



HAL
open science

Identifiants des documents numériques : ISBN, ISSN, URL, DOI, OpenURL...

Jean-Luc Archimbaud

► To cite this version:

Jean-Luc Archimbaud. Identifiants des documents numériques : ISBN, ISSN, URL, DOI, OpenURL....
2014. sic_01068135v1

HAL Id: sic_01068135

https://archivesic.ccsd.cnrs.fr/sic_01068135v1

Preprint submitted on 25 Sep 2014 (v1), last revised 26 Feb 2015 (v2)

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Identifiants des documents numériques : ISBN, ISSN, URL, DOI, OpenURL...

Auteur : [Jean-Luc Archimbaud](#) (Mathdoc) - Jean-Luc.Archimbaud@ujf-grenoble.fr - 22 sept 2014

Version en ligne (qui sera peut être plus récente) :

https://www.portail-math.fr/doc_utile/identifiants-docs-numeriques

Résumé

Description et caractéristiques des identifiants actuellement les plus utilisés pour les documents numériques, livres, publications en série (journaux, périodiques et collections) et articles, accessibles par réseau : ISBN, ISSN, URL, DOI, OpenURL, OAI, ARK, URI et de certains identifiants spécifiques à des bases d'informations connues.

Cible

Ce document est destiné aux personnes qui travaillent dans les laboratoires de recherche, en particulier de maths, et qui veulent connaître et comprendre le fonctionnement des principaux identifiants de documents numériques, par curiosité ou besoin professionnel, sans être des experts du domaine. Il se veut didactique et par conséquent il est parfois simplificateur.

Les identifiants

Pour de multiples raisons, il est très intéressant de pouvoir identifier chaque document numérique de manière unique.

Notre propos ici n'est pas d'expliquer ces avantages mais de décrire les principales méthodes et les codages associés utilisés actuellement pour identifier les documents numériques accessibles par réseau, en se limitant aux livres, publications en série (journaux, périodiques et collections) et articles scientifiques. Il se limite à ces documents numériques mais certaines de ces méthodes peuvent être utilisées pour d'autres objets. Les principes et les limitations restent les mêmes.

A noter aussi qu'il n'est pas exhaustif et se focalise sur ce qui est utilisé pour la documentation scientifique en maths.

L'identification semble une opération simple à réaliser, mais pour une utilité maximale, il faut que celle-ci permette :

1. à partir d'un identifiant, de désigner un seul document (et pas plusieurs documents différents, ni des versions différentes du document),
2. d'identifier tous les types de documents (livres, publications en série et articles scientifiques pour notre propos),
3. à toute personne d'identifier ses documents. La méthode doit être universelle, utilisable partout dans le monde, par tous à coût nul ou très minime. Elle doit donc être sans contrainte d'appartenance à une organisation,
4. d'accéder rapidement au document (pour un utilisateur sans connaissance technique du domaine de la documentation),
5. d'identifier le document de manière pérenne dans le temps. L'identifiant, par exemple, ne doit pas

- dépendre de l'existence d'une machine physique (car celle-ci peut être parfois inaccessible, voire disparaître),
6. d'identifier un document de manière unique. Un document ne doit pas avoir plusieurs identifiants différents. Ceci permet de pouvoir identifier tout ce qui se rapporte au document : si un document a plusieurs identifiants, on ne pourra pas obtenir ce résultat.

Ces 6 points seront appelés des fonctionnalités dans la suite. Ils sont donnés simplement pour mieux cerner le périmètre et les limitations de chaque type d'identification, limitations rarement exprimées. Ils peuvent permettre de mieux choisir l'identification à mettre en place pour l'ensemble de documents que l'on gère. Mais ils n'ont pas pour objectif de classer les différents types d'identification entre eux.

Dans la vie des documents, comme les articles, il y a généralement plusieurs versions (soumise, pré-publiée, publiée...) du même objet. Il serait intéressant d'avoir un identifiant qui ferait le lien entre ces versions, objet d'une septième fonctionnalité. Mais dans les systèmes de notation ci-dessous, aucun ne semble faire ce lien. Généralement les versions auront des identifiants différents, éventuellement même d'un type différent selon l'état du document.

Les méthodes et codages que nous allons décrire (avec une grille d'analyse par rapport aux 6 fonctionnalités ci-dessus), sont : ISBN, ISSN, URL, DOI, OpenURL, OAI, ARK, URI (ce sont actuellement les plus répandues) avant un tableau récapitulatif des 6 fonctionnalités pour ces identifiants et certains identifiants spécifiques à des bases d'informations connues. Une liste de quelques documentations sur le sujet termine ce texte.

ISBN

Monsieur Wikipedia nous dit :

*ISBN, International Standard Book Number (ISBN) ou Numéro international normalisé du livre (NINL) est un numéro international qui **permet d'identifier**, de manière unique, chaque édition de **chaque livre publié**, que son support soit numérique ou sur papier. Il est destiné à simplifier la gestion informatique pour tous les intervenants de la chaîne du livre (imprimeur, éditeur, libraire, bibliothèque, etc).*

ISBN n'identifie donc que les ouvrages et pas les journaux, ni les articles. À noter qu'il identifie une édition particulière. Ainsi un livre peut avoir plusieurs ISBN : un pour sa diffusion papier, un autre pour la version numérique (PDF)...

Codage des identifiants

- **ISBN10** : jusqu'en 2006 le codage, appelé ISBN10, a été sous la forme de 10 chiffres présentés en 4 segments A - B - C - K
 - A indique une zone géographique (de manière pas très logique). Exemples : 2 désigne les pays francophones (France, Belgique, Canada francophone, Luxembourg, Suisse francophone), 982 les pays du Pacifique Sud, 4 le Japon.
 - B identifie l'éditeur (de 1 à 7 chiffres).
 - C correspond au numéro d'ordre de l'ouvrage chez l'éditeur.
 - K est un code de vérification fonction de A, B et C.

Exemple : ISBN 2-271-05256-4 désigne un livre publié en France (A=2), par l'éditeur CNRS Edition qui a le code 271 (B=271). Ce livre est le 5256 ième chez cet éditeur.

