'organisation c'est-à-dire top down ou bottom-up. Cela influera sur la structure organisationnelle à mettre en place (quels acteurs devront être présents dans un groupe de coordination par exemple).

Quel est l'existant ?

Si des identifiants existent déjà, ils peuvent être partiellement conservés dans la nouvelle architecture. Cette possibilité doit être envisagée.

Qui informer ?

Une fois les nouveaux identifiants mis en place, il est primordial de communiquer sur la nouvelle politique d'identification

Fonctionnalités :

Comment et par qui seront assurées les fonctionnalités principales ?

Les fonctionnalités essentielles sont : l'assignation d'un identifiant, la validation, les redirections éventuelles et éventuellement la mise en correspondance avec les identifiants actuels.

Architecture :

Quelle sera la structure des identifiants ?

Cette structure doit tenir compte des pratiques dans le domaine, maintenir l'unicité et permettre éventuellement d'ajouter de nouveaux types d'identifiants par la suite.

Financement :

Quels sont les coûts et bénéfices ?

Le financement de la gestion des identifiants doit être évalué et clarifié afin qu'il n'y ait pas d'interruption dans la gestion.



Pour plus de détails consulter l'annexe 8.2 Checkliste ARE3NA.

8 Annexes

8.1 FAQ INSPIRE

Traduction de "[Implementation of Identifiers using URIs in INSPIRE](http://inspire.ec.europa.eu/index.cfm/pageid/5120)" <http://inspire.ec.europa.eu/index.cfm/pageid/5120> au 11/12/2014. Pas de garantie sur l'exactitude de cette traduction. Original ici : <http://cms.geobretagne.fr/content/faq-inspire-sur-les-iru-traduction-fr>

8.1.1 Pourquoi INSPIRE recommande-t-elle l'utilisation d'URIs dans le schéma http ?

A l'origine, INSPIRE n'était pas réduite au web, d'autres réseaux et mécanismes d'identification étaient considérés. Cependant, concernant l'implémentation, INSPIRE et le web ont évolué ensemble. Dans INSPIRE il n'y a plus d'attente concernant un autre réseau que le web. Sur le web, les URIs http sont devenus la principale façon de référencer les ressources du web. Pour conserver la connexion entre INSPIRE et le développement des infrastructures d'information associées, en particulier dans le contexte de l'e-government, l'utilisation des URIs http est considérée comme [une approche] prudente.

Les URIs http ne requièrent pas que l'URI puisse être déréférencé, même si supporter le déréférencement pour des ressources comme les objets spatiaux est une bonne pratique.

8.1.2 Quel est le bénéfice des URIs?

Les URI sont relativement faciles à déployer, et permettent de construire INSPIRE sur une infrastructure web puissante et scalable. Intégrer INSPIRE dans une architecture web conforte également la réutilisation des données spatiales issues d'INSPIRE par d'autres communautés car des technologies et des protocoles familiers sont utilisés.

Une conception appropriée des URI est importante pour contribuer à la durabilité des URIs. Voir le guide URI [D2.5 Annex H](#).

8.1.3 Comment les différents identifiants sont-ils encodés dans les données ?

En représentation GML:

- les URIs sont encodés dans la propriété *gml:identifiant* de chaque objet spatial ou de chaque série de données.
- les identifiants uniques d'objet sont encodés en tant qu'attribut *inspireId* de l'objet spatial, et utilisent un type de donnée *base:Identifiant* (note: puisque l'information *inspireId* est aussi encodée dans la propriété *gml:identifiant*, il est actuellement question d'abandonner la propriété *inspireId* dans l'encodage GML)

Les identifiants thématiques (identifiants de phénomènes du monde réel) sont encodés en tant qu'attributs standards de l'objet spatial, avec les modèles de données des annexes II/III; le nouveau type de donnée *base2:ThematicIdentifiers* a été introduit pour ces attributs.

Pour les autres formats, l'encodage des identifiants doit être spécifié au cas par cas en fonction des caractéristiques du format.

8.1.4 Quel est le rôle des URIs dans les services de téléchargement?

Les URIs sont utilisés pour identifier et référencer les séries de données spatiales. Déréférencer ces URIs aboutit au téléchargement des fichiers représentant les séries de données. Il peut y avoir plusieurs URIs pour un jeu de données, par exemple pour fournir le même jeu de données dans plusieurs systèmes de référence, ou en plusieurs langues.

Au-delà du téléchargement de séries de données, les services de téléchargement INSPIRE sont également recommandés pour donner accès aux objets spatiaux individuels, et dans ce cas le déréférencement de l'URI d'un objet spatial aboutit à une représentation de cet objet. Les détails de l'implémentation sont transparents pour l'utilisateur/le client, mais peuvent, par exemple, être implémentés par une redirection vers une requête *GetFeatureById* sur un service de téléchargement INSPIRE basé sur la spécification OGC Web Feature Service (WFS). Les URIs sont ainsi bien adaptées pour utiliser les données "in situ".

