HAL
science ouverte

HAL, une actualité des liens entre publications, données et codes sources de logiciels

12e édition des journées annuelles du réseau Mate-shs

16 mai 2025

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

cnrs

Inria

INRAE

CC BY

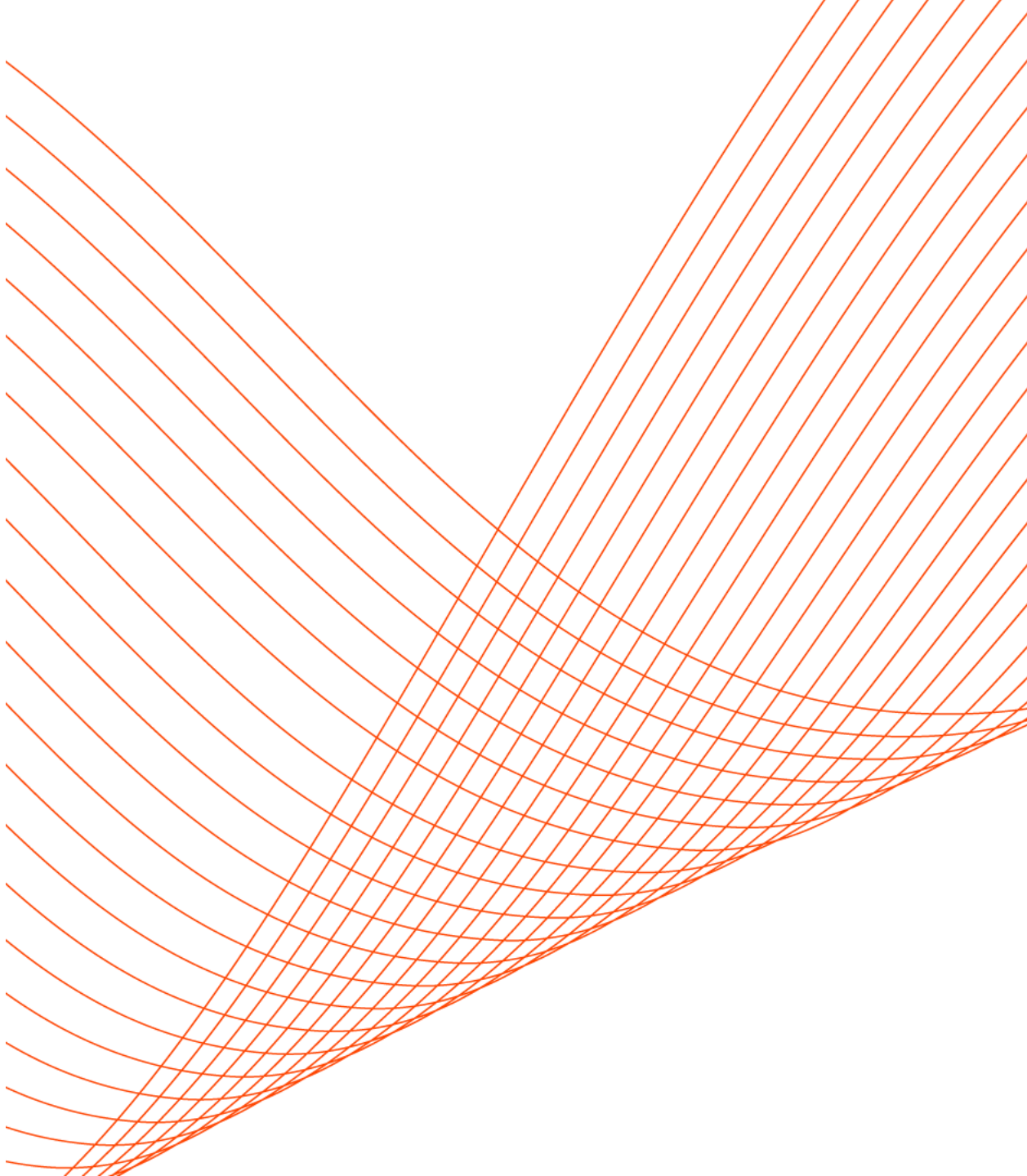
Au programme

HAL, une actualité des liens entre publications, données et codes sources

- Identifier les enjeux
- Comprendre les méthodes
- Présenter des résultats et les travaux en cours

01

Eléments de contexte



HAL
science ouverte

CCSD
Centre pour la Communication Scientifique Directe

EPI sciences
revues en accès libre

sciencesCONF
gestion de conférences

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

cnrs

Inria
inventors for the digital world

INRAE

HAL en bref

- Archive ouverte nationale pluridisciplinaire de la recherche française
  - Archive ouverte = Entrepôt numérique contenant des documents issus de la recherche scientifique
  - Auto-archivage (ou dépôt) par l'auteur
- Au service des chercheurs affiliés à des institutions académiques
- Objectif : partager en accès libre et gratuit les résultats de recherche
- Acteur majeur des deux Plans Nationaux pour la Science Ouverte

Deuxième plan national pour la science ouverte

- Généraliser l'accès ouvert aux publications
- Structurer, partager et ouvrir les données de la recherche
- Ouvrir et promouvoir les codes sources produits par la recherche
- Transformer les pratiques pour faire de la science ouverte le principe par défaut

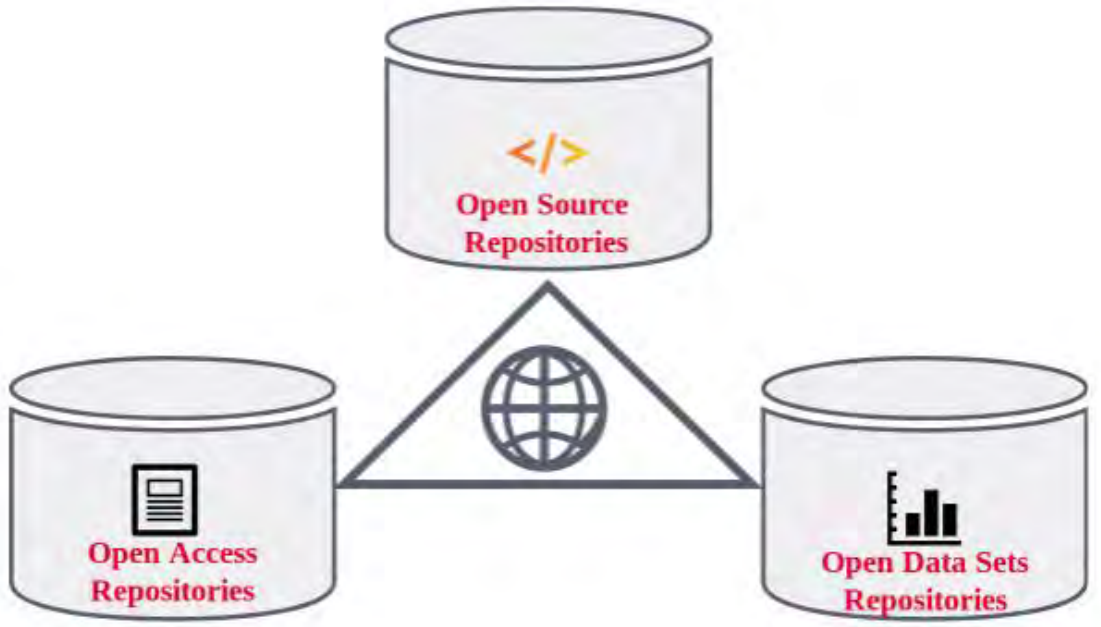
Construire un écosystème reliant codes, données et publications

⇒ Afin de créer un écosystème reliant codes, données et publications, la collaboration sera renforcée entre l'archive ouverte nationale HAL, la plateforme fédérée des données de la recherche Recherche Data Gouv, le secteur de l'édition scientifique et Software Heritage

Enjeu scientifique

Dans la quête pour la reproductibilité des résultats scientifiques et la transmission des connaissances scientifiques, trois piliers principaux :

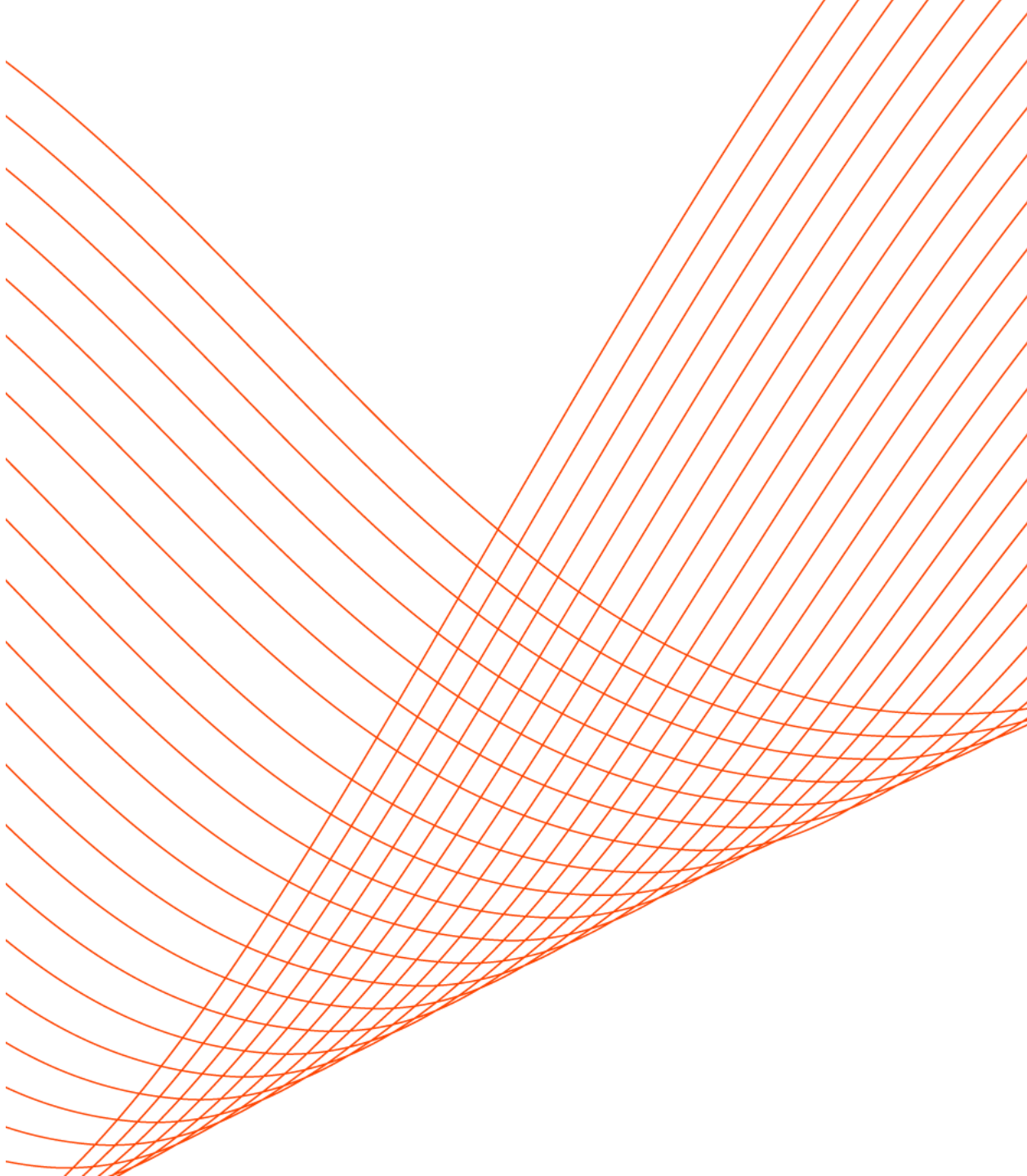
- des articles scientifiques décrivant les résultats établis par les chercheur·es
- les ensembles de données
- les logiciels utilisés ou produits



