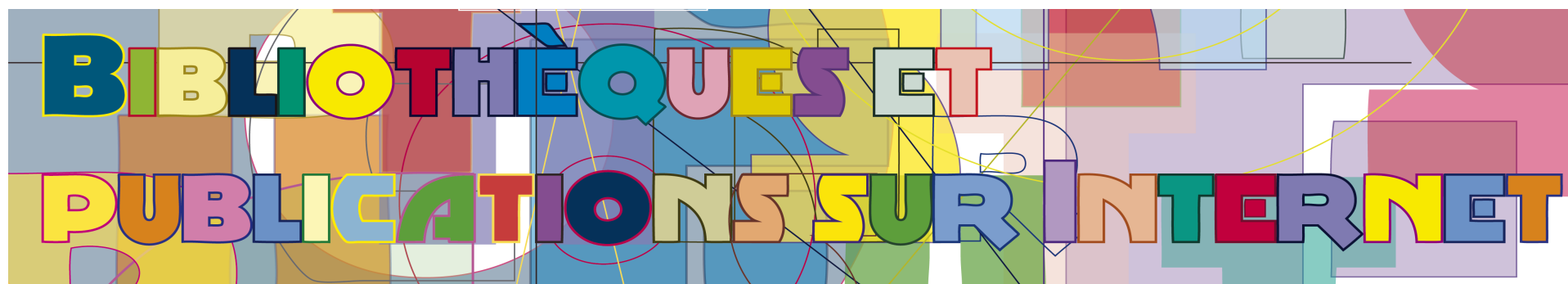


Les principes d'organisation du web de données : exemple d'une application scientifique

Gérald KEMBELLEC

Laboratoire Paragraphe, Université Paris 8

INTD - CNAM





Place des bibliothèques numériques

Les centres documentaires, notamment universitaires modernisent leurs interfaces en

- **informatisant** leurs catalogues;



Place des bibliothèques numériques

Les centres documentaires, notamment universitaires modernisent leurs interfaces en

- informatisant leurs catalogues;
- les **interconnectant**;



Place des bibliothèques numériques

Les centres documentaires, notamment universitaires, se sont adaptés à cette constatation en :

- informatisant leurs catalogues;
- les interconnectant;
- en ajoutant des **métas (glanage)** et options de **moissonnage**.

Notons l'effort de l'**ABES**
relayée par les **SCD** des universités
pour numériser et cataloguer
les thèses dans le projet **TEL**
via le dispositif **STAR**.

DE GROS EFFORTS...

Cette mise à disposition
n'induit pas une consultation
systématique des principaux intéressés:

- Etudiants;
- Enseignants;
- Chercheurs.

**... APPRÉCIÉS À LEUR JUSTE
VALEUR?**

Encore une place face à « **Amazoogle?** »

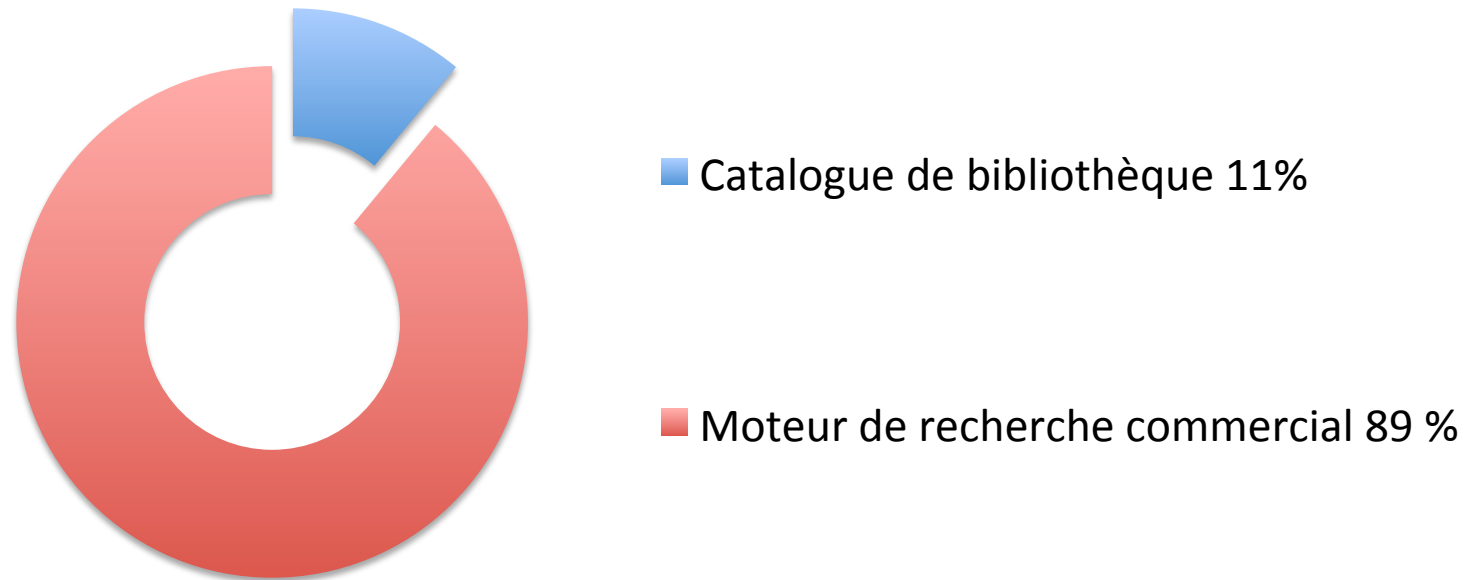
Les bibliothèques numériques *Entre modernité et désuétude*



«les catalogues
des bibliothèques
sont tombés
en disgrâce.»

Markey, K. (2007).
The online library catalog.
D-Lib Magazine, p. 13

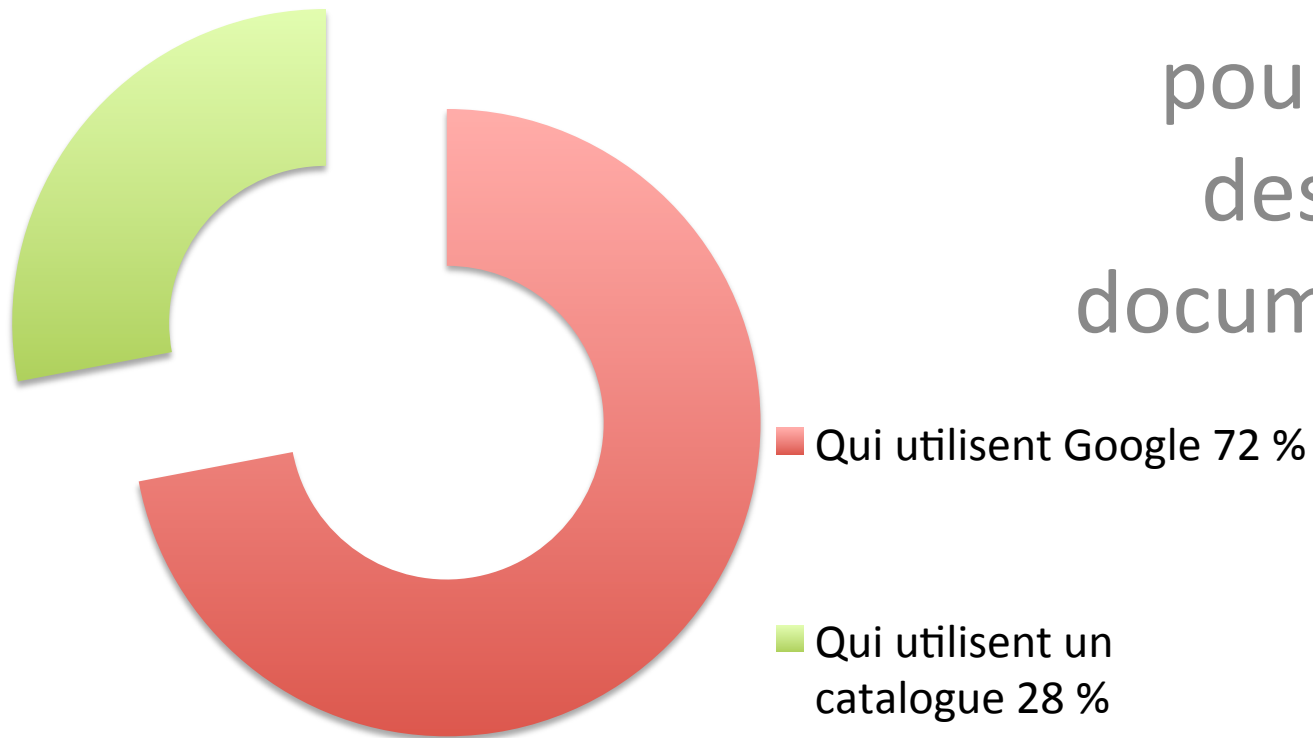
Recherche scientifique



Rares sont les étudiants, enseignants et chercheurs commençant leurs recherches par le catalogue de leur bibliothèque

Une étude de De Rosa menée en 2006 pour le compte de l'OCLC.

Les chercheurs utilisent de plus en plus Google pour trouver des sources documentaires



**SWAN, A. & BROWN, S. (2005).
OPEN ACCESS SELF-ARCHIVING AN AUTHOR STUDY**

De l'OAI au web de données : Bibliothèques
et publications sur Internet, Saint-Cloud,
Gérald KEMBELLEC

Lorcan Dempsey, directeur de la recherche à l'OCLC a qualifié notre époque d'ère d' **Amazoogole*** faisant allusion :

- au moteur de recherche **Google**
- à la librairie numérique **Amazon**

Qui sont devenus des réflexes conditionnés, même pour les chercheurs.

•
* De Rosa (2006)
College Students Perceptions
of Libraries and Information Resources

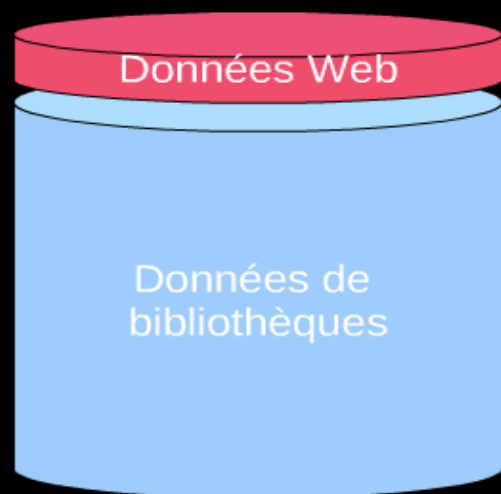
Déplacement du centre de gravité
de la RI vers les utilisateurs :
« *des nouvelles voies et de nouveaux champs
d'application
pour les catalogues en ligne. ** »

Nouveaux usages **DONC** nouveaux outils

* de Kaenel, I. & Iriarte, P. (2007).
Les catalogues des bibliothèques :
du web invisible au web social.

