# From Open Repositories to CRIS
# A Case Study

Joachim Schöpfel[1], Otmane Azeroual[2], Stéphane Chaudiron[1], Bernard Jacquemin[1], Eric Kergosien[1], Hélène Prost[3], Florence Thiault[4]

[1]*University of Lille, Gériico laboratory, Lille (France)*

[2]*German Center for Higher Education Research and Science Studies (DZHW), Berlin (Germany)*

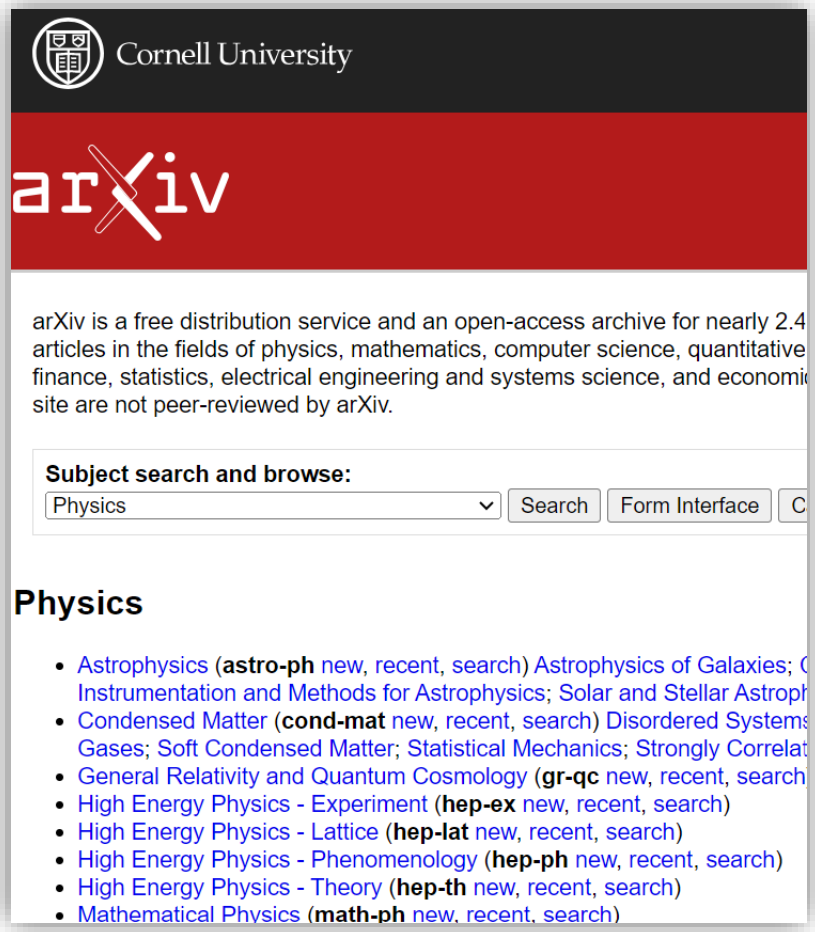
[3]*CNRS, Gériico laboratory, Lille (France)*

