

### Carte blanche CCSD

Journées CasuHAL 2025

25.06.2025

Yannick Barborini, Hélène Bégnis, Pascal Lubino









### Déroulé

01

Réalisations

02

Développements en cours

03

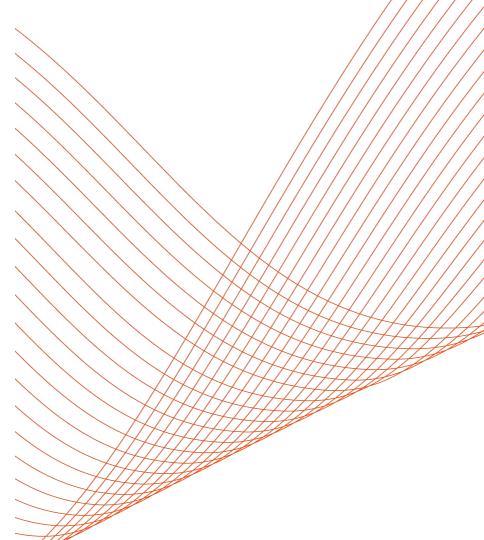
Validation technique de dépôts dans HAL

04

Accompagnement des utilisateurs de HAL



01





### **HALiance**

### **WP 3** - Extraction de métadonnées et alignement sur des données de référence

Récupérer des métadonnées et des identifiants dans les fichiers déposés et enrichir automatiquement la base HAL

### **WP4** - Gestion des métadonnées bibliographiques importées

Organiser et documenter le cycle de vie des métadonnées bibliographiques importées dans HAL

#### WP5 - Qualification des preprints

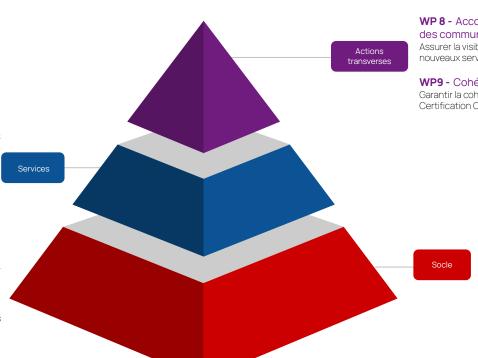
Articuler HAL à des dispositifs de relecture des preprints et rendre visible le cycle de la publication

### **WP6 -** Collecte ciblée et automatisée de publications

Mettre en œuvre une nouvelle voie d'alimentation de HAL par la collecte de publications scientifiques

#### WP7 - Lien Publications-Données de la recherche

Articuler les publications présentes dans HAL et les données de la recherche qui y sont liées



### **WP 8 -** Accompagnement et engagement des communautés

Assurer la visibilité et l'appropriation des nouveaux services HAL par ses utilisateurs

**WP9 -** Cohérence avec les principes FAIR Garantir la cohérence avec les principes FAIR & Certification CTS

#### WP1 - Infrastructure matérielle

Mettre à niveau et sécuriser l'infrastructure physique de HAL

#### WP2 - Infrastructure logicielle

Migrer, sécuriser et ouvrir le code logiciel



### **HALiance**

### **WP 3** - Extraction de métadonnées et alignement sur des données de référence

Récupérer des métadonnées et des identifiants dans les fichiers déposés et enrichir automatiquement la base HAL

### **WP4** - Gestion des métadonnées bibliographiques importées

Organiser et documenter le cycle de vie des métadonnées bibliographiques importées dans HAL

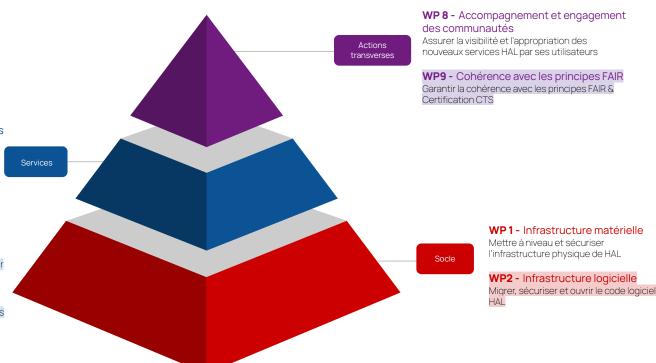
#### **WP5** - Qualification des preprints

Articuler HAL à des dispositifs de relecture des preprints et rendre visible le cycle de la publication

### **WP6 -** Collecte ciblée et automatisée de publications

Mettre en œuvre une nouvelle voie d'alimentation de HAL par la collecte de publications scientifiques

