

# Projet de base bibliographique ENS LSH

---

Ceci est le compte rendu d'une réunion qui s'est déroulée le 15/03/2005 sur les thèmes connexes des bases bibliographiques et bases de publication à l'Ecole. Y participaient :

- Serge Heiden (ICAR)
- Maud Ingarao (Desanti)
- Cécile Laube (Administration de la Recherche)
- Pierre Mounier (Cédille)
- Samantha Saïdi (Triangle)
- Bernard Teissier (Bibliothèque)

L'objectif de la réunion était de faire le point et de débattre des différentes bases de publication et bases bibliographiques existant à l'Ecole, des besoins en la matière, de l'état d'avancement des outils utilisés localement, ainsi que du contexte global dans lequel tout cela s'insère.

Un certain nombre de besoins ont été recensés, une perspective s'est dessinée, d'où est sortie une proposition d'action sur laquelle un consensus a été atteint entre les participants. Le présent document, amendé et validé par l'ensemble des participants, est le reflet de ce consensus.

## 1. Les besoins

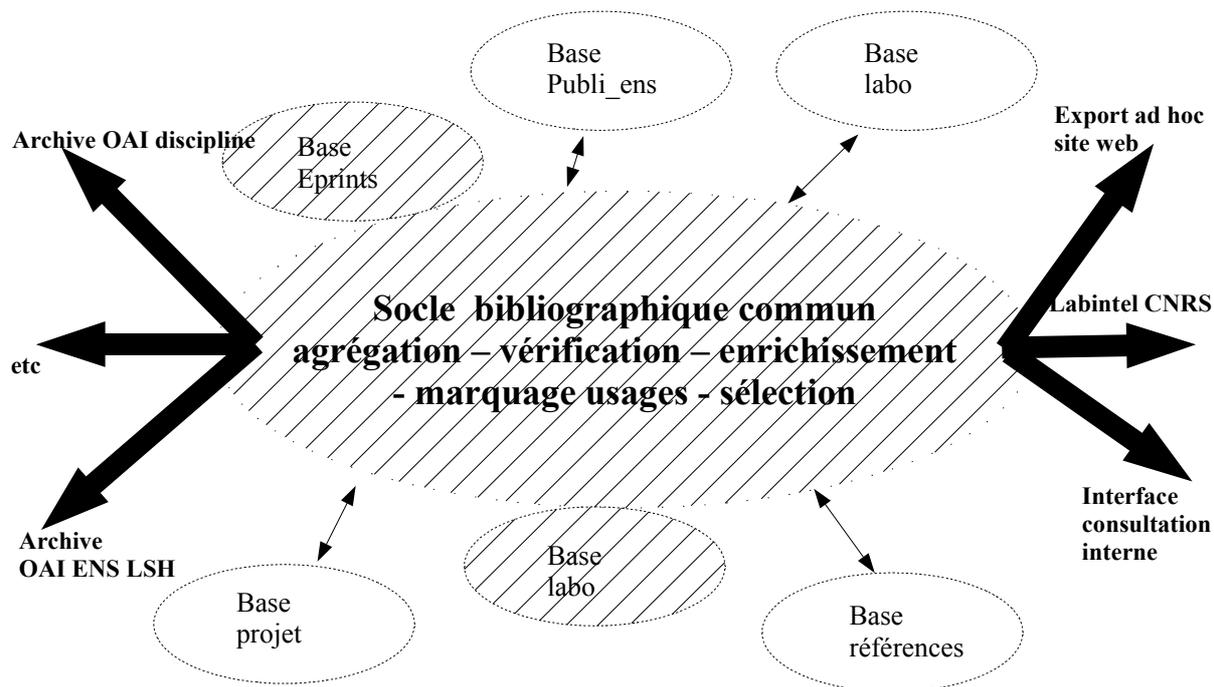
Un certain nombre de laboratoires, à commencer par Triangle, souhaitent disposer d'une base des publications de leurs chercheurs, qu'ils veulent pouvoir gérer comme ils l'entendent, en y incluant articles publiés diffusés à tous, communications non publiées disponibles seulement pour les membres du laboratoire, ou encore d'autre catégories de documents auxquels sont attachés des critères de diffusion divers.

