

La préservation de l'accès aux revues numériques

Fabrice Lecocq (lecocq@inist.fr)

Institut de l'Information Scientifique et
Technique
CNRS - UPS 076

Présentation Groupe PIN
25 janvier 2006

02/02/2006

1

Sommaire

- INIST / CNRS
- Contexte
- Présentation de D-Space
- Projets connexes
- Evolutions attendues
- Conclusion

02/02/2006

2

INIST / CNRS

02/02/2006

3

INIST - Principales activités

- Fourniture de Documents Primaires (FDP)
- Fabrication des bases de données Pascal et Francis
- Services en ligne
 - Article@INIST (lancé en 1997 ; 9,5 M)
 - Bases en ligne : BiblioSciences / TitaneSciences
- Portails
 - Généraliste : ConnectSciences
 - Thématiques : BiblioVie, BiblioSHS, STIC, Inserm...
- Veille, Infométrie, Multimedia, Edition de Revues, Formation, Expertises...

02/02/2006

4

INIST - Evolutions

- Passage de fournisseur de services INIST (FDP, Base de données) à intégrateur de contenus et services divers et variés, locaux ou distants
 - Portail d'intégration de services
 - Authentification, Electronic Right Management (ERM), proxy
 - Recherche fédérée, résolution de liens (linking), connecteurs
 - Méta-données, terminologie
 - TermSciences.inist.fr (novembre 2005)

02/02/2006

5

Positionnement institutionnel

- Unité Propre de Services du CNRS
- Appartenant à la Direction de l'Information Scientifique du CNRS (juin 2005)
- Forte relation avec les autres EPST (INSERM, INRA...)
- Coordination avec l'Université
- Relation avec l'ensemble de la Recherche publique ou privée, en France essentiellement, mais aussi au niveau Européen et mondial

02/02/2006

6

INIST - Organisation Interne

- Départements métiers
 - Edition Numérique et Fourniture de Documents (63 personnes)
 - Ingénierie de l'Information Scientifique (81)
 - Portails et services d'Information (61)
 - Systèmes d'Information (46)
- Direction générale, Affaires générales
- INIST-Diffusion : SA, filiale du CNRS

02/02/2006

7

Contexte



02/02/2006

8

Les nouvelles technologies

- Internet et la diffusion : accès en ligne et librairies électroniques (catalogues de bibliothèque, accès au texte intégral...)
 - Continuité de service, contrôle des accès...
 - Interconnexion de catalogue (Z39-50), prêt entre bibliothèques (ILL)
 - Des utilisateurs exigeants (environnement personnalisé, login unique ou Sigle-Sign-On, recherche fédérée...)

02/02/2006

9

La diffusion des savoirs

- Le mouvement de l'Open Access et les Archives Ouvertes (Open Archive) :
 - Mi-janvier : plus de 2000 revues soit 84 000 articles
Source : DOAJ (Directory of Open Access Journals ; www.doaj.org)
 - CNRS - Signataire de la Déclaration de Berlin sur le libre accès à la connaissance en Sciences Exactes, S de la Vie et Sciences humaines et sociales (octobre 2003)
 - Centre pour la Comm. Scientifique Directe (2000)
- Archives Institutionnelles : CNRS, INSERM, INRA, IFREMER...
 - Bases de textes intégraux, publications, thèses, rapports internes... accessibles à tous

02/02/2006

10

Les contenus documentaires

- Regroupement des institutions (Groupements de Commandes) pour négocier auprès des Editeurs
 - Respect des ayants-droits
 - Respect des éditeurs

- Statistiques, contrôle d'accès, gestion des droits, confidentialité

- Le passage des abonnements papier vers l'électronique doit s'accompagner d'une mise en place d'un système d'archivage des anciens numéros

02/02/2006

11

Exemple : Groupement de commandes « Elsevier »

- Abonnement à la Freedom Collection (1800 titres) et Cell Press (10 titres) pour la période 1995-2007
- Ayant droit : Couperin (Universités), CNRS, INSERM, INRA

- Archive, offre Editeur : Koninklijke Bibliotheek (payante)
- Deux centres d'archivage en France : ABES/CINES et INIST
- Fournitures par l'Editeur :
 - Contenus et métadonnées en XML
 - Etats d'abonnement de chaque institution
- Garanties par les centres d'archivage :
 - Sur les accès aux Revues
 - Sur les usages

02/02/2006

12

Archivage « on site »

- Les plates-formes commerciales
 - SDOS (Science Direct On Site) de la société Elsevier
 - EJOS (Electronic Journal On Site) de la société Endeavor
 - BePress (Berkeley Electronic Press) (13 sur 569)