- **ISBN13 (EAN-13)** : la plage de numérotation ISBN10 arrivant à saturation, à partir de 2007, une numérotation à 13 chiffres est imposée, ISBN13 qui découle du code à barre à la norme EAN-13. Le codage sur 13 chiffres est noté G - A - B - C - K.
 - G a la valeur 978 ou 979
 - A, 2 ou 3 chiffres, indique un pays
 - B, 4 ou 5 chiffres, indique un éditeur
 - C, 4 ou 5 chiffres, indique le numéro d'ouvrage chez l'éditeur
 - K, est une clé de contrôle fonction des 4 autres codes

G=978 indique que ce qui suit est un code ISBN10, ce qui permet d'intégrer les ouvrages identifiés avec le codage ISBN10 automatiquement avec la nouvelle notation. Ainsi ISBN 2-271-05256-4 devient ISBN 978-2-271-05256-8.

Les éditeurs qui avaient déjà obtenu un code B du format ISBN10 peuvent continuer à l'utiliser jusqu'à épuisement de leur plage d'adresses, avec A=978. Exemple : ISBN 978-2-9539187-9-3. Lorsque celui-ci est épuisé, ils basculent dans G=979.

G=979 est utilisé pour les nouveaux éditeurs ou pour ceux qui ont épuisé leur plage de codage ISBN10.

A=10 est le nouveau code pour les pays francophones (France, Belgique, Canada francophone, Luxembourg, Suisse francophone) dans ISBN13. Ainsi les nouveaux ouvrages édités dans un pays francophone ont maintenant un ISBN qui commence par 978-2 ou 979-10.

Affectation des identifiants

L'[Agence francophone pour la numérotation internationale du livre](#) (AFNIL) attribue les codes ISBN (et EAN) à tous les éditeurs français, belges francophones et des pays d'Afrique francophones. Le système d'attribution mondial est organisé par l'[International ISBN Agency](#) qui a son siège à Londres. Pour l'AFNIL, peuvent être éditeurs : maisons d'éditions, sociétés non spécialisées dans l'édition, associations, fondations, fédérations, syndicats, collectivités (exemples : mairies, écoles, collèges, lycées), auto-entrepreneurs, autres organismes.

Le service est gratuit.

Caractéristiques par rapport aux 6 fonctionnalités

Fonctionnalités présentes :

- L'affectation des numéros ISBN étant centralisé par zone géographique, le point 1 est validé : un identifiant désigne un seul document et pas plusieurs différents.
- L'obtention d'un identifiant est gratuit, ouvert à tous, partout dans le monde. Le point 3 est validé.
- L'identification dans le temps est pérenne. Le point 5 est validé.

Fonctionnalités presque présentes :

- On peut considérer qu'un identifiant ISBN est unique pour un livre, en ne prenant pas en compte la possibilité d'avoir un identifiant différent pour chaque publication du livre (imprimée, PDF en ligne...). Donc le point 6 est presque validé.

Fonctionnalités absentes :

- Le point 2 n'est pas rempli. Ce codage est réservé aux livres.
- À partir d'un numéro ISBN, pour accéder au livre, ce n'est pas immédiat. Des sites plus ou moins commerciaux tels que http://www.bookfinder.com/isbn_search/ ou <http://isbndb.com/> permettent, à partir d'un numéro ISBN, de trouver les caractéristiques du livre (titre, auteur...) mais leur base

est certainement incomplète. De même des bases de données bibliographiques manipulent ce numéro. Ensuite une recherche avec un moteur de recherche permettra certainement de trouver le livre en ligne. Donc le point 4 n'est pas impossible mais très difficile pour un non professionnel de la documentation.

Utilisation

Les numéros ISBN sont utilisés depuis très longtemps pour les livres imprimés commercialisés. Actuellement, le [dépôt légal](#), assuré en France par la BNF, est "obligatoire pour les livres, quel que soit leur procédé technique de production, d'édition ou de diffusion, dès lors qu'ils sont mis à la disposition d'un public qui excède le cercle de famille, à titre gratuit ou onéreux" (extrait du site de la BNF). Ce dépôt pour les livres demandant un numéro d'ISBN on peut en déduire que le numéro ISBN est obligatoire pour tous les ouvrages, imprimés ou numériques, même diffusés gratuitement. Mais cette obligation n'est pas toujours respectée en particulier pour les ouvrages numériques qui n'entrent pas dans le cadre d'une production formelle (initiative individuelle...).

ISSN

L'International Standard Serial Number (ISSN) **identifie une publication en série** (journal, périodique ou collection). Il peut aussi être affecté aux sites web, blogs et bases de données en ligne qui répondent à certains critères.

Un ISSN est attribué à une ressource continue sur un support donné. Il est lié au titre. Si la revue change de titre, un nouveau numéro doit être attribué. Si une revue est diffusée sur plusieurs supports, il faut autant de numéros que de support. Ainsi la revue "Les Annales de l'Institut Fourier" a comme code ISSN 0373-0956 pour la version papier et ISSN 1777-5310 pour la version électronique. L'instance française d'attribution des numéros ISSN recommande d'avoir un numéro ISSN pour les copies électroniques de périodiques qui ont cessé de paraître.

L'agence internationale ISSN gère aussi un autre identifiant, ISSN-L, qui permet de relier toutes les déclinaisons (version papier, version électronique...) d'une publication en série. Le numéro ISSN-L choisi est le numéro ISSN attribué à la plus ancienne version de la publication. Par exemple ISSN-L 0373-0956 identifie "Les Annales de l'Institut Fourier" avec l'information que cette revue a 2 versions qui ont 2 numéros ISSN différents : ISSN 1777-5310 et ISSN 0373-0956.

Codage des identifiants

Le code est formé de 2 séquences de 4 chiffres séparées par un tiret, précédées par ISSN. Exemple : ISSN 0994-7647. Le dernier chiffre est une clé de contrôle, qui peut prendre la valeur X pour 10.

