

Retour d'expérience



Collecte automatique de sommaires via les
entrepôts OAI des portails de revues.

Sophie Fotiadi, Sciences Po Lyon
sophie.fotiadi@univ-lyon2.fr



Collecte automatique de sommaires via les entrepôts OAI
Frédoc 2009



Somrev (Sommaires de revues)



- <http://www.somrev.info>
- Une base de sommaires de revues produite à SciencesPo Lyon depuis 1998
- De nombreuses évolutions : partenariats, liens vers le texte intégral, moissonnage,...

les 141 revues de Somrev

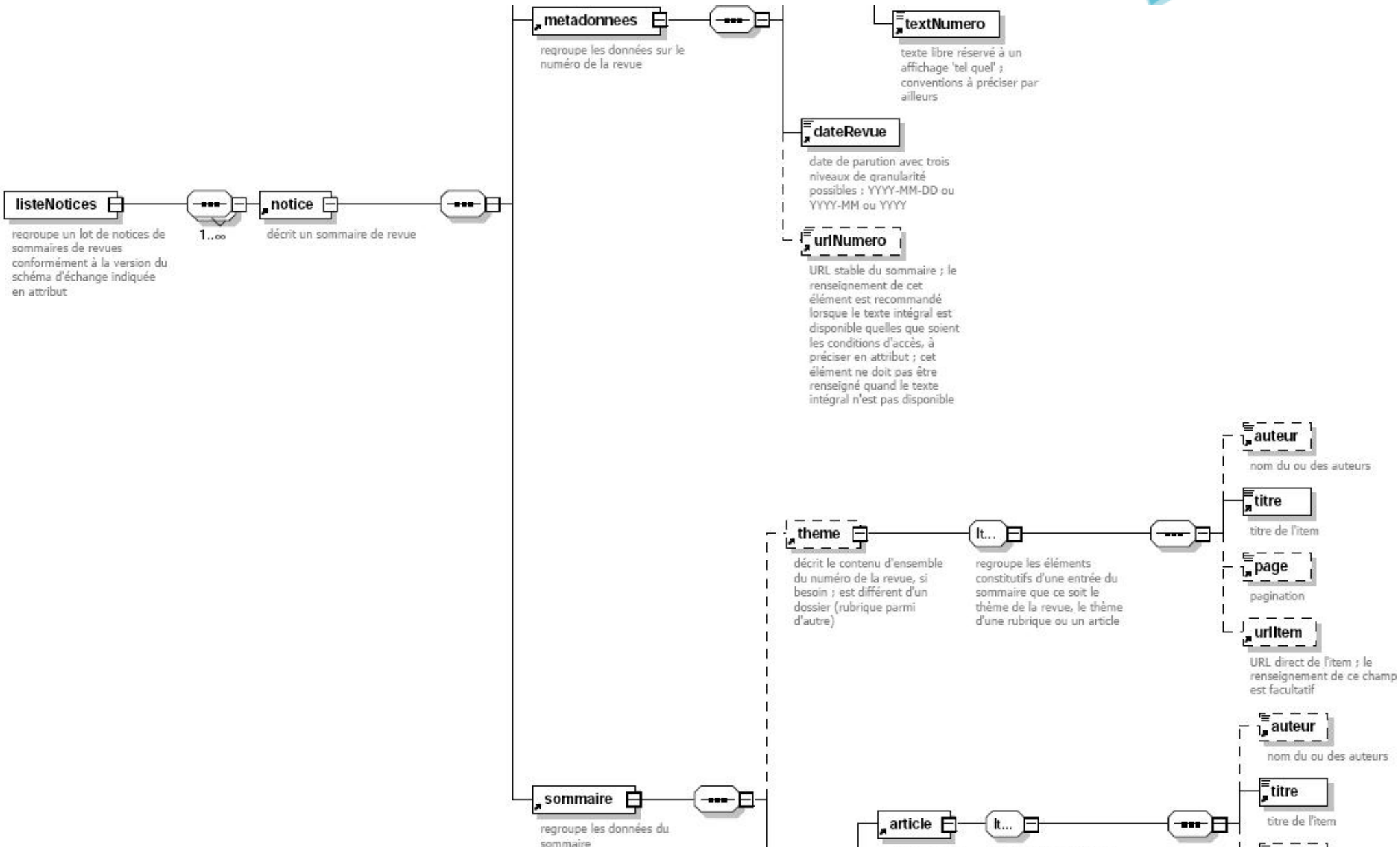
12 octobre 2009
9082 sommaires
dont 43% en ligne