- Les plates-formes Open Source
 - Première génération : E-prints (191), GreenStone, ETD-db (22), Open Publications System (16), BePress (13), Fedora (1)
 - Deuxième génération : D-Space (99), CDSWare (8)
 - Francophones : Archimede, Castore, HAL (5)
 - Atypique : LOCKSS

02/02/2006

Source : Institutional Archives Registry (<http://archives.eprints.org/eprints.php>)

13

E-Prints (www.eprints.org)

- Université de Southampton (2000) : Stevan Harnad / Robert Tansley
- Actuellement version 2.3.13
- Le plus utilisé : 197 sites sur 569
- Objectif : Open Access, Archives Institutionnelles auto-archivage, interface de recherche, RSS (flux de syndication), OAI-PMH (metadata harvesting)

- Métadonnées descriptives
- Peu orienté sur l'organisation des documents
- Peu orienté préservation

02/02/2006

14

Pour mémoire

- Fedora (Flexible Extensible Digital Object Repository) (www.fedora.info)
University of Virginia / Cornell University
- Greenstone (www.greenstone.org)
New Zealand Digital Library Project (avec l'UNESCO)
- CDSware (CERN Document Server Software) (cdsware.cern.ch)

02/02/2006

15

Monde francophone

- HAL : Hyper Article en Ligne (hal.ccsd.cnrs.fr)
CNRS / CCSD
- Archimede (www.bibl.ulaval.ca/archimede/)
Laval University Library
- Castore (www.castore.org) :
Ecole des Mines de Nantes / INRIA Rennes / Xerox

02/02/2006

16

Présentation de D-Space



02/02/2006

17

D-Space (www.dspace.org)

- Massachussets Institute of Technology/ Hewlett-Packard Labs. (R. Tansley / Smith Mac Kenzie...)
- Le 2^{ème} plus utilisé : 115 sites sur 569
- Objectif : «The DSpace digital repository system captures, stores, indexes, preserves, and distributes digital research material. »

02/02/2006

18

D-Space - Historique

- 2000-2002 : Collaboration HP / bibliothèque du MIT pour les besoins du MIT
 - Auto-archivage (création, indexation, recherche, affichage)
 - En peu d'étapes (workflow en 3 étapes)
 - Interface facile à utiliser
 - Publication dans un contexte Open Access (OAI)
 - Préservation des contenus numériques (ms-word, pdf, jpeg, gif, powerpoint)
- Nov 2002 : v1.0 mise en Open Source
- Mai 2003 : v1.1 (développement MIT / HP)

02/02/2006

19

D-Space - Historique (2)

- Mars 2004 : création du D-Space User Group & appel à contributions
- Août 2004 : 1.2 (dernier dévt MIT / HP)
- Février 2005 : 1.2.1 (premières contributions externes)
- Avril 2005 : annonce de la 1.3 alpha
- Octobre 2005 : 1.3.2

- 32 000 téléchargement en novembre 2005

02/02/2006

20

Objectifs

- Système d'archivage d'objets numériques
 - gestion, distribution et préservation à long terme d'objets numériques.
 - Organisation en collection
- Base de données des métadonnées associées
 - création, indexation, recherche...
- Plate-forme de dépôt et de diffusion de publications électroniques
 - Workflow de publication
 - OAI v2 (metadata), OpenURL (nommage)
- Droits, Délégation, Alertes, Espace personnalisé

02/02/2006

21

Historique en France

- « Aladin » : Accès Libre aux Archives du Dépôt Institutionnel Numérique de la MSH-Alpes
<https://dspace.msh-alpes.prd.fr/>
- Réunion à la MSH-Alpes (Grenoble) en nov 2004, création du groupe D-Space France : MSH-ALPES, Lille1, Lyon1, INIST, INSERM, CINES, Cemagref, INRA, IMAG
- INIST : Service « I-Revues » (couplage avec SPIP) (mai 2005)
- 2^{ème} Réunion du groupe D-Space France à l'INIST (Nancy) en juin 2005 ; Canada, Luxembourg
- GL7 (Grey Litterature) : Annonce de « LARA » par l'INIST
- INRA : « Dspace Avignon » (décembre 2005)

02/02/2006

22

Usages - quelques exemples

- Archives Institutionnelles (MSH, INRA...)
- Thèse en ligne
- Plate-forme de E-learning
- Bases d'images
 - Australian National University (dspace.anu.edu.au)
 - Tarlton Law Library, Austin, Texas