Source : Roberto Di Cosmo, Morane Gruenpeter, Bruno P Marmol, Alain Monteil, Laurent Romary, et al.. Curated Archiving of Research Software Artifacts : lessons learned from the French open archive (HAL). IDCC 2020 - International Digital Curation Conference, Feb 2020, Dublin, Ireland. ⟨10.2218/ijdc.v15i1.698⟩. https://inria.hal.science/hal-02475835, consulté le 25/05/2023

02

Le dépôt de données dans HAL



Répartition des documents en texte intégral déposés en 2024

167 751

documents en texte intégral déposés entre janvier et novembre 2024

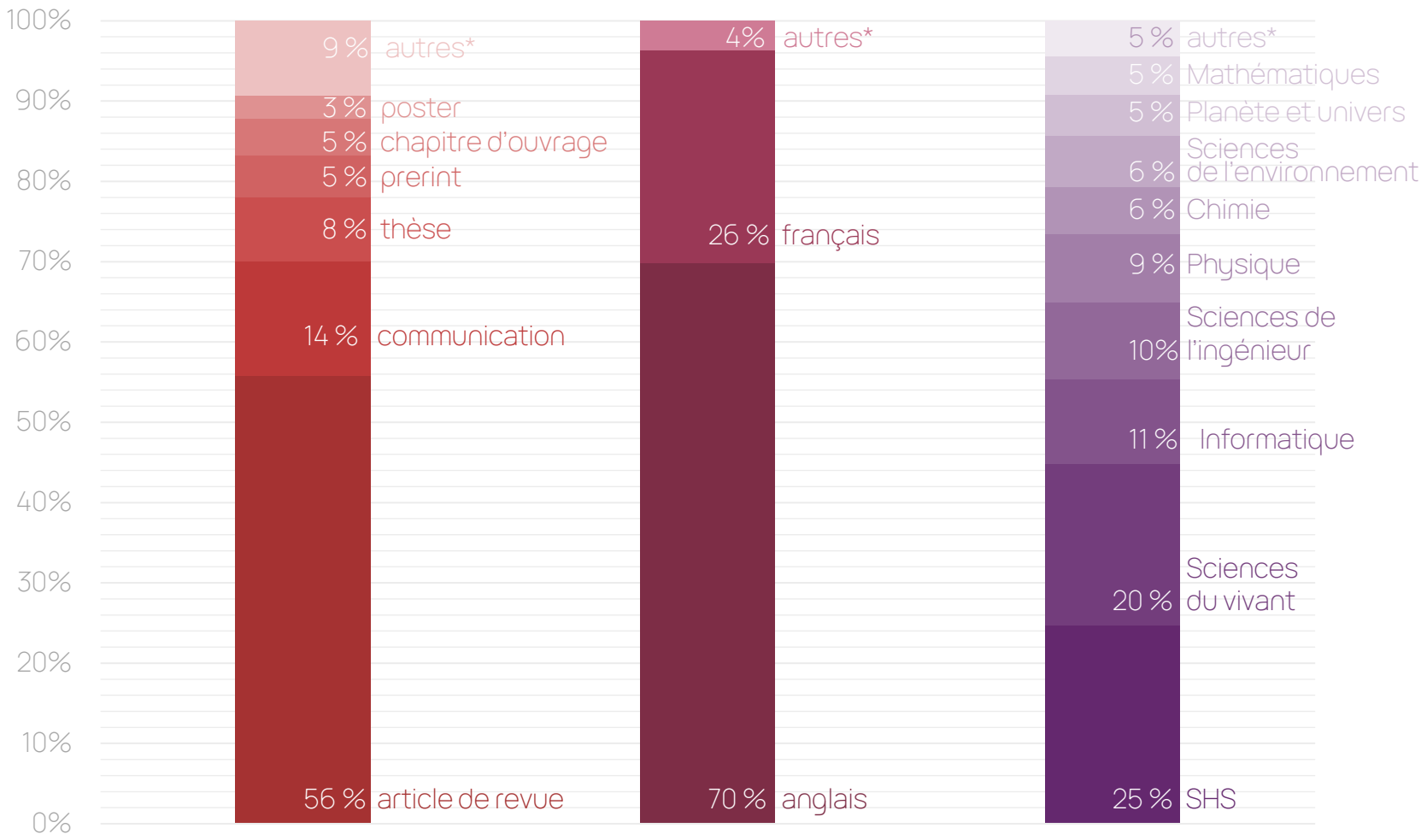
1 496 055

documents en texte intégral dans HAL

4 018 097

dépôts dans HAL

* Données au 9 mai 2025



Type de document

* Rapport, image, article de blog, HDR, ouvrage, vidéo, cours, acte de congrès, logiciel, dossier de revue, notice de dictionnaire.

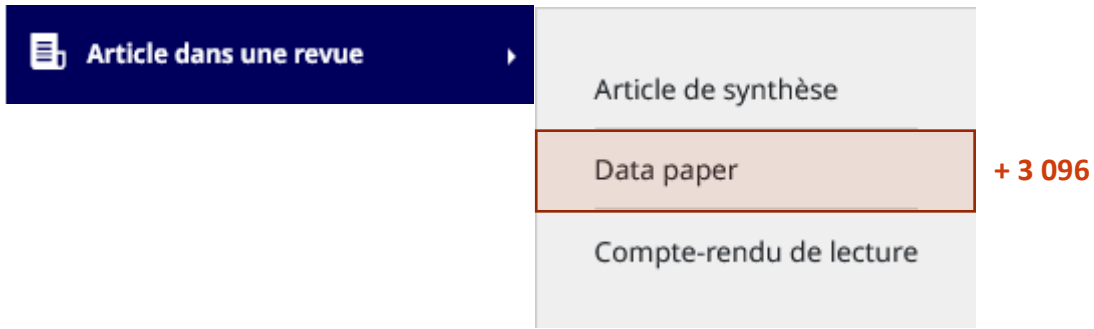
Langue du document

* Espagnol, italien, portugais, allemand, autres langues.

Discipline du document

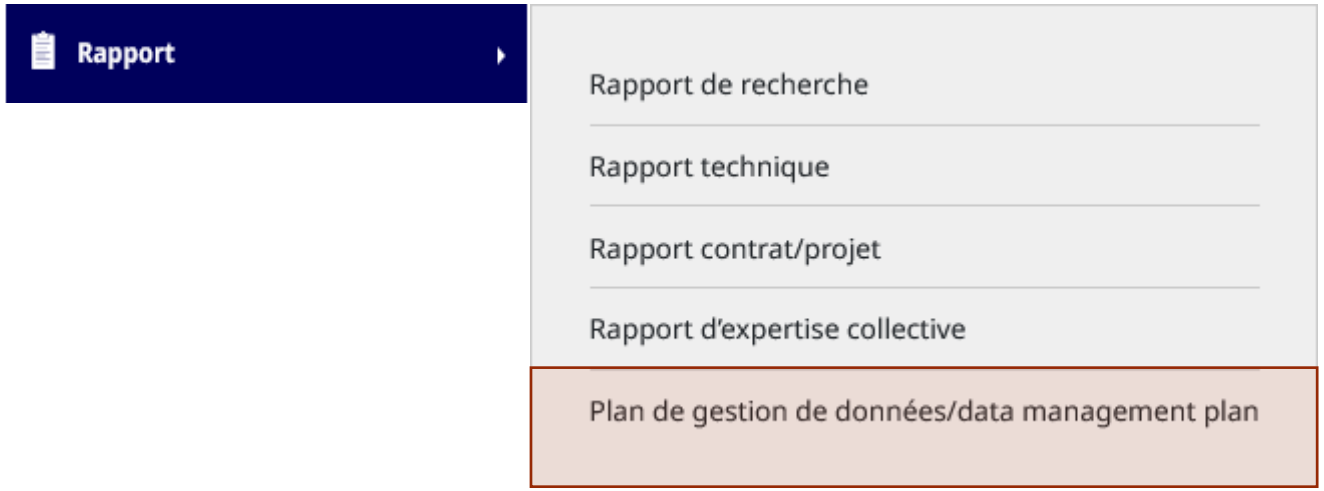
* Sciences cognitives, statistiques, économie et finance quantitative, science non linéaire.

