DATA SUMMIT LUXEMBOURG

Innovation throughout the Data Journey

11.12.2024

European Convention Center Luxembourg

DATA
SUMMIT
LUXEMBOURG

Walking Lunch

DATA
SUMMIT
LUXEMBOURG

Welcome back!

DATA SUMMIT LUXEMBOURG

Agenda: Afternoon

13:00	**The End-to-end journey – Part 1**	13:00-14:00 Workshop: Data Protection Training
14:15	Coffee break	14:00-15:00 Workshop: Data Stewardship Training
14:45	**The End-to-end journey – Part 2**	
15:50	Coffee break	
16:10	**A look into the future**	
17:00	Networking drinks	
18:00	**End of the Data Summit 2024**	

DATA SUMMIT LUXEMBOURG

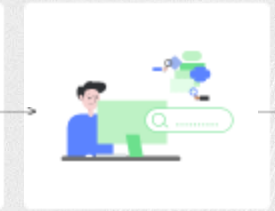
The End-To-End Journey – Part 1

DATA SUMMIT LUXEMBOURG

Data User Journey

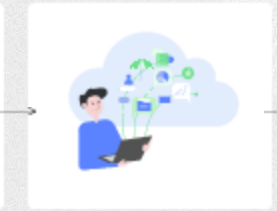
Step 0: **IDEATE**

Step 1: **DISCOVER**



Step 2: **REQUEST ACCESS**

Step 3: **ACCESS**



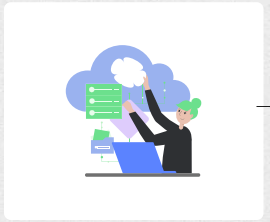
Step 4: **ANALYSE**

Step 5: **REVIEW RESULT**

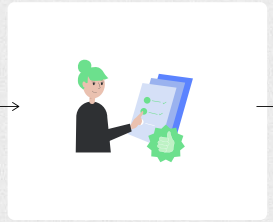
Step 6: **FINISH**

Data Provider Journey

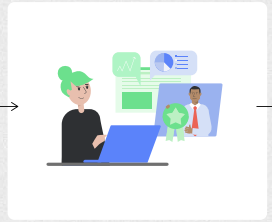
Step 0: **COLLECT**



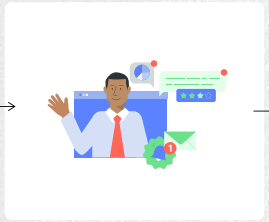
Step 1: **CATALOGUE**



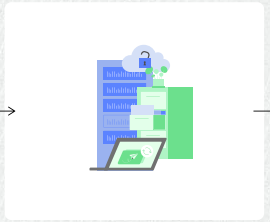
Step 2: **ENHANCE**



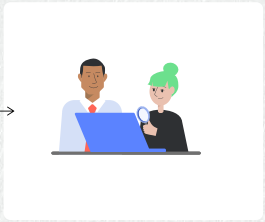
Step 3: **DECIDE ON ACCESS**



Step 4: **PROVIDE ACCESS**



Step 5: **REVIEW OUTPUT**



DATA SUMMIT LUXEMBOURG

Pinar Alper

Principal Data Steward
Luxembourg National Data Service

DATA SUMMIT LUXEMBOURG

Research Data Management

Supporting Luxembourg Institute of Science and technology (LIST) towards organisational solutions

- Data Management and Stewardship Support

- Focus on data management planning
- With state-of-the-art tools & templates

DATA SUMMIT LUXEMBOURG

Partner presentations

Supporting organisations and researchers in Data Management.

DATA SUMMIT LUXEMBOURG

Patrik Hitzelberger

Project Manager
Luxembourg Institute of Science and Technology (LIST)

SUPPORTING ORGANISATIONS AND RESEARCHERS IN DATA MANAGEMENT

PATRIK HITZELBERGER

QUALITY, DATA AND KNOWLEDGE MANAGEMENT TEAM

LUXEMBOURG INSTITUTE OF SCIENCE AND TECHNOLOGY | LIST

EVOLUTION OF RESEARCH DATA MANAGEMENT AT LIST

World

2003 — Berlin Declaration on Open Access

"Big Data"

2018/21 — Funding agencies requirements

LIST

2020 — RDM Task force /project

2023 — LIST RDM service

Supported by LNDS LUXEMBOURG NATIONAL DATA SERVICE

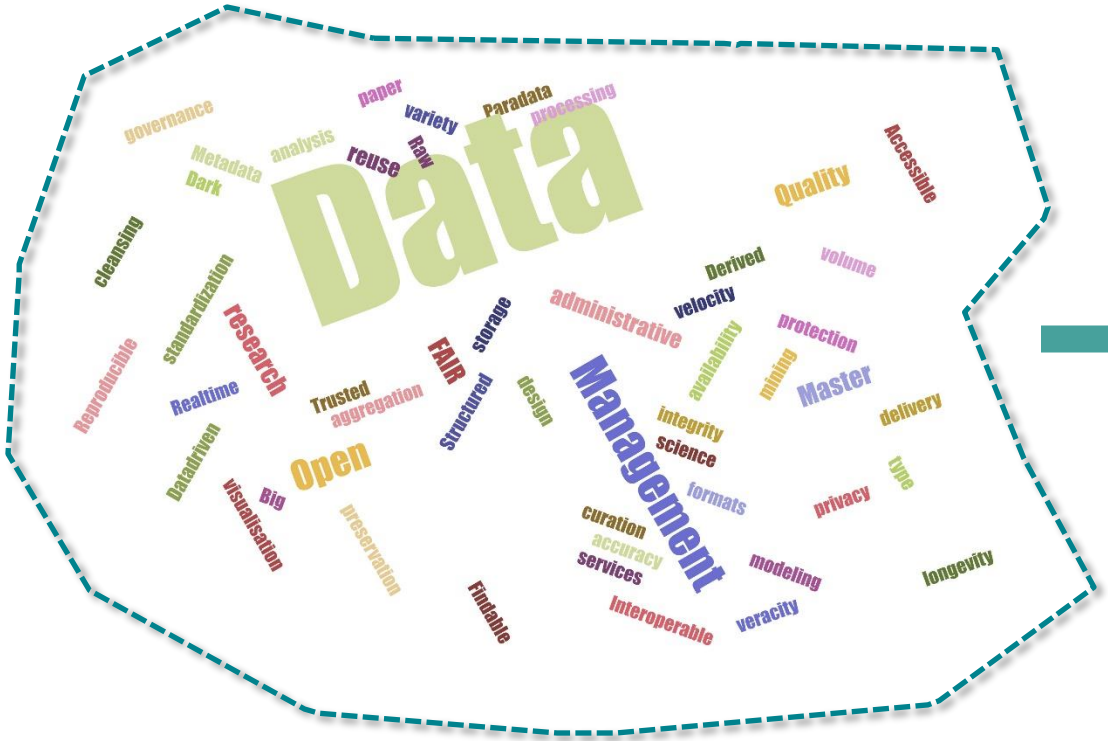
2024

Data Stewardship Wizard LIST

Datashare LIST

THE ULTIMATE OBJECTIVE

“Whatever” data is based on *managed* data



LUXEMBOURG INSTITUTE OF SCIENCE AND TECHNOLOGY LIST

LIST RESEARCH DATA MANAGEMENT

Support and tools for Researchers *and* the Organization

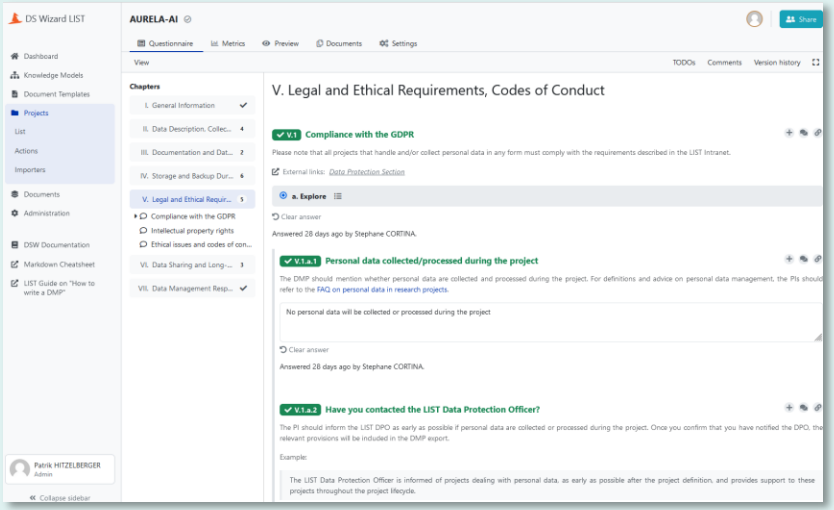
- Researcher support
- Organization Support
- Tools

EXAMPLE 1: DATA STEWARDSHIP WIZARD - LIST

Support Data Management *Planning*

- Edit DMPs based on intelligent templates
- Provide support throughout the whole Research Data Life Cycle
- Ensure standards and policies, long-term data-asset accessibility and sharing
  People leave, data stay
- Centralize all DMPs in one database

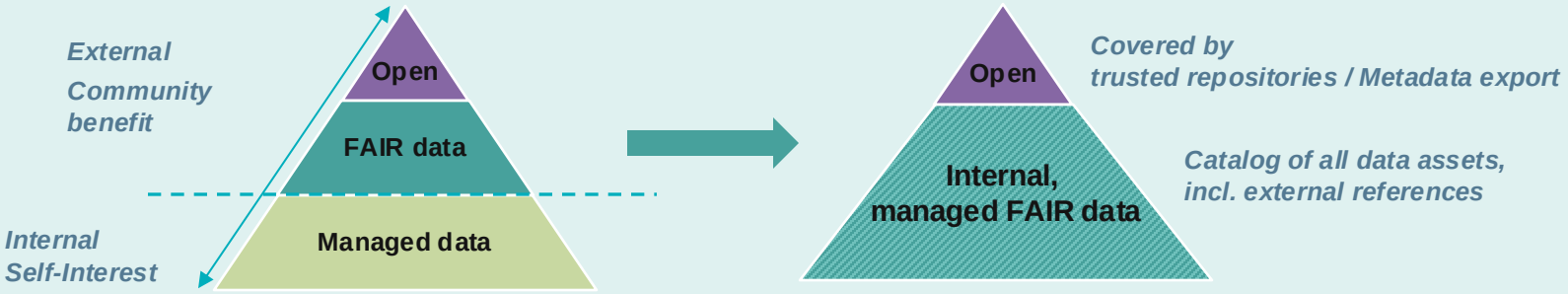
DSW



LUXEMBOURG INSTITUTE OF SCIENCE AND TECHNOLOGY LIST

EXAMPLE 2: DATASHARE LIST

Support Data *Sharing*



- LIST must persist *all* data in a FAIR manner
- Solution: Internal repository with external references : **Datashare LIST**

LUXEMBOURG INSTITUTE OF SCIENCE AND TECHNOLOGY LIST

Schema top left: Jones, S. Open, FAIR data and RDM. 2018. https://www.slideshare.net/sjDCC/open-fair-data-and-rdm. Available under Creative Commons Attribution License.

SUMMARY

LIST is moving from unmanaged data to managed FAIR data assets:

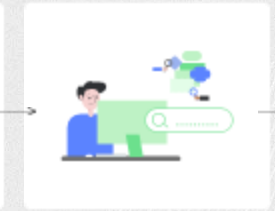
- **Commitment in the long run**
- **Individual and organizational focus with tools and policies**
- **The LUXEMBOURG NATIONAL DATA SERVICE has been a very helpful resource on this path**

DATA SUMMIT LUXEMBOURG

Data User Journey

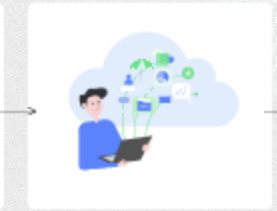
Step 0: **IDEATE**

Step 1: **DISCOVER**



Step 2: **REQUEST ACCESS**

Step 3: **ACCESS**



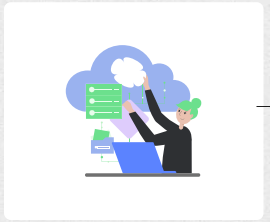
Step 4: **ANALYSE**

Step 5: **REVIEW RESULT**

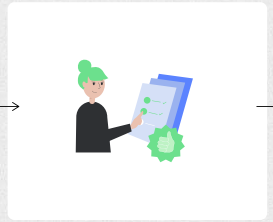
Step 6: **FINISH**

Data Provider Journey

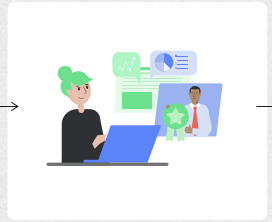
Step 0: **COLLECT**



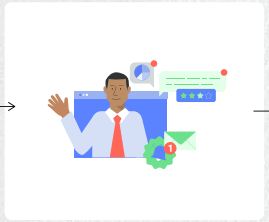
Step 1: **CATALOGUE**



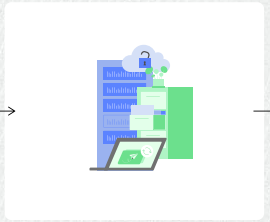
Step 2: **ENHANCE**



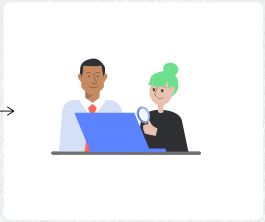
Step 3: **DECIDE ON ACCESS**



Step 4: **PROVIDE ACCESS**

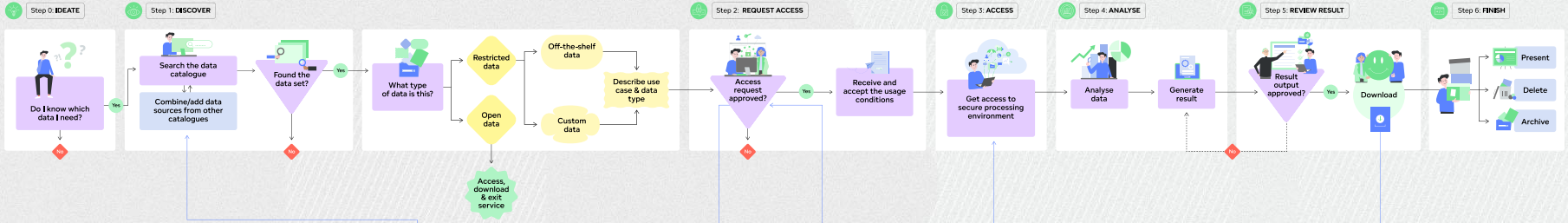


Step 5: **REVIEW OUTPUT**

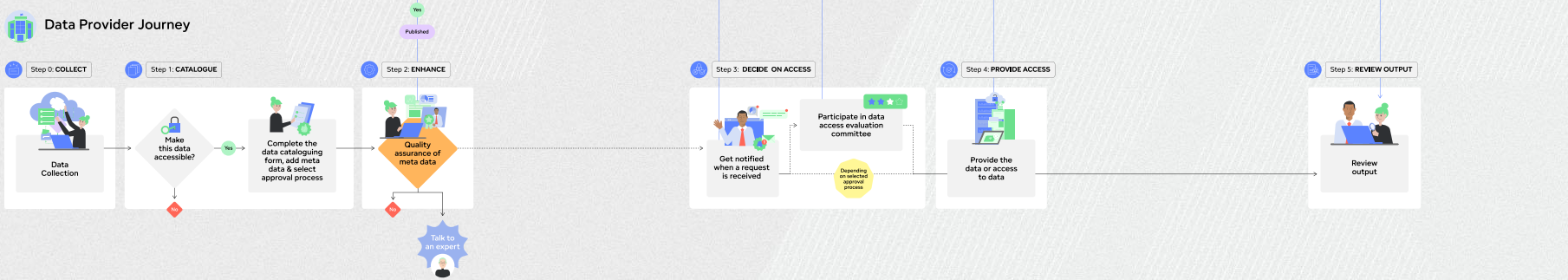


DATA SUMMIT LUXEMBOURG

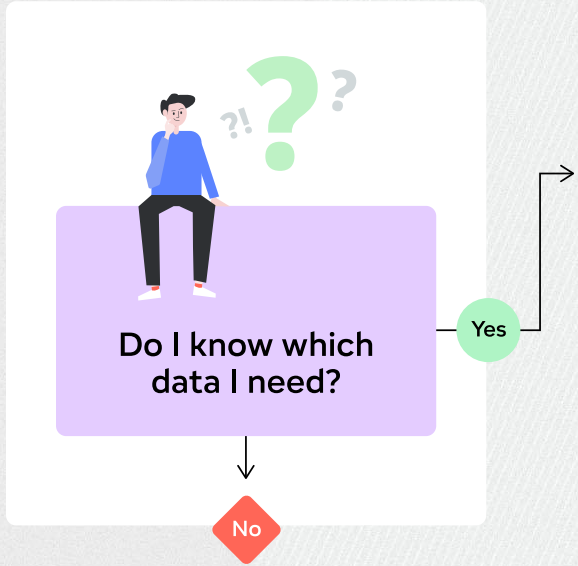
Data User Journey



Data Provider Journey



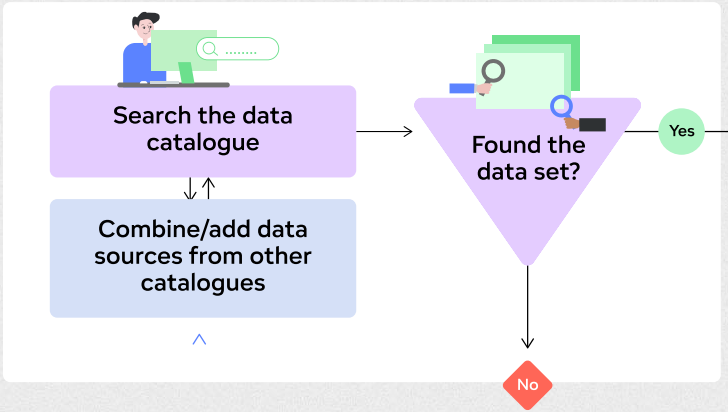
Step 0: **IDEATE**

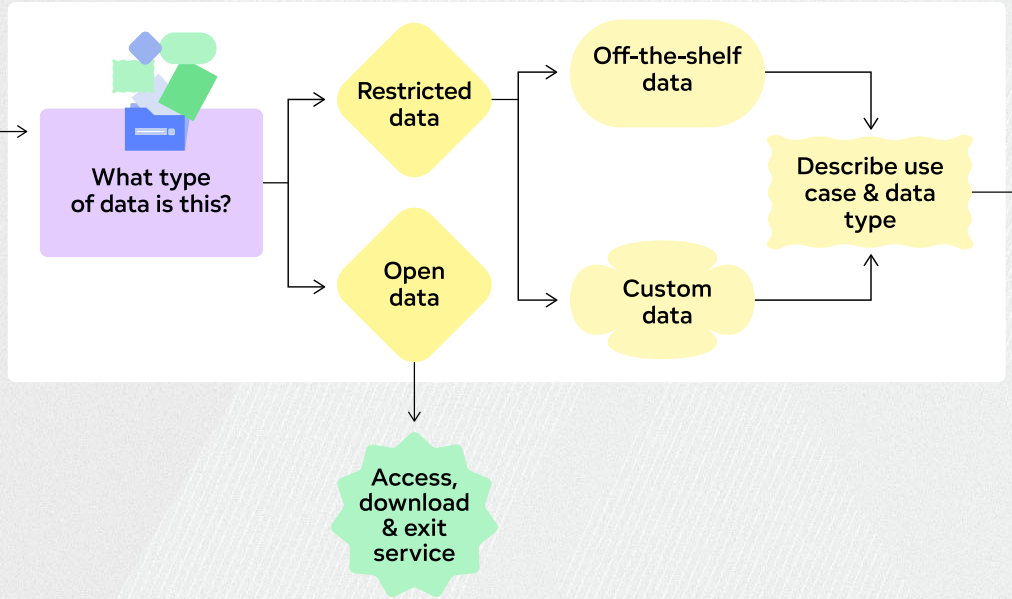


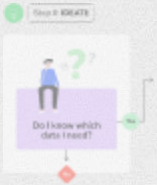
DATA SUMMIT LUXEMBOURG

Data User Journey

Step 1: **DISCOVER**







DATA SUMMIT LUXEMBOURG

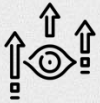
Danielle Welter

Principal Data Scientist
Luxembourg National Data Service

Data cataloguing – the first step to effective data reuse

Secure internal data with governance best practice

Facilitate data discovery through public metadata



Enhance transparency

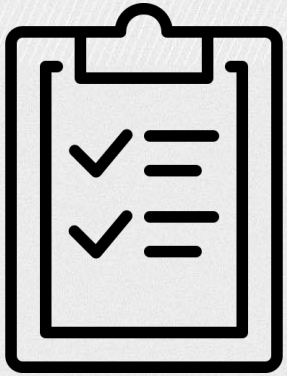
Foster innovation

DATA SUMMIT LUXEMBOURG

Data Cataloguing

Supporting the "Administration de l'Environnement"'s internal data inventorying for downstream metadata publication

