

Projet HAL-Inspire



Le projet HAL-Inspire, initié par la direction de l'IN2P3 en coopération avec le CERN et soutenu par la DIST-CNRS, vise à assurer de façon exhaustive et qualitative le référencement, la valorisation, l'archivage et le partage des publications du CNRS en physique des deux infinis.

Problématique :

- La production scientifique de l'IN2P3 doit être déposée dans l'archive ouverte française HAL, pour répondre aux prérogatives de la Commission Européenne et du MENESR.
- Les chercheurs de l'IN2P3 utilisent déjà massivement l'archive de prépublications ArXiv et la base de données de publications Inspire-HEP ; ils déposent rarement eux-mêmes dans HAL.
- Le nombre de professionnels de l'IST à l'IN2P3 décroît, ces derniers ne peuvent assurer seuls le dépôt HAL pour tout l'institut. Les laboratoires sans documentalistes n'ont donc souvent personne pour déposer leurs articles dans HAL et leur visibilité s'en trouve très atténuée.
- Inspire-HEP, développé au CERN, est un système d'information fonctionnel, fiable et reconnu qui moissonne dans le domaine de la physique les pré-publications dans ArXiv et les publications chez les éditeurs. Il intègre donc déjà une grande partie des publications de l'IN2P3.

Principe :

- Inspire-HEP est désormais la base de données centrale où se font le référencement, le traitement et la valorisation des publications de l'IN2P3. Les notices sont ensuite exportées de façon automatique dans l'archive nationale HAL.
- Les types de publications concernées sont, au début, les articles (ACL) et actes de conférences.
- Pour assurer un référencement exhaustif des publications affiliées à l'IN2P3, l'équipe d'Inspire-HEP a élargi la liste des domaines et des journaux moissonnés par leur système, et a créé une collection spécifique pour l'IN2P3. Le reste est soumis manuellement.
- Une équipe de documentalistes de l'IN2P3 participe au réseau mondial des catalogueurs d'Inspire-HEP et corrige, enrichit et valide les publications avec au moins une affiliation française.

Mise en place :

- Un accord de coopération a été signé en 2015 entre le CERN et le CNRS.
- Le projet a été présenté en 2016 aux directeurs d'unités IN2P3, afin d'obtenir leur adhésion à une approche mutualisée de la gestion des publications.
- Un chargé de la conduite du projet a été recruté pour 2 ans (2016 - 2018) et a été formé au CERN comme expert Inspire-HEP. Il coordonne le projet, anime l'équipe des catalogueurs, assure les tâches de catalogage complexes et sert d'interlocuteur entre le CERN et le CNRS.
- 9 documentalistes de l'IN2P3 ont été formés par lui au catalogage des publications, équipe de professionnels volontaires offrant une partie de leur temps de travail selon disponibilité.
- Une coopération technique a lieu entre les équipes informatiques du CCSD (HAL) et du CERN.
- Le traitement et la valorisation des publications ont débuté de façon exhaustive au 1.01.2017.
- Au cours du 1er semestre 2017, le dépôt des publications dans HAL encore effectué par les documentalistes de l'IN2P3 est progressivement arrêté.
- Le 3 juillet 2017, après une phase de bêta-tests, l'export groupé de notices d'Inspire-HEP vers HAL a été lancé. L'export sera dans un premier temps réalisé à intervalles réguliers. Ultérieurement, l'équipe d'Inspire-HEP mettra en place l'export en temps réel.

Catalogage (dans Inspire-HEP) :

- La condition pour qu'une notice soit exportée dans HAL est qu'elle ait été au préalable vérifiée et enrichie de métadonnées, puis validée par un catalogueur de l'IN2P3.
- Ne sont concernés que les articles ou actes publiés par un éditeur, avec DOI.
- Les métadonnées présentes et contrôlées sont notamment : tous les auteurs (même grandes collaborations), toutes les affiliations (même étrangères), les noms de collaboration et expérience, les liens vers le texte intégral (arXiv + éditeur), les références bibliographiques liées par hypertexte, autres (titre au format LaTeX, résumé, etc.).
- Un travail approfondi est mené sur la qualité du référencement des affiliations françaises dans Inspire-HEP. Les profils d'auteurs affiliés à une institution française sont également améliorés.

Archivage (dans HAL) :

- Les notices sont importées dans la base principale de HAL (non dans la collection HAL-IN2P3), mais s'affichent ensuite automatiquement dans les collections des structures concernées.
- Chaque affiliation française d'Inspire-HEP est reliée à son équivalente HAL, ainsi les affiliations françaises sont automatiquement affichées et référencées telles que dans AuréHAL.
- Le contributeur et propriétaire des notices importées d'Inspire-HEP est nommé "Inspire HEP" et l'adresse de contact est : inspire-hal-cataloger@cern.ch
- Toute personne souhaitant apporter une modification à une notice importée peut contacter cette adresse et demander un partage de propriété de la notice.
- Si une notice d'une publication est créée directement dans HAL avant l'importation de son équivalent Inspire-HEP, elle ne sera pas remplacée ni doublonnée (sur la base du DOI ou arXivID), l'import sera alors annulé pour cette notice.

Environnement :

- Sont également concernées par ce projet des publications d'autres instituts du CNRS (notamment l'INSU et l'INP), du CEA (IRFU, IPhT), de plusieurs établissements publics et universités.
- Ces structures vont bénéficier de l'application du projet HAL-Inspire en voyant une partie de leurs publications prises en charge, c'est-à-dire valorisées dans Inspire-HEP puis archivées dans HAL.
- Elles sont à ce titre informées de l'avancée du projet et invitées à en devenir partenaires.
- Afin d'assurer l'exhaustivité du référencement de ses publications, l'IN2P3 va promouvoir auprès des chercheurs le dépôt des prépublications dans ArXiv et l'utilisation des identifiants ORCID.
- Une application nommée LIMBRA (Listes et Métriques Bibliographiques Automatisées) est développée en parallèle du projet afin d'extraire d'une part des données bibliométriques à partir d'Inspire-HEP, d'autre part d'établir des listes ordonnées de publications pour chaque laboratoire. Cet outil est en cours de test par les documentalistes et est également proposé aux correspondants scientifiques des laboratoires de l'Institut (formations prévues).
- Une importante refonte de la base de données Inspire-HEP est prévue fin 2017. Elle permettra notamment l'envoi des fichiers pdf (fulltext) avec les notices, si les droits le permettent (licence CC, Scoap3). Les notices déjà présentes dans HAL seront mises à jour.

Informations utiles :

- Mathieu Grivès, chargé du projet : grives@lpccaen.in2p3.fr ; 02.31.45.25.22
- Sandrine Guesnon, chargée de mission IST pour l'IN2P3 et responsable du Réseau Démocrite (documentalistes de l'IN2P3) : guesnon@lpccaen.in2p3.fr ; 02.31.45.25.20
- Ursula Bassler, directrice adjointe de l'IN2P3 et chargée de l'IST : ursula.bassler@in2p3.fr ; 01.44.96.47.40
- Liens : [Inspire-HEP](#) (dont [notices validées](#)) ; [HAL](#) (dont [notices Inspire](#)) ; [ArXiv](#) ; [Limbra](#) ; [Portail des ressources documentaires IN2P3](#)

Mathieu Grivès, juillet 2017