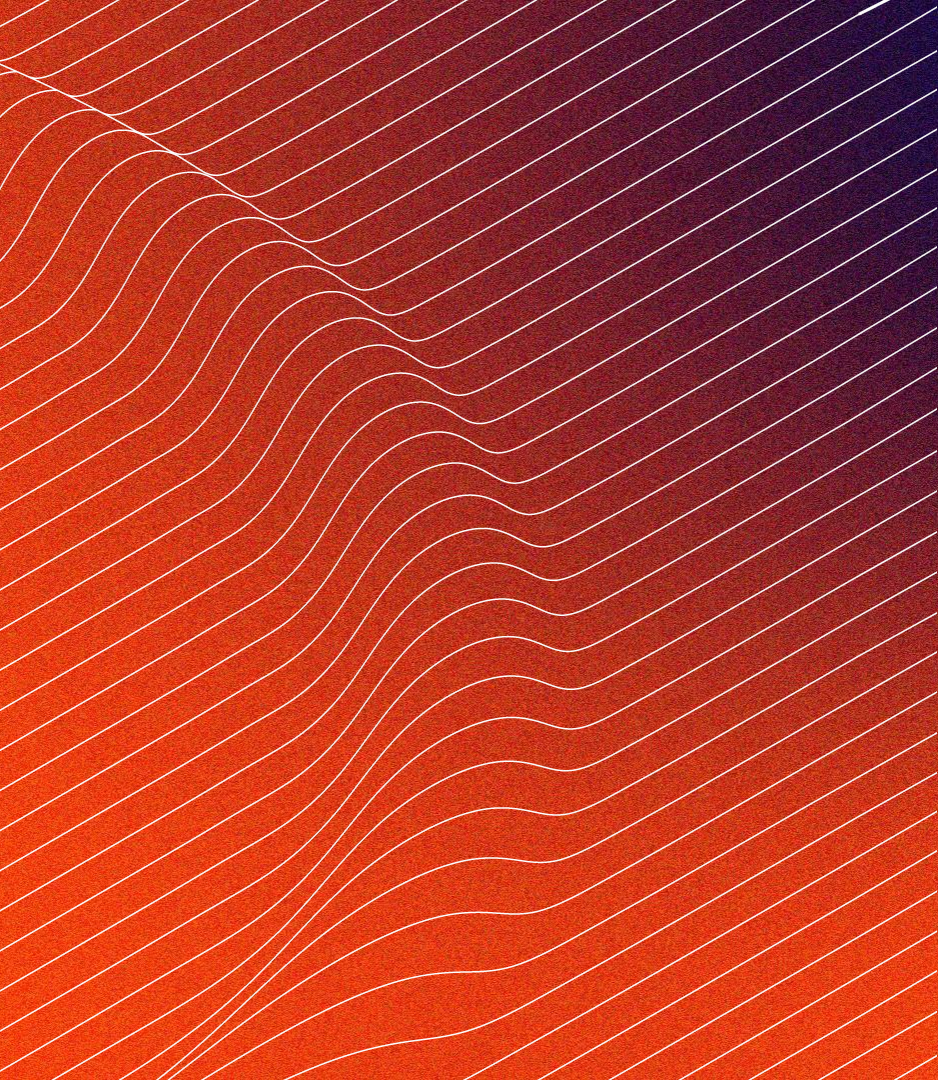


HAL open science

HAL : Strengthening Connections Between Publications, Data, and Software in the French National Open Science Ecosystem