Types de documents spécifiques



+ 3 096

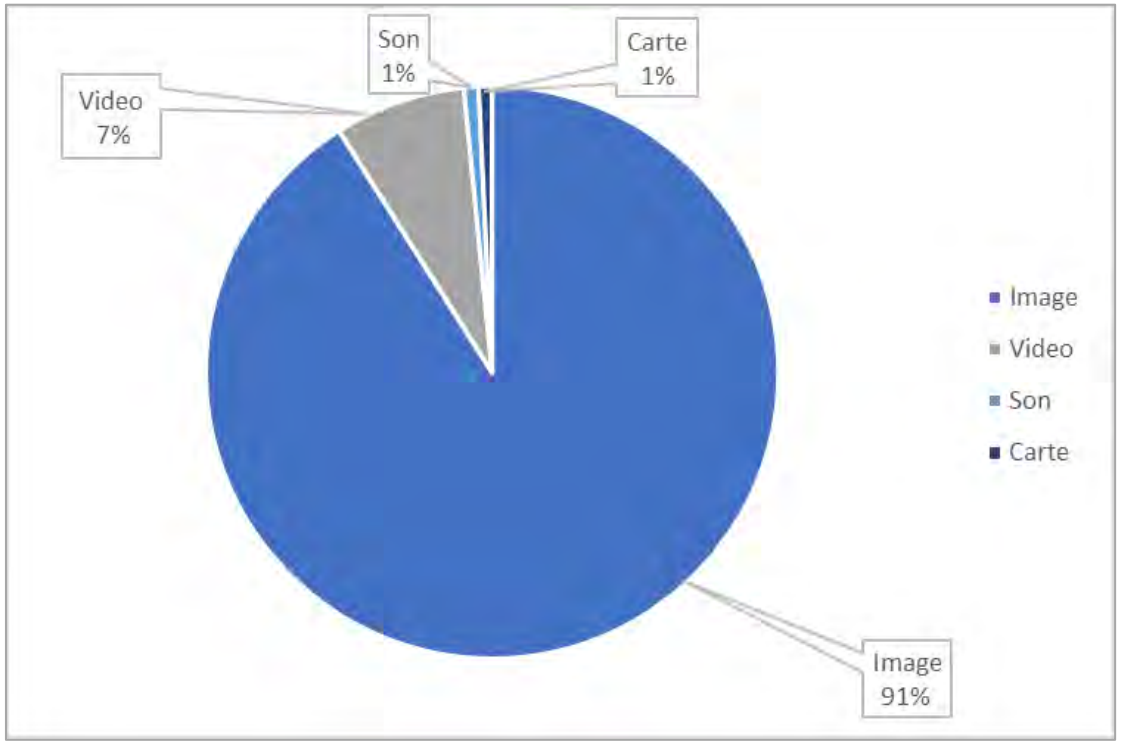
+ 1 797



+ 121

* Données au 9 mai 2025

Dépôt de données multimédia



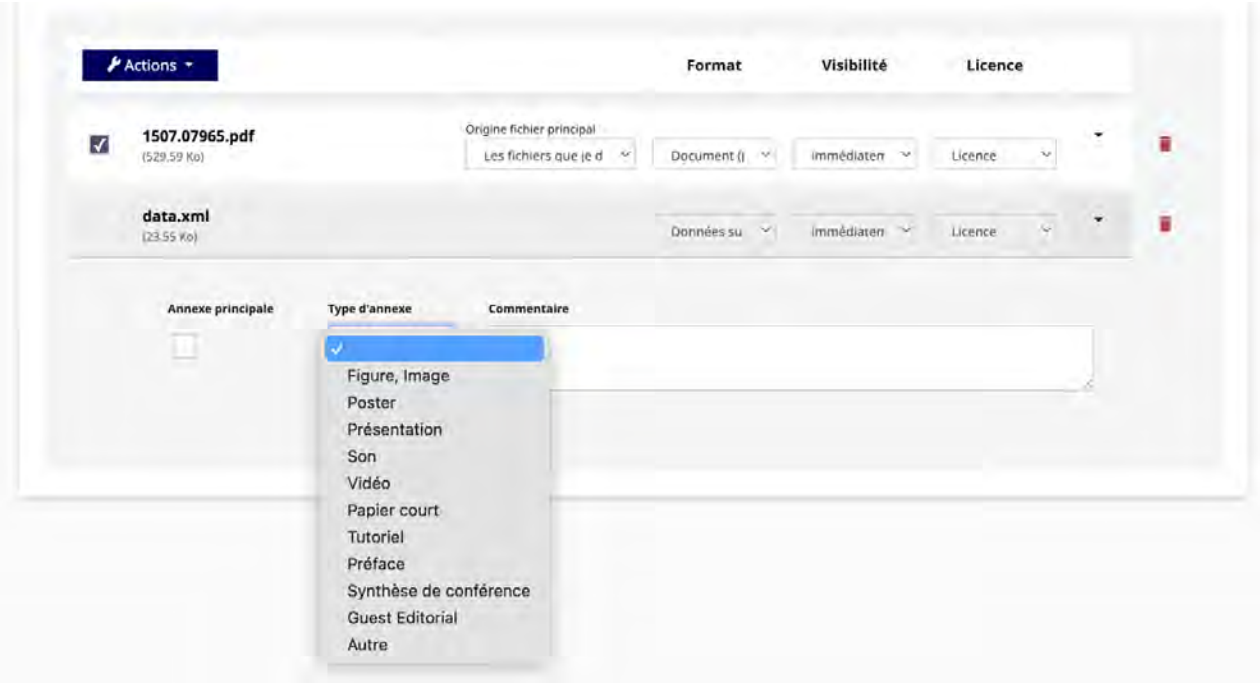
* Données au 9 mai 2025

Dépôt de données en annexes de publications

- Quoi ?
  - Figures, images, sons, videos
  - Fichiers de données
  - Autre
- Quand ?
  - Au moment du dépôt
    - ✓ Fichier principal + Fichiers annexes
    - ✓ Fichiers annexes uniquement
  - Une fois le dépôt en ligne
    - ✓ Fonctionnalité « Ajouter une annexe »
- Fonctionnalités de description des fichiers annexes :
  - Visibilité (embargo)
  - Licence de réutilisation
  - Type d'annexe
  - Commentaire

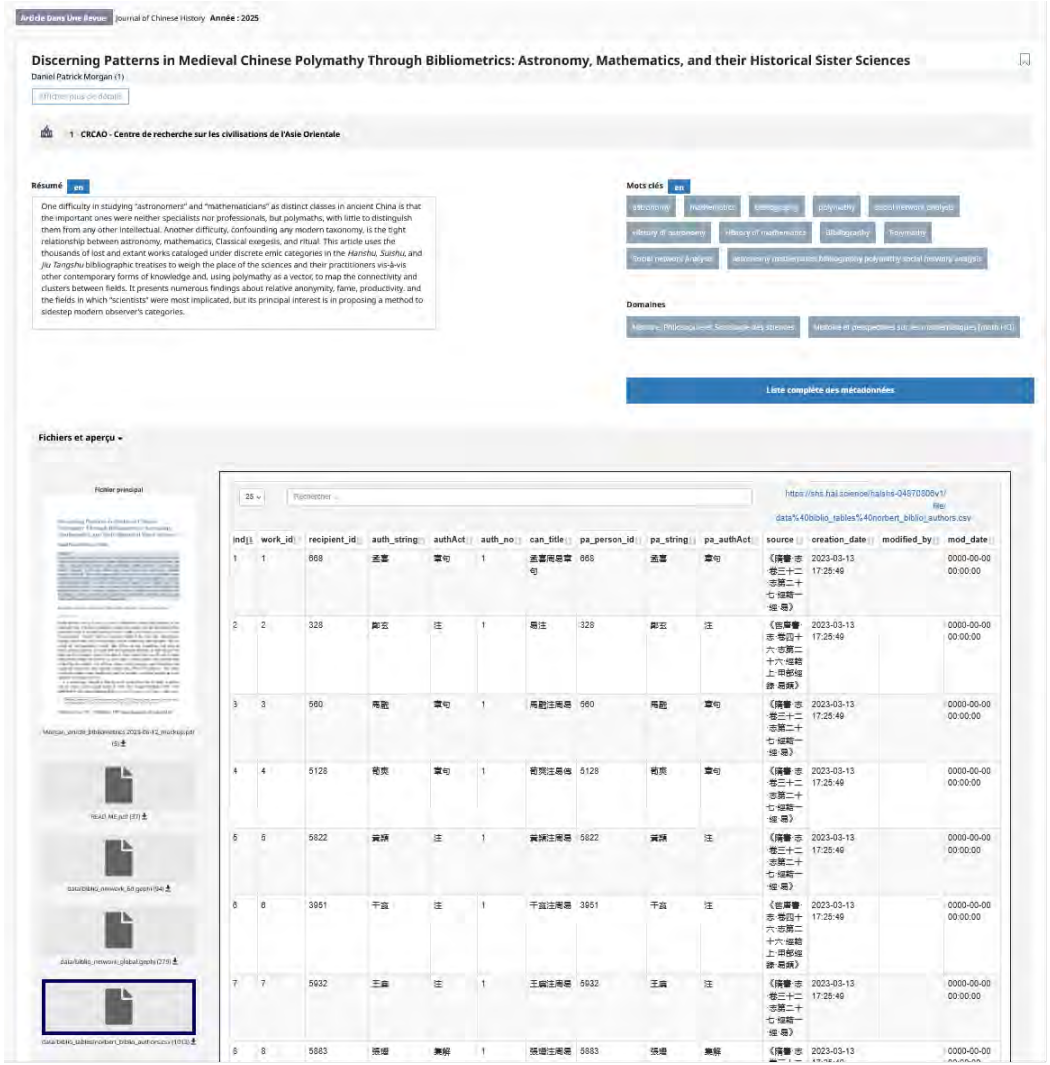
128 105
données supplémentaires
pour **72 833** dépôts

(+ de 50% des fichiers au format PDF)