- Data Cataloguing



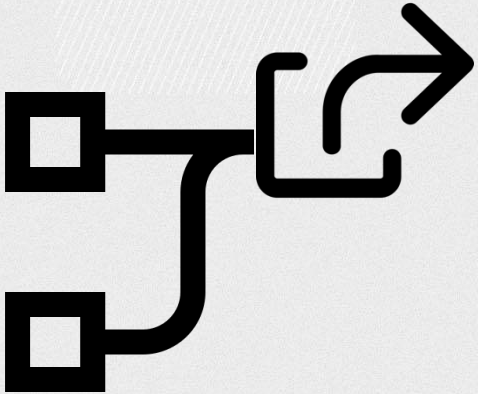
- Data inventory & cataloguing guidance
- Metadata model & process development

DATA SUMMIT LUXEMBOURG

Data Cataloguing

Leveraging existing data cataloguing efforts at the "Direction de la Santé" to facilitate metadata publication

- Data Cataloguing



- Curating internal metadata to match standards for public data catalogues

DATA SUMMIT LUXEMBOURG

Partner presentations

Unlocking insights: Crafting our administration's data catalogue

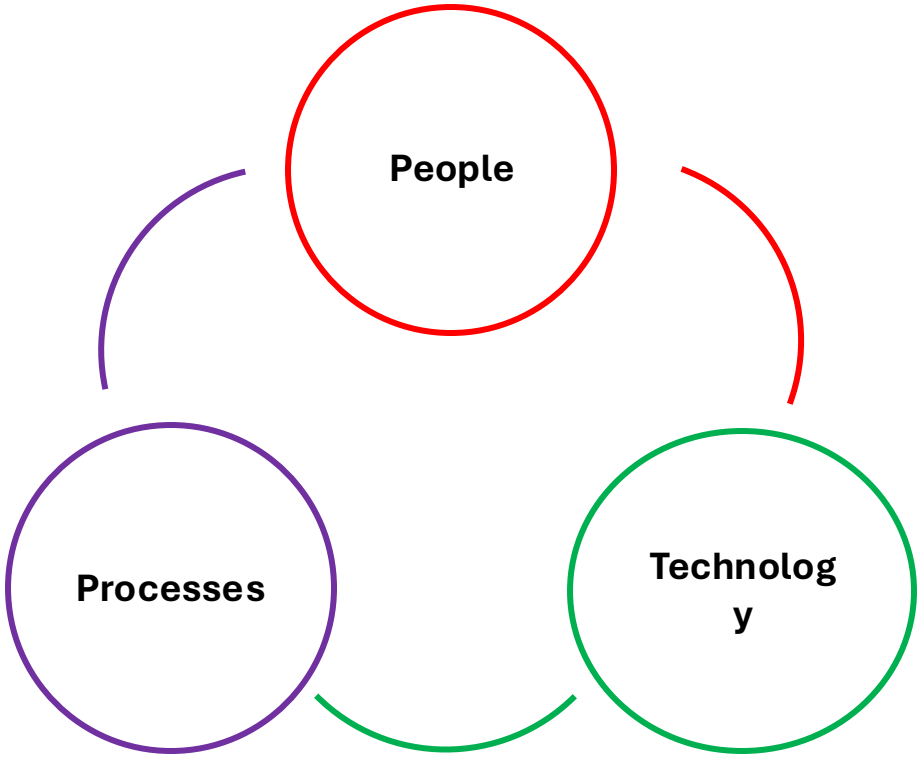
DATA SUMMIT LUXEMBOURG

Isabelle Naegelen

Responsable Système d'agréments &
Gouvernance de données
Administration de l'Environnement (AEV)

Introduction

- Goal
  - Improve data management for datasets at the Environment Agency, enabling structured access
- Scope
  - Building on 2023 pilot project: refining dataset definitions, creating workflows, and expanding metadata capture tool



Core Data Source Analysis

- Initial 66 narrowed to **27 relevant data sources** for cataloguing

What is the primary purpose of the data source?

Is the data primary data or derived from other data sources?

Does the data have links to other data sources?

Are there any defined vocabulary lists?

How many users does the data source have?

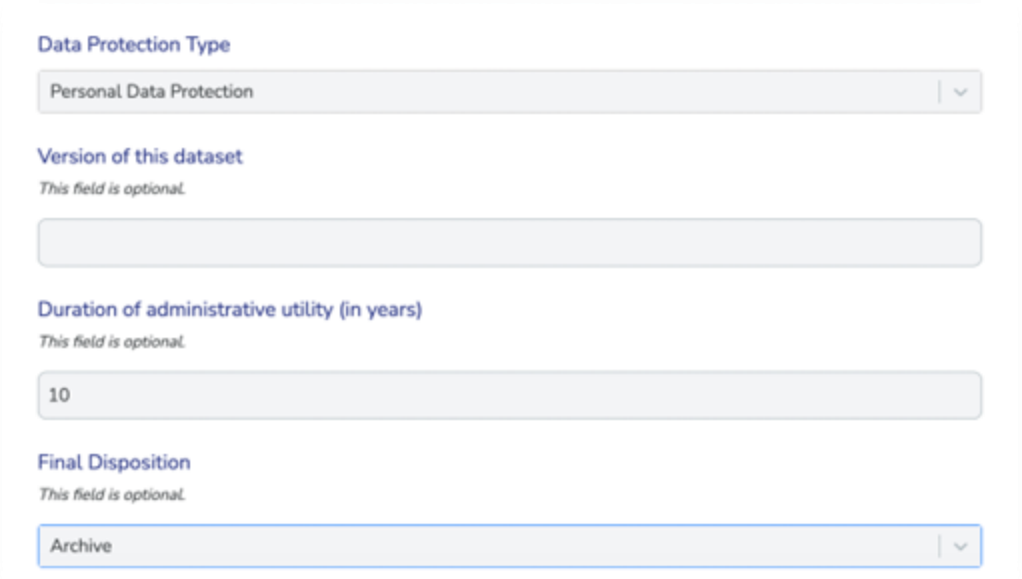
Does the data source constitute a “single dataset” or does it contain multiple datasets?

Workflow Design

- Adapted from AEV's financial approval procedures
- Approval flow stages
    1. Initial metadata entry by dataset expert
    2. Review and validation by group/department manager
    3. Leadership team review for publishing
    4. Publication through appropriate channels
- Broad departmental validation with minor adjustments based on feedback

Metadata Model Review

- Metadata model enhancements
  - Additional fields for areas of competence, data archiving details, and data usage duration

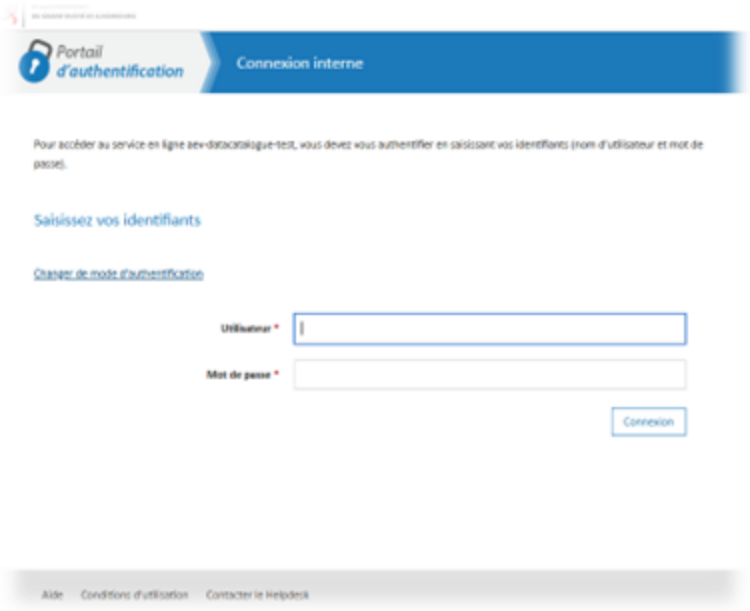


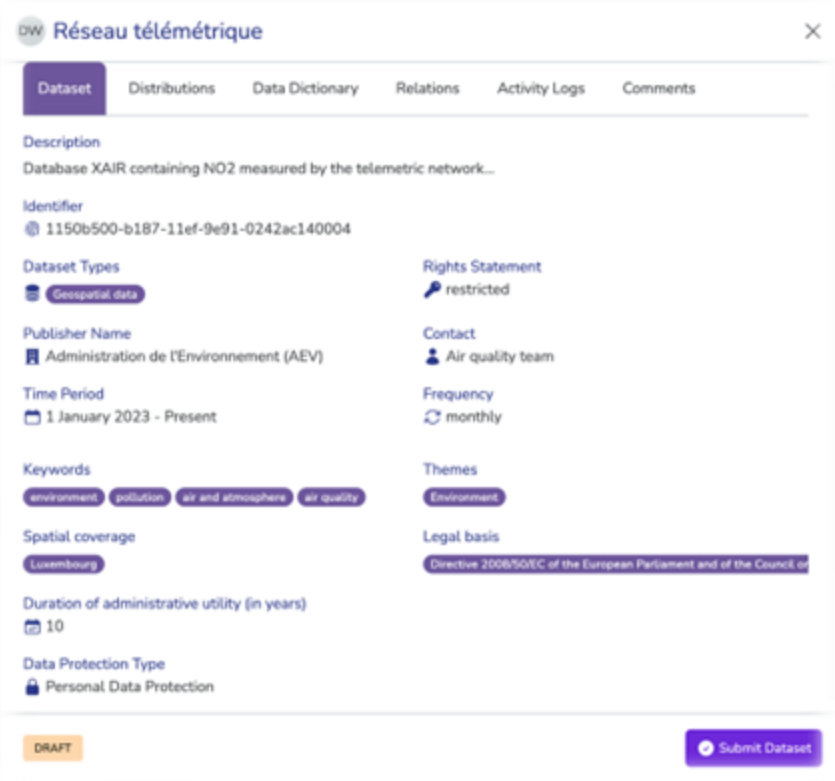
Dataset Workshops

- Purpose
  - In-depth workshops to refine dataset inventory, engage departments
- Structure
  - **9 workshops** covering dataset context, data processing, and feedback on approval flows
- Outcome
  - Successful engagement, leading to comprehensive inventory of **200+ datasets & sub-datasets** with standardised language and vocabulary

Metadata Capture System

- Expansion of the AEV-DCA pilot metadata capture system
  - Integrated OpenID connect (OIDC) for secure authentication
  - Activity logs for easy monitoring and reviewing of changes
  - Multiple dataset types
  - Rights statements at the distribution level
  - Grouping/nesting of keywords
  - Roles and status levels




DW
Réseau télémétrique
Dataset
Distributions
Data Dictionary
Relations
Activity Logs
Comments
Description
Database XAIR containing NO2 measured by the telemetric network...
Identifier
1150b500-b187-11ef-9e91-0242ac140004
Dataset Types
Geospatial data
Rights Statement
restricted
Publisher Name
Administration de l'Environnement (AEV)
Contact
Air quality team
Time Period
1 January 2023 - Present
Frequency
monthly
Keywords
environment
pollution
air and atmosphere
air quality
Themes
Environment
Spatial coverage
Luxembourg
Legal basis
Directive 2008/50/EC of the European Parliament and of the Council o
Duration of administrative utility (in years)
10
Data Protection Type
Personal Data Protection
DRAFT
Submit Dataset


IS
Cartographie stratégique du bruit - Grands axes routiers 2021
Dataset
Distributions
Data Dictionary
Relations
Activity Logs
Comments
Description
Les cartes stratégiques du bruit fournissent un inventaire de l'environnement sonore autour des infrastructures couvertes par la directive européenne 2002/49/CE du 25 juin 2002 relative à l'évaluation et à la gestion du bruit dans l'environnement : les grands axes routiers (de plus de 3 million... View more
Identifier
3bf84170-70fe-11ef-809f-0242ac120003
Dataset Types
Geospatial data
High-value dataset
Rights Statement
public
Publisher Name
Administration de l'environnement
Contact
Isabelle Naegelen
Time Period
1 January 2021 - 31 December 2021
Frequency
quinquennial
Keywords
environment
Themes
Environment
Spatial coverage
Luxembourg
APPROVED
2
Revert to Draft
Publish Dataset
Mark as Internal

Conclusion & Outlook

- Built a comprehensive data inventory, defined metadata approval procedures, enhanced metadata model

- Next steps
  - Implementation of metadata capture system, and procedures

DATA SUMMIT LUXEMBOURG

Partner presentations

From Data Chaos to Clarity: Implementing a Data Catalogue Strategy

DATA SUMMIT LUXEMBOURG

José Andrés Garcia
Head of Data & Analytics / Service IT
Pôle support à l'innovation,
Luxembourg Health Directorate

Meriem Duhamel
Data Engineer / Service IT
Pôle support à l'innovation,
Luxembourg Health Directorate

Initial situation

LE GOUVERNEMENT
DU GRAND-DUCHÉ DE LUXEMBOURG
Ministère de la Santé
et de la Sécurité sociale
Direction de la santé

- **Context:**
  - In 2021, the datateam at DiSa was created.
- **Initial situation :**

  - Many data sources and formats
  - Tooling disagrement
  - Low data quality, low trust
  - Siloed data management, multi-criteria
  - Data untraceable
  - Guarantee data protection, legislation, compliance, standards….

Challenges

LE GOUVERNEMENT
DU GRAND-DUCHÉ DE LUXEMBOURG
Ministère de la Santé
et de la Sécurité sociale
Direction de la santé

Data Governance

Raising Awareness & Building a Data-Driven Culture

Choose a data catalog solution

Technical complexity (data security, schema access ..)

Establish exhaustive contact with business users, in order to:

Discover existing data sources within the departments
Identify data owners and stewards
… and, in general, understand the what, why, and how of the data.

Approach

LE GOUVERNEMENT
DU GRAND-DUCHÉ DE LUXEMBOURG
Ministère de la Santé
et de la Sécurité sociale
Direction de la santé

- Data Strategy Set-up
- Collaboration to carry out the first data governance policy

- Define the scope of the data catalog
- Populate Data Glossary/Dictionary/Catalog
- Automatize the process

- Unlock the Power of Data Lineage
- Address Data Quality Issues
- Promote a data-driven culture

Benefits

LE GOUVERNEMENT
DU GRAND-DUCHÉ DE LUXEMBOURG
Ministère de la Santé
et de la Sécurité sociale
Direction de la santé

- Good communication with business for the follow-up

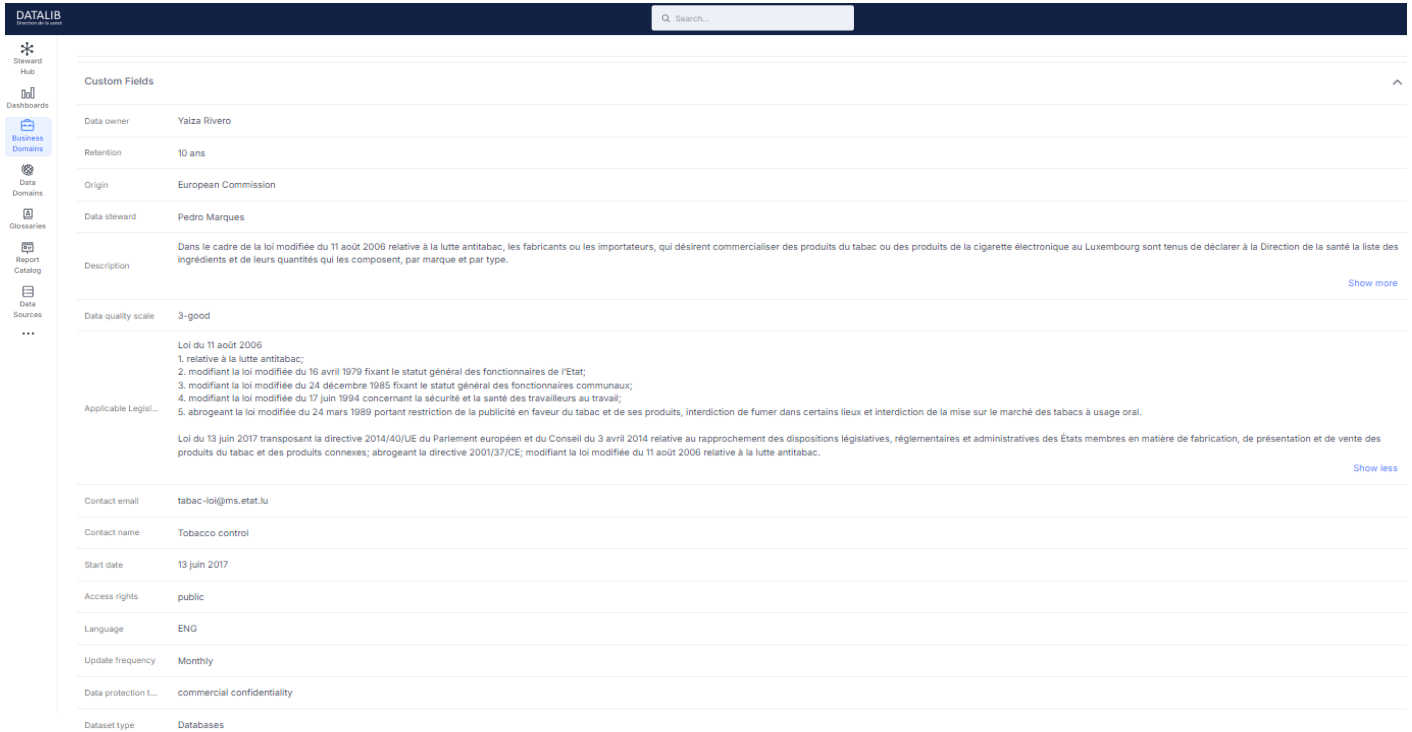
- Clear view of data accountability
- Working towards better data quality
- RCA & Impact analysis of data actions