**WP7 -** Lien Publications-Données de la recherche Articuler les publications présentes dans HAL et les données de la recherche qui y sont liées





### Évolution infrastructure Logicielle

Migration progressive du code de HAL vers un nouveau socle technique

→ Sécurité → Modernisation → Performances

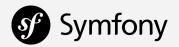


Equipe mixte (prestataire - CCSD)

Internalisation CCSD

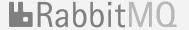


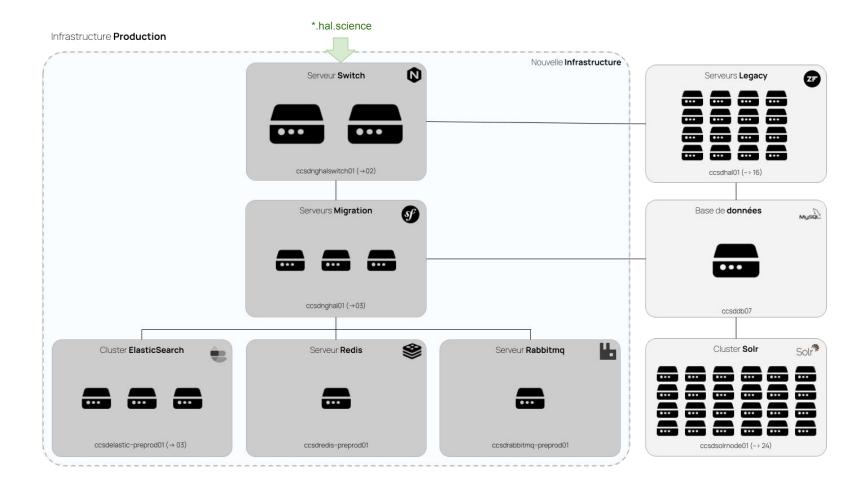
NGINX







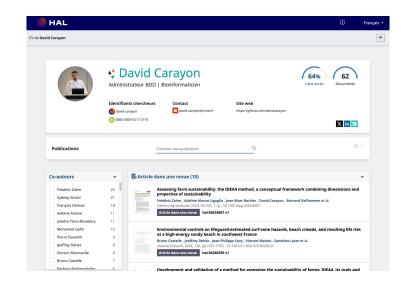


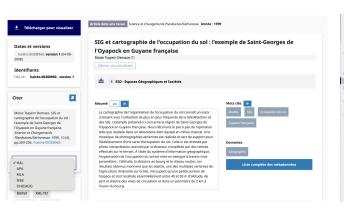


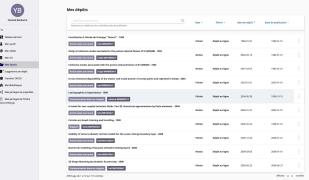


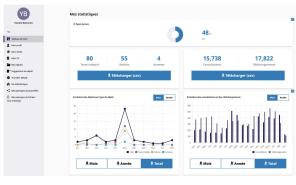
### Etat d'avancement

- CV en production avec intégration datasets, impression, ... (S1 2025)
  - export Word (S2 2025)
  - ouverture du code (S2 2025)
- HAL
  - Page d'un dépôt avec nouvelle citation (S2 2025)
  - Ecrans utilisateurs (S2 2025)
  - Démarrage migration des écrans d'administration









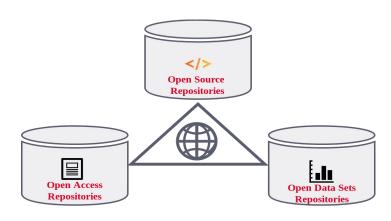


### Nouveau système de relations

### Enjeu scientifique

Dans la quête pour la **reproductibilité**, la réutilisation des résultats scientifiques et la **transmission** des connaissances scientifiques, **trois piliers de la Science Ouverte**:

- des **articles scientifiques** décrivant les résultats établis par les chercheur·es
- les ensembles de données
- les **logiciels** utilisés ou produits



**Three pillars of Open Science** Software Heritage CC-By 4.0 2019



### Lier les données aux publications : Formulaire de création

### 1 Choisir le type de relation

Basé sur une typologie simplifiée de DataCite (par exemple, isPartOf, isSupplementedBy)

### 2 Choisir le dépôt/archive

Prend en charge les principaux dépôts de données français :

