



FAIR-IMPACT

Expanding FAIR solutions across EOSC

FAIR Implementation Workshop : French Data Management Clusters

The "Recherche Data Gouv" ecosystem

Gilles MATHIEU, MESR, April 2024



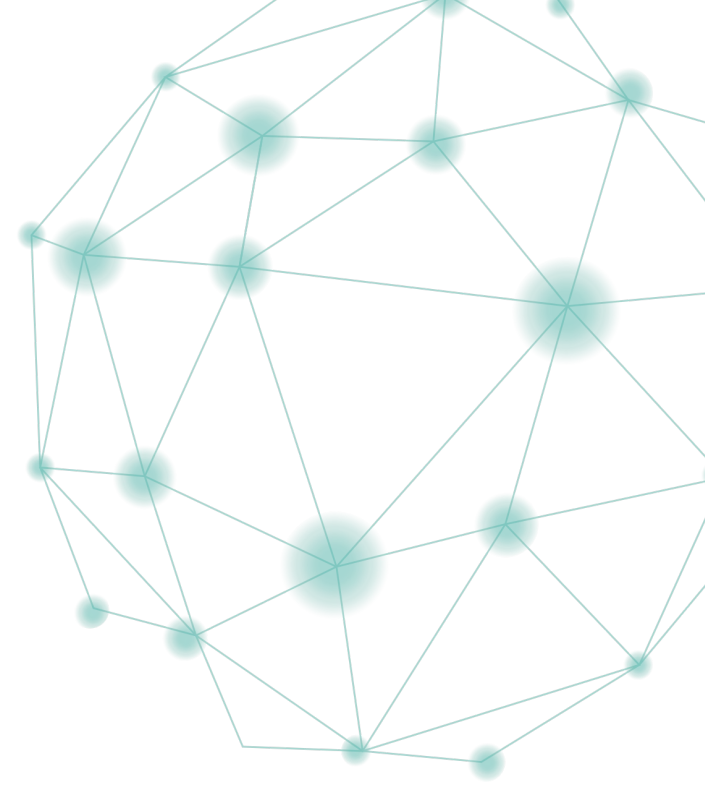
Funded by
the European Union

Recherche Data Gouv: An ecosystem for sharing and opening research data

April 2024



@RechercheDataGv



Recherche Data Gouv - from its genesis to the ecosystem

Ambition

Open up all research products and methods

"as open as possible, as closed as necessary"

- Publications
- Data
- Source codes

*... set out in the "Loi pour une République numérique" (2016), France
...reiterated by Open Science*

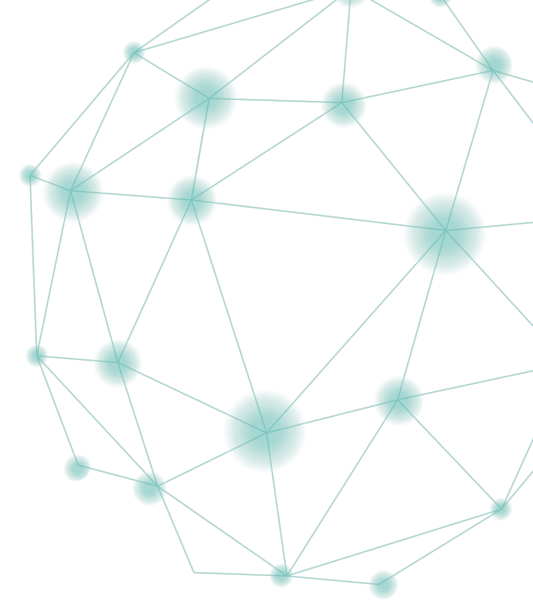
**More cumulative, robust and reproducible science
Science that is transparent and accessible to all**



Designing natively reusable data

Scientific work on FAIRisation of data :

The further away the re-use is, the more difficult it becomes to find and interpret the data.



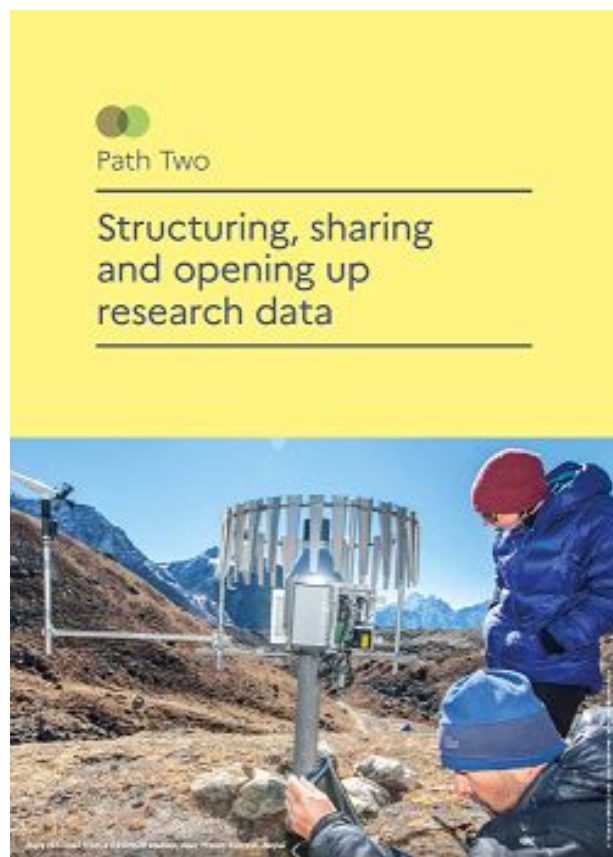
The first beneficiary of well-described data is the person who produced it!

