

Le projet archives ouvertes du **CNRS**

daniel.charnay@in2p3.fr



ArXiv, le modèle

❖ Initiative de Paul Ginsparg à Los Alamos (xxx)

- ◆ 4 000 manuscrits déposés par mois
- ◆ 400 000 manuscrits déposés à ce jour
- ◆ 300 000 consultations journalières
- ◆ des sites miroirs dans le monde entier



Australie, Brésil, Chine, Allemagne, Inde, Israël
Italie, Japon, Russie, Afrique du Sud, France
Corée du Sud, Espagne, Taiwan, Angleterre,...

Archives ouvertes, auto archivage

- ❖ **Le niveau attendu des documents et celui d'une publication soumise dans une revue à comité de lecture**
- ❖ **Les dépôts sont effectués par les auteurs eux-mêmes (ou par les bibliothécaires des laboratoires (choix institutionnel))**
- ❖ **Les documents sont consultables presque immédiatement et gratuitement du monde entier**
- ❖ **Ils sont centralisés dans quelques bases mondiales; la pérennité à long et très long terme des documents est assurée**
- ❖ **Miroirs dans le monde entier**
- ❖ **Complémentarité : beaucoup des documents sont par ailleurs publiés dans les revues traditionnelles.**

Les missions du CCSD

- ❖ **Une initiative du CNRS pour des archives scientifiques**
- ❖ **Une approche internationale :**
 - ◆ Ne pas construire une archive nationale mais s'intégrer dans le mouvement « open archives »
- ❖ **Une approche essentiellement « chercheur »**
 - ◆ Des archives alimentées directement par les chercheurs avec comme vocation première la réalisation d'un outil scientifique offrant l'accès au document intégral
 - ◆ Un outil administratif indirect grâce à la collecte « au passage » de méta données suffisantes
- ❖ **Une mission d'archivage à long terme**

Les services du CCSD

❖ Principalement deux serveurs :

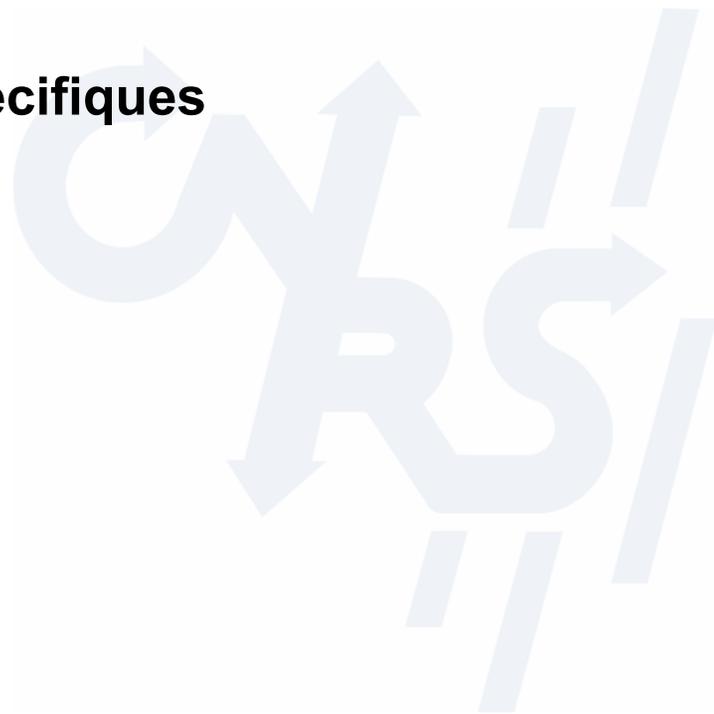
- ◆ **HAL** - Hyper Articles en Ligne interconnecté avec ArXiv
 - Thématique (SHS, MSH de Paris)
 - Démocrite@hal (IN2P3)
- ◆ **TEL** – Thèses en Ligne

❖ Historiquement quelques serveurs spécifiques

- ◆ Jean Nicod (sciences cognitives)
- ◆ ArchiveSIC (sciences de l'information)
- ◆ Cours en ligne, MemSIC ...