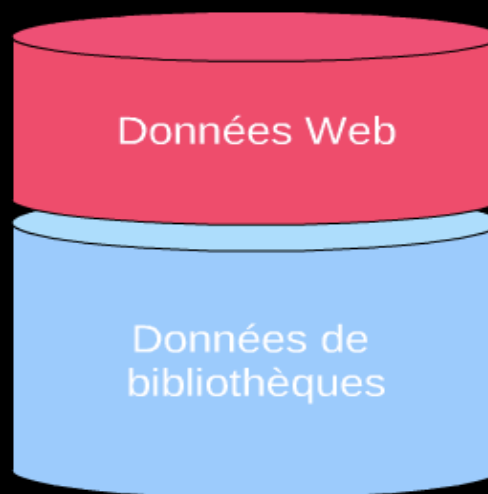
Les 3 âges des catalogues de bibliothèques sur le Web

Bermes, 2010



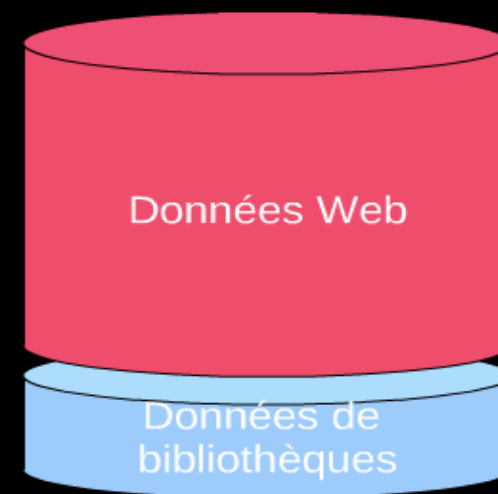
Opac Web :
seulement un accès
de surface à un silo
de données invisible

1990's



Next gen'
Catalogue :
fonctionnalités Web
2.0 + APIs pour
rendre les données
visibles

2000's



Web Catalogue :
catalogue
entièrement
conforme à
l'architecture du
Web

2010's ?

Pour modéliser un SRI adapté aux
besoins des des usagers de la
recherches...

Il faut connaître :

1. Le domaine de recherche

Pour modéliser un SRI adapté aux
besoins des des usagers de la
recherches...

Il faut connaître :

1. Le domaine de recherche
2. Les méthodes de recherche

Pour modéliser un SRI adapté aux
besoins des des usagers de la
recherches...

Il faut connaître :

1. Le domaine de recherche
2. Les méthodes de recherche
3. Les bases de connaissances

Pour modéliser un SRI adapté aux
besoins des des usagers de la
recherches...

Il faut connaître :

1. Le domaine de recherche
2. Les méthodes de recherche
3. Les bases de connaissances
4. Les outils de recherche / détection

Pour modéliser un SRI adapté aux
besoins des des usagers de la
recherches...

Il faut connaître :

1. Le domaine de recherche
2. Les méthodes de recherche
3. Les bases de connaissances
4. Les outils et méthodes de recherche / détection
5. SRI scientifique navigable

Pour modéliser un SRI adapté aux
besoins des des usagers de la
recherches...

Il faut connaître :

1. Le domaine de recherche
2. Les méthodes de recherche
3. Les bases de connaissances
4. Les outils et méthodes de recherche / détection
5. SRI scientifique navigable



*GED : Gestion Electronique
de Documents*

De plus en plus gros, de plus en plus gourmand...

Illustration : Luc Damas

La gestion du domaine de recherche

Le Lexique ?



Illustration : Luc Damas

La gestion du domaine de recherche

Le Trésor ?

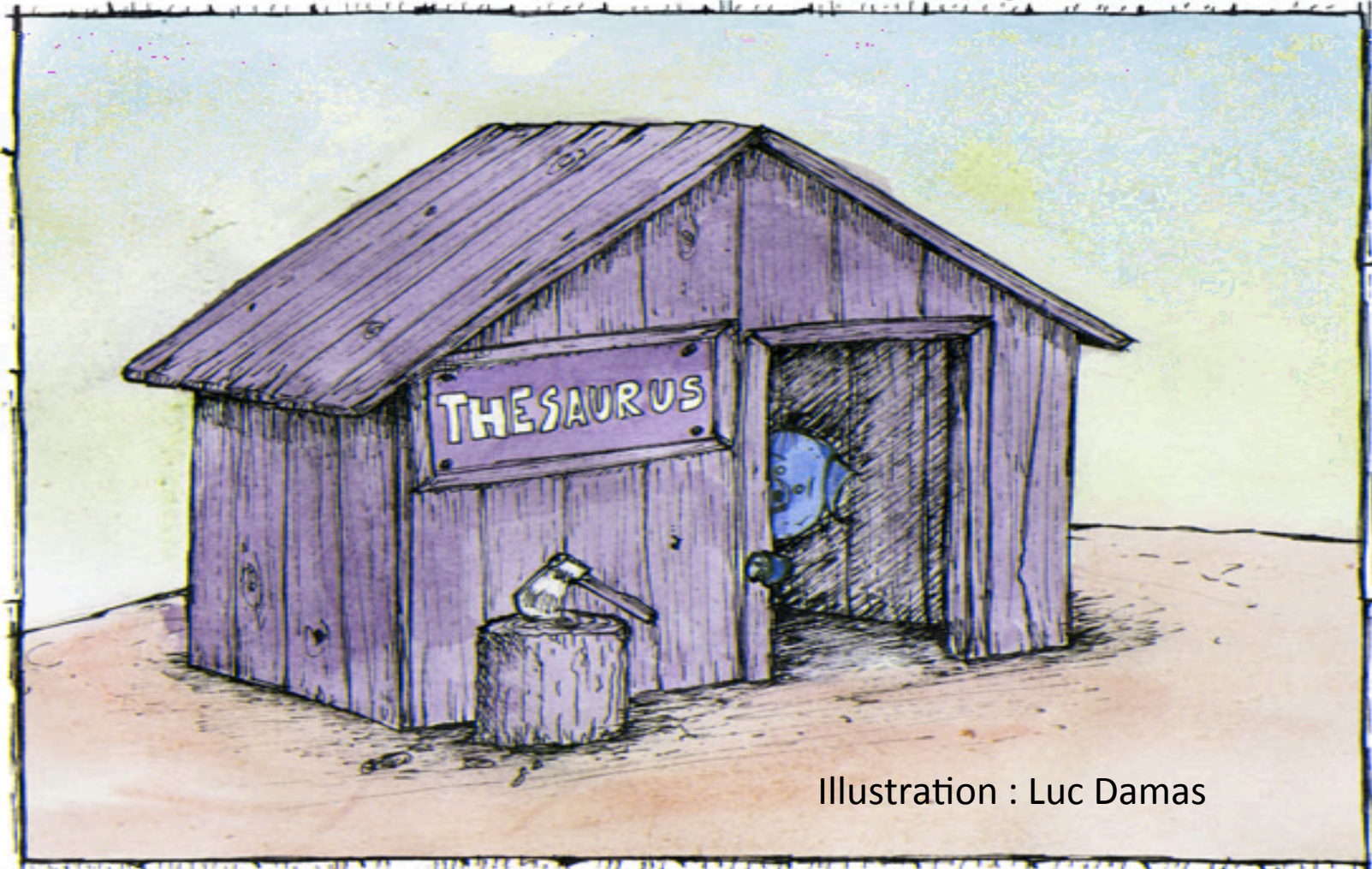


Illustration : Luc Damas

La gestion du domaine de recherche

L'Ontologie ?

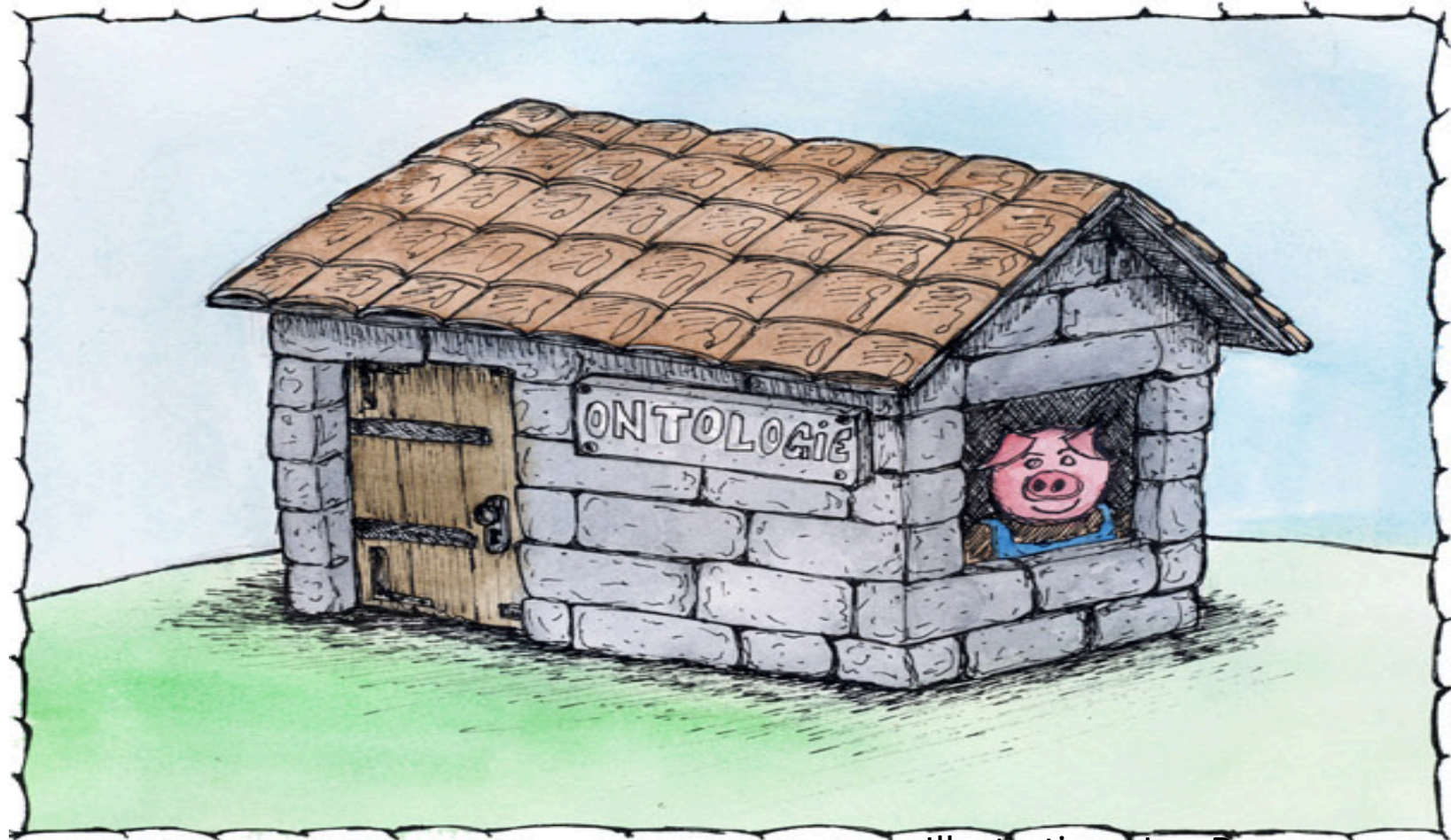


Illustration : Luc Damas

La gestion du domaine de recherche

De l'OAI au web de données : Bibliothèques
et publications sur Internet, Saint-Cloud,
Gérald KEMBELLEC



Illustration : Luc Damas

Pas de recette miracle

A minima une **Taxonomie** d'un domaine permet un classement

Un **Thésaurus** offre des relations sémantiques

Une **Ontologie** permet davantage de connexions et une recherche à facettes plus aisée...

