

# Identifiants des documents numériques : ISBN, ISSN, URL, Handle, DOI, OpenURL, OAI, ARK

V2

Article complet :

<http://jl.archimbaud.free.fr/Identifiants-docs-numeriques-JLA-janv-2015.pdf>

Jean-Luc Archimbaud - <http://jl.archimbaud.free.fr/>

Les ateliers de l'information – Grenoble – 3 mars 2015

# PLAN

---

- **Pourquoi suis-je là ?**
- **Objet de la présentation**
- **Les différents identifiants**
- **Tableau d'analyse**
- **Conclusions personnelles**

# Pourquoi suis je là ?

---

- **Ingénieur en informatique, CNRS : pas spécialiste en documentation**
- **Depuis 3 ans : cellule Mathdoc**
  - Beaucoup d'identifiants manipulés
  - Peu de connaissances approfondies sur ces identifiants
- **Rédaction d'un article (18 pages) été 2014 → sept 2014**
  - <http://jl.archimbaud.free.fr/Identifiants-docs-numeriques-JLA-janv-2015.pdf>
  - Diffusé documentalistes... : apprécié --> compléments → nouvelle version
- **Proposition de présentation aux Ateliers de l'information à Jean-Michel Mermet**

# Objet de la présentation

---

- **Constat : on a besoin d'identifier les documents numériques de manière universelle : accès, référencement...**
  - Le pourquoi n'est pas l'objet de ma présentation
- **Objet présentation : état des lieux des identifiants utilisés**
  - Et des mécanismes d'attribution...
  - Dans la communauté Ens Sup – Recherche
  - Pour articles (scientifiques), publications en série (journaux, périodiques, collections) et livres
  - → ISBN, ISSN, URL, DOI, Handle, OpenURL, OAI, ARK
    - article : URI, HAL, SUDOC-PPN, Zentralblatt&Mathscinet, ArXiv, Bibcode, Coden, PMIDID
  - Avec une grille d'analyse pour essayer de mieux les différencier (sans appréciation de valeur)
- **Survol du sujet : cf article pour plus d'informations**

# Grille d'analyse d'un système d'identification

---

- **1. Chaque identifiant désigne un seul document**
- **2. Permet d'identifier tous les types de documents**
  - Livres, publications en série et articles scientifiques pour notre propos
- **3. Ouvert à toute personne pour identifier ses documents**
  - Chaque auteur peut avoir un identifiant pour son document : sans contrainte d'appartenance à une organisation, à coût nul/minime
- **4. Permet d'accéder rapidement au document**
  - Pour un utilisateur non spécialiste de la documentation
- **5. Identifie le document de manière pérenne dans le temps**
  - Pas dépendant d'une machine...
- **6. Un document a un et un seul identifiant**
  - Permet de pouvoir identifier facilement tout ce qui se rapporte au document

# ISBN

ISBN 978-0747595823 -3



9 780747 595823 >

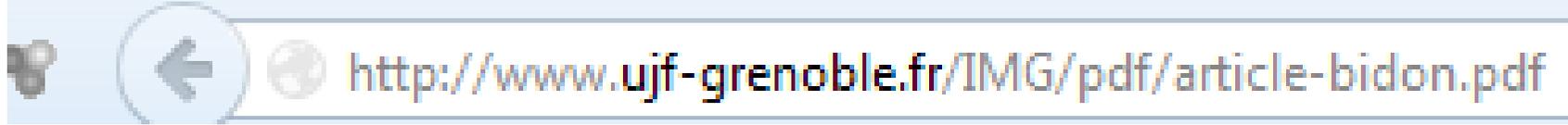
- **International Standard Book Number**
  - Pour chaque (édition d'un) livre publié (support papier/numérique)
- **ISBN10 (→ 2006) ISBN 2-271-05256-4**
  - Pays 2 (France), éditeur 271 (CNRS Edition), 5256 ième livre, 4 clé contrôle
- **ISBN14 (2006 →) ISBN 979-10-93727-04-2**
  - 979 (pas ISBN10), pays 10 (France), éditeur 933727, 4 ième livre, 2 clé contrôle
  - ISBN 978-2-271-05256-8 idem exemple ISBN10 précédent
- **FR : AFNIL (Agence Francophone pour la numérotation internationale du livre) délivre gratuitement un numéro aux éditeurs**
- **Repose sur la notion d'éditeur**
- **Pas uniquement document numérique**