- HAI
- Recherche Data Gouv
- Nakala

#### **8** Recherche assistée

Aide à localiser les ensembles de données dans les dépôts sélectionnés

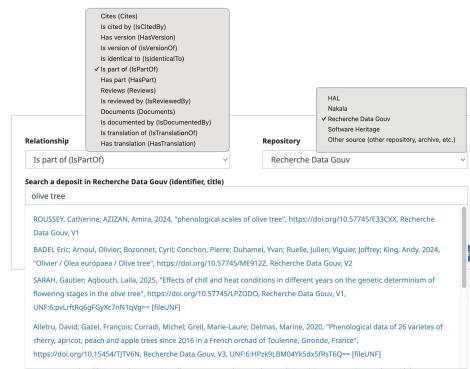
#### Validation des identifiants

Vérifie que les DOI ou autres identifiants persistants fournis sont valides

#### 6 Récupération des citations

Récupère automatiquement la citation de l'ensemble de données à partir de :

- Recherche Data Gouv
- Nakala
- Tout dépôt avec un DOI (via citation.doi.org)





### Visibilité et affichage des ressources associées

#### O Widget dédié sur la page d'accueil du dépôt

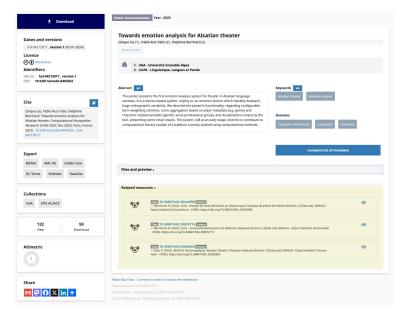
- Affichage et prévisualisation des données ou des logiciels associés
- Améliore la visibilité et l'accessibilité des résultats de recherche liés

#### O Champs spécifiques de l'API

- Nouveaux champs disponibles pour récupérer et exposer les relations
- Facilite l'intégration avec des plateformes et des outils externes

#### O Affichage dans les formats d'exportation

- Les relations sont incluses dans différents formats d'exportation (Dublin Core, TEI, RIOXX).
- Garantit l'interopérabilité et la cohérence des métadonnées entre les systèmes



#### Exemples de **Prévisualisation de données / logiciels**









# Échange automatisé de relations entre les données et les publications

#### O Les liens réciproques en toute simplicité

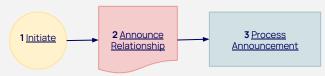
 La relation est partagée avec le référentiel de données afin d'établir un lien bidirectionnel.

#### O Activé par le protocole COAR Notify

- Initiative internationale visant à développer et à promouvoir un protocole normalisé, décentralisé et interopérable pour connecter les résultats de la recherche et les services entre les référentiels et les plateformes.
  - ⇒ https://coar-notify.net/
- Construit sur des normes Web ouvertes
  - Linked Data Notifications (LDN)
  - ActivityStreams



#### "Repository Announces Relationship" Workflow



```
"https://www.w3.org/ns/activitystreams"
 'https://coar-notify.net'
 "id": "https://research-organisation.org",
  "name": "Research Organisation",
  "type": 'Organization'
"context": (
 "id": "https://another-research-organisation.org/repository/datasets/item/201203421/"
  "ietf:cite-as": "https://doi.org/10.5555/999555666",
   "id": "https://another-research-organisation.org/repository/datasets/item/201203421/data_archive.zip"
    "mediaType": "application/zip"
   "type": [
     "Object"
     "sorg:Dataset"
 "type": [
    "sorg:AboutPage
  "as:object": "https://another-research-organisation.org/repository/datasets/item/201203421/",
  "as:relationship": "http://purl.org/vocab/frbr/core#supplement"
 "as:subject": "https://research-organisation.org/repository/item/201203/421/"
 "id": "urn:uuid:74FFB356-0632-44D9-B176-88BDA85758DC"
 "type": "Relationship"
"origin" - {
 "id": "https://research-organisation.org/repository"
 "inbox": "https://research-organisation.org/inbox/"
 "type": 'Service'
 "id": "https://another-research-organisation.org/repository",
  "inbox": "https://another-research-organisation.org/inbox/"
 "type": "Service"
"type": [
  "coar-notify:RelationshipAction"
```



# **État d'avancement :** liens entre les publications, les données et les logiciels

#### O Nouveau système de création de relations

- Opérationnel depuis début 2025
- Permet aux auteurs d'établir des relations explicites entre les publications, les données de recherche et les logiciels/codes sources
  - +33 000 relations déjà créées

