



## Lien entre HAL & NAKALA

Articulation entre les publications et les jeux de données.

---

Rencontres Huma-Num 2025

Yannick Barborini

Centre pour la Communication Scientifique Directe (CCSD) / CNRS



## HAL+, une infrastructure de recherche pour la communication scientifique



## EOSC-Pillar : Relations Publications - Données en SHS

Le projet **EOSC Pillar** vise à coordonner les efforts nationaux en matière de science ouverte en Autriche, en Belgique, en France, en Allemagne et en Italie. Il assure leur contribution et leur préparation à la mise en œuvre du European Open Science Cloud (EOSC).

L'initiative propose de partager les meilleures pratiques et susciter des possibilités d'approches interdisciplinaires au sein de l'EOSC.

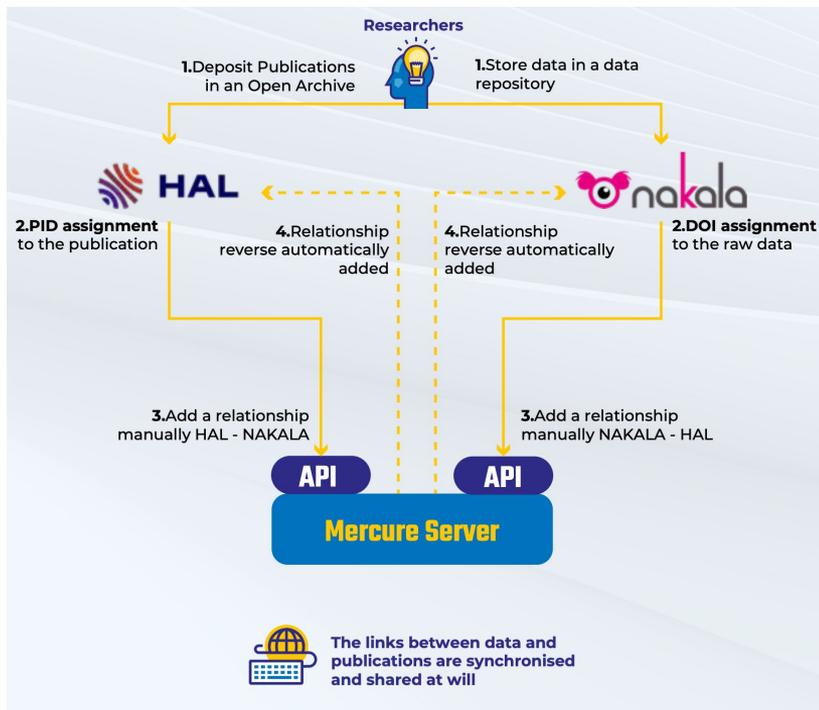
⇒ Task 6.5 : Réalisation d'une Preuve de Concept entre Nakala, l'entrepôt de données en SHS et les publications scientifiques présentes dans HAL



Juillet 2019 - Décembre 2022

Funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 857650

## Réalisation de la Preuve de Concept



- **PID** pour identifier les ressources
- **APIs** pour rechercher des ressources
- Vocabulaire standard pour typer les relations (et publier)  
⇒ **DataCite**
- Un outil de communication pour synchroniser et partager les liens  
⇒ protocole **Mercure**  
(<https://mercure.rocks/>)
- **Citation** de la publication et des données
- **Visualisation**

## HALiance

### Équipements structurants pour la recherche (Programme d'investissements d'avenir 3)

#### Objectifs

- Enrichir l'offre de services à destination de nos utilisateurs et fluidifier leur expérience
- Conforter l'archive HAL comme acteur central de l'écosystème de la recherche
- Aligner HAL de manière exemplaire et durable sur les recommandations et les critères d'excellence de la science ouverte

<i>Référence du projet</i>	 <b>21-ESRE-0047</b>
<i>Etablissement Coordinateur</i>	<b>CNRS Rhône Auvergne</b>
<i>Responsable du projet</i>	<b>Nathalie Fargier</b>
<i>Dates de démarrage et de clôture</i>	<b>2022-01-01 → 2027-06-30</b>
<a href="https://anr.fr/ProjetIA-21-ESRE-0047">https://anr.fr/ProjetIA-21-ESRE-0047</a>	

