# An open energy data lifecycle: organisation, methods and tools

**PhD defence**
**4 November 2025**

Presented by **Seun OSONUGA**

Before a jury of:

**Supervisors**
- Dr. Frederic WURTZ, CNRS/G2ELab
- Prof. Benoit DELINCHANT, Grenoble-INP UGA

**Invited guest**
- Aude de TOUCHET, Agence ORE

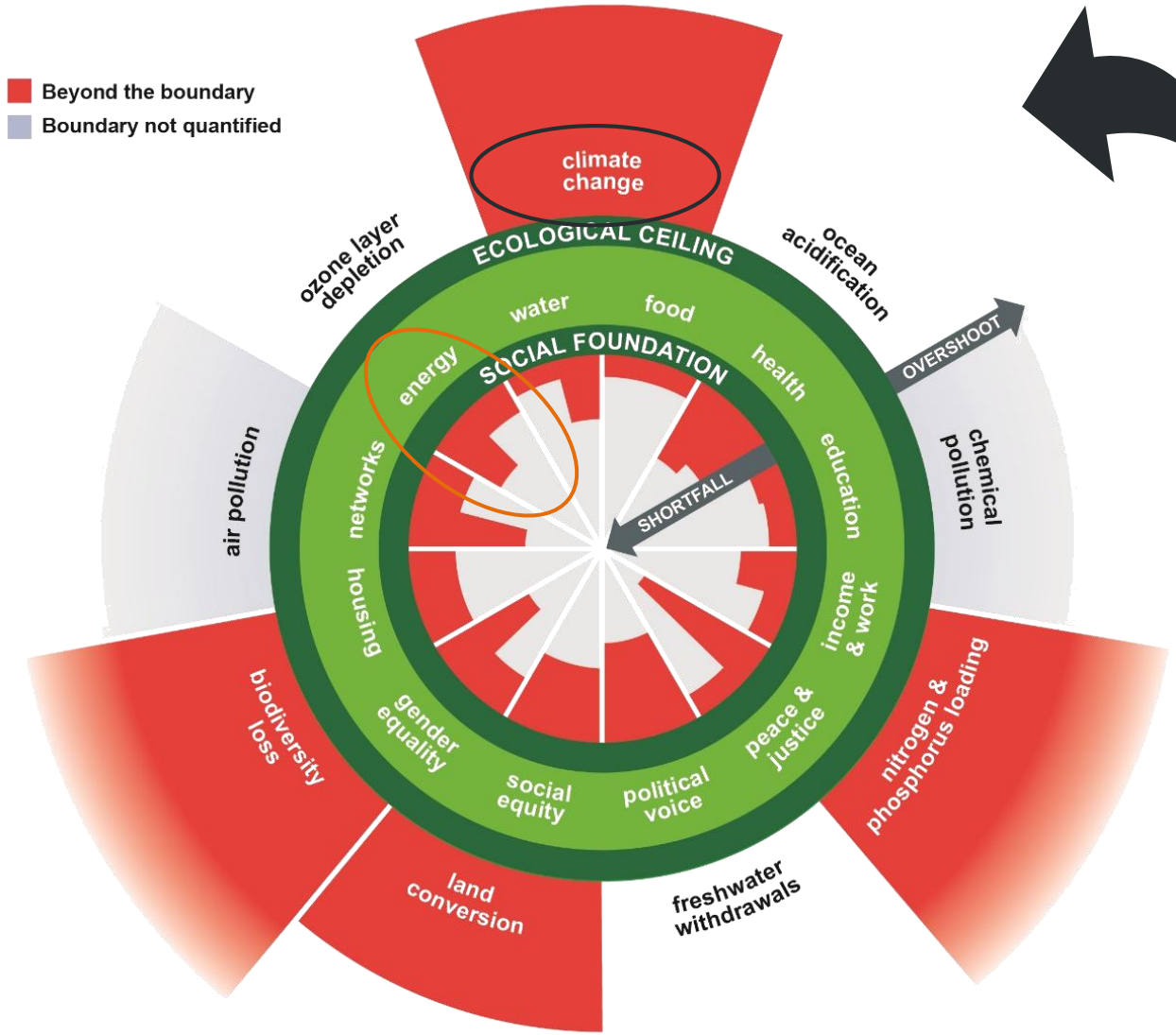
**Reviewers**
- Prof. Romain BOURDAIS, CentraleSupélec
- Prof. Bruno FRANÇOIS, Centrale Lille

**Jury president**
- Prof. Stephane Ploix, Grenoble-INP UGA

G2Elab Grenoble Génie Électrique Grenoble Electrical Engineering

UGA Université Grenoble Alpes | cnrs | GRENOBLE INP UGA

# The energy sector: a real enigma





Sources: Raworth, K., 2017. A Doughnut for the Anthropocene: humanity's compass in the 21st century. The Lancet Planetary Health 1, e48–e49. https://doi.org/10.1016/S2542-5196(17)30028-1 | Ge, M., Friedrich, J., Vigna, L., 2024. Where Do Emissions Come From? 4 Charts Explain Greenhouse Gas Emissions by Sector. https://www.wri.org/insights/4-charts-explain-greenhouse-gas-emissions-countries-and-sectors