[4]*University of Rennes 2, PREFICS laboratory, Rennes (France)*

May 15, 2024

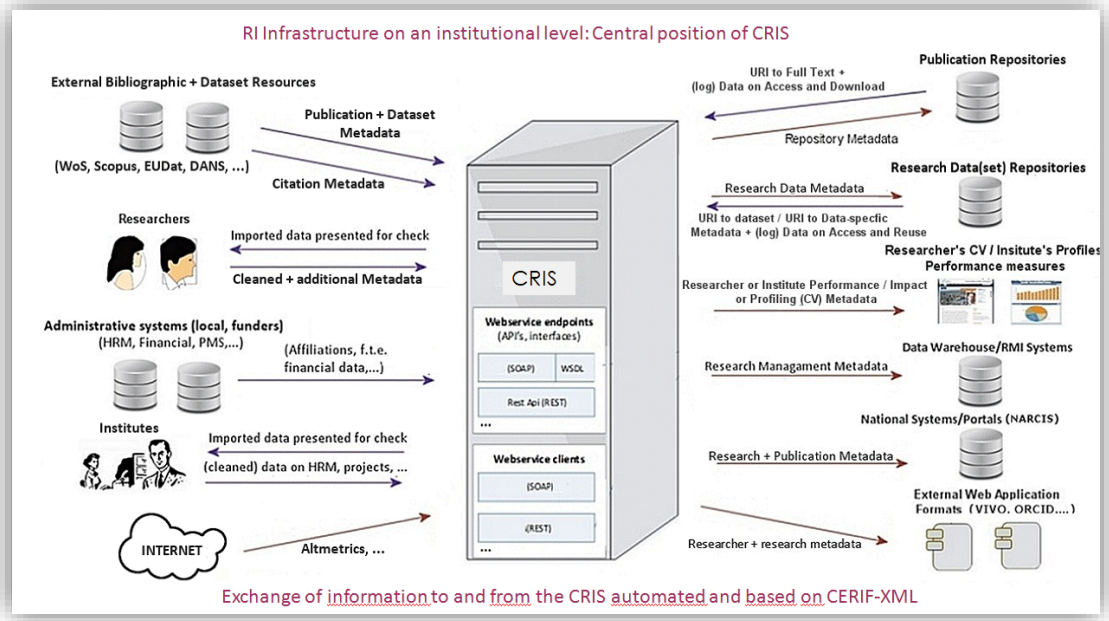
# (1) Convergence

# Open repositories



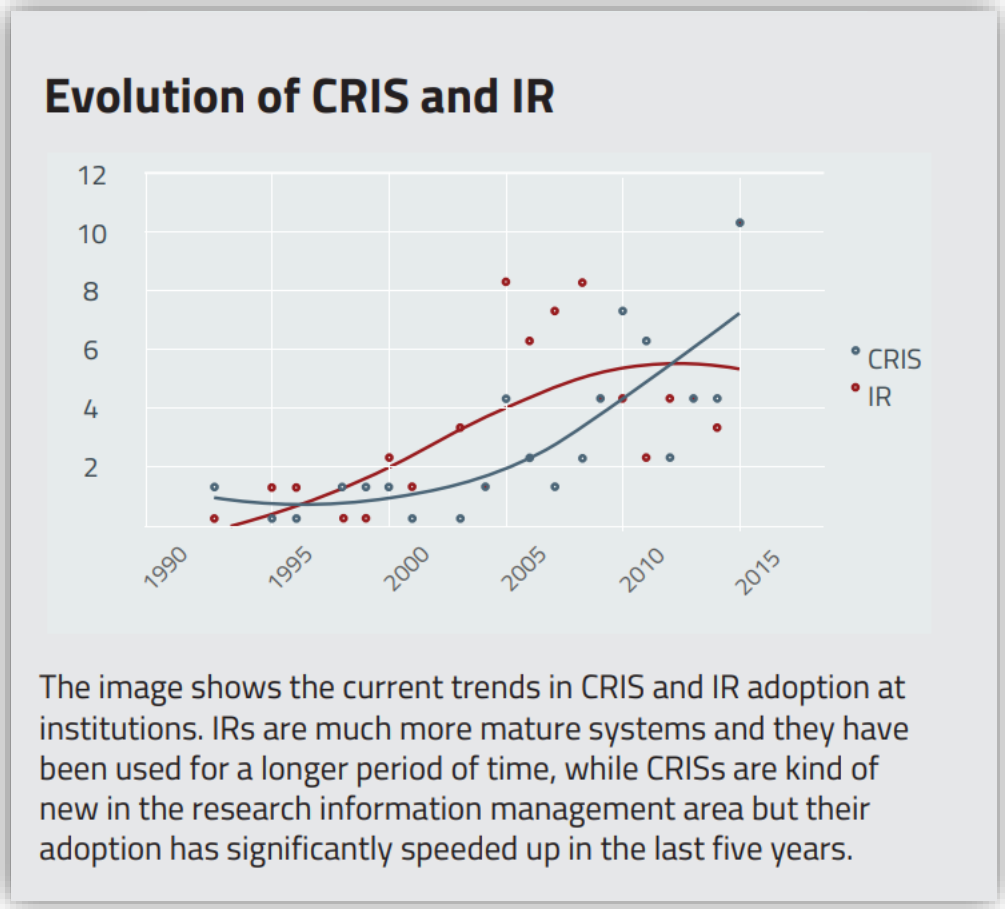

- *"Scholars need the tools and assistance to deposit their refereed journal articles in open electronic archives, a practice commonly called, self-archiving"* (Budapest Open Access Initiative 2002)
- 5,900 OA repositories worldwide, with 89% institutional repositories (IR) (source: OpenDOAR)

# Research information management systems




- Research information is "*information (...) relating to the conduct and communication of research*" (Barcelona Declaration on Open Research Information 2023)
- Research information management systems (RIMS, CRIS…) collect and store metadata on research activities and outputs (e.g., persons, publications, datasets, grants, projects, academic honors) (OCLC 2017)

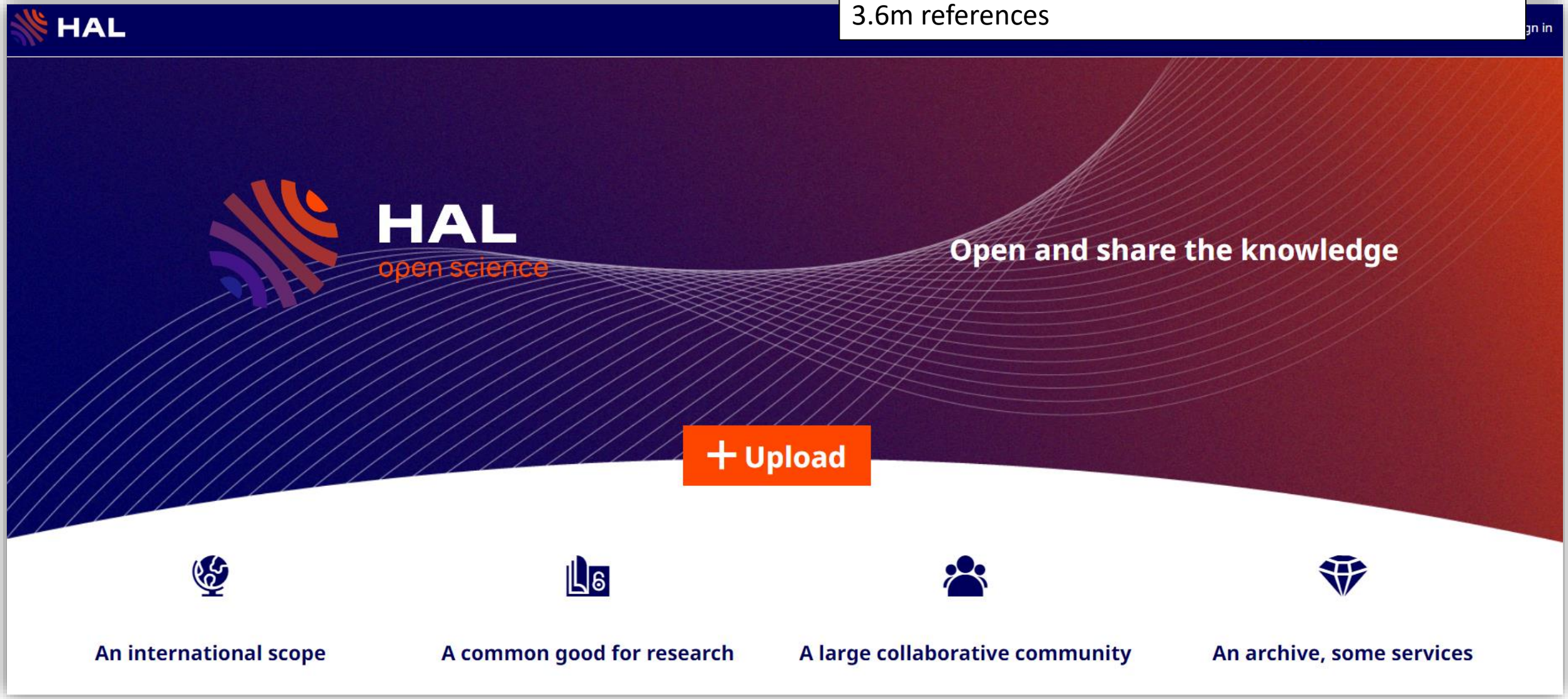
# Convergence with institutional repositories