Les numéros ISSN sont attribués séquentiellement, indépendamment du pays d'origine. Pour pallier le problème d'une numérotation identique affectée à plusieurs périodiques (7 chiffres, c'est très peu), un titre-clé est associé à chaque numéro ISSN créé. Un titre-clé est classiquement le nom du journal avec éventuellement entre parenthèses la région. Peut aussi être ajouté une période de publication (année de début - année de fin de publication), l'année de fin 9999 signifie que la revue continue de paraître.

Exemple : à l'ISSN 0994-7647 est associé le titre-clé "JOURNAL DU CNRS".

Affectation des identifiants

En France, l'ISSN est attribué par le "Centre national d'enregistrement des publications en série : ISSN France" à la BNF. Une [partie du site de la BNF](#) indique les modalités. Le service est gratuit.

Au niveau international, le CIEPS ou [ISSN International Centre](#), situé à Paris, coordonne l'affectation. Lors de la demande, le titre-clé est choisi et un ensemble d'informations (nom de la revue, pays, années de publication, éditeur, URL ou email...) sur la publication en série doivent être données. Ceci constitue une notice. Cet ensemble de notices est regroupé dans le Registre International de l'ISSN (plus d'1.7 million d'entrées à l'été 2014). Ce registre est accessible avec un abonnement, par différentes méthodes. Les

centres qui gèrent des bases bibliographiques sont abonnés à ce service.

Caractéristiques par rapport aux 6 fonctionnalités

Fonctionnalités présentes :

- L'obtention d'un identifiant est gratuit, ouvert à tous, partout dans le monde. Le point 3 est validé.
- Le code ISSN est pérenne. Le point 5 est validé.

Fonctionnalités presque présentes (mais selon l'interprétation, les 2 points ci-dessous peuvent basculer dans l'ensemble 'fonctionnalités présentes') :

- Un identifiant complet ISSN (numéro + titre-clé) désigne un seul journal. Donc le point un est rempli. Mais si on ne dispose que du numéro, il n'est pas rempli.
- Si la publication en série n'a pas eu de modification (changement de titre, de support...) elle n'a qu'un seul identifiant ISSN. Le point 6 est donc rempli. Par contre si elle a eu ce type de modification (ce qui est courant), elle aura plusieurs numéros ISSN. Donc le point 6 n'est dans ce cas pas complètement rempli.

Fonctionnalités absentes :

- Le point 2 n'est pas rempli. Ce codage est réservé aux publications en série.
- À partir d'un numéro ISSN, pour accéder à la publication en série, ce n'est pas immédiat. On peut éventuellement passer par un moteur de recherche. On peut aussi y accéder via une base de données bibliographiques abonnée au registre international ISSN. Donc le point 4 n'est pas impossible, mais très difficile pour un non professionnel de la documentation.

Utilisation

Les numéros ISSN sont très utilisés depuis longtemps dans les périodiques imprimés et les revues électroniques les ont aussi adoptés. Ils doivent figurer sur l'édition imprimée des périodiques et sur la page d'accueil des revues en ligne.

URL

L'URL, Uniform Resource Locator, est l'identifiant utilisé par tous les navigateurs internet. Il **indique la localisation de l'objet** ainsi que la méthode (le protocole de communication) pour y accéder.

Codage des identifiants

La structure est : "Méthode pour accéder" : "/" "localisation de l'objet".

Exemple : ce document a pour URL "https://www.portail-math.fr/doc_utilite/identifiants-docs-numeriques". Sa localisation est "www.portail-math.fr/doc_utilite/identifiants-uniques" : le document est le fichier "identifiants-uniques" dans le répertoire "doc_utilite" sur la machine "[www.portail-math](http://www.portail-math.fr)". La méthode pour accéder à l'objet est le protocole "https".

À noter que la partie répertoire/fichier peut être remplacée par un appel à un programme ou d'autres notations qui permettent d'accéder à l'objet.

Affectation des identifiants

Elle se fait en 2 parties :

- Le nom de domaine (portail-math.fr dans l'exemple) est délivré par un organisme officiel tel que l'[AFNIC](#) en France (Association Française pour le Nommage Internet en Coopération), qui gère entre autres le "top level domain" .fr. Au niveau international l'[ICANN](#) (Internet Corporation for Assigned Names and Numbers) coordonne l'attribution des noms de domaine.
- Ensuite ce sont les informaticiens du site qui fixent le protocole d'accès (comme https), les répertoires (comme doc_utilite) et le nom du document (comme identifiants-uniques).

Caractéristiques par rapport aux 6 fonctionnalités

Fonctionnalités présentes :

- L'affectation de ces URL est unique. Les noms de domaine sont uniques, ainsi que la partie noms de répertoire, documents, appels de fonction... qui est à la discrétion de chaque site. Le point 1 est validé.
- Les différents types de documents sont supportés par la notation URL. Le point 2 est vérifié.
- L'obtention d'un identifiant, concrètement d'un nom de domaine est simple et peu coûteux. De plus certains fournisseurs d'accès internet ou d'autres offrent gratuitement des espaces de stockage de documents accessibles avec la notation URL. Le point 3 est validé.
- L'URL, étant utilisée par les navigateurs, permet d'accéder très rapidement au document : le point 4 est validé.

Fonctionnalités absentes :

- L'URL est dépendante d'une machine dans un domaine (www.portail-math.fr dans l'exemple) et de la gestion des fichiers sur cette machine. Si le site change de nom de domaine, la machine est indisponible ou disparaît..., l'URL ne permettra plus d'accéder au document. Le point 5 n'est pas validé. On peut très fortement atténuer ce risque en stockant les documents numériques, en particulier les articles, sur des sites institutionnels qui feront le nécessaire pour conserver le système de nommage, comme les sites d'archives ouvertes (dont on peut penser que la durée de vie est supérieure à celle d'un laboratoire par exemple). De plus avec les techniques actuelles, derrière un seul nom de machine il peut y avoir un ensemble de serveurs répartis dans le monde par exemple (www.google.fr) et donc l'URL ne dépend pas d'une seule machine.
- L'URL n'identifie pas un document de manière unique. En effet, le document (comme cet article) peut être déposé sur un autre serveur et donc avoir deux identifiants, sans que l'on sache que les deux désignent un même objet. Le point 6 n'est pas validé.