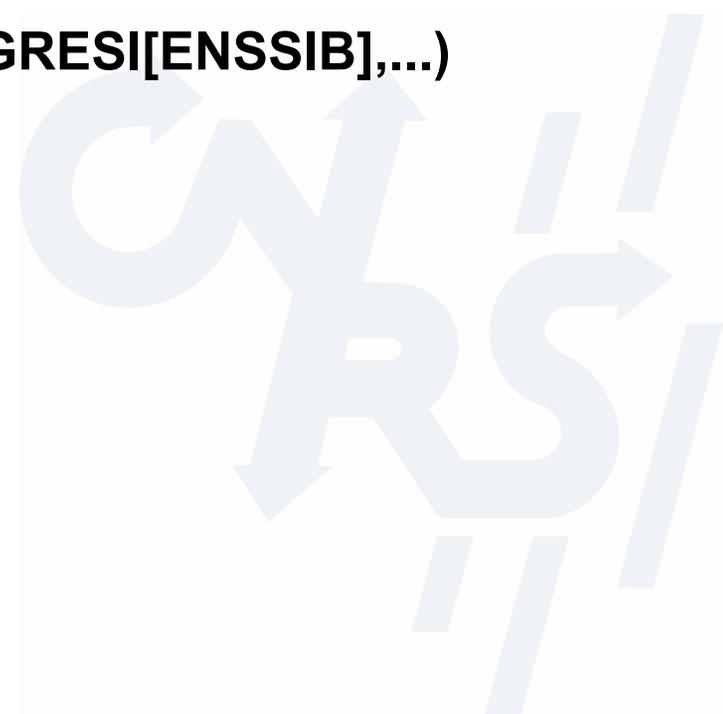
❖ Les miroirs

- ◆ ArXiv
- ◆ PhysNet



Collaborations

- ❖ ArXiv Cornell University (Paul Ginsparg)
- ❖ INRIA (accord cadre)
- ❖ MathDoc, l'UJF, l'IMAG
- ❖ Serveurs en SIC : **archivSic** (URFIST, GRESI[ENSSIB],...)
memSic (INIST)
- ❖ La MSH de Paris : le projet **Tématique**
- ❖ L'INRA ?
- ❖ Etc.



- ❖ **Un frontal pour tous les domaines scientifiques d'ArXiv**
 - ◆ Tous les documents déposés sur Hal le sont simultanément sur ArXiv si le domaine existe
 - Mêmes règles, mêmes formats, davantage de méta données
- ❖ **Un serveur multi-disciplinaire**
 - ◆ ArXiv et non ArXiv
- ❖ **Des méta données plus riches**
- ❖ **Des vues sélectives ou privées fondées sur un système de tampon (épi journaux, congrès)**
- ❖ **Trois types d'objets : articles, articles « anciens », et notices**
- ❖ **Ouvert depuis septembre 2002 (une centaine d'articles/mois déposés)**

- ❖ **Un système souple permettant en plus de la version « généraliste » des entrées sorties personnalisées**
- ❖ **Exemple Tématique en SHS :**
 - ◆ Plus de métadonnées, règles de dépôt différentes (autoarchivage et enrichissement de la notice par des documentalistes)
- ❖ **ou Démocrite@hal à l'IN2P3**
 - ◆ Système orienté « documentaliste »
 - ◆ Intégration de 20 000 notices et texte intégral depuis septembre 2004

TEL – Thèses En Ligne

- ❖ Un serveur de thèses multi-disciplinaire
- ❖ Une collaboration MathDoc CCSD
- ❖ Ouvert depuis juillet 2001
- ❖ Plus de 2000 thèses à ce jour (accroissement ~200 thèses/mois)
- ❖ Alimenté par les auteurs ou les laboratoires...
 - ◆ ...ou automatiquement (IMAG, IN2P3, PASTEL, projet 1000 thèses du XX^e siècle, etc.)
- ❖ Tris
 - ◆ Par universités, laboratoires
- ❖ Vues à intégrer (laboratoire, institution,...)
- ❖ Un modérateur par domaine vérifie l'existence réelle de la thèse
- ❖ Le serveur national de diffusion ?