# ISSN



- **International Standard Serial Number**
  - Publication en série (journal, périodique, collection)
  - 1 / titre ; 1 / support
- **ISSN 0994-7647 : Journal du CNRS**
  - 0994764 ième revue identifiée dans monde, 7 clé de contrôle
  - Plage étroite → des revues peuvent avoir le même numéro → titre(, années) ajouté au numéro pour mieux identifier
- **ISSN-L permet de relier les différents supports même revue**
- **FR : Centre national d'enregistrement des publications en série : ISSN France » à la BNF attribue un numéro gratuitement**
- **Pas uniquement document numérique**

# URL



- **Uniform Resource Locator**

- Utilisé par les navigateurs Web

- **<http://www.ujf-grenoble.fr/IMG/pdf/article-bidon.pdf>**

- http : méthode d'accès (protocole réseau)
- www.ujf-grenoble.fr : machine www dans le domaine ujf-grenoble.fr
- IMG/pdf : emplacement du document dans la machine www
- Article-bidon.pdf : nom du fichier contenant le document

- **FR : AFNIC (Association Française pour le Nommage Internet en Coopération) délivre les noms de domaines a un coût modique**

- L'administrateur du domaine gère : noms de machines, arborescences, noms de fichiers...

- **Identifiant Internet (pas d'origine monde de l'édition-documentation)**

- **Le plus ouvert, le plus facile à utiliser**

- **Mais le moins pérenne car**

- Domaine/machine peut disparaître (ou changer de nom), organisation des fichiers être modifié...

# HANDLE (HDL)

URI : <http://hdl.handle.net/2332/1241>

- **Systeme (organisation, services, logiciels) pour avoir des URL pérennes (indépendantes du lieu de stockage du document)**
- **Principe (simplifié)**
  - Le CNRI (Corporation for National Research Initiatives) délivre un identifiant (exemple 10760) à une organisation (50 \$ / an) et gère un serveur hdl.handle.net
  - L'organisation installe des logiciels add hoc pour assurer le service HDL sur un serveur chez elle, en attribuant un numéro à chaque document (exemple 4496)
  - Le document aura l'identifiant HDL : <http://handle.net.fr/10760/4496>
  - La machine handle.net.fr redirigera vers le serveur de l'organisation...
- **Pérenne si l'organisation ne disparaît pas et gère ce qu'il faut**

# DOI

Pour citer ce document :

| DOI : 10.4267/2042/54071

- **Digital Object Identifier**
- **Exemple : doi:10.5802/aif.2754**
  - Article appelé aif.2754 dans l'organisation ayant pour préfixe 10.5802
- **9 agences délivrent des préfixes aux organisations**
  - FR Ens Sup Rec : CrossRef, Datacite
- **Les organisations adhèrent à une agence et affectent des identifiants aux documents**
- **Comment ça marche ?**
  - Cas particulier d'HDL → système avec serveurs, logiciels, redirection...
- **Bien implanté mais cher**
  - Coût pour chaque document de l'ordre de 1 €

# OpenURL

---

- **Syntaxe qui permet de transporter des informations dans une URL**
  - Pour identifier un document, l'utilisateur demandant (éventuellement mot de passe)...
- **Serveurs (OpenURL) qui répondent à ces URL**
  - Où sont stockés des documents ou qui savent sur quels serveurs sont certains documents
  - Pas de serveur central, n'importe qui peut monter un serveur OpenURL
- **Exemple : <http://www.crossref.org/openurl?pid=JL.Archimbaud&issn=17775310&date=2011&volume=61&issue=7&spage=2681>**
  - Serveur OpenURL [www.crossref.org](http://www.crossref.org)
  - Login JL.Archimbaud
  - Demande document (article) dans ISSN 17775310 (revue) parue en 2011 dans volume 61, numéro 7 et page 2681
  - Serveur crossref redirige vers [http://aif.cedram.org/item?id=AIF\\_2011\\_\\_61\\_7\\_2681\\_0](http://aif.cedram.org/item?id=AIF_2011__61_7_2681_0)
- **Pas vraiment un identifiant mais utile pour sa définition des requêtes (--> standard)**
  - On inclut une question précise dans l'URL qui sera comprise par le serveur OpenURL
  - Nombreux serveurs OpenURL dans les universités