Open Repositories 2025, Chicago, USA

Yannick Barborini

Center for Direct Scientific Communication (CCSD) / CNRS

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

cnrs

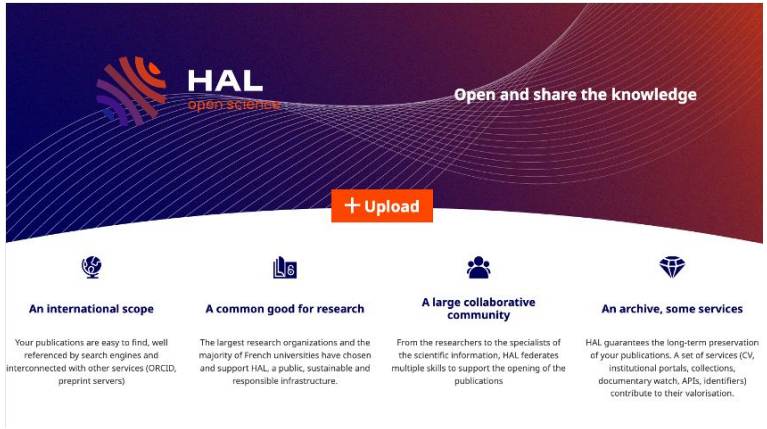
Inria

INRAE

HAL+, a Research Infrastructure for Scholarly Communication

HAL Overview

- **Founded in 2001** by a physicist, inspired by the **arXiv** model
- **Open repository** for sharing and preserving scholarly publications in all disciplines
- Serves as the **national multidisciplinary open repository** for French research
- Supported by all French universities and research institutions



Key Figures

1.5 M
Full-text Documents
Published & Unpublished Scholarly Works (50% Journal Articles)

70% English, 25% French
Content in Several Languages

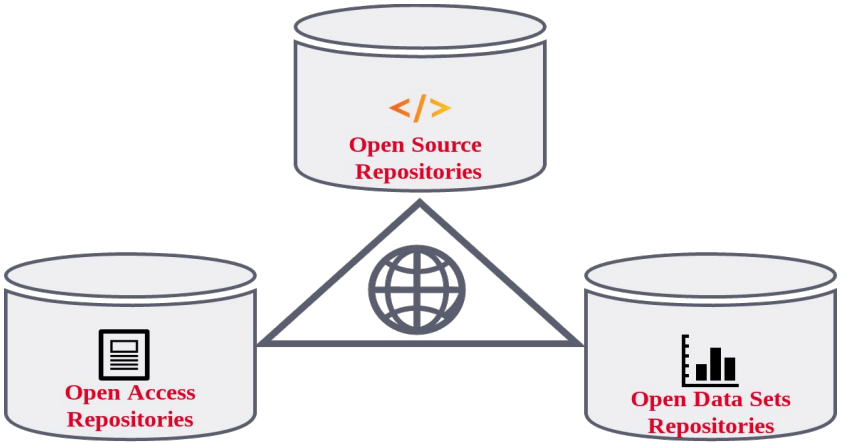
170,000
New Deposits in 2024
Consistent Annual Growth

100,000+
Active Users / Year
Growing Community of Researchers

200+ Million
Consultations & Downloads
Record Usage of Content

Enabling Reproducibility and Openness

To advance reproducible science and support knowledge sharing, it is essential to connect the **three pillars of Open Science**:



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

https://www.enseignementsup-recherche.gouv.fr/sites/default/files/2021-10/second-french-plan-for-open-science-13715.pdf

This work is carried out as part of the HALiance project, funded by the **French National Research Agency** (ANR). **PIA - Equipex+ HALiance - ANR-21-ESRE-0047**

Linking Data to Publications: Creation Form

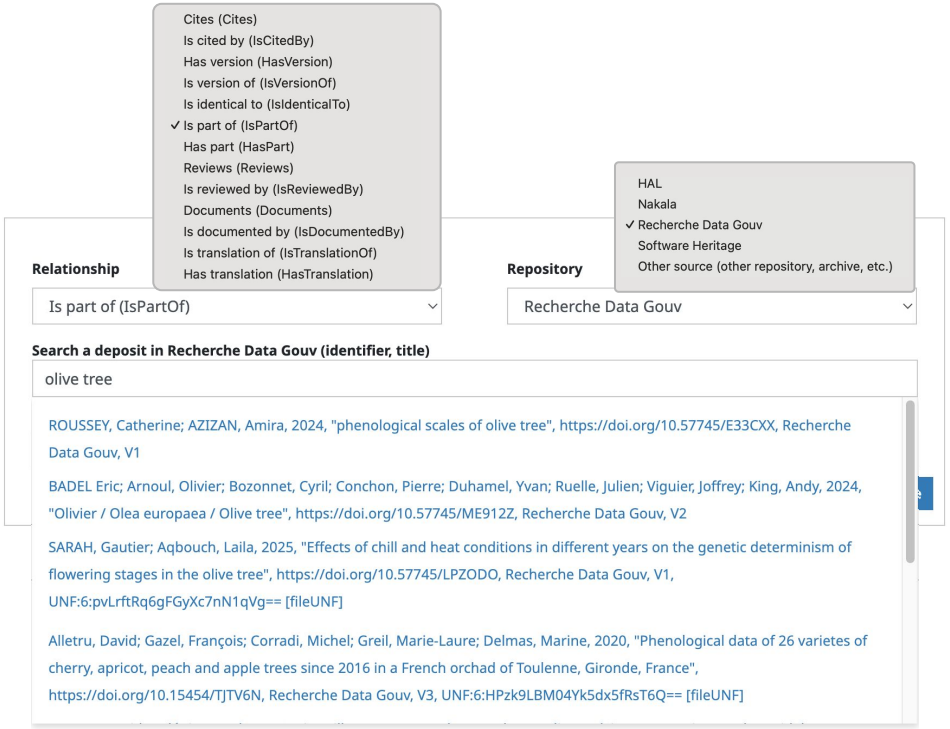
❶ Choose the Type of Relationship
Based on a simplified **DataCite typology** (e.g., isPartOf, isSupplementedBy)

