

Revue de la littérature

– Identifiants pérennes (PID) –
– Linked Data –
– Données de la recherche –

Nicolas Prongué & Julien A. Raemy
Projet ICOPAD – Haute école de gestion de Genève
Octobre 2017

Sommaire

1	Introduction.....	4
2	Partie A : identifiants / PID.....	5
2.1	Problématique.....	5
2.2	Background.....	5
2.2.1	Concepts fondamentaux.....	6
2.2.1.1	URI et schémas d'URI.....	6
2.2.1.2	URL.....	7
2.2.1.3	Domain Name System (DNS).....	7
2.2.1.4	Pérennité.....	7
2.2.1.5	Actionnabilité.....	7
2.2.2	Fonctionnement d'un système de PID traditionnel (avec résolveur).....	7
2.2.2.1	Résolution multiple.....	8
2.2.3	Pérennité par les archives du web.....	8
2.3	Brève typologie des identifiants.....	9
2.3.1	Critères distinctifs.....	9
2.3.2	ARK.....	9
2.3.3	PURL.....	9
2.3.4	URN.....	10
2.3.4.1	NBN (National Bibliography Number).....	11
2.3.4.2	UUID (Universally Unique Identifier).....	12
2.3.5	Handle.....	13
2.3.5.1	DOI en détail.....	13
2.3.5.2	ePIC en détail.....	14
2.3.6	URI / Cool URI.....	15
2.3.6.1	ORCID.....	15
2.3.6.2	ISNI.....	15
2.3.7	Autres.....	16
2.3.8	Tableau synthétisant les divers identifiants existants.....	17
2.4	Services.....	18
2.4.1	Politique pour l'utilisateur.....	18
2.4.2	Design des identifiants.....	19
2.4.2.1	Signification versus opacité.....	19
2.4.2.2	Caractères sensibles dans les URL.....	19
2.4.2.3	Longueur.....	19
2.4.3	Implémentation.....	20
3	Partie B : Linked Data et données de la recherche.....	21
3.1	Introduction.....	21
3.2	Périmètre.....	21
3.3	Bénéfices potentiels des Linked Data pour les données de la recherche.....	21
3.4	OpenScience - eScience.....	22
3.5	Projets.....	22
3.5.1	Université de Bielefeld - CITEC.....	22
3.5.2	Archives publiques de Taïwan.....	23
3.5.3	Research Graph.....	24
3.6	Vocabulaires LOD.....	25
3.7	Les Linked Data en art et design.....	25

3.7.1 Projets LOD.....	25
3.7.2 IIIF.....	26
4 Bibliographie.....	27

Table des illustrations

Illustration 1: Identifiants et cycle de vie des données (23, p. 5).....	5
Illustration 2: Délimitation entre domaines privé, partagé et public (31).....	6
Illustration 3: Fonctionnement d'un système d'identifiants pérennes (8).....	8
Illustration 4: Fonctionnement de PURL.....	10
Illustration 5: Processus d'attribution d'un NBN en Suisse.....	12
Illustration 6: Politique au sein de ANDS (33).....	19
Illustration 7: L'infrastructure Open Science à l'Université de Bielefeld (1).....	23
Illustration 8: Exemple d'une ressource enrichie avec des Linked Data dans data.odw.tw.....	24

1 Introduction

Ce document donne un aperçu de l'état de l'art des réalisations concernant les identifiants pérennes et le Linked Data, appliquées aux données de la recherche :

- Partie A : application des identifiants pérennes aux données de la recherche
- Partie B : application du Linked Data aux données de la recherche

Cette revue de la littérature est établie dans le cadre du projet ICOPAD à la Haute école de gestion de Genève.

2 Partie A : identifiants / PID

2.1 Problématique

A l'heure actuelle, les institutions en Suisse, et notamment les bibliothèques, ont affaire à de plus en plus de données d'une grande hétérogénéité. Alors que les métadonnées collectées et gérées jusqu'à présent se concentraient essentiellement sur les documents, une quantité croissante de données sont aujourd'hui saisies sur le contexte de ces documents : les personnes (auteurs), les institutions (détenteurs), les concepts (thématiques), les lieux (thématiques ou lieux de publication), les périodes (de création), etc.

Outre cette nouvelle variété d'entités décrites, les bibliothèques recherchent également des moyens d'établir des connexions : internes entre ces données et externes avec les données d'autres institutions.

L'un des prérequis d'une telle interconnexion est la présence d'identifiants fiables et pérennes, sur lesquels peuvent ensuite s'appuyer des services basés sur le Linked Data (voir chapitre 3). De nombreux systèmes d'identification pérenne existent actuellement, se distinguant par diverses caractéristiques, telles que leur périmètre d'application ou leur base technique. Certains sont prévus plutôt pour les publications, tandis que d'autres se centrent sur les personnes, les lieux, etc.

Cette partie a pour but de donner un aperçu général des possibilités d'identification et des critères distinctifs, afin de permettre à une institution de définir plus aisément une politique pour ses identifiants pérennes.

2.2 Background

Le processus de recherche est hautement collaboratif et implique un échange fréquent de données (Illustration 1) :

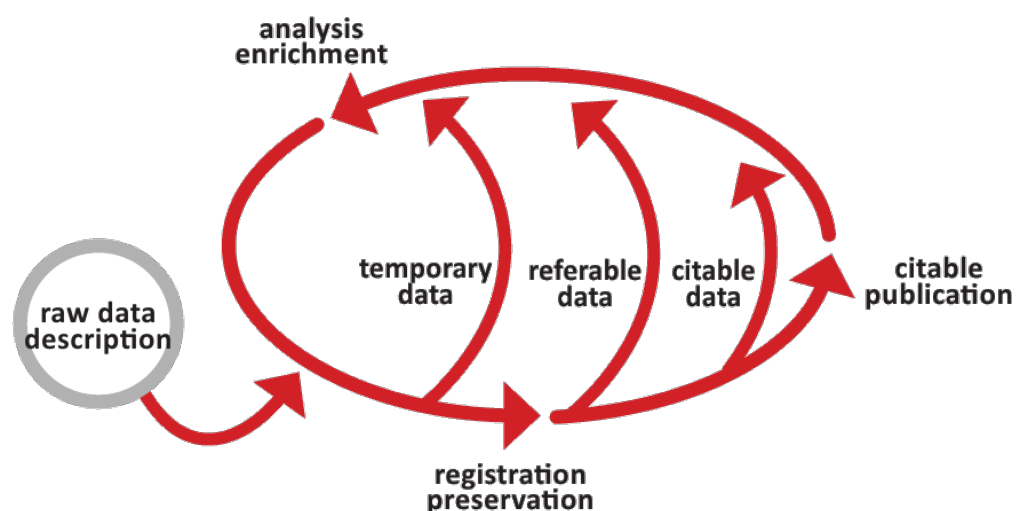


Illustration 1: Identifiants et cycle de vie des données (23, p. 5)

« a global context where resources are distributed across systems and standards, and the movement of data across disciplines and organizations is very intensive. » (8).

Les identifiants jouent un rôle essentiel lorsque les données passent du domaine privé au domaine public (Illustration 2). Elles acquièrent ainsi des caractéristiques nouvelles :

- accessibles par des utilisateurs ne les modifiant pas
- théoriquement stables, ou alors toute modification est documentée dans les métadonnées

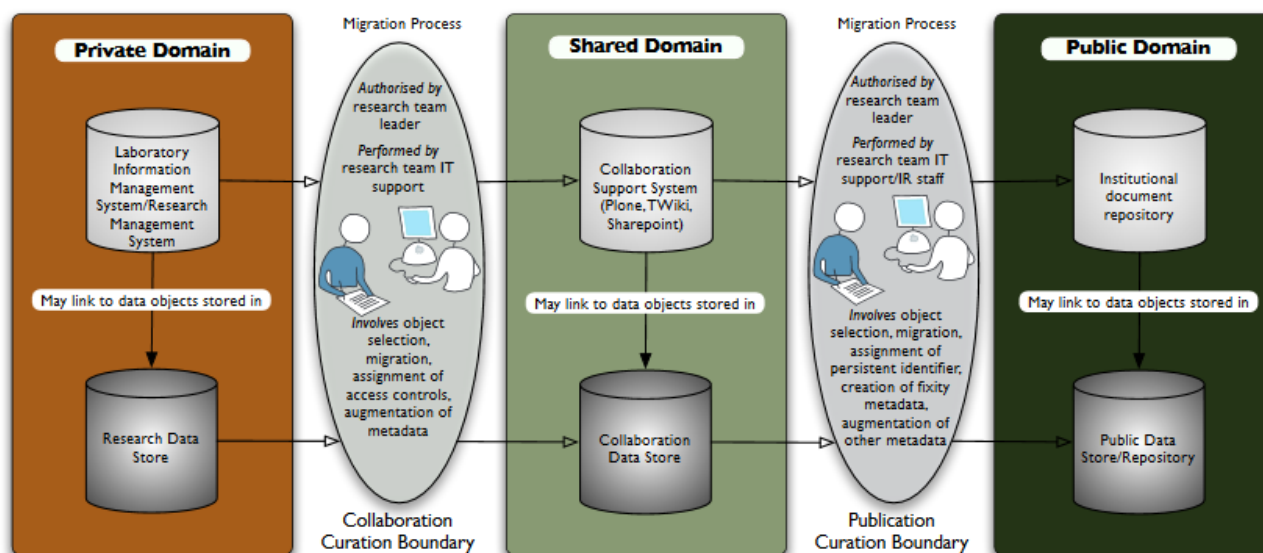


Illustration 2: Délimitation entre domaines privé, partagé et public (31)

- faciles à localiser

2.2.1 Concepts fondamentaux

2.2.1.1 URI et schémas d'URI

Un URI (Uniform Resource Identifier) est une notion très large désignant un moyen simple et extensible d'identification d'une ressource sous forme d'une chaîne de caractères. La ressource peut être directement accessible sur le web ou consister en un élément physique ou abstrait dont le web ne fournit que la description. Les URIs sont répartis en divers schémas : les schémas d'URI. Ceux-ci spécifient de quelle manière un identifiant sera associé à une ressource. Les schémas doivent être enregistrés auprès de l'IANA (Internet Assigned Numbers Authority) et ont été conçus pour être utilisés avec un protocole. Les schémas les plus connus sont entre autres http, https, ftp, urn, tel et mailto¹ (9, 10).

