



Daniel.Charnay@in2p3.fr
Franck.Laloe@ens.fr

Centre pour la Communication Scientifique Directe

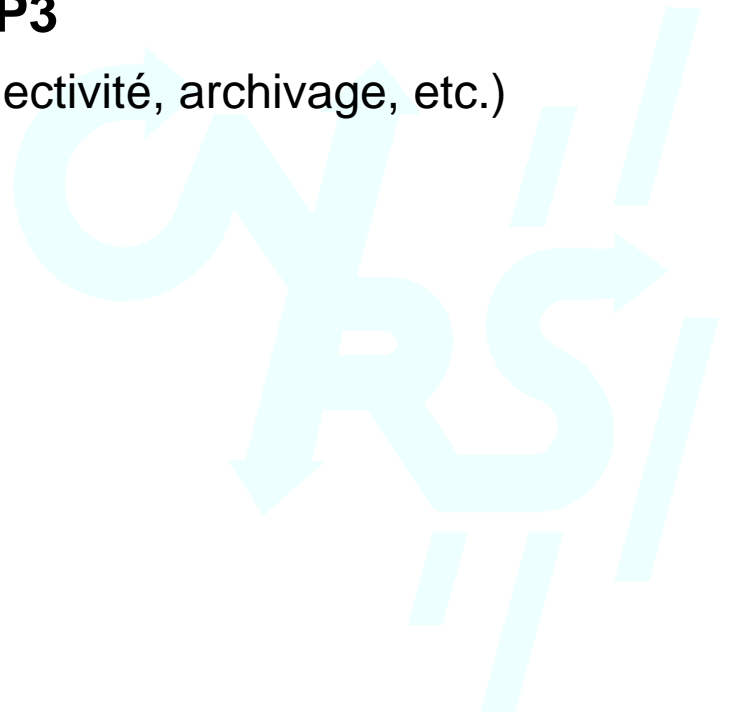


Les missions du CCSD

- ***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
 - Un outil administratif indirect grâce à la collecte « au passage » de méta données suffisantes
- **Une archivage universel à long terme**

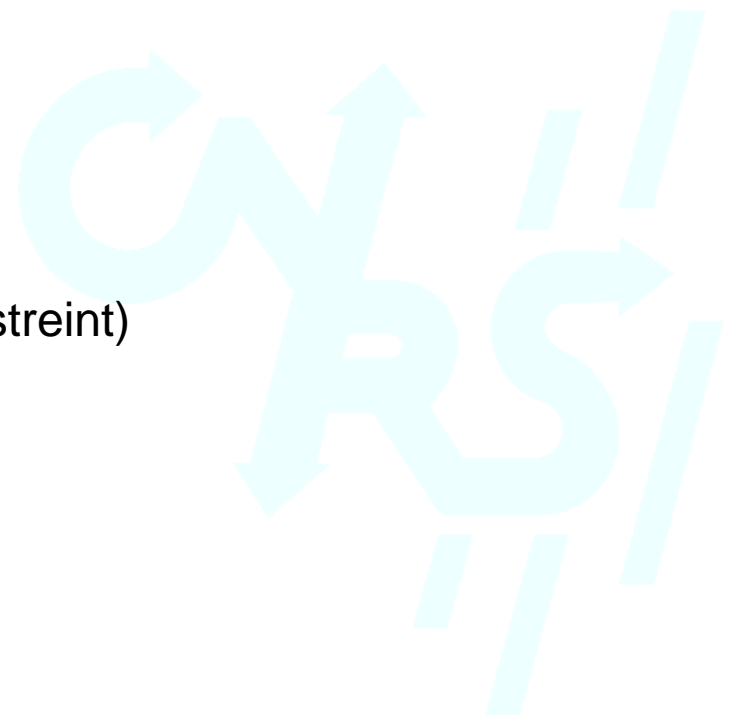
Le CCSD administrativement

- **Unité de service CNRS (SPM) créée en juillet 2000**
 - Sous l'impulsion des départements SPM et PNC
- **Localisée au Centre de Calcul de l'IN2P3**
 - Et bénéficiant de son infrastructure (connectivité, archivage, etc.)
- **2,5 ingénieurs de recherche**