8.1.5 Au-delà des objets spatiaux, quels autres types sont identifiés en utilisant les URIs ?

Un certain nombre de ressources partagées sont identifiées comme éléments d'INSPIRE, et sont référencées depuis les séries de données concernés par INSPIRE. Ces ressources sont habituellement maintenues dans un registre et sont identifiées par des URIs, par exemple:

- Des listes de codes, <http://inspire.ec.europa.eu/codelist/countrycode/de>
- Des systèmes de référence, <http://www.opengis.net/def/crs/EPSG/0/4258>

8.1.6 Y a-t-il des exemples d'URIs INSPIRE que je puisse étudier?

Oui, plusieurs États Membres sont en train de développer des séries de données utilisant les URIs, par exemple:

- le Royaume Uni: voir "Designing URI Sets for Location" (NdT : voir aussi <http://data.gov.uk/resources/uris>)

8.1.7 Qui alloue les URIs ?

Une fois un domaine enregistré au niveau national (ou par une organisation internationale), idéalement par l'entité de coordination appropriée, le fournisseur de données est en mesure de poursuivre l'allocation d'URI et de publier ses données.

8.1.8 INSPIRE va-t-il organiser un dispositif de distribution d'URI ?

Non, le dispositif existe puisqu'INSPIRE réutilise des ressources web existantes. Idéalement, chaque État membre (et peut-être l'EU pour les jeux de données vraiment paneuropéens) va établir son propre domaine et gérer ses propres ressources. Ceci réduira les coûts, et fera de sorte que les États membres ne dépendront pas de ressources externes.

8.1.9 Dans mon État membre, nous n'avons pas de mécanisme d'URI défini au niveau national - comment démarrer ?

Entrez d'abord en contact avec l'entité de coordination. Regardez aussi ce que d'autres États membres ont fait, et profitez des opportunités pour aborder ce sujet dans les groupes INSPIRE paneuropéens.

8.1.10 Puis-je utiliser des identifiants thématiques pour les objets spatiaux ?

Les identifiants thématiques sont utilisés en attributs des objets spatiaux et doivent être publiés lorsque disponibles mais, puisqu'ils décrivent des phénomènes du monde réel, ils sont peu adaptés à l'utilisation en tant qu'identifiant d'objet spatial.

8.1.11 Les URIs doivent-elles comporter des métadonnées ?

Il est possible d'inclure quelques métadonnées dans un URI ou en tant que lien vers l'URI (lequel peut être déréférencé depuis l'URI).

8.1.12 Puis-je versionner les URIs?

Oui, un mécanisme de version peut être ajouté à l'URI, et est recommandé pour inclure les règles dans le schéma d'URI.

8.1.13 Puis-je changer la structure des URI toutes les x années si j'en ai le besoin ??

S'il est possible de le faire, la persistance des identifiants est critique pour l'infrastructure et le développement de futurs services d'information. Les utilisateurs vont construire leurs services sur ces identifiants, et attendent que cela continue de fonctionner.

Les URIs peuvent être "redirigés" lorsque déréférencés, et une façon de découpler le schéma d'URI des détails d'implémentation est d'utiliser les capacités standards d'HTTP. Cependant, comme les redirections ont un impact sur les performances, elles doivent être utilisées avec précaution.

Comme tout changement significatif induira des coûts pour les fournisseurs de données et les utilisateurs, il est recommandé de concevoir une approche simple, et de la maintenir dans la durée si possible.

8.1.14 L'utilisation d'URIs signifie-t-elle que je ne peux pas protéger mes données des téléchargements non autorisés?

Vous pouvez toujours protéger vos données en utilisant les mécanismes de sécurité standards du web - même si cela va clairement limiter les bénéfices plus larges de la publication en faveur de la réutilisation.

8.1.15 Qu'obtiens-je en suivant une URI?

Cela dépend de la façon dont le fournisseur de données a configuré ses données. Il est possible de publier des données en plusieurs formats, HTML, GML, JSON, RDF, Shapefiles, GeoTIFF, etc. HTTP supporte la négociation de contenu entre le client et le serveur. La plupart des fournisseurs de données vont probablement commencer en respectant les exigences minimum, et étendront celles-ci avec le temps en fonction de la demande.

8.1.16 Comment les utilisateurs utilisent-ils mes données avec les URI?

De toutes les façons - souvent en utilisant des API Web et des applications, puisque les URI sont d'abord "lisibles par des machines". Ceci renforce le besoin de maintenir une approche simple et solide pour la publication de données.

8.1.17 Le service de découverte INSPIRE est conçu pour localiser les séries de données - comment un utilisateur trouve-t-il un objet particulier en utilisant les URI?

Si l'utilisateur a déjà l'URI d'un objet spatial, il peut continuer à l'utiliser pour référencer l'objet, y compris dans d'autres contextes.

Si une série de données est proposée par un service de téléchargement qui supporte l'accès direct, un objet spatial spécifique peut être identifié en utilisant des requêtes, c'est à dire en utilisant la localisation, les identifiants thématiques ou d'autres propriétés. Une fois identifiée, l'URI peut être utilisé pour référencer l'objet spatial.

Pour d'autres séries de données, Il peut être nécessaire de télécharger et d'inspecter la série de données entière pour identifier l'URI d'un objet spatial.