Formats des documents

- ❖ **Encouragement au dépôt des documents « source »**
 - ◆ Pérennité de l'archivage et conservation à long terme
- ❖ **TeX et LaTeX recommandés (sciences dures, ArXiv)**
- ❖ **Formats visualisables (PDF essentiellement)**
- ❖ **Autres formats possibles (selon les serveurs et les domaines)**
 - ◆ HTML, PS, RTF, DOC, audio ...
- ❖ **Migrations**
- ❖ **Quid de XML ?**
 - ◆ Difficulté à utiliser dans les domaines de physique et de mathématiques (formules, mathML ???)

Règles de dépôt

❖ Dépôt par l'auteur préalablement identifié

- ◆ Obtention libre d'un compte (« examen » du 1er dépôt d'un auteur)
- ◆ Modération avant publication
- ◆ Le document n'est **pas** examiné d'un point de vue scientifique
- ◆ Un spécialiste valide seulement que le document est relatif au domaine

❖ Tampons

- ◆ Un ou plusieurs tampon sont attribués à toute instance qui en fait la demande (université, société savante, journal, epst, laboratoire, etc.) . Des « vues » sélectives peuvent ainsi être proposées affichant seulement les articles « tamponnés »

❖ Versions

- ◆ L'auteur peut proposer plusieurs versions du même document (le tampon est dispensé par version)

Prépub, archives, éditeurs ...

- ❖ Une archive doit être considérée maintenant comme une base d'articles
- ❖ Cette base contient des articles tels qu'ils sont soumis à la revue
- ❖ Les revues (en physique particulièrement) prennent le manuscrit directement dans l'archive
- ❖ La référence de publication est ajoutée dans la notice lors de la publication (on l'espère 😊)
- ❖ Les auteurs peuvent spécifier dans leur contrat le maintien de l'article dans l'archive (en physique cela est souvent prévu)
- ❖ En physique, à part une ou deux revues prestigieuses mais à caractère moins professionnel, les éditeurs ne s'opposent pas au maintien de l'article dans l'archive

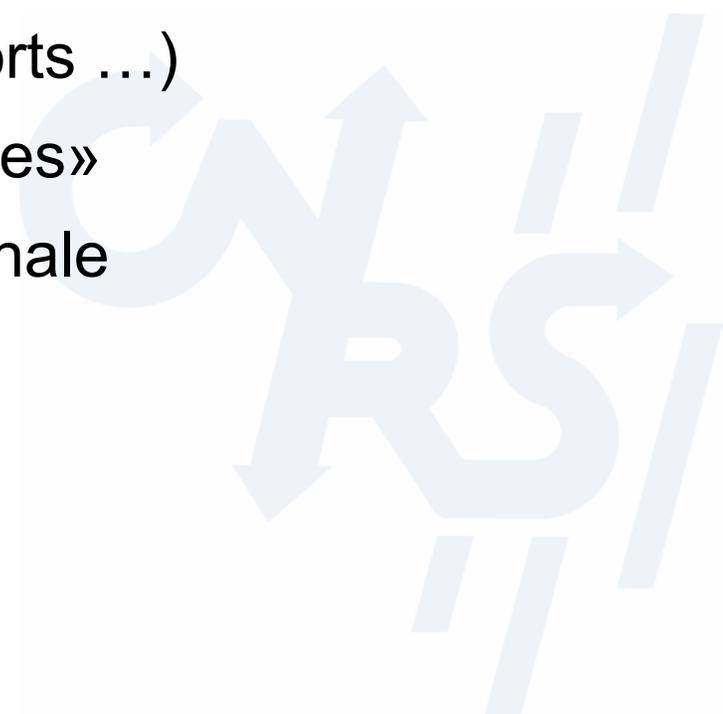
Futur

- ❖ **Etendre la couverture disciplinaire (en particulier vers les sciences humaines et sociales, mais aussi les sciences de la vie)**
- ❖ **Faire un travail de rétroconversion, à la fois des thèses et des articles**
- ❖ **Articuler mieux avec les revues scientifiques « classiques » (liens croisés, ajout de références pour tout le matériau publié)**
- ❖ **Favoriser l'émergence de revues scientifiques « nouvelles » (« overlay journals »), constituées de tables des matières et de liens pointant vers une archive générale et universelle**
- ❖ **Mettre tout le dispositif sous contrôle international pour assurer sa pérennité à très long terme (UNESCO ?)**

Héberger des archives au CCSD ?