Utilisation

Malgré ces inconvénients la notation URL est très largement utilisée, surtout par les non-professionnels du domaine de la documentation et pour les ressources qui ne sont pas déposées sur des sites de référence.

DOI

DOI, Digital Object Identifier, est un mécanisme d'**identification de ressources numériques** documentaires ou autres. C'est à la fois un **identifiant pérenne** (qui ne dépend pas de l'emplacement de la ressource en ligne), et un **mécanisme pour retrouver la ressource en ligne** (c'est à dire son URL) à partir de son identifiant DOI. Certaines métadonnées obligatoires décrivant la ressource sont demandées lors de l'affectation de chaque DOI (cf plus loin) permettant ainsi à ce mécanisme d'être une mini base bibliographique de ressources.

Comme exemples de ressources documentaires, un article, un livre, chaque chapitre d'un livre peuvent avoir un DOI.

Le mécanisme suit la spécification technique "Handle System" qui définit comment attribuer, gérer et

résoudre des identifiants persistants attribués à des objets numériques et à d'autres ressources Internet. Un DOI est donc un cas particulier d'identifiant Handle.

Le DOI vient du monde de l'édition traditionnel, l'AAP (Association of American Publishers) en est l'initiateur. Il est maintenant ouvert à d'autres organisations. Il est décrit entre autres dans les documents standards internet : [RFC 3650](#), [RFC 3651](#) et [RFC3652](#).

Codage des identifiants - service de localisation

Formellement, le code DOI est de la forme : doi:"préfixe"/"suffixe". La chaîne "doi:" est facultative. Le préfixe est attribué à une organisation (voir ci-après affectation des identifiants), qui fait fonction de bureau d'enregistrement où la ressource a été enregistrée. Le suffixe est l'identifiant de la ressource numérique dans cette organisation, identifiant dont la forme dépend de chaque organisation.

Exemples :

- doi: 10.5802/aif.2754 identifie un article désigné "aif.2754" dans l'organisation qui a le préfixe 10.5802, c'est à dire CEDRAM à [Mathdoc](#). "aif.2754" est l'identifiant choisi par Mathdoc dans CEDRAM. Il s'agit d'un article paru dans les Annales de l'Institut Fourier.
- doi:10.1016/j.physletb.2010.09.059 identifie un article désigné "j.physletb.2010.09.059" dans l'organisation qui a le préfixe "10.1016", c'est à dire l'éditeur [Elsevier](#). "j.physletb.2010.09.059" est l'identifiant choisi par Elsevier. Il s'agit d'un article de la revue Physics Letters B.
- doi:10.6096/BVET.IND.METEO identifie un article "BVET.IND.METEO" enregistré par l'[INIST](#), qui a le préfixe "10.6096".
- doi: 10.4000/cybergeo.23737 identifie un article "cybergeo.23737", un article de la revue Cybergeo enregistré par [Le Cléo](#) qui a le préfixe "10.4000".

Un serveur permet de faire le lien DOI - URL pour retrouver la ressource en ligne : <http://dx.doi.org/>

- On peut l'interroger en ligne en saisissant un identifiant DOI dans une fenêtre ou
- Accéder directement à la ressource en accolant le DOI à l'URL du serveur comme <http://dx.doi.org/10.5802/aif.2754> ou <http://dx.doi.org/10.1016/j.physletb.2010.09.059> pour reprendre 2 exemples précédents.

Pour cette raison, certains indiquent le DOI sous la forme : <http://dx.doi.org/>"préfixe"/"suffixe"

Certains navigateurs comme firefox permettent (à travers un plugin à installer) l'accès direct à une ressource en entrant l'URL doi:...

Affectation des identifiants

Environ 10 000 organisations (données de 2013) sont habilitées à délivrer des DOI. Ces organisations répondent à certains critères et acquittent une cotisation (forfaitaire ou par numéro attribué). Elles sont adhérentes à une des neuf (en 2014) "Registration agencies". Ces agences sont regroupées dans l'[International DOI Foundation](#) (IDF). Ces agences délivrent des numéros de préfixes à leurs membres qui peuvent ainsi être "bureau d'enregistrement". Les éditeurs traditionnels font partie des organisations (et donc membres d'une de ces agences) mais aussi d'autres institutions. Ainsi CEDRAM de Mathdoc (documentation scientifique en maths) et le Cléo (documentation scientifique en SHS) sont membres de l'agence CrossRef et l'INIST (documentation scientifique toutes disciplines) de DataCite. Ces trois organisations délivrent donc des DOI, chacune pour des ressources de leur domaine.

Le service est ainsi payant et chaque bureau d'enregistrement est réservé à des utilisateurs spécifiques. De plus chaque bureau d'enregistrement a sa politique pour accepter ou non l'affectation de DOI suivant le type du document. Par exemple CrossRef n'accepterait pas l'affectation de DOI aux articles dans HAL, alors que d'autres bureaux le feraient peut-être.

Lors de l'affectation d'un code DOI, des informations sur la ressource numérique sont collectées : le nom complet, le nom court, l'ISSN, l'URL... pour une revue, le titre, la langue, la date, les auteurs, l'URL... pour un article...

À noter que pour qu'un identifiant DOI soit pérenne, il faut que le producteur de la ressource informe le bureau d'enregistrement si la ressource change d'URL...

Caractéristiques par rapport aux 6 fonctionnalités

Fonctionnalités présentes :

- Un identifiant DOI désigne un seul document. Le point 1 est validé.
- Le point 2 est rempli. Ce codage permet d'identifier tous types de documents en particulier livres, publications en série et articles scientifiques.
- À partir d'un DOI, en particulier avec la notation <http://dx.doi.org/>"préfixe"/"suffixe" il est très simple d'accéder au document, même sans connaissance en documentation. Le point 4 est satisfait.
- Le code DOI est pérenne. Le point 5 est validé.
- Un identifiant DOI désigne un seul document de manière unique. Une ressource documentaire ne peut être déclarée que par un seul bureau d'enregistrement. Le point 6 est rempli.