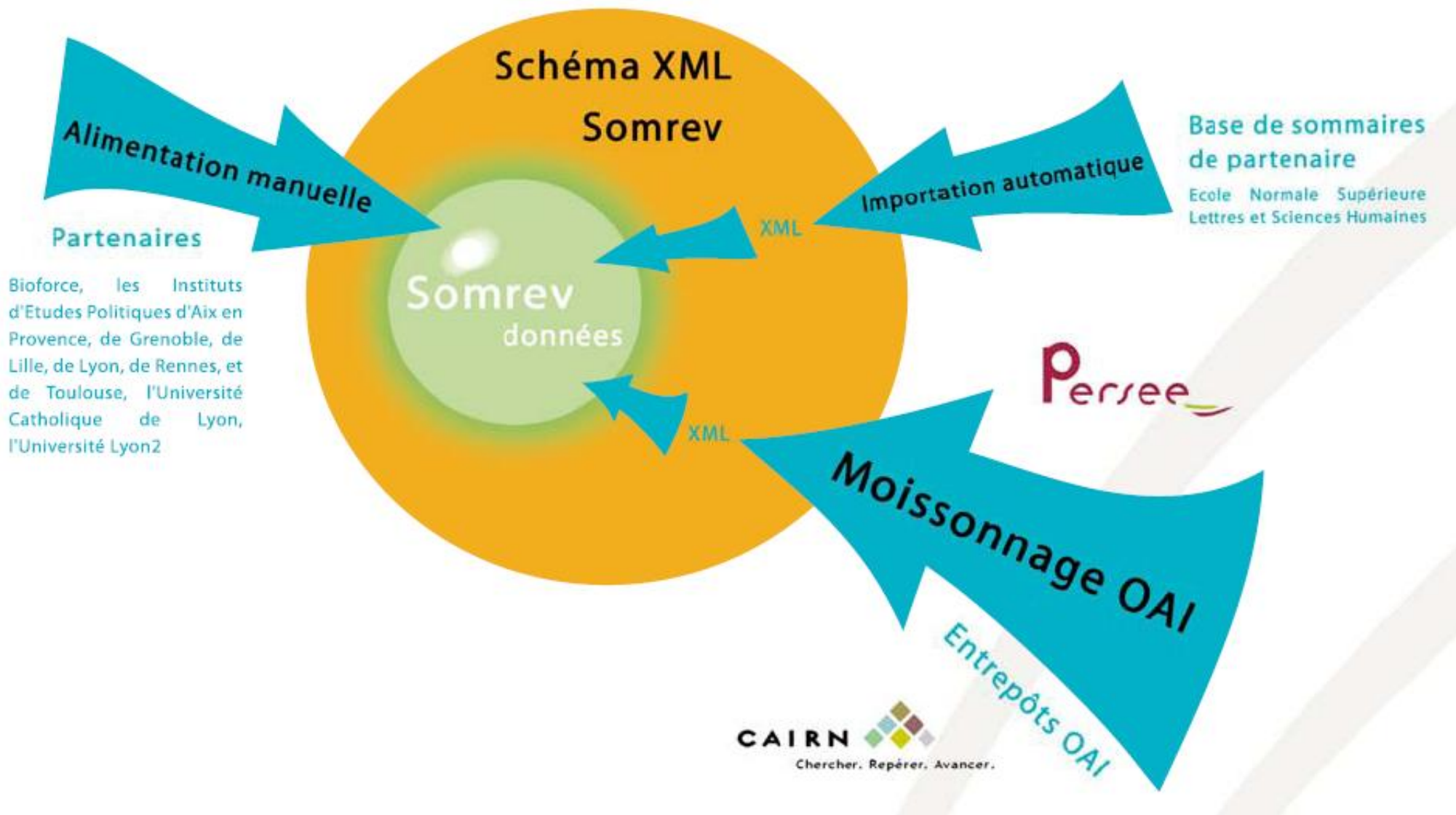
Collecte automatique de sommaires via les entrepôts OAI
Frédoc 2009