# OAI

---

- **Open Archive Initiative**

- Origine : faciliter les échanges et interrogation d'archives

- **Mécanisme**

- Entrepôts OAI : serveurs fournisseurs de documents
- Moissonneurs OAI : serveurs qui interrogent les entrepôts
- OAI-PMH : protocole (→ logiciels) d'échanges
- Organisation OAI gère un registre international des entrepôts et moissonneurs :
  - Vérifie régulièrement que les serveurs respectent OAI-PMH

- **Exemple identifiant : oai:hal.archives-ouvertes.fr:hal-00819683**

- Entrepôt : hal.archives-ouvertes.fr, doc : hal-00819683

- **Très répandu dans archives ouvertes (HAL...) et documents en open access**

- Comme méthode d'échanges : documents mais aussi métadonnées...

# ARK

Permalien sur ce document

<http://gallica.bnf.fr/ark:/12148/bpt6k114>

- **Archival Resource Key**
- **ark:/12148/bpt6k107371t**
  - 12148 : identifiant d'une organisation (la BNF), délivré par la CDL (California Digital Library)
  - Bpt6k107371t : identifiant du document dans l'organisation
  - Autre forme <http://gallica.bnf.fr/ark:/12148/bpt6k107371t> : utilisable avec un navigateur
- **Utilisé par**
  - Nombreuses bibliothèques (surtout américaines ?)
  - BNF (Bibliothèque Nationale de France)

# Tableau d'analyse

	1 id → 1 doc	Tous types de docs	Ouvert à tous	Accès rapide	Id pérenne	1 doc → 1 id
ISBN	OUI	NON	OUI	NON	OUI	Presque
ISSN	Presque	NON	OUI	NON	OUI	Presque
URL	OUI	OUI	OUI	OUI	NON	NON
HDL	OUI	OUI	Presque	OUI	OUI	OUI
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

# Conclusions personnelles

---

- **Identifiant (de document) unique universel ? : non**
  - Car personne ne peut l'imposer
  - Ce n'est pas une impossibilité technique
    - Sauf difficulté pour la migration de l'existant
  - Les identifiants sont des objets commerciaux et de pouvoir
  - Certaines sociétés/associations/..
    - Vivent de l'attribution d'identifiants
    - Maîtrisent l'organisation de l'attribution des identifiants
    - Ont développé des logiciels, bases de données, moteurs de recherche... basés sur certains identifiants
- **Chaque identifiant existant a actuellement son utilité**
  - Certains sont historiques
- **Si on un choix à faire : le faire en connaissant les différentes options**

# **Identifiants des documents numériques : ISBN, ISSN, URL, Handle, DOI, OpenURL, OAI, ARK**

V2

Article complet :

<http://jl.archimbaud.free.fr/Identifiants-docs-numeriques-JLA-janv-2015.pdf>  
Jean-Luc Archimbaud - <http://jl.archimbaud.free.fr/>

Les ateliers de l'information – Grenoble – 3 mars 2015

Licence de cette présentation : Creative Common BY-NC-ND

# PLAN

---

- **Pourquoi suis-je là ?**
- **Objet de la présentation**
- **Les différents identifiants**
- **Tableau d'analyse**
- **Conclusions personnelles**

# Pourquoi suis je là ?

---

- **Ingénieur en informatique, CNRS : pas spécialiste en documentation**
- **Depuis 3 ans : cellule Mathdoc**
  - Beaucoup d'identifiants manipulés
  - Peu de connaissances approfondies sur ces identifiants
- **Rédaction d'un article (18 pages) été 2014 → sept 2014**
  - <http://jl.archimbaud.free.fr/Identifiants-docs-numeriques-JLA-janv-2015.pdf>
  - Diffusé documentalistes... : apprécié --> compléments → nouvelle version
- **Proposition de présentation aux Ateliers de l'information à Jean-Michel Mermet**

# Objet de la présentation

---

- **Constat : on a besoin d'identifier les documents numériques de manière universelle : accès, référencement...**
  - Le pourquoi n'est pas l'objet de ma présentation
- **Objet présentation : état des lieux des identifiants utilisés**
  - Et des mécanismes d'attribution...
  - Dans la communauté Ens Sup – Recherche
  - Pour articles (scientifiques), publications en série (journaux, périodiques, collections) et livres
  - → ISBN, ISSN, URL, DOI, Handle, OpenURL, OAI, ARK
    - article : URI, HAL, SUDOC-PPN, Zentralblatt&Mathscinet, ArXiv, Bibcode, Coden, PMDID
  - Avec une grille d'analyse pour essayer de mieux les différencier (sans appréciation de valeur)
- **Survol du sujet : cf article pour plus d'informations**