#### O Système de transfert de notifications

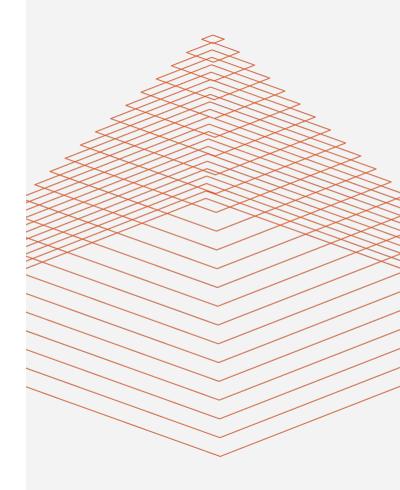
- Actuellement en cours de test
- Construit en utilisant le protocole COAR Notify pour l'interopérabilité et la décentralisation

#### O Intégrations à venir (en cours)

 $^{
m recherche.data.gouv.fr}$  - Référentiel national de données ouvertes pour la recherche en France

Ona - Dépôt de données pour les SHS

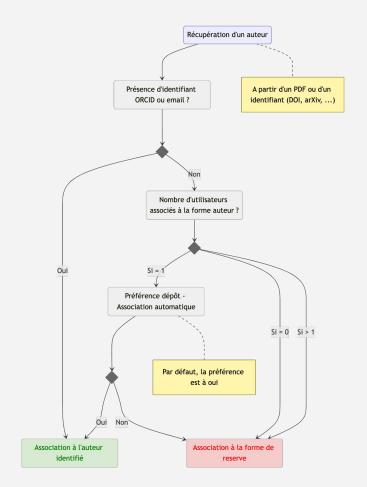






### Autres réalisations

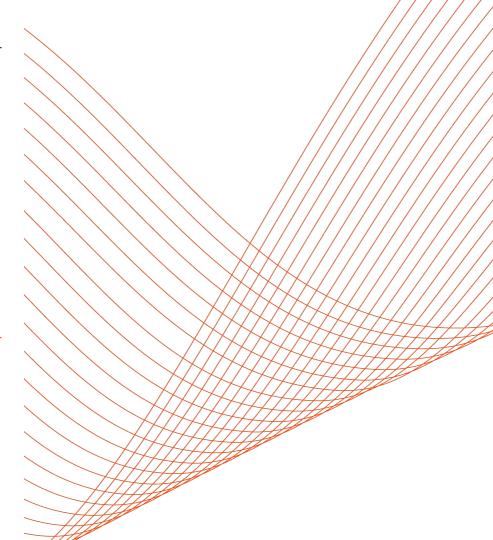
- Evolution service suggestion par rapport retours enquête 2024 (nouveau critère co-auteurs + textes)
   ⇒ 45 000 utilisateurs du service pour ~20 000 dépôts effectués
- Mise en place du nouveau référentiel des financeurs intégrant les donnée de l'ADEME et de l'ANSES. (MAJ des projets européens en cours)
- Amélioration lien HAL → ORCID
  - mapping de la typologie
  - Titre transmis dans la langue du texte intégral
- Intégration PACTOLS
- Amélioration accessibilité
- Evolution système d'identification auteurs (voir logigramme du nouveau fonctionnement)





02

### Développements en cours



# Septembre 2025

### Evolution du service de suggestion

- Actuellement le service de suggestion ne propose que les publications identifiées comme OA avec le fichier récupérable automatiquement

Corpus de **137 644 dépôts potentiels** depuis 2014 65 445 notices seules dans HAL, 72 199 non présents

- Nouveau service permettant de suggérer toutes les publications identifiées dont le texte intégral n'est pas présent dans HAL
  - Avec récupération automatique du fichier (identique service actuel)
  - Avec upload du fichier
  - ⇒ Dépôt simplifié : ajout simple du fichier + renseignement de la licence

Corpus de **253 453 dépôts potentiels** depuis 2014 103 024 notices seules dans HAL, 150 429 non présents



### Evolution du formulaire pour la sélection de la licence

- Actuellement, la licence peut être renseignée :
  - au niveau des fichiers, pour préciser la licence de chaque document
  - au niveau des métadonnées, pour définir une licence par défaut pour l'ensemble des fichiers déposés

#### - Ce qui va changer

- La licence devient une métadonnée obligatoire

Pour les dépôts existants sans licence :

→ L'autorisation donnée lors du dépôt (mise en ligne et diffusion via HAL) vaudra licence implicite.