Prévisualiser des données en tant que fichiers annexes

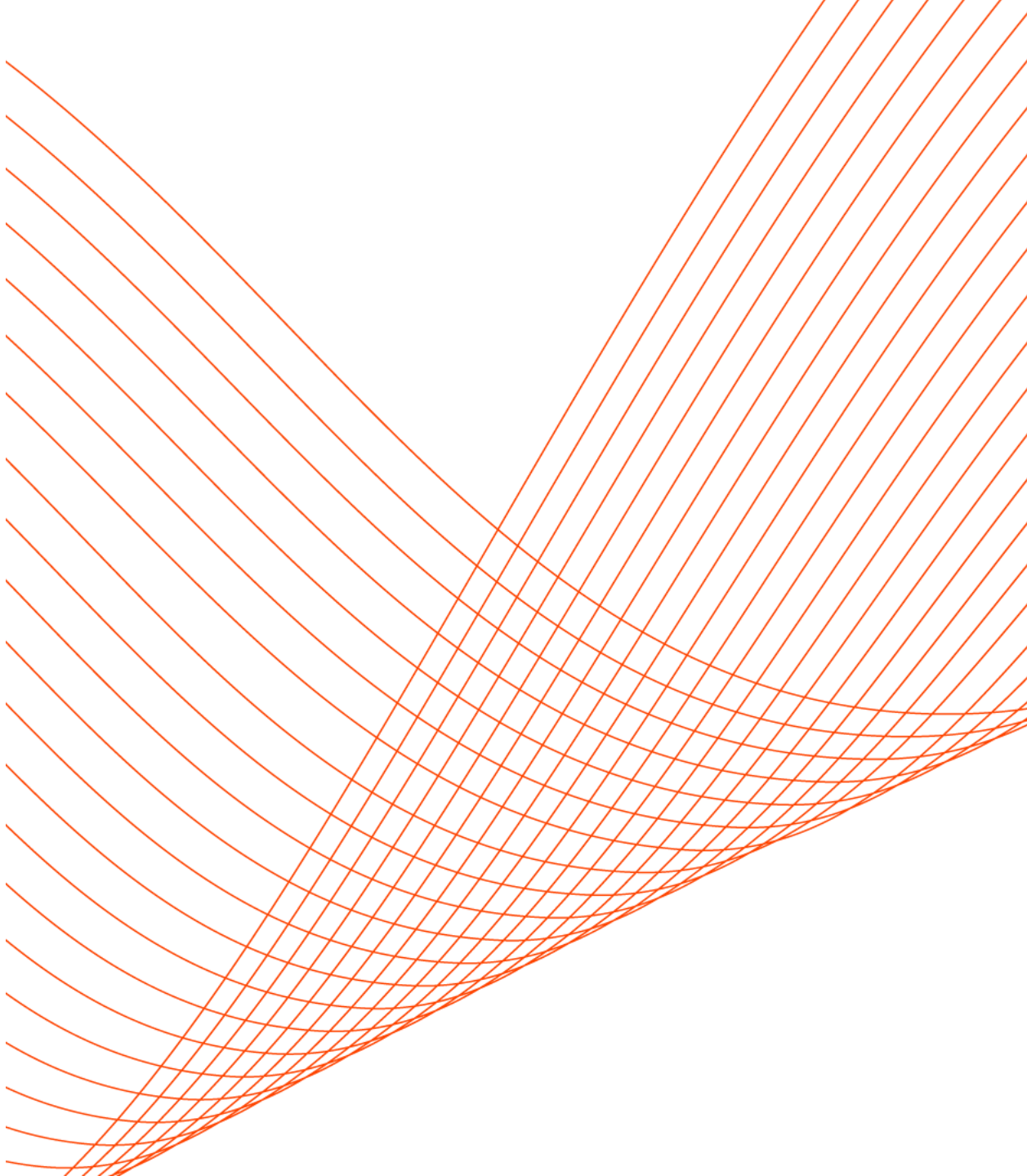
- Accès aux données supplémentaires sur la page publique d'un dépôt
- Prévisualisation possible (pour les formats pdf, csv, text, mp4, img, zip)
- Pas de citation disponible pour les fichiers annexes



Vue du dépôt : https://hal.science/halshs-04870806

03

Articuler publications, données et codes sources de logiciels : un état des lieux

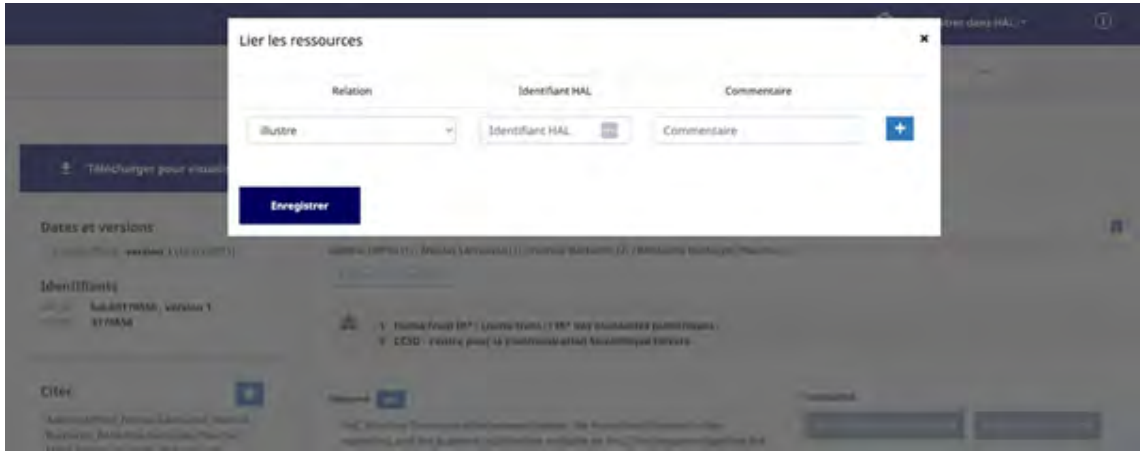


Avant la refonte, plusieurs services co-existent

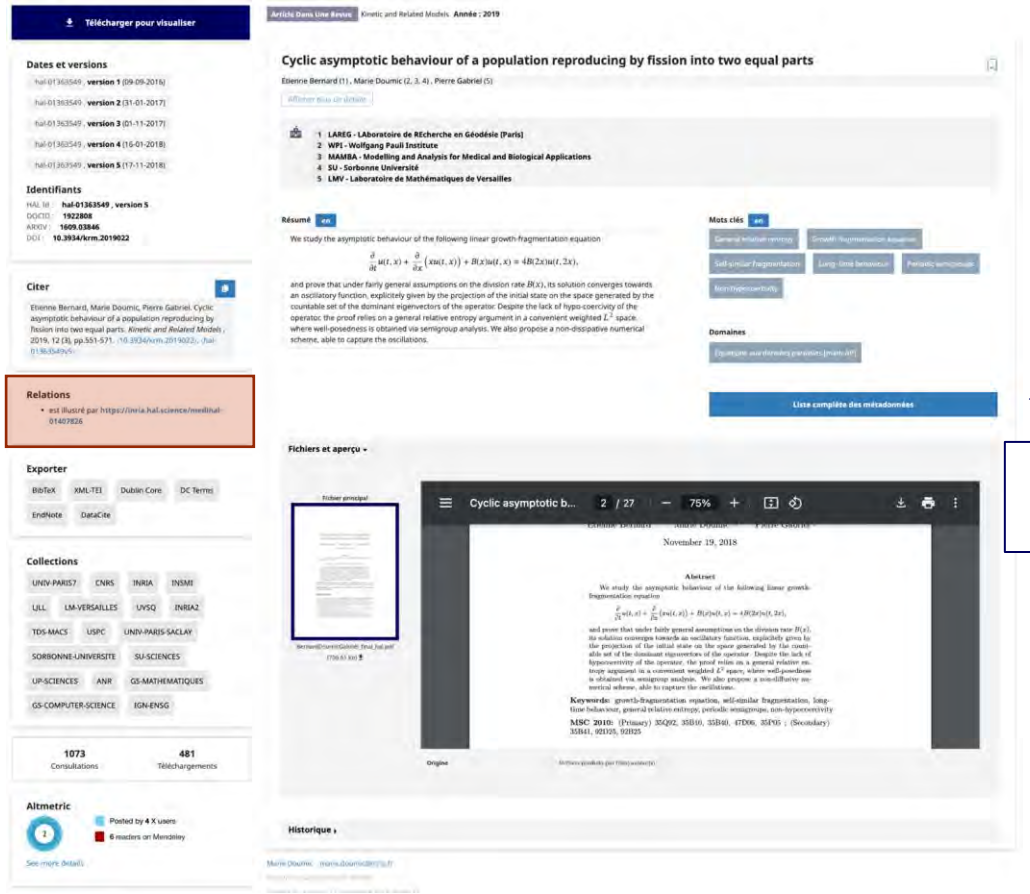
- Dans le dépôt HAL, ajout du DOI d'un jeu de données associé à la publication et archivé dans un entrepôt extérieur à HAL (= > création d'un lien vers le jeu de données)

Type de dépôt	Publication	Donnée (Image, vidéo, son)	Logiciel
Métadonnées	Données associées (DOI) Logiciels associés (SWHID)	Publications associées (HAL, DOI, ...) Logiciels associés (SWHID)	Publications associées (HAL, DOI, ...) Données associées (DOI)