- Metadata sharing with LNDS , IGSS, ObSanté…

DataCatalog

DATALIB
Direction de la santé

LE GOUVERNEMENT
DU GRAND-DUCHÉ DE LUXEMBOURG
Ministère de la Santé
et de la Sécurité sociale
Direction de la santé



Still a long journey

LE GOUVERNEMENT
DU GRAND-DUCHÉ DE LUXEMBOURG
Ministère de la Santé
et de la Sécurité sociale
Direction de la santé

Keep up with infrastructure

Data security management

Historical data management

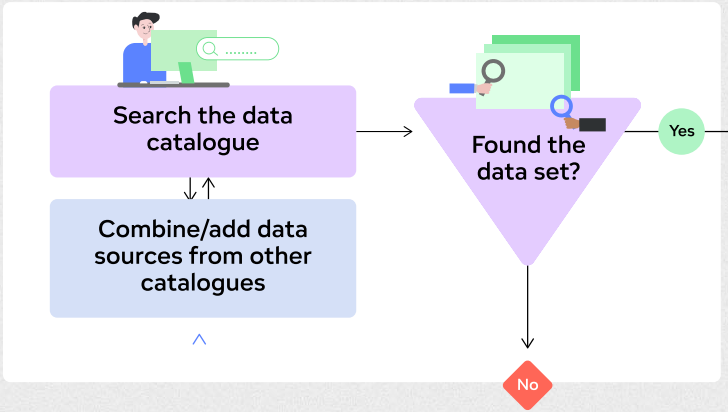
Data quality planning and trust

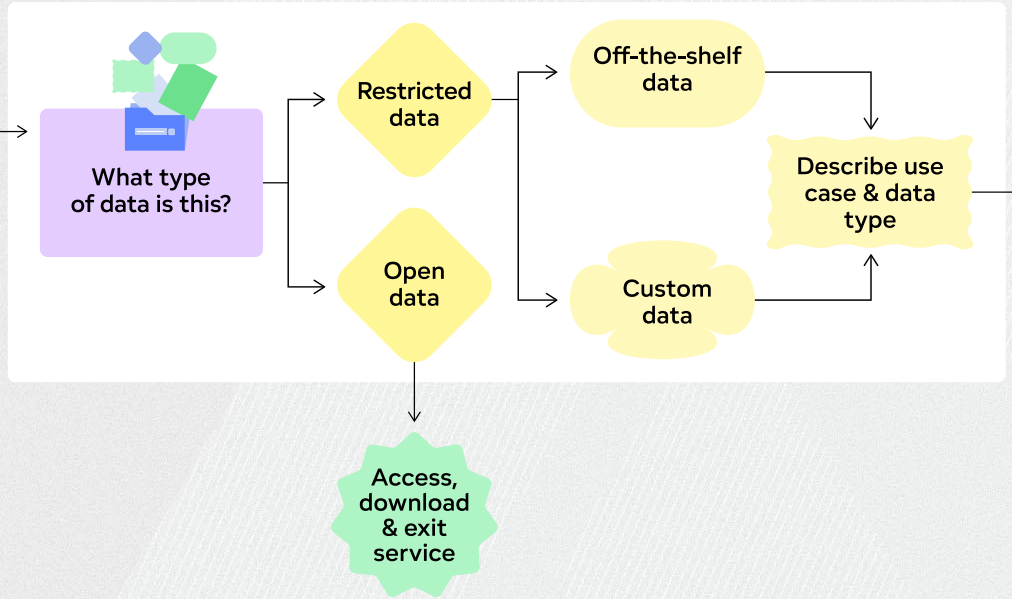
Data LifeCycle management

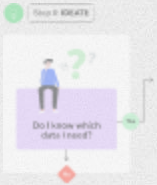
DATA SUMMIT LUXEMBOURG

Data User Journey

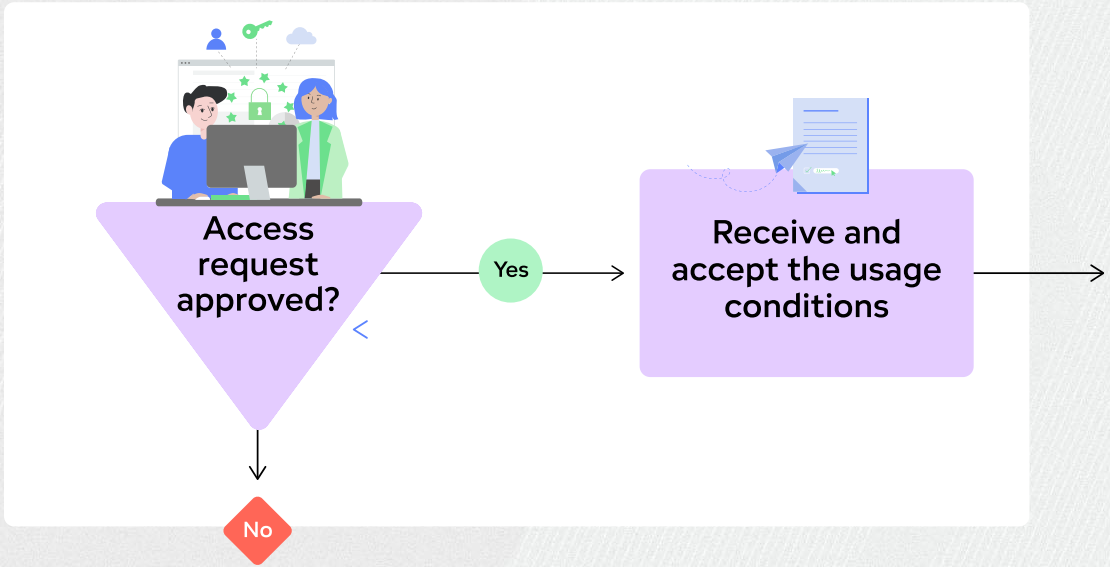
Step 1: **DISCOVER**

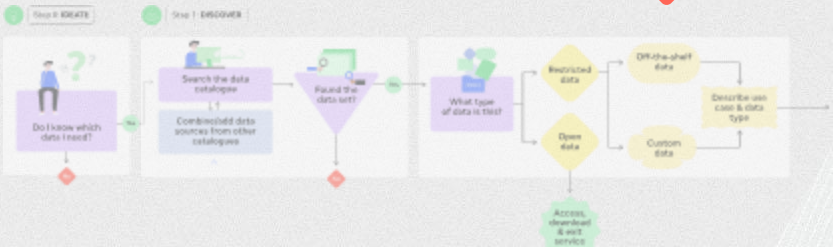






Step 2: **REQUEST ACCESS**



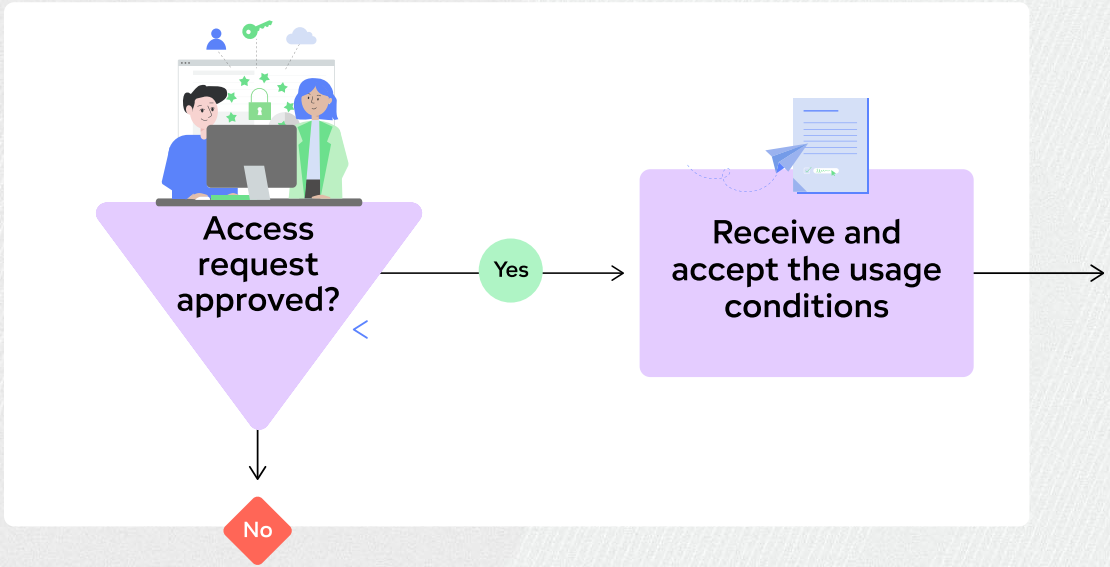


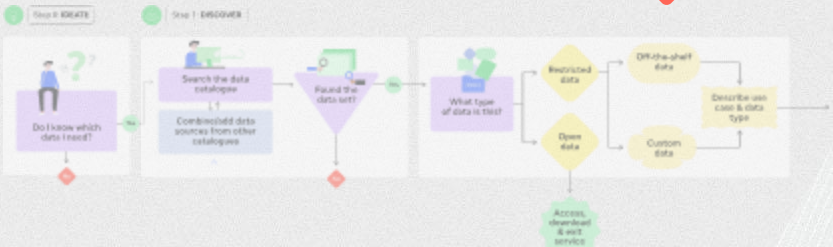
DATA SUMMIT LUXEMBOURG

Pinar Alper

Principal Data Steward
Luxembourg National Data Service

Step 2: **REQUEST ACCESS**





DATA SUMMIT LUXEMBOURG

Obtaining Data Access

Supporting LISER in the XBorder Project

- Access Request Optimisation

- Innovative solution to link data from multiple countries
- Develop input for
  - Access request form
  - Data protection concept

DATA SUMMIT LUXEMBOURG

Partner presentations

Challenges in Requesting Data Access

DATA SUMMIT LUXEMBOURG

Andrea Albanese

Research Scientist, Labour Market
Luxembourg Institute of Socio-Economic Research (LISER)

LISER
LUXEMBOURG INSTITUTE OF SOCIO-ECONOMIC RESEARCH

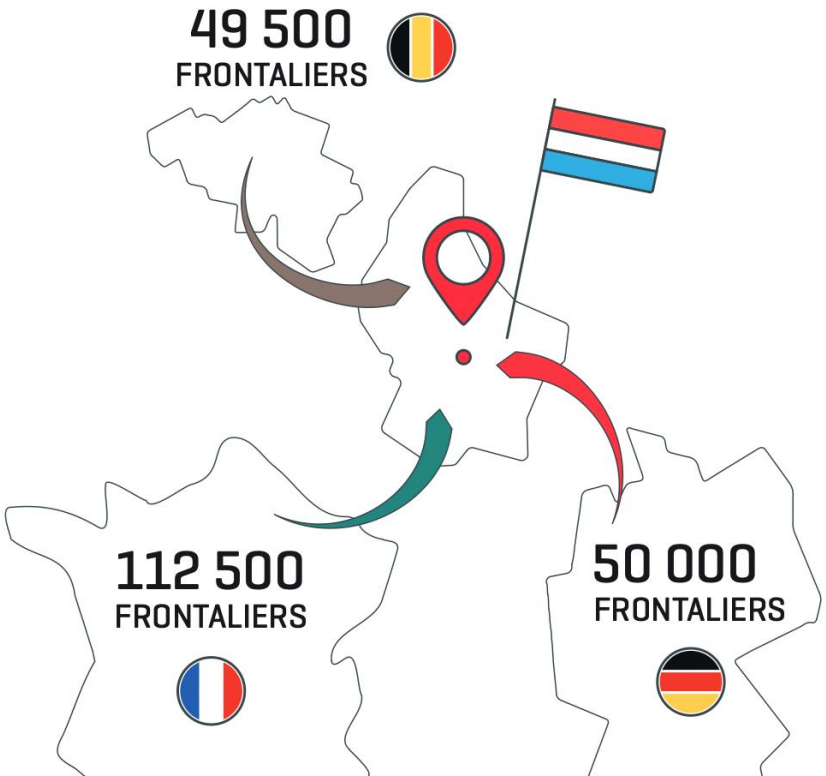
Data Access: Navigating the Complexities

Andrea ALBANESE

Research Scientist
Labour Market Department
Luxembourg Institute of Socio-Economic Research
11 December 2024

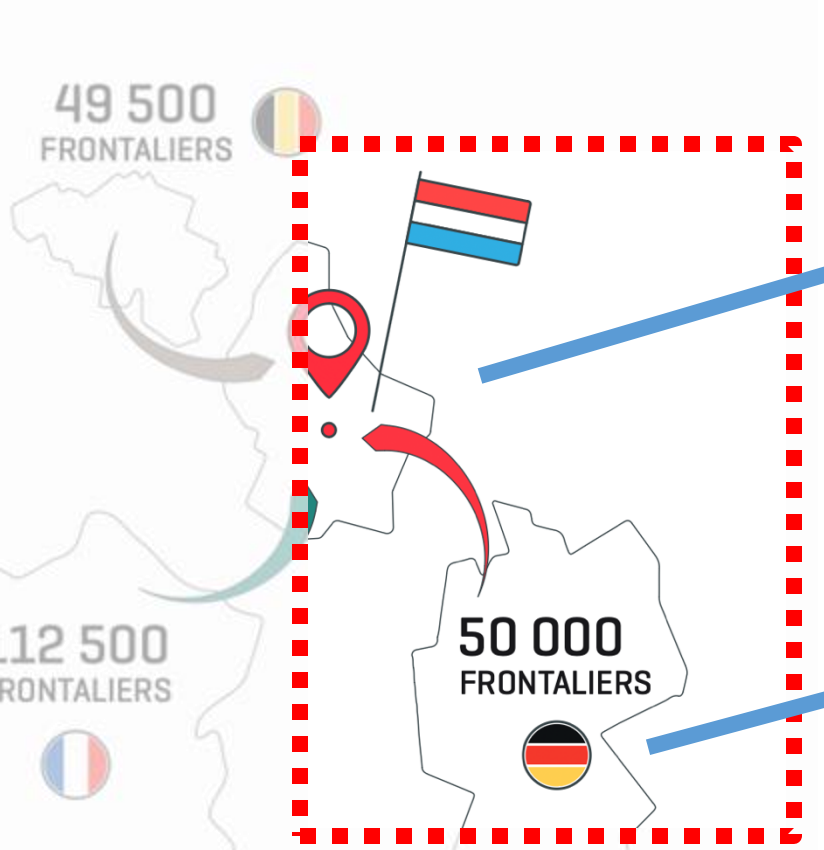
Data Summit 2024, ECCL Luxembourg

A unique opportunity to study **international labor mobility**



Luxembourg is the
1st destination in EU
for cross-border workers

The cross-border project



Le gouvernement luxembourgeois

Inspection générale de la sécurité sociale

?

IAB INSTITUTE FOR EMPLOYMENT RESEARCH

The Research Institute of the Federal Employment Agency

Foreseen process – Where are we ?

STEP 1
Find matching pseudonyms
for the population of interest

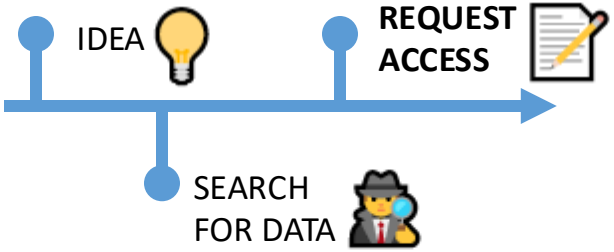
1. IGSS and IAB send encrypted pseudonyms to LNDS
2. LNDS matches encrypted pseudonyms by proximity
3. LNDS provides IGSS and IAB a linkage key for connecting the data

STEP 2
Analyse pseudonymised data
In a secure processing environment

4. IGSS and IAB create final datasets and send to Secure Processing Environment (SPE)
5. LISER analyses data in Secure Processing Environment (SPE)

Foreseen process – Where are we ?

STEP 0
Data Discovery & Access Request

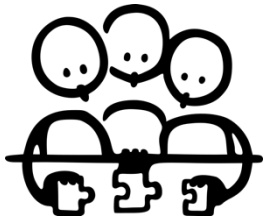


STEP 1
Find matching pseudo
for the population of i

1. IGSS and IAB pseudonyms to LN
2. LNDS matches enc pseudonyms by pr
3. LNDS informs IGSS of which subset to data

Challenges in obtaining data access

- At LISER, form filling is always a **collective endeavour**



  - Researchers & data office collect inputs from the legal experts, the IT and the DPO
  - Form sign-off is by the organizational signing official (CEO)

- Full compliance with provider requirements is key:

  - IT infrastructure
  - login security
  - safe-rooms / sharing offices
  - screen filters, etc

Trade-off: preserving anonymity vs project feasibility

- Access requests can be a very **time-consuming** exercise

- Finding the right compromise between privacy and usefulness of data is a tricky balancing act

Trade-off: preserving anonymity vs project feasibility

- Obtaining access takes months, maybe even year(s)
  - Lengthy processing of the application and data preparation
  - May require several iterations between the different stakeholders involved
- Data provider requests justification to ensure anonymity of individuals
  - Some variables (e.g. health information) are more sensitive than others
  - Crossing all the variables should not lead to too small cells: variables often aggregated
- The anonymization process may enter in conflict with the research methods used for causal evaluation in observational studies where details are needed
  - E.g. many policies rely on thresholds fixed by the policymaker to define individual eligibility to the policy
    - ➢ e.g. being born after an exact date, having an income below an exact threshold, living in a certain neighbourhood etc…
    - ➢ Methods for causal inference rely on these details to implement credible analysis e.g. Regression Discontinuity Design

What goes into an access application form

ACCESS APPLICATION FORM

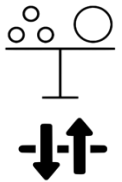
- *Institutional contacts*
- *Project proposal: objectives and methods*
- *Information on prospective data users*
- *Data access expiry date*
- *Description of (micro) data requested*
  - *Sample size and design*
  - *Variables requested and justifications*
  - *Aggregation for anonymisation and new variable creation*
  - *Addition of extra variables provided by the researchers*
  - *Desired delivery times and planned updates*

What goes into an access application form

- Data protection compliance
  - Data storage: technical and organizational data protection measures
  - Data privacy impact assessment (where necessary)
- Data access can occur in various ways, so data protection measures differ
  - Often rely on a secure processing environment
  - Data access for end-user:
    - A safe room in the institute with regulated access
    - If not, often extra security measures e.g. screen filters, only certain IP
    - More sensitive data: access only at the data provider offices (which can be abroad)

What can LNDS do?