Find and reuse data, manage it better, save time and reduce the digital footprint

Citation rate for publications with shared data: +25% to +75%

Shared data recognised as a research output

French second plan for open science



“5. Create Recherche Data Gouv which will offer sovereign and certified (Core Trust Seal) storage and signposting for research data, and a range of support services for researchers”

Commitments

Support the structuring, preservation, sharing and opening up of research data, as well as the discovery of research data

Promote practices for and eco-responsible data management practices

Encourage the re-use of research data

Federate local and thematic initiatives

Second national plan for open science

<https://www.ouvrirlascience.fr/second-national-plan-for-open-science/>

Data, algorithms and source code policy

<https://www.ouvrirlascience.fr/politique-des-donnees-des-algorithmes-et-des-codes-sources-feuille-de-route-2021-2024/>



Create *Recherche Data Gouv*

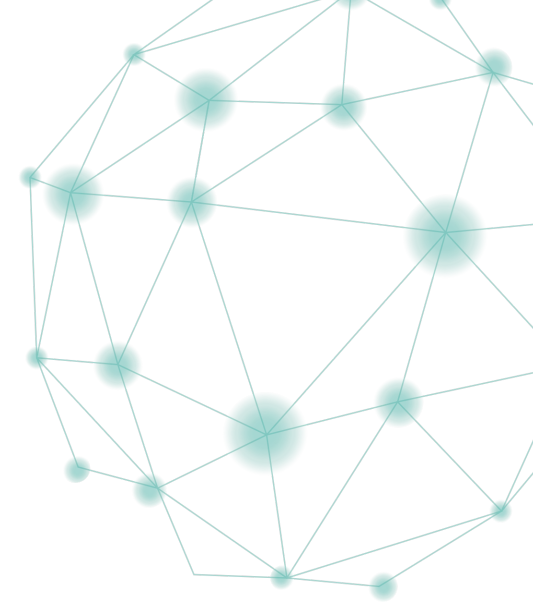
Working towards recherche.data.gouv.fr

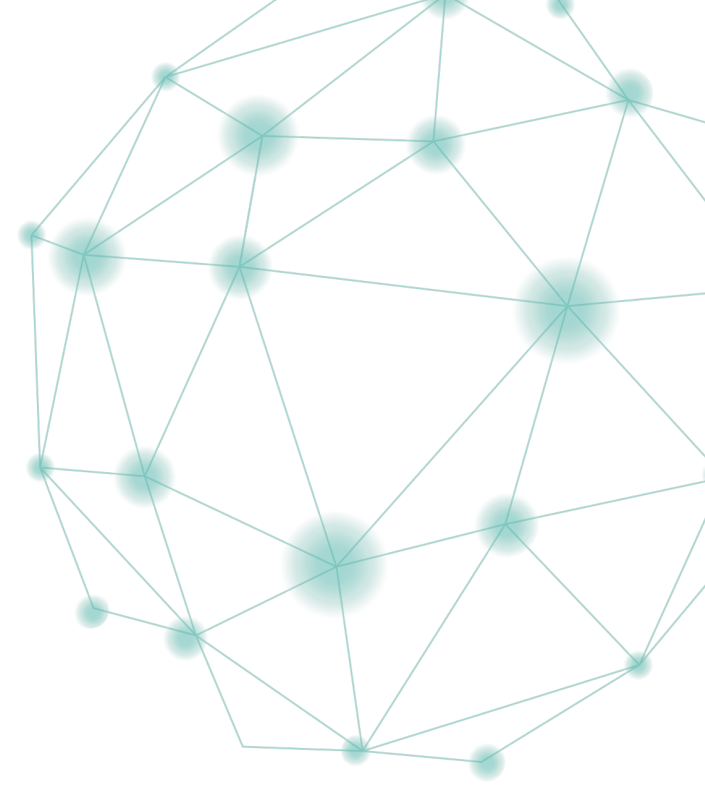
Study and prefiguration phases (2020-2021)

- Feasibility study for a generic repository: user needs
- Comparative study of other national data services
Australia, Norway, Netherlands, Canada, United Kingdom, Germany
- Ambitions for the service and implementation scenarios

Lessons learned from the study and prefiguration phases

- Towards a single ecosystem – « stronger together »
- The emphasis should be put on quality rather than quantity
- Support should be a central element of the system





