

# Open Archive Initiative

Sara Aubry - sara.aubry@bnf.fr  
Laurent Duploux - laurent.duploux@bnf.fr

*Département de la bibliothèque numérique  
Bibliothèque nationale de France*

*Groupe PIN - 15 janvier 2004*

1

## PLAN

### ✓ Origines de l'OAI

- ✓ Protocole OAI-PMH
  - modèle de données
  - acteurs
  - fonctionnement
  
- ✓ Quelques exemples d'applications

2

## **Oct. 1999 : rencontre à Santa Fé**

- ✓ Autour du web et du développement des systèmes de publication électronique
  - la recherche devient pluridisciplinaire
  - les réseaux haut débit et l'informatique personnelle se développent
  - le web devient un outil d'échange et de diffusion
  - les bibliothèques et les musées cherchent à rendre visibles leurs collections

3

## **Oct. 1999 : rencontre à Santa Fé**

- les archives ouvertes ont du succès ...
  - > arXiv : <http://arxiv.org>
  - > CogPrints : <http://cogprints.ecs.soton.ac.uk>
  - > RePEc (Research Papers in Economics) : <http://repec.org>
  - > NDLTD (Networked Digital Library of Theses and Dissertations) : <http://www.ndltd.org>
  - > NCSTRL (Networked Computer Science Technical Reference Library) : <http://www.ncstrl.org>
  - > ... mais ils sont indépendants : il faut les rendre interopérables

4

## ***Fév. 2000 : convention de Santa Fé***

- ✓ « Un cadre général pour la fédération de contenus sur le web »
  
- ✓ Deux alternatives
  - interroger simultanément des bases hétérogènes et réparties (Z39.50)
  - collecter massivement les métadonnées dans un ou plusieurs services centraux
    - › le premier prototype (Universal Preprint Service - UPS) utilise l'archive NCSTRL, le protocole Dienst et les métadonnées Open Archive Metadata Set (OAMS)

5

## ***L'Open Archive Initiative s'organise***

- ✓ Août/septembre 2000
  - l'OAI se formalise
    - › création d'un comité scientifique et d'un comité technique
    - › coordination assurée par Herbert Van de Sompel (Los Alamos) et Carl Lagoze (Cornell)
  
  - l'OAI est subventionnée
    - › Digital Library Federation / Coalition for Networked Information / National Science Foundation

6

## ***Sa mission***

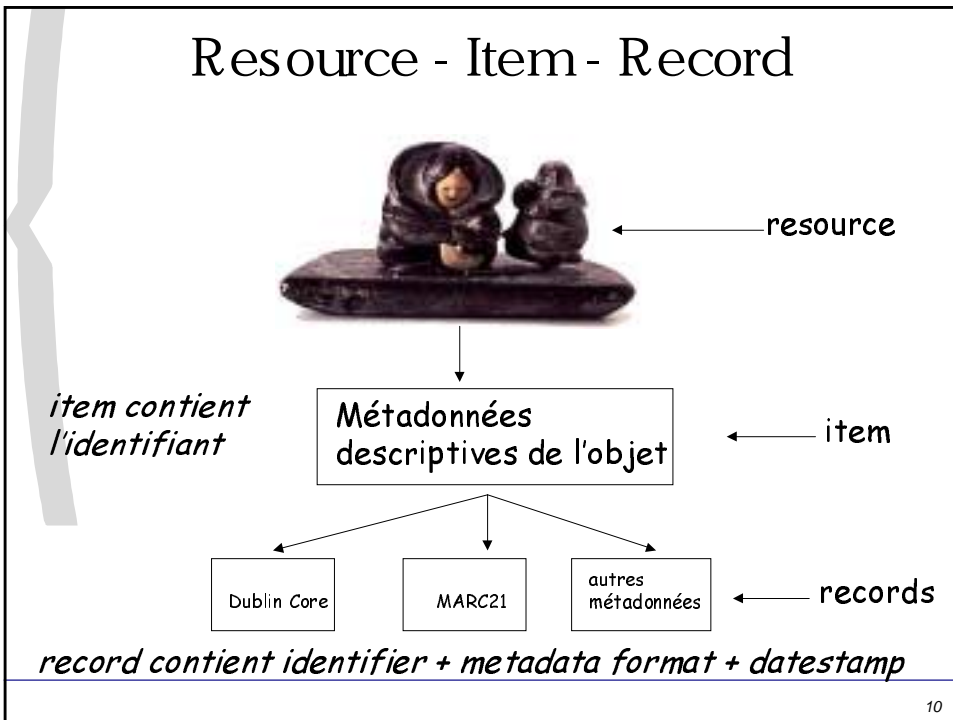
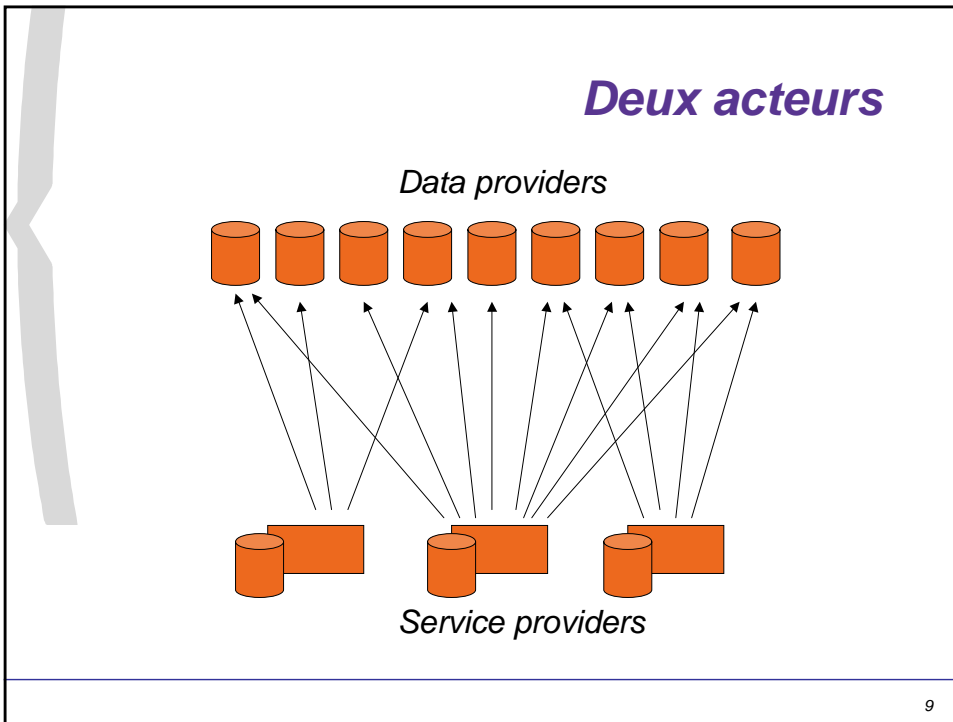
- ✓ Passe de
  - « l'OAI est un forum qui permet de soulever et de résoudre les problèmes que pose l'interopérabilité des archives d'e-prints »
  