- Data from IR to CRIS
- Data from CRIS to IR
- CRIS with IR functionalities
- IR with CRIS functionalities
- ‘‘*CRIS acting as repositories, repositories with extended data models, a wide range of interoperability features between co-existing CRISs and repositories and even a new species in the ecosystem that claims to be both a repository and a CRIS*’’ (Ribeiro et al. 2016)

# (2) The HAL ecosystem

Funding from government, RPOs, and RFOs
1.3m papers
3.6m references



# Mission

- Launched in 2002 as a French public open repository
  - Technical and political support by French HEP community
  - Partnership with arXiv
- Political support by government (National Plan for Open Science), universities, and research organisations
- Political support by national research agency (ANR) and research assessment council (HCERES)
- “*HAL is a multidisciplinary open archive providing free access to published and unpublished research results. It serves researchers affiliated with academic institutions, whether public or private*”

# Ecosystem HAL

**In the repository**

- 148 institutional portals
  - Universities, research organisations…
- Thousands of collections
  - Laboratories, projects…
- Management of identifiers
  - Persons, organisations, research projects

**On top of the repository**

- 33 overlay journals
- Conference management system with 13,000 events
- Interconnection with local repositories
- Partnerships with journal platforms

# (3) Quantitative evidence

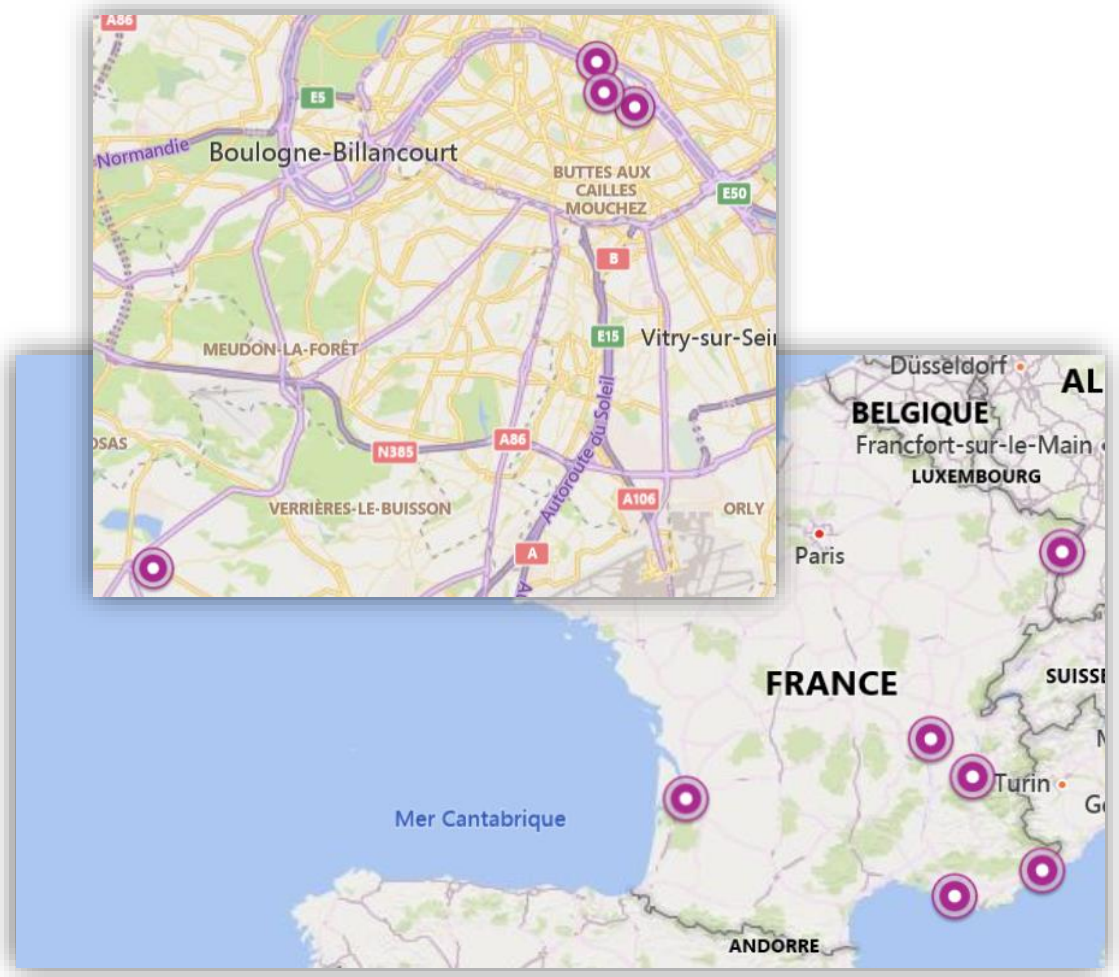
HAL/LO research project 2019-2022 with funding from GIS « Reseau URFIST »

Scientometric data

# Research project on HAL - sample

Research laboratories of the French excellence initiative

- 10 universities
- >1,200 laboratories
- >1m deposits on HAL
- All domains
- OST 2021 : 45% of French papers