# What is this data used for?

## Operational uses

- System planning and forecasting at different levels
- Energy system digitisation
- Leveraging AI for operations

## Knowledge and innovation

- Modelling of technical and human aspects of our energy systems
- Observing trends in energy use (individual, regional, or global)
- More open and reproducible science

Data serves to meet our **energy security** goals globally while **lowering** the **environmental**, **economic**, and **social** costs of our global energy system

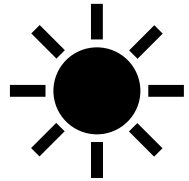


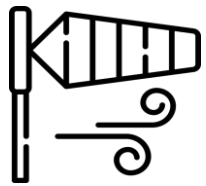
Source: Images by BulentYILDIZ de Pixabay

# But what is energy data?

## DATA HOLDERS

METEO FRANCE
À VOS CÔTÉS, DANS UN CLIMAT QUI CHANGE

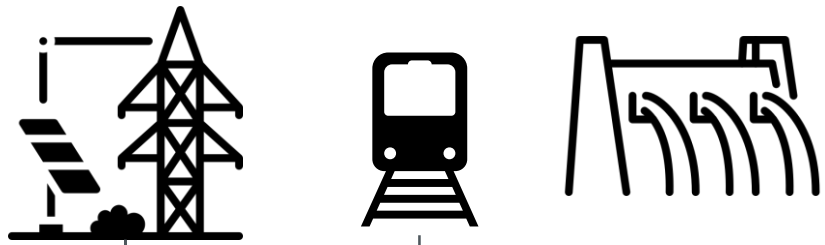




AGENCE ORE
Ensemble, maîtrisons la donnée d'énergie

Rte

GRDF



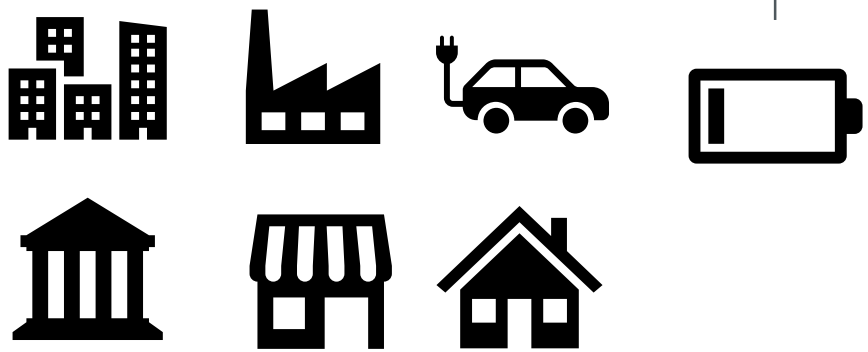




RENAULT

ENGIE

ENEDIS

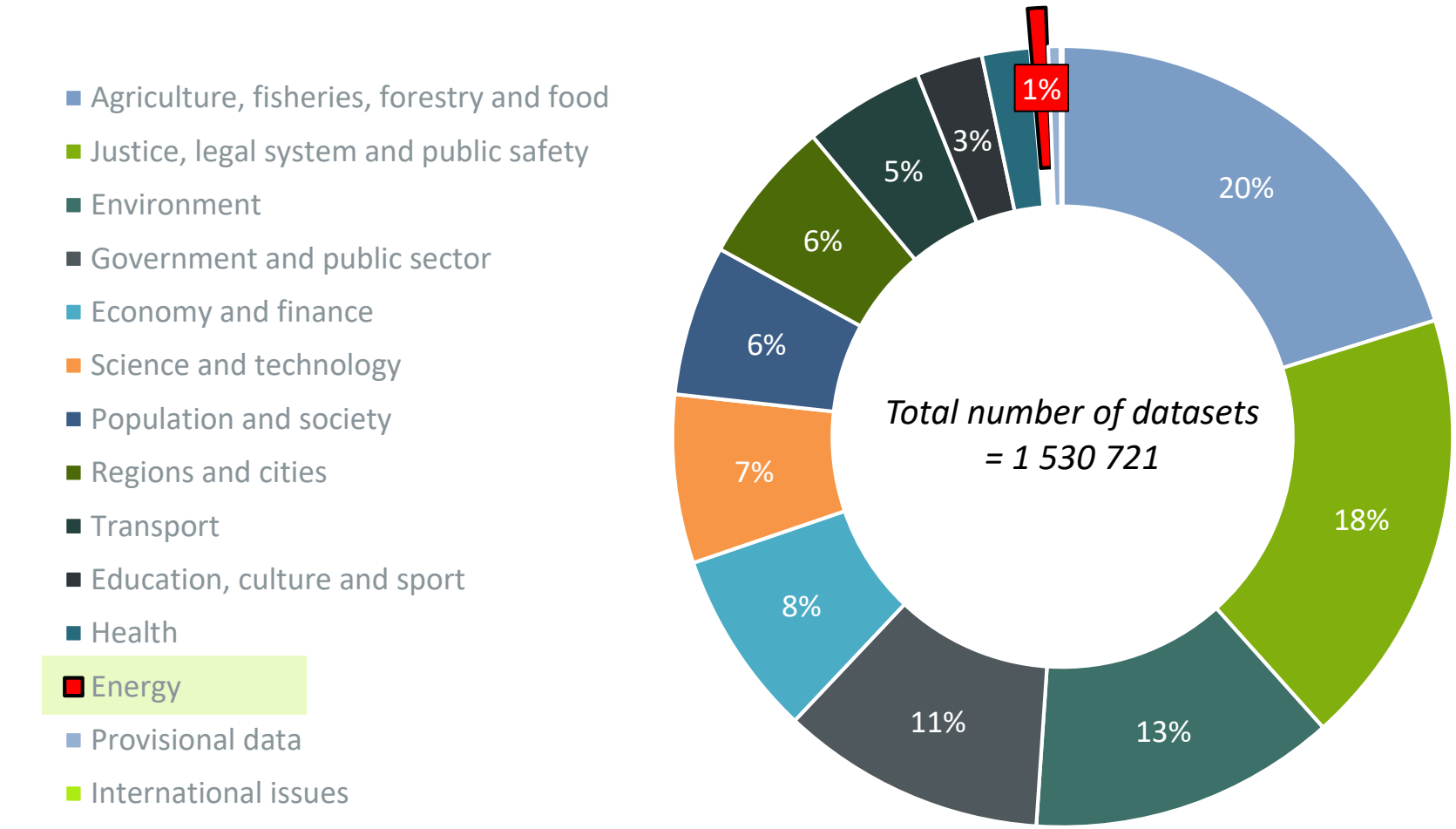


## TIERS

- **Commons**: Precipitation data, Solar radiation data
- **Shared and large-scale**: Data on the electricity grid, dam status
- **Private (highly variable)**: Individual building data, EV usage data

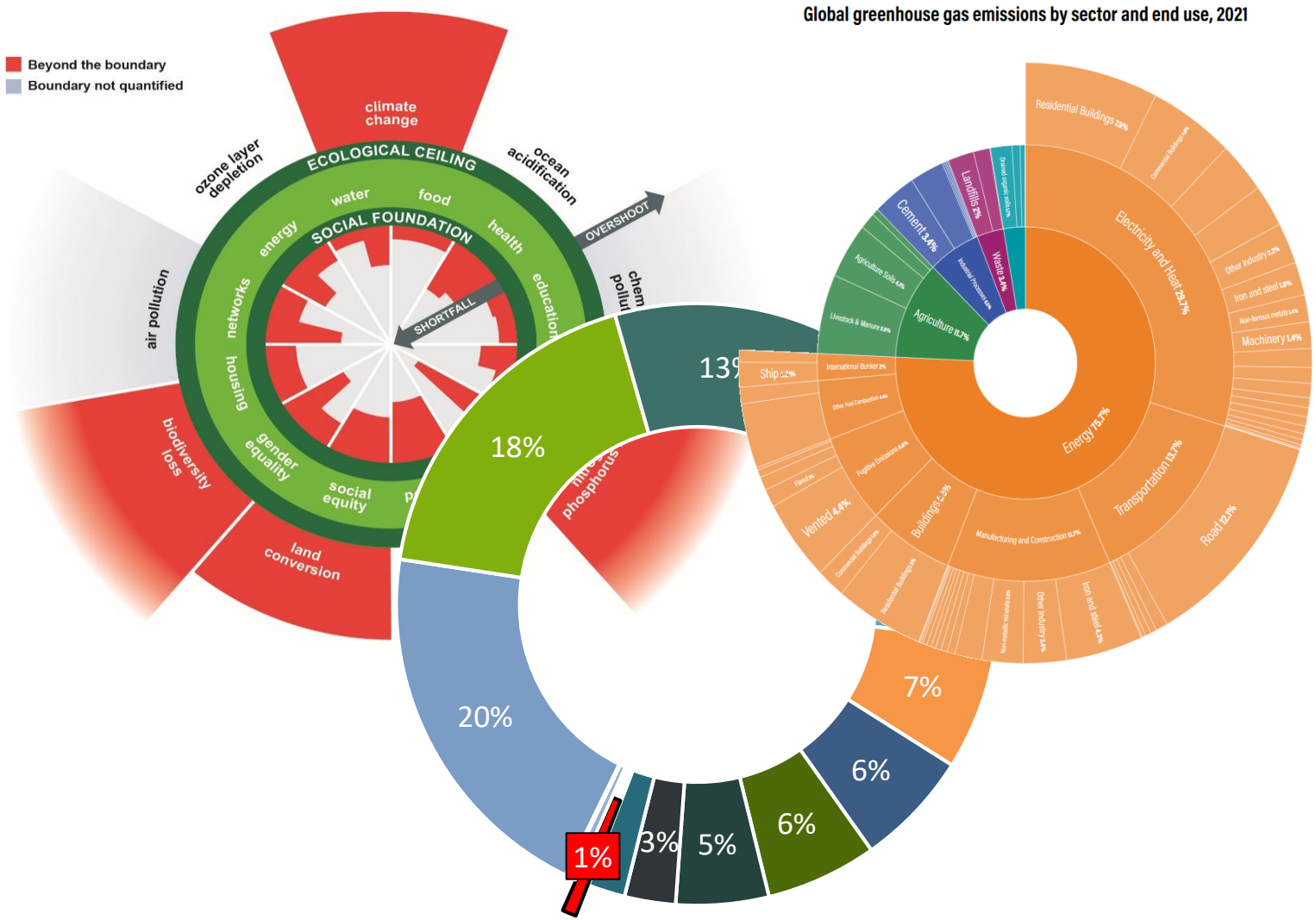
# The energy sector: a real enigma

**Percentage of total datasets by theme on the European open data catalogue (data.europa.eu)**




Source: "Data in figures | Data.Europa.Eu". 2017. Accessed 28 October 2025. https://data.europa.eu/catalogue-statistics/currentState/category?locale=en.