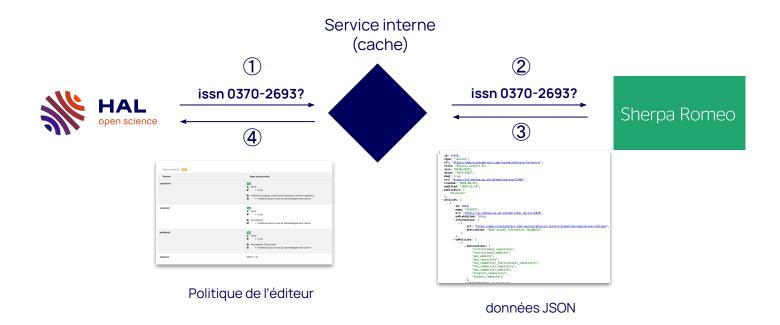
#### Pour les nouveaux dépôts :

- → La licence devra être spécifiée au niveau des fichiers
- → Si elle ne l'est pas, une sélection sera demandée au moment de la soumission (comme c'est déjà le cas pour les dépôts ANR)



### Evolution service de politique des éditeurs

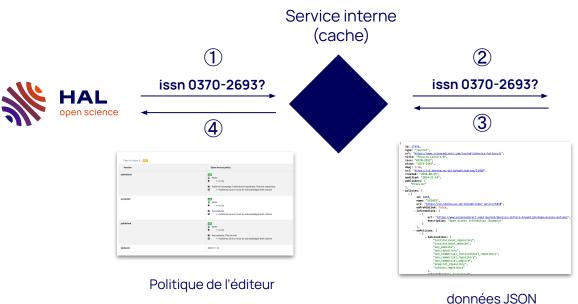
⇒ Service Actuel





### Evolution service de politique des éditeurs

⇒ Futur Service









### **SoFAIR**: Making Software FAIR

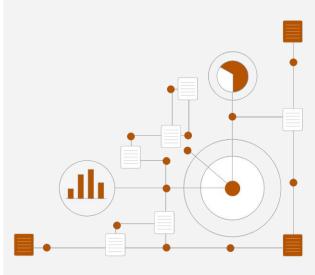
#### Vision à long terme

- Encourager les bonnes pratiques en matière de conservation des logiciels de recherche
- SoFAIR établira un workflow assisté par machine, intégré dans des infrastructures universitaires ouvertes, afin d'aider les chercheurs à identifier, décrire, enregistrer, relier et archiver les logiciels de recherche.

#### **Ambition**

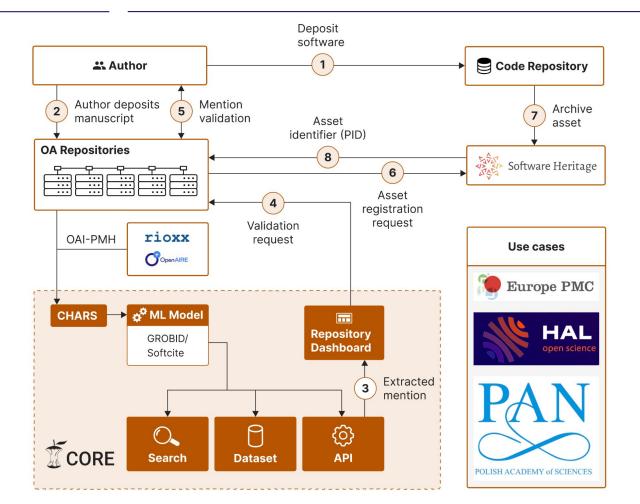
- Développer un workflow assisté par l'apprentissage automatique pour le cycle de vie des logiciels couvrant toutes les étapes depuis 1)
   l'identification des mentions de logiciels dans les publications, 2) leur validation par les auteurs, 3) leur enregistrement avec des PID et leur archivage.
- Intégrer ce workflow dans les infrastructures universitaires, en mettant la solution à la disposition des entrepôts ouverts





https://sofair.org/ https://sofairoa.github.io/documentation/

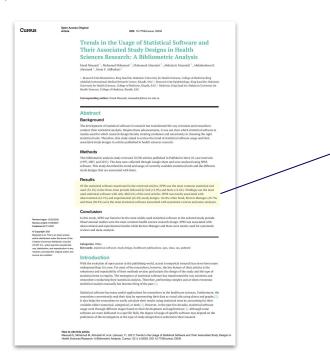






### **Extraction** de mention de logiciels (étape 3)

L'objectif du module Softcite de GROBID est de reconnaître toute mention de logiciel dans les documents textuels, XML et PDF.