# HALiance - Workpackages

## WP3 - 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

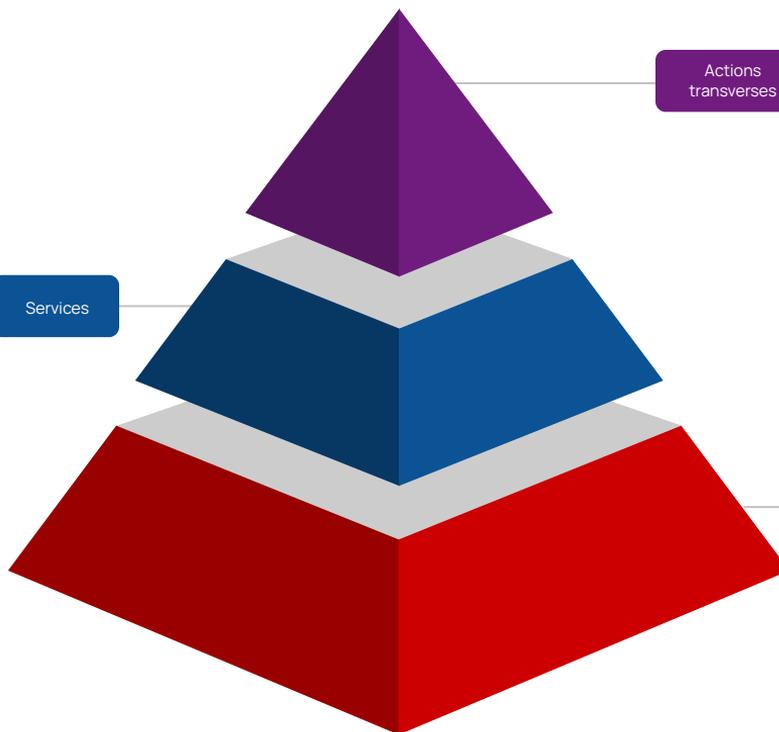
Articuler HAL à des dispositifs de relecture des preprints et rendre visible le cycle de la publication

## 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



## WP8 - 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 HAL

# Lien Publications-Données de la recherche

## WP3 - 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

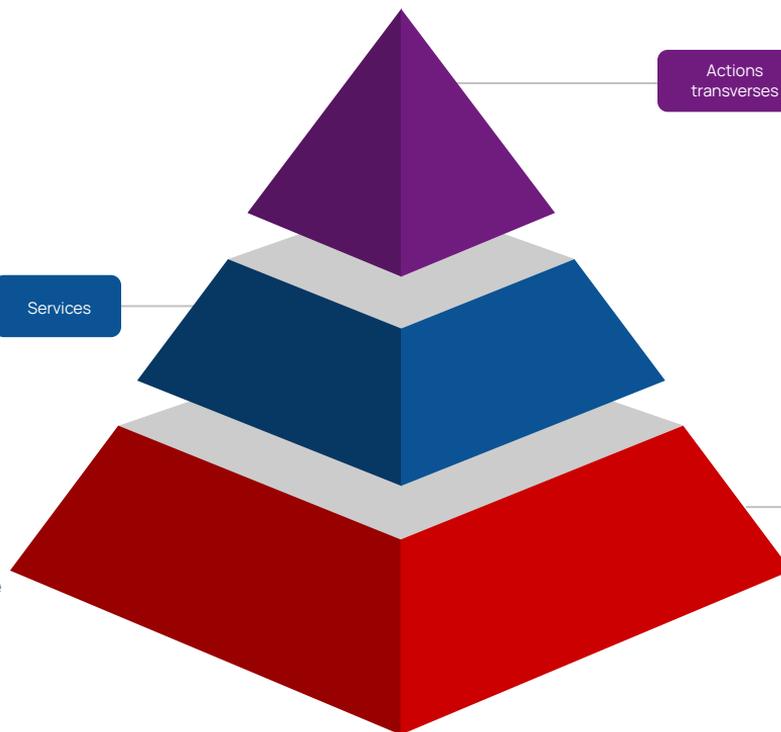
Articuler HAL à des dispositifs de relecture des preprints et rendre visible le cycle de la publication