❖ Avantages

- ◆ Sauvegarde à long terme
- ◆ Migrations technologiques
- ◆ Suivi des standards (OAI, exports ...)
- ◆ Vues « privées et personnalisées »
- ◆ Visibilité nationale et internationale
- ◆ Gratuit



Structure de HAL



Les objectifs recherchés 1

- ❖ **Archive multidisciplinaire orientée « texte intégral »**
 - ◆ Une seule base de données pour l'ensemble des disciplines
 - ◆ Dépôt du texte intégral mais aussi de notices et de rétro dépôts
- ❖ **Orientée « chercheurs »**
 - ◆ Autoarchivage, mais possibilité d'enrichissement par des professionnels
- ❖ **Centralisée mais ...**
- ❖ **... Multi interfaces**
 - ◆ Préservation de l'identité
 - ◆ Méta données « variables » au-delà de la DTD minimum
 - ◆ Saisies et sorties personnalisées
- ❖ **Internationale**
 - ◆ Interconnectée avec les autres systèmes de références
 - ArXiv en physique / mathématiques / informatique / biologie / ...

Les objectifs recherchés 2

❖ Mesurabilité de la production scientifique

- ◆ Des laboratoires, des institutions et plus globalement de toutes les tutelles
- ◆ À disposition des déposants mais aussi des institutions
 - Génération des listes de publications dans tous formats
 - (XML, TeX, DOC/RTF, PDF, ...)

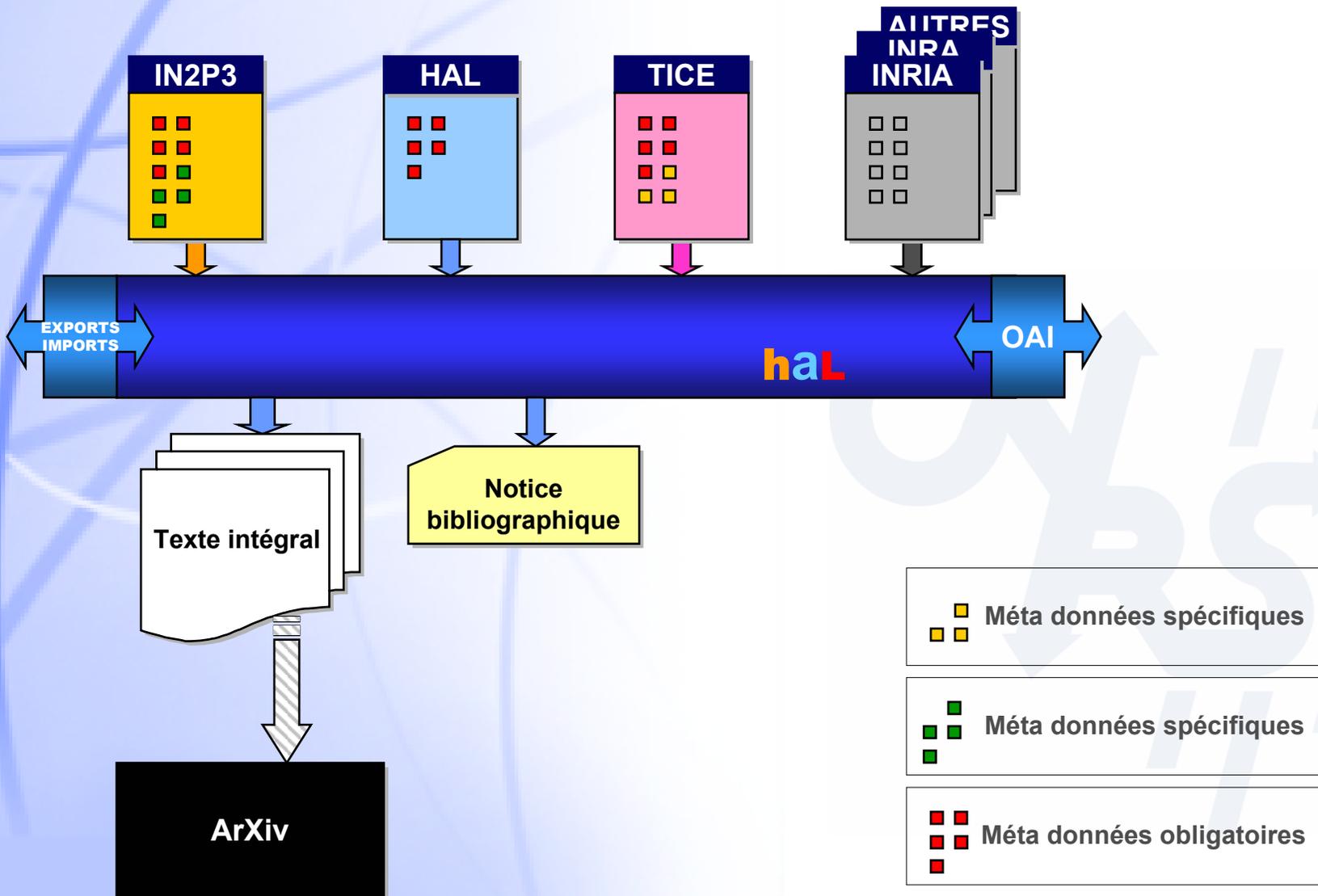
❖ Transferts vers les bases « administratives »