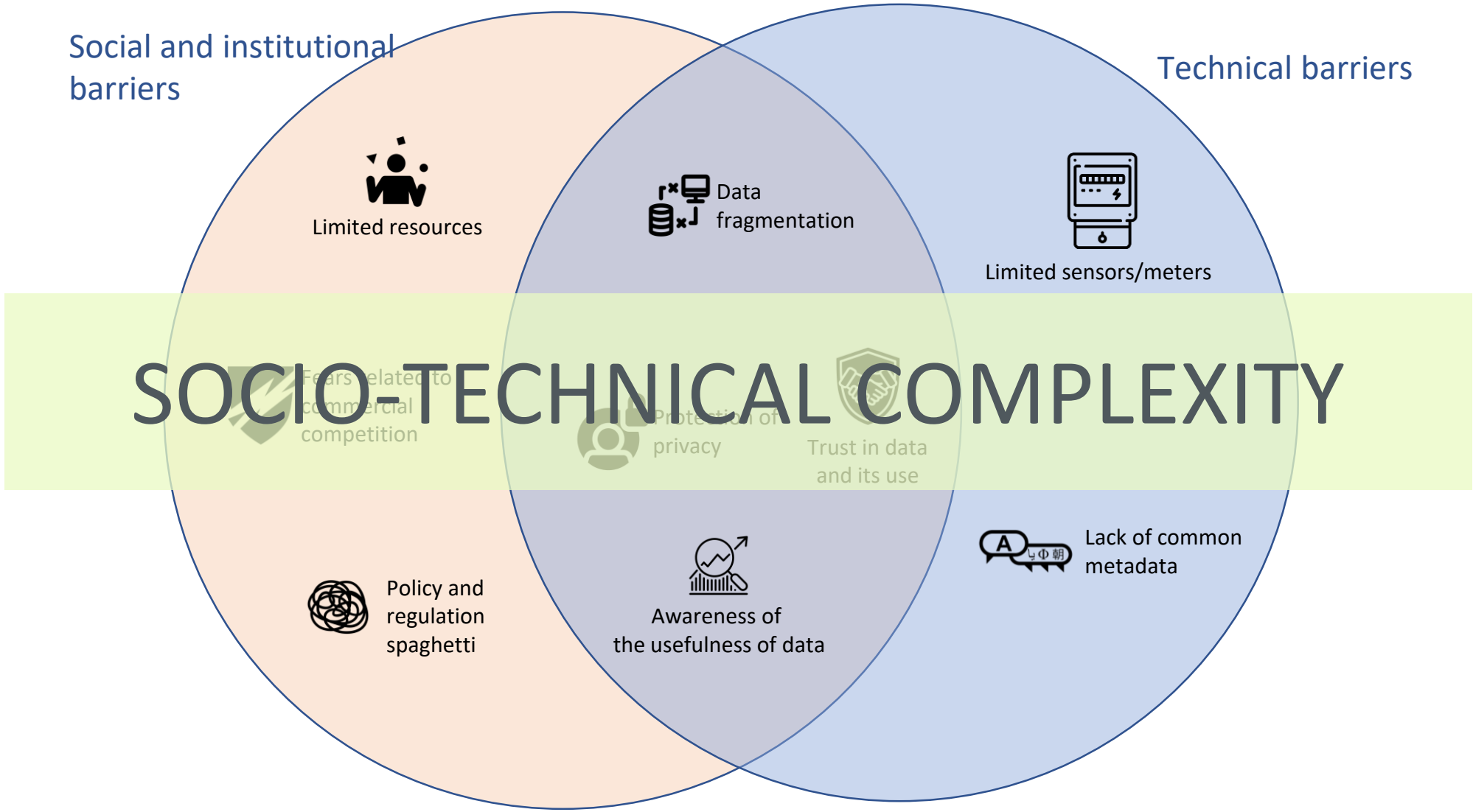
# The energy sector: a real enigma







# So why isn't it currently shared?



# My thesis seeks to answer three questions

## Our questions for today...

1. **HOW** can energy data be made **ACCESSIBLE**?
2. Will this be **VALID** for the **DIVERSITY** of energy data?
3. How will this **FACILITATE INTERACTIONS** between stakeholders?