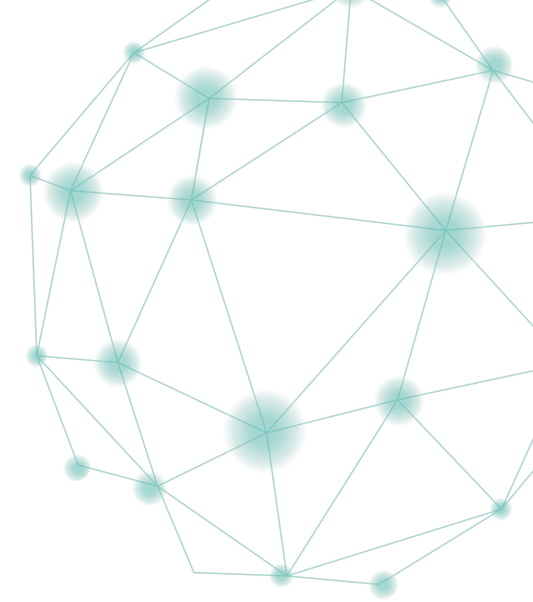
An ecosystem for sharing and opening research data

Aims

Leaving no research team without any solution to open or share their data

Make support a central element of the system

Recognise, federate and expand existing initiatives



A network of support services close to the research community

- Raising awareness, training or supporting research teams throughout the data life cycle

A sovereign solution for depositing, publishing and reporting data

- A federated national research data platform : data repository and data registry
- A complement to thematic repositories
- An alternative to publishing on commercial platforms or platforms that do not meet trust criteria
- Become an *European open science cloud* service

Global view of the ecosystem

19+ DATA MANAGEMENT CLUSTERS

Generalist expertise which is geographically close to the research teams it supports on all data issues

4 INSTITUTIONAL REFERENCE CENTRES

Provide specific guidance according to institutional policies and strategies

1 DATA REPOSITORY

A mutualised repository for all institutions to deposit and publish data

6+ THEMATIC REFERENCE CENTRES

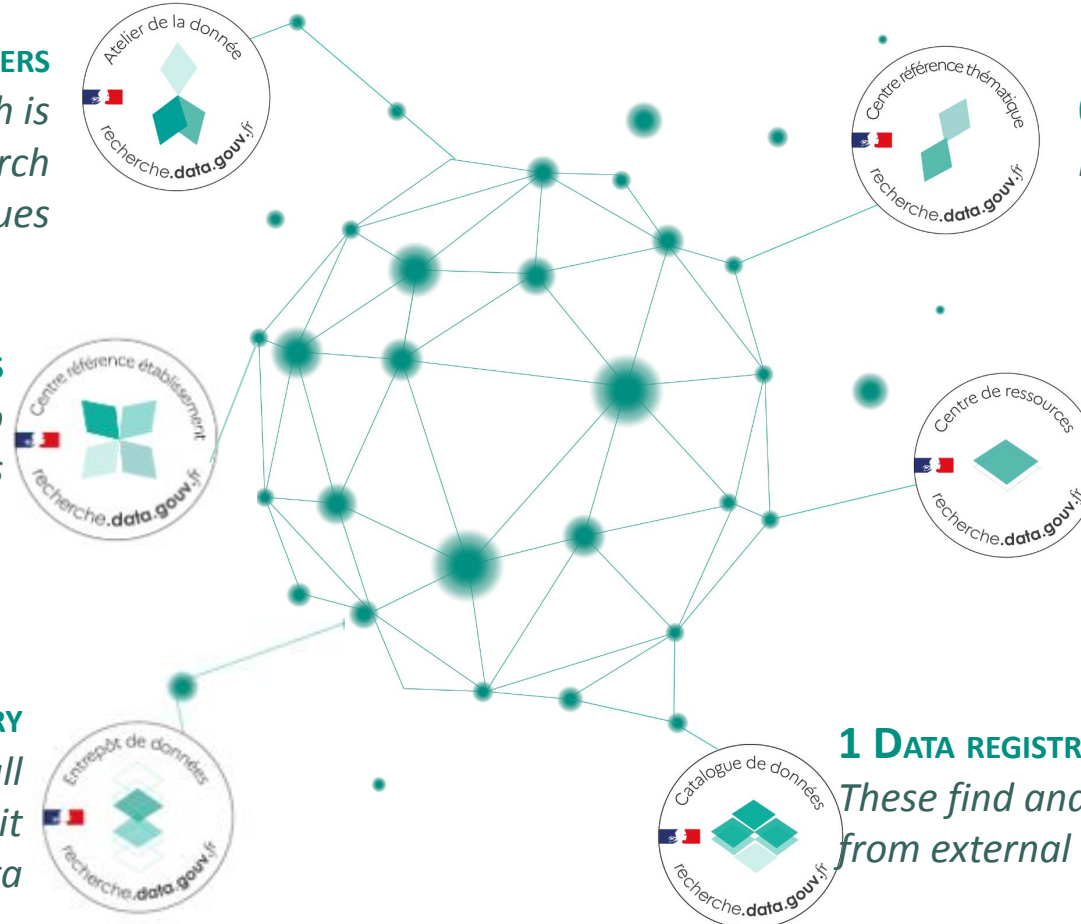
Expertise in a given scientific field

4 RESOURCE CENTRES

These support the data management clusters and capitalise on their practices

1 DATA REGISTRY

These find and harvest data from from external repositories



Data management clusters

<https://recherche.data.gouv.fr/en/page/data-management-clusters-generalist-services-throughout-france>