La gestion du domaine de recherche

De l'OAI au web de données : Bibliothèques
et publications sur Internet, Saint-Cloud,
Gérald KEMBELLEC

Pour modéliser un SRI adapté aux
besoins des des usagers de la
recherches...

Il faut connaître :

1. Le domaine de recherche
- 2. Les méthodes de recherche**
3. Les bases de connaissances
4. Les outils et méthodes de recherche / détection
5. SRI scientifique navigable

L'ACCÈS AU DOMAINE DE CONNAISSANCE



© CCSD Centre pour la communication scientifique directe - <http://ccsd.cnrs.fr>

Accueil Déposer Consulter **Rechercher** S'abonner

version française RSS

.. Rechercher > Recherche simple .. ?

Veuillez décrire votre requête dans le formulaire ci-dessous

N'importe lequel des champs ci-dessus + ?

Rechercher Annuler

Trier les résultats suivant

date de publication, écriture desc.

Nombre de documents trouvés : aucun document

- Via un moteur de recherche
- Via un portail
- Ou les deux

De l'OAI au web de données : Bibliothèques
et publications sur Internet, Saint-Cloud,
Gérald KEMBELLEC

L'ACCÈS AU DOMAINE DE CONNAISSANCE

Index thématique

Matériel informatique

Alimentation • Boîtier • Carte graphique • Carte mère • Carte son • Carte TV • Clavier • Clé USB • Disque dur • Écran • Graveur • Hub ou switch • Imprimante • Lecteur de CD • Lecteur de disquette • Lecteur de DVD • Mémoire vive • Microprocesseur • Palette graphique • Processeur • Modem • Onduleur • Ordinateur • Ordinateur portable • PDA • Routeur • Scanner • Souris • Ventirad • Webcam

Systèmes d'exploitation

BeOS • BSD • DOS • GCOS • GNU/Linux • Mac OS • MVS • OS/2 • Solaris • TPF • Unix • Windows • Systèmes temps réel • VxWorks

Logiciels propriétaires et libres

3D Studio Max • Blender • Firefox • Flash • Gimp • GNOME • Internet Explorer • iTunes • KDE • LaTeX • Microsoft Office • Mozilla • MPlayer • OpenOffice.org • Paint.NET • Photoshop • Visual Studio • Winamp

Internet, société de l'information et culture informatique

Blog • Brevet logiciel • Cyberculture • Droits numériques • E-gouvernement • E-inclusion • Filtrage d'internet • Format ouvert • Fracture numérique • Geek • Monde virtuel • Monitoring de site web • Moteur de recherche • Nerd • Site Web • Société de la connaissance • Système d'information • Technologies de l'information et de la communication • Wiki

Normes et Technologies

AAC • ADSL • ATX • Baladodiffusion • Bluetooth • DVI • Ethernet • FTTH • ICC • ISO 9660 • JPEG • MIDI • MIME • Modèle OSI • MP3 • Microformats • MPEG • OpenDocument • PCI Express • PCMCIA • POSIX • Poste-à-poste • RFC • Standards du Web • TCP/IP • Unicode • USB • VGA • Wi-Fi • WiMAX • x64 • x86

Programmation

Algorithmique • Bogue • Compilateur • Débogueur • Génie logiciel • Langages (ActionScript • ASP • Assembleur • C++ • Java • Perl • PHP • Python • Scheme) • Programmation orientée objet • Programmation parallèle

Informatique théorique

Algorithmique • Méthode formelle • Théorie des graphes • Théorie de la complexité • Théorie de la calculabilité • Théorie des langages • Théorie de l'information

Sécurité informatique, cryptographie et Vie privée

Criminalité informatique • Cryptographie • Fuite d'information • Logiciel espion • Logiciel malveillant • Pare-feu • Pirate informatique • Signature numérique • Stéganographie • Virus

Organismes et organisations

Organismes internationaux : Apache • BSA • Fondation Mozilla • FSF • ICANN • IEEE • IETF • ISO • OMG • OSI • Unicode • W3C
Organismes français : Cigref • CNIL • Syntec • Munci • April

Constructeurs informatiques et éditeurs de logiciels

3dfx • Adobe • AMD • Apple • ATI • Bull • Dell • Google • Hewlett Packard • IBM • Intel Corporation • Microsoft • Motorola • NEC • Norton • NVidia • Silicon Graphics • Sun Microsystems • Xerox • Yahoo!

Personnalités

Charles Babbage • Tim Berners-Lee • Noam Chomsky • Bill Gates • Steve Jobs • Donald Ervin Knuth • Jay Miner • Gordon E. Moore • Blaise Pascal • Claude Shannon • Richard Stallman • Linus Torvalds • Bui Tuong Phong • Alan Mathison Turing • Steve Wozniak

- Via un moteur de recherche
- Via un portail
- Ou les deux

De l'OAI au web de données : Bibliothèques
et publications sur Internet, Saint-Cloud,
Gérald KEMBELLEC

L'ACCÈS AU DOMAINE DE CONNAISSANCE

Google annuaire Recherche Google [Aide sur l'Annuaire](#)

Informatique

Catégories

Actualité et médias (41)	Imagerie numérique (89)	Polices de caractères (15)
Associations et clubs (137)	Images (186)	Programmation (795)
Boutiques en ligne (55)	Intelligence artificielle (111)	Réalité virtuelle (30)
Calcul parallèle (14)	Internet (2562)	Robotique (79)
Commerce et économie (615)	Jeux (1606)	Santé (287)
Conception assistée par ordinateur (19)	Logiciel libre (195)	Sciences de l'informatique (114)
Conférences (11)	Logiciels (2886)	Sécurité (189)
Discussions (19)	Matériel (406)	Systèmes d'exploitation (505)
Emploi (17)	Multimédia (32)	Terminologie (20)
Ergonomie (50)	Musique (91)	Transferts de données (54)
Formation (405)	Newsgroups (63)	Usenet (63)
Formats de données (152)	Plateformes (94)	Vidéotex (19)

Catégorie apparentée :
[Société > Fêtes et jours fériés > Journées mondiales > Informatique](#) (3)

- Via un moteur de recherche
- Via un portail
- Ou les deux

De l'OAI au web de données : Bibliothèques
et publications sur Internet, Saint-Cloud,
Gérald KEMBELLEC

Modèles **psycho-cognitifs** de RI

Plusieurs modèles de recherche d'information proposés depuis les années 90 :

- Modèle initial de Guthrie (1988)
- Modèle « évaluation, sélection, traitement » (Dinet, Rouet 2002)
- Modèle « Information Search Process » (Kuhlthau, 1991)
- Modèle « TIMS » (Dillon 95)
- Modèle « berrypicking » (sérendipité . . .)(Bates, 1993)
- Modèle « information-seeking » (Marchionini 1997).

Modèles **psycho-cognitifs** de RI

Ce qu'il faut retenir :

Le processus de recherche, de sélection d'information est toujours un processus **cognitif** complexe découpé en tâches impactées par un facteur **psychologique** fort :

surcharge informationnelle, déficit ergonomique, bruit,
silence

=

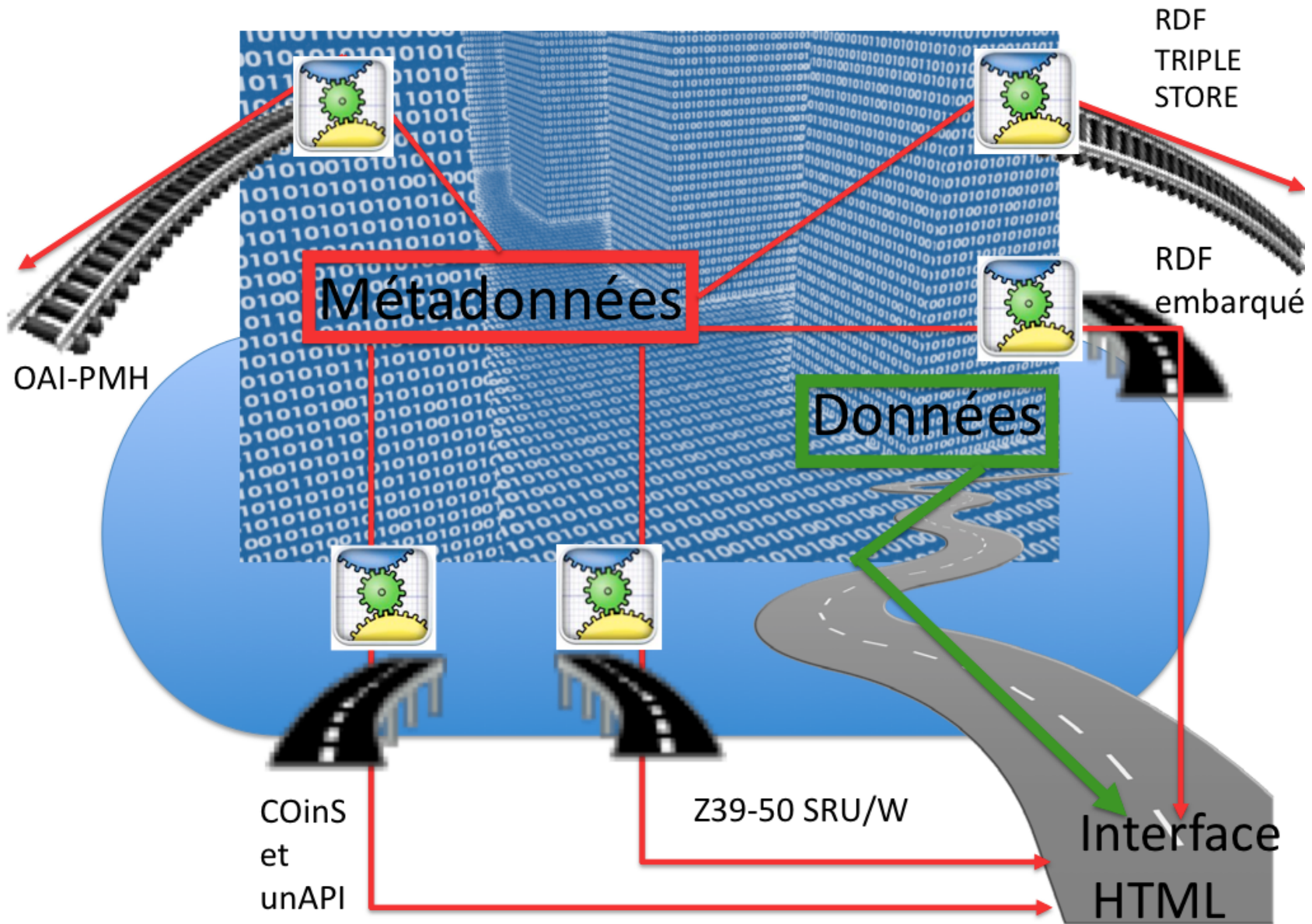
Découragement, abandon ou changement de SRI

Pour modéliser un SRI adapté aux besoins des des usagers de la recherches...