02/02/2006

23

Usages - INIST

- Edition / Diffusion de Revues (« Editeurs »)
 - Service I-Revues de l'INIST (revues + congrès)
- Littérature grise
 - LARA, Libre Accès aux Rapports de recherche (INIST, en cours de mise en production)
 - Open-Sigle (INIST, à l'étude) (base bibliographique)
- Intranet, communication interne
 - Mémoire de l'INIST
 - Contrats et marchés

02/02/2006

24

Organisation des objets

D-Space

- Communauté
- Sous-communauté
- Collection
- Item (méta-données)
- Bundle
- Bistream

Exemple I-Revue (INIST)

- Revue
- Volume
- N°
- Article
- Fichiers

02/02/2006

25

Organisation - Détail

- Niveau Communauté
 - Nombre de communautés illimité
 - Chaque communauté est constituée de 0, 1 ou plusieurs sous communautés et/ou de 1 ou plusieurs collections
- Niveau Sous-communauté (facultatif)
 - Nombre de sous-communautés illimité
 - Chaque sous-communauté est constituée de 1 ou plusieurs collections
- Niveau Collection
 - Nombre de collections illimité
 - Chaque collection contient un nombre illimité d'item

02/02/2006

26

Attributs

- Chaque communauté, sous-communauté et collection possède sa propre page d'accueil
 - Logo, Copyright...

- Pour les collections
 - Chaque collection peut définir son propre workflow de dépôt
 - Chaque collection peut posséder son ou ses administrateurs

02/02/2006

27

Les objets numériques

- Items : niveau auquel sont associées des métadonnées descriptives ; peut appartenir à plusieurs collections

- Bundle

- Bistream : niveau le plus fin avec des formats divers que l'on qualifie de
 - Supported (soutenu)
 - Known (connu)
 - Unsupported (non soutenu)

02/02/2006

28

Les acteurs

- « E-peoples »
 - Utilisé pour la soumission, le workflow, l'affichage (cas revues non Open Access), l'administration, l'abonnement aux collections
- Groupes
 - Ensemble de e-peoples
 - Utilisé pour la définition des autorisations : read, add, remove, write, administration des collections, workflow

02/02/2006

29

Workflow de dépôt

- Gestion du workflow
 - Processus de dépôt est défini pour chaque collection
 - Les tâches sont assignées par des alertes mail
 - La tâche se fait via une interface web
- Les étapes d'édition & les actions
 - 1 - Acceptation ou rejet de la soumission
 - 2 - Édition des métadonnées - Acceptation ou rejet de la soumission
 - 3 - Édition des métadonnées
 - Transmission pour stockage et mise en ligne

02/02/2006

30

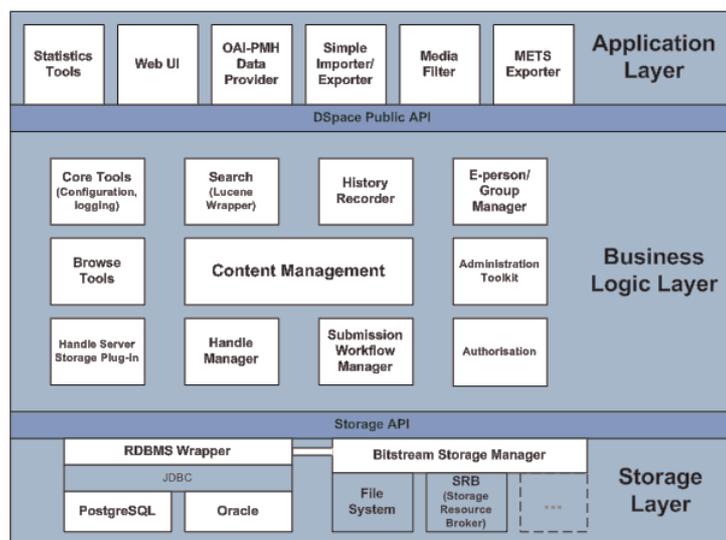
Fonctions d'administration

- Les 3 actions possibles
 - Gestion des déposants
 - Édition, suppression, ajout de soumissionnaires
 - Gestion des items
 - Éditer les métadonnées d'un item
 - Retirer un item de la collection
 - Lien avec d'autres collections
 - Inclure des items d'une autre collection

02/02/2006

31

Architecture



02/02/2006

32

Installation

- Distribution Open Source
- Technologies : Java, Tomcat, XML, PostgreSQL (SGBD), Lucene Jakarta (moteur de recherche)...
- Tourne sous Linux, Solaris, HP-UX, Windows

- SGBD PostgreSQL pour
 - L'organisation des communautés, des collections
 - Les métadonnées
 - La liste des e-people et des droits*Implique maintenance et sauvegardes*
- Stockage dans un système de fichiers