Certains schémas tel que Handle (hdl) ne sont pas enregistrés formellement auprès de l'IANA mais peuvent être utilisé de manière informelle sous le schéma général info (exemple : info:hdl:102.100.100/15)².

Un URI se comprend donc comme un terme englobant entre autres les URL et les URN, les premiers s'utilisant en principe avec le schéma http: et les seconds avec le schéma urn:. Plus généralement, on peut distinguer les URI en deux catégories, les localisateurs et les noms, où un URI peut adopter l'une ou les deux caractéristiques simultanément (11).

1 La liste complète est accessible sous : <https://www.iana.org/assignments/uri-schemes/uri-schemes.xhtml> (consulté le 26 juillet 2017)

2 La liste complète des schémas disponibles dans info: est accessible sous : <http://info-uri.info/registry/> (consulté le 03 août 2017)

2.2.1.2 URL

Un URL donne accès à une ressource directement. Il se peut alors – par hasard – que l'URL soit identique à l'URI. Ce cas de figure est néanmoins impossible si la ressource identifiée est un élément physique (par exemple une personne).

2.2.1.3 Domain Name System (DNS)

- Standardisation en 1987³
- Associe des adresses IP, sous la forme "xxx.xxx.xxx.xxx" (IPv4) en nom de domaine web.
- Les noms de domaine sont distribués hiérarchiquement pour leur résolution. Les TLD (Top Level Domain) sont d'abord résolus (exemples : .org, .ch, .com, .de). Ensuite, à l'intérieur d'un TLD, les sous-domaines sont résolus (exemple : wikipedia.org).

2.2.1.4 Pérennité

La pérennité peut être temporelle. Dans ce sens, John Kunze, fortement impliqué dans le développement de ARK, décrit un identifiant pérenne comme « an identifier [that] is valid for long enough » (12). En effet, on ne peut parler en termes d'éternité, mais plutôt en terme de période suffisamment longue.

La pérennité peut aussi concerner le contenu de la ressource. Certaines ressources abstraites sont dynamiques et évoluent (exemple : un article Wikipedia), alors que d'autres ressources concrètes sont figées (un document daté et disponible en PDF).

La pérennité d'un système d'identifiant est subjective, comme l'affirme E. Bermès (13, p. 4) :

« Ainsi un établissement comme la Bibliothèque nationale de France accorde une grande valeur à l'indépendance à la fois technique et budgétaire de son système d'identifiants, car cette indépendance lui garantit une liberté de mise en œuvre qui va être favorable, dans le contexte d'un tel établissement, à la pérennité. En revanche, un petit éditeur pourra préférer un système très contraint, car cette contrainte va lui apporter un confort technique (en lui fournissant des outils) et une sécurité globale (en proposant par exemple un système de continuité si l'éditeur disparaît) qui sont indispensables à une véritable pérennité. »

La pérennité peut toucher divers aspects d'un identifiant (14) :

- Le nom lui-même de l'identifiant
- L'association entre l'identifiant et la ressource identifiée
- Le service de résolution existants
- L'entité (institution, personne) responsable de l'identifiant

2.2.1.5 Actionnabilité

Un identifiant peut être actionnable, c'est à dire qu'il permet d'accéder à la ressource ou à une description de celle-ci. Le protocole HTTP, utilisable grâce à un navigateur web, rend ce déréférencement possible.

2.2.2 Fonctionnement d'un système de PID traditionnel (avec résolveur)

Un système d'identifiants pérennes traditionnel se base sur un résolveur, qui redirige depuis un identifiant vers une localisation. Cela présuppose une infrastructure technique, mais également sociale. Trois parties prenantes sont présupposées (15, p. 20) :

- Une autorité attribuant l'identifiant
- Un service de résolution

3 RFC 1034(1987): <https://tools.ietf.org/html/rfc1034> (consulté le 26 juillet 2017)
RFC 1035(1987): <https://tools.ietf.org/html/rfc1035> (consulté le 26 juillet 2017)

- Un fournisseur de contenu

Néanmoins, un système doit être conçu de manière agnostique à toute technologie, en vue d'une meilleure pérennité : l'identifiant doit survivre aux évolutions techniques !

« *persistence is a function of organizations, not technology* » (16)

L'infrastructure sociale sous-jacente est la base du paradigme de l'identifiant pérenne, différenciant cette approche notamment de l'approche des URI du web sémantique.

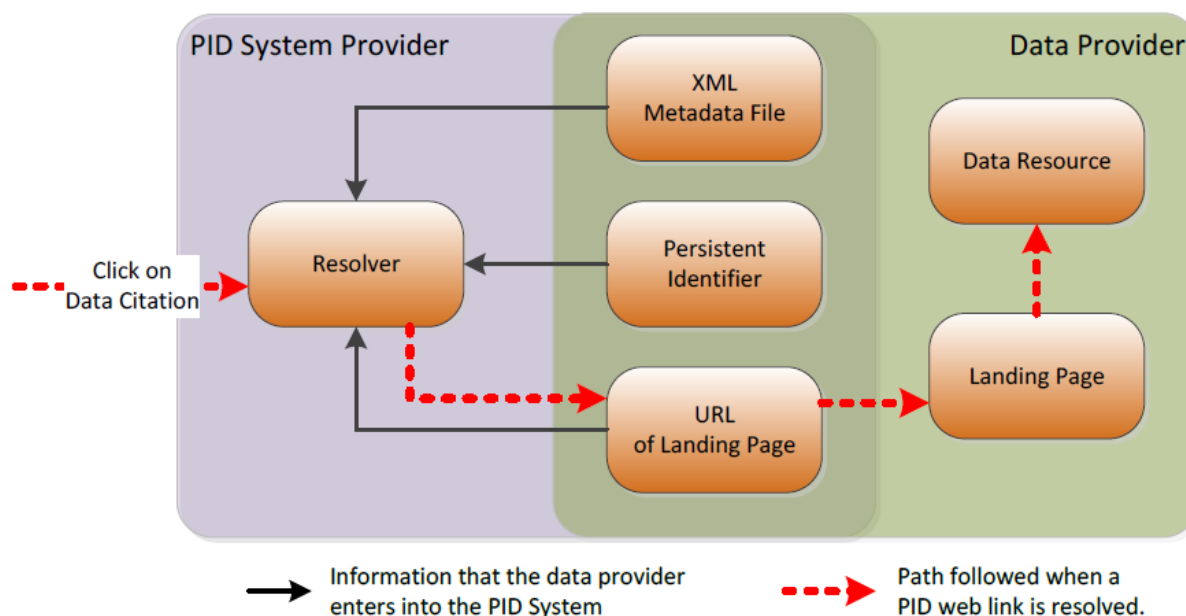


Illustration 3: Fonctionnement d'un système d'identifiants pérennes (8)

Le graphique ci-dessous (Illustration 3) explique l'échange de données entre le fournisseur du service PID et le fournisseur de données.

Un système de PID se base donc en principe sur un système de résolution garantissant la pérennité des URI. Chaque système possède son propre serveur de résolution, en raison de disparités organisationnelles et technologiques. Certaines institutions garantissent des miroirs de certains résolveurs, afin d'assurer la disponibilité du service.

Pour résoudre un PID, il faut donc connaître l'adresse du serveur en question ou – si aucun résolveur n'existe (cas des ISBN) – une source fiable et exhaustive pouvant contenir l'information recherchée. Les services <http://identifiers.org/> et <http://n2t.org/> proposent une résolution de divers PID par construction de l'URL (exemples: <http://identifiers.org/isbn:2070518426> ou <http://n2t.org/doi:10.5281/zenodo.838689>), où l'on spécifie le suffixe du type de PID suivi de deux points et de l'identifiant lui-même.

Les désavantages du système de PID avec résolveur sont entre autres (17) :

- la dépendance au résolveur et à sa pérennité
- le fait que l'éditeur doit se procurer un PID, sans quoi l'identification pérenne n'est pas possible
- La contribution économique que ce système implique en général

2.2.2.1 Résolution multiple

Un résolveur peut, pour un identifiant, comporter plusieurs *locators* différents. S'il est intelligent, il fournit le locator le plus approprié en fonction de l'utilisateur (localisation du fichier, langue, format, etc.).