8.2 Checkliste ARE3NA

Traduit de "ARE3NA (D.TD.04) – Vasilescu et al. (2015): Governance of Persistent Identifiers"

8.2.1 Gouvernance

- La gouvernance devrait être inclure toutes les parties prenantes (internes et externes), soit directement ou indirectement.
- La structure de gouvernance devrait être flexible et il devrait être possible d'ajouter des groupes et des groupes de travail de façon ponctuelle.
- La gouvernance devrait inclure des représentants habilités à prendre des décisions et de veiller à leur mise en œuvre, mais aussi l'expertise technique suffisante pour prendre les bonnes décisions. La gouvernance devrait inclure une représentation structurelle et clairement définir le processus de prise de décision.
- Il devrait y avoir une définition claire des rôles et des responsabilités de chacun.
- Il devrait y avoir une définition claire des objectifs et des moyens mis en oeuvre pour l'atteindre mais aussi de la manière dont l'environnement externe est pris en compte.
- Afin d'assurer le bon fonctionnement et l'efficacité, la structure de gouvernance devrait être conçue en tenant compte de la prise de décision mieux adapté au sein du domaine.
- La gouvernance devrait conduire la maturité des différentes communautés.
- Toutes les organisations qui gèrent des identifiants persistants devraient viser à réduire la complexité, la redondance et inefficacités en adoptant des processus standardisés, structures de données, et technologies.
- Toutes les organisations qui gèrent des identifiants persistants devraient avoir une politique écrite indiquant comment les identifiants sont créés et gérés, qui est responsable de la fiabilité, de la persistance et de la disponibilité. Cette politique devrait être mise à la disposition d'autres parties prenantes, de préférence sous une forme normalisée qui rendrait plus facile la comparaison ou la mise en correspondance entre les différentes approches.
- Une organisation doit indiquer clairement dans leur politique quels genres de ressources sont identifiés et qui a le droit d'attribuer des identificateurs.

En cas de ressources partagées, il devrait être clair et convenu qui est le responsable principal de ces ressources. Ce doit être documenté dans la politique. Il est inévitable que ces ressources partagées aient plusieurs identifiants. Cependant, ces ressources ne doivent pas avoir des identifiants en double (utiliser des règles de conception adéquates), et les liens entre les identifiants différents doivent être créés (en utilisant l'alias ou un ensemble de lien). Si possible l'affectation d'identifiant doit être automatisée.

- ❑ Toutes les organisations de gestion des identifiants devraient être préciser dans leur politique d'identifiants la politique de la confidentialité des données ou la réglementation applicable.
- ❑ Toutes les organisations de gestion des identifiants doivent mentionner dans leur politique l'ensemble des normes utilisées, les règles pour la définition, la création et l'entretien des identifiants.
- ❑ Toutes les organisations qui gèrent des identifiants devraient décrire dans le management de la qualité dans la politique des identifiants.
- ❑ Toutes les organisations qui gèrent des identifiants devraient préciser dans la politique des conditions d'accès nécessaires pour accéder aux ressources.
- ❑ Toutes les organisations qui gèrent des identifiants devraient préciser dans les détails de la politique sur la façon dont l'assurer la conservation à long terme des identifiants et des ressources.
- ❑ Une organisation devrait définir clairement le processus de gestion du changement avec les rôles et les responsabilités de chacun.
- ❑ Une organisation devrait rédiger un guide pour l'accord de niveaux de service (LSA).

8.2.2 Opérations

- ❑ Une organisation doit être clair au moins sur le processus suivant: enregistrement, validation, redirection, la préservation à long terme et si nécessaire la mise en correspondances.
- ❑ Une organisation doit suivre les directives mises en place par la Commission européenne concernant la définition des opérations de base et mettre en place des rôles clairs pour chacune de ces opérations. L'organisme gestionnaire des identifiants est responsable de la persistance, la fiabilité du système d'identifiants et la gestion du changement.

La Commission européenne devrait donner des orientations sur la définition des opérations et d'assurer l'unicité des domaines.

- ❑ Une organisation devrait clarifier qui finance la gestion des identifiants.
- ❑ Une organisation devrait réaliser une étude financière pour la gestion des identifiants

8.3.3 Architecture

- ❑ Pour les règles de conception, le propriétaire des données doit avoir la liberté de choisir le design de l'identifiant. Néanmoins, il est conseillé de s'inspirer des pratiques existantes dans la communauté. (Ex: The 10 rules for persistent URIs (Commission , 2012)) .

- ❑ Le choix entre les URI et DOI ou d'autres normes peut être fait par propriétaire des données. Le propriétaire ne devrait pas être exclu de la stratégie globale, stratégie globale pour PID ne devrait pas exclure l'un d'eux .
- ❑ Pour l'architecture les organisations devraient utiliser des normes pour la mise en œuvre des PID pour assurer l'interopérabilité.
- ❑ Il est conseillé de mettre les processus de gestion en place pour assurer que les identificateurs persistants et les données associées sont toujours synchronisés, de sorte que lorsque les enregistrements de base de données changent les identificateurs persistants pour ces enregistrements sont examinés rapidement.