Fonctionnalités absentes :

- L'obtention d'un identifiant DOI n'est pas gratuit et nécessite de répondre à des critères imposés par au moins une organisation bureau d'enregistrement. Le point 3 n'est pas validé.

Utilisation

Des chiffres d'avril 2013 font état de 85 millions de numéros DOI attribués à travers 9 500 organisations. Ce codage est donc très utilisé pour les documents publiés par les éditeurs ou organisations institutionnelles connues. Vu le coût par numéro ce n'est pas un identifiant utilisé pour tous les documents.

OpenURL

D'après mes lectures, on parle d'OpenURL de manières très diverses, selon la date de la rédaction, l'approche théorique ou pragmatique et le degré d'optimisme de l'auteur sur le déploiement de ce mécanisme. Ce qui suit est une définition personnelle, inspirée de ce que j'ai lu, mais qui est criticable.

L'objectif d'OpenURL est de **délivrer aux utilisateurs des ressources** (ou des informations sur les ressources) **auxquelles ils ont le droit d'accéder**. Ceci peut s'appliquer à des objets numériques divers, nous nous limiterons dans ce qui suit aux ressources documentaires.

Concrètement OpenURL est basé sur 2 éléments :

- une syntaxe qui permet, dans une URL, de transporter des informations sur une ressource documentaire (métadonnées qui permettent d'identifier la ressource) et éventuellement des informations sur la personne qui émet la demande, pour l'identifier et en déduire les droits de cette personne sur la ressource. L'URL est alors appelé OpenURL. La syntaxe des informations sur la ressource dans l'URL (la description bibliographique sous forme de métadonnées) est définie par la norme NISO Z39.88.
- un mécanisme qui permet, avec cette URL, de délivrer à la personne tout ou partie de la ressource ou des informations pour accéder à la ressource selon ses droits. Ce mécanisme est assuré par un

serveur (appelé link resolver, OpenURL link server, serveur de liens...). En l'absence d'information sur la personne dans l'URL, l'identification du demandeur par le serveur peut être effectuée par numéro IP, login-mot de passe, certificat électronique, cookie ou autres méthodes d'identification pratiquées sur le web. Dans la syntaxe, la première partie de l'OpenURL est ainsi le nom de ce serveur et la seconde partie les informations sur la ressource recherchée et éventuellement le demandeur.

La norme NISO Z39.88 permet de spécifier les métadonnées, par exemple : le titre d'un article, d'une revue, les numéros ISSN ou DOI, l'auteur, la date de publication (cf exemple ci-dessous). Chaque serveur OpenURL suit cette norme ce qui peut permettre de rendre interopérables des services documentaires hétérogènes qui échangent des métadonnées via OpenURL et utilisent des résolveurs locaux pour faire aboutir les requêtes.

OpenURL n'est pas à la base un identifiant mais la bonne combinaison de métadonnées indiquées peut en faire un identifiant.

Exemple de serveur OpenURL : CrossRef

CrossRef demande d'être identifié sur son site pour utiliser son service OpenURL en ligne. J'ai donc ouvert un compte CrossRef "Jean-Luc.Archimbaud@ujf-grenoble.fr", sans mot de passe, ce qui donne droit à un service minimum, c'est à dire l'accès aux informations publiques.

Ensuite dans mon navigateur en entrant l'URL :

<http://www.crossref.org/openurl?pid=Jean-Luc.Archimbaud@ujf-grenoble.fr&issn=17775310&date=2011&volume=61&issue=7&spage=2681>

Le serveur CrossRef me redirige vers :

http://aif.cedram.org/item?id=AIF_2011_61_7_2681_0

qui est un article paru dans la revue Annales de l'Institut Fourier, ISSN 1777/5310, volume 61, numéro 7, en 2011, à la page 2681.

Si j'avais eu un vrai compte CrossRef avec un mot de passe, j'aurais indiqué ce mot de passe dans le champ pid sous la forme username:password.

Affectation des identifiants

Il n'y a pas une organisation officielle d'affectation d'identifiant ou de préfixe d'identifiant. N'importe qui peut ouvrir un serveur OpenURL.

Caractéristiques par rapport aux 6 fonctionnalités

Fonctionnalités présentes :

- Le point 2 est rempli. Ce codage permet d'identifier tous types de documents en particulier livres, publications en série et articles scientifiques.
- À partir d'un code OpenURL, ce code étant de la forme d'une URL, on accède très rapidement au document. Le point 4 est satisfait.

Fonctionnalités absentes :

- Un identifiant OpenURL peut désigner plusieurs documents. En effet la description du document peut être limitée à un argument (nom de l'auteur par exemple) qui désignera ainsi plusieurs documents. Le point 1 n'est pas satisfait.
- Un document doit être référencé dans un serveur OpenURL pour disposer d'un code OpenURL. Or il n'y a pas de serveur international OpenURL où l'on puisse enregistrer n'importe quel

document. Le point 3 n'est pas satisfait.

- Le code OpenURL reposant sur des serveurs quelconques sans organisation ni contrainte pour ces serveurs, ils peuvent changer de nom, disparaître... La notation OpenURL n'est pas pérenne. Le point 5 n'est pas validé.
- Un document peut avoir plusieurs identifiants OpenURL différents, il suffit qu'il soit 'déclaré' dans plusieurs serveurs OpenURL. Le point 6 n'est rempli.

Utilisation

OpenURL est principalement un codage standardisé pour interroger un serveur qui donne accès à un certain nombre de ressources (selon la configuration du serveur), ou d'informations sur les ressources, selon l'identité du demandeur (et donc de ses droits). Or il n'y a pas de serveur OpenURL international, ouvert à tous. Ainsi ce n'est pas un identifiant standard mondial.