## Presentation plan

1. Introduction
2. **Formalisation of the data lifecycle**
3. Experimentation based on six real-life case studies
   - i. Experimental setup
   - **ii. Impact of data typologies**
   - **iii. Supporting interactions between stakeholders**
4. Conclusions, limitations, and future work

# Agenda

- What a data lifecycle is?
- Common steps in most data lifecycles
- Suitability for the energy sector
- Proposition of a new data lifecycle

- Introduction
- **Formalisation of the data lifecycle**
- Experimentation based on six real-life case studies
  - *Experimental setup*
  - *Impact of data typologies*
  - *Support for interactions between stakeholders*
- Conclusions, limitations and prospects

G2Elab
Grenoble Génie Électrique
Grenoble Electrical Engineering

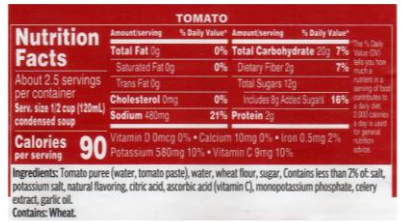
# What does "data availability" mean?

**Easy to find**

**Understandable**

No metadata (bad)


Natural language metadata (good)




Controlled vocabulary metadata (better)

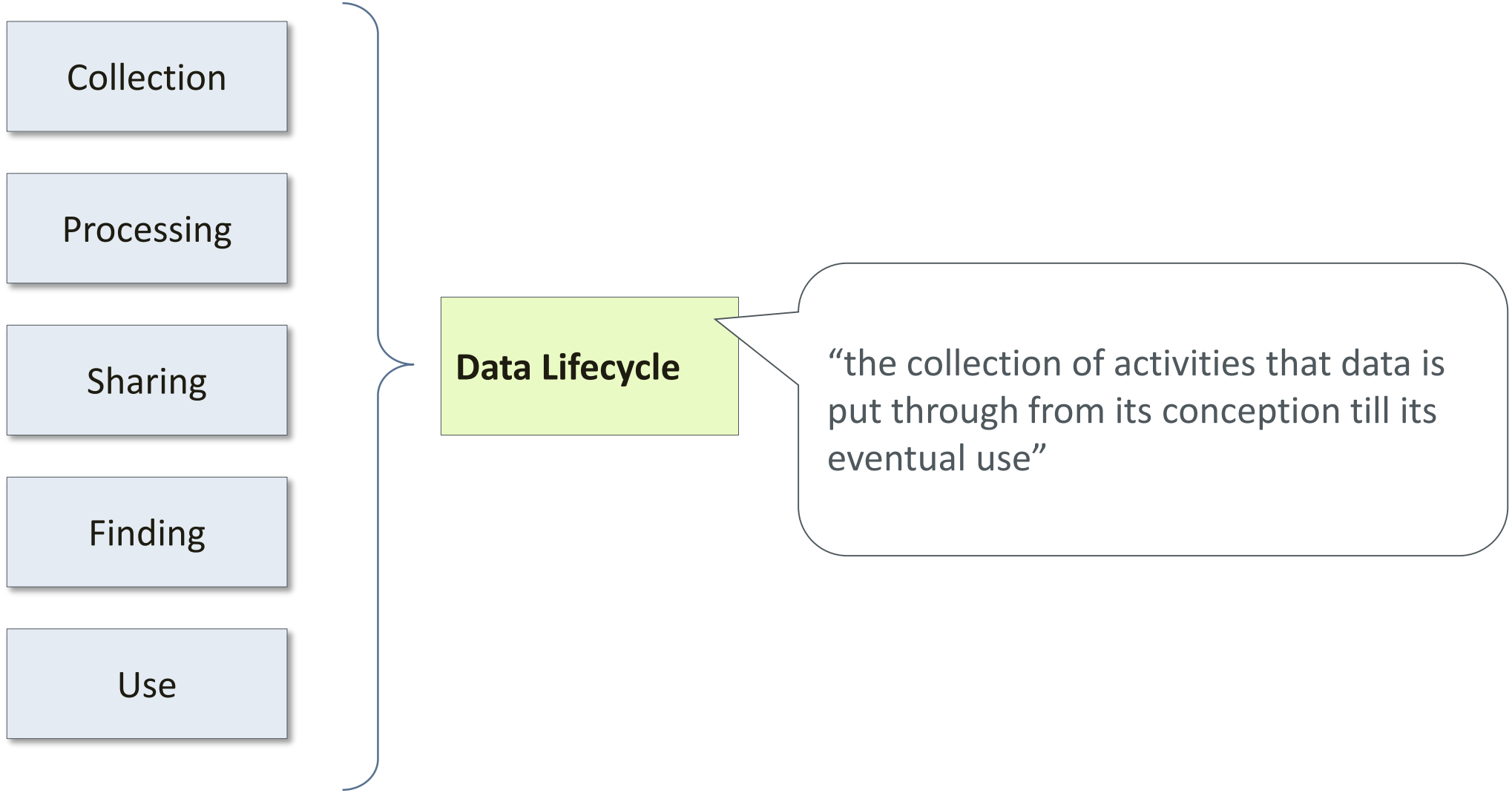
=

**Licence for use**

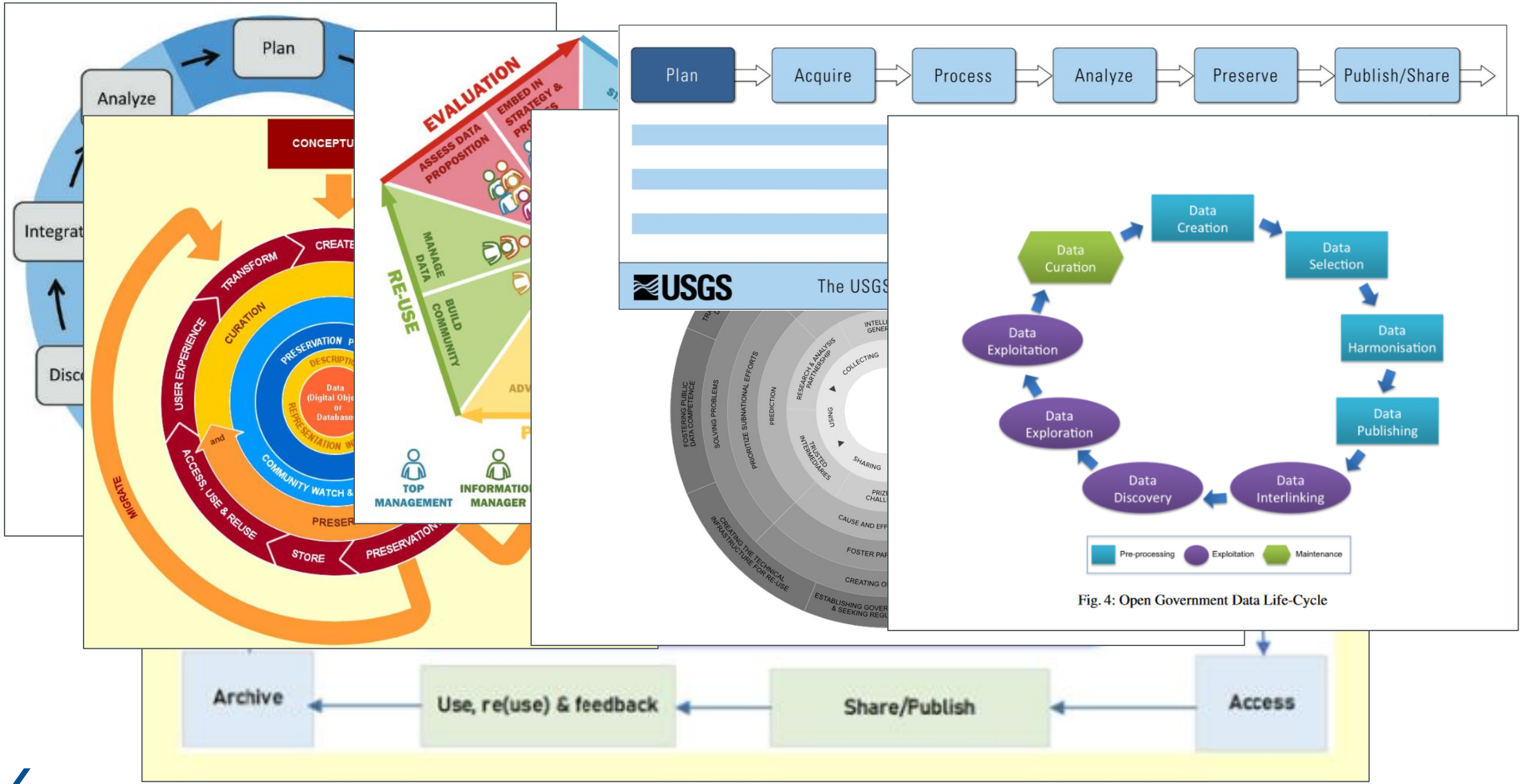



Sources: Shelves: Image par (Joenomias) Menno de Jong de Pixabay, Tins with metadata: Poulain, P., 2024. MDverse: Shedding Light on the Dark Matter of Molecular Dynamics Simulations. https://doi.org/10.5281/zenodo.12541180 ; Cash register: Wikicommons