2.2.3 Pérennité par les archives du web

Une autre approche est celle de la citation d'une ressource au sein d'une archive du web, garantissant ainsi une pérennité plus longue. Cette dernière dépend néanmoins directement de la durée de vie de l'archive en question. La responsabilité de la pérennité revient donc selon ce système aux archives, alors qu'elle réside chez l'éditeur et le service de résolution dans les systèmes de PID.

L'un des principaux défis de cette approche réside dans la représentation des éléments liés à une page web et influant l'affichage de la page (17).

Les archives du web peuvent garantir une certaine pérennité dans une optique de citation, mais ne résolvent pas le problème de l'attribution d'identifiant à des ressources.

2.3 Brève typologie des identifiants

2.3.1 Critères distinctifs

Les divers identifiants peuvent se différencier selon les critères ci-dessous (18, p. 5) :

- Périmètre d'application : données, publications, livres, personnes, etc.
- Technologie sous-jacente
- Coût
- Standardisation
- Usage effectif
- Métadonnées collectées
- Gouvernance et infrastructure sociale
 - hébergement décentralisé (ARK) ou centralisé (Handle)

2.3.2 ARK

- Date de 2001
- Maintenu par l'Université de Californie
- Est capable de séparer un identifiant unique assigné à une ressource des potentielles multiples adresses pouvant faire office de proxys vers la ressource finale (19)
- Particularités des spécifications :
 - NMAH (Name Mapping Authority Hostport) : adresse du serveur qui résout les identifiants
 - Qualificatif : suffixe ajouté à l'identifiant et permettant d'affiner la localisation à l'intérieur d'une ressource (qualificatif de granularité) ou de demander un service ou un format d'affichage spécifique (qualificatif de service).
- Peut identifier divers types d'entités (événements, personnes, publications, etc.)
- Peut intégrer d'autres schémas d'identifiants
- Implémentations
 - BNF

2.3.3 PURL

- Développé par OCLC autour de l'année 1996, et maintenu par Internet Archive depuis 2016
- Solution décentralisée, consistant en un serveur hébergeant le résolveur. Chacun peut installer le serveur PURL librement chez lui et mettre en place un service PURL.

- En complément, un serveur hébergé par Internet Archive est à disposition pour créer ses propres PURL⁴.
- Système non standardisé au niveau de l'IETF (donc reste uniquement une solution technique)

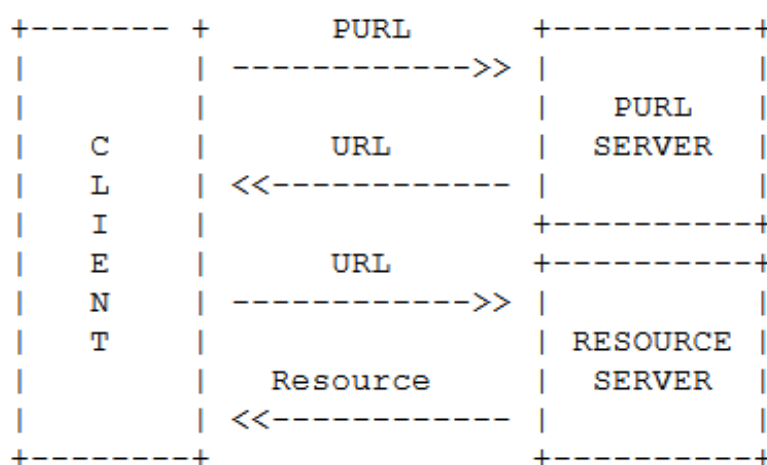


Illustration 4: Fonctionnement de PURL

- Un PURL ne pointe pas directement vers la localisation de la ressource, mais vers un service de résolution (Illustration 4).
- Le protocole HTTP est la solution technique utilisée pour la résolution (redirect)
- Un PURL est composé de trois parties (ex : `http://purl.oclc.org/keith/home`):
 - Le protocole (`http`);
 - l'adresse de résolution, ou nom de domaine (`purl.oclc.org`)
Cette adresse est résolue par le DNS
 - le nom (`keith/home`)
Le nom est résolu par le résolveur PURL

Critique :

- Des identifiants PURL ne se résolvent plus car des serveurs n'ont pas été maintenus et sont désormais déconnectés du réseau.

2.3.4 URN

- Système standardisé en 1997 par l'IETF, norme révisée ensuite en 2002 puis 2017⁵
- Standard indépendant de toute technologie, dont pas obligatoirement lié à un résolveur
- 62 namespaces définis aujourd'hui (03.07.17)⁶
 - Chaque namespace est isolé des autres et utilise un résolveur différent
 - Désavantage : nécessite de rédiger les spécifications du namespace
- Est l'un des schémas URI enregistrés formellement

4 Accessible sous <https://archive.org/services/purl/> (consulté le 26 juillet 2017)

5 RFC 2141(1997): <https://tools.ietf.org/html/rfc2141> (consulté le 26 juillet 2017)

RFC 8141(2017): <https://tools.ietf.org/html/rfc8141> (consulté le 26 juillet 2017)

6 <https://www.iana.org/assignments/urn-namespaces/urn-namespaces.xhtml> (consulté le 25 juillet 2017)

- Implémentations
 - ISSN
 - ISBN
 - ISAN
 - NBN (National Bibliography Numbers)
 - UUID (Universally Unique Identifier)
- URN architecture assumes a DNS-based Resolution Discovery Service (RDS) to find the service appropriate to the given URN scheme

2.3.4.1 NBN (National Bibliography Number)

Description (20) :

- Portée des NBN : « NBNs have traditionally been given to documents that do not have a publisher-assigned identifier, but are cataloged to the national bibliography. NBNs can be seen as a fall-back mechanism: if no other, better established identifier such as ISBN can be given, an NBN is assigned. »

En pratique, il s'agit souvent de ressources électroniques auxquelles n'a été attribué aucun ISBN.

- Chaque pays élabore donc ses propres règles pour la création des NBN.
- Chaque pays génère ses propres identifiants NBN. Pour éviter les doublons, un code de pays est ajouté en préfixe. Chaque pays peut avoir une implémentation différente du NBN selon ses besoins ; il n'existe pas de coordination internationale.

Projet suisse (21) :

- Mise à disposition par la Bibliothèque nationale suisse d'un mini programme sous forme de feuille Excel (complétée par du code VBA) permettant de générer des NBN.
- L'obtention d'URN est gratuite.
- La Bibliothèque coordonne le service en Suisse, mais utilise le service de résolution de la Bibliothèque nationale allemande. L'Autriche fait de même.
- Un URN de type "urn:nbn:ch:[...]" ne peut être attribué qu'à un document (revue, site web, article, livre, affiche, etc.).
- Les URN sont attribués soit par la BN, soit par des archives de longue durée (du domaine universitaire suisse) autorisées par la BN.
- L'Illustration 5 illustre le processus d'attribution d'un URN.

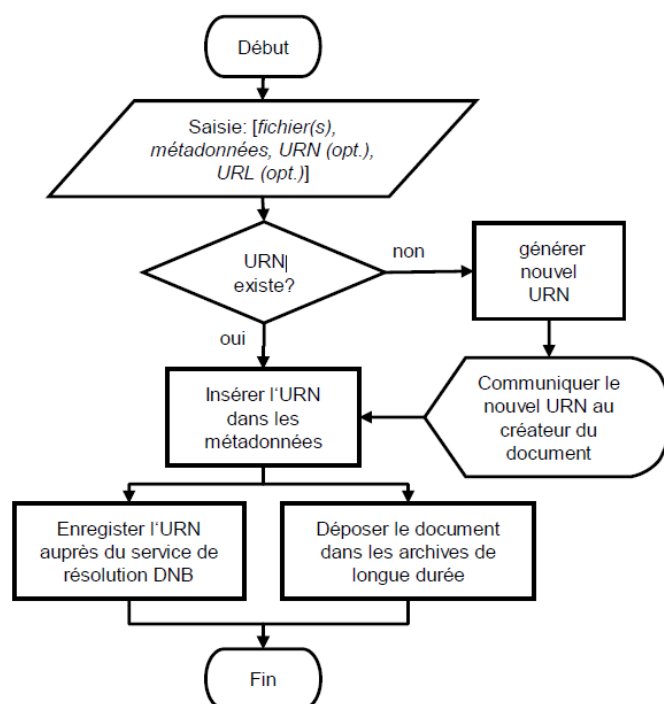


Illustration 5: Processus d'attribution d'un NBN en Suisse

Projet italien (19) :

- Inspiration du système hiérarchique DNS pour déléguer la gestion de partie du NBN italien (nbn:it:). Chaque autorité de second niveau gère les identifiants du niveau en question (ex. : nbn:it:frd:).
- Chaque « noeud » est un résolveur, et moissonne les données de ses propres sous-niveaux. Un noeud peut également interroger les données d'un noeud frère, constituant ainsi un système redondant dont le but est la fiabilité.
- Des politiques concernant l'attribution de NBN sont créées pour chaque noeud. Thèmes : comment générer des PI bien formés, quelles ressources méritent un PI, quelle granularité, audit de dépôt pour attribution de PI.
- Les institutions attribuent des PI à des dépôts en moissonnant leurs données (DC, MPEG21, METS) via OAI-PMH.