Somrev : un schéma XML de données



Somrev : une alimentation multiple



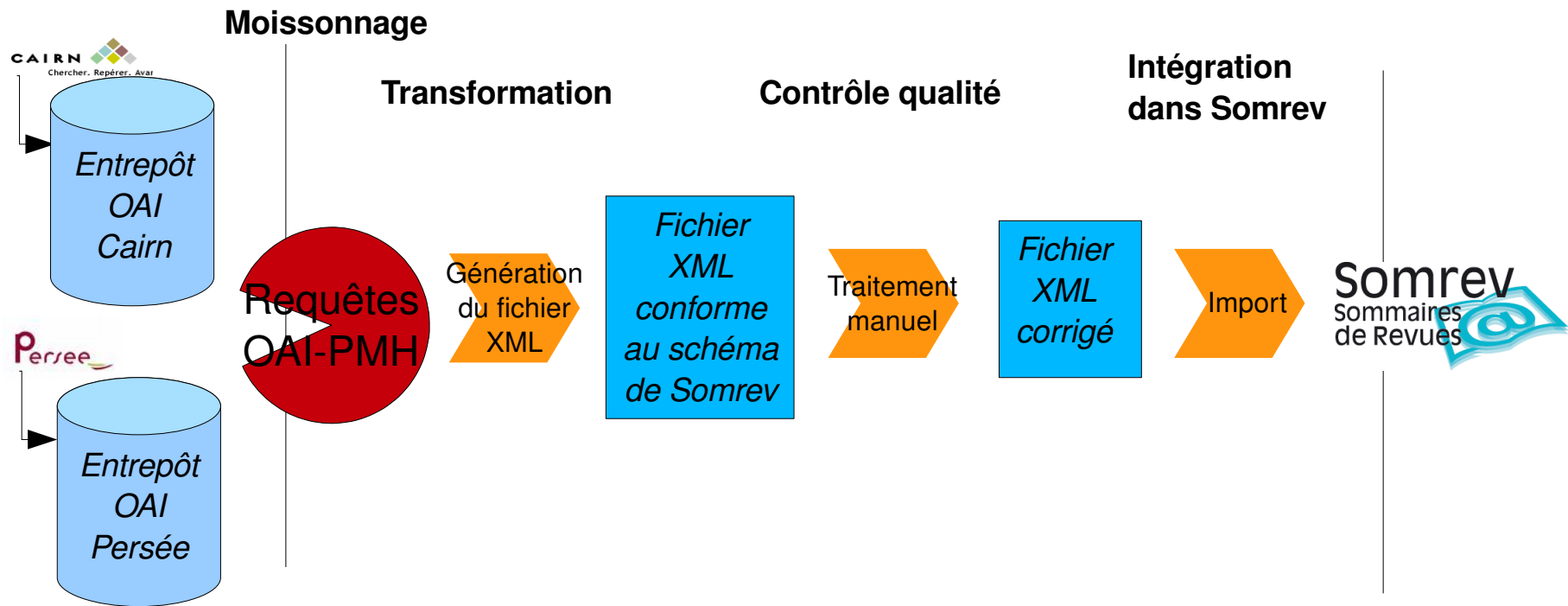
Moissonnage : chaîne de traitement



Fournisseur de données

Traitements des données

Fournisseur de services



Systeme d'import automatique



- Mis en place à partir du schéma XML
- Script php d'intégration des données dans la base MySQL Somrev
- A partir de 2007 pour Persée
[2009 : Nouvelle possibilité proposée : utiliser l'OAI light harvester de Persée]
- A partir de 2008 pour l'ENS-LSH pour des échanges (dans les deux sens).



Collecte automatique de sommaires via les entrepôts OAI
Frédoc 2009



Somrev et les portails de revues



2073 sommaires importés
depuis Persée



439 sommaires importés depuis
Cairn



Collecte automatique de sommaires via les entrepôts OAI
Frédoc 2009



Un moissonneur...



- Interface de moissonnage de sommaires développée par un stagiaire (juin 2008)
- Mis en place en liaison avec les portails de revues
- Utilisation de la bibliothèque SimpleXML de php pour la manipulation des arbres XML, avec SHTML, CSS et Javascript pour le script de récupération et d'intégration des sommaires.

...des entrepôts OAI Exemple de Cairn



- Travail préalable avec le portail (ex. informations sur les pages, sur l'accès libre ou restreint,...)
- Utilisation des métadonnées DC (Dublin Core) et Mets (Metadata Encoding and Transmission Standard)
- Récupération des informations articles, (dont résumés) et de la structure des documents.



Collecte automatique de sommaires via les entrepôts OAI
Frédoc 2009



<http://oai.cairn.info/>



- La liste des revues disponibles sur Cairn...