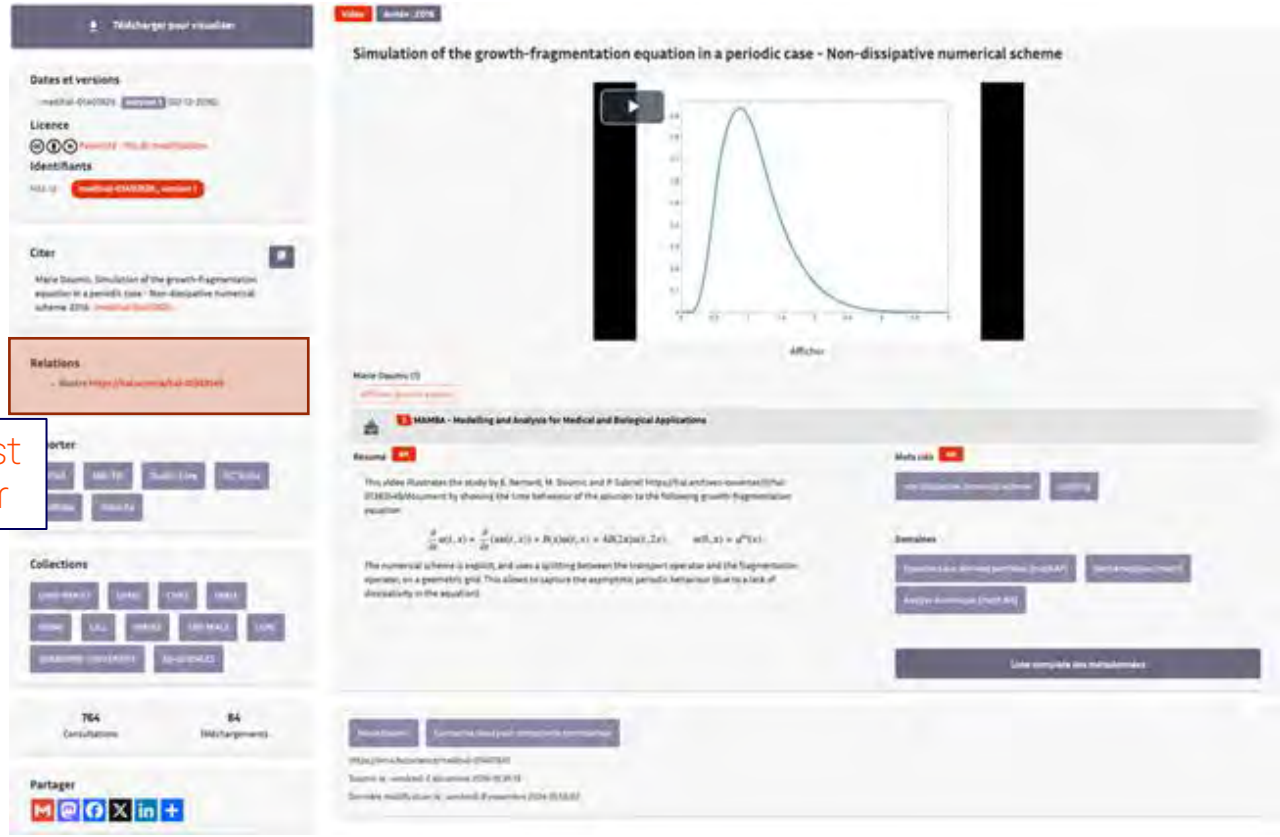
- Création de relations entre des dépôts HAL
  - Fonctionnalité « Lier les ressources »
    - Ex : Une publication déposée dans HAL mise en relation avec une image déposée dans MédiHAL
  - Typage de la relation (Liste fermée : *hasPart – isPartOf, illustrate – isIllustratedBy, etc.)*
  - Création automatique des relations réciproques



Système de relations entre dépôts HAL



Illustre - Est illustré par



Ancienne vue du dépôt : https://hal.science/hal-01363549

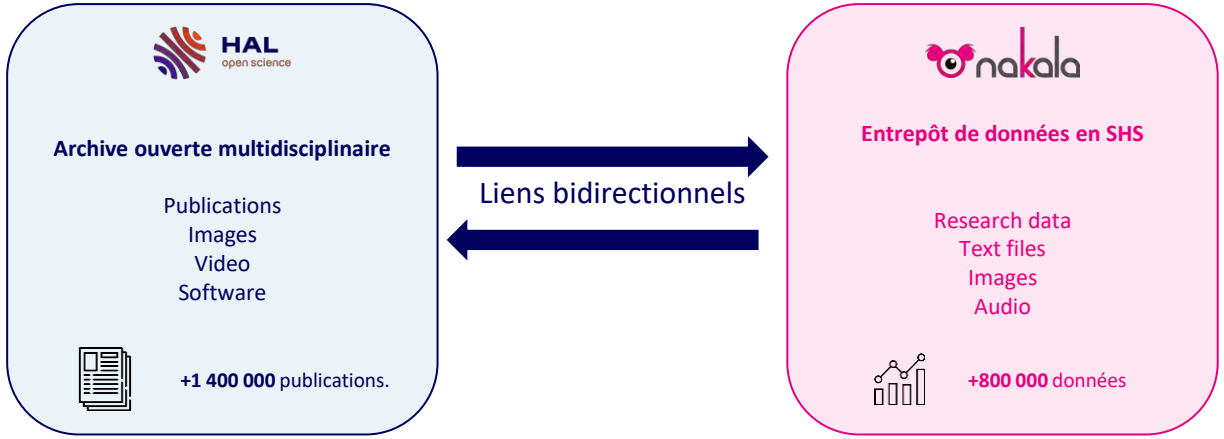
Ancienne vue du dépôt : https://hal.science/medihal-01407826

Récapitulatif

	Dépôt de fichiers annexes	Relations créées lors du dépôt	Service "Lier mes ressources"
Identifiants acceptés	Non	données: **DOI** logiciels : **SWHID** publications : **HAL, DOI, PMC, ...**	**HAL**
Contrôle des identifiants	Non	Non	**Oui**
Typage des relations	Non	Non	**hasFormat - isFormatOf, hasPart - isPartOf, hasVersion - isVersionOf, illustrate - isIllustratedBy, references - isReferencedBy, requires - isRequiredBy, conformsTo**
Affichage des relations	Prévisualisation de certains types de fichiers	Dans la liste complète des métadonnées	Dans un widget sur la page d'un dépôt
API	fileAnnexes_s	researchData_s	related_s
Exposition dans la TEI	‹ **ref** type="**annex**" subtype="sous-type" target="Lien" ›	‹ **ref** type="**researchdata**" target="identifiant" ›	‹ **relatedItem** target="identifiant HAL" type="type" **/** ›

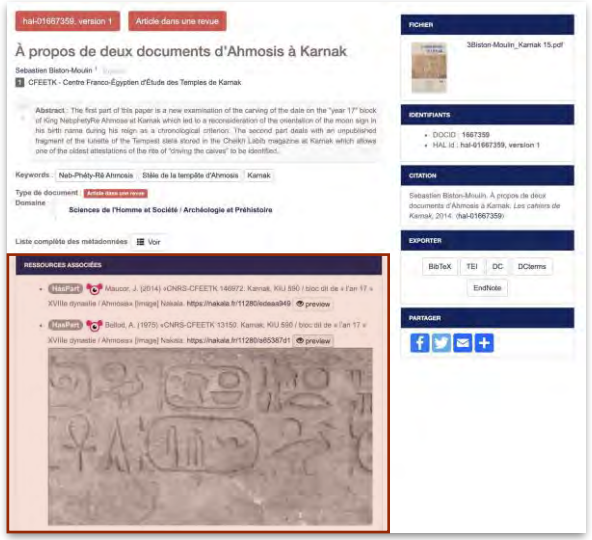
Projet EOSC Pillar (2019 – 2022)

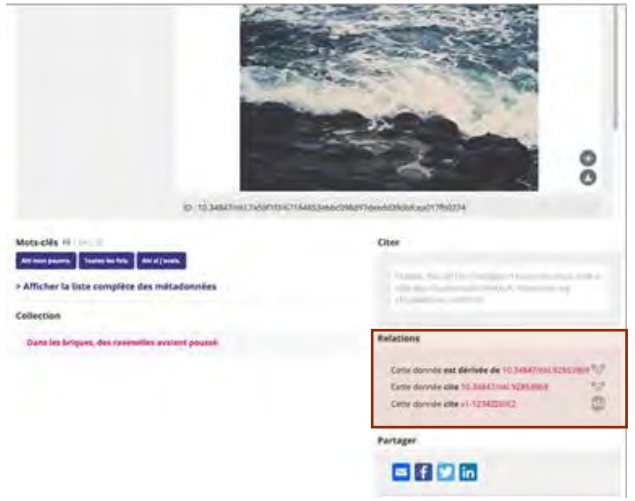
- Objectif global : coordonner les efforts nationaux en matière de science ouverte (Autriche, Belgique, France, Allemagne, Italie)
- Collaboration CCSD - Huma-Num
  - Réalisation d'une Preuve de Concept entre Nakala, l'entrepôt de données en SHS, et les publications scientifiques présentes dans HAL



POC HAL – Nakala : lier les publications dans HAL et les données dans Nakala

- Ajout d'une relation à partir de HAL
  - PID pour identifier les ressources
  - APIs pour rechercher des ressources (Aide à la saisie et autocomplétion via API Nakala)
  - Vocabulaire standard pour typer les relations
    - DataCite
  - Un outil de communication pour synchroniser et partager les liens
    - Protocole Mercure (https://mercure.rocks/)
  - Citation de la publication et des données
  - Visualisation





Cf Adeline Joffres, Nicolas Larrousse, Alice Ball, Yannick Barborini, Bénédicte Kuntziger-Planche. FAIR principles in data life for humanities. EOSC-Pillar Final Conference, Oct 2022, Paris, France. ⟨10.5281/zenodo.7156780⟩. ⟨hal-03805632⟩

04

Articuler publications, données et codes sources de logiciels : des évolutions dans le cadre du projet HALiance

Informations sur le projet

- HALiance - HAL Interoperability, Accessibility and Network Centric Environment

- 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

Référence du projet	**21-ESRE-0047**
Etablissement Coordinateur	**CNRS Rhône Auvergne**
Responsable du projet	**Nathalie Fargier**
Dates de démarrage et de clôture	**2022-01-01 → 2027-06-30**
https://anr.fr/ProjetIA-21-ESRE-0047	

HALiance – Work Packages

		Work Package	Actions
SOCLE	1	Infrastructure matérielle	• Mise à niveau de l'infrastructure matérielle • Redondance du système de sauvegarde
SOCLE	2	Infrastructure logicielle	• Migration et sécurisation du code logiciel • Ouverture du code source
SERVICES	3	Extraction de métadonnées et alignement sur des données de référence	• Optimisation de la récupération des entités nommées des PDF • Alignement sur les données de références AuréHAL
SERVICES	4	Gestion des métadonnées bibliographiques importées	• Traçabilité des métadonnées • Mise en place de règles de gestion de conflit
SERVICES	5	Qualification des preprints	• Articulation de HAL avec les serveurs de preprints • Visibilité du cycle de la publication
SERVICES	6	Collecte ciblée et automatisée de publications	• Mise en œuvre d'une nouvelle voie d'alimentation de HAL en texte intégral
SERVICES	7	Lien Publications-Données de la recherche	• Articulation entre les publications présentes dans HAL et les données de la recherche
ACTIONS TRANSVERSES	8	Accompagnement et engagement des communautés	• Visibilité et appropriation des nouveaux services HAL par ses utilisateurs (communication, documentation, formation, ...)
ACTIONS TRANSVERSES	9	Cohérence avec les principes FAIR	• Mise en place d'indicateurs FAIR • Démarche de certification Core Trust Seal