Les services du CCSD

- **Les serveurs généralistes :**
 - HAL - Hyper Articles en Ligne
 - TEL – Thèses en Ligne
- **Les serveurs spécifiques**
 - Jean Nicod (sciences cognitives)
 - ArchiveSIC (sciences de l'information)
 - Prélude (Education Nationale) (accès restreint)
- **Les miroirs**
 - ArXiv
 - PhysNet



Collaborations

- **Pour le développement**

- ArXiv - Paul Ginsparg et son équipe : [HAL](#)
- MathDoc : [TEL](#)

- **Les premiers serveurs**

- [Institut Jean Nicod](#)
- l'URFIST, le GRESI de l'ENSSIB, le CCI/CRECI de l'Université Paris 7 : [ArchivSic](#)
- CERN : CDS et [HAL](#), frontal vers ArXiv

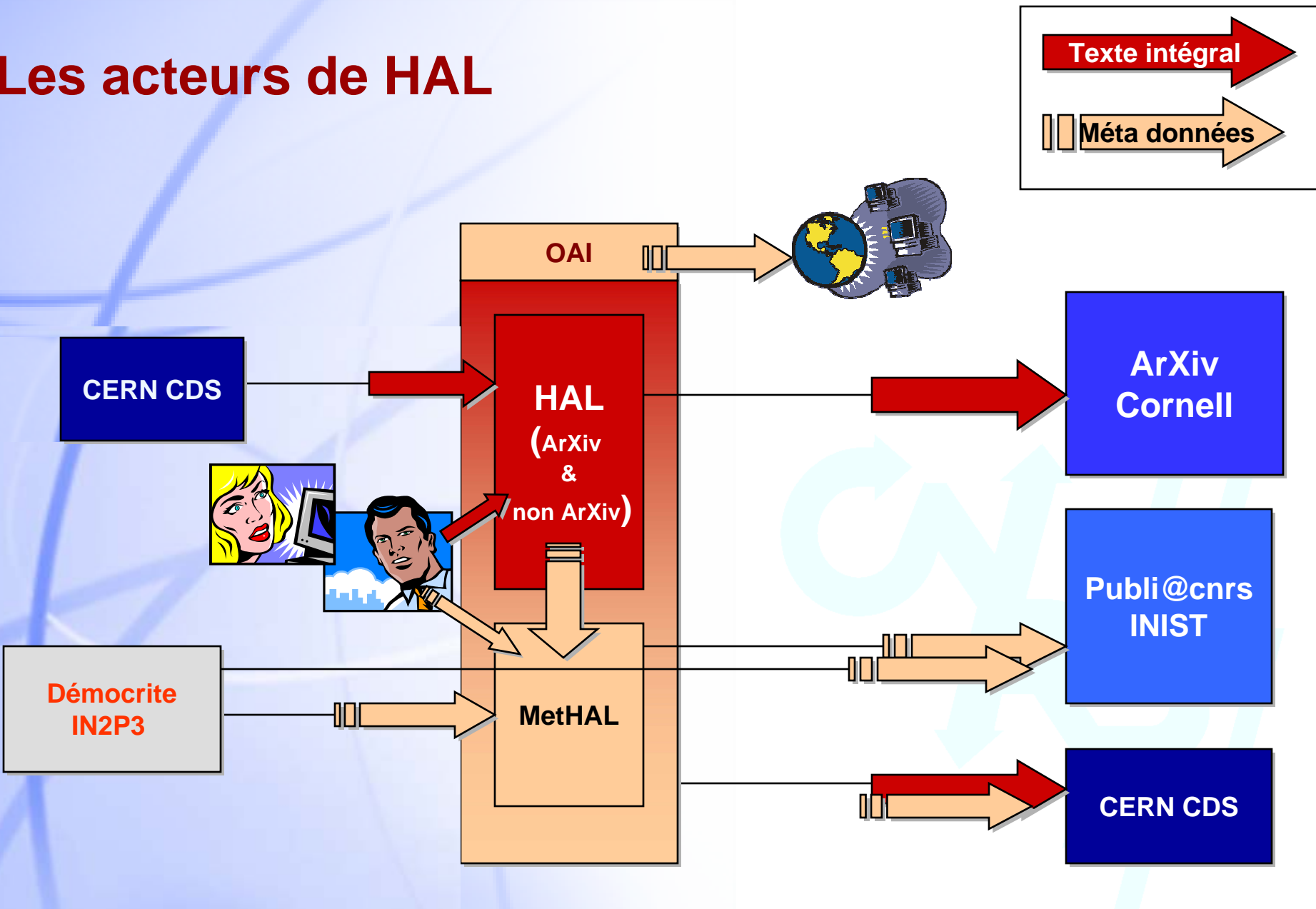
- **Les transferts vers d'autres bases**