Projet allemand (22):

- Collaboration entre la Deutsche Nationalbibliothek et la Universitäts- und Landesbibliothek Sachsen-Anhalt
- Implémentation d'URN sous la forme de NBN pour les publications du VD17
- Exemple: <http://nbn-resolving.de/urn:nbn:de:gbv:3:1-78197-p0015-7>
- Possibilité de citer à la granularité d'une page, même si l'identifiant lui-même se réfère à un objet numérisé dans son ensemble. Ci-dessus, la partie "p0015-7" se compose de la pagination (p0015) et d'un numéro de contrôle (7). L'identifiant de la ressource elle-même est "urn:nbn:de:gbv:3:1-78197".

2.3.4.2 UUID (Universally Unique Identifier)

- Identifiant utilisé en informatique pour les composantes logicielles telles que des plugins.
- Normalisé : RFC 4122 (2005), ISO/IEC 9834-8:2008
- Le GUID (Globally Unique Identifier) est une variante du UUID créée par Microsoft.

2.3.5 Handle

- Système implémenté pour la première fois en 1994, et standardisé en 2003⁷
- Système conçu par la CNRI (Corporation for National Research Initiatives)⁸
- Son application principale est sans aucun doute le DOI
- Spécifications :
 - Le standard définit l'identifiant sous forme d'un préfixe, contenant l'autorité de nommage, et d'un suffixe, contenant le nom local de la ressource (exemple : 20.1000/100).
 - Le préfixe est lui-même subdivisé par des points en différents domaines (par exemple "102." vaut pour l'Australie, "10." pour les DOI, etc.).
 - Un identifiant Handle est opaque ; il ne trahit donc rien de la ressource qu'il désigne. Ces informations sont stockées exclusivement dans les métadonnées.
 - Ne nécessite pas l'utilisation de DNS, à l'inverse de PURL et ARK, car il n'utilise pas de nom de domaine.
- Global Handle Registry (GHR)
 - Service de base du système Handle, recensant tous les préfixes utilisés
 - Sous la responsabilité de la fondation DONA⁹ basée à Genève, mais est géré en pratique par la CNRI.
 - Indique pour chaque préfixe à quel résolveur local la demande doit être relayée ; il ne résout pas directement toutes les requêtes !
- L'enregistrement d'un préfixe au sein du registre global est gratuit la plupart du temps (23, p. 10).
- Handle Proxy Server, pour résolution par URL : [http://hdl.handle.net/\[id\]](http://hdl.handle.net/[id])
- Quelques institutions peuvent accéder et gérer le GHR : les Multi-Primary Administrator (MPA)(24)
 - IDF (DOI) : préfixe 10
 - ITU : préfixe 11
 - CNRI : préfixe 20
 - GWDG (ePIC) : préfixe 21
 - Coalition in China (ETIRI, CDI, CHC) : préfixe 86

2.3.5.1 DOI en détail

- Introduit dès l'année 2000, syntaxe standardisée par NISO dès 2010, ensemble du système normalisé par l'ISO en 2012¹⁰
- Périmètre : principalement publications scientifiques et données de la recherche
- "doi" pourrait être enregistré comme namespace URN, étant donné qu'il remplit toutes les exigences. Ceci impliquerait néanmoins une redirection supplémentaire pour la résolution d'identifiant, ainsi que des lourdeurs administratives (11).
- Coût par identifiant, permettant d'entretenir l'infrastructure de résolution
- Un DOI redirige en principe vers une « landing page » et non pas directement vers l'objet numérique.

7 RFC 3650(2003): <https://www.ietf.org/rfc/rfc3650> (consulté le 26 juillet 2017)

8 <https://www.cnri.reston.va.us/> (consulté le 8 septembre 2017)

9 <https://www.dona.net/> (consulté le 25 juillet 2017)

10 NISO Z39.84-2005 (R2010), ISO 26324:2012(F)

- « "DOI" est interprété comme "identifiant numérique d'objet" (et non "identifiant d'objet numérique") » (25)
- Structure d'un DOI sur l'exemple <http://doi.org/10.15291/libellarium.v9i2.256>
 - Préfixe (10.15291) composé d'un indicateur du répertoire suivi du code du déclarant
 - Suffixe (libellarium.v9i2.256) de longueur variable, pouvant également incorporer un autre type d'identifiant (exemple : 10.1038/issn.1476-4687)
- Métadonnées
 - Un DOI est associé à des métadonnées. Une institution doit toujours fournir au moins le noyau DOI (ou DOI Kernel Metadata Declaration) (26, chap. 4) de métadonnées à son agence d'enregistrement.
 - Le noyau de métadonnées peut être complété à volonté par des métadonnées complémentaires.
- L'Australian National Data Service (ANDS) propose par exemple un service en ligne¹¹ gratuit permettant aux institutions partenaires d'obtenir des DOI pour leurs publications ou données, que ce soit de manière manuelle ou automatique.

Organisation

- La gouvernance du système DOI est assurée par l'IDF (International Digital Object Identifier Foundation, Inc.), une institution à but non-lucratif basée aux Etats-Unis (26, chap. 7).
- L'IDF a une dizaine d'agences d'enregistrement, dont notamment Crossref et Datacite.
- Ces agences d'enregistrement
 - ont des membres pouvant assigner des DOI. Le CERN et l'ETH Zürich sont membres de Datacite¹².
 - allouent des préfixes de DOI, enregistrent les DOI et fournissent l'infrastructure nécessaire à leurs membres
 - participent au financement de IDF par des cotisations, mais peuvent ensuite appliquer un modèle de prix différent à leurs propres membres (par exemple la gratuité)
 - peuvent être commerciales. Datacite et Crossref sont à but non lucratif.
 - DataCite a pour but une amélioration des possibilités de citation des jeux de données et encouragent dans cette optique l'attribution de DOI.
 - CrossRef est un consortium d'éditeurs attribuant des DOI aux publications scientifiques qu'ils produisent.

Critiques du système DOI

- Les DOI ont une excellente acceptation au sein des communautés scientifiques. Les chercheurs connaissent ce système d'identification, et apprécient en obtenir pour leurs publications et leurs jeux de données, y voyant là des opportunités de visibilité de leur activité scientifique (citation).
- Les DOI peuvent rediriger les utilisateurs plutôt vers les sites des éditeurs, où la ressource est payante, que vers les sites des bibliothèques, où la ressource peut être gratuite. DOAI¹³ est une alternative au résolveur officiel, menant de préférence vers des ressources en Open Access. DOAI s'appuie sur les données du moteur de recherche BASE (Bielefeld Academic Search Engine).

2.3.5.2 ePIC en détail

- Consortium fondé en 2009 par un groupe de six institutions de recherche européennes. L'adhésion requiert une participation aux coûts (cotisation) par les membres ePIC (27).

11 Cite my data : <http://www.ands.org.au/online-services/doi-service> (consulté le 31 juillet 2017)

12 Seule la bibliothèque de l'ETH peut être considéré comme membre en Suisse à cause du statut particulier du CERN.

13 Accessible sous : <http://doai.io/> (consulté le 10 août 2017)

- ePIC vise l'attribution d'identifiants pérennes – sur la base du système Handle – aux données de la recherche se trouvant encore à un stade précoce du cycle scientifique, devant être citables mais dont la préservation à long terme n'est pas encore décidée. ePIC se destine aux institutions de recherche européennes.
- Il n'est pas forcément nécessaire de devenir membre pour obtenir un préfixe ePIC et maintenir un service de résolution.
- Possibilité d'identifier un fragment d'une ressource, sans créer un identifiant spécifique au fragment (utilisation d'extension et de base)
- Structure d'un identifiant ePIC : 21.012/xyz-123
 - 21 est le préfixe ePIC au sein du système Handle
 - 012 est le préfixe de l'institution au sein du consortium ePIC
 - xyz-123 est le suffixe, identifiant la ressource
- La suppression d'une ressource référencé par un identifiant ePIC est permise. L'identifiant doit alors renvoyer aux métadonnées, qui sont conservées

2.3.6 URI / Cool URI

- Approche Linked Data
- Une URI « cool » est une URI pérenne, qui ne changera pour aucune raison exceptée l'insolvabilité du fournisseur de données en regard de son nom de domaine (28). En pratique cependant, de nombreuses raisons mènent les institutions à changer leurs URI dites pérennes.
- Peut être généré et attribué par tous (pas forcément lié à une autorité d'enregistrement telle qu'IDF). Dans ce cas, c'est une solution gratuite, qui n'implique pas de structure de responsabilité induite par le système des identifiants pérennes.

La pérennité peut aussi être garantie par une autorité d'enregistrement qui gère un type d'identifiant (exemple : ISNI).

- Peuvent être dynamiques : le contenu d'un identifiant à un moment X ne sera pas forcément le même une année plus tard (actualisation)
- Un groupe de travail s'est réuni en 2011 et a publié un manifeste pour une conception coordonnée des identifiants selon l'approche des PID et selon l'approche du Linked Open Data.
- Implémentations
 - ORCID
 - ISNI

2.3.6.1 ORCID

- utilisé pour les personnes vivantes dans le monde académique
- ne stocke pas d'informations sensibles sur les personnes telles que la date de naissance, le genre ou l'adresse postale
- Numéro composé de 16 chiffres en quatre groupes (XXXX-XXXX-XXXX-XXXX)

2.3.6.2 ISNI

- Normalisé par ISO 27729
- Périmètre : entités publiques, à savoir personnes réelles ou fictives et organisations, issues de la chaîne de production de contenus intellectuels et artistiques (29).
- L'objectif est l'identification stricte des entités afin d'établir un pont entre les identifiants locaux/propriétaires. ISNI ne comprend donc pas d'informations détaillées ; celles-ci se trouvent sur les portails alimentant ISNI.