Of the statistical software mentioned in the retrieved articles, SPSS was the most common statistical tool used (52.1%) in the three-time periods followed by SAS (12.9%) and Stata (12.6%). WinBugs was the least used statistical software with only 40 (0.6%) of the total articles. SPSS was mostly associated with observational (61.1%) and experimental (65.3%) study designs. On the other hand, Review Manager (43.7%) and Stata (38.3%) were the most statistical software associated with systematic reviews and meta-analyses.



### **Extraction** de mention de logiciels (étape 3)

L'objectif du module Softcite de GROBID est de reconnaître toute mention de logiciel dans les documents textuels, XML et PDF.

Entity type: software Software type: environment Raw name: SPSS Of the statistical software mentioned in the retrieved articles, SPSS was the most common statistical tool used (52.1%) in the three-time periods followed by SAS (12.9%) conf: 0.9026 and Stata (12.6%). WinBugs was the least used statistical software with only 40(0.6%) of the total articles. SPSS was mostly associated with observational (61.1%) and Mention-level: used (1.000) experimental (65.3%) study designs. On the other hand, Review Manager (43.7%) and Document-level: used (1.000) Stata (38.3%) were the most statistical software associated with systematic reviews and SPSS Statistics is a statistical software suite meta-analyses. developed by IBM for data management, advanced analytics, multivariate analysis, business intelligence, criminal investigation. Long produced by SPSS Inc., it was acquired by IBM in 2009. Current versions (post 2015) have the brand name: **IBM SPSS Statistics.** Wikidata statements References: W

**SPSS** 

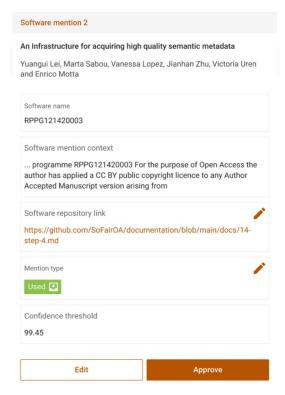


### Validation des mentions de logiciels (étape 5)

- Les auteurs peuvent valider, corriger les mentions logicielles
- Utilisation de notification COAR Notify

#### Après validation

- Suggestion de dépôt du logiciel dans HAL avec transfert dans SWH
- Création de relation entre le logiciel et la publication





03

Déploiement de la délégation de validation technique des dépôts



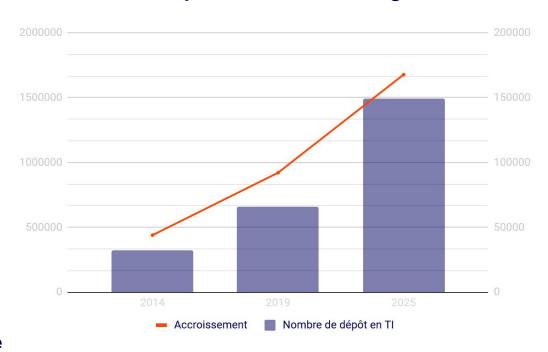
### Des politiques à impact sur le nombre de dépôts en texte intégral

HAL est la voie privilégiée des établissements de l'ESR pour mettre en œuvre l'accès ouvert aux publications.

# **167 751** dépôts en 2024

#### Défis:

- Absorber l'augmentation du volume annuel de dépôts dans HAL
- Garantir un délai rapide de mise en ligne des dépôts avec fichier





### Les étapes du projet

2022 à 2025

#### 1. Clarifier le vocabulaire

Validation = Mise en ligne du dépôt avec fichier et/ou annexe

Facultativement Curation des données.

### 2. Délimiter le périmètre de prise en charge

La validation est liée à l'utilisation d'un portail

### 3. Mettre en place une coordination interne au CCSD

Pour organiser et mener le déploiement national de la prise en charge de la validation technique



#### Axes actuels de travail

• Incitation : organisation d'actions destinées à informer les établissements [au long cours]

- Rendez-vous préalables et individuels avec le CCSD
- 4 ateliers de retours d'expériences en 2024-2025 : ASNR (ex IRSN), Sorbonne Université, ISARA et Université Gustave Eiffel (septembre)



#### Axes actuels de travail

- Incitation
- Formation : consolidation des actions de formation adaptées aux besoins des établissements et de leurs personnels *[en cours]*

- 65 personnels formés en présentiel lors de 7 sessions en 2024
- Ouverture du chantier d'hybridation de la formation (auto-formation, visio, présentiel)



#### Axes actuels de travail

- Incitation | Formation
- Recommandations: élaboration de consignes et bonnes pratiques pour garantir aux déposants et aux valideurs une pratique homogène et cohérente [perspective fin 2025].