- Navigate the process, fulfill application requirements
- Improve project proposals
  - Anonymity vs utility trade off
  - Innovative solutions to complex cases
  - Stepping in as partner with secure environment
- Navigate the legal and data protection requirements



- Help partners "pilot" projects of complex cases that can be re-usable

Thank you very much

Andrea.Albanese@liser.lu

Luxembourg Institute of Socio-Economic Research

DATA SUMMIT LUXEMBOURG

Basheer Kalash

Senior Data Scientist
Luxembourg National Data Service

DATA SUMMIT LUXEMBOURG

Synthetic Census Data

Improve accessibility to official statistics

- Synthetic Data Factory

✓ Assessed methods and tools used by national statistical offices

✓ Recommended a method and tool

🔧 Currently, implementing on STATEC 2021 census data

LNDS
LUXEMBOURG NATIONAL DATA SERVICE

STATEC

DATA SUMMIT LUXEMBOURG

Partner presentations

Synthetic Data: Accelerating Innovation While Ensuring Privacy

DATA SUMMIT LUXEMBOURG

Claude Lamboray

Responsable Méthodologie
STATEC

Synthetic Data: Accelerating Innovation While Ensuring Privacy

Data Summit Luxembourg

11.12.2024

Claude Lamboray

STATEC

Objective

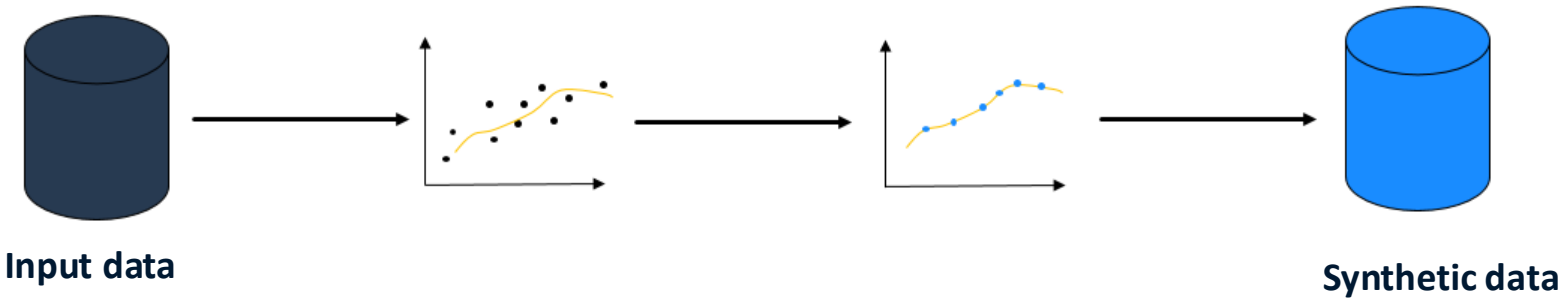
- STATEC produces and disseminates official statistics

- STATEC guarantees the non-disclosure of confidential data
  - Data are published in an aggregated form
  - Access to microdata for scientific purposes within STATEC premises

- **Can we enhance user access to more granular data while preserving privacy?**

Synthetic data for official statistics

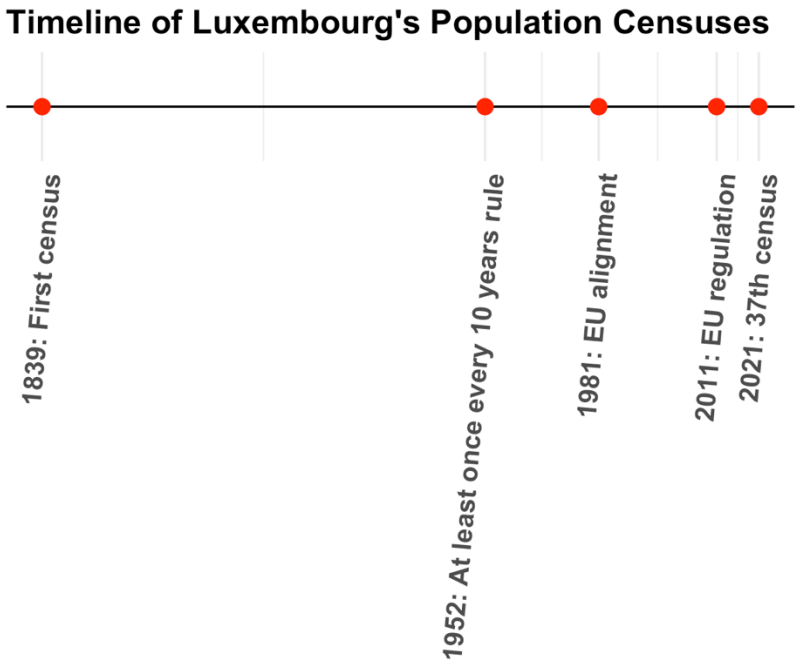
- Synthetic data:
  - mimic the properties of the initial data set
  - while preserving privacy



Synthetic data for official statistics

- Use cases:
  - 👥 Releasing synthetic microdata to the public
  - 📊 Testing analysis
  - 📘 Education
  - 💻 Testing technology

Census data

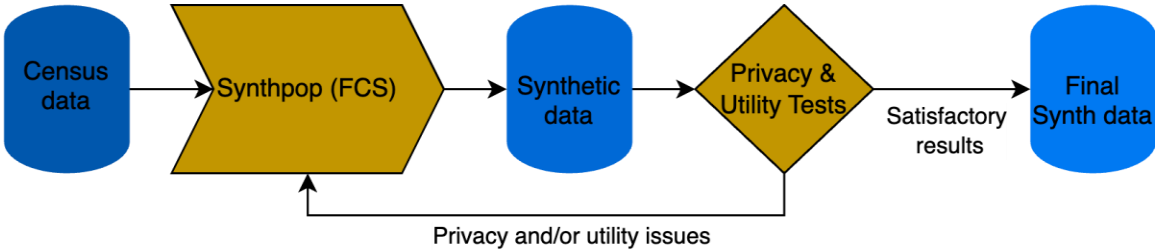

Timeline of Luxembourg's Population Censuses
1839: First census
1952: At least once every 10 years rule
1981: EU alignment
2011: EU regulation
2021: 37th census


STATISTIQUES.lu
Data
News
Publications
Surveys
About
Census
Census
01
History and overview
02
Evolution of the population
03
Territorial distribution
04
Distribution per km²
05
Nationalities
06
The migratory background
07
Urban structure
08
Linguistic diversity
09
Geographical distribution of immigrants
10
Home/work commuting zones within Luxembourg
11
Residents between education, employment and retirement
12
Active residents
13
Overview of residents' housing
14
Being a senior in Luxembourg
15
Education levels of the Luxembourg population

The approach

1. **Analysis of the existing methodologies and tools**
   - Full Conditional Specification (FCS) approaches
   - Existing tools (R package `synthpop`)

2. **Application on STATEC census data**



Outlook

- Finalise the analysis on the real census data
- Assess the utility and privacy of the obtained synthetic data
- Draft a final report describing the approach and the results

DATA SUMMIT LUXEMBOURG

Hadiza Mahaman

Data Scientist
Luxembourg National Data Service

DATA SUMMIT LUXEMBOURG

Partner presentations

Improving Mobility through improving Data Quality

DATA SUMMIT LUXEMBOURG

Optimizing Data Quality

Effective Transport Management

- Data Quality and Curation Service

LNDS

LUXEMBOURG NATIONAL DATA SERVICE

- Implementation of scalable methodologies for efficient NeTEx data processing.
- Monitoring of datasets to ensure consistent data quality over time.

DATA SUMMIT LUXEMBOURG

Gil Georges

Program Project manager
Ministry of Mobility and Public Works (MMTP)

DATA SUMMIT LUXEMBOURG

Partner presentations

Understanding Weather – Accelerate access to high quality Meteo Data

DATA SUMMIT LUXEMBOURG

Bert Verdonck
CEO
Luxembourg National Data Service

Dr. Grégory Nain
Co-founder, Head of Operations
DataThings S.A

Transform data for standardised, fast and efficient analysis

Proof of concept: feedback requested

- Data Collection, Enrichment & Merging
- Data Quality & Curation
- GreyCat platform from:

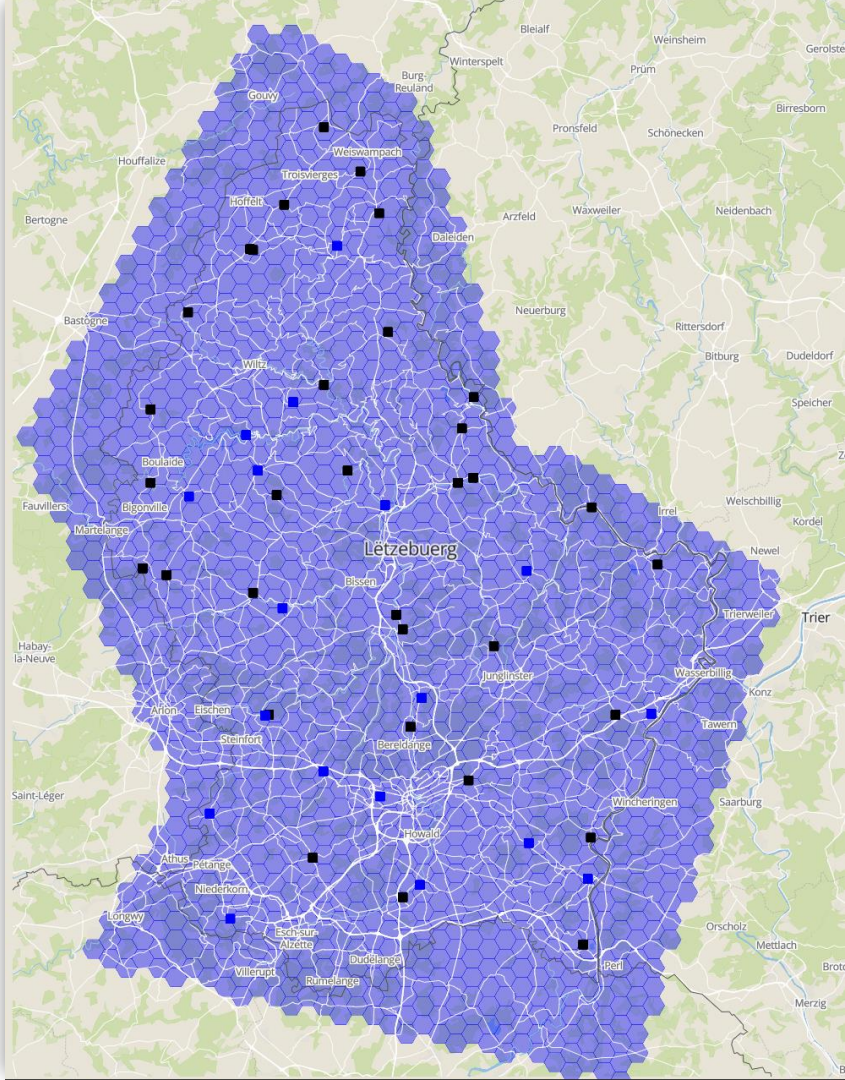
- Prototype a generic Luxembourg Meteo Mesh API
- Ensure infrastructure is efficient to sustain
  - Low to no code
  - Low infrastructure costs

Luxembourg meteo map

- Weather stations: 50 public stations
- Sensors: up to 28 values per station
- Pace: up to 12 values/hour (5 min.)
- History: 5 years
- Large variation in .csv formats
- 50 stations x 5 years x 5 minutes = 26M records/measurement

→ Interpolated map (5 closest neighbours)

→ Uniform API for consistent access



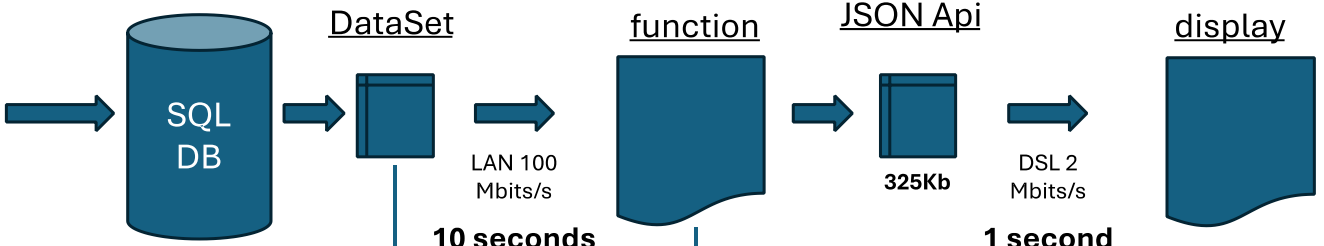
Demo …

Relational DB approach

SQL Query

Specifies what data to extract

```
SELECT station_id, time, temp
FROM measures
WHERE
 date >= '2023-01-01T00:00:00'
 AND
 date < '2024-01-01T00:00:00'
 AND
 station_id in (32, 21, 17...)
```



Measures

station_id	time	temp
66	2023-04-24T16:35:20	12
215	2023-04-24T16:35:25	13.7
272	2023-04-24T16:35:30	13
268	2023-04-24T16:35:35	13.5
95	2023-04-24T16:35:40	12.5
135	2023-04-24T16:35:45	13.2
278	2023-04-24T16:35:50	12.9
229	2023-04-24T16:35:55	12.6
276	2023-04-24T16:36:00	14.8
29	2023-04-24T16:36:05	13.5
145	2023-04-24T16:36:10	13.2
268	2023-04-24T16:36:15	15.6
95	2023-04-24T16:36:20	12.6
...	...	...

Date: 2 tests
Station: 25 tests (avg.)
27 tests per line

157,7 million lines
4,258 billion tests

21 seconds with
2GHz processor
+ 128Gb RAM

3 columns: 24 bytes
- Date: 8 bytes
- StationId: 8 bytes
- Temp: 8 bytes

5,256 million lines

=> **120,15 Mb**
to transfer to program

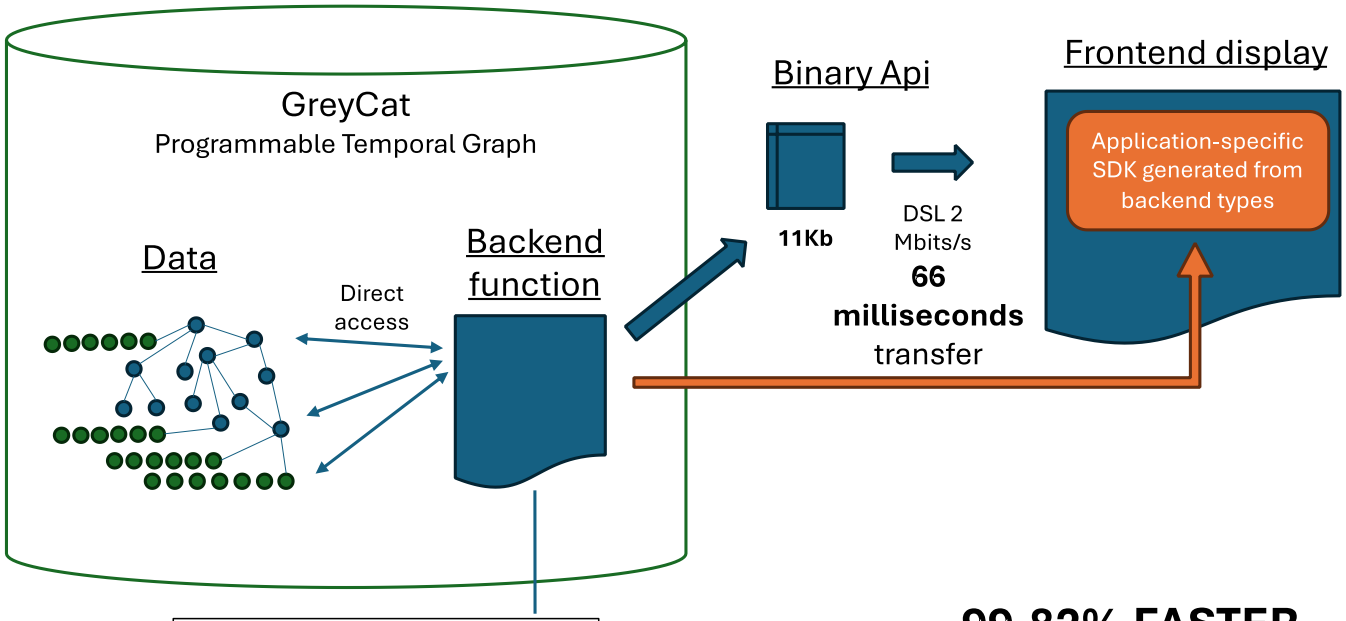
Parsing 5,256 million lines
Compute stations' average
For each of 1524 cells
- Compute distance to stations
- Take top 5
- Compute weighted average

17 seconds execution

TOTAL
Execution 49 seconds
Transfer 120Mb

Infrastructure
1 x 950€/month
1 x 550€/month
=> 18.000€/year

GreyCat approach



Navigate graph
Compute stations' average
For each of 1524 cells
- Compute distance to stations
- Take top 5
- Compute weighted average

20 milliseconds execution

TOTAL
Execution 86 milliseconds
Transfer 11Kb

Infrastructure
1 x 82€/month
=> 984€/year

99,82% FASTER

99.999...% LESS DATA TRANSFER

94,53% CHEAPER

Comparison in 5 years from now (+5 years of data)

Relational DB approach

41 seconds (+21s)
256Gb RAM (+128Gb)

TOTAL
Execution 1min11s (+21s)
Transfer 120Mb

Infrastructure
1 x 1150€/month (+200€)
1 x 550€/month
=> 20.400€/year (+2400€)

GreyCat approach

24 milliseconds (+4ms)
8Gb RAM

TOTAL
Execution 90 milliseconds (+4ms)
Transfer 11Kb

Infrastructure
1 x 82€/month
=> 984€/year

Refactoring of Luxembourg meteo data

- GreyCat analytics: directly hosted with the data
- Removed need for query execution and data extraction
- Infrastructure sized to execution time, not to the data size
- End-to-end frugality leads to substantial cost savings
- Next step: collect feedback and identify potential users
- Next step: enlarge scope to include restricted data