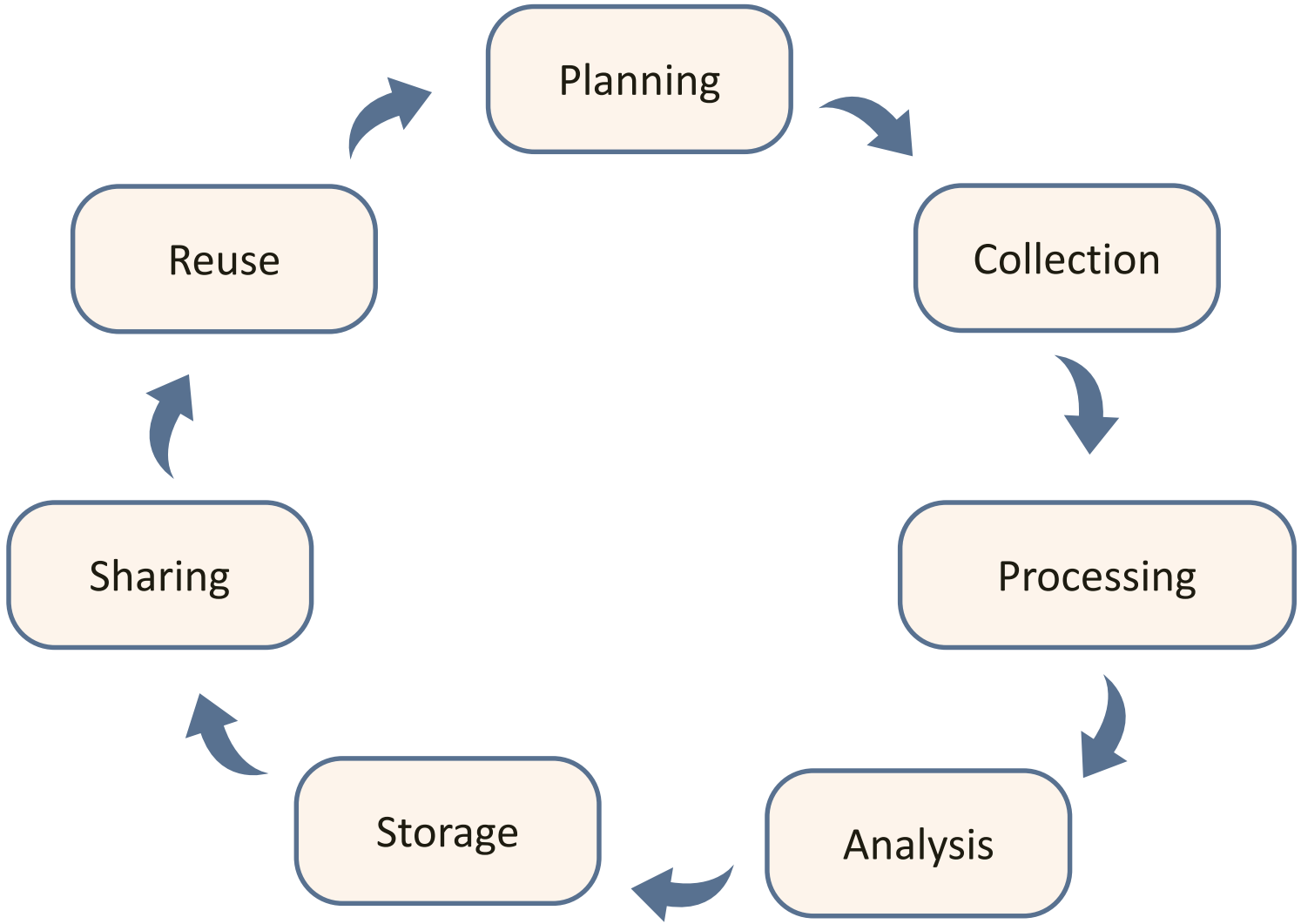
# What is a data lifecycle?



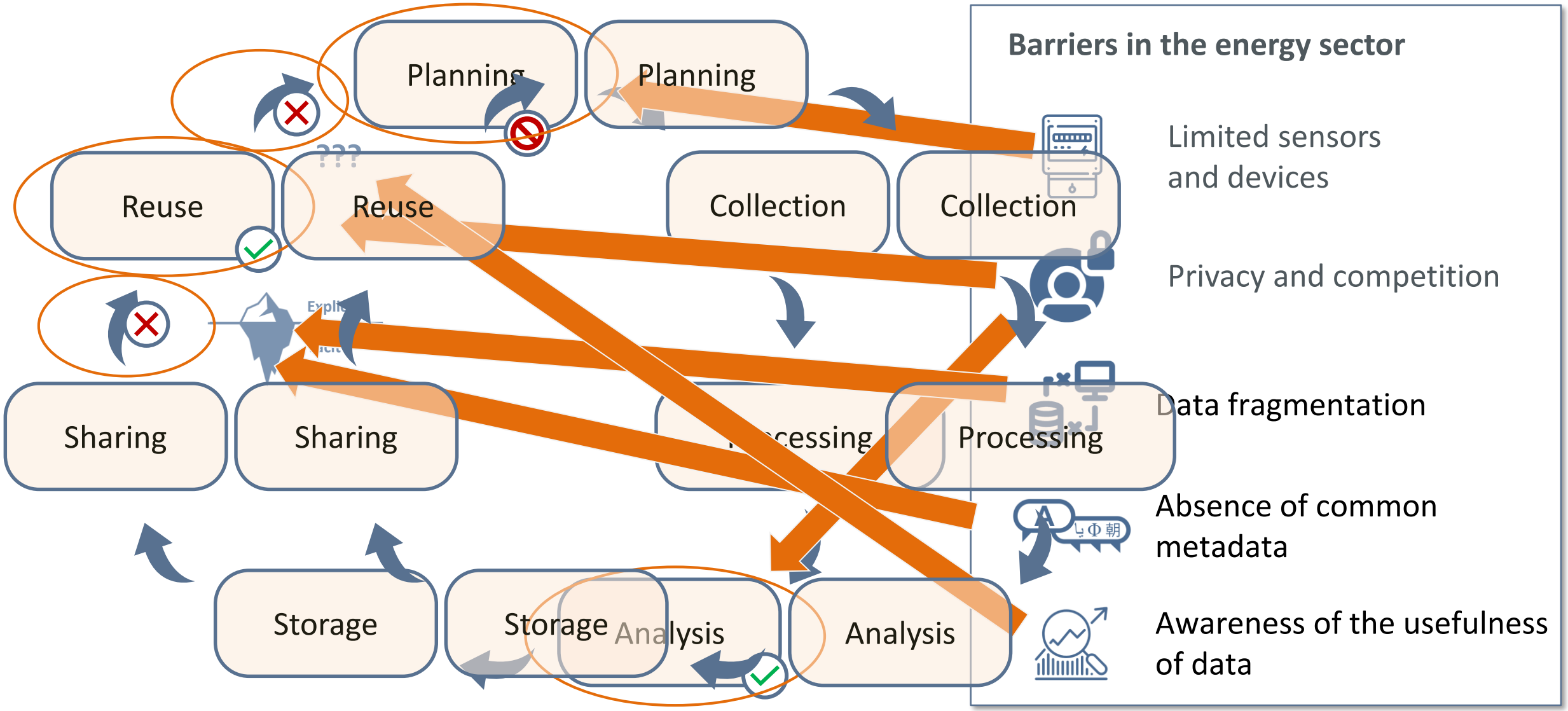
# There is a cacophony of data lifecycles in literature