- En interne : identification de situations de validation basées sur la prévalence des types de documents et des problématiques.
- Expertise du réseau : les consignes et recommandations de Sorbonne Université comme source complémentaire



#### Axes actuels de travail

- Incitation | Formation | Recommandations
- Réseau et partenariat : mise en place d'un réseau national de valideurs techniques animé par le CCSD [perspective 2026]

+ 16



**42** 

**152** 

portails institutionnels en 2024 - 2025 portails institutionnels membres du réseau portails institutionnels

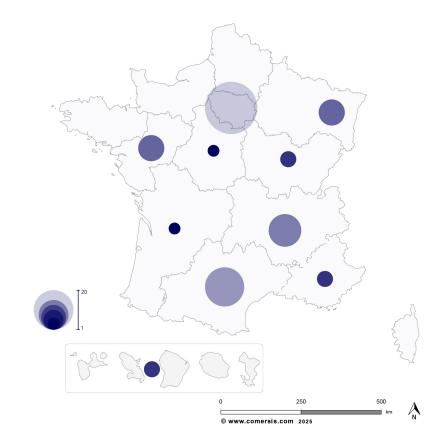


### Comment s'organise-t-elle aujourd'hui?

### CCSD...

## 42 établissements acteurs de la validation

(INIST, INRIA, INRAE, Sorbonne U, Nantes Université, Université de Lorraine, AMU, ASNR, ISARA, UJM, etc.)





# Quelles informations propose le CCSD pour aider à la prise de décision ?

📊 Éléments clés pour **définir le périmètre de la validation** 

### Données chiffrées de volumétrie

• corpus partagé vs corpus exclusif de dépôts à traiter

### Courbes de tendance d'activités

- Identification des pics annuels (saisonnalité)
- Identification des pics hebdomadaires (jours de forte activité)
- **Estimation du temps** à dédier



### Prêt à vous lancer? 🚀

#### ∨os contacts

- Sébastien Mazzarese : sebastien.mazzarese@ccsd.cnrs.fr
- Hélène Bégnis : helene.begnis@ccsd.cnrs.fr

### Notre accompagnement

- Rendez-vous individuel préalable avec le CCSD
- Formation par le CCSD

### **X** Vos ressources

- Documentation de HAL → Base de connaissances
- Convention de partenariat HAL 2026 2028 PSO Licences et PSO Stratégie de non cession exclusive des droits
- Les différents quides du CoSo

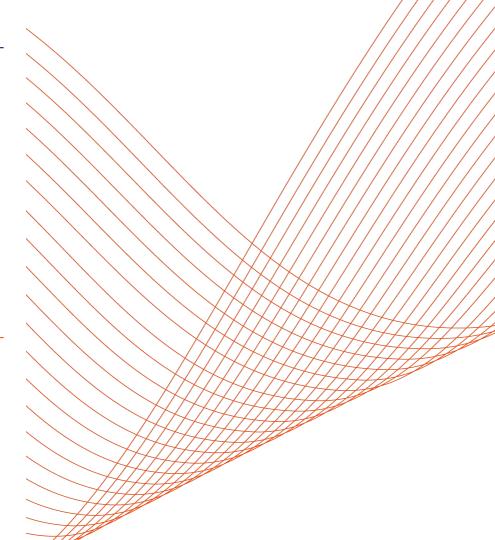
### Motre réseau

- Liste de discussion moderation-hal
- Webinaire(s) REX, atelier(s)



04

Productions de ressources pour accompagner les utilisateurs





Objectif : proposer et valoriser des ressources diversifiées, adaptées aux différents publics et qui permettent de (mieux, bien) utiliser HAL

#### Constats:

- les utilisateurs de HAL (tous profils confondus) ont une connaissance perfectible de notre offre en matière de ressources documentaires et pédagogiques
- les utilisateurs expriment le besoin de disposer de ressources/formations + nombreuses et diversifiées pour être plus efficaces dans leur usage de HAL

(Source : enquête recueil des besoins de formations, mai 2023 → 714 réponses / 57% de chercheurs / 43% de pro de l'IST)



#### Valoriser et faire connaître

# Sur quoi continuons-nous de travailler?

- Refonte de la doc' de HAL
- Les Essentiels
- UsuHAL
- HAL formation



doc.hal.science







#### Refonte de la doc'HAL

#### Objectif:

**®** Mettre en place une **documentation utilisateurs** (aussi) à jour (que possible), accessible depuis le site HAL, **rédigée pour les utilisateurs** dans leur grande **diversité**.