Par contre, de nombreux serveurs de ce type, appelés en France "résolveurs de liens", sont installés, en particulier dans les universités avec un périmètre limité. Ils peuvent être interrogés avec la notation OpenURL. Ils ne permettent d'accéder qu'aux documents déclarés dans le serveur, c'est à dire principalement aux ressources qui font l'objet d'abonnements souscrits par l'université et parfois à des ressources libres. Ils sont basés sur des logiciels commerciaux dont l'organisation et les options techniques (description des ressources, système d'identification des utilisateurs...) sont spécifiques à chaque logiciel.

OAI

De même que OpenURL, OAI a plusieurs définitions et n'est pas vraiment un système d'identification (néanmoins il peut être utilisé pour cette fonction).

D'après mes lectures, OAI, Open Archives Initiative, est à la base un projet pour faciliter les échanges et **l'interrogation d'archives de documents numériques** tel que HAL. Il est maintenant visiblement utilisé pour d'autres bases d'informations que les archives.

Pour cela, un protocole a été défini : OAI-PMH, Open Archives Initiative Protocol for Metadata Harvesting, pour permettre la récupération et l'échange de métadonnées de différentes archives libres ou autres bases d'informations. Le protocole est basé sur le dialogue de 2 acteurs appelés : les fournisseurs de données (les sites qui hébergent les données) et les fournisseurs de service (aussi appelés moissonneurs, qui sont des portails d'accès à plusieurs archives avec un moteur de recherche). Les premiers vont donner accès à leur catalogue à travers ce qui est appelé un "entrepôt OAI", un outil chargé de répondre aux requêtes formulées par les seconds. Un moteur de recherche a été développé spécifiquement pour les fournisseurs de service : BASE.

Dans le protocole OAI-PMH chaque enregistrement de métadonnées (donc chaque ressource documentaire) est désigné par un identifiant OAI, lié au fournisseur de service.

Codage des identifiants

L'identifiant a la forme : oai: "nom d'un fournisseur de données, d'une archive" : "identifiant de la ressource dans l'archive". La notation est ainsi similaire au DOI.

Exemples : oai:arXiv.org:hep-th/9901001 (hep-th/9901001 est un identifiant de arXiv) , oai:hal.archives-ouvertes.fr:hal-00819683 (hal-00819683 est un identifiant de HAL).

Affectation des identifiants

De même que OpenURL, n'importe qui peut ouvrir un "entrepôt OAI" ou un moissonneur OAI. Par contre l'Open Archives Initiative maintient [un registre des fournisseurs de données](#) (les serveurs HAL et arXiv sont enregistrés dans ce registre par exemple) et [un registre des fournisseurs de service](#) OAI dans

lequel on peut enregistrer son service. Lors de l'enregistrement dans ces registres, puis régulièrement, un ensemble de tests est effectué pour vérifier que le fournisseur implémente correctement le protocole OAI-PMH.

Caractéristiques par rapport aux 6 fonctionnalités

Fonctionnalités présentes :

- Un identifiant OAI désigne un seul document. Le point 1 est validé.
- On peut utiliser l'identification OAI pour les livres, publications en série et articles scientifiques. Le point 2 est validé.

Fonctionnalités presque présentes :

- La pérennité de l'identifiant est liée à celle de l'archive ouverte. Elle n'est donc pas totalement assurée mais néanmoins les serveurs d'archives ouvertes sont destinés à durer donc on peut considérer qu'il y a peu de chance que le serveur disparaisse. Le point 5 est presque rempli.

Fonctionnalités absentes :

- La méthode n'est pas universelle. Pour avoir un identifiant OAI la ressource doit être déclarée dans un entrepôt de données OAI, et il n'y a pas d'entrepôt ouvert à tout. Le point 3 n'est pas validé.
- Pour un utilisateur qui n'est pas compétent en documentation, il est difficile de retrouver un document avec son identifiant OAI. Le point 4 n'est pas rempli.
- Un document peut être déposé dans plusieurs archives ou bases d'informations,, il aura alors plusieurs OAI. Le point 6 n'est pas validé.

Utilisation

OAI est très répandu dans le monde des archives ouvertes et dans des entrepôts de données en accès ouvert. HAL par exemple, affiche un identifiant OAI dans les métadonnées de chaque article déposé.

ARK

ARK (Archival Resource Key) est un système d'identification qui peut **identifier tout type d'objet numérique mais aussi physique**. La [California Digital Library](#), CDL, est à l'origine du système et elle le maintient. Elle l'utilise également.

Codage des identifiants

L'identifiant est de la forme `ark:/NAAN/Name` avec "NAAN", Name Assigning Authority Number, numéro de 5 chiffres de l'organisation qui va attribuer le nom suivant, "Name" le nom de l'objet. Exemple : `ark:/52327/61248`, `ark:/13030/654xz321/s3/f8.05v.tiff`

On transforme cet identifiant pour en faire une URL en ajoutant le nom du serveur qui va donner accès à l'objet avec l'identifiant ARK, serveur qui est appelé NMA, Name Mapping Authority. Exemple : <http://collections.banq.qc.ca/ark:/52327/61248> indique que le serveur `collections.banq.qc.ca`, la Bibliothèque et Archives Nationales du Québec va résoudre l'identifiant ARK pour retourner le document.

<http://gallica.bnf.fr/ark:/12148/bpt6k107371t> pointe sur un document avec l'identifiant `ark:/12148/bpt6k10737`, accessible via le serveur <http://gallica.bnf.fr/>

Un identifiant ARK est parfois noté sous cette forme URL.

Affectation des identifiants

La CDL affecte gratuitement un NAAN, Name Assigning Authority Number aux organisations qui en font la demande. Ensuite l'organisation attribue les Names. Par exemple, la Bibliothèque et Archives Nationales du Québec a pour NAAN 52327. Une [liste des organismes utilisant des identifiants ARK](#) est maintenu par la CDL.

Caractéristiques par rapport aux 6 fonctionnalités

Fonctionnalités présentes :

- Un identifiant ARK désigne un seul document. Le point 1 est validé.
- On peut utiliser l'identification ARK pour les livres, publications en série et articles scientifiques. 2 est validé.