❷ Select the Repository/Archive
Supports major French data repositories:
- HAL
- Recherche Data Gouv
- Nakala

❸ Assisted Search
Helps locate datasets within selected repositories

❹ Identifier Validation
Checks that provided DOIs or other persistent IDs are valid

❺ Citation Retrieval
Automatically retrieves dataset citation from:
- Recherche Data Gouv
- Nakala
- Any repository with a DOI (via **citation.doi.org**)

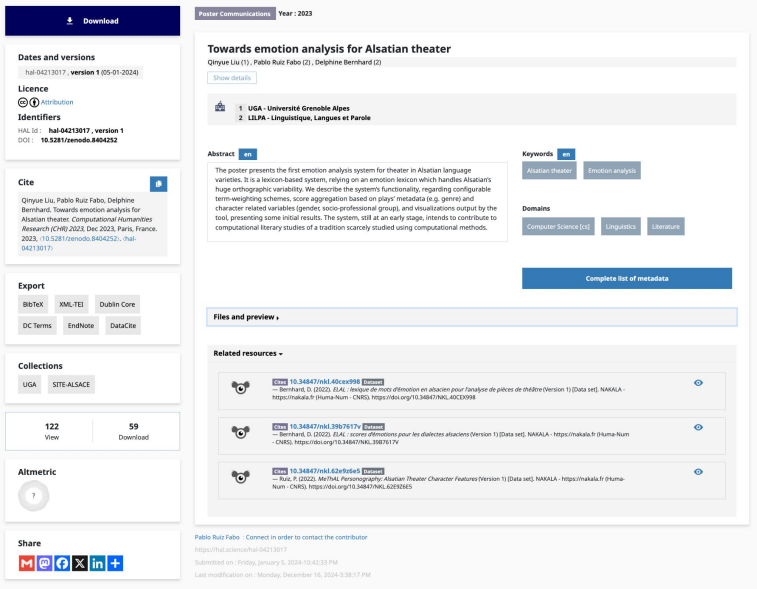


Visibility and Display of Related Resources

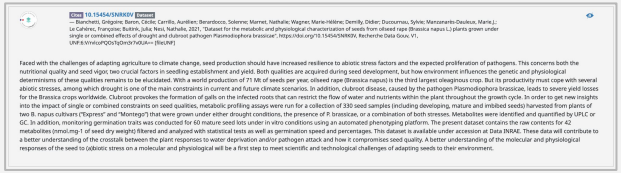
0 **Dedicated Widget** on the Deposit Landing Page
- Clearly displays and previews associated datasets or software
- Improves visibility and accessibility of linked research outputs

0 **Specific API fields**
- New fields available to retrieve and expose relationships
- Facilitates integration with external platforms and tools

0 **Display in export formats**
- Relationships included in different export formats (Dublin Core, TEI, RIOXX)
- Ensures interoperability and consistent metadata across systems