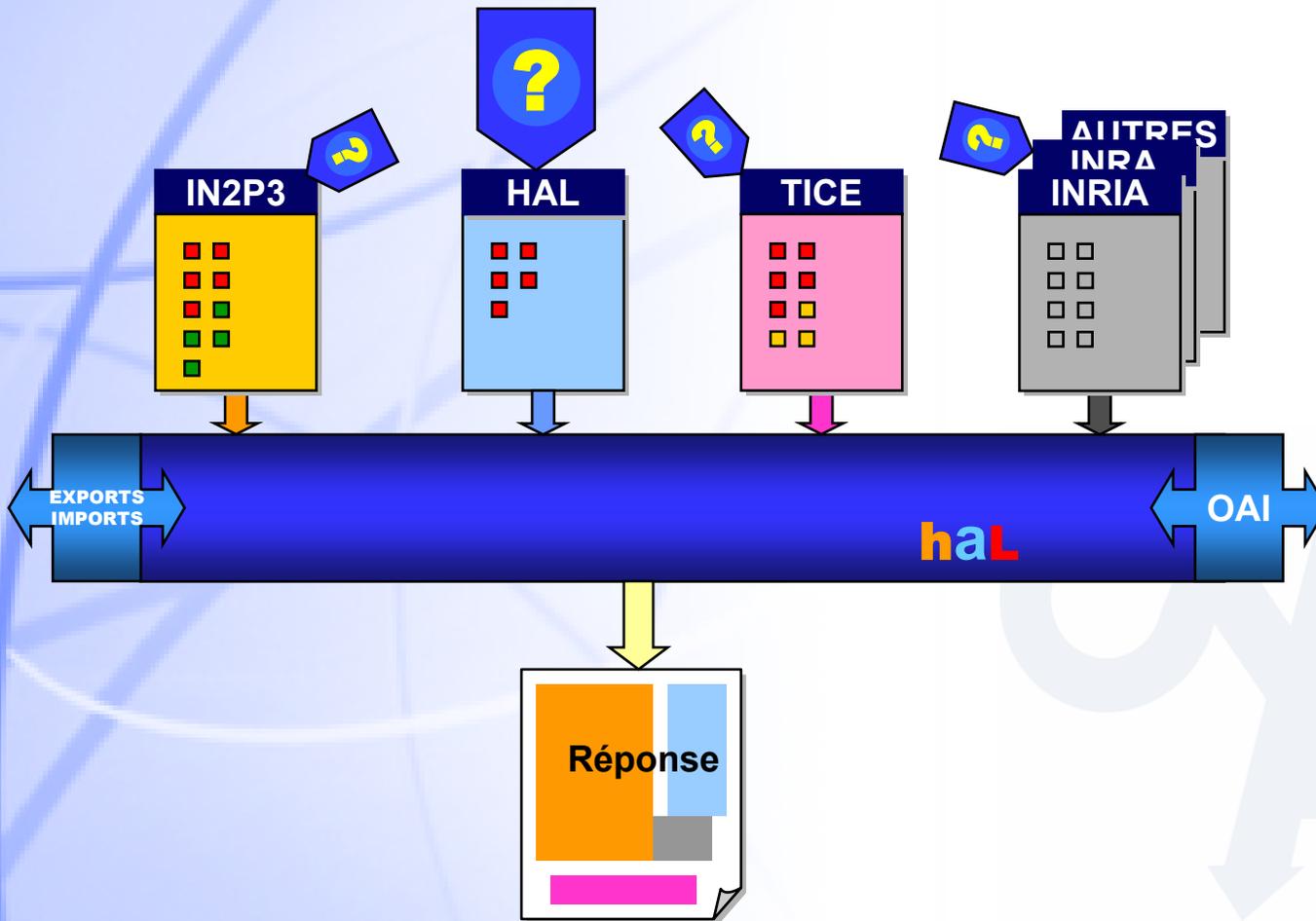
- ◆ Alimentation de publi-cnrs (département PNC)
- ◆ Pré alimentation de CRAC (fiche chercheur CNRS)
- ◆ ...