- Gouvernance : ISNI International Agency (ISNI-IA), organisation à but non-lucratif

2.3.7 Autres

ISRC : International Standard Recording Code

- Identifiant international standardisé (ISO 3901:2001) pour enregistrements sonores et audio-visuels musicaux.
- Identifie les enregistrements et non les médias physiques de stockage, tels que des CD ou des DVD.

GND : Gemeinsame Normdatei

- Fichier d'autorités encyclopédique, dont la coordination et le développement sont assurés par la Bibliothèque nationale allemande.
- De nombreuses institutions culturelles participent à la saisie et à l'enrichissement des données de la GND.
- Les identifiants sont attribués en tant que « Cool URIs ».

2.3.8 Tableau synthétisant les divers identifiants existants

	Périmètre d'application	Technologie sous-jacente	Standardisation	Coûts pour la publication	Usage effectif	Métadonnées collectées	Gouvernance, infrastructure sociale	Hébergement (centralisé ou décentralisé)
ARK	libre	ARK	IETF (ébauche expirée)	Aucun	Moyennement répandu	Aucune (décentralisé)	Autorité internationale délivrant les numéros d'autorités nommantes	Décentralisé
PURL	libre	PURL	Aucune	Aucun	Moyennement répandu	Aucune	Solution technique développée par la communauté, sous l'égide de Internet Society	Décentralisé
NBN	Publications recensées dans une bibliographie nationale	URN	IETF RFC 3188	Aucun	Moyennement répandu	Dépend du pays (si un résolveur existe ou non)	L'autorité compétente est la bibliothèque nationale du pays.	Hébergement facultatif (mais si un résolveur existe : centralisé au niveau du pays)
DOI	Publications scientifiques ou jeux de données issus de la recherche	Handle	ISO 26324:2012 (Handle : IETF RFC 3650)	Cotisations de la part des agences d'enregistrement	Très répandu	Oui, certains éléments de base	Une autorité internationale, quelques agences d'enregistrement selon les domaines	Centralisé
ePIC	Données de la recherche	Handle	(Handle : IETF RFC 3650)	Cotisation de la part des institutions membres	Peu répandu	Oui, certains éléments (non connus, à chercher)	Un consortium et ses membres	Centralisé auprès d'un des six membres ePIC
ORCID	Personnes vivantes actives dans le domaine académique	Cool URI	Aucune (compatible ISO 27729:2012 - ISNI)	Cotisation de la part des institutions membres	Très répandu	Oui, un set prédéfini	Une société internationale et ses membres	Centralisé
ISNI	Personnes réelles ou fictives et organisations	Cool URI	ISO 27729:2012	Cotisations de la part des agences d'enregistrement	Très répandu	Oui, un set prédéfini	Une autorité internationale avec plusieurs agences d'enregistrement	Centralisé

2.4 Services

Un service d'identifiants pérennes est géré via un identifiant management system. Un tel service prend en compte deux aspects (30) :

1. La curation des identifiants : création, actualisation, suppression
2. La résolution : accès à la ressource et/ou à ses métadonnées

L'implémentation d'un système d'identifiants pérennes au sein d'une institution peut se faire selon deux approches (31) :

1. Le chercheur fait lui-même la demande d'identifiant pour son jeu de données, via une interface adéquate. L'institution doit dans ce cas s'engager à maintenir le jeu de données disponible de manière pérenne, ou au moins ses métadonnées.
2. L'équipe administrant les données de la recherche attribue les DOI aux jeux de données. Dans ce cas, cette action doit être communiquée au chercheur afin qu'il puisse citer ses données.

2.4.1 Politique pour l'utilisateur

*« [...] policy is far more important in guaranteeing persistence of identifiers than technology. »
(14)*

Une politique décrit l'usage voulu d'un service mis en place et émet de recommandations et un engagement de l'institution. Elle peut aborder entre autres les points suivants :

- Services mis à disposition
- Manière de citer les identifiants du service
- Ressources recevant un identifiant pérenne
- Type d'identifiant par type de ressource
- Moment de demande et de l'attribution d'un identifiant pérenne pour une ressource
- Rôles des personnes dans la gestion des identifiants
- Pérennité garantie (pendant combien d'années) des identifiants

Par exemple, le Australian National Data Service (ANDS) a développé deux services d'identifiants pérennes, s'articulant selon la politique synthétisée dans l'illustration 6¹⁴. ANDS n'attribue des DOIs que si le jeu de données remplit certaines conditions (publication scientifique, préservation à long terme, métadonnées suffisantes, etc.), sans quoi il attribue un identifiant Handle propre à l'institution. L'avantage de ce système est qu'un chercheur peut obtenir un identifiant dans une phase précoce du processus scientifique, et peut ainsi établir des citations de ses jeux de données avant de les avoir déposés dans un dépôt agréé.

14 Exemple de politique pour le service DOI (<http://www.ands.org.au/online-services/doi-service/doi-policy-statement>, consulté le 2 août 2017) et pour le service Handle (<http://www.ands.org.au/online-services/handle-service/handle-policy-statement>, consulté le 2 août 2017).

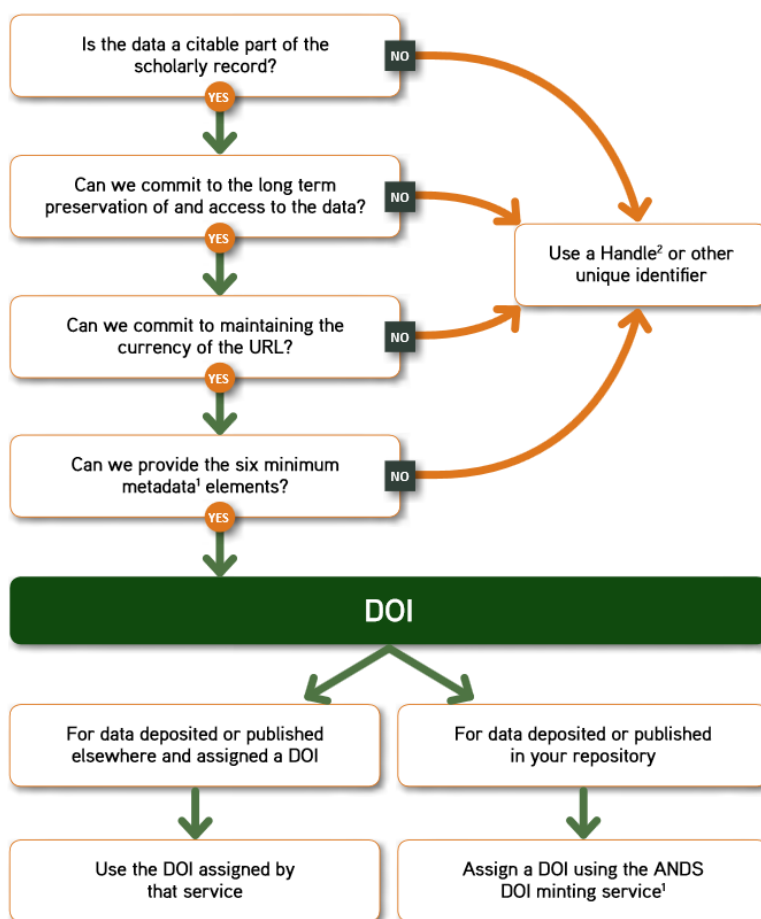


Illustration 6: Politique au sein de ANDS (33)

« chat entities should be assigned persistent identifiers, how should those identifiers be exposed to services, and what guarantees does the provider make on how long various facets of the identifiers will persist » (<http://www.dlib.org/dlib/january09/nicholas/01nicholas.html>)

2.4.2 Design des identifiants

2.4.2.1 Signification versus opacité

Un identifiant opaque : il ne trahit rien de la ressource et se compose d'une suite abstraite de caractères

- Avantages : ne se base pas sur des attributs de la ressource (titre, auteur, etc.) pouvant changer avec le temps
- Inconvénients : il ne se mémorise pas facilement, n'est pas significatif, se reproduit difficilement de manière manuelle

2.4.2.2 Caractères sensibles dans les URL

De nombreux identifiants sont généralement réutilisés sous forme d'URL. Certains caractères sensibles dans les URL doivent donc être évités dans les identifiants (30, chap. 3.1.3), tels que l'esperluette ou l'espace. Les lettres I et O devraient être omises afin d'éviter toute confusion avec les caractères l et 0.

2.4.2.3 Longueur

Une longueur uniforme facilite ensuite la vérification de la validité de l'identifiant (30, chap. 3.1.3.4).

2.4.3 Implémentation

Nicholas, Ward et Blinco (14) proposent une checklist pour l'implémentation d'un service de gestion des identifiants garantissant la pérennité :

1. Identifier le modèle des données (quelles entités existe-t-il et comment sont-elles reliées ?)
2. Déterminer à quelles entités des PID doivent être attribués
3. Dissocier la gestion des identifiants de la gestion des ressources (la suppression d'une ressource électronique ne doit pas impliquer la suppression du PID)
4. Définir les workflows de gestion des identifiants et les intégrer à la gestion de l'information (ex. : si une ressource change d'URL, un processus automatique effectue le changement dans le résolveur)
5. Développer des services tirant profit des PID (gestion interne et accès externes)
6. Déterminer la notion de pérennité et s'engager en conséquence au moyen d'une politique

Ces étapes impliquent des interactions avec le fournisseur de données/ressources, le gestionnaire de données / IT et l'utilisateur. Ce travail a été réalisé dans le cadre du projet PILIN en Australie, qui a débouché sur la création de ANDS¹⁵ (Australian National Data Service).