DATA
SUMMIT
LUXEMBOURG

Coffee Break

DATA SUMMIT LUXEMBOURG

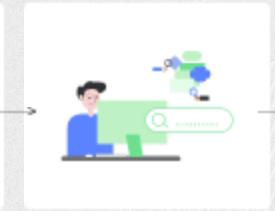
The End-To-End Journey – Part 2

DATA SUMMIT LUXEMBOURG

Data User Journey

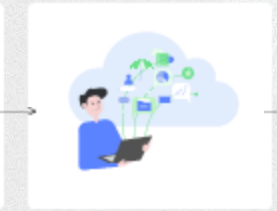
Step 0: **IDEATE**

Step 1: **DISCOVER**



Step 2: **REQUEST ACCESS**

Step 3: **ACCESS**



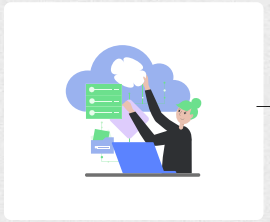
Step 4: **ANALYSE**

Step 5: **REVIEW RESULT**

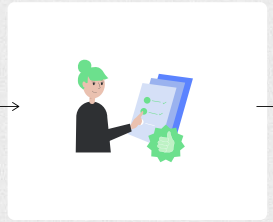
Step 6: **FINISH**

Data Provider Journey

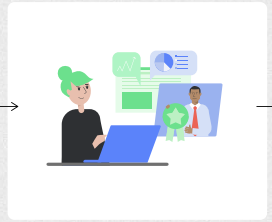
Step 0: **COLLECT**



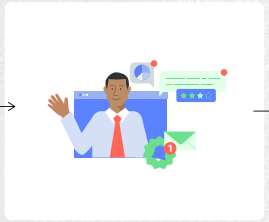
Step 1: **CATALOGUE**



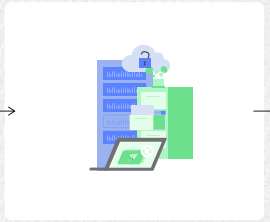
Step 2: **ENHANCE**



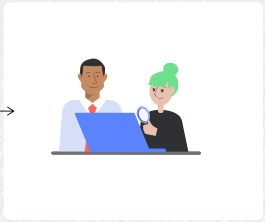
Step 3: **DECIDE ON ACCESS**



Step 4: **PROVIDE ACCESS**

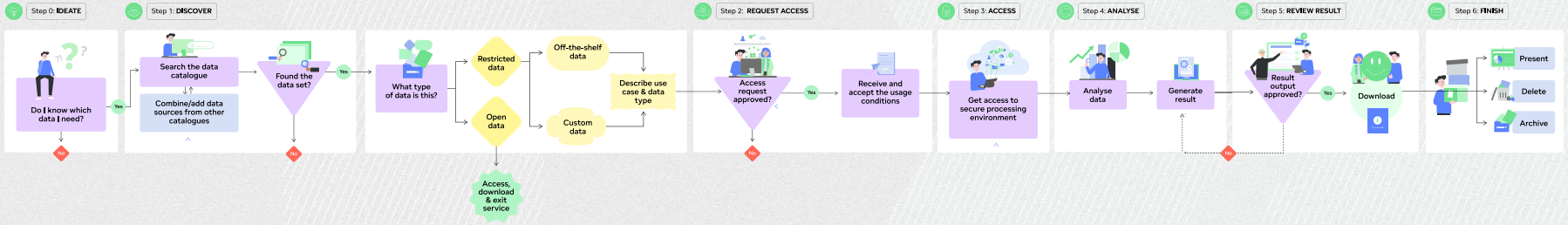


Step 5: **REVIEW OUTPUT**

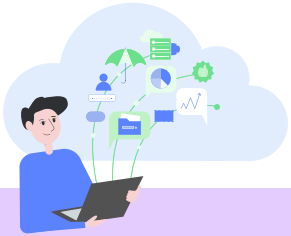


DATA SUMMIT LUXEMBOURG

Data User Journey



Step 3: **ACCESS**



Get access to secure processing environment

DATA SUMMIT LUXEMBOURG

LNDS Presents

Unveiling New Opportunities for Enabling Secondary use of data – Identifier-Management & Pseudonymisation (IPMS)

DATA SUMMIT LUXEMBOURG

Alemayehu Taye
Senior Data Scientist
Luxembourg National Data Service

Xavier Adam
Principal Data Scientist
Luxembourg National Data Service

Clement GORLT
Lead Cybersecurity
INCERT

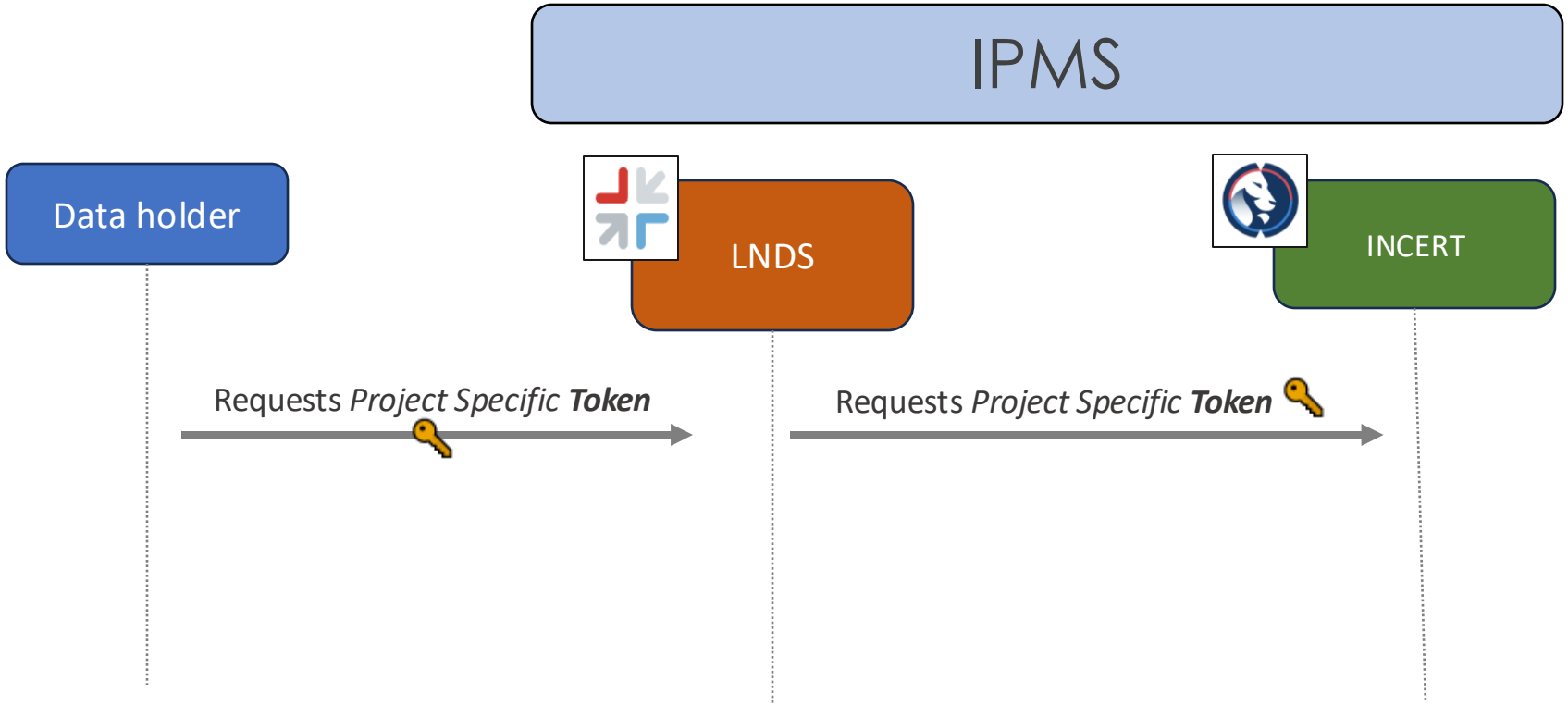
DATA
SUMMIT
LUXEMBOURG

LNDS

LUXEMBOURG NATIONAL DATA SERVICE

INCERT

IPMS - In details • Getting a token



IPMS - In details • Getting a token

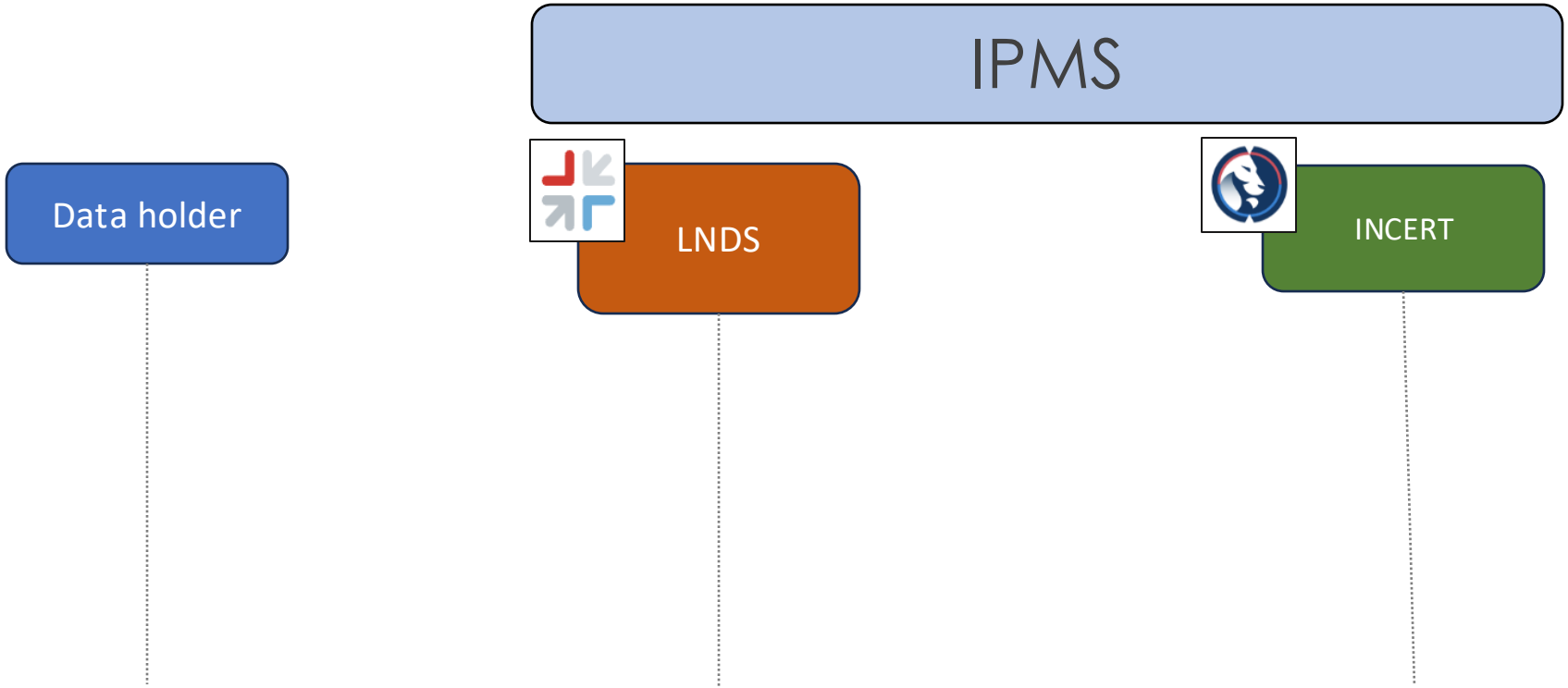
What's in a *Project - Specific Token* **?**

- Project ID
- Pseudonymisation Configuration
  - *Method (**Reversible** vs **Non-reversible**)*
  - *Treatment (Always the same or Always different)*
  - *Output format (Alphanumeric, Numeric, Alphabet)*

Segregated roles - INCERT:

1. **INCERT** will never know who the request is for.

IPMS - In details • Request Pseudonyms



IPMS - In details • Request Pseudonyms

Segregated roles - LNDS

2. **LNDS** will never see the Identifiers (sensitive data)!

IPMS - In details • Scope

IPMS - Scope

- Identifiers/Pseudonyms are ***never stored.*** Only the ***encryption key*** is kept if requested in the ***reversibility policy***
- Data holders can have as ***many projects/tokens*** as needed!
- At the request of Data Holder A, a token with the same project ID could be generated for Data Holder B to allow ***collaboration***.

DATA SUMMIT LUXEMBOURG

LNDS Presents

Protecting the analysis of confidential data – the challenge, a concept and a solution

DATA SUMMIT LUXEMBOURG

Todor Kondic
Senior Data Engineer
Luxembourg National Data Service

Karim Chaouch
Senior DevOps Engineer
Luxembourg National Data Service

LNDS

Secure Processing Environment (SPE)

Protecting the analysis of confidential data – the challenge, a concept and a solution

The Challenge

How to enable **modern analysis** while protecting data **confidentiality**?

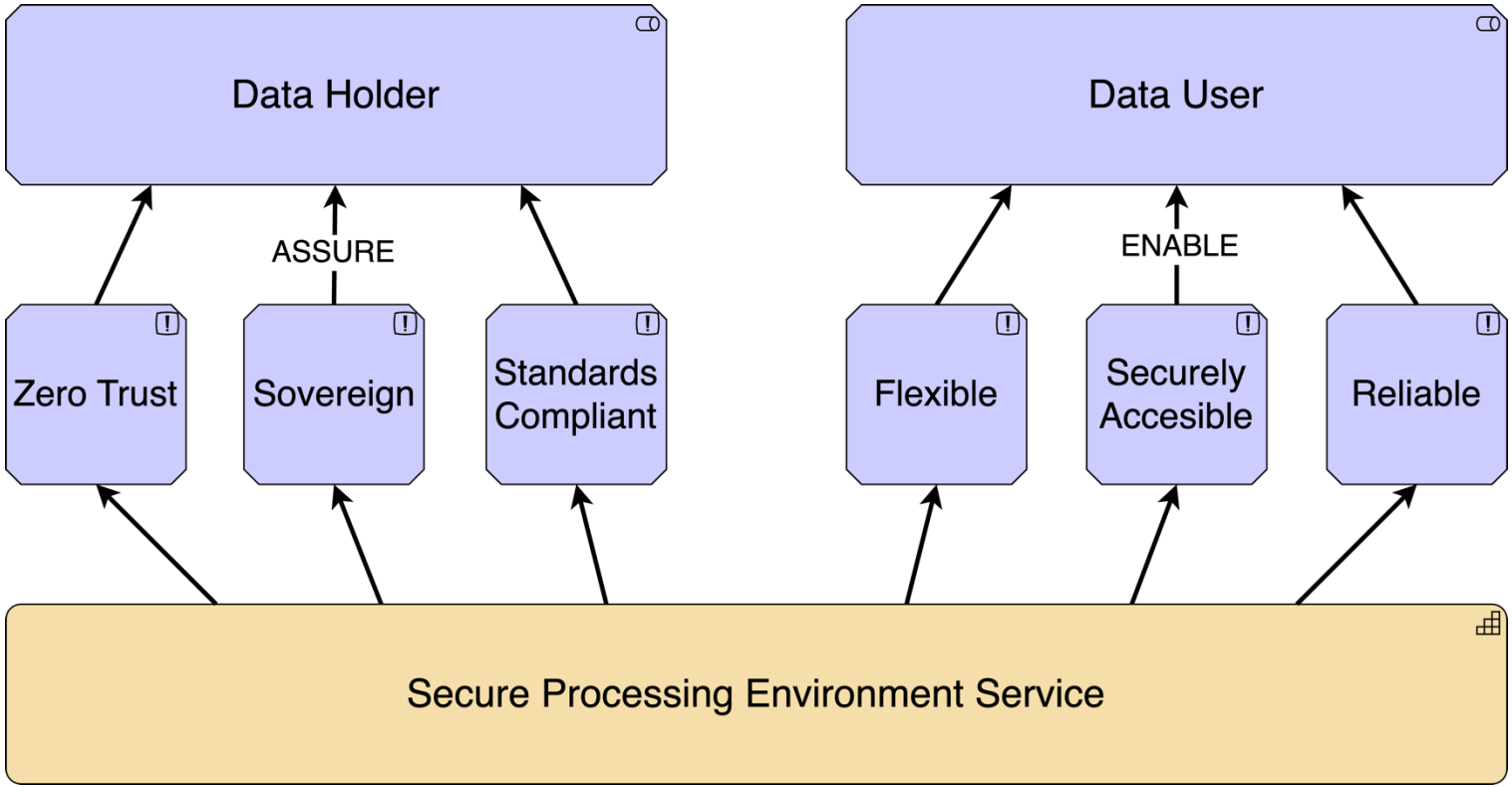
- Big data
- Flexible compute resources
- Diverse software needs