Il faut connaître :

1. Le domaine de recherche
2. Les méthodes de recherche
3. **Les bases de connaissances**
4. Les outils et méthodes de recherche / détection
5. Les outils de génération de bibliographie

Urbanisation du SI Documentaire



De l'OAI au web de données : Bibliothèques
et publications sur Internet, Saint-Cloud,
Gérald KEMBELLEC

Pour modéliser un SRI adapté aux besoins des des usagers de la recherches...

Il faut connaître :

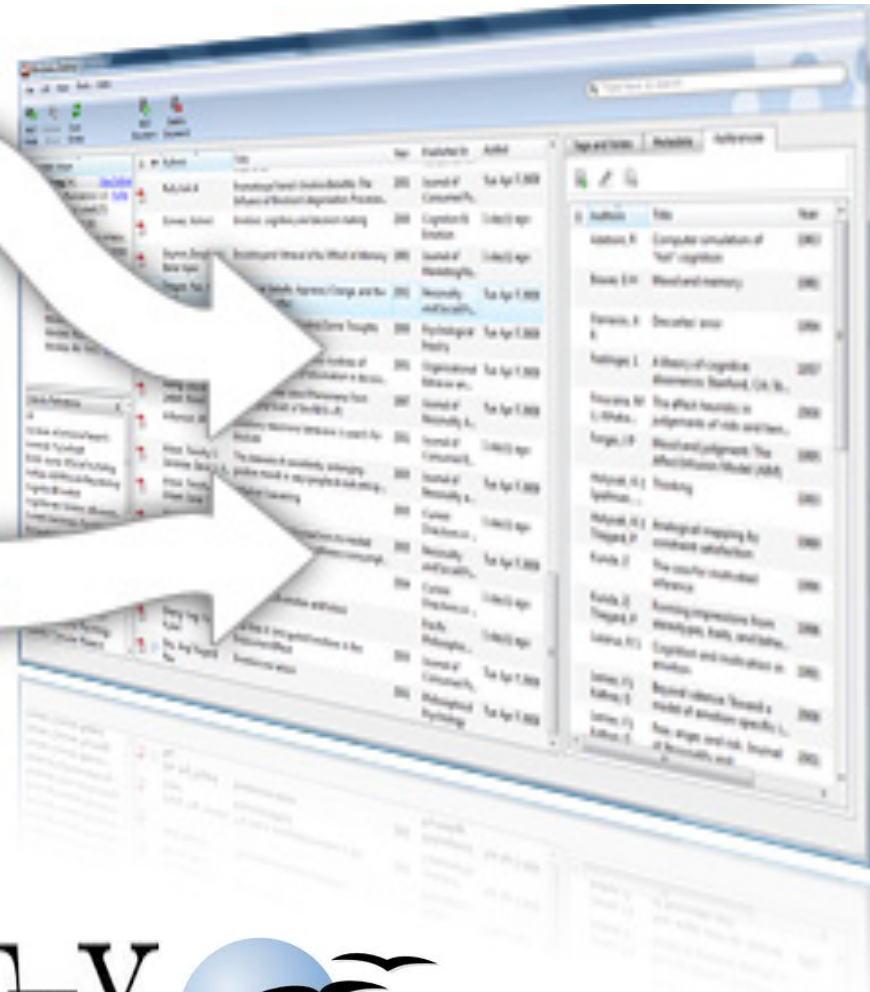
1. Le domaine de recherche
2. Les méthodes de recherche
3. Les bases de connaissances
- 4. Les outils et méthodes de recherche / détection**
5. SRI scientifique navigable

Automatiser les tâches techniques



Sync

Import



Avec les
bons
outils



LATEX



... et les bons formats d'exposition du web de données

Deux méthodes :

Le glanage (*gleaning*):

détection contextuelle sur une page Web

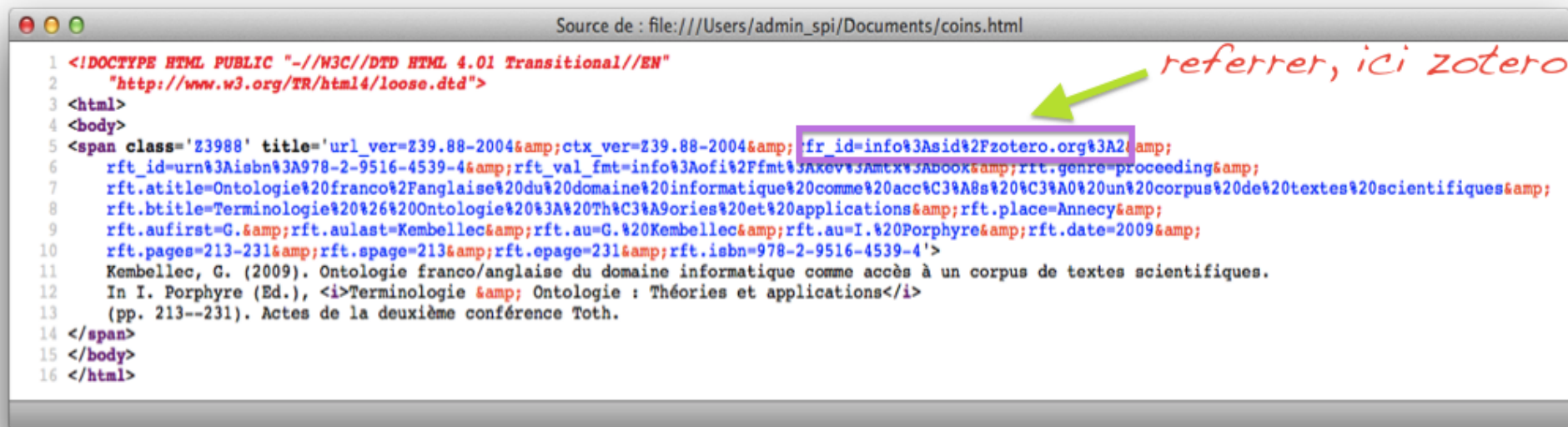
Le moissonnage (*harvesting*):

Requête sur des collections

... et les bons formats d'exposition du web de données

COinS

Le glanage (*gleaning*) avec
Zotero, Mendeley



```
Source de : file:///Users/admin_spi/Documents/coins.html
1 <!DOCTYPE HTML PUBLIC "-//W3C//DTD HTML 4.01 Transitional//EN"
2 "http://www.w3.org/TR/html4/loose.dtd">
3 <html>
4 <body>
5 <span class='Z3988' title='url_ver=Z39.88-2004&amp;ctx_ver=Z39.88-2004&amp;fr_id=info%3Asid%2Fzotero.org%3A2&amp;
6 rft_id=urn%3Aisbn%3A978-2-9516-4539-4&amp;rft_val_fmt=info%3Aofi%2Ffmt%3Akev%3Amtx%3Abook&amp;rft.genre=proceeding&amp;
7 rft.atitle=Ontologie%20franco%2Fanglaise%20du%20domaine%20informatique%20comme%20acc%3%A8s%20%20%20un%20corpus%20de%20textes%20scientifiques&amp;
8 rft.btitle=Terminologie%20%26%20Ontologie%20%3A%20Th%3A9ories%20et%20applications&amp;rft.place=Annecy&amp;
9 rft.aufirst=G.&amp;rft.aulast=Kembellec&amp;rft.au=G.%20Kembellec&amp;rft.au=I.%20Porphyre&amp;rft.date=2009&amp;
10 rft.pages=213-231&amp;rft.spage=213&amp;rft.epage=231&amp;rft.isbn=978-2-9516-4539-4'>
11 Kembellec, G. (2009). Ontologie franco/anglaise du domaine informatique comme accès à un corpus de textes scientifiques.
12 In I. Porphyre (Ed.), <i>Terminologie &amp; Ontologie : Théories et applications</i>
13 (pp. 213--231). Actes de la deuxième conférence Toth.
14 </span>
15 </body>
16 </html>
```

... et les bons formats d'exposition du web de données

Dublin Core
Embedded
metadata

Le glanage (*gleaning*)

```
Source de : http://www.geraldkembellec.fr/html_embedded_metadata.html
1 <!DOCTYPE html PUBLIC "-//W3C//DTD XHTML 1.0 Strict//EN" "http://www.w3.org/TR/xhtml1/DTD/xhtml1-strict.dtd">
2 <html xmlns="http://www.w3.org/1999/xhtml" xml:lang="fr" lang="fr">
3 <head>
4 <meta http-equiv="Content-type" content="text/html; charset=utf-8" />
5 <meta http-equiv="Content-language" content="fr" />
6 <title>Ontologie franco/anglaise du domaine informatique comme accès à un corpus de textes scientifiques</title>
7 <link rel="schema.DC" href="http://purl.org/dc/elements/1.1/" />
8 <link rel="schema.MARCREL" href="http://www.loc.gov/loc/terms/relators/" />
9 <meta name="DC.type" content="Text" />
10 <meta name="DC.identifier" scheme="URI" content="http://ontology.univ-savoie.fr/toth/TOTH2008_actes.pdf" />
11 <meta name="DC.identifier" scheme="ISBN" content="978-2-9516-4539-4" />
12 <meta name="DC.title" content="Ontologie franco/anglaise du domaine informatique comme accès à un corpus de textes scientifiques" />
13 <meta name="DC.publisher" content="Porphyre Éditions" />
14 <meta name="DC.location" content="Porphyre Éditions" />
15 <meta name="DC.language" scheme="RFC3066" content="fr" />
16 <meta name="DC.type" content="inproceedings" />
17 <meta name="author" content="Kembellec, Gérald" />
18 <meta name="DC.creator" content="Kembellec, Gérald" />
19 <meta name="keywords" content="ontologie, informatique, CLIR, IR" />
20 <meta name="keywords" xml:lang="fr" lang="fr" content="ontologie, informatique, CLIR, IR" />
21 <meta name="DC.subject" xml:lang="fr" lang="fr" content="ontologie, informatique, CLIR, IR" />
22 <meta name="DC.date" scheme="W3CDTF" content="2008-05-01" />
23 <meta name="DC.relation.isPartOf" content="Actes de la deuxième conférence Toth" />
24 <meta name="DC.source" content="Actes de la deuxième conférence Toth" />
25 <meta name="DC.coverage" content="France" />
26 <meta name="citation_conference" content="Terminologie et Ontologie : Théories et applications" />
27 <meta name="citation_publisher" content="Porphyre Éditions" />
28 <meta name="citation_authors" content="Kembellec, Gérald" />
29 <meta name="citation_title" content="Ontologie franco/anglaise du domaine informatique comme accès à un corpus de textes scientifiques" />
30 <meta name="citation_date" content="2008-05-01" />
31 <meta name="citation_year" content="2008" />
32 <meta name="citation_Event_place" content="Annecy" />
33 <meta name="citation_isbn" content="978-2-9516-4539-4" />
34 <meta name="citation_firstpage" content="213" />
35 <meta name="citation_lastpage" content="231" />
36 <meta name="citation_language" content="fr" />
37 <meta name="citation_keywords" content="ontologie, informatique, CLIR, IR" />
38 <meta name="citation_pdf_url" content="http://ontology.univ-savoie.fr/toth/TOTH2008_actes.pdf" />
39 </head>
40 <body>
41 Kembellec, G. (2009). Ontologie franco/anglaise du domaine informatique comme accès à un corpus de textes scientifiques.
42 In I. Porphyre (Ed.), <i>Terminologie & Ontologie : Théories et applications</i>
43 (pp. 213--231). Actes de la deuxième conférence Toth.
44 </body>
45 </html>
```