15 Accessible sous <http://www.ands.org.au/> (consulté le 8 août 2017)

3 Partie B : Linked Data et données de la recherche

3.1 Introduction

Cette partie A donne des éléments de réponse sur l'utilisation des Linked (Open) Data dans le cadre des données de la recherche (research data), ainsi qu'un bref aperçu de son utilisation dans les domaines de l'art et du design.

L'utilisation des PID dans le cadre des données de la recherche (par exemple assigner un DOI à un set de données) est une tendance de plus en plus importante ; que ce soit pour assurer la pérennité de ces données, pour simplifier la citation par les pairs ou encore pour les rendre plus accessibles. Les LOD ont un énorme potentiel à faire avancer la recherche scientifique, comme par exemple en tirant de nouvelles informations grâce aux interférences possibles avec les technologies du web sémantique.

Les institutions de mémoire, les GLAM (galleries, libraries, archives, museums), sont les plus à même d'enrichir les collections scientifiques et culturelles par le biais des LOD car elles ont les connaissances des fonds et ont tout intérêt à permettre une interopérabilité entre différents dépôts de données. Les LOD augmentent ainsi la valeur ajoutée des collections tout en supprimant les redondances dans le stockage des métadonnées descriptives dans différents lieux.

3.2 Périmètre

Les technologies du web sémantique sont de plus en plus souvent utilisées pour les différentes étapes du cycle de la recherche. Distinguons ici trois acteurs importants :

1. Le chercheur : il crée, réutilise, traite, transforme, analyse les données. Il s'intéresse d'abord à la structure interne et au contenu des données, éléments spécifiques à sa discipline. Il peut utiliser le Linked Data comme outil pour partager ou analyser ces données spécifiques, dont il est le spécialiste. C'est le cas en particulier dans les domaines des sciences de la vie.
2. L'archiviste : il évalue, préserve et migre les données. Pour ses activités, le Linked Data n'offre pas de bénéfices directs (voir chapitre 3.3).
3. Le bibliothécaire : il publie et valorise les données, toutes disciplines confondues. Il s'intéresse aux points communs entre les jeux de données, à un niveau macro. Il peut utiliser le Linked Data pour mieux exposer les données et créer des interconnexions afin que le chercheur les trouve plus facilement. Tout au plus, il peut recommander des outils et des bonnes pratiques au chercheur dans ce domaine.

Les activités du chercheur (1) varient selon la discipline concernée, et celles de l'archiviste (2) n'utilisent pas directement le Linked Data. Ce document se concentre donc sur les activités du bibliothécaire (3) en lien avec le Linked Data dans le domaine des données de la recherche.

3.3 Bénéfices potentiels des Linked Data pour les données de la recherche

- Meilleures possibilités de recherche de sets de données
- Flexibilité de la représentation des données en RDF permettant de prendre en compte la grande hétérogénéité des données de la recherche
- Meilleure exposition des données, dans l'optique Open Science
- Encouragement à la "data driven research" grâce à l'exploitation des données par des outils du web sémantique
- Intégration avec des données externes (ex : DBpedia). Annotation/enrichissement des données par des sources externes directement par le chercheur (ex : ajout d'URI DBpedia en lien).

3.4 OpenScience - eScience

L'utilisation des Linked Data dans le domaine des données de la recherche est compris dans la mouvance et la vision de l'Open Science où les recherches scientifiques sont disponibles de manière ouverte et pérenne à chaque étape d'un projet. Le partage de ces résultats est autant une nécessité pour la communauté scientifique que pour les institutions elles-mêmes (1).

Selon Jim Gray (Microsoft Research), l'eScience serait le quatrième paradigme dans la manière de rechercher et d'étudier la science. L'eScience combine les trois autres, et anciens, paradigmes (théorie, expérimental et simulation). Les bibliothèques ont un rôle à jouer pour s'assurer du bon fonctionnement de la recherche dans cette mouvance car elles ont une longue tradition dans la dissémination et la valorisation de la connaissance. Le monde scientifique peut avoir confiance dans les bibliothèques car elles ont tendance à durer. Ainsi, le risque de voir disparaître des bibliothèques face à des projets ou initiatives actuelles est moins fort (2).

3.5 Projets

Les trois projets ou initiatives succinctement présentés dans cette section proposent pour la plupart des DOI comme identifiants ou/et une couche sémantique en Linked Data pour faire le lien vers des sources internes ou externes.

3.5.1 Université de Bielefeld - CITEC

A l'Université de Bielefeld, les LOD sont utilisés dans le cadre du centre interdisciplinaire de recherche CITEC (The Center of Excellence Cognitive Interaction Technology¹⁶) pour satisfaire trois exigences qu'ils se sont fixées : **simplicité** de publier les données, **utilité** des données publiées pour la communauté scientifique et que tout puisse être facilement **cité** (1). L'objectif de CITEC est le développement d'un écosystème technique et de services favorisant l'Open Science.

L'infrastructure (Illustration 7) fonctionne autant comme un dépôt institutionnel où un DOI est assigné à chaque set de données que comme un 'Linked Data hub' où la couche sémantique supplémentaire provient de sources internes (informations concernant l'université, les institutions et les chercheurs) et externes (DBpedia, ontologies du domaine).

Les composantes de l'infrastructure sont les suivantes :

- Publication Management : correspond au dépôt institutionnel, contenant des articles scientifiques et publications
- Data Management : gestion des données de la recherche
 - La plateforme CITEC Open Research met à disposition des chercheurs des outils de gestion des

16 Accessible sous : <http://cit-ec.de/> (consulté le 10 août 2017)

- données (wikis, gestion de projet, ticketing) et de publication dans un dépôt approprié
- Le toolkit CITEC permet l'intégration des diverses sources, notamment sets de données depuis CITEC Open Research et publications depuis PUB, et l'interconnexion à l'aide des Linked Data et des DOI. Des pages web sont créées par le toolkit sur la base du CMS Drupal.

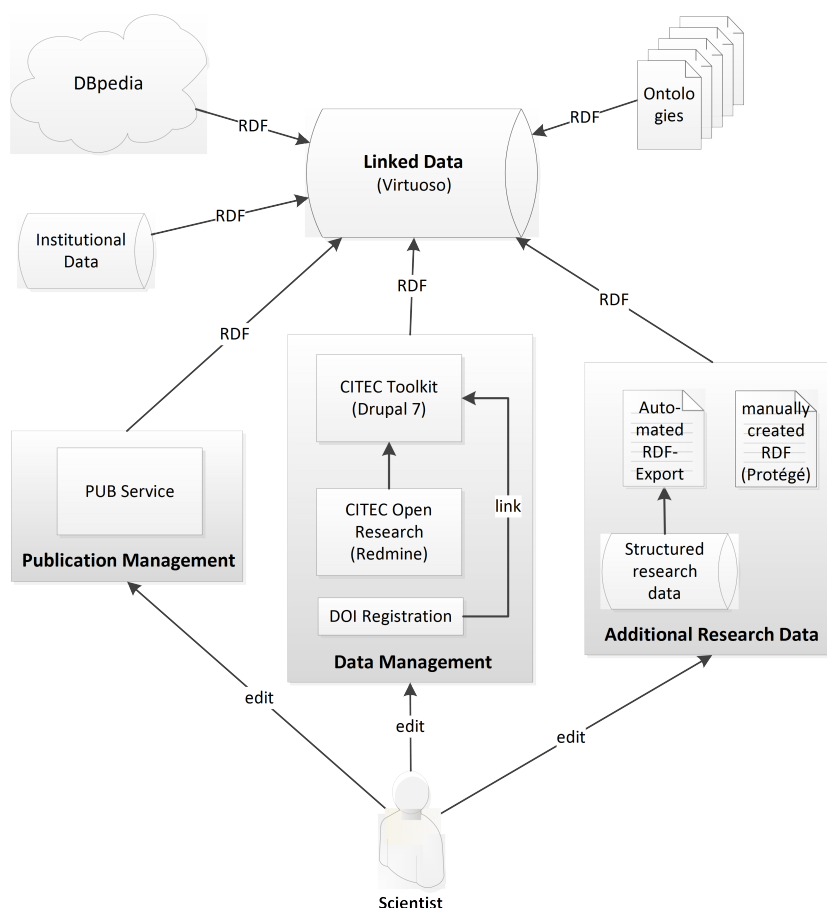


Illustration 7: L'infrastructure Open Science à l'Université de Bielefeld (1)

L'établissement de la couche sémantique se fait par collecte des données des différents pans de l'infrastructure. Pour PUB et la base de données des collaborateurs de l'université, le RDF est généré au moyen de XSLT. Pour le toolkit, il est créé grâce à un module Drupal. Les vocabulaires VIVO et DOAP sont utilisés pour la représentation des informations. Une interconnexion avec DBpedia a été choisie, car la source est encyclopédique et très bien connectée.