<http://oai.cairn.info/oai.php?verb=ListSets>

- La liste des numéros d'une revue

<http://oai.cairn.info/oai.php?verb=ListIdentifiers&metadataPrefix=mets&set=DOCSI>

- La table des matières d'un numéro d'une revue (structure)

http://oai.cairn.info/oai.php?verb=GetRecord&metadataPrefix=mets&identifiant=oai:cairn.info:DOCSI_463

- Les métadonnées des articles (via le Dublin Core), avec les résumés notamment

http://oai.cairn.info/oai.php?verb=GetRecord&metadataPrefix=oai_dc&identifiant=oai:cairn.info:DOCSI_463_0001



Chercher. Repérer. Avancer.

Archive OAI-PMH

Ce site distribue les données bibliographiques des articles de revues diffusés en texte intégral sur www.cairn.info conformément au protocole **OAI-PMH** version 2.0.

Le moissonnage de ces données bibliographiques est libre et ne requiert aucune formalité préalable.

Chaîne de traitement dans notre interface : copies d'écran



Les prochaines pages montrent, une à une, les différentes étapes de notre chaîne de traitement (entièrement accessible depuis un navigateur web) pour le moissonnage du dernier numéro de « Documentaliste-Sciences de l'information » via l'entrepôt OAI de Cairn



Collecte automatique de sommaires via les entrepôts OAI
Frédoc 2009



Le moissonnage OAI

Etape 1. Sélection



- Une page web liste les revues sélectionnées sur l'entrepôt OAI

[lecture de la liste des ensembles de l'OAI de Cairn – en vert les numéros déjà dans Somrev]

Choix des sommaires à récupérer :

Selection actuelle : IEPLYON. Les revues en **rouge** sont les revues n'ayant pas de condition renseignée pour le calcul Vol/Num. Les numéros en **vert** sont déjà dans somrev.

Toutes les revues | IEPLYON

<input type="checkbox"/> L C D	A contrario
<input type="checkbox"/> L C D	Afrique contemporaine
<input type="checkbox"/> L C D	L'Année sociologique
<input type="checkbox"/> L C D	Actes de la recherche en sciences sociales
<input type="checkbox"/> L C D	Cahiers internationaux de sociologie
<input type="checkbox"/> L C D	Le Courrier des pays de l'Est
<input type="checkbox"/> L C D	Critique internationale
<input type="checkbox"/> L C D	Documentaliste-Sciences de l'Information
<input type="checkbox"/> [SOM.]	Volume 46, 3 - La relation clients-professionnels de l'info-doc
<input type="checkbox"/> [SOM.]	Volume 46, 2 - Records management
<input type="checkbox"/> [SOM.]	Volume 46, 1 - Web 2.0 et information-documentation
<input type="checkbox"/> [SOM.]	Volume 45, 4 - Pleins feux sur la veille

Documentaliste-Sciences de l'Information, Volume 46, 3 (A.D.B.S., fre)

! \ - Accès restreint ! Vous devez être identifié ou être sur l'un des postes de votre établissement (qui doit être abonné) pour accéder au texte intégral de l'article.

Titre : La relation clients-professionnels de l'info-doc

Quel avenir pour l'intermédiation documentaire ? (GAYON Élisabeth) (p. 1-1) (Gratuit)

André Chonez. 1927-2009 (RAUZIER Jean-Michel) (p. 2-2) (Gratuit)

Réingénierie de thésaurus : une étude de cas (CYROT Catherine, PREUSS Christian) (p. 4-13) (Payant)

Méthodes techniques et outils (BATTISTI Michèle, CHAUDIRON Stéphane) (p. 14-19) (Payant)

Métiers et compétences (VALDERRAMA Asuncion, BOULOGNE Arlette) (p. 20-21) (Payant)

Droit de l'information (CHOISY Stéphanie, BATTISTI Michèle, VERCKEN Gilles) (p. 22-26) (Payant)

La relation clients-professionnels de l'info-doc : vers une nouvelle intermédiation ?