- Health records
- Trade secrets
- …

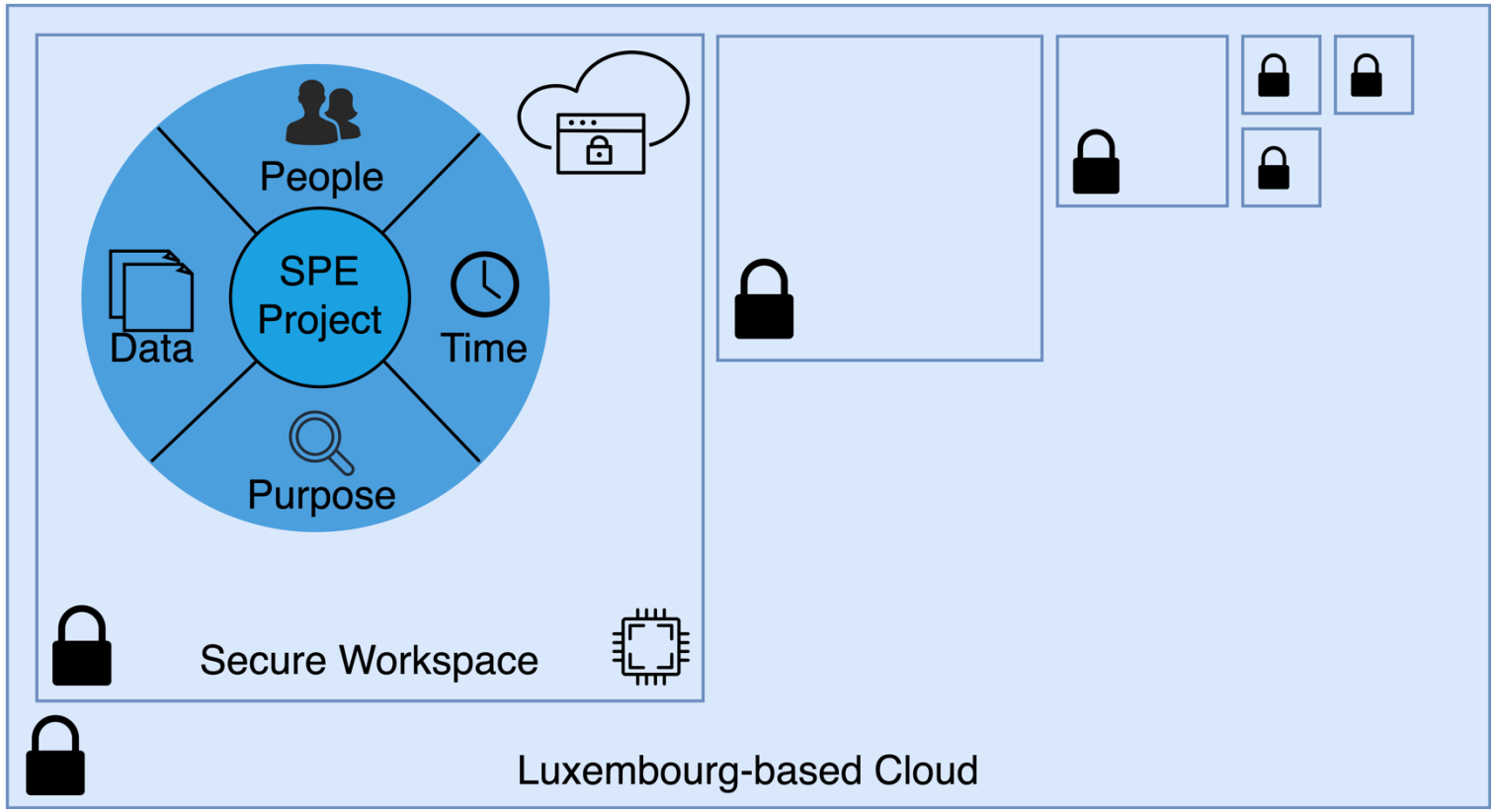
… can’t do it on a laptop

… a burden for IT departments

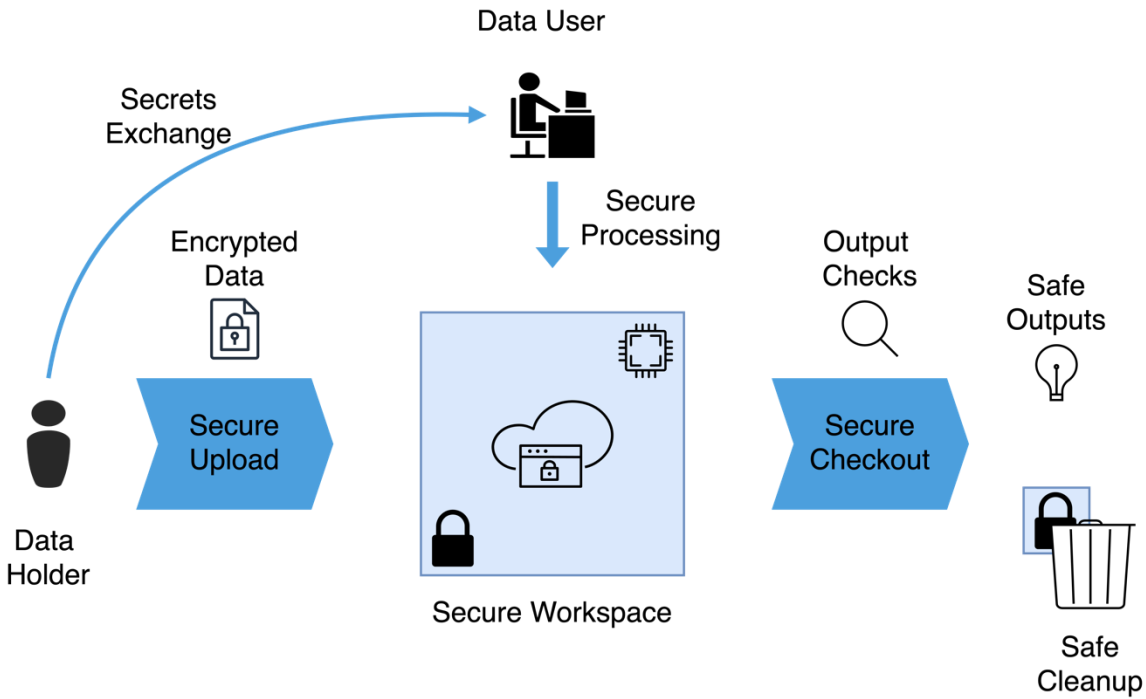
Guiding Principles



The Concept



Process



Secure Workspace

Don’t walk alone …

Learn and discover together with our partners

Doheem …

ctie
Centre des technologies de l'information de l'État

Luxembourg Microdata Platform on Labour and Social Protection

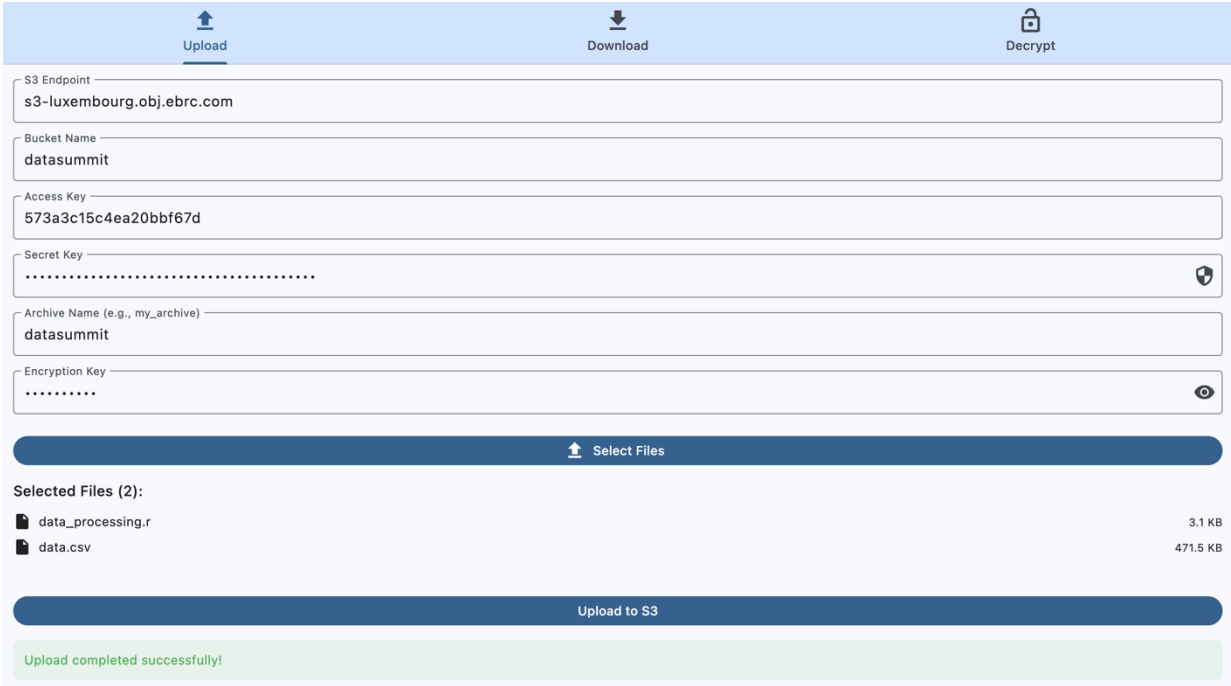
… and in EU

TEHDAS 2

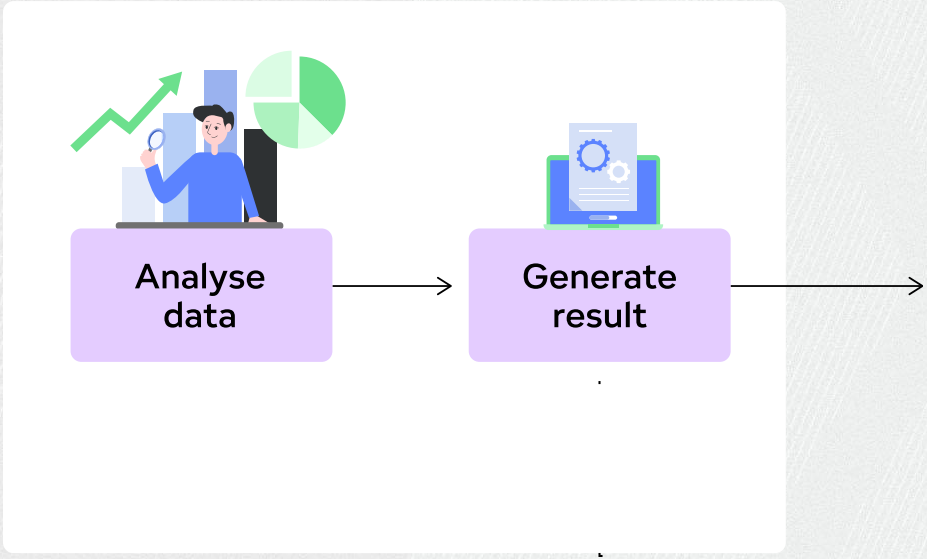
eosc | ENTRUST
European Network of Trusted Research Environments

Demo: secure data upload

- Secure S3 bucket
- Client-side encryption
- Strong encryption key



Step 4: **ANALYSE**



DATA SUMMIT LUXEMBOURG

Kerstin Neininger

Senior Data Scientist
Luxembourg National Data Service

DATA SUMMIT LUXEMBOURG

The last 5%: Identifying connectivity gaps in Luxembourg

Bring high-speed internet to 100% of buildings

LNDS Services

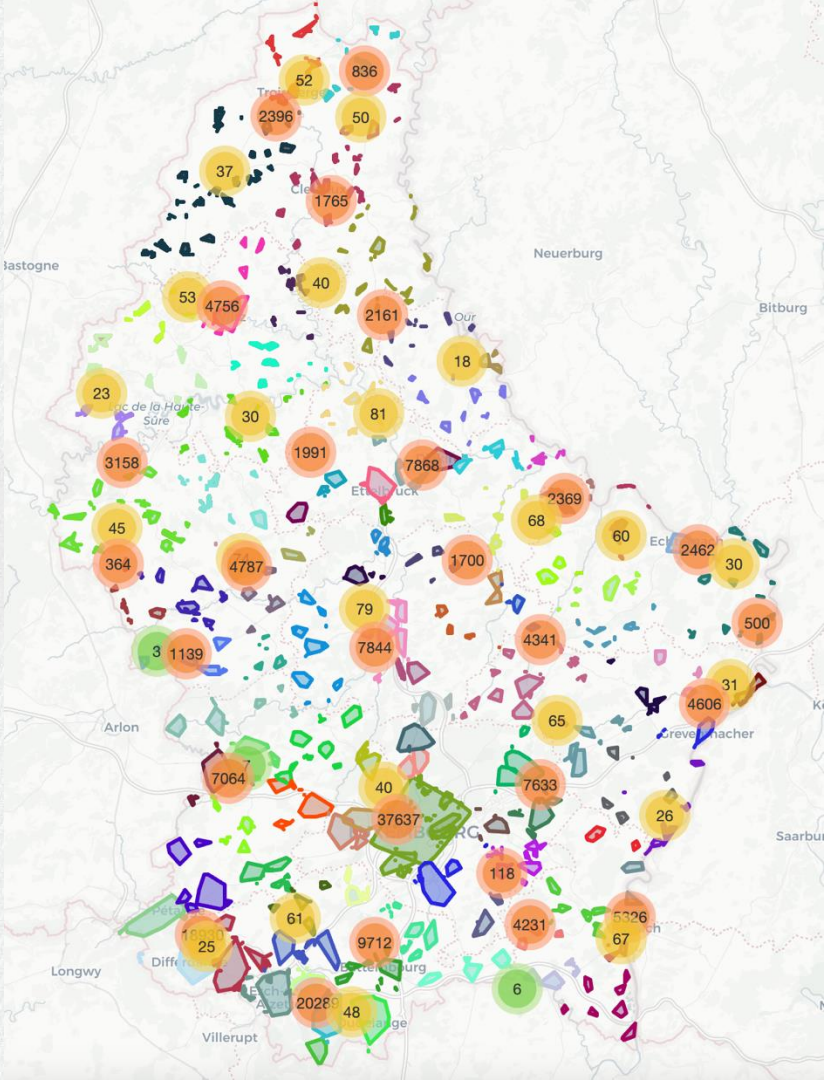
- Data Enrichment
- Data Analysis and Visualisation

INSTITUT LUXEMBOURGEOIS DE RÉGULATION

data. public.lu

MY CONNECTIVITY

"Connecting Every Commune"



DATA SUMMIT LUXEMBOURG

Partner presentations

A data-driven approach to tackle the missing 5% of ultra-high-speed broadband in Luxembourg

DATA SUMMIT LUXEMBOURG

Julien Larios

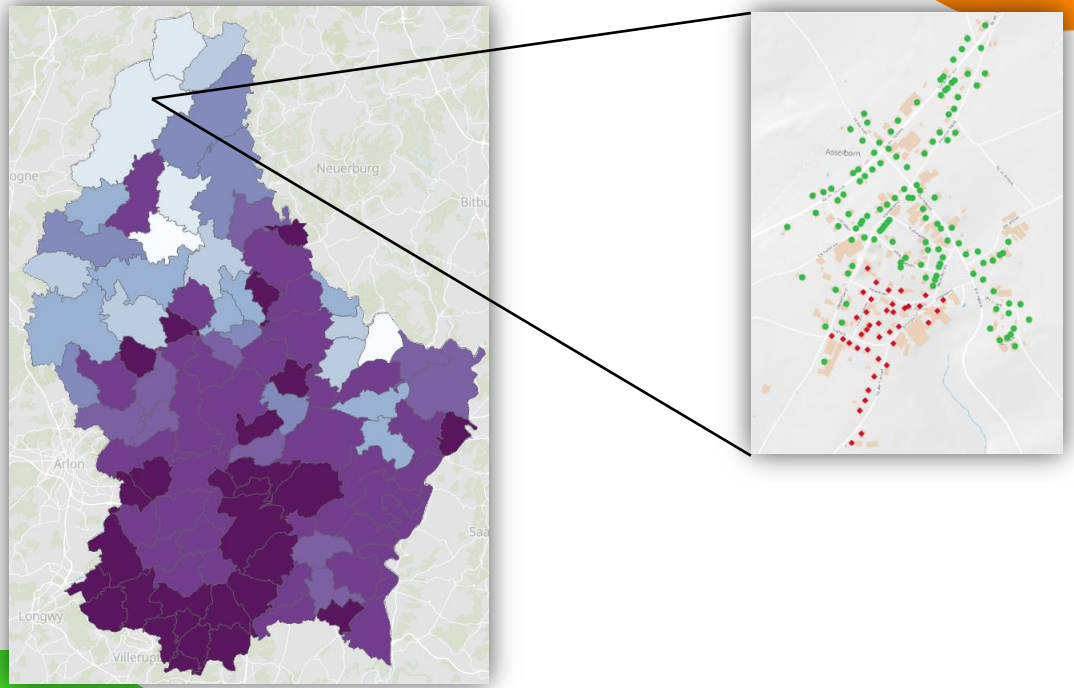
CEO

MyConnectivity

A data-driven approach for the missing 5% of ultra-high-speed broadband in Luxembourg

LNDS Data Summit 2024

The missing 5%

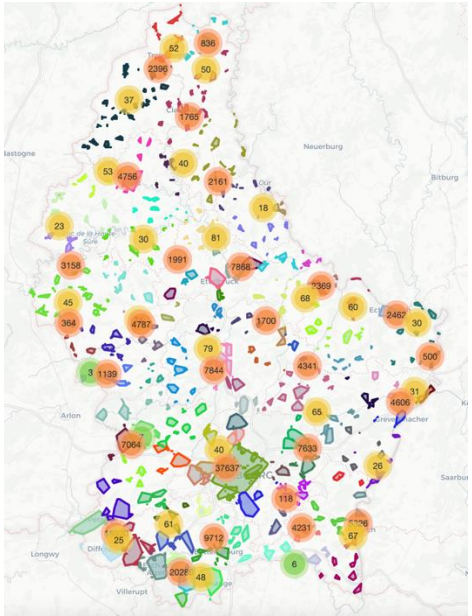


How to identify?
How to qualify?
How to prioritize?

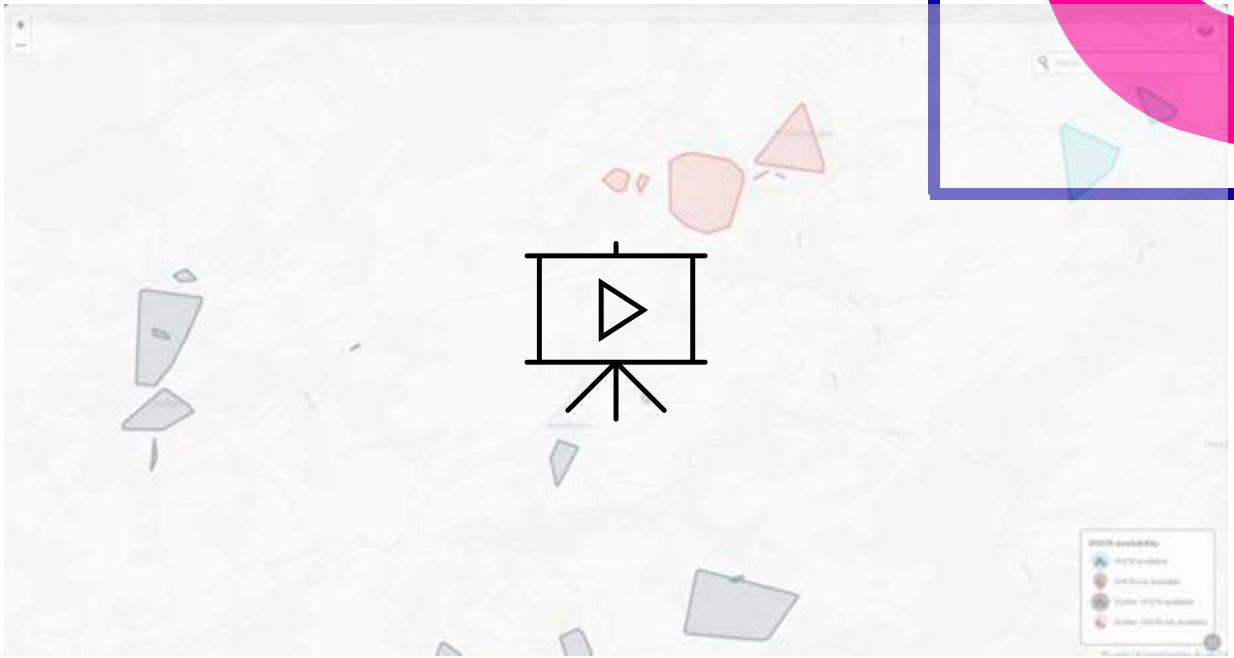
Challenges & process of the project

- Challenges :
  - Accessing certain datasets
  - Combining several datasets, such as :
    - 5G coverage speeds, Address cadaster, ILR’s inventory

- Process :
  - 7 different LNDS services supported the end-to-end process
  - Stakeholders involvment along the entire project



Impact



Who are we ?

a neutral team in Luxembourg

with a mission to

accelerate connectivity topics.

100% of Luxembourg connected to very-high-capacity networks.

THE GOVERNMENT
OF THE GRAND DUCHY OF LUXEMBOURG
Ministry of State

Department of Media, Connectivity and Digital Policy

LU-CIX GIE

MyConnectivity in action

THANK YOU!

https://myco.lu/

Julien Larios

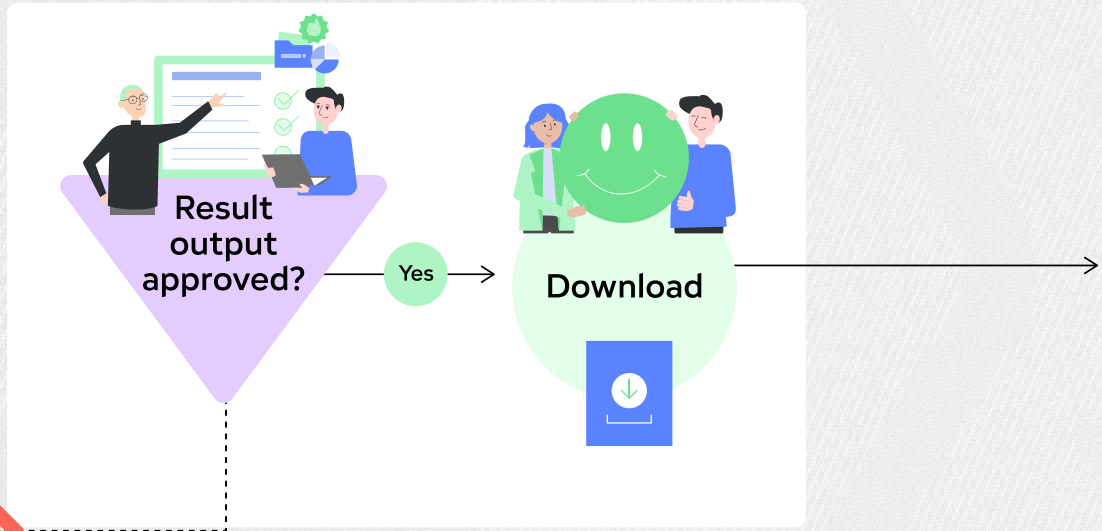
CEO

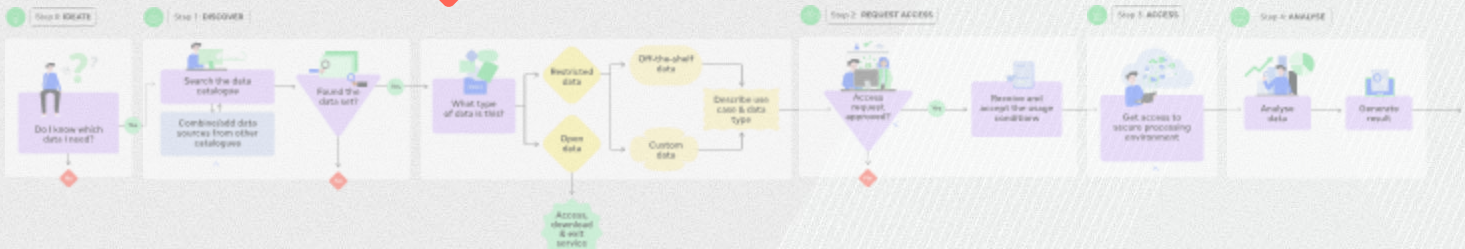
Julien.Larios@myconnectivity.lu

MyConnectivity G.I.E.