As of March 2023:

14 labelled clusters
5 on the pathway to labelling
~10 ongoing projects
Ongoing call for further labelling

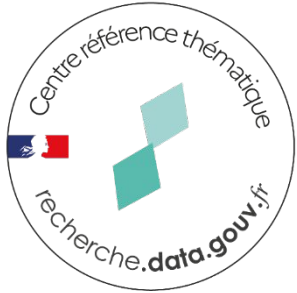
90+ FTE
350+ people involved
Around 80 institutions involved

**Generalist expertise which is geographically close
to the research teams it supports on all data issues**





Thematic reference centres



Advanced expertise in a given scientific field

Definition and dissemination of **best practices and international standards for data management, processing and dissemination** by scientific field

The first landscape put forward:

Research infrastructures that structure data management and dissemination for their scientific community



Resource centres



Training and e-training resource centre Recherche Data Gouv

Sharing and certification of **teaching tools**
along with the development and provision of
generalist e-training



Shared tools for the clusters resource centre Recherche Data Gouv

Development and maintenance of **shared tools**
for the **data management clusters**



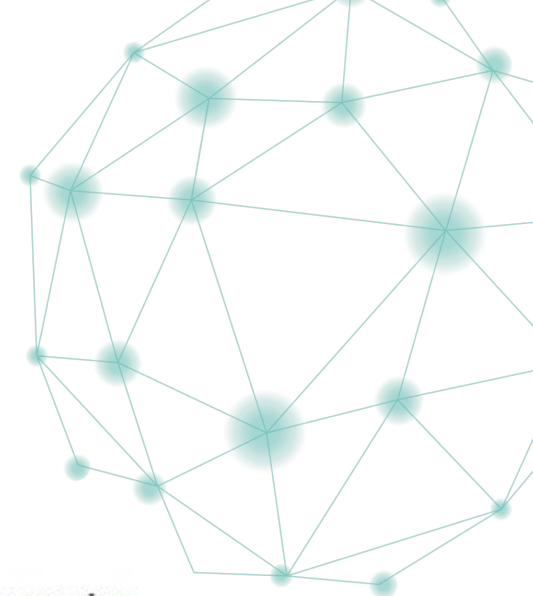
Skills resource centre Recherche Data Gouv

A system aimed at **increasing the skills of**
the **data management clusters**



Repository-registry resource centre Recherche Data Gouv

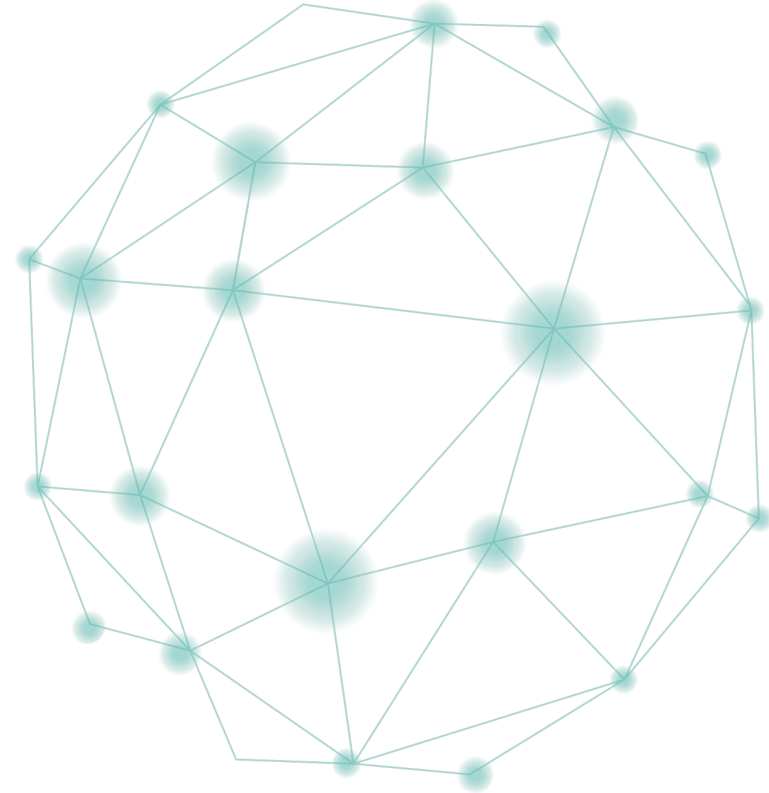
Development and maintenance of the
repository-registry along with user support