# Grille d'analyse d'un système d'identification

---

- **1. Chaque identifiant désigne un seul document**
- **2. Permet d'identifier tous les types de documents**
  - Livres, publications en série et articles scientifiques pour notre propos
- **3. Ouvert à toute personne pour identifier ses documents**
  - Chaque auteur peut avoir un identifiant pour son document : sans contrainte d'appartenance à une organisation, à coût nul/minime
- **4. Permet d'accéder rapidement au document**
  - Pour un utilisateur non spécialiste de la documentation
- **5. Identifie le document de manière pérenne dans le temps**
  - Pas dépendant d'une machine...
- **6. Un document a un et un seul identifiant**
  - Permet de pouvoir identifier facilement tout ce qui se rapporte au document

# ISBN

ISBN 978-0747595823 -3



9 780747 595823 >

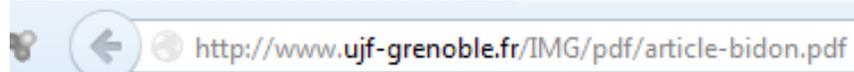
- **International Standard Book Number**
  - Pour chaque (édition d'un) livre publié (support papier/numérique)
- **ISBN10 (→ 2006) ISBN 2-271-05256-4**
  - Pays 2 (France), éditeur 271 (CNRS Edition), 5256 ième livre, 4 clé contrôle
- **ISBN14 (2006 →) ISBN 979-10-93727-04-2**
  - 979 (pas ISBN10), pays 10 (France), éditeur 933727, 4 ième livre, 2 clé contrôle
  - ISBN 978-2-271-05256-8 idem exemple ISBN10 précédent
- **FR : AFNIL (Agence Francophone pour la numérotation internationale du livre) délivre gratuitement un numéro aux éditeurs**
- **Repose sur la notion d'éditeur**
- **Pas uniquement document numérique**

# ISSN



- **International Standard Serial Number**
  - Publication en série (journal, périodique, collection)
  - 1 / titre ; 1 / support
- **ISSN 0994-7647 : Journal du CNRS**
  - 0994764 ième revue identifiée dans monde, 7 clé de contrôle
  - Plage étroite → des revues peuvent avoir le même numéro → titre(, années) ajouté au numéro pour mieux identifier
- **ISSN-L permet de relier les différents supports même revue**
- **FR : Centre national d'enregistrement des publications en série : ISSN France » à la BNF attribue un numéro gratuitement**
- **Pas uniquement document numérique**

# URL



- **Uniform Ressource Locator**
  - Utilisé par les navigateurs Web
- **<http://www.ujf-grenoble.fr/IMG/pdf/article-bidon.pdf>**
  - http : méthode d'accès (protocole réseau)
  - www.ujf-grenoble.fr : machine www dans le domaine ujf-grenoble.fr
  - IMG/pdf : emplacement du document dans la machine www
  - Article-bidon.pdf : nom du fichier contenant le document
- **FR : AFNIC (Association Française pour le Nomage Internet en Coopération) délivre les noms de domaines a un coût modique**
  - L'administrateur du domaine gère : noms de machines, arborescences, noms de fichiers...
- **Identifiant Internet (pas d'origine monde de l'édition-documentation)**
- **Le plus ouvert, le plus facile à utiliser**
- **Mais le moins pérenne car**
  - Domaine/machine peut disparaître (ou changer de nom), organisation des fichiers être modifié...

# HANDLE (HDL)

URI : <http://hdl.handle.net/2332/1241>

- **Système (organisation, services, logiciels) pour avoir des URL pérennes (indépendantes du lieu de stockage du document)**
- **Principe (simplifié)**
  - Le CNRI (Corporation for National Research Initiatives) délivre un identifiant (exemple 10760) à une organisation (50 \$ / an) et gère un serveur hdl.handle.net
  - L'organisation installe des logiciels add hoc pour assurer le service HDL sur un serveur chez elle, en attribuant un numéro à chaque document (exemple 4496)
  - Le document aura l'identifiant HDL : <http://handle.net.fr/10760/4496>
  - La machine handle.net.fr redirigera vers le serveur de l'organisation...
- **Pérenne si l'organisation ne disparaît pas et gère ce qu'il faut**