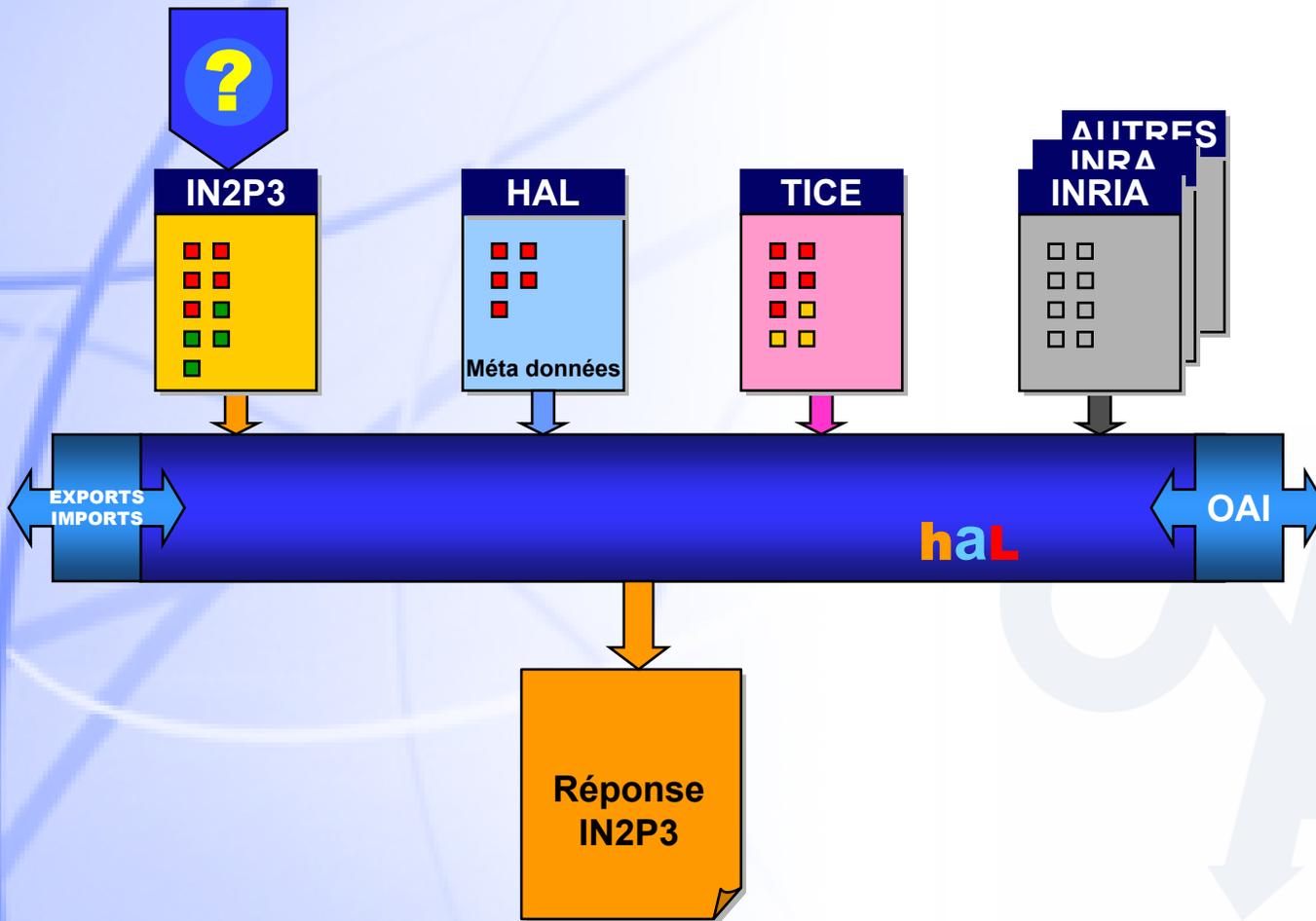
❖ Pérennité

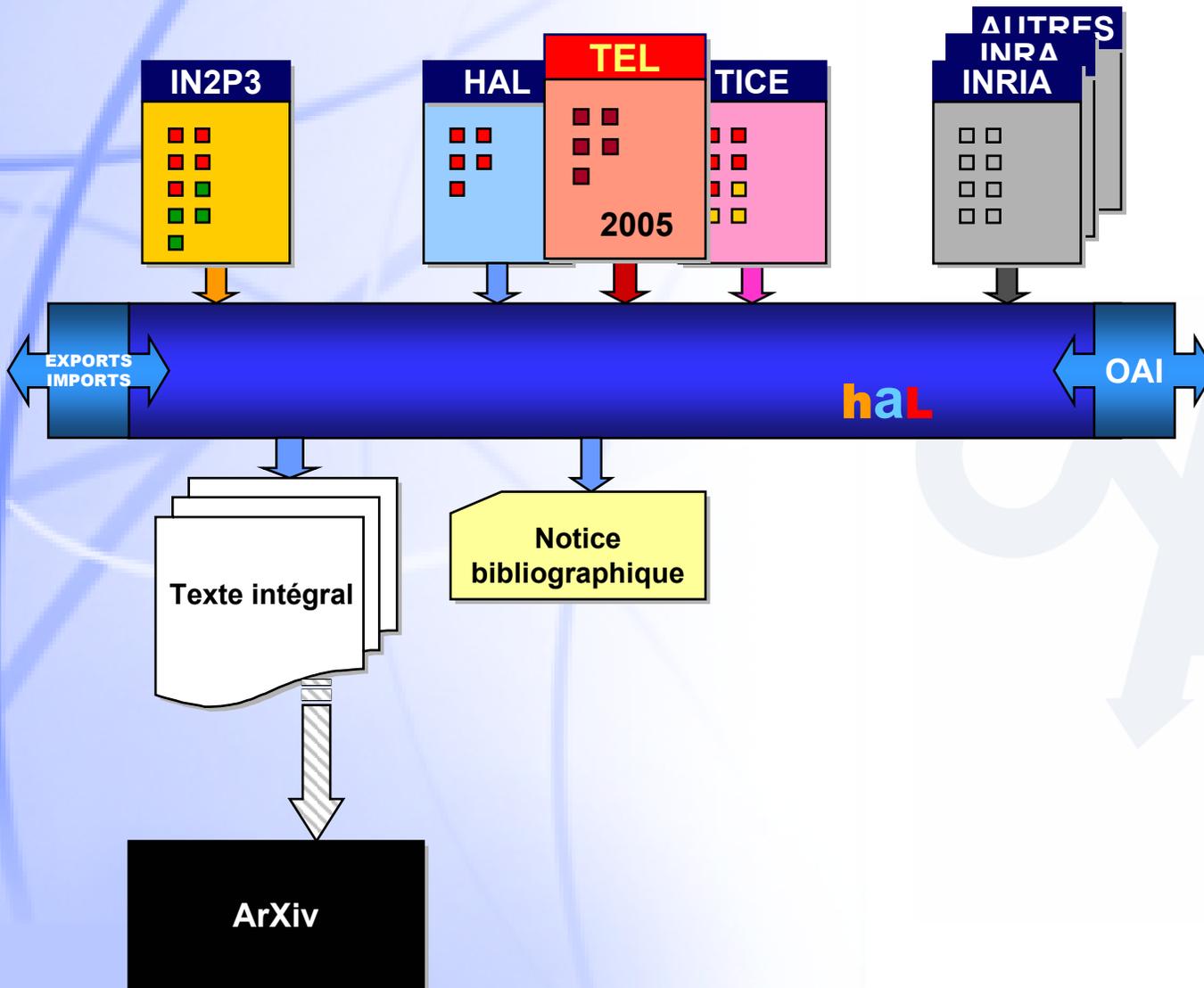
- ◆ Archivage à long terme
- ◆ Stabilité des URLs