HALiance – Work Packages

		Work Package	Actions
SOCLE	1	Infrastructure matérielle	• Mise à niveau de l'infrastructure matérielle • Redondance du système de sauvegarde
SOCLE	2	Infrastructure logicielle	• Migration et sécurisation du code logiciel • Ouverture du code source
SERVICES	3	Extraction de métadonnées et alignement sur des données de référence	• Optimisation de la récupération des entités nommées des PDF • Alignement sur les données de références AuréHAL
SERVICES	4	Gestion des métadonnées bibliographiques importées	• Traçabilité des métadonnées • Mise en place de règles de gestion de conflit
SERVICES	5	Qualification des preprints	• Articulation de HAL avec les serveurs de preprints • Visibilité du cycle de la publication
SERVICES	6	Collecte ciblée et automatisée de publications	• Mise en œuvre d'une nouvelle voie d'alimentation de HAL en texte intégral
SERVICES	7	Lien Publications-Données de la recherche	• Articulation entre les publications présentes dans HAL et les données de la recherche
ACTIONS TRANSVERSES	8	Accompagnement et engagement des communautés	• Visibilité et appropriation des nouveaux services HAL par ses utilisateurs (communication, documentation, formation, ...)
ACTIONS TRANSVERSES	9	Cohérence avec les principes FAIR	• Mise en place d'indicateurs FAIR • Démarche de certification Core Trust Seal

Work Package 7 : Lien Publications – Données de la recherche

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

- Livrables :
  - HAL est interopérable avec Nakala (SHS) et les entrepôts Dataverse (notamment Recherche Data Gouv)
  - Une solution automatisée repère et associe à la publication déposée dans HAL l'identifiant et la citation des données associées
  - Un service permet de déposer un jeu de données associé à une publication et de le transférer vers l'entrepôt de données idoine

Partenaire INRAE

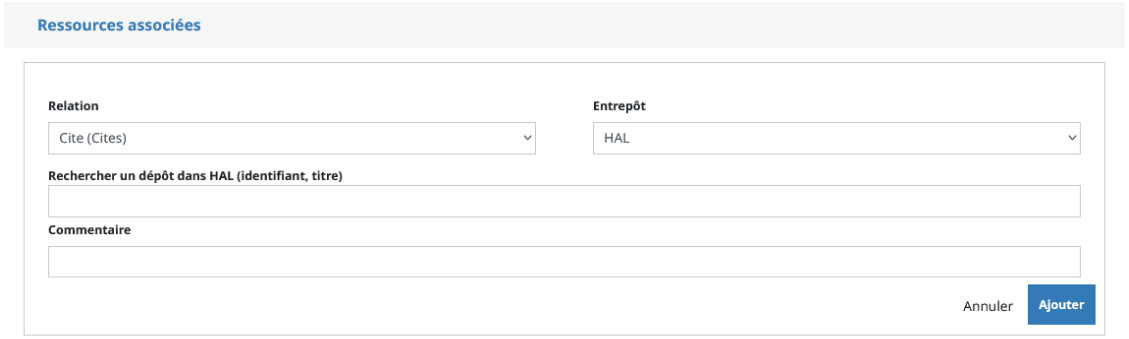
Interopérabilité entre HAL et Dataverse

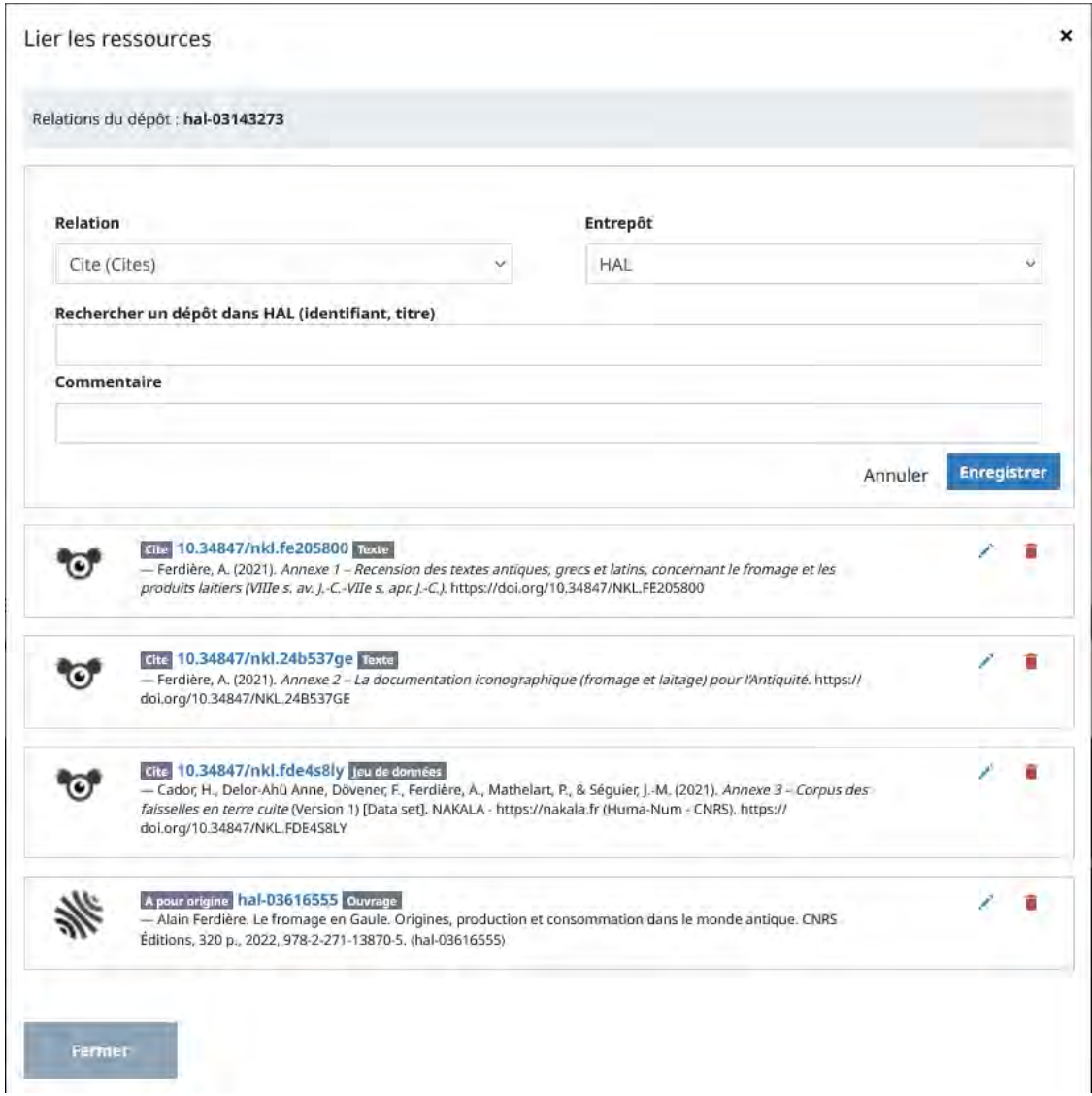
Partenaire Huma-Num IR*

Interopérabilité entre HAL et Nakala

Refonte du système de relations

- Un livrable d'HALiance : Mettre en place un nouveau service de création de relations publications-données pour remplacer les services existants
- Un module unique permettant de créer les relations publications - données – logiciels
- Accessible :
  - Depuis le formulaire de dépôt
  - Depuis le service « Lier mes ressources » d'un dépôt en ligne