D'autres laboratoires, comme ICAR souhaitent disposer d'un outil de gestion des bibliographies de projets ou d'équipes, dotés ou non des documents eux-mêmes.

Les personnes qui se sont investies dans la mise en place, l'alimentation et la gestion de la base Eprints de l'Ecole (Archive ENS-LSH) souhaitent disposer désormais d'un relais institutionnel pour deux raisons : tout d'abord, il est souhaitable qu'une évolution se fasse vers un système plus élaboré et plus stable que le logiciel utilisé actuellement ; ensuite il devient impératif, étant donné les récentes initiatives en la matière, que l'Ecole dispose d'une importante base de publications afin de maintenir ou d'accroître sa visibilité au sein des instruments de recherche documentaire pour les publications scientifiques.

## 2. La perspective

Il apparaît à tous que la perspective de recourir au logiciel SIM d'Archimed constitue une bonne base de réflexion. L'important est d'y recourir de manière modulaire, c'est-à-dire, très exactement, de s'en servir comme concentrateur par agrégation d'informations bibliographiques conçues et saisies en amont au sein de bases autonomes, qu'elles soient gérées ou non par le SIM. Chacune de ces références, agrégée au sein d'une base bibliographique commune, pourra éventuellement faire l'objet d'un traitement (de type enrichissement ou vérification) en fonction de l'utilisation qui en sera faite (et donc du niveau d'exigence auquel elle doit répondre). Elle sera finalement marquée/labellisée pour l'une ou l'autre des utilisations qu'offre le système. Schématiquement, on peut représenter le processus de la manière suivante :



Un point critique est la communication entre la base commune et les bases spécifiques. Une discussion doit avoir lieu sur des formats et une structure bibliographique commune pour une interopérabilité minimale. Par défaut, le SIM peut fournir une interface spécifique à des sous-ensembles bibliographiques qui seraient gérés directement dans le système selon le modèle commun. Un autre point doit être précisé : la communication entre base commune et bases spécifiques se fait dans les deux sens : la base spécifique alimente la base commune, mais elle bénéficie d'elle en retour (services d'import et d'enrichissement-validation documentaire des références). Dans tous les cas, la base commune n'est pas publique en tant qu'entité, mais fournit des services aux bases partenaires et produit des flux d'export ciblés pour des usages spécifiques.

L'intérêt d'une telle organisation n'est pas négligeable : elle permet d'articuler l'autonomie des utilisateurs (collectifs de recherche) dans le choix des outils, l'organisation de la saisie et la gestion de leur base avec une centralisation permettant la mutualisation de l'effort de saisie et de vérification et la mobilisation de l'ensemble des références pour des usages non prévus par chacun des collectifs.

Concrètement, cela signifie :

- la possibilité d'éviter la double saisie
- la possibilité pour chacun des collectifs de répondre immédiatement à ses besoins sans devoir attendre que le système central soit totalement fonctionnel
- la prise en compte dans l'organisation même du système de la récupération de l'existant
- la possibilité de mobiliser rapidement des masses importantes de références qui peuvent être

fortement visibles de l'extérieur (Sudoc, moteurs scientifiques, chercheurs et enseignants)

### 3. Propositions d'action

Plusieurs chantiers doivent être ouverts.