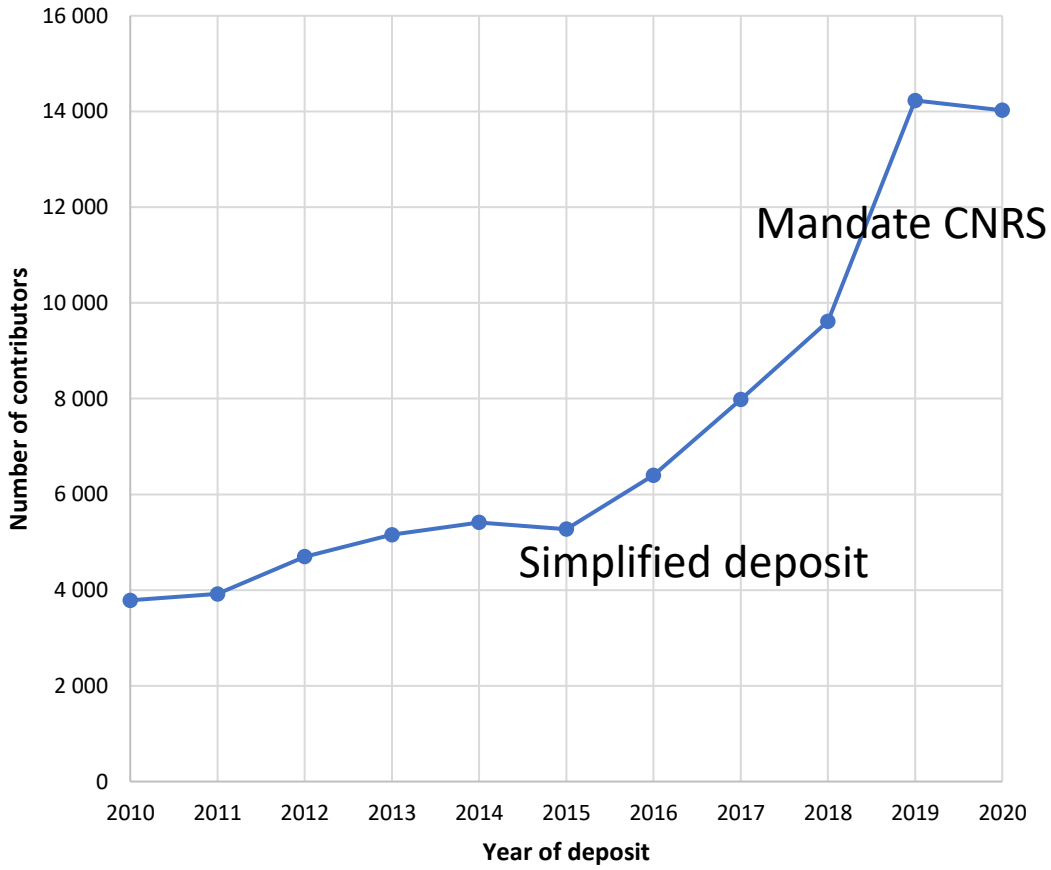
# Growing number of contributors and deposits

Progress is strongly correlated with the number of laboratories using HAL and the number of deposits

($r > 0.9$)

The average number of deposits per contributor rose by 36%

The average number of deposits per laboratory has tripled. This growth is probably not due to improved research performance, but to increased use of HAL by laboratories, simplified procedures, support from libraries, etc.