Fonctionnalités presque présentes :

- L'identifiant ARK sous la forme `ark:/12148/bpt6k107371t` ne permet pas à un utilisateur non averti en documentation de retrouver facilement le document. Par contre avec la forme <http://gallica.bnf.fr/ark:/12148/bpt6k107371t> utilisable avec tous les navigateurs c'est évident. Le point 4 est presque rempli.

Fonctionnalités absentes :

- La méthode n'est pas universelle. Pour avoir un identifiant ARK il faut que la ressource soit sur un serveur d'une organisation NAAN . Le point 3 n'est pas validé.
- Un document peut être déposé dans plusieurs serveurs d'organisation NAAN, il aura plusieurs identifiant ARK. Le point 6 n'est pas validé.

Utilisation

La CDL utilise évidemment le système mais aussi de nombreuses bibliothèques universitaires. En France, plusieurs organisations sont NAAN. La BNF (NAAN : 12148) utilise les identifiant ARK pour ses documents numériques et ses notices bibliographiques.

URI

URI (Uniform Resource Identifier) est une **syntaxe générique pour une identification pérenne des ressources dans l'Internet**. Elle a été promue par les instances liées à l'Internet et au Web (IETF et W3C) et décrite dans le [RFC3986](#). Ce RFC donne aussi des recommandations sur le processus pour résoudre ces identifiants.

Un identifiant URI est composé de 3 parties :

- Un préfixe, qui indique le contexte - le système dans lequel l'identifiant est attribué. Exemples : `http`, `DOI`.
- Un élément qui désigne l'autorité - l'organisation qui a attribué l'identifiant au sein de ce système. Cette autorité est désignée selon une convention propre au système. Exemples : nom de domaine pour l'URL, préfixe identifiant une organisation pour le DOI.
- Le nom de la ressource, chaîne de caractères qui identifie la ressource de manière unique, au sein

de ce système et pour cette autorité.

Plusieurs identifiants décrits avant (URL, DOI...) respectent cette syntaxe et sont donc des URI.

Tableau récapitulatif des 6 fonctionnalités pour les identifiants décrits

Je rappelle que ce tableau n'est pas destiné à mettre en avant la "meilleure" méthode d'identification, chacune est utilisée et a son intérêt, mais à permettre de les caractériser et de les différencier, en prenant du recul.

À noter que la fonctionnalité "Identification pérenne" est impossible à évaluer avec certitude car elle demande de "prévoir l'avenir". Il est nécessaire que chacun l'évalue après une réflexion, en faisant appel à son bon sens, car en première approximation ça peut être trompeur. Prenons deux exemples :

- L'URL est souvent présentée comme non pérenne, ce que je fais d'ailleurs dans le tableau. Mais la pérennité est très variable selon l'URL. Si l'URL référence par exemple un document qui est sur une page personnelle sur un site web de laboratoire (laboratoire qui peut disparaître, changer de nom, changer de politique par rapport aux pages personnelles, changer d'organisation de son Web...) ou chez un fournisseur d'accès internet qui offre un hébergement gratuit, l'espérance de pérennité est faible. Par contre si c'est sur un site d'archives ouvertes comme HAL, on peut supposer que l'URL a beaucoup plus de chances d'être pérenne.
- Il n'est pas garanti qu'une identification fournie par une instance, même payante, soit pérenne à 100 %. En effet, l'instance peut disparaître ou changer son mode de fonctionnement (tarifs, restrictions d'utilisation...) par exemple.

	Un identifiant désigne une seule ressource documentaire (1)	Tous les types de ressources documentaires sont identifiables (2)	Toute ressource documentaire peut obtenir un identifiant (3)	Accès rapide à la ressource documentaire (4)	Identification pérenne (5)	Une ressource documentaire a un et un seul identifiant (6)
ISBN	OUI	NON	OUI	NON	OUI	PRESQUE
ISSN	PRESQUE	NON	OUI	NON	OUI	PRESQUE
URL	OUI	OUI	OUI	OUI	NON	NON
DOI	OUI	OUI	NON	OUI	OUI	OUI
OpenURL	NON	OUI	NON	OUI	NON	NON
OAI	OUI	OUI	NON	NON	PRESQUE	NON
ARK	OUI	OUI	NON	PRESQUE	PRESQUE	NON

Identifiants spécifiques à des bases d'informations connues

Chaque base de données documentaires ou base bibliographique a besoin d'identifier de manière unique les objets qu'elle manipule. Elle crée donc un identifiant spécifique à sa base ou s'appuie sur un identifiant utilisé dans un domaine similaire au sien. Si cette base est très utilisée, l'utilisation de cet identifiant se répand alors à l'extérieur de la base en question et devient un identifiant de référence, rivalisant avec les identifiants présentés avant.

Ci-dessous en voici quelques-uns connus, particulièrement lié à la documentation en maths.