- INIST : Collecte et transfert des notices
- CERN CDS



- **Un frontal pour tous les domaines scientifiques d'ArXiv**
 - Tous les documents déposés sur Hal le sont simultanément sur ArXiv
 - ◆ Mêmes règles, mêmes formats, davantage de méta données
- **Un serveur multi-disciplinaire**
 - Archivage des documents pour les disciplines absentes chez ArXiv
- **Des méta données plus riches**
- **Une extension « méta seulement » (MetHal)**
- **Des vues sélectives ou privées fondées sur un système de tampon**
- **Ouvert depuis septembre 2002**
- **Alimentation en méta données de la base de l'INIST**

Les acteurs de HAL

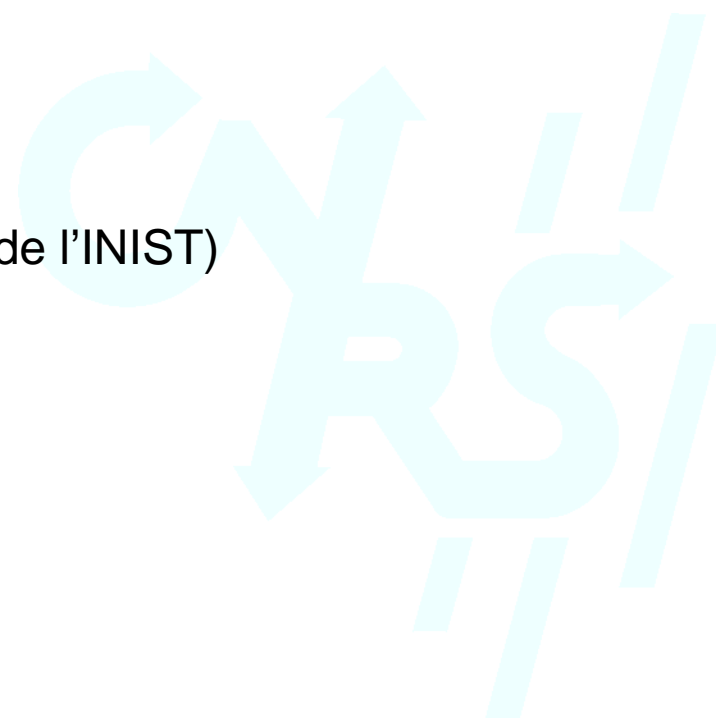


TEL – Thèses En Ligne

- **Un serveur de thèses multi-disciplinaire**
- **Une collaboration MathDoc CCSD**
- **Ouvert depuis juillet 2001**
- **500 thèses environ à ce jour**
- **Alimenté par les auteurs ou les laboratoires...**
 - ...ou automatiquement (via Démocrite de l'IN2P3 par exemple)
- **Tris**
 - Par universités, laboratoires
- **Vues personnalisées (laboratoire, institution,...)**
- **Un « contrôleur » par domaine vérifie l'existence réelle de la thèse**