*Liens vers les vocabulaires
de description de métas*

... et les bons formats d'exposition du web de données

unAPI

Le glanage (*gleaning*)

```
Source de : http://www.geraldkembellec.fr/unAPI/index.html
1 <!DOCTYPE html PUBLIC "-//W3C//DTD HTML 4.01 Transitional//EN" "http://www.w3.org/TR/html4/loose.dtd">
2 <html>
3   <head profile="http://a9.com/-/spec/opensearch/1.1/">
4     <title>Ontologie franco/anglaise du domaine informatique comme accès à un corpus de textes scientifiques</title>
5     <meta http-equiv="content-language" content="fr">
6     <meta http-equiv="content-type" content="text/html; charset=ISO-8859-1">
7     <meta http-equiv="Content-Style-Type" content="text/css">
8     <link rel="stylesheet" href="http://www.geraldkembellec.fr/refbase/css/style.css" type="text/css" title="CSS Definition">
9     <link rel="unapi-server" type="application/xml" title="unAPI" href="http://www.geraldkembellec.fr/refbase/unapi.php">
10    <script language="JavaScript" type="text/javascript" src="refbase/javascript/common.js"></script>
11    <script language="JavaScript" type="text/javascript" src="/refbase/javascript/prototype.js"></script>
12    <script language="JavaScript" type="text/javascript" src="/refbase/javascript/scriptaculous.js?load=effects,controls"></script>
13  </head>
14  <body>
15    <table>
16      <tbody>
17        <tr>
18          <td colspan="8" class="smaller" align="center">
19            <a href="http://www.geraldkembellec.fr/refbase/show.php?record=58" title="copy this URL to directly link to this record"></a>
20            <div class="unapi">
21              <abbr class="unapi-id" title="http://www.geraldkembellec.fr/refbase/show.php?record=58">{:abbr}</abbr>
22            </div>
23          </td>
24          <td id="ref58" class="citation" valign="top">
25            Kembellec, G. (2009). Ontologie franco/anglaise du domaine informatique comme accès à un corpus de textes scientifiques.
26            In I. Porphyre (Ed.), <i>Terminologie & Ontologie : Théories et applications</i>
27            (pp. 213--231). Actes de la deuxième conférence Toth.
28          </td>
29        </tr>
30      </tbody>
31    </table>
32  </body>
33 </html>
```

URL du résolveur

identifiant de l'objet

... et les bons formats d'exposition du web de données

OAI-PMH

Le moissonnage (*harvesting*) avec ORI-OAI

The screenshot shows an XML document with the following structure and annotations:

- Request (Requête):** Lines 6-7 show the request parameters: `<request verb="GetRecord" metadataPrefix="oai_dc" identifier="oai:hal.archives-ouvertes.fr:hal-00628355">`. The `verb` attribute is annotated with "Verbe".
- Response (Réponse):** Lines 8-44 show the response structure, including a `<record>` element with a `<header>` and `<metadata>` section.
- Annotations:**
 - "format de retour" points to the `metadataPrefix="oai_dc"` attribute in the request.
 - "identifiant de la fiche" points to the `identifier="oai:hal.archives-ouvertes.fr:hal-00628355"` attribute in the request.
 - "adresse du document primaire" points to the `<dcterms:identifier>http://hal.archives-ouvertes.fr/docs/00/62/83/55/PDF/TOTh2008_actes-2.pdf</dcterms:identifier>` element in the metadata.

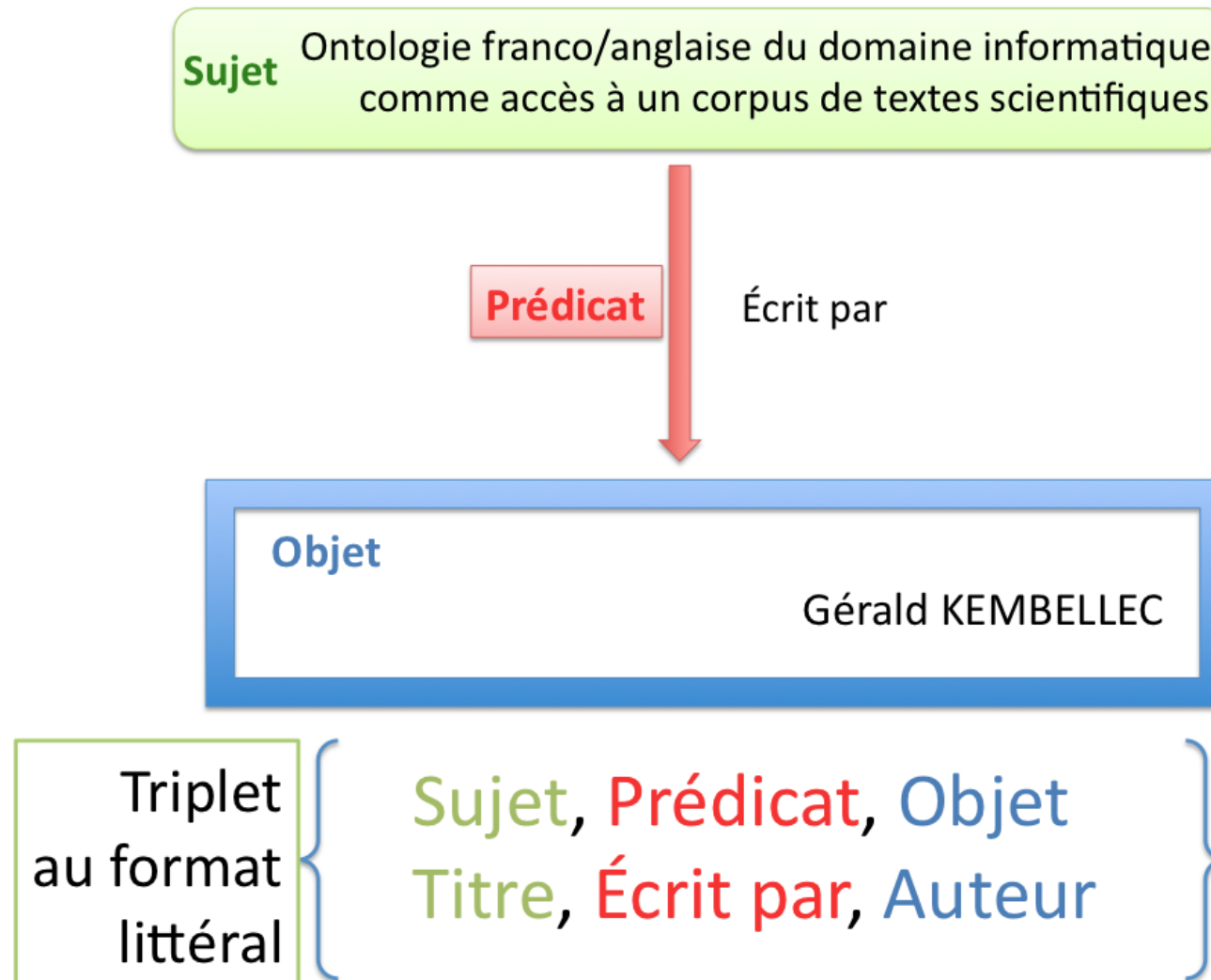
... et les bons formats d'exposition du web de données

RDF Triple Store

Le moissonnage (*harvesting*) en SPARQL

```
Source de : file:///Users/admin_spi/Documents/RDF_paper.xml
1 <?xml version="1.0" encoding="UTF-8"?>
2 <rdf:RDF
3   xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
4   xmlns:dc="http://purl.org/dc/elements/1.0/"
5   xmlns:foaf="http://xmlns.com/foaf/0.1/"
6   xmlns:dcq="http://purl.org/dc/qualifiers/1.0/"
7 <rdf:Description rdf:about="http://hal.archives-ouvertes.fr/hal-00628355/">
8   <dc:title>
9     Ontologie franco/anglaise du domaine informatique comme accès à un corpus de textes scientifiques
10  </dc:title>
11  <dc:creator>
12    <foaf:Person>
13      <foaf:name>Gérald Kembellec</foaf:name>
14      <foaf:surname>Kembellec</foaf:surname>
15      <foaf:givenname>Gérald</foaf:givenname>
16      <foaf:homepage rdf:resource="http://dblp.l3s.de/d2r/resource/authors/Gérald_Kembellec"/>
17    </foaf:Person>
18  </dc:creator>
19  <dc:subject>Ontologie de domaine; CLIR; Bibliothèque numérique; KBS</dc:subject>
20  <dc:publisher>Institut Porphyre</dc:publisher>
21  <dc:date rdf:datatype="http://www.w3.org/2001/XMLSchema#gYear">2008</dc:date>
22 </rdf:Description>
23 </rdf:RDF>
```