A shared data repository

for higher education and research institutions

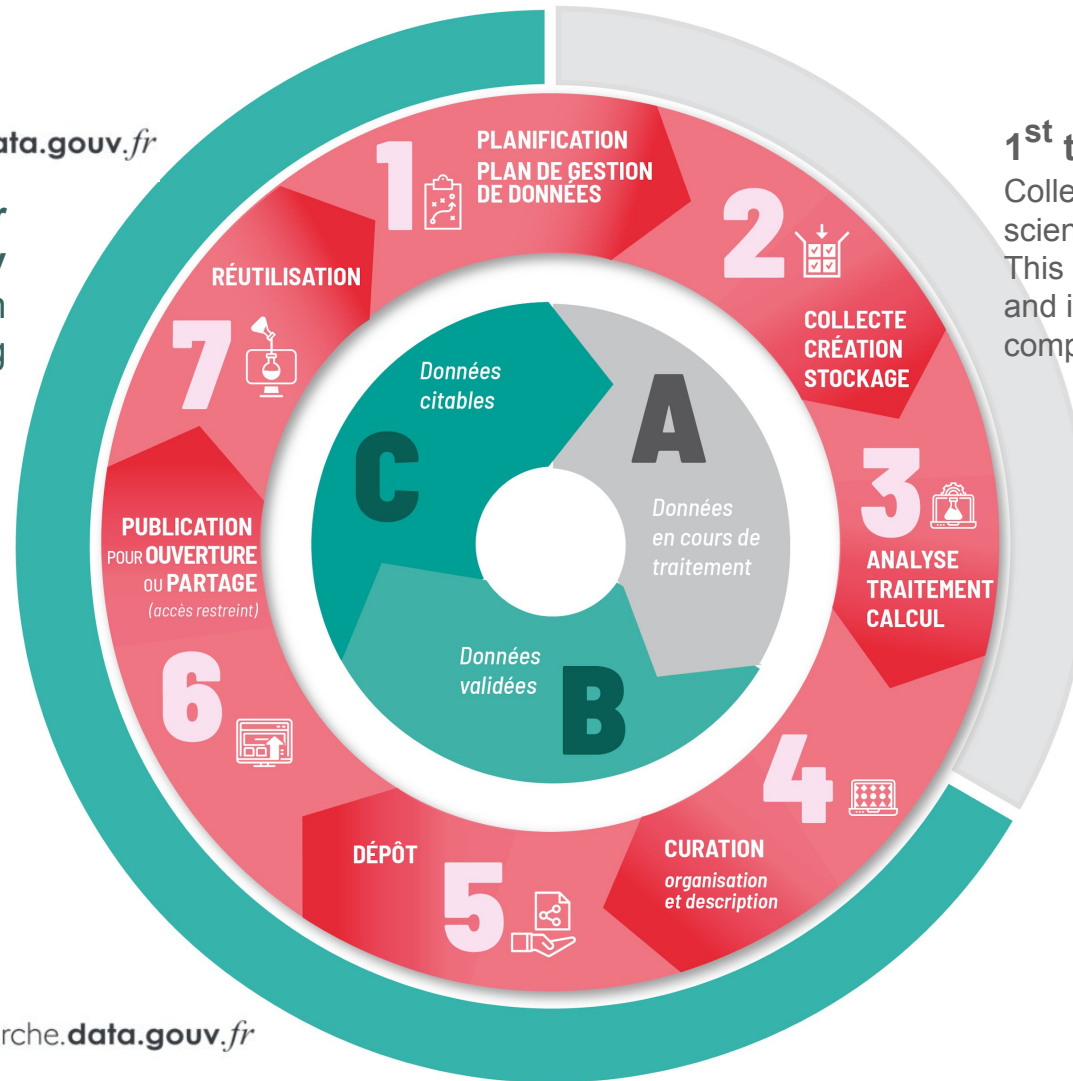
- A national offer developed and maintained by institutions at the service of the entire research community
- Research teams deposit and publish their data directly
- Spaces for institutions to manage and moderate their data
- Institutions are relieved of the cost of maintaining a technical offering
- A state-of-the-art offering
- Critical mass to achieve an international dimension





 recherche.data.gouv.fr

3rd tier
Recherche Data Gouv
exposes citable data, through
opening or sharing



1st tier

Collect, create, store and compute scientific data.
This is done within the laboratories and in collaboration with storage and computing service providers