3.5.2 Archives publiques de Taïwan

Une grande majorité des archives publiques de Taïwan (plus de 840 000 ressources numériques en XML) sont actuellement disponibles sur un dépôt de données (bêta) en LOD¹⁷. Tout d'abord, les informations ont été formatées en CSV puis transformées en RDF. Le processus XML-CSV-RDF a été choisi pour garantir autant un accès facile pour les machines que pour les humains. Une ontologie a été également créée pour décrire ces données.

17 Accessible sous : <http://data.odw.tw/> (consulté le 09 août 2017)

The screenshot shows a dataset page on data.odw.tw. The main title is 'Metadata as Linked Data for Research Data Repositories'. The page includes a sidebar with social media links (Social, Google+, Twitter, Facebook) and other access options (JSON-LD, Turtle, XML via CKAN API). The main content area displays the following metadata:

METADATA	
rdf:type	r4r:Article
r4r:locateAt	http://data.odw.tw/article/20170308
dc:coverage	<ul style="list-style-type: none"> • 台北, 台灣 • Taipei, Taiwan
dc:creator	<ul style="list-style-type: none"> • 李承鑫, 黃韋菁, 莊庭瑞 • Cheng-Jen Lee, Andrea Wei-Ching Huang and Tyng-Ruey Chuang
dc:date	2017-03-08
dc:description	Data repository has long become an effective mechanism for integrating and managing the research resources. And how to improve the reuse value of research resources is one of the key issue regarding data repository. Recently, publishing those resources in the form of linked data with referenced interlinks and rich semantics on the world wide web draws much attention from data repository owners. In this study, we first converted 840,000 XML-formatted CC-licensed digital resources from the Union Catalog of Digital Archives Taiwan (http://catalog.digitalarchives.tw/) into human-editable CSV tabular data. Those tabular data are then transformed into linked data in RDF (Resource Description Framework). The XML-CSV-RDF process ensures accessibility to resources for both human and machine. An ontology (voc.4odw) is also designed to describe each resource and its provenance. Two types of linked data resources are generated: one type is simply described by Dublin Core's

Illustration 8: Exemple d'une ressource enrichie avec des Linked Data dans data.odw.tw

Tous ces sets de données sont disponibles en ligne sur une infrastructure construite avec CKAN (Comprehensive Knowledge Archive Network) et les données peuvent être exportées en JSON-LD, en Turtle et en RDF/XML. Un accès SPARQL (Virtuoso) a été intégré pour les requêtes avancées (3)¹⁸.

3.5.3 Research Graph

- Base de connaissance internationale interconnectant les données de la recherche, notamment autour des quatre entités d'intérêt suivantes :
 1. Les chercheurs
 2. Les publications
 3. Les programmes/fonds de recherche
 4. Les jeux de données
- Données normalisées au sein d'un modèle de graph, et rassemblées grâce à des services de mash-up
- Initiative menée dans le cadre du projet RD-Switchboard mené par le Data Description Registry Interoperability (DDRI) working group de la Research Data Alliance¹⁹.
- Fort ancrage australien dans le projet, mais également présence de GESIS (Benjamin Zapilko) ou

¹⁸ Illustration 8 est un exemple d'un jeu de données (un poster).

¹⁹ Accessible sous : <https://www.rd-alliance.org/groups/data-description-registry-interoperability.html> (consulté le 09 août 2017)

encore du CERN.

Linked Data dans le cadre d'un projet (encore en cours) (4)

- Mapping du modèle Research Graph au modèle VIVO ISF
- Transformation des données du Research Graph en triplets VIVO
- Création d'une interface VIVO pour le Research Graph

3.6 Vocabulaires LOD

DCAT – Data Catalog Vocabulary²⁰ :

- vocabulaire pour la description de jeux de données, en vue de faciliter leur interopérabilité sur le web
- recommandation du W3C Recommendation (2014)

VIVO²¹ :

- vocabulaire complexe permettant de représenter l'environnement d'une institution académique, notamment les personnes, les organes institutionnels, les événements, les documents, les rôles, etc.
- est basé sur divers vocabulaires établis, tels que bibo, vcard ou foaf, et sur d'autres vocabulaires plus spécifiques (obo, Ontology for Biomedical Investigations). L'intégration de l'ensemble de ces vocabulaires forme le "VIVO Integrated Semantic Framework" (VIVO-ISF), structurellement logique (5).
- VIVO est un projet soutenu par les cotisations de ses membres, et dont le but est le développement de logiciels Open Source et de l'ontologie VIVO dans le domaine académique.

DOAP - Description of a Project²² :

- vocabulaire permettant de décrire un projet
- développé par un indépendant, Edd Wilder-James (ou Edd Dumbill)

Data Cube²³ :

- vocabulaire pour la description de données, notamment statistiques
- permet par exemple de décrire un jeu de données, des observations, une période de référence, et surtout la structure interne du jeu de données (types de mesure, attributs, etc.).
- recommandation du W3C Recommendation (2014)

OAI ORE²⁴ :

- vocabulaire pour la description, et surtout l'échange de métadonnées sur les ressources web
- permet de regrouper des ressources en agrégations (par exemple un article possédant un pré-print, un post-print, une landing page, un set de métadonnées, etc.)

3.7 Les Linked Data en art et design

3.7.1 Projets LOD

20 Accessible sous : <http://www.w3.org/ns/dcat> (consulté le 09 août 2017)

21 Accessible sous : <http://vivoweb.org/ontology/core> (consulté le 09 août 2017)

22 Accessible sous : <http://usefulinc.com/ns/doap#> (consulté le 09 août 2017)

23 Accessible sous : <http://purl.org/linked-data/cube> (consulté le 15 août 2017)

24 Accessible sous : <http://www.openarchives.org/ore/terms/> (consulté le 15 août 2017)

Les projets en LOD dans les sciences humaines émanent majoritairement de grandes institutions d'Europe et d'Amérique du Nord telles que le Rijksmuseum, le Getty Research Institute²⁵, le National Gallery of Art ou encore Europeana²⁶ qui a notamment ouvert tous leurs sets de données en licence CC0 et qui peuvent être manipulées par le biais d'une API ou avec des requêtes SPARQL (6).

→ Voir projet LODZ

3.7.2 IIIF

Les universitaires s'appuient de plus en plus sur les dépôts d'images numériques pour leur recherche car ils contiennent un accès facilité aux objets culturels, patrimoniaux et scientifiques tels que des manuscrits, des peintures, des images en trois dimensions, des journaux, des cartes ou autres matériaux fragiles et rares. Les bibliothèques, les centres d'archives et les musées ont renforcé l'accès en ligne à leurs collections numériques, mais généralement ces dernières sont contenues dans des silos bâtis avec des serveurs et interfaces utilisateur incompatibles avec d'autres systèmes, bloquant ainsi une interopérabilité facilitée entre différentes institutions et limitant les possibilités d'interaction des usagers finaux. Les institutions travaillant indépendamment ont bien souvent du mal à maintenir des dépôts d'images numériques face à un environnement où les usagers finaux ont des besoins en constante évolution et dans un contexte de migration des contenus et de la progression rapide des technologies.

Depuis 2012, l'initiative International Image Interoperability Framework (IIIF) permet aux institutions détenant des images numériques²⁷ et fonds numérisés de ne publier et stocker qu'une fois leurs fichiers et métadonnées sérialisées en JSON-LD grâce à l'utilisation des APIs de présentation et d'image. Ceci permet une interopérabilité entre différentes institutions qui se conforment aux spécifications. Cela offre ainsi aux usagers la possibilité de manipuler, annoter et comparer des images²⁸. (7).

25 Dans le domaine des Linked Data, le Getty est surtout connu pour avoir développé des vocabulaires (ULAN, TGN, AAT, Iconography Authority): <http://www.getty.edu/research/tools/vocabularies/lod/> (consulté le 10 août 2017)

26 <http://labs.europeana.eu/api/linked-open-data-introduction> (consulté le 10 août 2017)