- ✓ à
  - « l'OAI a pour but de faciliter les échanges, l'interopérabilité et l'extensibilité des archives contenant tout type de contenu numérique »

7

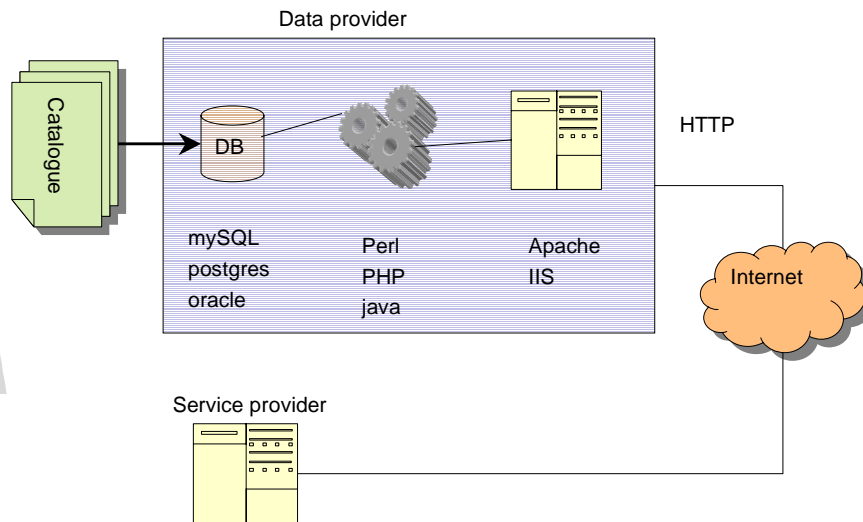
## ***Le protocole de l'OAI***

- ✓ L'OAI met au point un protocole de collecte des métadonnées
  - Open Archive Initiative Protocol for Metadata Harvesting (OAI-PMH) simple basé sur des standards (HTTP, XML)
- ✓ Déjà plusieurs versions
  - v1 en janvier 2001 révisée en juin 2001
  - v2 en juin 2002
- ✓ Objectif
  - découvrir, présenter et analyser le contenu d'une archive

8



## Mise en œuvre d'un data provider



11

## OAI-PMH : Les requêtes informatives

- ✓ Identify information sur le repository
  - repositoryName
  - baseURL
  - protocolVersion
  - earliestDatestamp
  - deletedRecord
  - granularity
  - adminEmail
  - compression
  - description

12

## OAI-PMH : Les requêtes informatives

- ✓ ListMetadataFormats formats supportés par le repository
  - metadataPrefix
  - schema
  - metadataNamespace
- ✓ ListSets
  - setSpec
  - setName
- ✓ GetRecord
  - header (unique identifier, timestamp, setSpec)
  - metadata