OAI et exports

- **OAI : supporté sur tous les services**
 - HAL (2.0)
 - TEL (2.0)
- **Export**
 - TEL vers MathDoc (XML ~ OAI ++)
 - HAL vers article@cnsr (XML selon DTD de l'INIST)
 - HAL vers CERN (XML)



Règles de dépôt

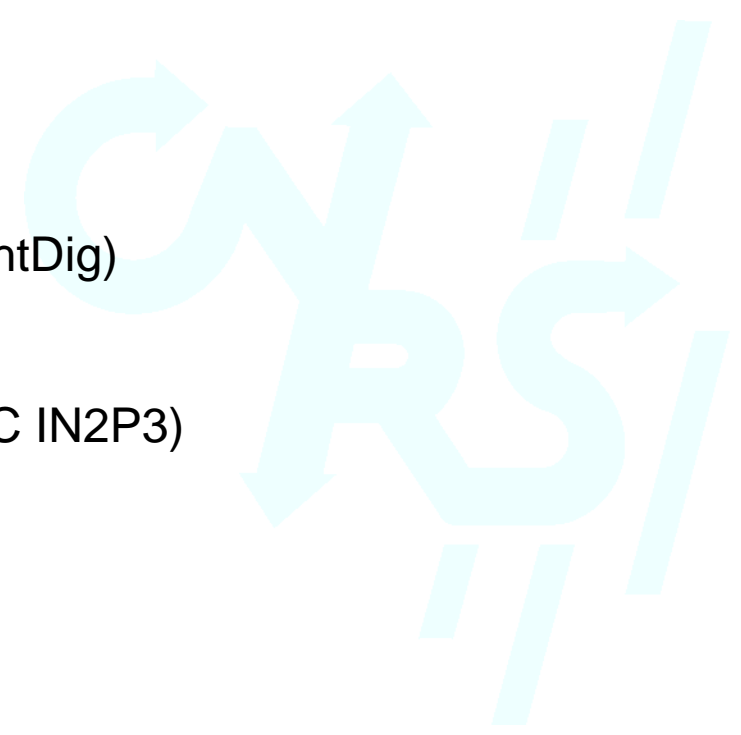
- **Dépôt par l'auteur préalablement identifié**
 - Obtention libre d'un compte depuis un domaine reconnu (.edu, .univ-xxx, .cern.ch, ...)
- **Modération avant publication**
 - Le document n'est **pas** examiné d'un point de vue scientifique
 - Un spécialiste valide seulement que le document est relatif au domaine
- **Tampons**
 - Un ou plusieurs tampons sont attribués à toute instance qui en fait la demande (université, société savante, journal (*overlay journals*), 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)

Formats des documents

- **Encouragement au dépôt des documents « source »**
 - Pérennité de l'archivage et conservation à long terme
- **TeX et LaTeX recommandés (ArXiv)**
 - Extraction automatique des auteurs (en cours). Problème des publications en physique des particules (500 auteurs et plus pour un article !)
- **Formats visualisables (PDF essentiellement)**
- **Autres formats possibles (variable selon le serveur et le domaine)**
 - HTML, PS, RTF, DOC, audio ...
- **Quid de XML ?**
 - Difficulté à utiliser dans les domaines de physique et de mathématiques (formules, mathML ???)

Technologies

- **HAL**
 - Développement spécifique CCSD
 - PHP, MySQL
- **TEL et serveurs spécifiques :**
 - Eprint 1.1 (Perl, MySQL)
 - Francisation, contrôle, indexeur *full text* (htDig)
- **Hardware**
 - Linux & NAS – Robot de sauvegarde (CC IN2P3)



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



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**

URL's

- www.ccsd.cnrs.fr
- tel.ccsd.cnrs.fr
- hal.ccsd.cnrs.fr
- archiveSIC.ccsd.cnrs.fr
- JeanNicod.ccsd.cnrs.fr
- arxiv.ccsd.cnrs.fr
- www.ccsd.cnrs.fr/physnet





Daniel.Charnay@in2p3.fr
Franck.Laloe@ens.fr

**Centre pour la
Communication
Scientifique
Directe**