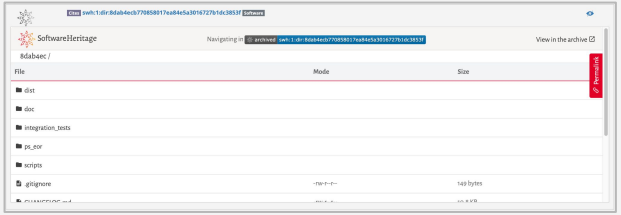


Example of **Data and Software Preview**



https://hal.science/hal-03336685v1

https://hal.science/hal-04213017v1



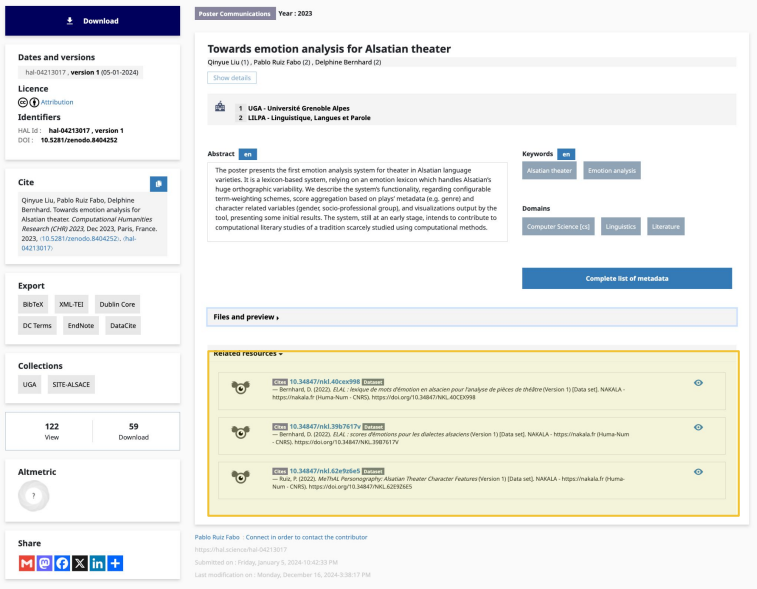
https://hal.science/hal-04344205v1

Visibility and Display of Related Resources

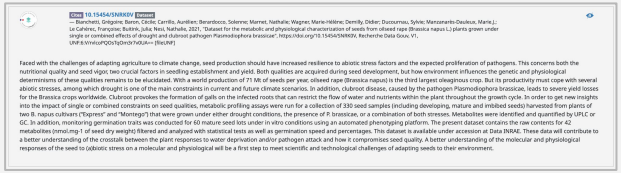
0 **Dedicated Widget** on the Deposit Landing Page
- Clearly displays and previews associated datasets or software
- Improves visibility and accessibility of linked research outputs

0 **Specific API fields**
- New fields available to retrieve and expose relationships
- Facilitates integration with external platforms and tools

0 **Display in export formats**
- Relationships included in different export formats (Dublin Core, TEI, RIOXX)
- Ensures interoperability and consistent metadata across systems

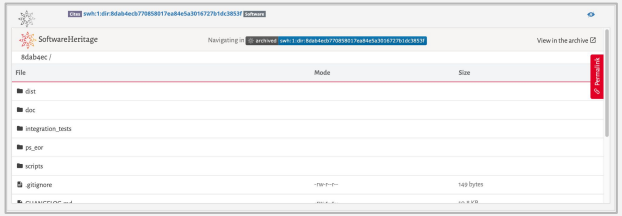


Example of **Data and Software Preview**



https://hal.science/hal-03336685v1

https://hal.science/hal-04213017v1



https://hal.science/hal-04344205v1

Automated Exchange of Relationships Between Data and Publications

0 **Reciprocal Linking Made Simple**

The relationship is shared with the data repository to establish a two-way link

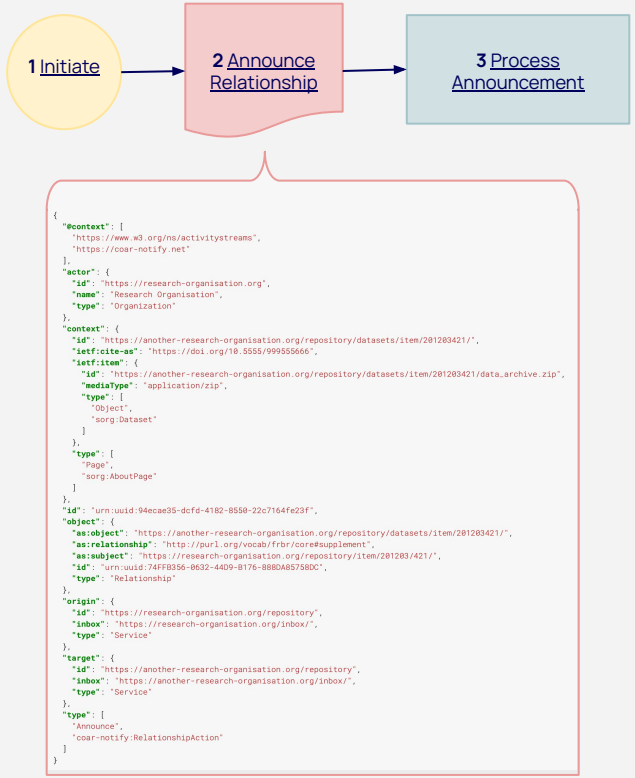
0 Enabled by the **COAR Notify protocol**