02/02/2006

33

L'interopérabilité

- Descriptif des métadonnées en Dublin Core Qualifié
- Propagation des contenus par OAI-PMH (Open Archive Initiative Protocol for Metadata Harvesting version 2.0)
- Supporte Unicode
- Module d'Import/Export propriétaire

- En cours : Implementation OAIS
DSpace METS Document Profile for Submission Information Packages
<http://cwspace.mit.edu/docs/xsd/METS/SIP/profilev0p9p1/>

02/02/2006

34

Pérennité du nommage

- Développement CNRI (Corporation for National Research Initiative) (www.handle.net)
 - Architecture à 2 niveaux
 - Global Handle Registry (top level)
 - Local Handle Registry (lower level)
- <http://hdl.handle.net/2042/1> devient
<http://documents.irevues.inist.fr/handle/2042/1>
- Actuellement, un préfixe par instance D-Space
 - Délégation d'administration de type « Start of Authority DNS » : pas prévue pour l'instant

02/02/2006

35

D-Space et le stockage

- Depuis la 1.3, couplage avec SRB (Storage Resource Broker)
 - Virtualisation du stockage, réplication : « Any Data, Anywhere, Any Time ; Management, Collaboration, Publication and Preservation for Distributed Data Collections »
 - Projet du San Diego Supercomputer Center et de l'Université de San Diego (www.sdsc.edu/DICE/SRB/)
- Utilisation du modèle LOCKSS (Lots of Copies Keeps Stuff Safe - lockss.stanford.edu) à l'étude pour D-Space 2
- Réflexion métier sur la représentation des données (métadonnées spécifiques, schéma XML...)

02/02/2006

36

Projets connexes



02/02/2006

37

D-Space et extensions (1)

- Utilisation de D-Space pour les thèses
- Workflow avec des étapes supplémentaires
 - Champs métadonnées supplémentaires
 - Visibilité restreinte temporaire

 - TAPIR (Theses Alive Plug-in for Institutional Repositories) de la Edinburgh University Library (sourceforge.net/projects/tapir-eul/)

 - CWSpace (Computer Sciences – Université de Louvain)

02/02/2006

38

D-Space et extensions (2)

- Plate-forme de diffusion de Revues
 - Moving Wall, droits d'accès
 - Problème des vues (par thématique, par revues...)
 - Habillage, kiosque
- Plate-forme de mise en forme des Revues
 - Utilisation du logiciel LODEL en amont de D-Space
- Plate-forme d'édition de Revues
 - Workflow plus riche (reviewing du comité de lecture)
 - DPubs (dpubs.org) (Cornell University Library), projet en cours mais prend du retard

02/02/2006

39

Prototype en cours

- SRW/U (Search/Retrieve WebServices)
 - Ex Z39-50 Next Generation (2001) (Zing)
 - Version stable (1.1) en février 2004
 - Voir l'article « SRW/U with OAI »
D-lib magazine Février 2005 (www.dlib.org)
- OLAC (Open Language Archive Community)
 - Voir l'article « A Metadata Search Engine for Digital Language Archives » D-lib magazine Février 2005

02/02/2006

40

Evolutions attendues



02/02/2006

41

D-Space 2

- Internationalisation
- Architecture modulaire pour faciliter :
 - la personnalisation
 - l'extension (système de plug-in)
 - la possibilité d'adapter D-Space à un besoin particulier (thèses, images...)
 - des releases plus fréquentes et un logiciel de meilleure qualité
- Implémentation de LOCKSS ?

02/02/2006

42

D-Space 2

- Interopérabilité
 - Design avec Web Services
- Implémentation de nouveaux standards
 - Modèle OAIS (pour la structure des données et les processus de préservation) ; import/export plus standardisé ; voir METS, MPEG-21
 - SRW/U Search pour l'interrogation ?
 - Meilleur support de l'OAI
 - Structure de Métadonnées plus riche
 - RDF ? (car peut supporter plusieurs schémas de métadonnées)
 - Extensible Dublin Core ?