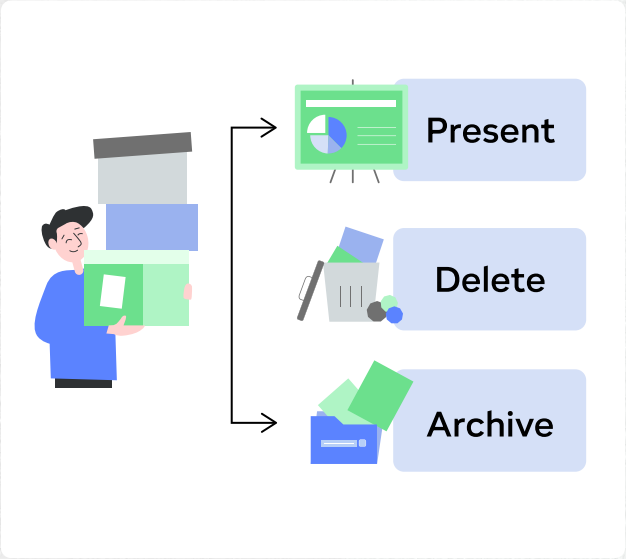
9, Rue du Laboratoire,
L-1911 Luxembourg

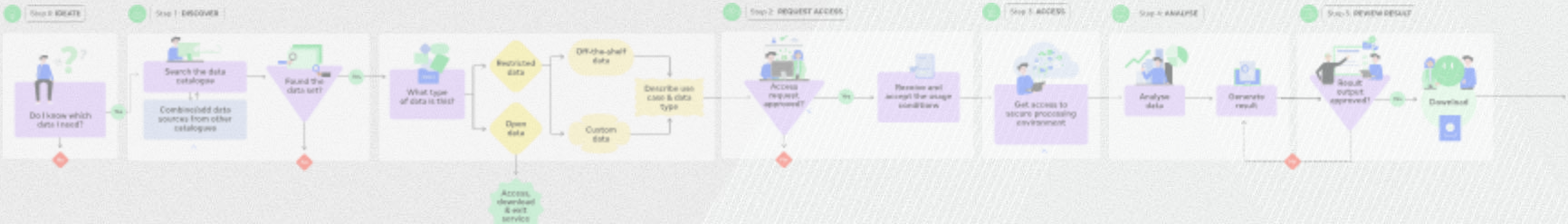
Step 5: **REVIEW RESULT**





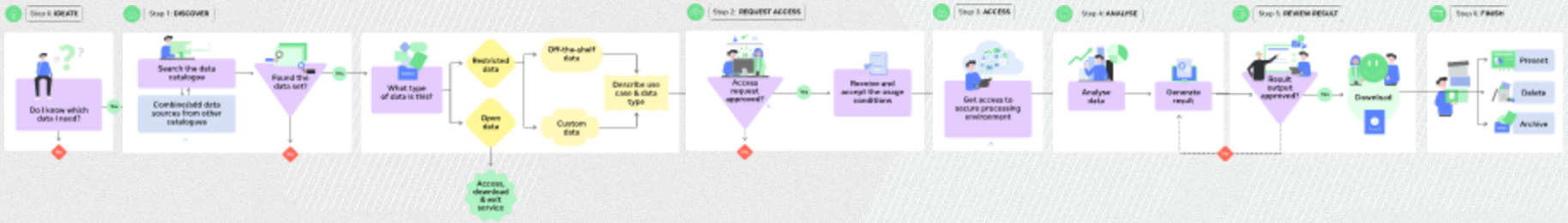
Step 6: **FINISH**





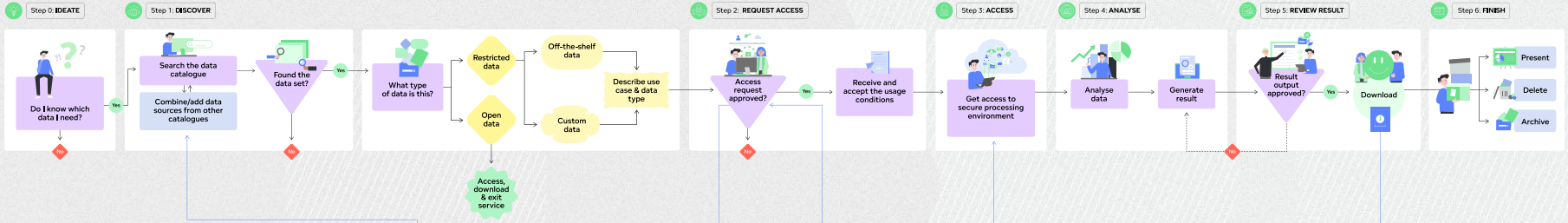
DATA SUMMIT LUXEMBOURG

Data User Journey

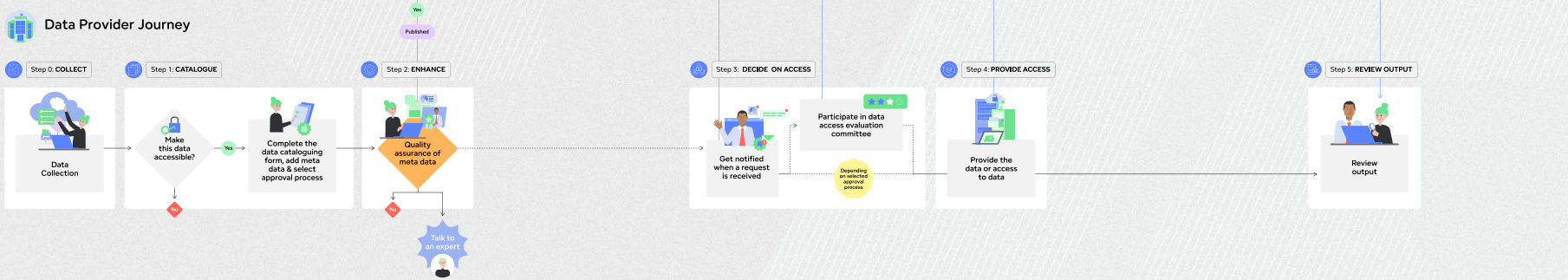


DATA SUMMIT LUXEMBOURG

Data User Journey



Data Provider Journey

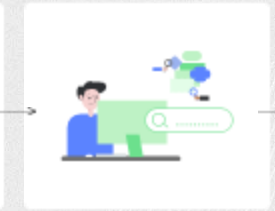


DATA SUMMIT LUXEMBOURG

Data User Journey

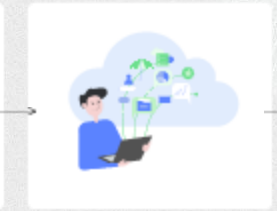
Step 0: **IDEATE**

Step 1: **DISCOVER**



Step 2: **REQUEST ACCESS**

Step 3: **ACCESS**



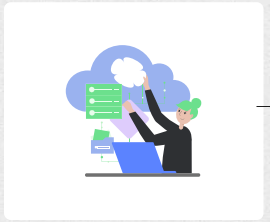
Step 4: **ANALYSE**

Step 5: **REVIEW RESULT**

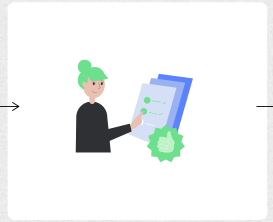
Step 6: **FINISH**

Data Provider Journey

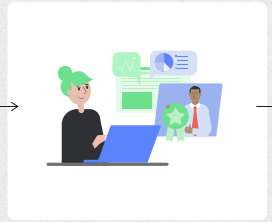
Step 0: **COLLECT**



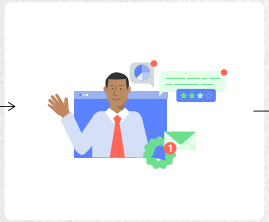
Step 1: **CATALOGUE**



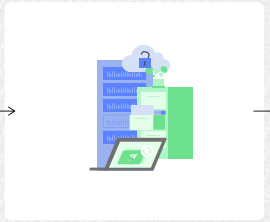
Step 2: **ENHANCE**



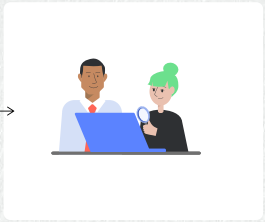
Step 3: **DECIDE ON ACCESS**



Step 4: **PROVIDE ACCESS**



Step 5: **REVIEW OUTPUT**



DATA SUMMIT LUXEMBOURG

Data Sharing initiatives in the Luxembourg ecosystem

DATASPACE 4HEALTH
LUXEMBOURG

Luxembourg launched Dataspace4Health:

a groundbreaking open ecosystem designed to revolutionize secure and compliant health data exchange

Visit our website

NTT DATA · HÔPITAUX ROBERT SCHUMAN · UNIVERSITÄT LUXEMBURG · LUXEMBOURG INSTITUTE OF HEALTH · AGENCE eSanté LUXEMBOURG · LNDS · LUXINNOVATION #MakingInnovationHappen

Follow us on LinkedIn

Luxembourg launched Dataspace4Health: a groundbreaking open ecosystem designed to revolutionize secure and compliant health data exchange

The need

The project is motivated by the need for a new approach to health data sharing that respects the GDPR and the patients' rights.

Currently health data is often siloed, fragmented and underutilized, which limits the potential for innovation and research.

Solution

The project explores how to use Gaia-X as a European framework for data spaces that ensures data protection, security and interoperability.

This initiative aims to address both business and patient needs by ensuring compliance with regulations, unlocking the value of health data through responsible monetization, and exploring the potential of data and AI to develop new treatment options.

Outcomes

- The built ecosystem is use-case driven and has been validated through initial use cases in diabetes and oncology, paving the way for future applications in other healthcare sectors.
- Enhanced patient care includes improved diagnoses and treatments, better understanding of diseases, and effective preventive measures.
- A decision support system based on AI that can prevent diabetes complications might be a key step in the treatment of Diabetes. In one use-case, that reliable decision support system uses thousands of anonymised patient records linked to a Digital Twin of patients from LIH to provide personalised treatment.
- It is a Gaia-X Lighthouse candidate in healthcare for Europe, setting the standard for secure and compliant health data exchange across the continent.

> This visionary project paves the way for secure and compliant health data exchange, ultimately leading to a healthier future for all.

Led by NTT DATA, Hôpitaux Robert Schuman, Luxembourg Institute of Health, the University of Luxembourg, Agence eSante, LNDS, Luxinnovation.

Visit our website

Follow us on LinkedIn

DATA SUMMIT LUXEMBOURG

Partner presentations

The End-to-End Journey concludes

GDI: From Data to Discovery: Empowering Clinical Insights

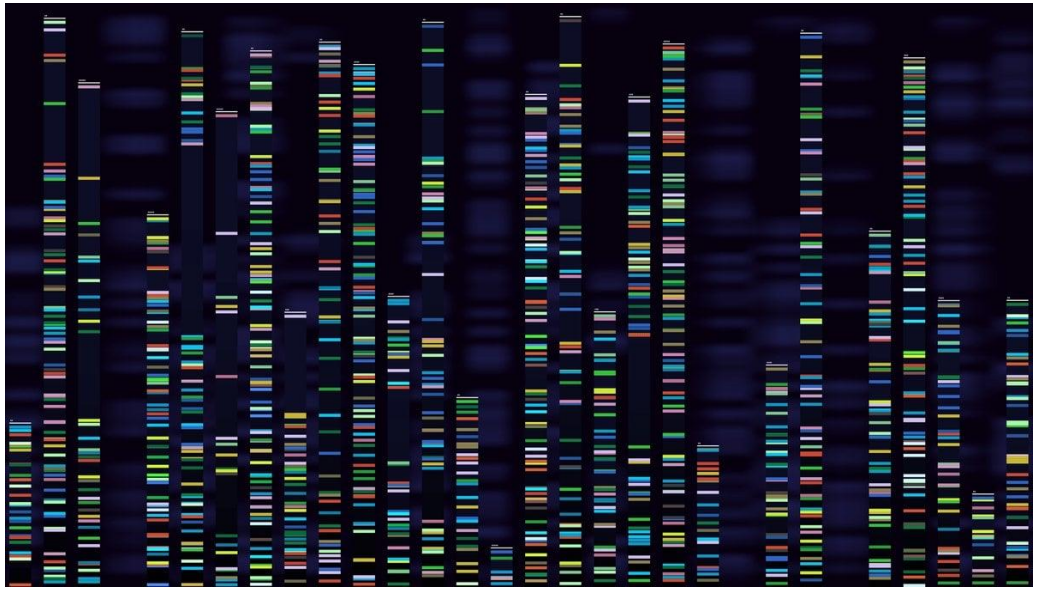
DATA SUMMIT LUXEMBOURG

Bruno Rodrigues

PhD, Responsable du Service Statistiques et du Service Stratégie data
Ministère de la Recherche et de l'Enseignement supérieur (MESR)

Genomic medicine

A genome is the complete set of an organism's genetic material.



Genomic medicine

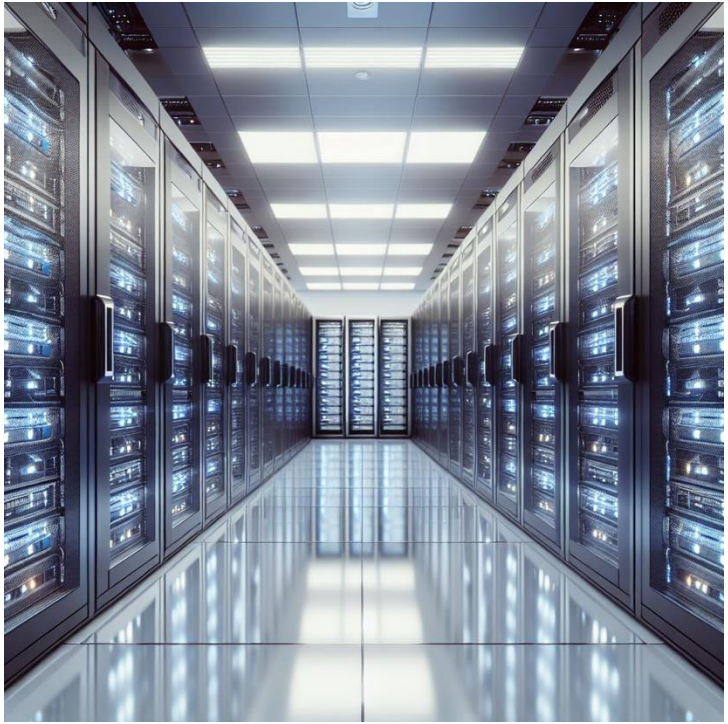
LE GOUVERNEMENT
DU GRAND-DUCHÉ DE LUXEMBOURG

Understand the genome can lead to better personalized treatments

Genomic medicine

LE GOUVERNEMENT
DU GRAND-DUCHÉ DE LUXEMBOURG

Massive data, AI and high performance computing are needed



Genomic medicine

Especially true for rare diseases: European collaboration is vital!

source: https://eu-rd-platform.jrc.ec.europa.eu/_en

1+ Million Genome initiative (1+MG)

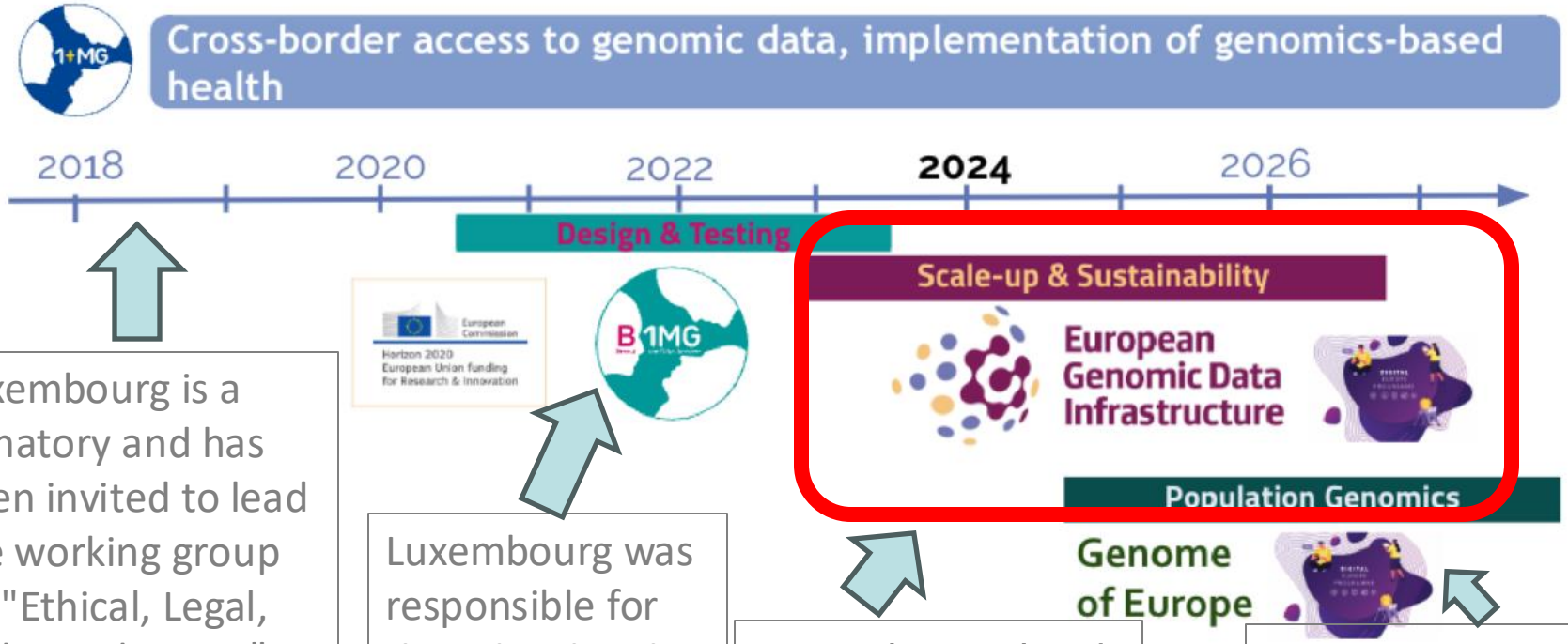
LE GOUVERNEMENT
DU GRAND-DUCHÉ DE LUXEMBOURG

European governmental initiative to give cross-border access to genomic and health data for personalised medicine

- Access to genomic and health data in a national secure processing environment
  - To support clinical decision making
  - To research for better prevention, diagnosis and treatment
  - To support policy making for personalised medicine

The 1+MG iniative

LE GOUVERNEMENT
DU GRAND-DUCHÉ DE LUXEMBOURG



Luxembourg is a signatory and has been invited to lead the working group on "Ethical, Legal, and Social Issues" (ELSI).

Luxembourg was responsible for the ethical and legal aspects of the project.