Dépôt

❖ Simplification

◆ 4 étapes

1. Méta données
2. Auteurs, laboratoires, tutelles et association
3. [téléchargement du document]
4. Récapitulation contrôle, dépôt

▣ métadonnées ▣ auteurs ▣ fichiers ▣ récapitulation

en arrière possible

Ajout / Modification d'un auteur

champs obligatoires

Prénom: Rozy

Nom: Martin

Autre(s) prénom(s):

Initiales:

Courriel: rozy.martin@ipno.fr

URL page perso:

Information sur le laboratoire sélectionné :

Sigle	IPNO
Intitulé	Institut de Physique Nucléaire d'Orsay
Institution(s)	CNRS : UMR 8608
Université(s)	Université Paris Sud - Paris XI
URL du laboratoire	http://ipnweb.in2p3.fr
Adresse postale	IPN - 15, rue Georges Clemenceau - 91406 ORSAY CEDEX
Pays du laboratoire	France

AUTEUR(S)

modifier supprimer l'auteur sélectionné

Rozy Martin **2**

Emile Dubois

Ajouter Ajouter une liste d'auteurs

LABORATOIRE(S)

modifier supprimer le laboratoire sélectionné

1 - LKB (Lhomond)

2 - IPNO

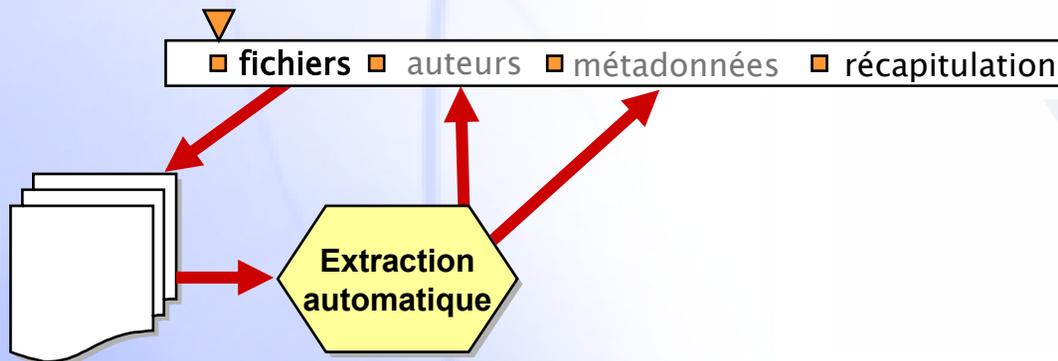
Ajouter un laboratoire pré-remplissé Ajouter

associer les éléments sélectionnés

Dépôt

❖ Choix des étapes

- ◆ L'ordre des étapes peut être paramétré au niveau d'une interface particulière
 - Par exemple on peut décider de commencer par le téléchargement du fichier, afin d'en extraire automatiquement ...
 - ...La liste d'auteurs
 - ...Le résumé
 - ...Les références bibliographiques
 - ...Etc.
 - Ces possibles automatismes dépendent fortement du niveau de structuration du document source, (donc de la discipline ?)



Exemples de déploiements institutionnels de HAL

- ❖ **Sciences physiques : Démocrite**
- ❖ **Sciences humaines et sociales : Tématic**
- ❖ **INRIA : en cours**



URL'S

❖ <http://ccsd.cnrs.fr>

◆ <http://hal.ccsd.cnrs.fr>

◆ <http://archive-edutice.ccsd.cnrs.fr>

◆ <http://democrite.ccsd.cnrs.fr>

◆ <http://hal.inria.fr> / <http://inria.ccsd.cnrs.fr> (en cours)



<http://ccsd.cnrs.fr>
<http://hal.ccsd.cnrs.fr>