02/02/2006

43

D-Space 2

- Scalabilité
 - Fichiers de plus de 2 Go
 - Nombres d'items portés à 10 millions ou plus
 - Taille globale : plusieurs To
- Augmentation du nombre d'utilisateurs simultanés
- Replication / Mirroring de D-Space
- Version control

02/02/2006

44

Pour patienter : la 1.4

- Flux RSS (flux de syndication, alertes)
- Intégration de CAS (Central Authentication System – Université de Yale) (http://www.esup-portail.org/consortium/espace/SSO_1B/cas/)
- Licence Creative Commons
- Système d'annotation (Post-it)

La 1.4 était annoncée pour fin 2005/début 2006

Pour suivre : <http://wiki.dspace.org/RoadMap>

02/02/2006

45

Sites D-Space

- Le site principal www.dspace.org
- Mailing liste dspace-general@mit.edu
<http://mailman.mit.edu/mailman/listinfo/dspace-general>
- Le site d'échange : wiki.dspace.org
- [Sourceforge.net/projects/dspace](http://sourceforge.net/projects/dspace) (pour administrateurs systèmes, développeurs) : bug report, archives du projet

02/02/2006

46

La communauté D-Space

- First User Group Meeting – MIT (mars 2004)
- International Workshop – Bangalore (mars 2005)
- 2d User Group Meeting – Cambridge Juillet 2005
www.lib.cam.ac.uk/dspace/usergroup2005/programme.htm
- Mise en place d'un conseil de gouvernance

- User Group Meeting - Sydney 23 Janvier 2006
http://www.apsr.edu.au/Open_Repositories_2006/
- User Group Meeting - Bergen Avril 2006

02/02/2006

47

La communauté D-Space Fr

- Mailing list
dspace-fr@services.cnrs.fr

- Site dspace-fr
<https://dspace.msh-alpes.prd.fr/handle/1801/437/>

02/02/2006

48

Open Repository

- Offre de Service de BioMed Central (payante)
- Environnement D-Space préconfiguré et hébergé sur le site de BioMed Central
 - Offre Standard
 - Offre Silver
 - Offre Gold

www.openrepository.com

02/02/2006

49

Conclusion

- Développé à base de logiciel libre (code ouvert)
- Interopérabilité : OAI, OpenURL, Import/Export
- Utilisation de standard : Unicode, DC Qualifié
- Permet d'appliquer une politique d'Etablissement

... un bon départ pour la préservation

02/02/2006

50

Bibliographie générale (1)

- <http://www.erudit.org/documentation/etude/chapitre05.html>
Conception d'un portail de production, de diffusion et de gestion de publications électroniques - Chapitre 5 : L'archivage ; Guylaine Beaudry et Gérard Boismenu ; Groupe interinstitutionnel sur la publication électronique (Québec) ; 2000 (article d'introduction)
- <http://cisti-icist.nrc-cnrc.gc.ca/darf-ran/darf-ran-fra.pdf>
Recherche sur l'archivage numérique - Recommandations émanant de la consultation dans le cadre des recherches sur l'archivage numérique ; Glen Newton & Christine St-Jacques - CNRC Conseil National de Recherches Canada / Institut Canadien de l'Information Scientifique et Technique ; Ottawa ; septembre 2004 (voir les Annexes)
- http://digipat.stis.fgov.be/fr/digipat_docs.asp
Préservation Numérique ; Elena Phalet ; Service d'Information Scientifique et Technique - Politique Scientifique Fédérale - Bruxelles ; juin 2004 + mise à jour juillet 2005 (panorama des projets ; très nombreux liens)

02/02/2006

51

Bibliographie générale (2)

- <http://www.dlib.org/dlib/november05/rosenthal/11rosenthal.html>
Requirements for Digital Preservation Systems - A Bottom-Up Approach ; David S.H. Rosenthal ; Stanford University Libraries ; Novembre 2005
- <http://www.ariadne.ac.uk/issue43/prudlo/>
E-Archiving: An Overview of Some Repository Management Software Tools (LOCKSS, EPrints, D-Space) ; Marion Prudlo ; Avril 2005
- <http://www.soros.org/openaccess/software/>
A guide to Institutional Repository Software v3 ; Open Society Institute ; Aout 2004
- <http://www.dspace.org/implement/leadirs.pdf>
Creating an Institutional Repository: LEADIRS Workbook (Learning About Digital IR Seminars) ; Chapter 3 - Choosing an Institutional Repository Software Platform ; Mary R. Barton ; MIT Libraries ; 2005

02/02/2006

52

Bibliographie D-Space

- DSpace : Aspects organisationnels
<http://hdl.handle.net/1801/439>
- DSpace : Un Système d'Archivage d'Objets Numériques
<http://hdl.handle.net/1801/440>

02/02/2006

53

Glossaire

- OAI-PMH Open Archives Initiative - Protocol for Metadata Harvesting
- RSS Really Simple Syndication
- ERM Electronic Right Management
- SRB Storage Resource Broker
- CAS Central Authentication System
- ILL Inter Library Loan

02/02/2006

54