## 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



Actions  
transverses

## WP8 - 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

Services

Socle

## WP1 - Infrastructure matérielle

Mettre à niveau et sécuriser l'infrastructure physique de HAL

## WP2 - Infrastructure logicielle

Migrer, sécuriser et ouvrir le code logiciel HAL

## 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 :

- HAL
- Recherche Data Gouv
- Nakala

### 3 Recherche assistée

Aide à localiser les ensembles de données dans les dépôts sélectionnés

### 4 Validation des identifiants

Vérifie que les DOI ou autres identifiants persistants fournis sont valides

### 5 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)

**Relationship**

- Cites (Cites)
- Is cited by (IsCitedBy)
- Has version (HasVersion)
- Is version of (IsVersionOf)
- Is identical to (IsIdenticalTo)
- Is part of (IsPartOf)
- Has part (HasPart)
- Reviews (Reviews)
- Is reviewed by (IsReviewedBy)
- Documents (Documents)
- Is documented by (IsDocumentedBy)
- Is translation of (IsTranslationOf)
- Has translation (HasTranslation)

**Repository**

- HAL
- Nakala
- Recherche Data Gouv
- Software Heritage
- Other source (other repository, archive, etc.)

Is part of (IsPartOf) ▼

Recherche Data Gouv ▼

**Search a deposit in Recherche Data Gouv (identifier, title)**

olive tree

ROUSSEY, Catherine; AZIZAN, Amira, 2024, "phenological scales of olive tree", <https://doi.org/10.57745/E33CXX>, Recherche Data Gouv, V1

BADEL Eric; Arnoul, Olivier; Bozonnet, Cyril; Conchon, Pierre; Duhamel, Yvan; Ruelle, Julien; Viguier, Joffrey; King, Andy, 2024, "Olivier / Olea europaea / Olive tree", <https://doi.org/10.57745/ME912Z>, Recherche Data Gouv, V2

SARAH, Gautier; Aqbouch, Laila, 2025, "Effects of chill and heat conditions in different years on the genetic determinism of flowering stages in the olive tree", <https://doi.org/10.57745/LPZODO>, Recherche Data Gouv, V1, UNF:6:pvLrftRq6gFgyXc7nN1qVg== [fileUNF]

Alletru, David; Gazel, François; Corradi, Michel; Greil, Marie-Laure; Delmas, Marine, 2020, "Phenological data of 26 varieties of cherry, apricot, peach and apple trees since 2016 in a French orchard of Toulence, Gironde, France", <https://doi.org/10.15454/TJTV6N>, Recherche Data Gouv, V3, UNF:6:HPzk9LBM04Yk5dx5fRsT6Q== [fileUNF]

# Visibilité et affichage des ressources associées

## 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

## Champs spécifiques de l'API

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

## 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

The screenshot shows a HAL record page for a poster. It includes a 'Download' button, 'Dates and versions' (hal-04213017, version 1 05-01-2024), 'Licence' (CC BY Attribution), and 'Identifiers' (HAL ID: hal-04213017, DOI: 10.5281/zenodo.8404232). The 'Cite' section provides the full citation for the poster. 'Export' options include BibTex, XML/TEI, Dublin Core, DC Terms, EndNote, and DataCite. 'Collections' include UGA and STE-ALSACE. 'Collections' statistics show 122 views and 59 downloads. 'Altmetric' shows a question mark. 'Share' options include M, @, F, X, LinkedIn, and +. The 'Poster Communications' section shows the year 2023 and the title 'Towards emotion analysis for Alsatian theater'. The 'Abstract' section contains the poster's text. 'Keywords' include Alsatian Theater and Emotion analysis. 'Domains' include Computer Science [cs], Linguistics, and Literature. A 'Complete list of metadata' button is visible. 'Files and preview' and 'Related resources' sections are also present.

## Exemples de Prévisualisation de données / logiciels

H-ID	Periodicité	création	Discipline(s)	éditoriales	bibliographiques	contribution	intégraux	clés	clés	Résumés
308	Annuelle	2011-2020	Art Littérature Linguistique	Yes	Yes	Ponctuellement	French (FR), French (FR)	Yes	Yes	Yes
340	Semestrielle	2011-2020	Droit	Yes	Yes	Jamais	French (FR), French (FR)	Yes	Yes	Yes
367	Semestrielle	2011-2020	Information et communication	Yes	Yes	Pour tous les numéros.	French (FR), French (FR)	Yes	Yes	Yes

The screenshot shows a HAL record for software heritage. It includes a 'SoftwareHeritage' section with a 'View in the archive' button. Below this is a table with columns for 'File', 'Mode', and 'Size'. The table lists files like 'dot', 'dot', 'Integration\_tents', 'ps\_dot', 'scripts', and 'glsigne' with their respective sizes.