La relation clients-professionnels de l'info-doc. Vers une nouvelle intermédiation ? (p. 28-29) (Gratuit)

Le client au c?ur des « stratégies documentaires » (MELOT Agnès) (p. 30-31) (Gratuit)

Évolutions de la relation avec les usagers (BRETELLE-DESMAZIÈRES Danièle, MESGUICH Véronique, MEN Hervé Le, BOUDRY Christophe, WIEGANDT Caroline, BISBROUCK Marie-Françoise, SOGN Françoise) (p. 32-44) (Payant)

Quels usagers ? Quels usages ? (RANJARD Sophie, GAYON Élisabeth, ARMAND Françoise, DUBOIS Sophie, COUSSEAU Françoise, COUSSOT Danièle) (p. 46-57) (Payant)

Quelle relation demain? (POISSENOT Claude, OCTOBRE Sylvie, BRISSET Isabelle, PEYRELONG Marie-France, JACKSON Catherine, BARS Sylvie Le, BOUCHER Thierry, MARTIN France) (p. 58-69) (Payant)

Sources et ressources (MELOT Agnès) (p. 70-71) (Payant)

Notes de lecture

Notes de lecture (p. 72-77) (Gratuit)

XML

```
<?xml version="1.0"?>
<listeNotices schemaVersion="0.2">
  <notice>
    <identification>
      <identifiant>oai:cairn.info:DOCSI_463</identifiant>
      <dateCreation>2009-10-12</dateCreation>
      <dateMaj>2009-02-01</dateMaj>
    </identification>
    <metadonnees>
      <issn>
        </issn>
      <numero>463</numero>
      <textNumero>Volume 46, 3, 2009</textNumero>
      <dateRevue granularity="YYYY">2009</dateRevue>
      <urlNumero acces="extranet">http://www.cairn.info/sommaire.php?ID_REVUE=DOCSI&ID_NUMPUBLIE=DOCSI_463</urlNumero>
    </metadonnees>
    <resume>
      <theme>
        <titre>La relation clients-professionnels de l'info-doc</titre>
      </theme>
      <article>
        <titre>Quel avenir pour l'intermédiation documentaire ?</titre>
```



Etape 2 : Récupération d'un fichier XML



Nom du fichier xml en sortie :

Outil permettant la génération d'un fichier XML, respectant le schéma de Somrev, pour les revues ou numéros sélectionnés.

[Récupération des données moissonnées sur l'entrepôt OAI dans un format XML]



Collecte automatique de sommaires via les entrepôts OAI
Frédoc 2009



Etape 3.

Traitement du fichier XML (1/2)



Edition du fichier : sommairesxml/DOCSI_46_3

```
<?xml version="1.0"?>
<listeNotices schemaVersion="0.2">
<notice>
  <identification>
    <identifiant>oai:cairn.info:DOCSI_463</identifiant>
    <dateCreation>2009-10-12</dateCreation>
    <dateMaj>2009-02-01</dateMaj>
  </identification>
  <metadonnees>
    <issn>
    </issn>
    <numero>463</numero>
    <textNumero>Volume 46, 3, 2009</textNumero>
    <dateRevue granularite="YYYY">2009</dateRevue>
    <urlNumero acces="extranet">http://www.cairn.info/sommaire.php?ID_REVUE=DOCSI&ID_NUMPUBLIE=DOCSI_463</urlNumero>
  </metadonnees>
  <resume>
    <theme>
      <titre>La relation clients-professionnels de l'info-doc</titre>
    </theme>
    <article>
      <titre>Quel avenir pour l'intermédiation documentaire ?</titre>
      <urlItem acces="internet">http://www.cairn.info/article.php?ID_ARTICLE=DOCSI_463_0001</urlItem>
      <auteur>Gayon Elisabeth</auteur>
      <page>1-1</page>
    </article>
    <article>
      <titre>André Chonez. 1927-2009</titre>
      <urlItem acces="internet">http://www.cairn.info/article.php?ID_ARTICLE=DOCSI_463_0002</urlItem>
      <auteur>Rauzier Jean-Michel</auteur>
      <page>2-2</page>
    </article>
  </resume>
</notice>
</listeNotices>
```

Sauvegarder

Réinitialiser

Fermer

Etat :

Fichier Chargé

Etape 3. Traitement du fichier XML (2/2)



- Il s'agit de notre contrôle qualité
- A ce niveau nous augmentons la granularité de la date du numéro qui par défaut est souvent une année
- En fonction des revues et des périodes nous avons souvent des corrections manuelles à apporter sur le contenu (accents, casse,...).