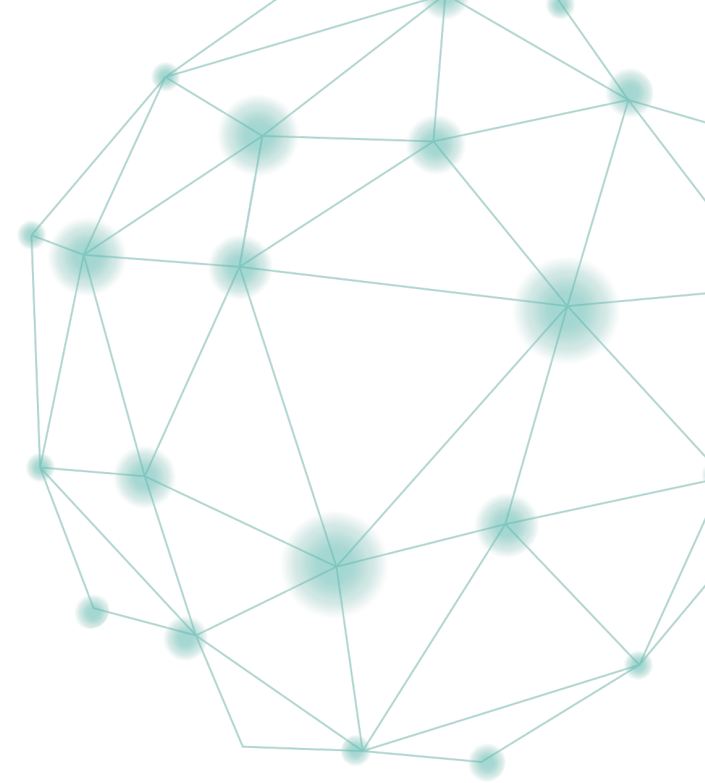


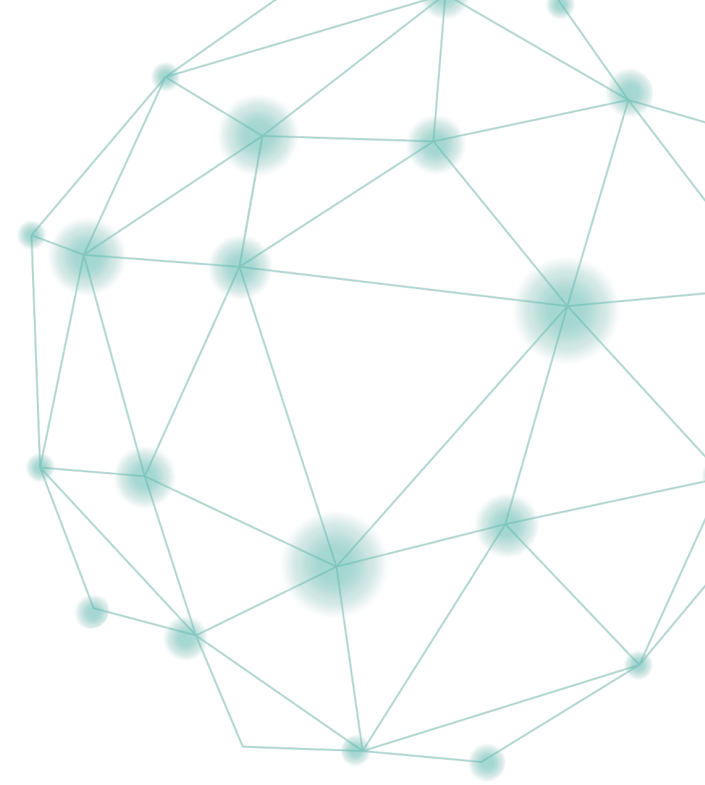
Co-construction Collective governance

- Project partners



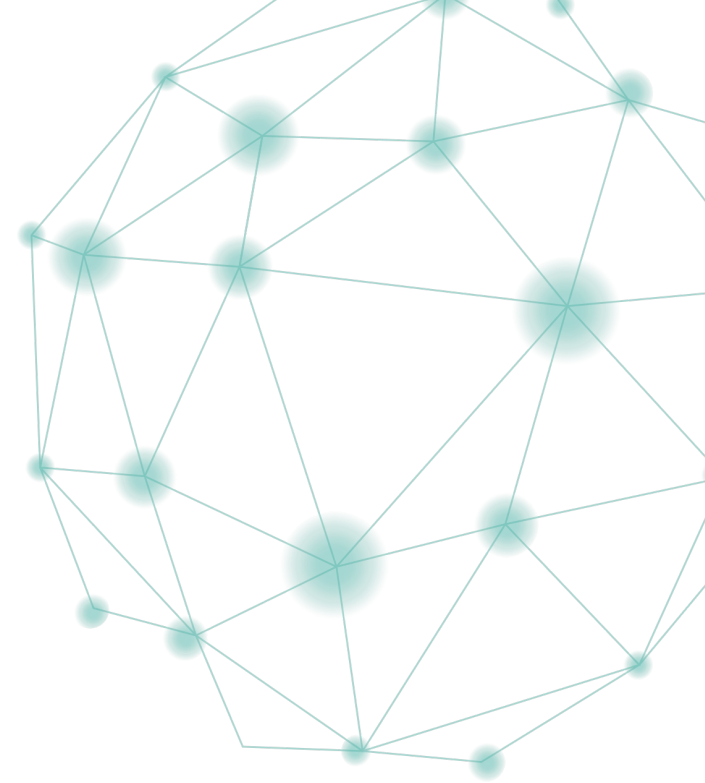
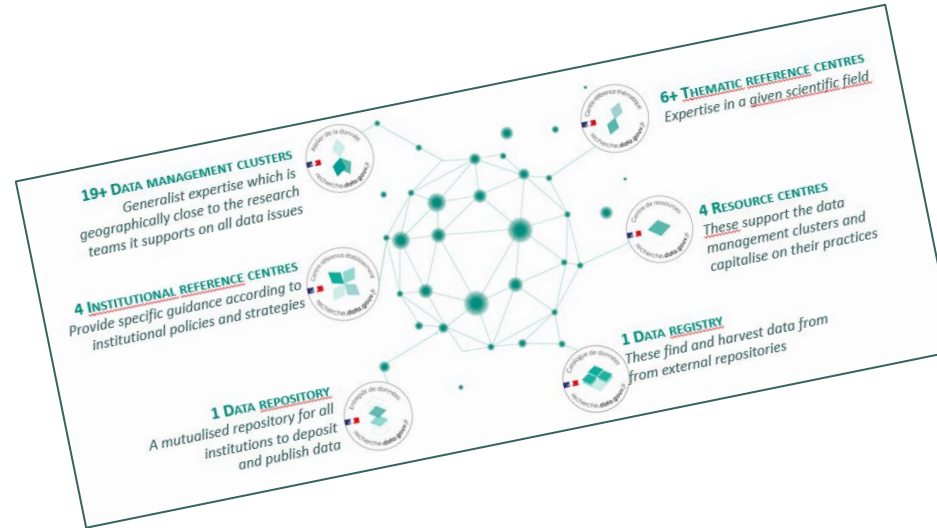
- Representatives of the higher education and research community