- **International initiative** to develop and promote a standardized, decentralized, and interoperable protocol for connecting research outputs and services across repositories and platforms
  ⇒ https://coar-notify.net/

- **Built on Open Web Standards**
  - Linked Data Notifications (LDN)
  - ActivityStreams

notify COAR

"Repository Announces Relationship" Workflow



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

Progress Update: Linking Publications, Data, and Software

New Relationship Creation System

- **Operational since early 2025**
- Enables authors to establish explicit relationships between Publications, Research data and Software/ source code
- **+33,000 relationships** already created

Notification Transfer System

- Currently undergoing testing
- Built using the **COAR Notify Protocol** for interoperability and decentralization

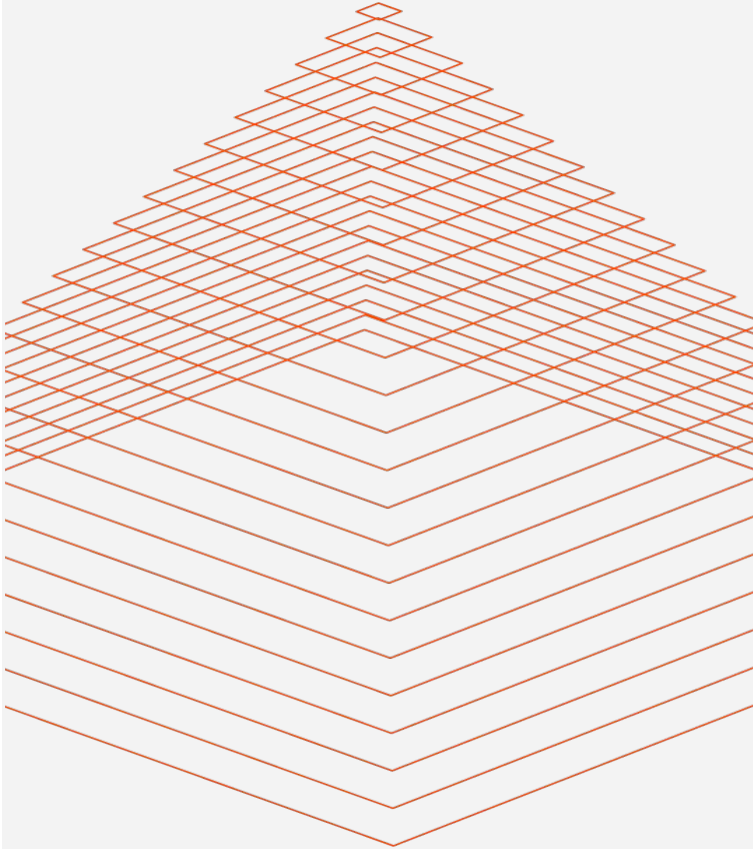
Communication Fully Operational

nakala

Upcoming Integrations (in progress)

recherche.data.gouv.fr

Software Heritage



What's Next ?

1 Expand Interoperability

- Identify and collaborate with new repositories interested in exchanging relationships via the **COAR Notify protocol**
  ⇒Toward broader adoption across disciplines and platforms

2 Leverage Machine Learning for Discovery

- Apply machine learning tools to detect mentions of data and software in publications (**GROBID, SoftCite, DataStet**)
- Propose author validation of these automatically detected relationships

SoFAIR — Making Software FAIR: A machine-assisted workflow for the research software lifecycle

3 Enrich HAL with External Services

- Integrate external services to supplement HAL metadata with pre-existing relationships

Scholexplorer — The Data Literature Interlinking Service

- Accepts validated metadata about links between publications and data
- Enhances discoverability and metadata completeness

HAL
open science

Thank you !

Yannick Barborini
yannick.barborini@ccsd.cnrs.fr

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

cnrs

Inria

INRAE