13

## OAI-PMH : Les listes

- ✓ ListIdentifiers
    - header
  - ✓ ListRecords
    - header
    - metadata
- ont pour arguments**
- from , until
  - metadataPrefix
  - set
  - resumptionToken

14

## *Différents types de data providers*

- ✓ Archives disciplinaires, banques de données
  - physique
    - › arXiv : <http://arxiv.org>
    - › PhysNet : <http://physnet.uni-oldenburg.de/PhysNet/>
  - chimie
    - › ChemWeb : <http://www.chemweb.com>
  - informatique
    - › NCSTRL : <http://www.ncstrl.org>
  - neurosciences
    - › CogPrints : <http://cogprints.ecs.soton.ac.uk>
  - sciences de la vie
    - › PubMed Central : <http://www.pubmedcentral.nih.gov>

15

## *Différents types de data providers*

- ✓ Editeurs commerciaux
  - American Physical Society : <http://www.aps.org>
  - BioMed Central : <http://www.biomedcentral.com>
- ✓ Archives institutionnelles et universitaires
  - CERN Document Server (CDS) : <http://cds.cern.ch>
  - DSpace (MIT) : <http://dspace.mit.edu>
  - SHERPA (UK) : <http://www.sherpa.ac.uk>
- ✓ Bibliothèques numériques
  - American Memory (LC) : <http://memory.loc.gov>
  - Université du Michigan : <http://www.umdl.umich.edu>

16



## *Différents types de data providers*

- ✓ En France
  - Revue Productions Animales de l'INRA, thèses de Lyon II, publications de l'ENS-LSH, ...
  - Centre pour la Communication Scientifique Directe : <http://ccsd.cnrs.fr>
  - Bientôt Gallica !
  
- ✓ Liste des DP enregistrés (134 en janvier 2004)
  - <http://www.openarchives.org/Register/BrowseSites.pl>

17

## *Service providers les plus connus*

- ✓ Recherche bibliographique
  - ARC : <http://arc.cs.odu.edu>
  - OAIster : <http://oaister.umd.umich.edu/o/oaister/>
  - my.OAI (personnalisation) : <http://www.myoai.com>
  
  - RLG Cultural Materials : <http://www.rlg.org/culturalres/>
  - Scirus (Elsevier) : <http://www.scirus.com>
  
  - UNICAT : <http://www.kbr.be/unicat/>
  
- ✓ Références et citations
  - Citebase : <http://citebase.eprints.org>

18

## ***Services providers particuliers***

- ✓ Intermédiaire avec les moteurs de recherche
  - DP9 : <http://arc.cs.odu.edu:8080/dp9/index.jsp>
- ✓ Agrégateur (cache/proxy)
  - Celestial : <http://celestial.eprints.org>
- ✓ En France
  - BNSA : <http://bnsa.aquitaine.fr>
- ✓ Liste des SP enregistrés (16 en janvier 2004)
  - <http://www.openarchives.org/service/listproviders.html>

19

## ***Des outils compatibles OAI-PMH***

- ✓ Permettent de gérer des archives en ligne
  - e-prints (Southampton) : <http://www.eprints.org>
  - DSpace (MIT) : <http://dspace.org>
  - CDSware (CERN) : <http://cdsware.cern.ch>

20

## Conclusion

- ✓ OAI est devenu incontournable pour
  - la mise en place d'archives ouvertes, de catalogues collectifs et de sites portails
  - l'échange de métadonnées
  
- ✓ Travaux en cours
  - gestion des droits
  - analyse des usages
  - développement des services

21

## Sites utiles

- ✓ Open Archive Initiative
  - site officiel : <http://www.openarchives.org>
  - listes de discussion OAI-general et OAI-implementers
  
- ✓ Au niveau européen
  - Open Archive Forum : <http://www.oaforum.org>
  - Open Archives Activities and Experiences in Europe : <http://www.dlib.org/dlib/january03/dobratz/>
  
- ✓ Open Access
  - pages de H. Bosc, S. Jérôme et J.P. Schmitt : <http://www.tours.inra.fr/tours/doc/comsci.htm>
  - liste de discussion September 1998 American Scientist Forum

22

## Démonstration

- ✓ Repository Explorer
  - <http://oai.dlib.vt.edu/cgi-bin/Explorer/oai2.0/testoai>
  - <http://celestial.eprints.org/cgi-bin/status>
- ✓ Exemple de DIVA (archive universitaire suédoise / 2000 records)
  - <http://publications.uu.se/portal/OAI>
- ✓ Identify [ex](#)
- ✓ ListMetadataFormats [ex](#)
- ✓ ListSets [ex](#)
- ✓ ListIdentifiers [janvier 2003](#)
- ✓ ListRecords [janvier 2003](#)
- ✓ GetRecord (oai:DiVA.se:umu-3) [oai\\_dc](#) / [marc21](#)