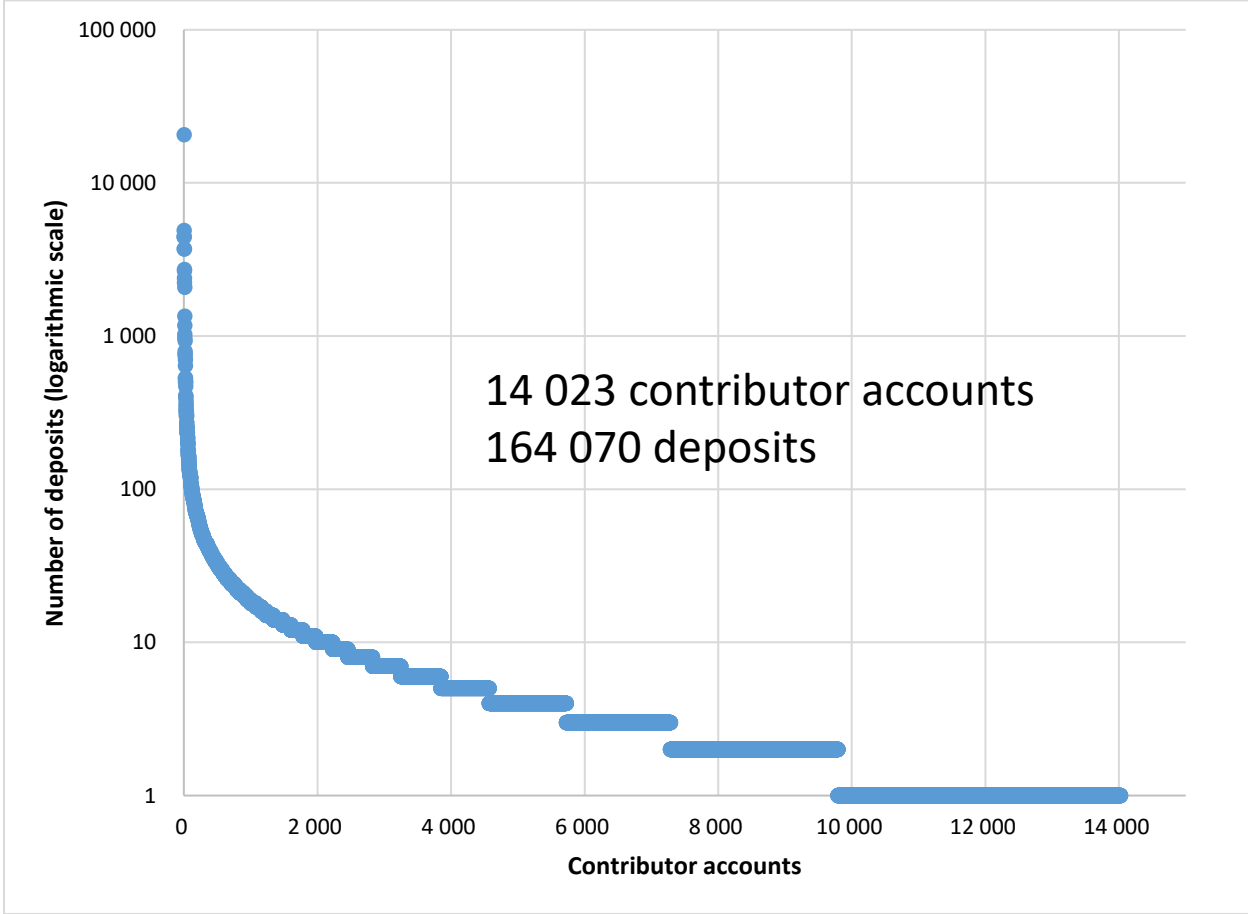



Number of contributor accounts (2010 - 2020)

# A variable contribution

## The deposit curve is a Pareto distribution

| | | |
|---|---|---|
| 20% of papers were submitted by 83% of contributors | 50% of all papers were submitted by less than 1% of contributors | Many contributor accounts only submitted one, two or three papers in 2020 |




Distribution of deposits by contributor accounts (2020)

# Differents types of contributors

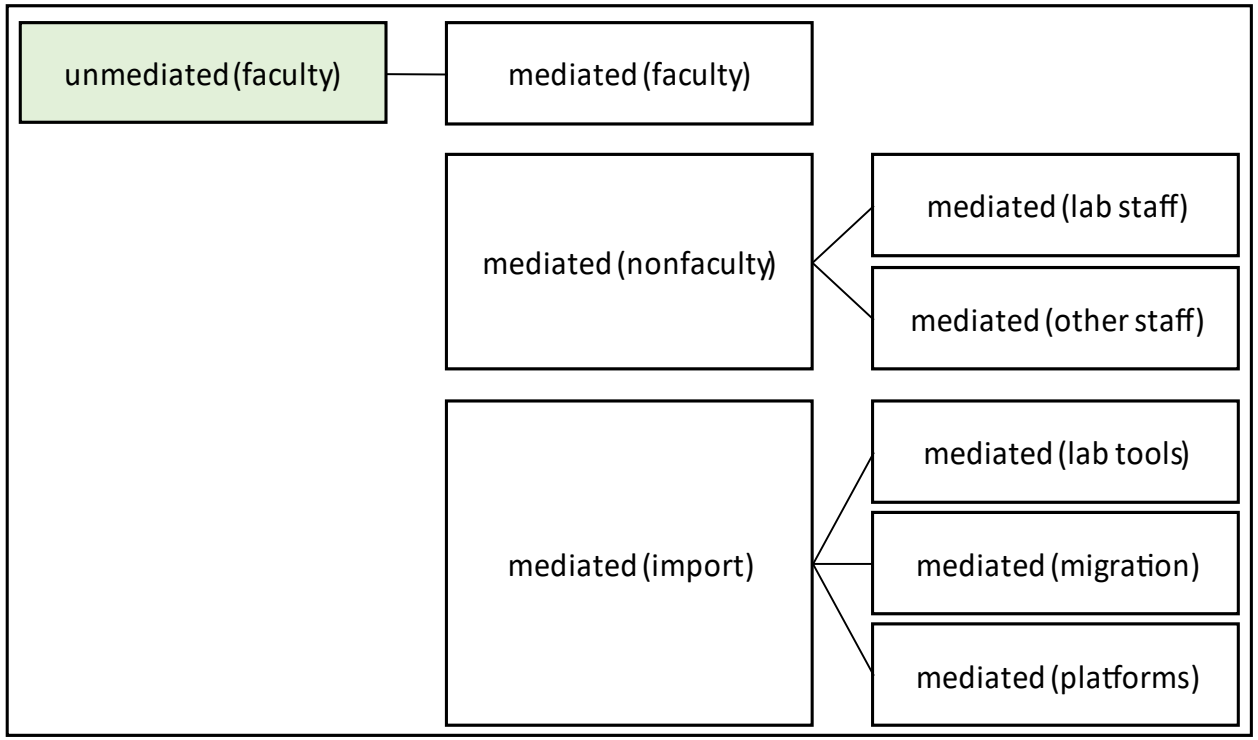
Self-archiving is just one way of feeding HAL

Deposits by others

- Colleagues
- Technical lab staff
- Other, external staff

Importing records

- Import via lab tools
- Import via migration flows
- Import via other platforms (publishers)

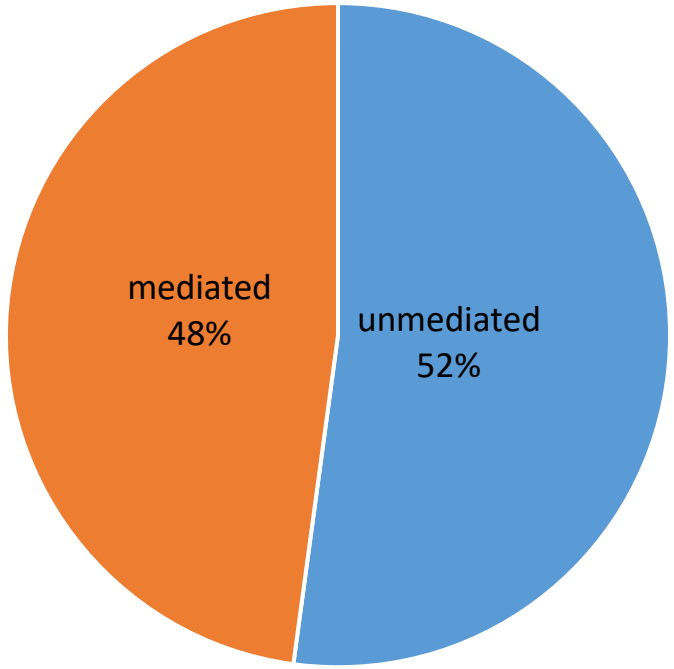


# The relative importance of self-archiving

The part of mediated contributor accounts can be estimated at 1,2 %

This small number of contributors represents 78,510 deposits, or 48% of the 165,070 deposits made in 2020