En France

- [HAL](#)
Pour les articles déposés, HAL utilise la notation "**hal-XXXXXXXX** - version N" avec XXXXXXXX un numéro de séquence de dépôt et N le numéro de version de l'article (exemple : hal-00819683, version 1). Pour l'identification des thèses, "hal" est remplacé par "tel" et pour les cours par "cel". À noter que :
 - Il suffit de faire précéder l'identifiant par "<http://hal.archives-ouvertes.fr/>" pour avoir l'URL du document. Exemple : <http://hal.archives-ouvertes.fr/hal-00819683>
 - Lors de l'affichage d'un document, HAL précise 3 identifiants pour le document : le sien, l'URL et l'OAI. Exemple :
hal-00819683, version 1
<http://hal.archives-ouvertes.fr/hal-00819683>
oai:hal.archives-ouvertes.fr:hal-00819683
- Le [catalogue SUDOC](#) utilise pour ses notices, l'**identifiant PPN**, Pica Production Number, conséquence du logiciel utilisé : PICA de la société OCLC. PPN est donc utilisé dans d'autres bibliothèques dans le monde qui utilisent ce logiciel.
 - L'identifiant PPN est composé de 9 caractères. Les 8 premiers chiffres correspondent au numéro séquentiel de la notice dans la base. Le dernier caractère est une clé de contrôle, sous la forme d'un chiffre ou d'un X. Exemple : PPN 039563782.
 - Il est possible d'utiliser cet identifiant pour accéder directement à une notice dans le sudoc : <http://www.sudoc.fr/{PPN}>. Exemple pour la notice de PPN 039563782 : <http://www.sudoc.fr/039563782>.
 - Dans le SUDOC, un PPN peut aussi être affecté à un auteur, identifiant qui permet, via [le service IdReF](#) de faire le lien entre un auteur et toutes ses publications référencées.
 - Dans ses notices, pour les revues, le SUDOC stocke l'identifiant ISSN.
 - La base bibliographique [WorldCat](#) de l'[OCLC](#) utilise aussi cet identifiant à 9 caractères appelé alors **Numéro OCLC**.
- [NUMDAM](#) et [CEDRAM](#) utilisent un identifiant interne mais sans l'afficher. NUMDAM affiche l'URL du document (exemple : http://www.numdam.org/item?id=BSMF_1872-1873__1__27__1 et CEDRAM le DOI.
- Le [CFP](#) n'affiche pas d'identifiant interne. Pour une revue il indique l'ISSN et le PPN.

Dans les bases bibliographiques internationales

- [Zentralblatt](#) utilise un identifiant interne sous forme de 2 nombres de 4 chiffres notés X.Y par exemple Zbl 1267.14016. Cet identifiant est dans un champ appelé "**an**".
Pour accéder à la notice de la ressource précédente il suffit d'utiliser l'URL : <https://www.zbmath.org/?q=an:1267.14016>.
Comme identifiant affiché, ZB affiche le DOI (<http://dx.doi.org/10.1016/j.geomphys.2012.11.003>

pour l'exemple précédent).

- [MathSciNet](#) utilise un identifiant noté **MR**, sous forme de 7 chiffres, comme MR3180943. Pour accéder à une notice de la ressource précédente il suffit d'utiliser l'URL : <http://www.ams.org/mathscinet-getitem?mr=3180943>.

Dans les autres bases internationales

- [ArXiv](#) utilise depuis mars 2007 un identifiant de la forme : **arXiv:YYMM.NNNNvV** avec YY année, MM mois, NNNN le numéro de séquence de la soumission dans le mois, vV un numéro de version. Exemple : arXiv:1305.0178 qui est le 178ième article déposé en mai 2013. Pour accéder à la notice de cet article : <http://arxiv.org/abs/1305.0178>

Dans d'autres communautés scientifiques que les maths, quelques exemples

- **Bibcode** est un identifiant utilisé par plusieurs bases bibliographiques de l'astronomie comme [SIMBAD](#). Il est composé de 19 caractères : YYYYYJJJJVVVVMPPPPA avec YYYYY l'année et JJJJJ un code de revue, VVVV le numéro de volume, M la section, PPPP la page du début, A l'initiale du nom de l'auteur principal. 1974AJ.....79..819H est un article publié en 1974 dans la revue The Astronomical Journal, volume 79, page 819 avec comme auteur principal W. D Heintz.
- **CODEN** est un identifiant de publication à six caractères, particulièrement utilisé en chimie. Il est actuellement attribué par le [Chemical Abstracts Service \(CAS\)](#) à toute publication qui en fait la demande.
- Dans le domaine des sciences de la vie, biologie, médecine, [PubMed](#) (moteur de recherche) utilise un identifiant **PMID** (PubMed identifier), [PubMed Central](#) un identifiant **PMCID**. À noter que l'outil [pmid2doi](#), permet de faire la conversion PMID vers DOI.

Quelques documentations en ligne sur le sujet

ARK : [CDL ARK](#) : <https://wiki.ucop.edu/display/Curation/ARK>

ARK : [ARK à la BNF](#) : http://www.bnf.fr/fr/professionnels/issn_isbn_autres_numeros/a.ark.html

DOI : [article Wikipédia](#) en anglais : http://en.wikipedia.org/wiki/Digital_object_identifier

DOI : [DOI Handbook](#) International DOI Foundation (documentation très complète en ligne) : http://www.doi.org/doi_handbook/TOC.html

DOI : [INIST - Attribution de DOI](#) : <http://www.inist.fr/?Attribution-de-DOI>

DOI : [Le Digital Object Identifier : Le système du DOI, Catherine Lupovici, 1998](#) : <http://bbf.enssib.fr/consulter/bbf-1998-03-0049-007>. Attention ce document date un peu.

ISBN, ISSN... : [ISBN, ISSN et autres identifiants](#) (site de la BNF) : http://www.bnf.fr/fr/professionnels/issn_isbn_autres_numeros.html

ISBN : [Manuel d'utilisation de l'ISBN \(Agence Internationale de l'ISBN, Londres 2012\)](#) : <http://www.afnil.org/pdf/Manuel%20d%27utilisation%20de%20l%27ISBN.pdf>

ISSN : [Manuel de l'ISSN \(Centre international de l'ISSN, Juin 2012\)](https://www.portail-math.fr/sites/default/files/ISSNManual2012-FRA.pdf) : <https://www.portail-math.fr/sites/default/files/ISSNManual2012-FRA.pdf>

ISSN : [Comprendre l'ISSN](http://www.issn.org/fr/comprendre-lissn/) (site du Centre International d'Enregistrement des Publications en série) : <http://www.issn.org/fr/comprendre-lissn/>

OAI : [serveur de l'Open Access Initiative](http://www.openarchives.org/) : <http://www.openarchives.org/>

OpenURL : [article Wikipédia](http://en.wikipedia.org/wiki/OpenURL) en anglais : <http://en.wikipedia.org/wiki/OpenURL>

URI : [RFC3986](http://tools.ietf.org/html/rfc3986) : <http://tools.ietf.org/html/rfc3986>