Etape 4. Import dans la base Somrev



IMPORT DE NOTICES DEPUIS UN FICHER XML : ÉTAPE 3/3

Traitement de la notice no.1

Résumé des métadonnées :

Identifiant Koha : 163

Volume :

Numéro : 463

TexteNuméro : Volume 46, 3, 2009

Titre Numéro : La relation clients-professionnels de l'info-doc

Date Revue : 2009-00-00

Date création : 2009-10-12

Date Mäj : 2009-02-01

URL | accès : http://www.cairn.info/sommaire.php?ID_REVUE=DOCSI&ID_NUMPUBLIE=DOCSI_463 | extranet

- [OK] Insertion de l'article : Quel avenir pour l'intermédiation documentaire ?
- [OK] Insertion de l'article : André Chonez. 1927-2009
- [OK] Insertion de l'article : Réingénierie de thésaurus : une étude de cas
- [OK] Insertion de l'article : Méthodes techniques et outils
- [OK] Insertion de l'article : Métiers et compétences
- [OK] Insertion de l'article : Droit de l'information
- [OK] Insertion de la rubrique : La relation clients-professionnels de l'info-doc : vers une nouvelle intermédiation ?
- [OK] Insertion de l'article : La relation clients-professionnels de l'info-doc. Vers une nouvelle intermédiation ?
- [OK] Insertion de l'article : Le client au cœur des « stratégies documentaires »
- [OK] Insertion de l'article : Évolutions de la relation avec les usagers
- [OK] Insertion de l'article : Quels usagers ? Quels usages ?
- [OK] Insertion de l'article : Quelle relation demain?
- [OK] Insertion de l'article : Sources et ressources
- [OK] Insertion de la rubrique : Notes de lecture
- [OK] Insertion de l'article : Notes de lecture

Résultat du traitement :
sommaire disponible en ligne

Somrev
Sommaires
de Revues



Somrev
Sommaires
de Revues

BASES DE DONNEES

navigation sur le site Webdoc ▼ valider

Somrev : base de signalement de sommaires de périodiques

sommaire de la revue

Revue Documentaliste - Sciences de l'information
numéro Volume 46, 3, 2009
titre La relation clients-professionnels de l'info-doc
[accès réservé](#)

- Sommaire**
- Quel avenir pour l'intermédiation documentaire ? - Gayon Élisabeth p.1-1 [en ligne](#)
 - André Chonez. 1927-2009 - Rauzier Jean-Michel p.2-2 [en ligne](#)
 - Réingénierie de thésaurus : une étude de cas - Cyrot Catherine, Preuss Christian p.4-13 [accès réservé](#)
 - Méthodes techniques et outils - Battisti Michèle, Chaudiron Stéphane p.14-19 [accès réservé](#)
 - Métiers et compétences - Valderrama Asuncion, Boulogne Arlette p.20-21 [accès réservé](#)
 - Droit de l'information - Choisy Stéphanie, Battisti Michèle, Vercken Gilles p.22-26 [accès réservé](#)

La relation clients-professionnels de l'info-doc : vers une nouvelle intermédiation ?

Modifications possibles dans l'interface de production 'manuelle'



formulaire de saisie - nouvelle base Somrev

Documentaliste - Sciences de l'information (depuis 2001)

(Volume/Tome 46) (No 3) Volume 46, 3, 2009

chrono : 2009-07-00

Titre : La relation clients-professionnels de l'info-doc

http://www.cairn.info/sommaire.php?ID_REVUE=DOCSI&ID_NUMPUBLIE=DOCSI_463 accès réservé

[modifier le numéro](#)

• Quel avenir pour l'intermédiation documentaire ? Gayon Élisabeth	1-1 Web	modifier - ajou
• André Chonez. 1927-2009 Rauzier Jean-Michel	2-2 Web	modifier - ajou
• Réingénierie de thésaurus : une étude de cas Cyrot Catherine, Preuss Christian	4-13 Web réservé	modifier - ajou
• Méthodes techniques et outils Battisti Michèle, Chaudiron Stéphane	14-19 Web réservé	modifier - ajou
• Métiers et compétences Valderrama Asuncion, Boulogne Arlette	20-21 Web réservé	modifier - ajou
• Droit de l'information Choisy Stéphanie, Battisti Michèle, Vercken Gilles	22-26 Web réservé	modifier - ajou

La relation clients-professionnels de l'info-doc : vers une nouvelle intermédiation ?

• La relation clients-professionnels de l'info-doc. Vers une nouvelle intermédiation ?	28-29 Web	modifier - ajouter - ajouter sous ce thème
----------------------------------------------------------------------------------------	---------------------------	--------------------------------------------------------------------------------------------

Retour d'expérience : Un bilan positif... à suivre!



- Récupération de plus de 2500 sommaires avec lien vers le texte intégral
- Semi-automatique : vérification, corrections, validation des données...
- Echanges avec les portails de revues
- En cours : mise en place pour Revues.org...