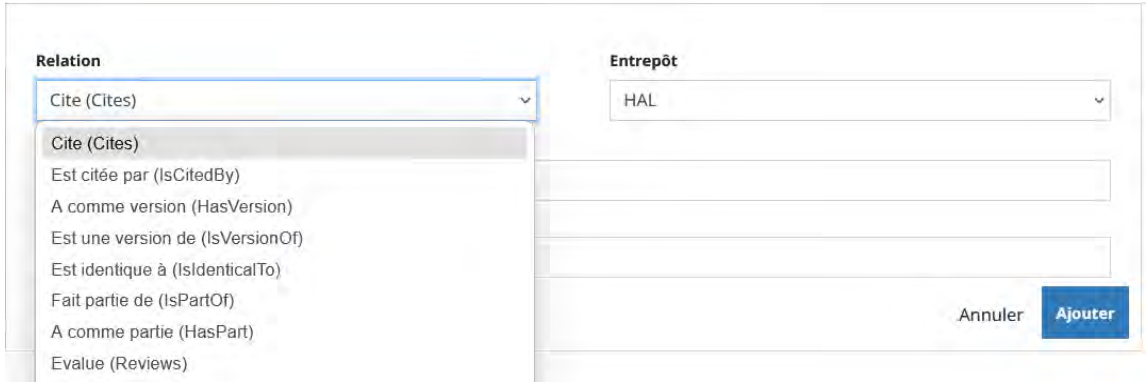


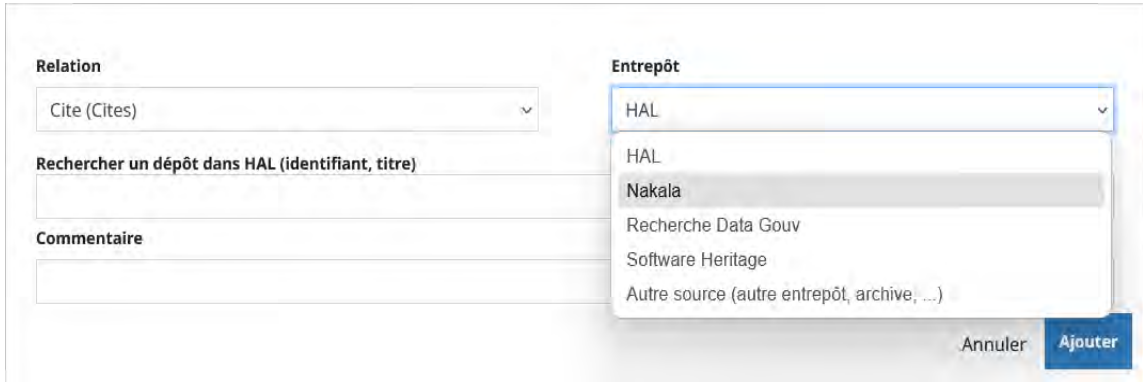


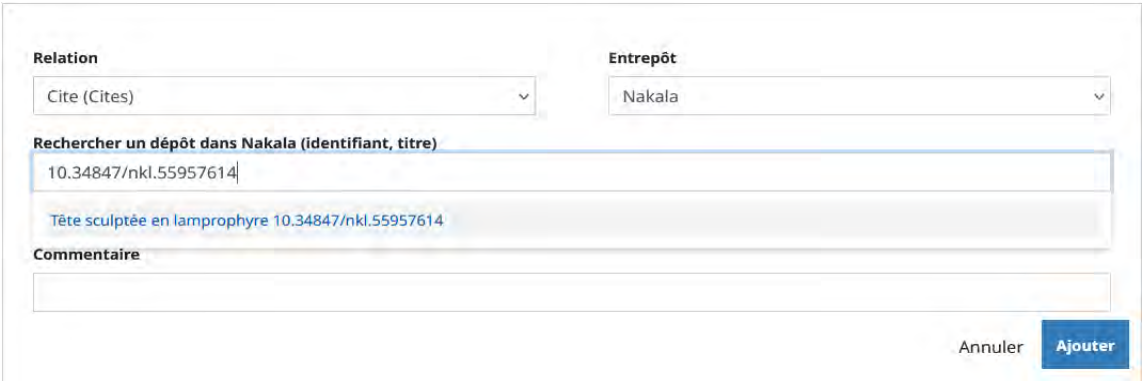
Vue du dépôt : https://hal.science/hal-03143273v1

Création de relations

- Qualification du type de relation
  - Typologie Datacite (liste réduite des types de liens : *cite*, *a comme version*, *fait partie de*…)
- Sélection de l'entrepôt
- Aide à la saisie permettant de rechercher une donnée dans les entrepôts
  - Nakala
  - Recherche Data Gouv
- Contrôle sur les identifiants (API du site DOI.org)
- Récupération des citations
  - Nakala
  - Recherche Data Gouv
  - Autre entrepôt (si saisie d'un DOI, utilisation OpenAlex)





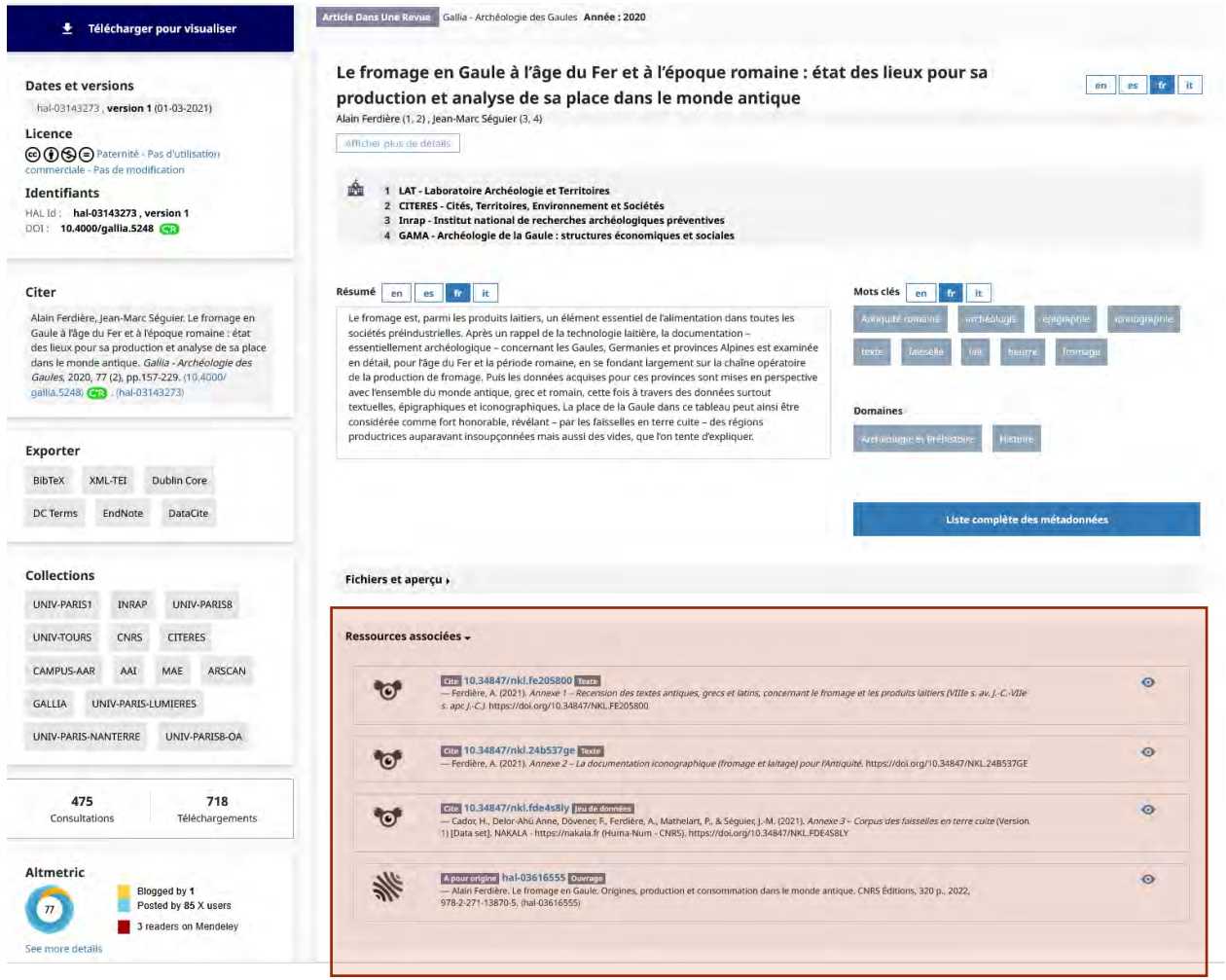


Affichage et exposition des relations

- Widget dédié sur la page publique d'un dépôt
  - Citation
  - Affichage du lien direct vers la ressource
  - Type de relation (*Cite, a comme version, etc.*)
  - Typologie documentaire
- Champs spécifiques dans l'API
  - relatedData_s
  - relatedPublication_s
  - relatedSoftware_s
- Exposition dans les formats d'export (TEI)

```
<relatedItem target="https://doi.org/10.34847/nkl.fe205800" type="Cites" subtype="http://purl.org/coar/resource_type/c_18cf" />
<relatedItem target="https://doi.org/10.34847/nkl.24b537ge" type="Cites" subtype="http://purl.org/coar/resource_type/c_18cf" />
<relatedItem target="https://doi.org/10.34847/nkl.fde4s8ly" type="Cites" subtype="http://purl.org/coar/resource_type/c_ddb1" />
<relatedItem target="https://hal.science/hal-03616555v1" type="IsOriginalFormOf" subtype="http://purl.org/coar/resource_type/c_2f33" />
```

Vue de la TEI du dépôt : https://hal.science/hal-03143273v1/tei

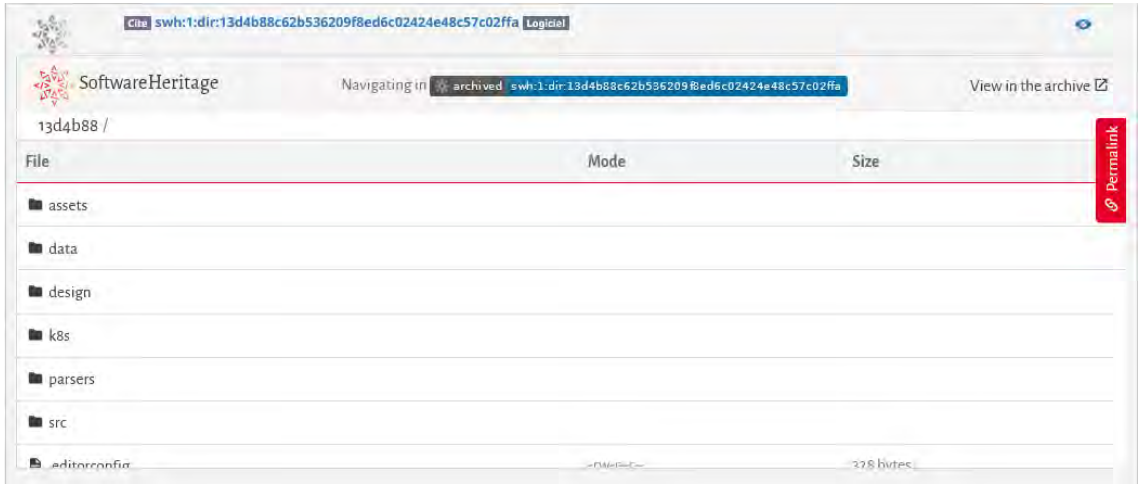


Vue du dépôt : https://hal.science/hal-03143273v1

Prévisualisation des ressources associées

Prévisualisation dans HAL : Bibracte, A. M. (2022). Perle de verre (Version 1). NAKALA - https://nakala.fr (Huma-Num - CNRS). https://doi.org/10.34847/NKL.DF0F9C4U

Prévisualisation dans HAL : Vercasson, Allison; Gaucel, Sébastien; Destercke, Sébastien; Gontard, Nathalie; Guillard, Valérie; Angellier-Coussy, Hélène, 2024, "How to assess the layers' thicknesses in polymer-coated cardboards?", https://doi.org/10.57745/D63VYY, Recherche Data Gouv, V1, UNF:6:SC5jWms99vucb25mLdpGYA== [fileUNF]