In other words, self-archiving on HAL corresponds to just over half of all 2020 deposits



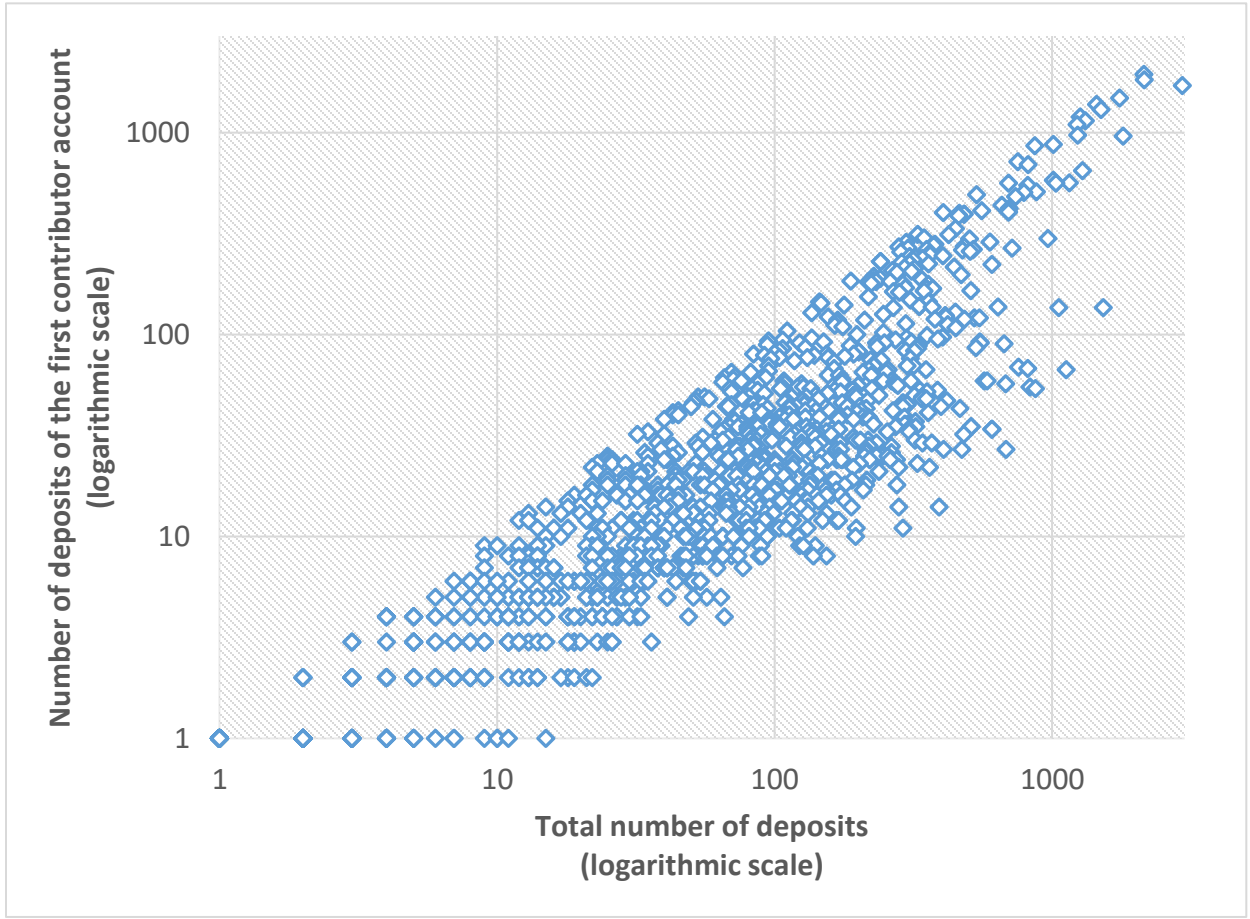
# Correlation and differences

Correlation between deposits and contributors: r=.54

Correlation between deposits of the *first contributor* and all deposits: r=.86

No significant differences between universities

But some significant projects (migration INRAE)



# (4) Qualitative evidence

HAL/LO research project 2019-2022 with funding from GIS « Reseau URFIST »