27 Une API spécifique aux ressources audiovisuelles est également en train d'être élaborée par la communauté IIIF.

28 Accessible sous: <https://iiif.io/> (consulté le 10 août 2017)

4 Bibliographie

1. WILJES, Cord, JAHN, Najko, LIER, Florian, PAUL-STUEVE, Thilo, VOMPRAS, Johanna, PIETSCH, Christian et CIMIANO, Philipp. Towards Linked Research Data: An Institutional Approach. In : *3rd Workshop on Semantic Publishing* [en ligne]. Montpellier, France, 26 mai 2013. pp. 27-38. [Consulté le 3 août 2017]. Disponible à l'adresse : <https://pub.uni-bielefeld.de/publication/2580621>
2. BRASE, Jan. DataCite and linked data. *JLIS.it*. 15 janvier 2013. Vol. 4, n° 1, pp. 365. DOI <http://dx.doi.org/10.4403/jlis.it-5493>.
3. LEE, Cheng-Jen, WEI-CHING HUANG, Andrea et CHUANG, Tyng-Ruey. Metadata as Linked Data for Research Data Repositories. In : *International Symposium on Grids & Clouds 2017*. Taipei, Taiwan, 9 mars 2017.
4. MICHAEL, Conlon et AMIR, Aryani. Creating an open linked data model for Research Graph using VIVO Ontology. In : *Open Repository Conference* [en ligne]. Brisbane : (preprint), 24 juillet 2017. [Consulté le 9 août 2017]. Disponible à l'adresse : https://figshare.com/articles/Creating_an_open_linked_data_model_for_Research_Graph_using_VIVO_Ontology/4757917
5. VIVO 1.9.x documentation. *DuraSpace wiki* [en ligne]. 18 avril 2017. [Consulté le 9 août 2017]. Disponible à l'adresse : <https://wiki.duraspace.org/display/VIVODOC19x/>
6. MAYER, Allana. Linked Open Data for Artistic and Cultural Resources. *Art Documentation: Journal of the Art Libraries Society of North America*. 1 mars 2015. Vol. 34, n° 1, pp. 2-14. DOI 10.1086/680561.
7. SNYDMAN, Stuart, SANDERSON, Robert et CRAMER, Tom. The International Image Interoperability Framework (IIIF): A community & technology approach for web-based images. In : *Archiving Conference* [en ligne]. Los Angeles, CA, mai 2015. pp. 16-21. Disponible à l'adresse : <https://purl.stanford.edu/df650pk4327>
8. CEOS. *Persistent Identifiers Best Practices* [en ligne]. janvier 2016. Committee on Earth Observation Satellites. [Consulté le 24 juillet 2017]. Disponible à l'adresse : <https://earth.esa.int/documents/1656065/2265358/CEOS-Persistent-Identifier-Best-Practices>
9. BERNERS-LEE, Tim, FIELDING, Roy et MASINTER, Larry. Uniform Resource Identifier (URI): Generic Syntax (RFC 3986). *The Internet Engineering Task Force* [en ligne]. janvier 2005. [Consulté le 21 juin 2014]. Disponible à l'adresse : <http://www.ietf.org/rfc/rfc3986>
10. W3C. Architecture of the World Wide Web: volume one. *World Wide Web Consortium* [en ligne]. 15 décembre 2004. [Consulté le 26 juillet 2017]. Disponible à l'adresse : <https://www.w3.org/TR/2004/REC-webarch-20041215/>
11. INTERNATIONAL DOI FOUNDATION. DOI system and internet identifier specifications. *Digital Object Identifier system* [en ligne]. 6 mai 2015. [Consulté le 26 juillet 2017]. Disponible à l'adresse : <https://www.doi.org/factsheets/DOIIdentifierSpecs.html>
12. HAKALA, Juha. Persistent identifiers: an overview. *Technology watch report* [en ligne]. 13 octobre 2010. [Consulté le 19 juillet 2017]. Disponible à l'adresse : <http://www.metadaten-twr.org/2010/10/13/persistent-identifiers-an-overview/>

13. BERMÈS, Emmanuelle. *Des identifiants pérennes pour les ressources numériques: l'expérience de la BNF* [en ligne]. 5 mai 2006. Bibliothèque nationale de France. [Consulté le 4 août 2017]. Disponible à l'adresse : http://www.bnf.fr/documents/ark_presentation_bermes_2006.pdf
14. NICHOLAS, Nick, WARD, Nigel et BLINCO, Kerry. A policy checklist for enabling persistence of identifiers. *D-Lib magazine* [en ligne]. janvier 2009. Vol. 15, n° 1/2. [Consulté le 7 août 2017]. DOI 10.1045/january2009-nicholas. Disponible à l'adresse : <http://www.dlib.org/dlib/january09/nicholas/01nicholas.html>
15. BAZZANELLA, Barbara, BORTOLI, Stefano et BOUQUET, Paolo. Can persistent identifiers be cool? *International journal of digital curation*. 14 juin 2013. Vol. 8, n° 1, pp. 14-28. DOI 10.2218/ijdc.v8i1.246.
16. OCLC. Introduction to Persistent Uniform Resource Locators. In : *Internet Society* [en ligne]. Montréal : Internet Society, 24 juin 1996. [Consulté le 26 juillet 2017]. Disponible à l'adresse : https://www.isoc.org/inet96/proceedings/a4/a4_1.htm
17. ZIERAU, Eld, NYVANG, Caroline et HVID KROMANN, Thomas. Persistent Web References - Best Practices and New Suggestions. In : *Proceedings of the 13th International Conference on Digital Preservation* [en ligne]. Bern, Switzerland, octobre 2016. pp. 237-246. [Consulté le 3 août 2017]. Disponible à l'adresse : http://www.ipres2016.ch/frontend/organizers/media/IPRES2016/_PDF/IPR16.Proceedings_4_Web_Broschuere_Link.pdf
18. SIMONS, Natasha. Unpacking persistent identifiers for research. *VALA Tech Camp* [en ligne]. Melbourne. 13 juillet 2017. [Consulté le 25 juillet 2017]. Disponible à l'adresse : <https://www.slideshare.net/AustralianNationalDataService/unpacking-persistent-identifiers-for-research>
19. BELLINI, Emanuele, CIRINNÀ, Chiara, LANCIA, Maurizio, LUNGHI, Maurizio, PUCCINELLI, Roberto, SACCONI, Massimiliano, SEBASTIANI, Brunella et SPASIANO, Marco. Persistent identifier distributed system for digital libraries. In : *World Library and Information Congress 2009, 75th IFLA General Conference and Assembly* [en ligne]. Milan : IFLA, 23 août 2009. pp. 9. [Consulté le 31 juillet 2017]. Disponible à l'adresse : <https://www.ifla.org/past-wlic/2009/193-lunghi-en.pdf>
20. HAKALA, Juha et INTERNET SOCIETY. Using National Bibliography Numbers as Uniform Resource Names (RFC 3188). *Internet Engineering Task Force* [en ligne]. octobre 2001. [Consulté le 31 juillet 2017]. Disponible à l'adresse : <https://tools.ietf.org/html/rfc3188>
21. BIBLIOTHÈQUE NATIONALE SUISSE. *e-Helvetic: manuel URN* [en ligne]. 27 octobre 2016. Bibliothèque nationale suisse. [Consulté le 15 août 2017]. Disponible à l'adresse : http://www.nb.admin.ch/nb_professionnel/01693/01695/01706/index.html
22. SOMMER, Dorothea. Persistent identifiers: the « URN granular » project of the German National Library and the University and State Library Halle. *LIBER quarterly* [en ligne]. 9 février 2010. Vol. 19, n° 3-4. [Consulté le 8 août 2017]. DOI 10.18352/lq.7965. Disponible à l'adresse : <http://www.liberquarterly.eu/articles/10.18352/lq.7965/>
23. SCHWARDMANN, Ulrich. Trainingskurs Arbeiten mit PIDs: Teil 1. *RDA DE Training Workshop* [en ligne]. Dresden. 8 juin 2017. [Consulté le 24 juillet 2017]. Disponible à l'adresse : http://www.forschungsdaten.org/images/8/81/2017-06-08-Trainingskurs_Arbeiten_mit_PIDs_Teil1-Schwardmann.pdf
24. LANNOM, Larry. DONA Foundation, Administering the Global Handle Registry (GHR). [en ligne]. 4 octobre 2015. [Consulté le 25 juillet 2017]. Disponible à l'adresse : <https://zenodo.org/record/31784#.WXcXmHq68ek>
25. ORGANISATION INTERNATIONALE DE NORMALISATION. *Information et documentation: système d'identifiant numérique d'objet*. 1ère éd. Genève : ISO, 2012. ISO, 26324:2012(F).

26. INTERNATIONAL DOI FOUNDATION. Digital Object Identifier system handbook. *Digital Object Identifier system* [en ligne]. 22 février 2016. [Consulté le 7 août 2017]. Disponible à l'adresse : <https://www.doi.org/hb.html>
27. EPIC. Persistent Identifiers for eResearch. [en ligne]. 2017. [Consulté le 7 août 2017]. Disponible à l'adresse : <http://www.pidconsortium.eu/>
28. BERNERS-LEE, Tim. Cool URIs don't change. *World Wide Web Consortium* [en ligne]. 1998. [Consulté le 25 juillet 2017]. Disponible à l'adresse : <https://www.w3.org/Provider/Style/URI.html.en>
29. ISNI. FAQ. *International Standard Name Identifier* [en ligne]. 2017. [Consulté le 4 août 2017]. Disponible à l'adresse : <http://www.isni.org/content/faq>
30. ANDS. Persistent identifiers: expert level. *Australian National Data Service* [en ligne]. 2017. [Consulté le 2 août 2017]. Disponible à l'adresse : <http://www.ands.org.au/guides/persistent-identifiers-expert>
31. SIMONS, Natasha. Implementing DOIs for research data. *D-Lib magazine* [en ligne]. mai 2012. Vol. 18, n° 5/6. [Consulté le 31 juillet 2017]. DOI doi:10.1045/may2012-simons. Disponible à l'adresse : <http://www.dlib.org/dlib/may12/simons/05simons.html>
32. TRELOAR, Andrew. *Private research, shared research, publication and the boundary transitions* [en ligne]. 19 mars 2012. [Consulté le 3 août 2017]. Disponible à l'adresse : http://andrew.treloar.net/research/diagrams/data_curation_continuum.pdf
33. ANDS. *What identifier do I need: a pathfinder for data managers* [en ligne]. 2017. Australian National Data Service. [Consulté le 2 août 2017]. Disponible à l'adresse : http://www.ands.org.au/_data/assets/pdf_file/0010/386884/id_decision_tree.pdf