Lessons learnt & key to success

Diversity... or entropy?



- More than 35 different groups
- 1200+ interactions possibilities
- Hundreds of people involved
- Dozens of common topics
- Interdependent – sometimes overlapping - perimeters
- Distinct entities, each with their own policies

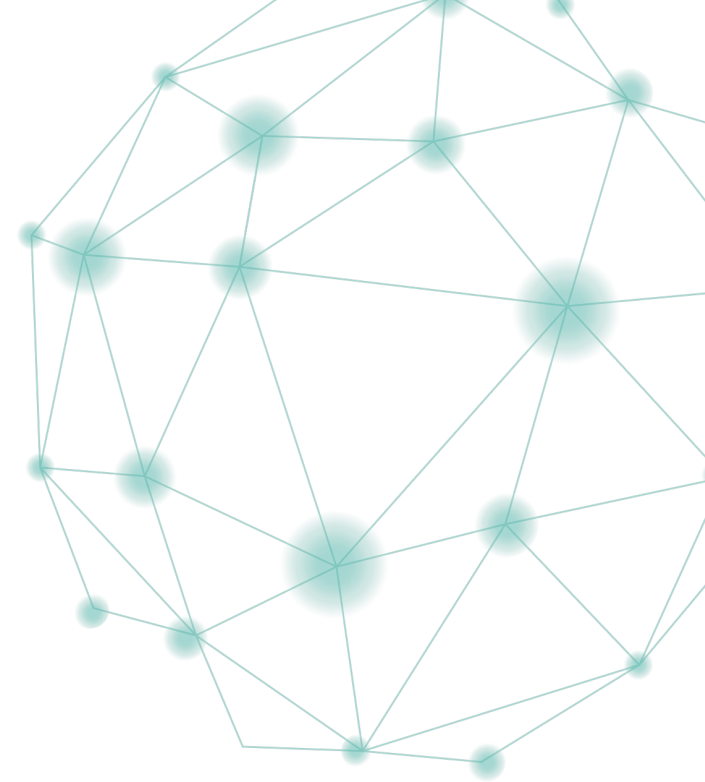
How does that work?

Example #1 : Lanscape Integration

The French Research Community didn't wait for *Recherche Data Gouv* to work on data management