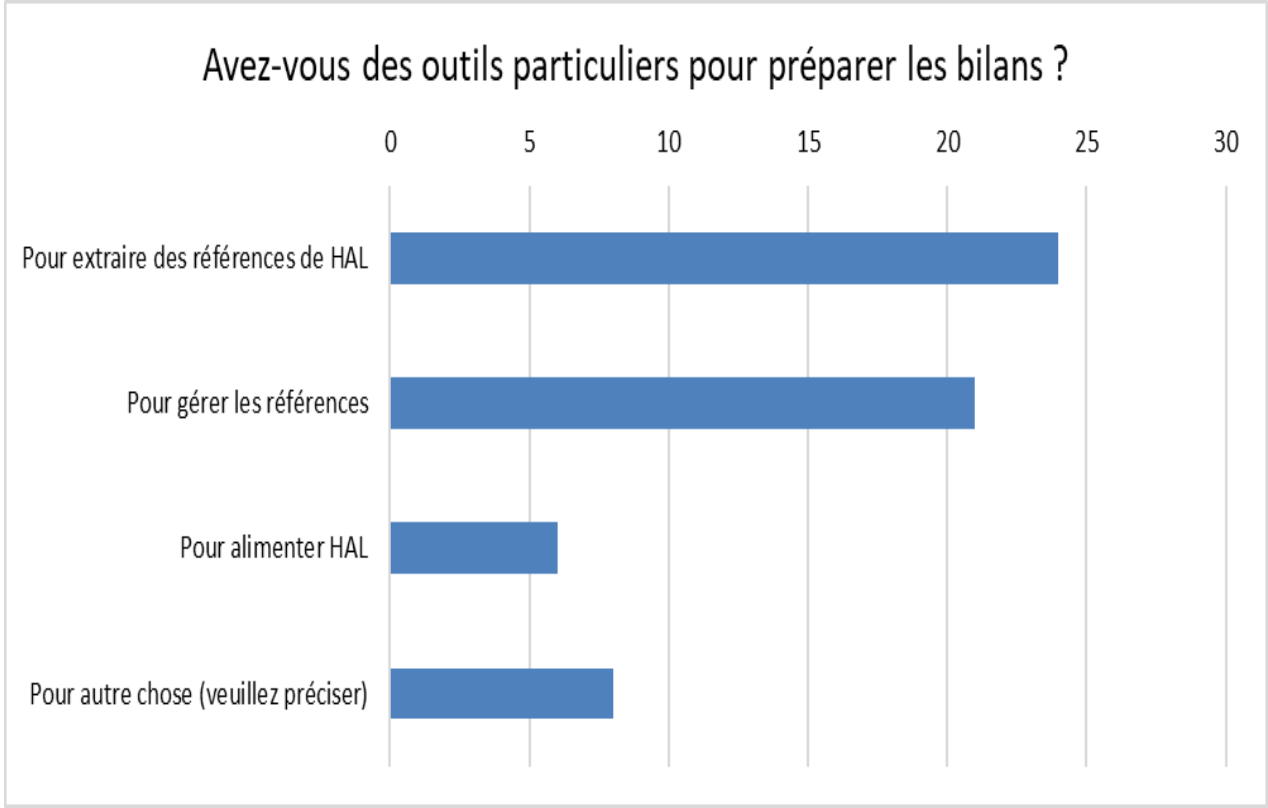
Survey and interview data

# Three discourse types

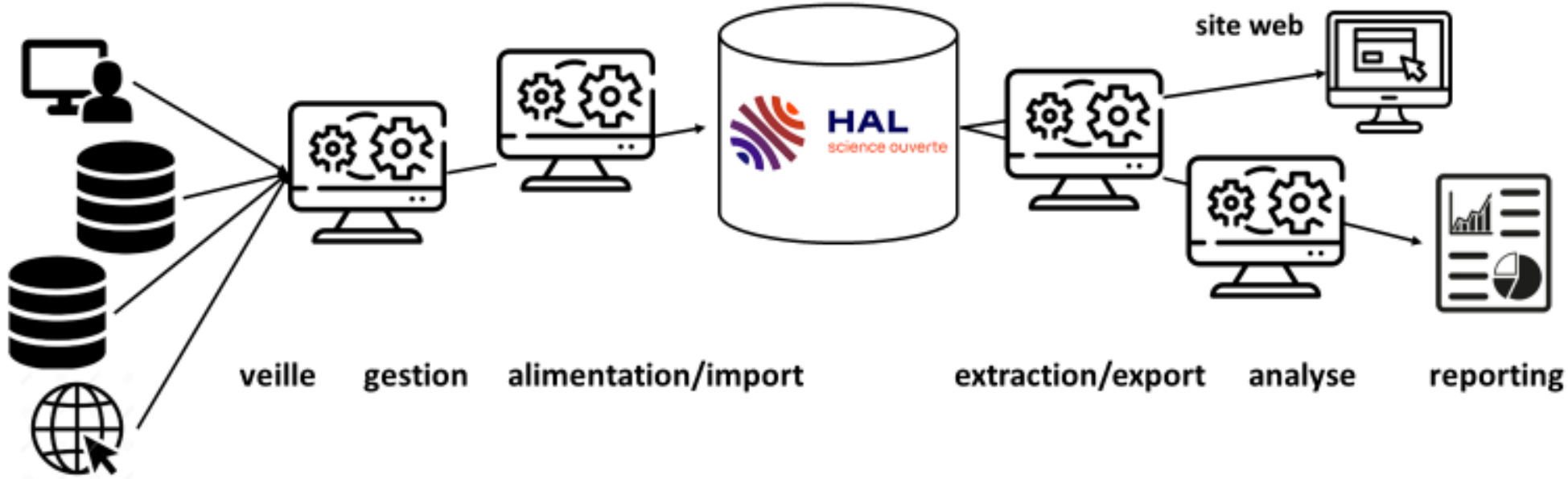
| | |
|---|---|
| An administrative discourse that emphasises institutional incentives and requests from the supervisory authorities | • *"Cela devient une obligation pour les chercheurs CNRS"* |
| A political or even ideological discourse which underlines the interest of using HAL for better research, more responsible and more open to society | • *"Une démarche très romantique, dans le sens où c'est un bel idéal (...) un peu utopique pour l'instant"* |
| A utilitarian and pragmatic discourse which highlights the use of HAL for the smooth running and development of the lab | • *"Pour les bilans annuels du laboratoire. Nous avons besoin de l'information sur le nombre de publiants car notre budget en dépend en partie. On utilise HAL aussi pour assurer le suivi de l'utilisation d'un certain nombre de ressources"* |

# An ecosystem of tools

| | |
|---|---|
| External tools developed for HAL | • Upstream (import)<br>• Downstream (export) |
| Other external tools | • Data bases, search engines<br>• Reference managers |
| In-house tools | • Data bases<br>• Bibliometric tools |