## Échange automatisé de relations entre les données et les publications

### 0 Les liens réciproques en toute simplicité

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

### 0 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



```

{
  "@context": [
    "https://www.w3.org/ns/activitystreams",
    "https://coar-notify.net"
  ],
  "actor": {
    "id": "https://research-organisation.org",
    "name": "Research Organisation",
    "type": "Organization"
  },
  "context": {
    "id": "https://another-research-organisation.org/repository/datasets/item/201203421/",
    "set:item": "https://doi.org/10.5555/99955666",
    "set:item": {
      "id": "https://another-research-organisation.org/repository/datasets/item/201203421/data.archive.zip",
      "mediaType": "application/zip",
      "type": [
        "Object",
        "org:Dataset"
      ]
    }
  },
  "type": [
    "Page",
    "org:AboutPage"
  ]
},
{
  "id": "urn:uuid:94ecae35-dcfd-4182-8558-22c7164fe29f",
  "object": {
    "as:object": "https://another-research-organisation.org/repository/datasets/item/201203421/",
    "as:relationship": "https://purl.org/vocab/frbr/core#supplement",
    "as:subject": "https://research-organisation.org/repository/item/201203421/",
    "id": "urn:uuid:74FFB356-6632-44D9-B176-88B0A85758DC",
    "type": "Relationship"
  },
  "origin": {
    "id": "https://research-organisation.org/repository",
    "inbox": "https://research-organisation.org/inbox/",
    "type": "Service"
  },
  "target": {
    "id": "https://another-research-organisation.org/repository",
    "inbox": "https://another-research-organisation.org/inbox/",
    "type": "Service"
  },
  "type": [
    "Announce",
    "coar-notify:RelationshipAction"
  ]
}

```

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

### 0 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**

### 0 Système de transfert de notifications

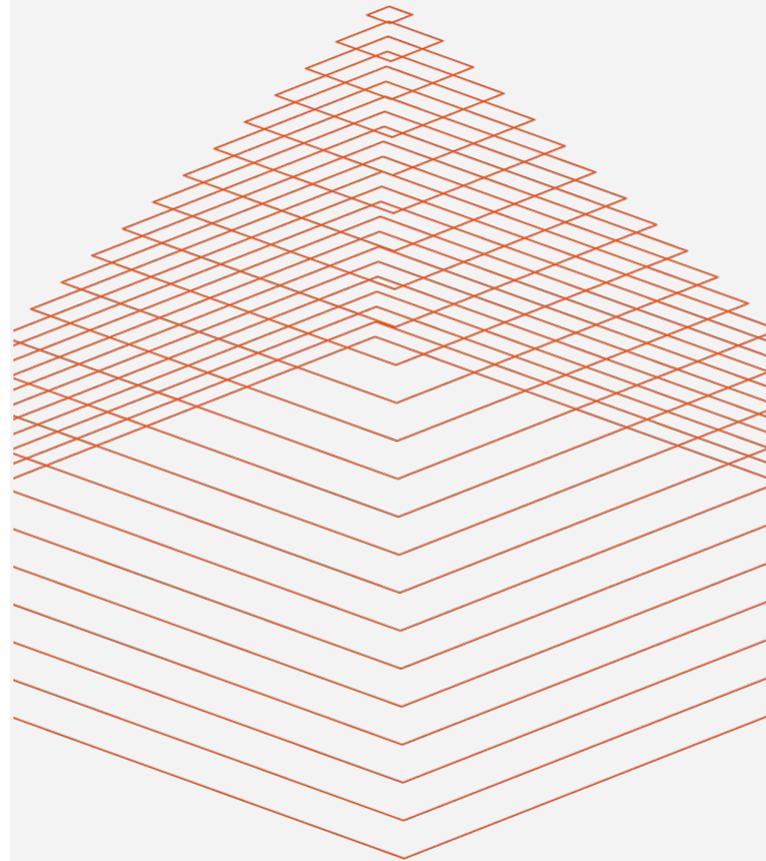
- Actuellement en cours de test
- Construit en utilisant le protocole COAR Notify pour l'interopérabilité et la décentralisation