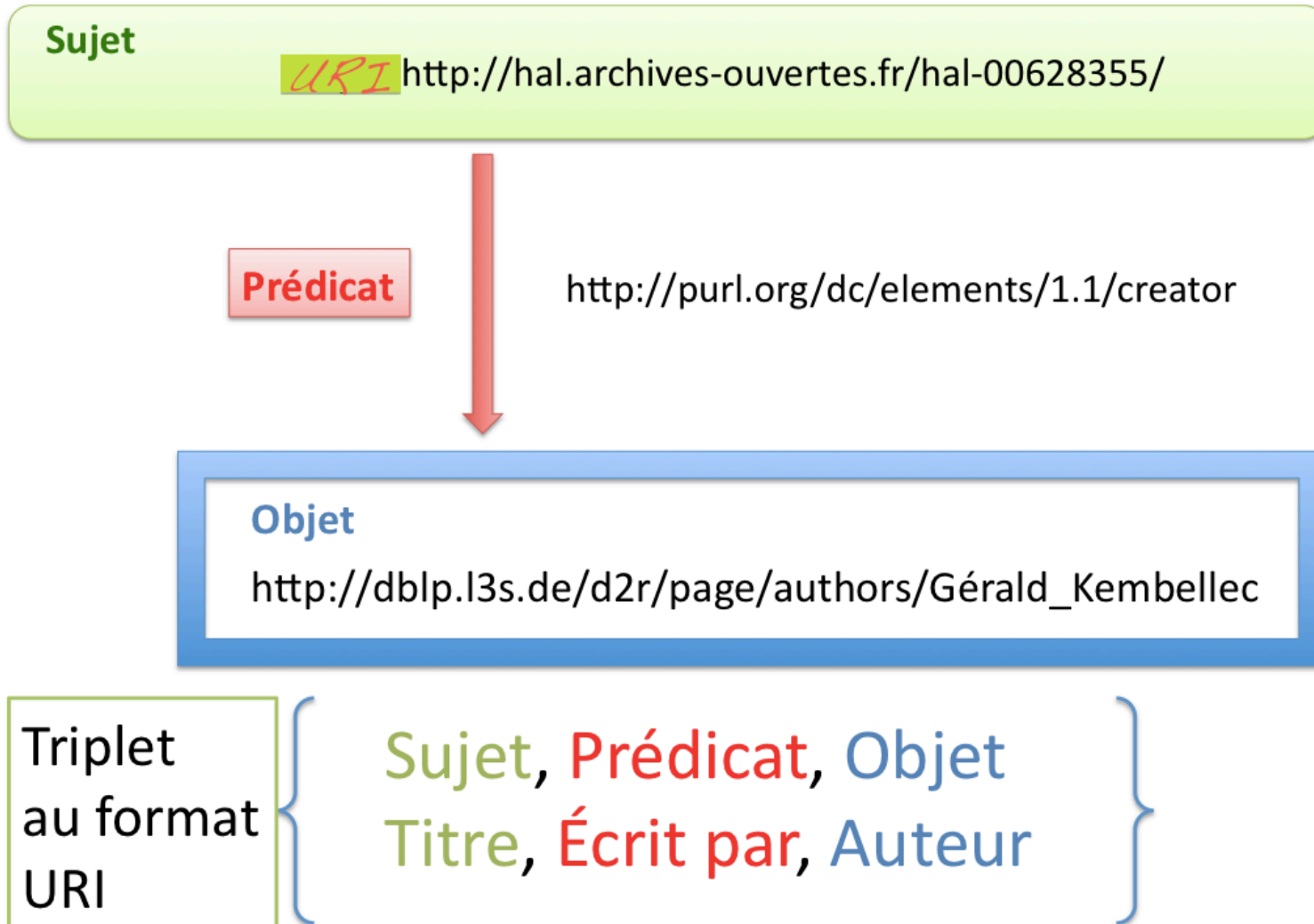

Exemple de triplet RDF en format littéral



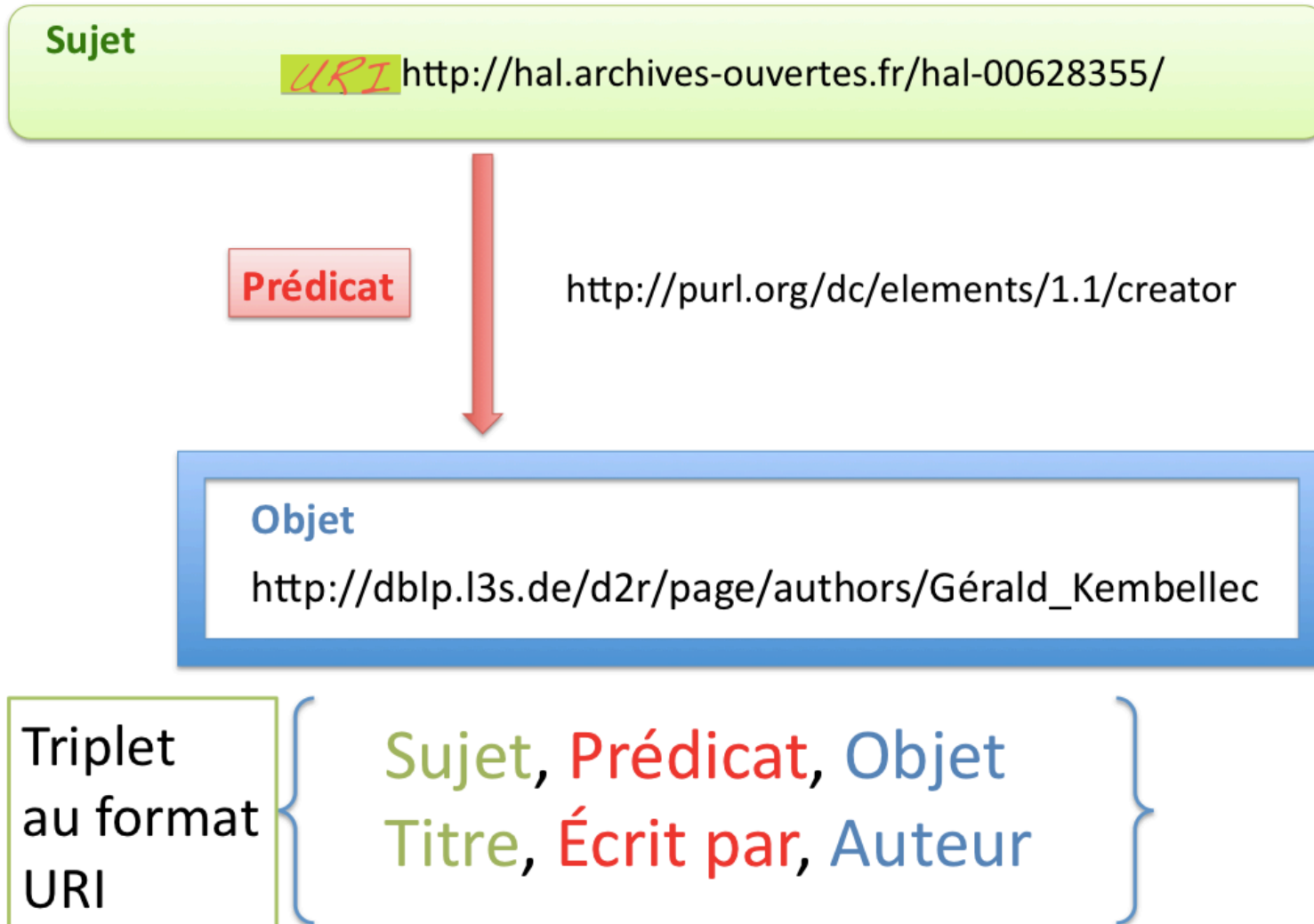
Se lit en littéral: « **Ontologie franco/anglaise ... Scientifiques** **est écrit par** Gérald KEMBELLEC »

De l'OAI au web de données : Bibliothèques
et publications sur Internet, Saint-Cloud,
Gérald KEMBELLEC