#### Moyens mis en oeuvre :

- Constitution d'un GT CCSD x CasuHAL pour la refonte de l'arborescence
- Création d'une équipe « doc'HAL » au sein du support en responsabilité de sa rédaction et de son suivi
- Mise en place de points de suivi réguliers
- Elaboration de consignes de rédaction pour plus d'efficacité et pérenniser l'activité



doc.hal.science



#### Refonte de la doc'HAL

- **©** Finaliser la migration (été 2025)
- Poursuivre la mise à jour des pages (au long cours)
- Engager la traduction en anglais de la section « Déposer » (T4 2025)
- Poursuivre les efforts concernant l'échange d'information entre le pôle développement et le pôle support (au long cours)



#### Les Essentiels de HAL

### Objectif:

**⊚** Mettre à disposition des **fiches pratiques synthétiques** qui présentent les **bénéfices** de certaines fonctionnalités de HAL d'abord aux chercheur·e·s et ensuite aux utilisateurs professionnels.

#### Mise en oeuvre:

- Elaboration d'une ligne éditoriale
- Identification des thèmes prioritaires
- Recherche de sources scientifiques pour appuyer l'argumentation
- 5 fiches pratiques produites







La boîte à outils UsuHAL

#### Objectif:

© Produire un outil de **diffusion de ressources documentaires et pédagogiques** à l'usage des utilisateurs professionnels et actifs de HAL pour renforcer l'usage de la plateforme.

#### Mise en oeuvre:

- Elaboration d'une ligne éditoriale
- Mise en place d'un comité et d'un planning éditorial
- Articulation avec l'équipe Doc'HAL et l'équipe formations
- 15 numéros envoyés depuis mai 2024 aux listes de diffusion du CCSD







### Deux posters

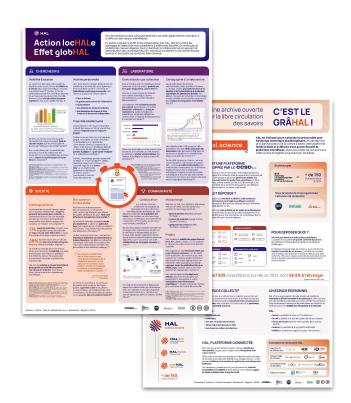
### Objectif:

© Créer des **posters scientifiques** qui présentent HAL du **point de vue des chercheur·e·s** 

#### Mise en oeuvre:

- Atelier de création collaboratif avec les ambassadeur·ice·s HAL
- Respect des standards scientifiques

   (bibliographie, sources, illustrations, indicateurs, formats)
- Diffusion (bientôt)





### Développement de HAL Formation

Un espace pour les ressources et les formations à HAL

**6** Mettre à disposition des utilisateurs de HAL un espace accueillant toutes les **ressources pédagogiques** et les **formations** produites par le CCSD

X Plateforme LMS / Moodle / hébergé et maintenu par un prestataire

Ouverture en novembre 2024

Renforcement de l'équipe d'ingénierie pédagogique

Plan d'action 2024 du CCSD (action 11) - 2025 (action 15)



### Développement de HAL Formation

#### Où en est-on?

### Réalisations

Un accès centralisé aux sessions de formation (53 sessions synchrones depuis 11/2024 + auto-formation)

Un accès facilité aux ressources pédagogiques :

- 7 séries de vidéos tutoriels
- 5 fiches pratiques
- 2 replays de webinaire

### Perspectives

- Formation multimodale à la validation technique
- Ressources pour déposant : dépôt, idHAL, licences





#### Le CCSD recrute!

Pour l'assistance aux utilisateur-ice-s

1 chargé.e de validation et support, formateur·ice HAL

1 chargé.e de validation et support, **référent·e documentation utilisateurs** 

1 chargé.e de validation des dépôts, support HAL et Sciencesconf

### Rejoignez le CCSD!





### Vos prochains rendez-vous

avec le CCSD

[30 juin de 14h à 15h30] Webinaire Parlons Science ouverte "Science ouverte, impact et société : regards croisés de PathOS et COMMONS"

### [Septembre 2025]

- 15 et 18 septembre : Webinaire et retours d'expérience (ISARA et Gustave Eiffel) sur la prise en charge de la validation technique
- 23 septembre : **Webinaire PSO** sur les usages de la science ouverte

[Octobre 2025] Congrès de l'ADBU à Nantes



Merci!

https://hal.science

https://ccsd.cnrs.fr