# Scope and challenges of this ecosystem




*qualité des sources*
*exhaustivité*
*accessibilité*
*fiabilité*

*qualité des métadonnées*
*(richesse, homogénéité)*

*interopérabilité*

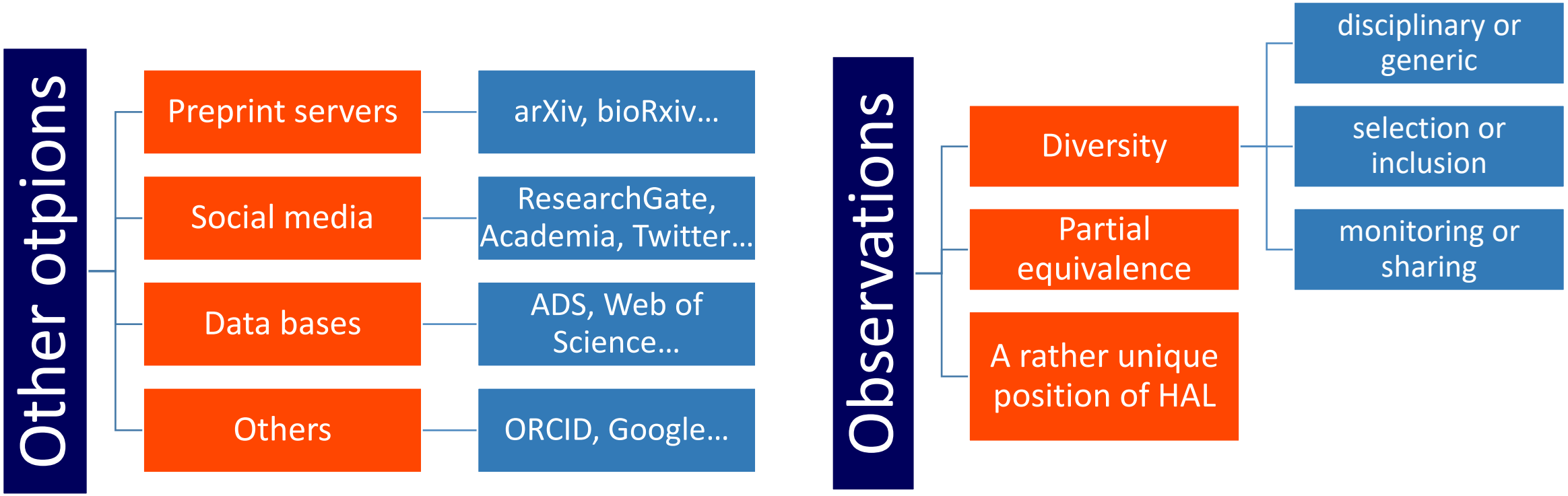
*qualité des métadonnées*
*(granularité, fiabilité)*

*interopérabilité*

Gestionnaire de références

Système d'information recherche

# Functional equivalents of HAL



# (5) Perspectives

# The impact of research assessment

**Mandatory procedures**

- Institutional assessment: national research assessment council (HCERES)
- Individual assessment: research performing organisations (CNRS, INRAE)

**Incentives**

- No other accepted publication sources than HAL
- Budget decisions by some universities (bonus/malus)
- Obviously, there is a direct and immediate impact of assessment on individual and collective information behaviour

# New challenges

This transformation is not specific to France

The new challenges are those of research infrastructures and, in particular, those of CRIS

Metadata quality

Data and metadata FAIRness

Responsible research assessment

Ethics

# The transformation of HAL

- *"Chaque fois qu'on a besoin d'un indicateur, on utilise HAL"*
- HAL has not become a CRIS (and probably will not), but
  - HAL is the most important (central) data source for research assessment (publications)
  - HAL develops new services and functionalities similar to a CRIS
- The challenge of openness and interoperability
- Metadata quality: completeness, reliability, sources, PIDs…
  - *Garbage in, garbage out*
  - Quality control?
  - Data cleansing?
- Ethics
  - Date type (personal and other sensitive data)
  - Access, usage, and potential impact of data

# References


- Azeroual, O., & Schöpfel, J. (2019). Quality issues of CRIS data: An exploratory investigation with universities from twelve countries. *Publications*, *7*(1). https://doi.org/10.3390/publications7010014
- Schöpfel, J., & Azeroual, O. (2021). Current research information systems and institutional repositories: From data ingestion to convergence and merger. In D. Baker & L. Ellis (Eds.), *Future Directions in Digital Information* (pp. 19–37). https://doi.org/10.1016/B978-0-12-822144-0.00002-1
- Schöpfel, J., Chaudiron, S., Jacquemin, B., Kergosien, E., Prost, H., & Thiault, F. (2023). The Transformation of the Green Road to Open Access. *Publications*, *11*(2), 29. https://doi.org/10.3390/publications11020029
- Schöpfel, J., Thiault, F., Prost, H., Jacquemin, B., & Kergosien, É. (2023). L'utilisation de HAL par les laboratoires de recherche. *Balisages*, (6). https://doi.org/10.35562/balisages.1166
- Schöpfel, J., Prost, H., Jacquemin, B., Kergosien, É., & Thiault, F. (2024). La valorisation de HAL : Finalités, outils et process. *JDMDH Preprint*. https://hal.univ-lille.fr/GERIICO/hal-04450898v1
- Schöpfel, J., Kergosien, É., Prost, H., Jacquemin, B., & Thiault, F. (2024). L'utilisation de HAL : Pratiques, avis et alternatives. *Preprint JDMDH*. https://hal.univ-lille.fr/GERIICO/hal-04523829v1

euro CRIS
THE INTERNATIONAL ORGANISATION FOR RESEARCH INFORMATION

# Thank you

joachim.schopfel@univ-lille.fr