Luxembourg leads several important functions.

Luxembourg leads the ELSI work and contributes data.

Genomic Data Infrastructure

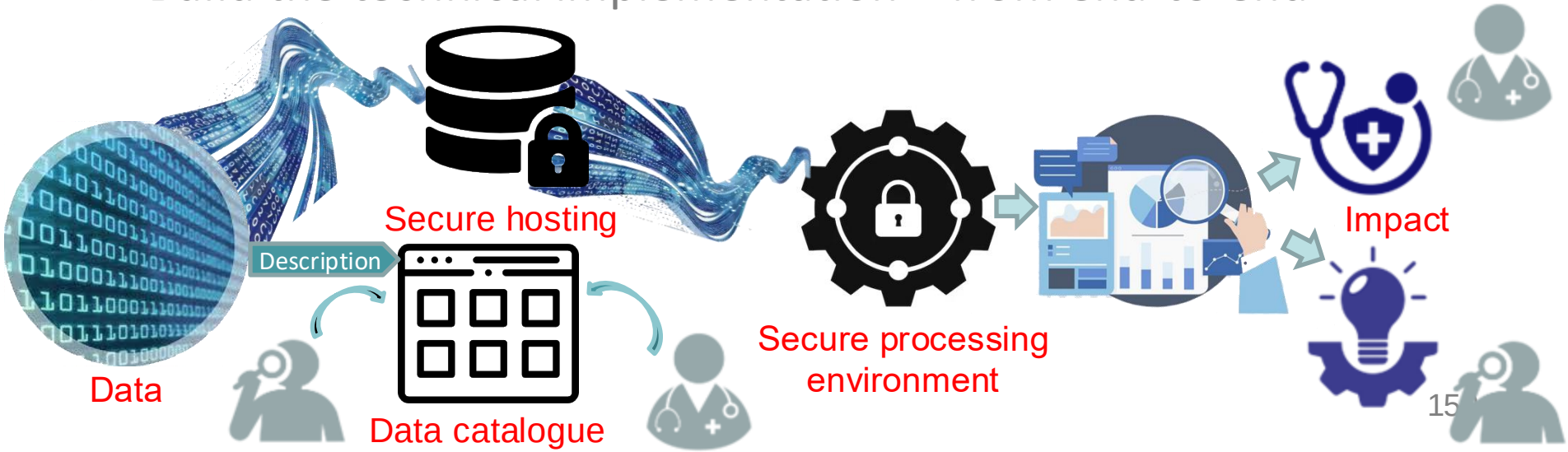
LE GOUVERNEMENT
DU GRAND-DUCHÉ DE LUXEMBOURG

- Project started in November 2022 (end: October 2026)
- Objectives:
  - Deploy a federated data infrastructure for the secondary use of genomic and clinical data across Europe
  - Establish a sustainable data governance policy and coordination mechanism
  - Contribute to improving the interoperability of genomic and clinical data
- Luxembourg:
  - Leads the sustainability pillar, aiming to create a durable legal framework and financial model.
  - Leads the European Operations work package and develops the user portal

1+MG – realising the vision

LE GOUVERNEMENT
DU GRAND-DUCHÉ DE LUXEMBOURG

- Build a trusted and trustworthy data governance and provide for the necessary legal framework
- Gain patients and citizens' buy-in for a co-creation to enable better healthcare in Luxembourg
- Build the technical implementation – from end-to-end



Entities involved in the project for national implementation

LE GOUVERNEMENT
DU GRAND-DUCHÉ DE LUXEMBOURG

LE GOUVERNEMENT
DU GRAND-DUCHÉ DE LUXEMBOURG
Ministère de la Santé
et de la Sécurité sociale

LE GOUVERNEMENT
DU GRAND-DUCHÉ DE LUXEMBOURG
Ministère de la Recherche
et de l'Enseignement supérieur

LNDS

LUX PROVIDE
(discussions préliminaires)

LUXEMBOURG INSTITUTE OF HEALTH

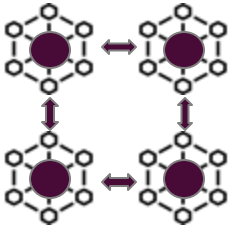
LNS LUXEMBOURG

The GDI by the end of 2026

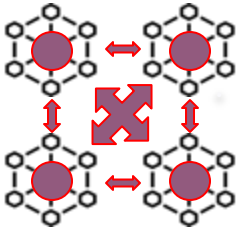
LE GOUVERNEMENT
DU GRAND-DUCHÉ DE LUXEMBOURG

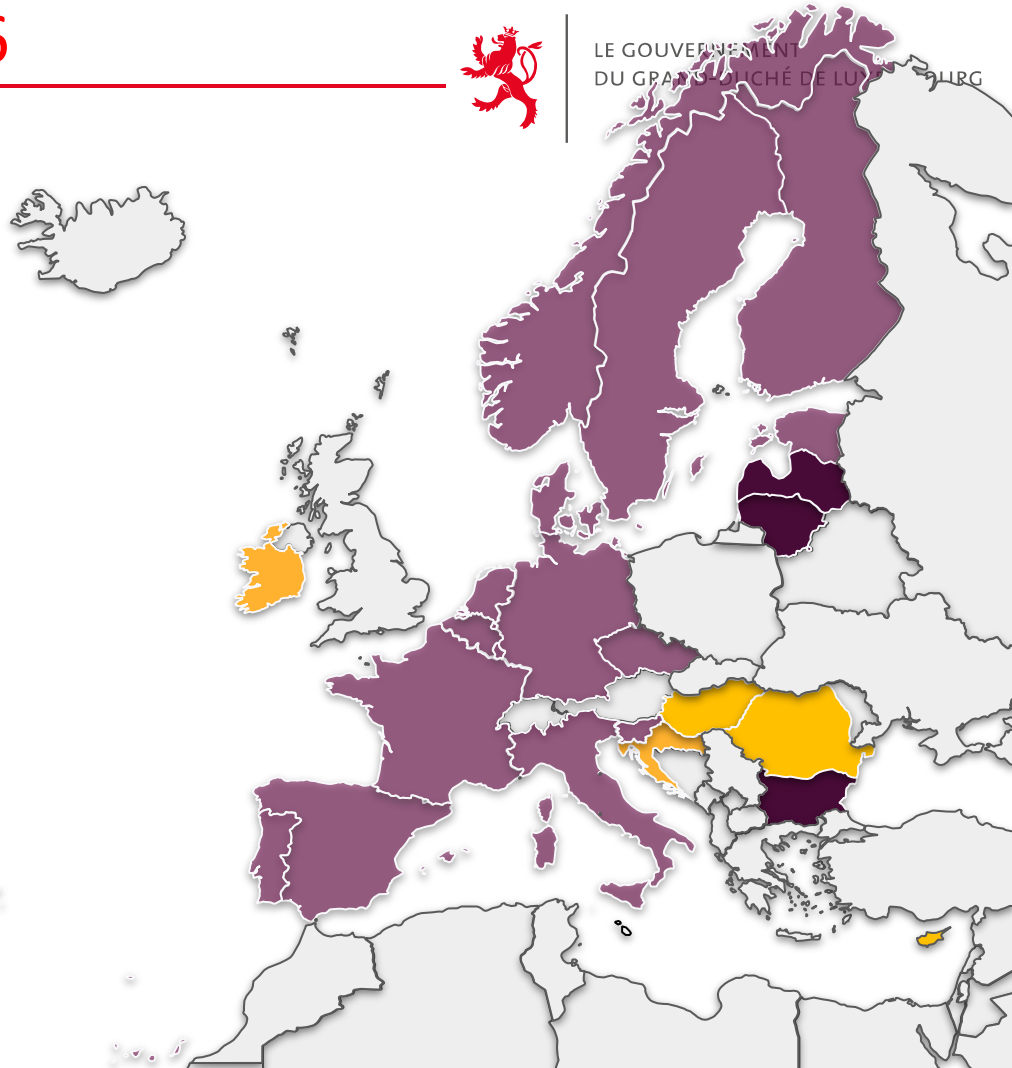
***Onboarding*: 6**
Addressing national challenges for deployment

***Deployment*: 3**
Operational national node addressing challenges for connections at the EU level



***Operational*: 15**
Fully deployed national node interconnected with EU services, capable of providing access to cross-border data in accordance with the 1+MG or EHDS governance.





Genomic Data Infrastructure

LE GOUVERNEMENT
DU GRAND-DUCHÉ DE LUXEMBOURG

- Pivotal moment in the project

Creation of the a legal entity and its headquarters of the infrastructure as in a Member State
(Luxembourg is currently the designated host).

Genomic Data Infrastructure

LE GOUVERNEMENT
DU GRAND-DUCHÉ DE LUXEMBOURG

- Pivotal moment in the project

Creation of the legal headquarters of the infrastructure likely as an EDIC hosted in a Member State.

EDIC: European Digital Infrastructure Consortium (EDIC) is an instrument made available to Member States under the Digital Decade Policy Programme 2030 to speed up and simplify the setup and implementation of multi-country projects. EDICs will enable the achievement of the Digital Decade general objectives and targets.

Genome EDIC

LE GOUVERNEMENT
DU GRAND-DUCHÉ DE LUXEMBOURG

- A new entity operationalising the genomic data infrastructure of the Genomic Data Infrastructure project.
- It will promote the secondary use of genomic data to enhance research, personalized healthcare, and the development of new health policies.
- The data will be accessible across all states joining Genome EDIC, enabling doctors to use data on rare diseases across Europe for their patients.
- Only researchers and doctors who need the data for their work will have access through a secure computing environment.

Welcome the Genome EDIC:
an opportunity for Luxembourg

LE GOUVERNEMENT
DU GRAND-DUCHÉ DE LUXEMBOURG

- The first country to host a data space with a very high impact
- Positions Luxembourg as a leader in the digital sector
- Strengthens Luxembourg as a hub for highly skilled workers
- Enhances our national data infrastructures (LNDS, Meluxina, Meluxina-Q...)
- Numerous interactions with EHDS: the national implementation of EHDS will bring significant economies of scale to Genome EDIC!

DATA
SUMMIT
LUXEMBOURG

Coffee Break

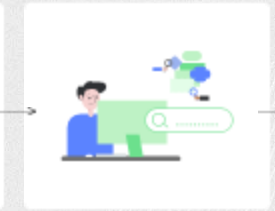
DATA SUMMIT LUXEMBOURG

Data User Journey

Data User Journey

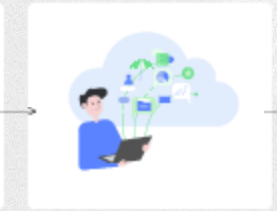
Step 0: **IDEATE**

Step 1: **DISCOVER**



Step 2: **REQUEST ACCESS**

Step 3: **ACCESS**



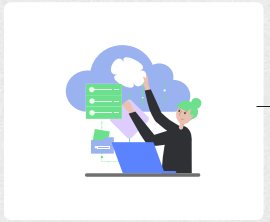
Step 4: **ANALYSE**

Step 5: **REVIEW RESULT**

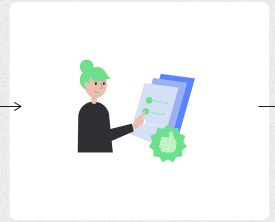
Step 6: **FINISH**

Data Provider Journey

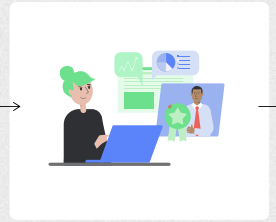
Step 0: **COLLECT**



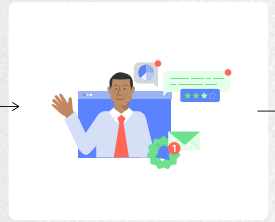
Step 1: **CATALOGUE**



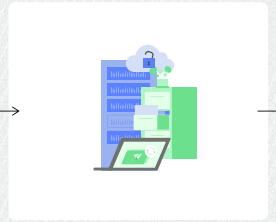
Step 2: **ENHANCE**



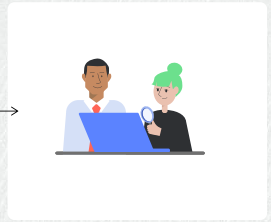
Step 3: **DECIDE ON ACCESS**



Step 4: **PROVIDE ACCESS**



Step 5: **REVIEW OUTPUT**



DATA SUMMIT LUXEMBOURG

Keynote

A Day in the Life of Alice in the world of Web 3.0 Data Space

DATA SUMMIT LUXEMBOURG

Alex Tourski

Founder
Post-Platforms Foundation

Alice

One day from life of Alice In Wonder Web 3.0 Data Space

See the video-recording of the keynote, created by Post-Platforms Foundation, here.

Alex Tourski
Alex.tourski@postplatforms.org

DATA
SUMMIT
LUXEMBOURG

Looking back at Data Summit 2024

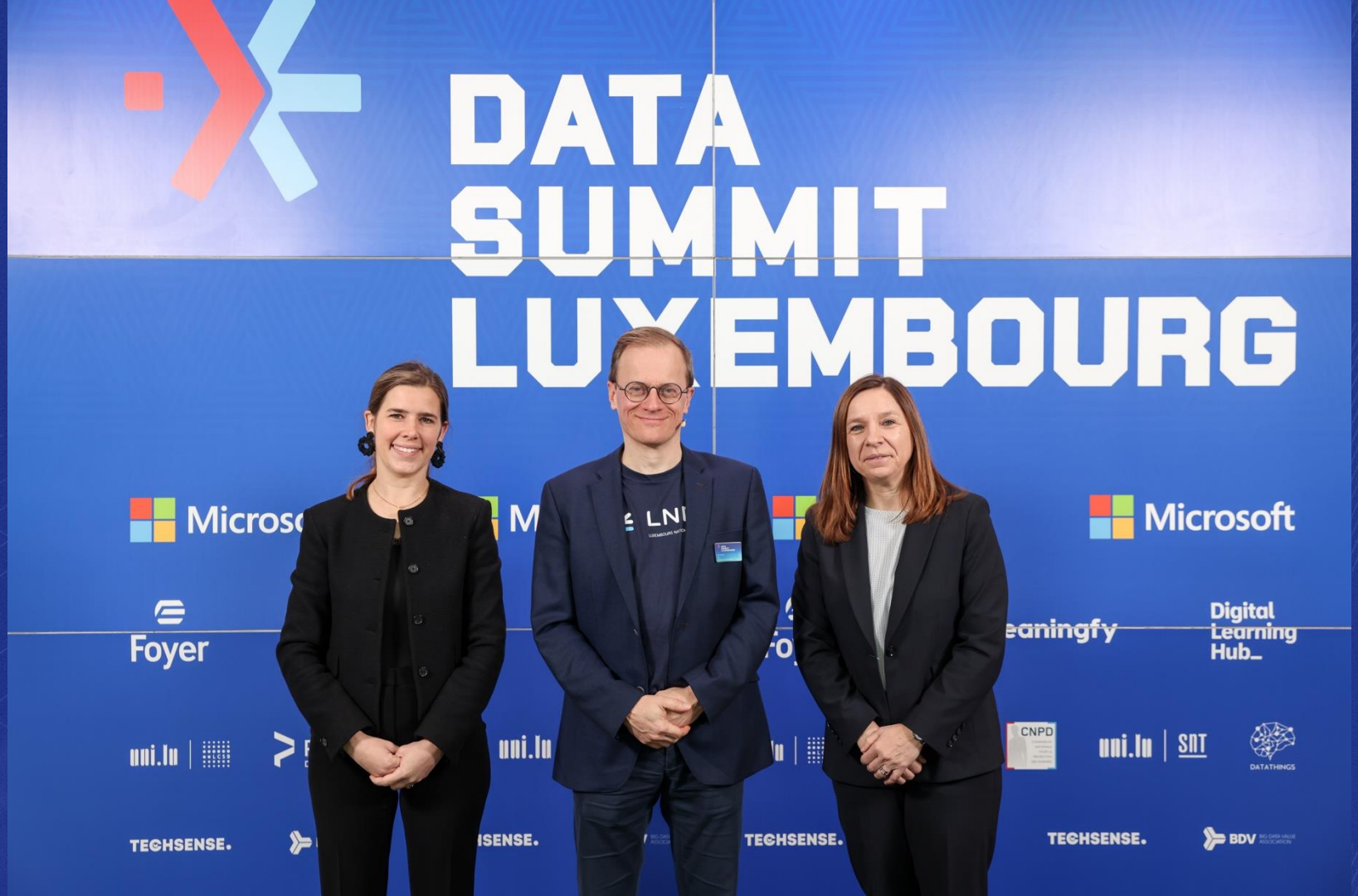

Data
Movers
Meaningful
LNDS


LNC


Big THANK YOU to our Partners & Sponsors!
Microsoft
Meaningfy
Foyer
Digital Learning Hub
Deloitte
LUXEMBOURG
TECHNESE
Innovation throughout the
DATA JOURNEY


DATA SUMMIT LUXEMBOURG
Luxembourg National Data Service
Stéphanie Obertin
Minister for Digitalisation
Minister for Research and Higher Education
Innovation throughout the
DATA JOURNEY


DATA SUMMIT LUXEMBOURG
Innovation throughout the
DATA JOURNEY


DATA
SUMMIT
LUXEMBOURG
Microsoft
Foyer
Digital Learning Hub
CNPD
DATATHINGS
TECHSENSE.
TECHSENSE.
TECHSENSE.
BDV


DATA SUMMIT LUXEMBOURG
Luxembourg National Data Service
Elisabeth Margue
Minister of Justice
Minister Delegate to the Prime Minister for Media and Connectivity
Innovation throughout the DATA JOURNEY


DATA SUMMIT LUXEMBOURG
Luxembourg National Data Service
are we cruising?
MODERATOR
Bert Verdonck
CEO
Luxembourg National Data Service
Max Gindt
Attaché
Ministry of State - Department for Media, Connectivity and Digital Policy
Björn De Vidts
CEO
Athumi
Ana Garcia Robles
Secretary General
Big Data Value Association (BDVA)
Malte Beyer-Katzenberger
Policy Officer
European Commission DG-CONNECT
Innovation throughout the
DATA JOURNEY


LNDS


LNDS


Data
Matters
DATA SUMMIT
LUXEMBOURG
Innovation throughout the
DATA JOURNEY

DATA SUMMIT LUXEMBOURG

3 next steps

Datasummit.lu

Datasummit.lu

LNDS.lu

DATA
SUMMIT
LUXEMBOURG

Cocktail

DATA SUMMIT LUXEMBOURG

Big THANK YOU to our Partners & Sponsors!

Gold Sponsor

Microsoft

Silver Sponsors

Meaningfy

Foyer

Digital Learning Hub

Deloitte.

Bronze Sponsors

CNPD Commission Nationale pour la Protection des Données

DATATHINGS

uni.lu | LCSB

uni.lu | SNT

Portus Data Centers

Other Sponsors & Partners

TECHSENSE.

BDV Big Data Value Association

LuxData

Publications Office of the European Union

The Government of the Grand Duchy of Luxembourg

Luxembourg Institute of Health

Luxembourg Institute of Socio-Economic Research

DATA SUMMIT LUXEMBOURG

Save the date!
02.12.2025

European Convention Center Luxembourg

DATA SUMMIT LUXEMBOURG

Be part of Data Summit Luxembourg 2025!

Scan the QR code

DATA SUMMIT LUXEMBOURG

DATA
SUMMIT
LUXEMBOURG

Goodbye!