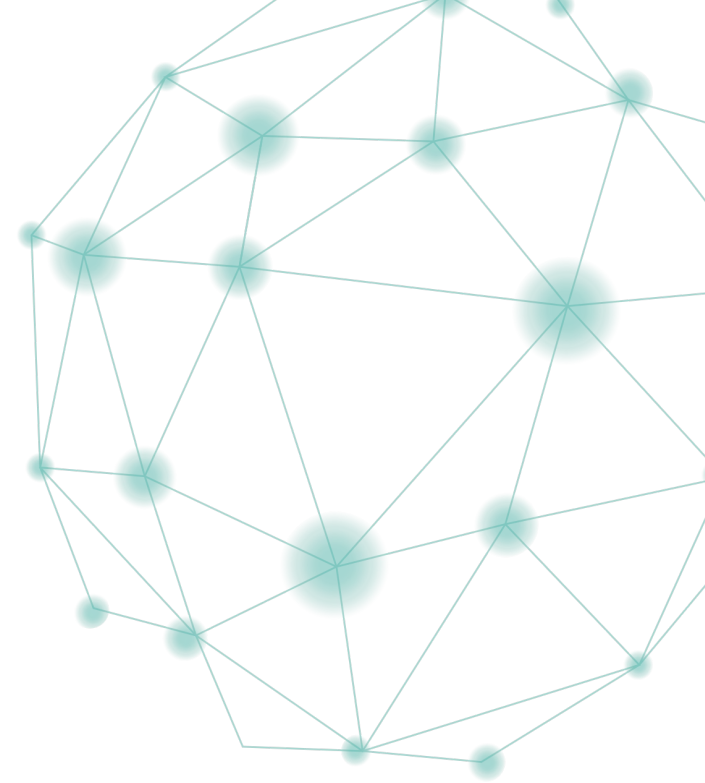
- **Thematic Reference Centres**
 - Research Infrastructures with an established service offer
- **Institutional Reference Centres**
 - Institutions that have had a support offer set up for many years
- **Resource Centres**
 - Existing structures focusing on a specific subject
- **Labelling Data Management Clusters**
 - A way of recognising an already established service

Key idea is not to do everything from scratch once again, but to allow existing structures to work better together



Example #2 : The Data Management Clusters Network

- **Within Data Management Clusters**
 - Diversity of profiles : Researchers, Teachers, Engineers, Librarians, etc.
 - Complementary competencies
 - Size heterogeneity: from a few individuals to dozens of people.
- **Data Management Clusters Network**
 - Sharing knowledge and competencies
 - Link people together and enrich interactions between the clusters
 - Ensure uniform information sharing amongst all clusters





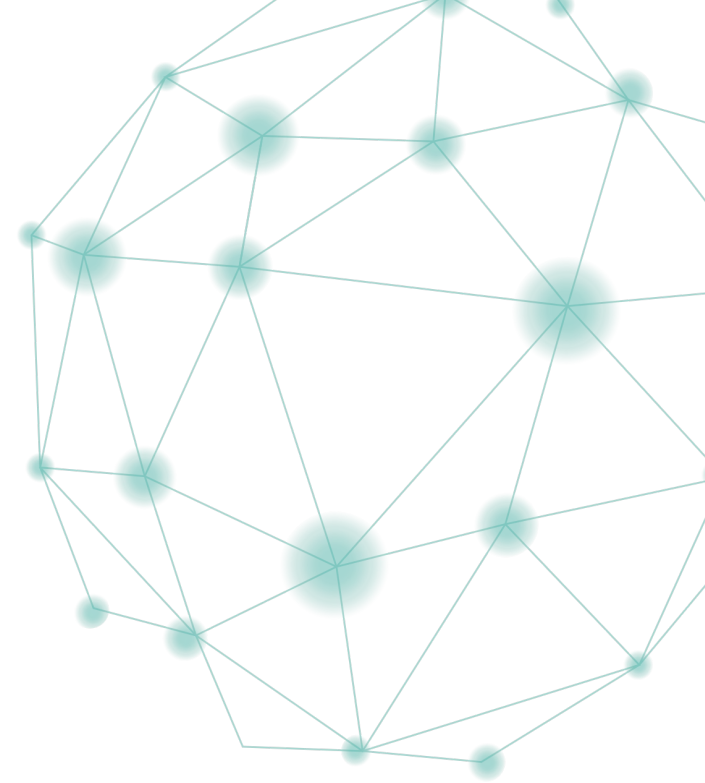
Example #2 : The Data Management Clusters Network

Different working groups

- **Objectives** : share and mutualise resources developed by the clusters, and produce new resources useful for everyone
 - WG#1 common tools for the clusters
 - WG#2 training for people staffing the clusters
 - WG#3 "Toolbox" and practical information for researchers
 - WG#4 Setting up a network of data contact point in research labs
 - WG#5 Codes and Software support
 - WG#6 links and communication between the different ecosystem modules

Example #2 : The Data Management Clusters Network

- **The network is autonomous**
 - Formed and handled by the data management clusters
 - Coordinated by a committee of 3 people rotating regularly
- **A way to facilitate interaction with the rest of the ecosystem**
 - Reference Centres de références
 - Resource Centres de ressources
 - Other stakeholders and coordinating bodies



How does that work?

3 keys to success

- **Bring together without standardising, homogenize without impoverishing**
 - Take into account what already exists
 - Capitalise on the systems in place
 - Focus on what binds things together
- **Avoid pyramids and silos**
 - Organise things transversally
 - Build a network
 - Promote bottom-up initiatives
- **Become a demonstrator of efficiency**
 - Provide a concrete and measurable service
 - Focus on common interests
 - Prefer agility



Follow the news

<https://recherche.data.gouv.fr/en>

 @RechercheDataGv

 <https://www.linkedin.com/company/recherche-data-gouv/>

Thank you!

eosc | FAIR-IMPACT



@fairimpact_eu /company/fair-impact-eu-project



Funded by
the European Union



@RechercheDataGv