- Chantier « **structure bibliographique** » : un groupe de travail doit produire dans des délais raisonnables un consensus sur une structure bibliographique générique qui devra être utilisée par toutes les bases partenaires pour communiquer avec la base commune. Il pourra s'appuyer en particulier sur les initiatives existantes au sein de l'Ecole. Au delà, il pourra prolonger ses activités sur un travail de veille et de recommandation.
- Chantier « **base commune** » : Il faut déterminer la structure de la base commune, réaliser les développements informatiques nécessaires, déterminer l'organisation du travail de gestion documentaire de la base, affecter les ressources humaines nécessaires.
- Chantier « **formats d'échange** » : le choix d'un certain nombre de formats d'échanges de données aussi bien pour les communications entre bases que pour la production de flux de données externes doit être fait. Il faut déterminer et réaliser les développements informatiques nécessaires.
- Chantier « **flux de sortie** » : il faut déterminer le type et le format des flux de sortie qui seront proposés, réaliser les développements informatiques nécessaires. Il faut organiser l'articulation documentaire/éditorial sur certains flux de sortie, en particulier ceux sur lesquels va être apposé le label institutionnel ; aider les laboratoires, les équipes, les chercheurs à utiliser les outils qui leur permettent d'opérer cette articulation pour eux-mêmes.
- Chantier « **identifiants numériques** » : le besoin d'un service d'attribution d'identifiants uniques aux objets numériques produits localement est criant. Les réponses à ces besoins, du type serveur Open URL, entrent légitimement dans la problématique de base bibliographique telle qu'elle est exprimée ici.

Rappelons que l'outil de gestion électronique de documents est acquis et qu'il est déjà fonctionnel sur au moins une base (base des vidéos). Le catalogue d'ENS Editions est aussi en cours de réalisation avec le SIM. Malgré tout, il faut prévoir la mobilisation de forces humaines dotées de compétences documentaires et informatiques (développement), ou, le cas échéant, imaginer des schémas de pilotage en interne de développements informatiques externalisés.

### 4. Conclusion

Plusieurs bases existent déjà ; d'autres sont sur le point de se développer. Il ne faudrait pas excessivement tarder, à la fois pour des raisons internes (en particulier sur le chantier « structure bibliographique » qui demande que des discussions soient dès à présent engagées) que du fait de récents mouvements dans notre environnement. Un certain nombre de signes récents manifestent le succès de l'implantation du protocole OAI dans l'espace institutionnel français :

- l'adoption par le portail Sudoc du protocole pour agréger les références en provenance de différentes bases : <http://www.portail-sudoc.abes.fr/>
- l'adoption du même protocole par le groupe de travail CENS-CAIRN-Erudit-revorg-Persée pour l'échange de données entre leurs plate-formes : par exemple : <http://oai.persee.fr/oai.do?verb=Identify>
- la multiplication à la fois des outils de publication capable de répondre au protocole ([GNU Eprints](#), [Dspace](#), [SDX](#), [Lodel](#), [Spip](#)) pour ceux que nous connaissons) et des bases OAI existantes, en particulier du fait du développement du [CCSD](#), mais aussi de l'investissement d'institutions de recherche et d'enseignement supérieur comme la nôtre ([Archive ENS-LSH](#)).

- L'utilisation par les moteurs de recherche spécialisés comme [In extenso](#) et probablement [Google scholar](#) du protocole pour indexer les publications scientifiques.
- L'apparition d'initiatives proposant des services de « moissonnage » de données SHS dans l'espace francophone : <http://www.msh-reseau.prd.fr/oai/>
- L'engagement politique réitéré des grands organismes de recherche nationaux pour appuyer le mouvement des archives ouvertes : <http://www2.cnrs.fr/presse/communique/640.htm>

Du coup, c'est un certain type de circulation de l'information (en particulier bibliographique) qui risque de prendre de l'importance à partir de maintenant : le système que nous proposons, reposant sur une double articulation en entrée et sortie de données a des chances d'être très pertinent dans notre environnement.

Nous pensons que ce projet, à l'articulation des nouvelles technologies et des missions de la bibliothèque de l'École doit être porté pour la mise en place d'un premier lot de fonctionnalités avant le prochain quadriennal, permettant de prévoir plus précisément des financements pour la suite.