# DOI

Pour citer ce document :

| DOI : 10.4267/2042/54071

- **Digital Object Identifier**
- **Exemple : doi:10.5802/aif.2754**
  - Article appelé aif.2754 dans l'organisation ayant pour préfixe 10.5802
- **9 agences délivrent des préfixes aux organisations**
  - FR Ens Sup Rec : CrossRef, Datacite
- **Les organisations adhèrent à une agence et affectent des identifiants aux documents**
- **Comment ça marche ?**
  - Cas particulier d'HDL → système avec serveurs, logiciels, redirection...
- **Bien implanté mais cher**
  - Coût pour chaque document de l'ordre de 1 €

# OpenURL

---

- **Syntaxe qui permet de transporter des informations dans une URL**
  - Pour identifier un document, l'utilisateur demandant (éventuellement mot de passe)...
- **Serveurs (OpenURL) qui répondent à ces URL**
  - Où sont stockés des documents ou qui savent sur quels serveurs sont certains documents
  - Pas de serveur central, n'importe qui peut monter un serveur OpenURL
- **Exemple : <http://www.crossref.org/openurl?pid=JL.Archimbaud&issn=17775310&date=2011&volume=61&issue=7&spage=2681>**
  - Serveur OpenURL [www.crossref.org](http://www.crossref.org)
  - Login JL.Archimbaud
  - Demande document (article) dans ISSN 17775310 (revue) parue en 2011 dans volume 61, numéro 7 et page 2681
  - Serveur crossref redirige vers [http://aif.cedram.org/item?id=AIF\\_2011\\_\\_61\\_7\\_2681\\_0](http://aif.cedram.org/item?id=AIF_2011__61_7_2681_0)
- **Pas vraiment un identifiant mais utile pour sa définition des requêtes (--> standard)**
  - On inclut une question précise dans l'URL qui sera comprise par le serveur OpenURL
  - Nombreux serveurs OpenURL dans les universités

# OAI

---

- **Open Archive Initiative**
  - Origine : faciliter les échanges et interrogation d'archives
- **Mécanisme**
  - Entrepôts OAI : serveurs fournisseurs de documents
  - Moissonneurs OAI : serveurs qui interrogent les entrepôts
  - OAI-PMH : protocole (→ logiciels) d'échanges
  - Organisation OAI gère un registre international des entrepôts et moissonneurs :
    - Vérifie régulièrement que les serveurs respectent OAI-PMH
- **Exemple identifiant : oai:hal.archives-ouvertes.fr:hal-00819683**
  - Entrepôt : hal.archives-ouvertes.fr, doc : hal-00819683
- **Très répandu dans archives ouvertes (HAL...) et documents en open access**
  - Comme méthode d'échanges : documents mais aussi métadonnées...

---

- **Archival Resource Key**

- **ark:/12148/bpt6k107371t**

- 12148 : identifiant d'une organisation (la BNF), délivré par la CDL (California Digital Library)
- Bpt6k107371t : identifiant du document dans l'organisation
- Autre forme <http://gallica.bnf.fr/ark:/12148/bpt6k107371t> : utilisable avec un navigateur

- **Utilisé par**

- Nombreuses bibliothèques (surtout américaines ?)
- BNF (Bibliothèque Nationale de France)

## Tableau d'analyse

	1 id → 1 doc	Tous types de docs	Ouvert à tous	Accès rapide	Id pérenne	1 doc → 1 id
ISBN	OUI	NON	OUI	NON	OUI	Presque
ISSN	Presque	NON	OUI	NON	OUI	Presque
URL	OUI	OUI	OUI	OUI	NON	NON
HDL	OUI	OUI	Presque	OUI	OUI	OUI
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

# Conclusions personnelles

---

- **Identifiant (de document) unique universel ? : non**
  - Car personne ne peut l'imposer
  - Ce n'est pas une impossibilité technique
    - Sauf difficulté pour la migration de l'existant
  - Les identifiants sont des objets commerciaux et de pouvoir
  - Certaines sociétés/associations/..
    - Vivent de l'attribution d'identifiants
    - Maîtrisent l'organisation de l'attribution des identifiants
    - Ont développé des logiciels, bases de données, moteurs de recherche... basés sur certains identifiants
- **Chaque identifiant existant a actuellement son utilité**
  - Certains sont historiques
- **Si on un choix à faire : le faire en connaissant les différentes options**