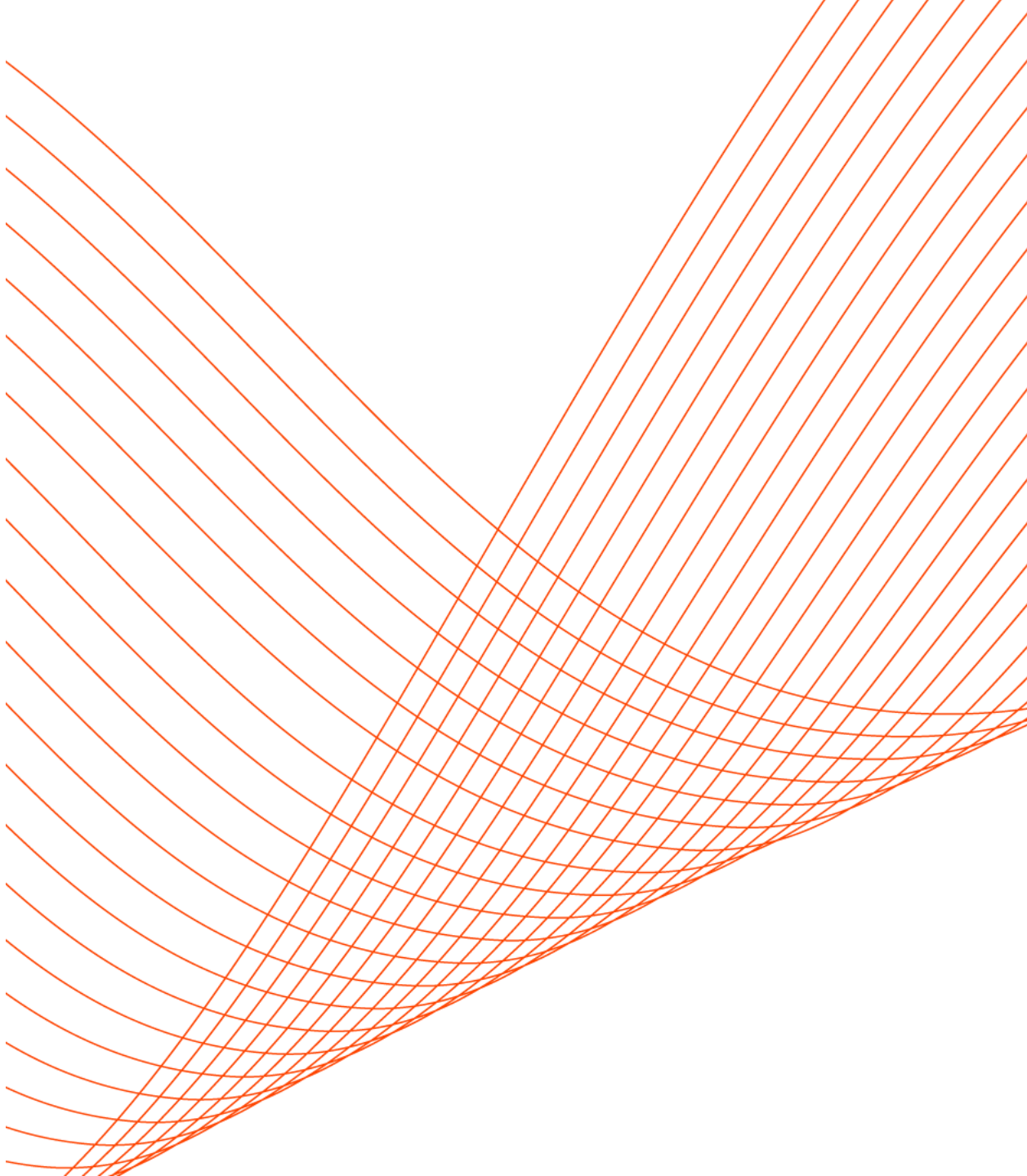


Prévisualisation dans HAL : swh:1:dir:13d4b88c62b536209f8ed6c02424e48c57c02ffa

Prévisualisation dans HAL : Vincent Guichard. Bibracte, Centre archéologique européen, proposition de programme scientifique pour les années 2006-2008. [Rapport de recherche] Bibracte, Centre archéologique européen. 2005, pp.44. ⟨halshs-01301719⟩

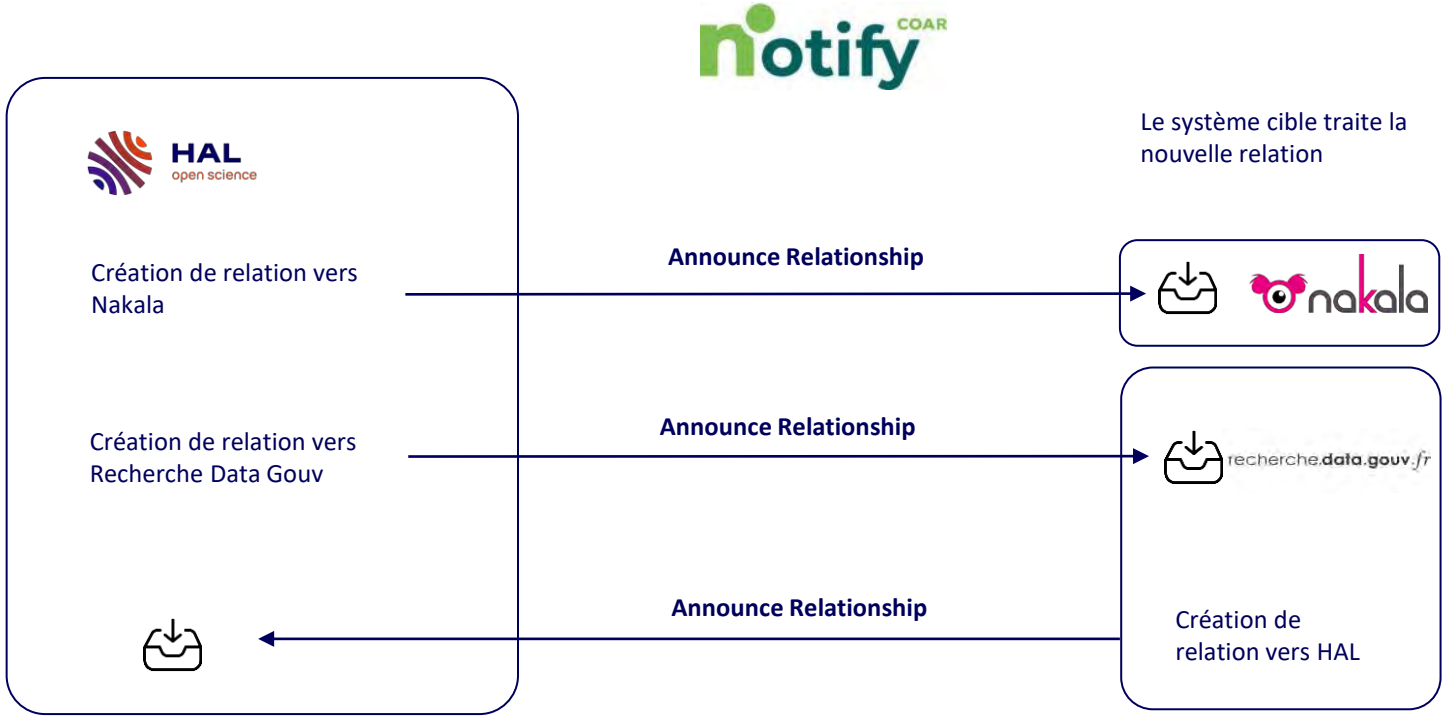
04

Travaux en cours



Signaler et échanger des relations

- Un livrable d'HALiance : Mettre en place un mécanisme de transfert des relations créées dans HAL vers les entrepôts de données partenaires



https://coar-notify.net/catalogue/workflows/repository-relationship-repository/


```
{
    "@context": [
        "https://www.w3.org/ns/activitystreams",
        "https://coar-notify.net"
    ],
    "origin": {
        "type": "Service",
        "id": "https://hal.halpreprod.archives-ouvertes.fr/",
        "inbox": "https://inbox-preprod.archives-ouvertes.fr/"
    },
    "id": "urn:uuid:e3c0c5af-6301-4753-8db5-0afe67373935",
    "actor": {
        "id": "https://hal.science",
        "name": "HAL Open Archive",
        "type": "Service"
    },
    "object": {
        "type": "Relationship",
        "as:subject": "https://hal.halpreprod.archives-ouvertes.fr/hal-03685684v1",
        "as:object": "https://hal.halpreprod.archives-ouvertes.fr/hal-03653005",
        "as:relationship": "Cites"
    },
    "context": {
        "type": [
            "sorg:AboutPage"
        ],
        "id": "https://hal.halpreprod.archives-ouvertes.fr/hal-03653005",
        "ietf:cite-as": "https://hal.halpreprod.archives-ouvertes.fr/hal-03653005"
    },
    "target": {
        "id": "https://hal.halpreprod.archives-ouvertes.fr/",
        "inbox": "https://inbox-preprod.archives-ouvertes.fr/",
        "type": "Service"
    },
    "type": [
        "Announce",
        "coar-notify:RelationshipAction"
    ]
}
```


Modèle de notification structurée au format JSON

Signaler et échanger des relations

- Transfert des relations créées dans HAL vers les entrepôts de données partenaires

Cas d'usage :
Un chercheur dépose une publication dans HAL, il indique l'identifiant des données associées dans Nakala.
La relation sera alors transmise à Nakala pour créer la relation réciproque et associer la donnée et la publication.

- Mise en place d'un mécanisme d'intégration à HAL des relations créées dans les entrepôts de données

Cas d'usage :
Un chercheur dépose un jeu de données dans RDG et indique la publication HAL associée.
Cette relation sera automatiquement récupérée et viendra enrichir la publication HAL.

HAL
science ouverte

Merci de votre attention

Sébastien Mazzarese (Coordinateur des portails HAL et Responsable de l'équipe d'assistance aux utilisateurs)

Crédit : Yannick Barborini (Responsable Conception et Développements HAL et directeur technique du projet HALiance)

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

cnrs

Inria

INRAE