Exemple du *même* triplet RDF en URI

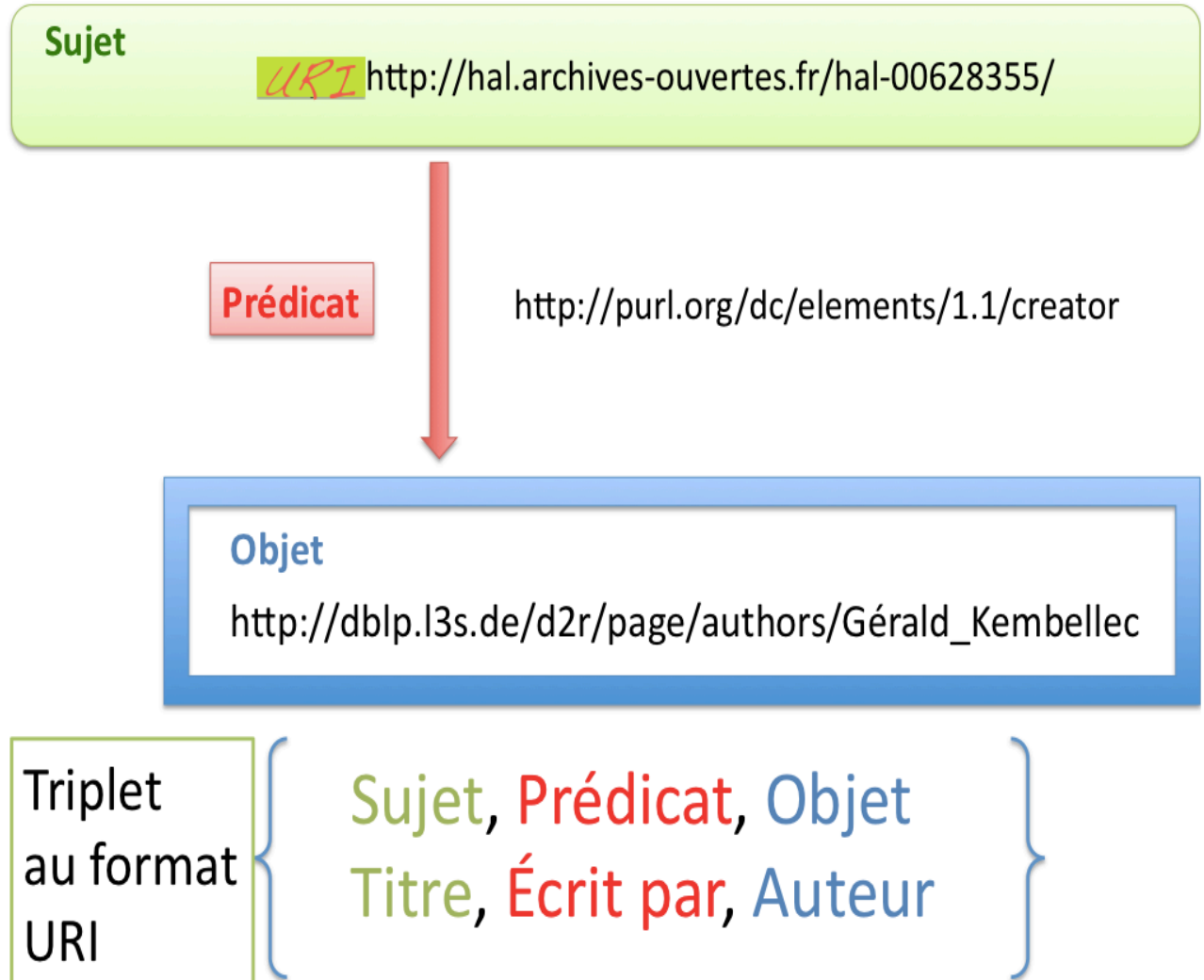


Exemple du *même* triplet RDF en URI



Exemple du *même* triplet RDF en URI

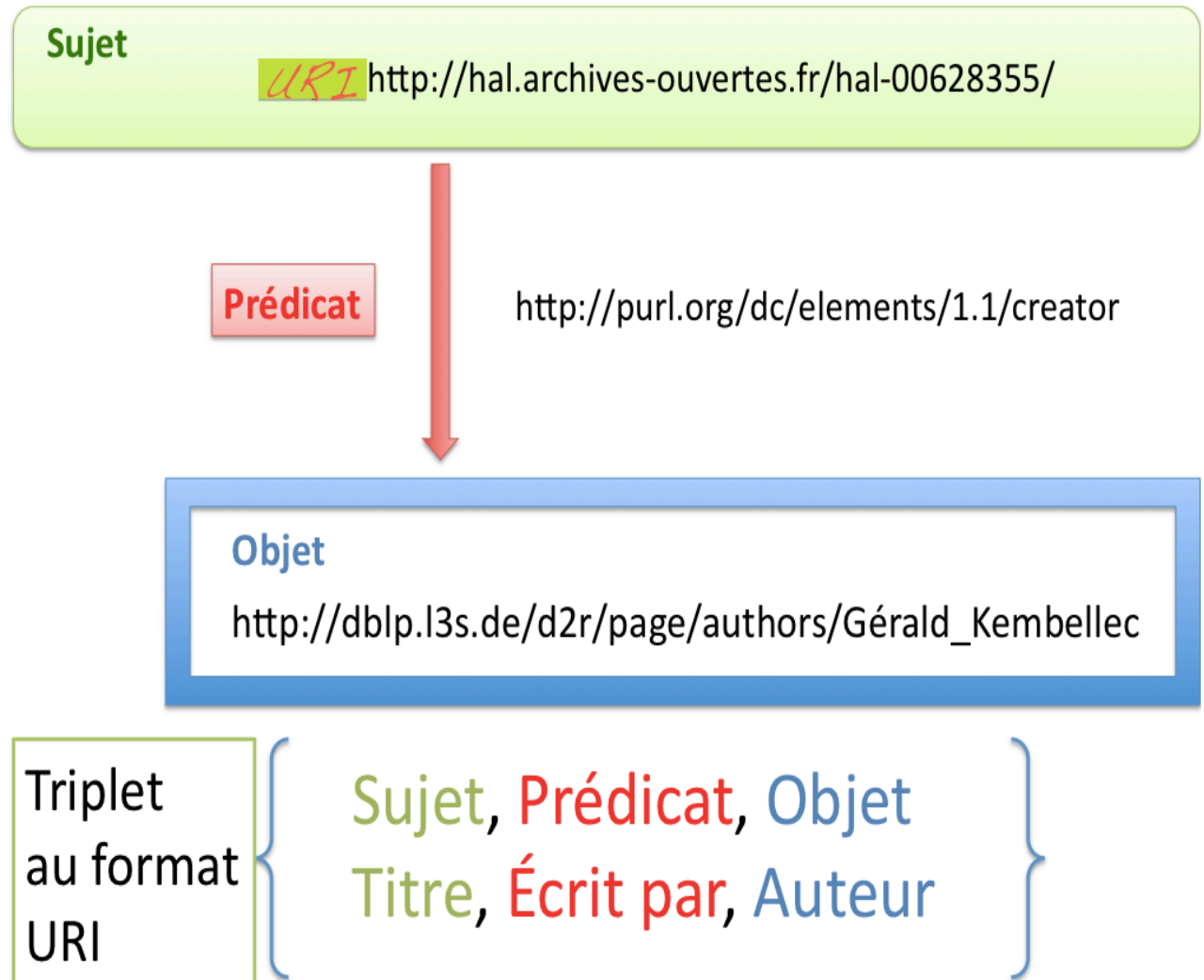
1. L'article sur HAL



Exemple du *même* triplet RDF en URI

1. L'article sur HAL

2. Est écrit
ou en format
dublin core
« creator »



Exemple du *même* triplet RDF en URI

1. L'article sur HAL

Sujet

URI <http://hal.archives-ouvertes.fr/hal-00628355/>

2. Est écrit
ou en format
dublin core
« creator »

Prédicat

<http://purl.org/dc/elements/1.1/creator>

3. L'auteur présenté
à l'URI
(adresse) indiquée

Objet

http://dblp.l3s.de/d2r/page/authors/Gérald_Kembellec

Triplet
au format
URI

Sujet, Prédicat, Objet
Titre, Écrit par, Auteur

... et les bons formats d'exposition du web de données

Z 39-50 SRU-W

Le moissonnage (*harvesting*) par URL ou client

```
1 <?xml version="1.0"?>
2 <zs:searchRetrieveResponse xmlns:zs="http://www.loc.gov/zing/srw/">
3   <zs:version>1.1</zs:version>
4   <zs:numberOfRecords>4</zs:numberOfRecords>
5   <zs:records>
6     <zs:record>
7       <zs:recordSchema>info:srw/schema/1/dc-v1.1</zs:recordSchema>
8       <zs:recordPacking>xml</zs:recordPacking>
9       <zs:recordData>
10        <srw_dc:dc xmlns:srw_dc="info:srw/schema/1/dc-schema"
11          xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
12          xmlns="http://purl.org/dc/elements/1.1/"
13          xsi:schemaLocation="info:srw/schema/1/dc-schema http://www.loc.gov/standards/sru/resources/dc-schema.xsd">
14          <title>A framework for web science </title>
15          <creator>Berners-Lee, Tim.</creator>
16          <type>text</type>
17          <publisher>Boston : Now,</publisher>
18          <date>c2006.</date>
19          <language>eng</language>
20          <description>Includes bibliographical references (p. 115-134).</description>
21          <description>Originally published as Foundations and trends in web science ; vol. 1, issue 1.</description>
22          <subject>World Wide Web--Technological innovations.</subject>
23          <subject>Web services.</subject>
24          <subject>World Wide Web--Social aspects.</subject>
25          <identifier>URN:ISBN:1933019336 (pbk.)</identifier>
26          <identifier>URN:ISBN:9781933019338 (pbk.)</identifier>
27        </srw_dc:dc>
28      </zs:recordData>
29      <zs:recordPosition>1</zs:recordPosition>
30    </zs:record>
31  </zs:records>
32 </zs:searchRetrieveResponse>
```

Source de : file:///Users/admin_spi/Documents/srw_dc.xml

Ligne 29, Colonne 53

Pour modéliser un SRI adapté aux besoins des des usagers de la recherches...

Il faut connaître :

1. Le domaine de recherche
2. Les méthodes de recherche
3. Les bases de connaissances
4. Les outils et méthodes de recherche / détection
5. Les outils de génération de bibliographie

Modélisation de l'outil

Domaine de connaissance : L'informatique



Arbre de concepts

Taxonomie

Thésaurus

Ontologie

traitement de l'information (3231)

traitement de texte (3236)

traitement des données (3236)

traitement des images

USE traitement des données (3236)

traitement électronique des données

USE traitement des données (3236)

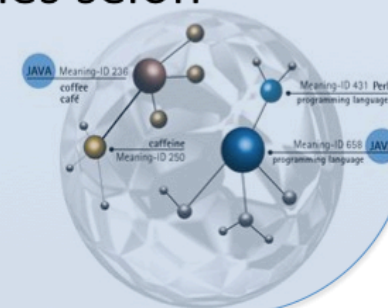
transfert d'information

USE échange d'information (3231)

translittération (3231)

Liste organisée de termes normalisés servant à l'indexation des documents

Graphe d'objets liés selon une sémantique normalisée



Représentation graphique, Zhang 2008

- Le paradigme QB

(Query searching and Browsing / Requête et navigation)

- Le paradigme BQ

(Browsing and Query searching / Navigation et requête).

- Le paradigme BO

(Browsing Only / Navigation seule).

Représentation graphique Zhang étendue (proposition)

- Le paradigme QBQ-S

(Query, Browsing, Query-Searching
Requête, navigation – requête)

créer un compte en 3 clics !

Navigateur d'Ontologie Informatique

Visualisation d'une ontologie de domaine relative à l'informatique par un Topic Mapper

systèmes à usage spécifique et orientés applicatifs Vous allez choisir une traduction alternative du noeud sélectionné

Faites un choix
systèmes à usage spécifique et orientés applicatifs

Chercher

Session

Articles relatifs

• choose a library

C.3
NEC 7720
RDBM

Description du noeud : C.3

nom français
usages spéciaux et applications systèmes

english name
special-purpose and application-based systems
Proposer une autre traduction

compose
C..organisation du système des ordinateurs

isRelatedTo
J.7.les ordinateurs dans d'autres systèmes

isComposedBy
C.3.a.microprocesseur / micro-ordinateur demandes

UNE TAXONOMIE ACCESSIBLE PAR MOTEUR

En langage naturel

De l'OAI au web de données : Bibliothèques
et publications sur Internet, Saint-Cloud,
Gérald KEMBELLEC

créer un compte en 3 clics !

Navigateur d'Ontologie Informatique

Visualisation d'une ontologie de domaine relative à l'informatique par un Topic Mapper

systèmes à usage spécifique et orientés applicatifs Vous allez choisir une traduction alternative du noeud sélectionné

Faites un choix
systèmes à usage spécifique et orientés applicatifs

Chercher

Session

 login
 passwd

Articles relatifs

- choose a library

Description du noeud : C.3

nom français
usages spéciaux et applications systèmes

english name
special-purpose and application-based systems

Proposer une autre traduction

compose
C..organisation du système des ordinateurs

isRelatedTo
J.7.les ordinateurs dans d'autres systèmes

isComposedBy
C.3.a.microprocesseur / micro-ordinateur demandes

UNE TAXONOMIE NAVIGABLE

Pour le contexte de recherche

De l'OAI au web de données : Bibliothèques
et publications sur Internet, Saint-Cloud,
Gérald KEMBELLEC

créer un compte en 3 clics !