### 0 Intégrations à venir (en cours)

recherche.[data.gouv.fr](https://data.gouv.fr) - Référentiel national de données ouvertes pour la recherche en France

 nakala - Dépôt de données pour les SHS

 Software Heritage - Archives universelles de logiciels



## Prochaines étapes

### ○ Développer l'interopérabilité

- Identifier et collaborer avec de nouveaux référentiels intéressés par l'échange de relations via le protocole COAR Notify.  
⇒ Vers une adoption plus large à travers les disciplines et les plateformes

### ○ Tirer parti de l'apprentissage automatique pour la découverte

- Appliquer des outils d'apprentissage automatique pour détecter les mentions de données et de logiciels dans les publications (**GROBID**, **SoftCite**, **DataStet**).
- Proposer une validation par l'auteur de ces relations détectées automatiquement



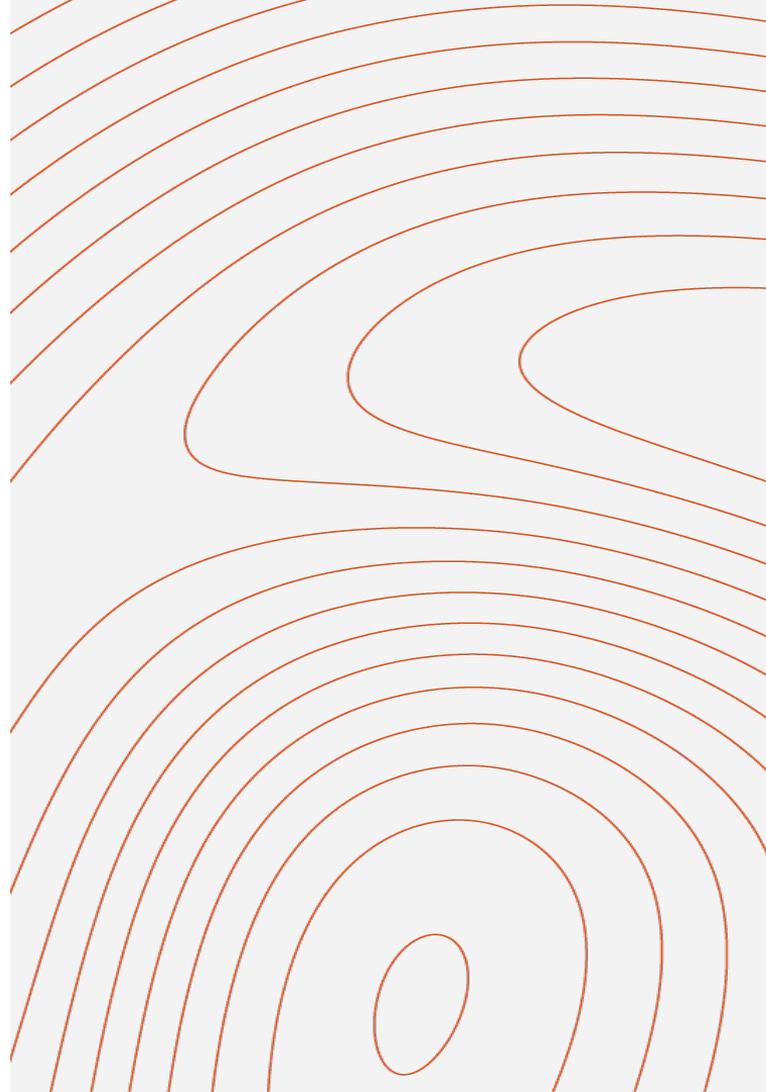
Making Software FAIR: un flux de travail assisté par machine pour le cycle de vie des logiciels de recherche

### ○ Enrichir HAL avec des services externes

- Intégrer des services externes pour compléter les métadonnées de HAL avec des relations préexistantes



- Accepte des métadonnées validées sur les liens entre les publications et les données
- Améliore la découvrabilité et l'exhaustivité des métadonnées





**HAL**  
open science

**Merci !**

**Yannick Barborini**

*yannick.barborini@ccsd.cnrs.fr*

  
**MINISTÈRE  
DE L'ENSEIGNEMENT  
SUPÉRIEUR  
ET DE LA RECHERCHE**  
*Liberté  
Égalité  
Fraternité*



*Inria*

**INRAE**