Navigateur d'Ontologie Informatique

Visualisation d'une ontologie de domaine relative à l'informatique par un Topic Mapper

Vous allez choisir une traduction alternative du noeud sélectionné

systèmes à usage spécifique et orientés applicatifs ▾
 Faites un choix
 systèmes à usage spécifique et orientés applicatifs

Chercher

Session

 login
 passwd

Articles relatifs

- choose a library

 ▾

Description du noeud : C.3

nom français
usages spéciaux et applications systèmes

english name
special-purpose and application-based systems

Proposer une autre traduction

compose
C..organisation du système des ordinateurs

isRelatedTo
J.7.les ordinateurs dans d'autres systèmes

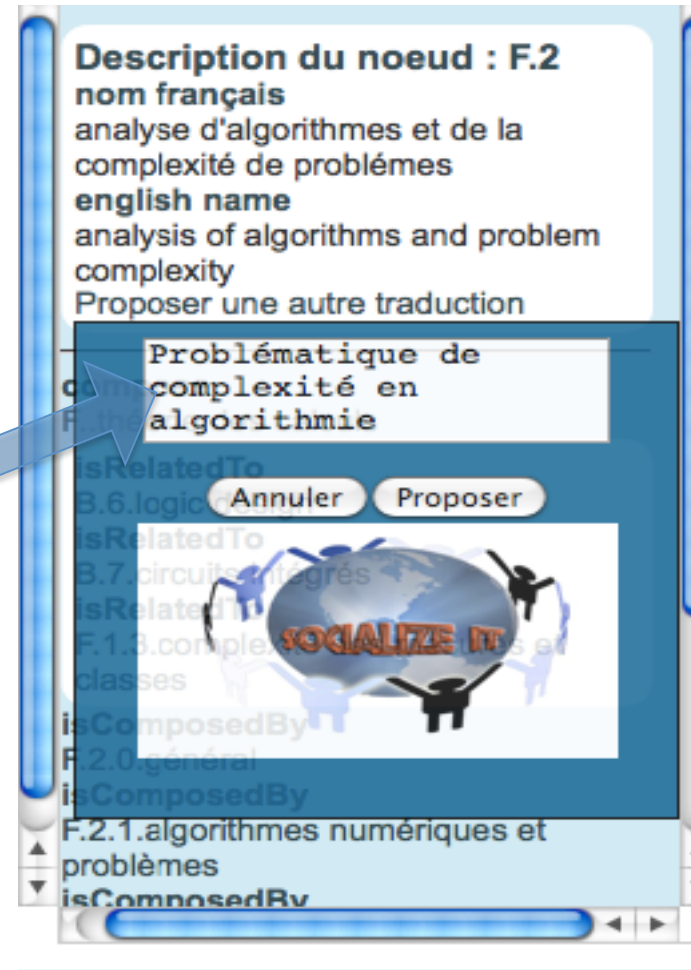
isComposedBy
C.3.a.microprocesseur / micro-ordinateur demandes

UNE POSSIBILITÉ DE FOCUS

Pour affiner la recherche et/ou rebondir

De l'OAI au web de données : Bibliothèques
et publications sur Internet, Saint-Cloud,
Gérald KEMBELLEC

Possibilité de proposer des traductions différentes et d'ajouter des métadonnées



Social Tagging, enrichissement commun et personnel

Folksonomie

créer un compte en 3 clics !

Navigateur d'Ontologie Informatique

Visualisation d'une ontologie de domaine relative à l'informatique par un Topic Mapper

systèmes à usage spécifique et orientés applicatifs Vous allez choisir une traduction alternative du noeud sélectionné

Faites un choix
systèmes à usage spécifique et orientés applicatifs

Chercher

Session

 login
 passwd

Description du noeud : C.3

nom français
usages spéciaux et applications systèmes

english name
special-purpose and application-based systems
Proposer une autre traduction

compose
C..organisation du système des ordinateurs

isRelatedTo
J.7.les ordinateurs dans d'autres systèmes

isComposedBy
C.3.a.microprocesseur / micro-ordinateur demandes

Articles relatifs

- choose a library

C.3
NEC 7720
RDBM

Accès à une base locale issue de DBLP

De l'OAI au web de données : Bibliothèques
et publications sur Internet, Saint-Cloud,
Gérald KEMBELLEC

créer un compte en 3 clics !

Navigateur d'Ontologie Informatique

Visualisation d'une ontologie de domaine relative à l'informatique par un Topic Mapper

systèmes à usage spécifique et orientés applicatifs Vous allez choisir une traduction alternative du noeud sélectionné

Faites un choix
systèmes à usage spécifique et orientés applicatifs

Chercher

Session

 login
 passwd

Description du noeud : C.3

nom français
usages spéciaux et applications systèmes

english name
special-purpose and application-based systems
Proposer une autre traduction

compose
C..organisation du système des ordinateurs

isRelatedTo
J.7.les ordinateurs dans d'autres systèmes

isComposedBy
C.3.a.microprocesseur / micro-ordinateur demandes

Articles relatifs

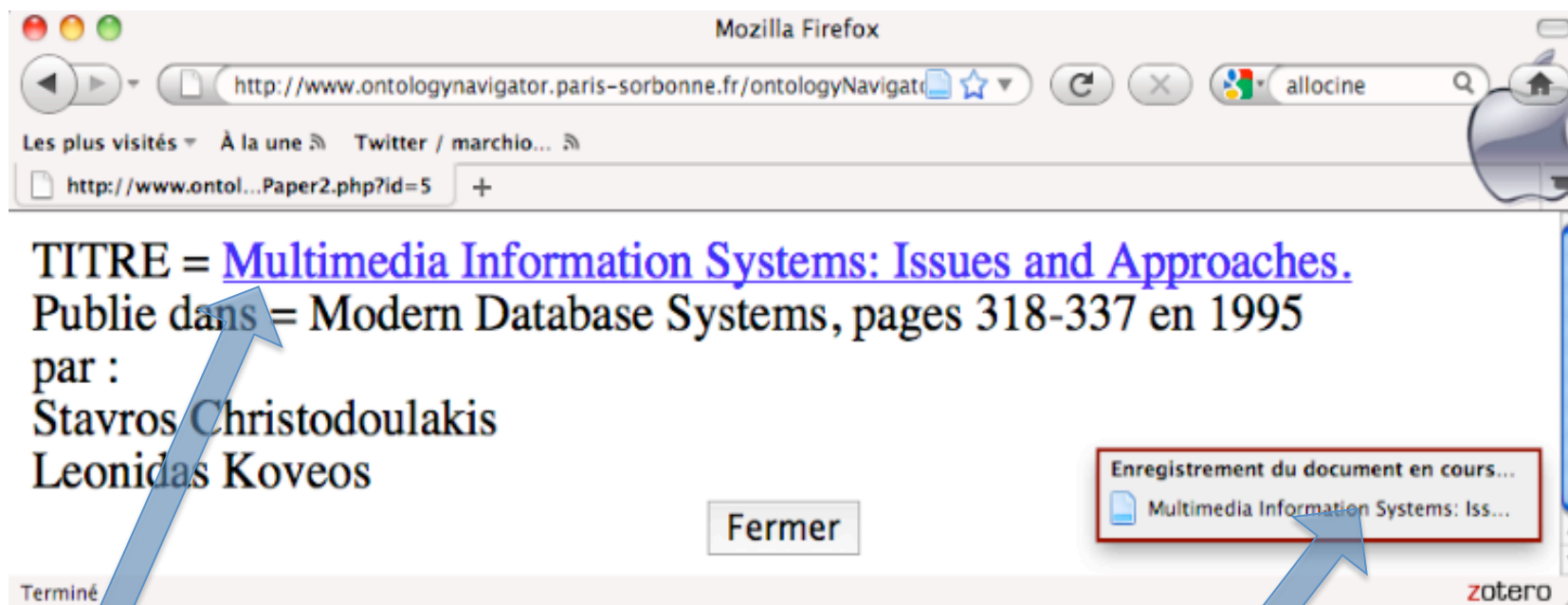
• choose a library

local library

- local library
- Google Scholar
- ACM Library
- BASE Lab
- MS academic
- Worlcat
- CiteSeer
- DBLP
- CSbib
- del.icio.us
- HAL
- EduTice (CNRS)
- arXiv
- IEEE Xplore
- crossRef

ET AUX BASES DE CONNAISSANCES EXTERNES

De l'OAI au web de données : Bibliothèques et publications sur Internet, Saint-Cloud,
Gérald KEMBELLEC



aux **métadonnées normalisées COiNS (Z39-88)**

pour le **GLANAGE bibliographique** avec Zotero,
Mendeley....

UN HYPERLIEN DIRECT VERS GOOGLE SCHOLAR

TRAVAUX FUTURS

- Proposition d'un système de recommandation (assistance à la sélection/décision)
- Méthode de recherche fédérées sur les bases externes (Z39-50, SRU-W, OAI-PMH, SPARQL)...

...dès que l'offre les bases de connaissances spécialisées en ligne seront moissonnables de

manière normalisée.

De l'OAI au web de données : Bibliothèques
et publications sur Internet, Saint-Clément,
Gérald KEMBELLEC

Gérald KEMBELLEC

Pour modéliser un SRI efficace en bibliothèque numérique

EN

CONCLUSION

**Tenir compte des modèles (psycho) cognitifs
établis pour :**

- Faire progresser les usagers en méthode de RI
- Les aider à maîtriser le domaine de connaissance par la visualisation
- Leur donner confiance en eux (trouver facilement)

EN

Pour modéliser un SRI efficace en bibliothèque numérique

CONCLUSION

Normaliser la présentation des métas :

- Être compatible avec les outils de moissonnage et glanage
- Automatiser le processus d'intégration des notices

Merci pour votre attention

Merci pour votre
attention

des questions ?

- Bates, M. (1993). **The design of browsing and berrypicking techniques for the online search interface.** *Online Information Review*, 13, 407–424
- Dinet, J. et Rouet, J. (2002). **La recherche d'information : processus cognitifs, facteurs de difficultés et dimension de l'expertise.** In C. Paganelli, ed., *Interaction homme-machine et recherche d'information*, 133–161, Hermès
- de Kaenel, I. et Iriarte, P. (2007). **Les catalogues des bibliothèques : du web invisible au web social.** *Revue électronique de science de l'information*, 1
- Kuhlthau, C.C. (1991). **Inside the search process : Information seeking from the user's perspective.** *Journal of the American Society for Information Science*, 42, 361–371.
- de Rosa, C. (2006). **College Students Perceptions of Libraries and Information Resources. A Report to the OCLC Membership. OCLC.**
- Dillon, A. (1996). **TIMS : A framework for the design of usable electronic text.** *Cognitive Aspects of Electronic Text Processing*, 99–120
- Guthrie, J. (1988). **Locating information in documents : examination of a cognitive model.** *Reading Research Quarterly*, 23, 178–199.
- Marchionini, G. (1997). **Information seeking in electronic environments.** Cambridge University Press
- Markey, K. (2007). **The online library catalog.** *D-Lib Magazine*, p. 13
- Swan, A. et Brown, S. (2005). **Open access self-archiving an author study.** Tech. rep., Key Perspectives Limited, <http://cogprints.org/4385/01/jisc2.pdf>, (consulté le 01/07/12).
- Zhang, W. (2002). **Developing web-enhanced learning for information fluency : A liberal arts college.** *Reference & User Services Quarterly*, 41
- Zhang, J. et Marchionini, G. (2004). **Coupling browse and search in highly interactive user interfaces : a study of the relation browser++.** In *Digital Libraries, 2004. Proceedings of the 2004 Joint ACM/IEEE Conference on*
- Zhang, J. (2008). **Visualization for information retrieval, vol. 23.** Springer.