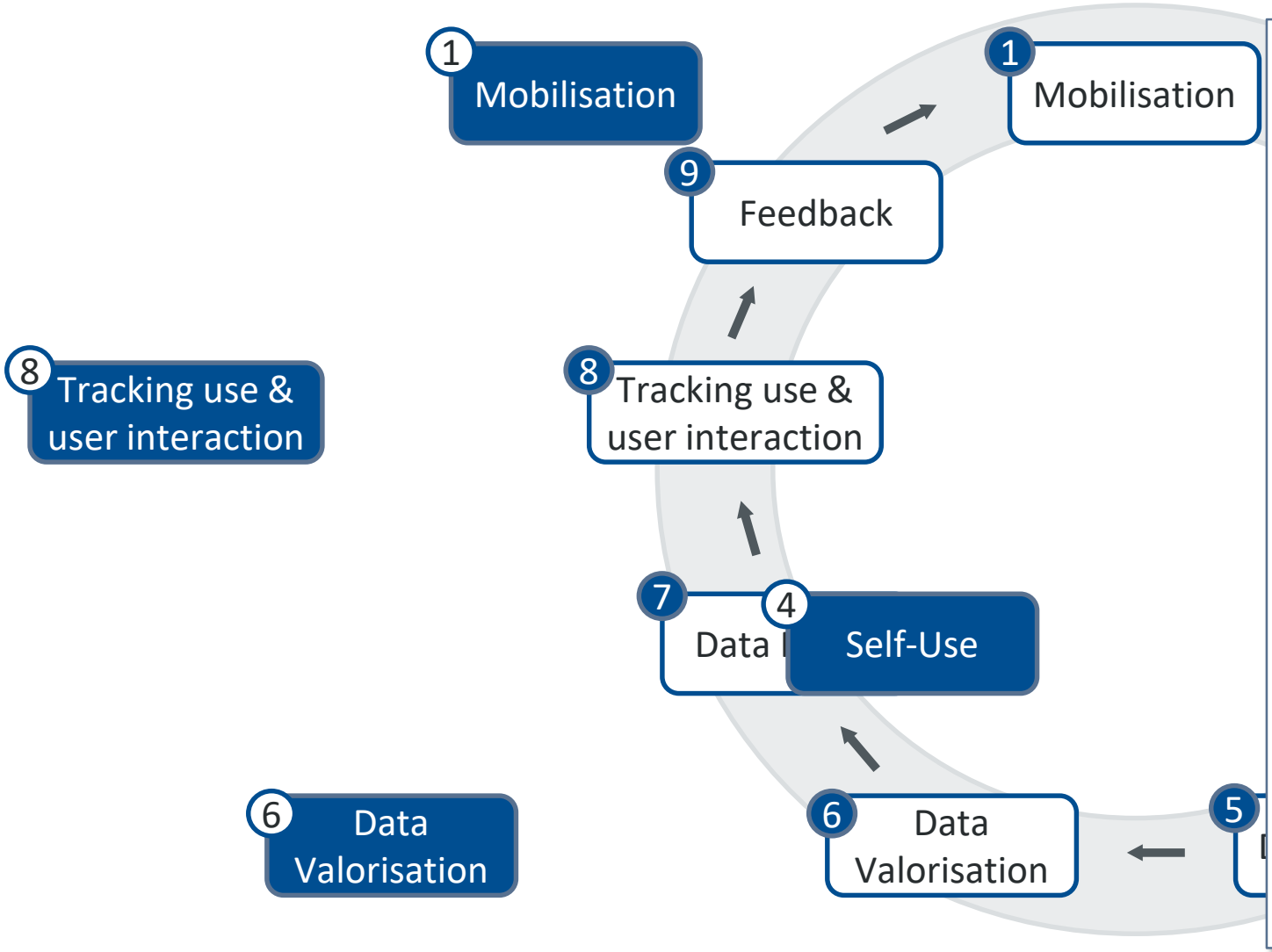
# The most common steps among data lifecycles are...



# However, some of these steps are not adapted to the energy sector



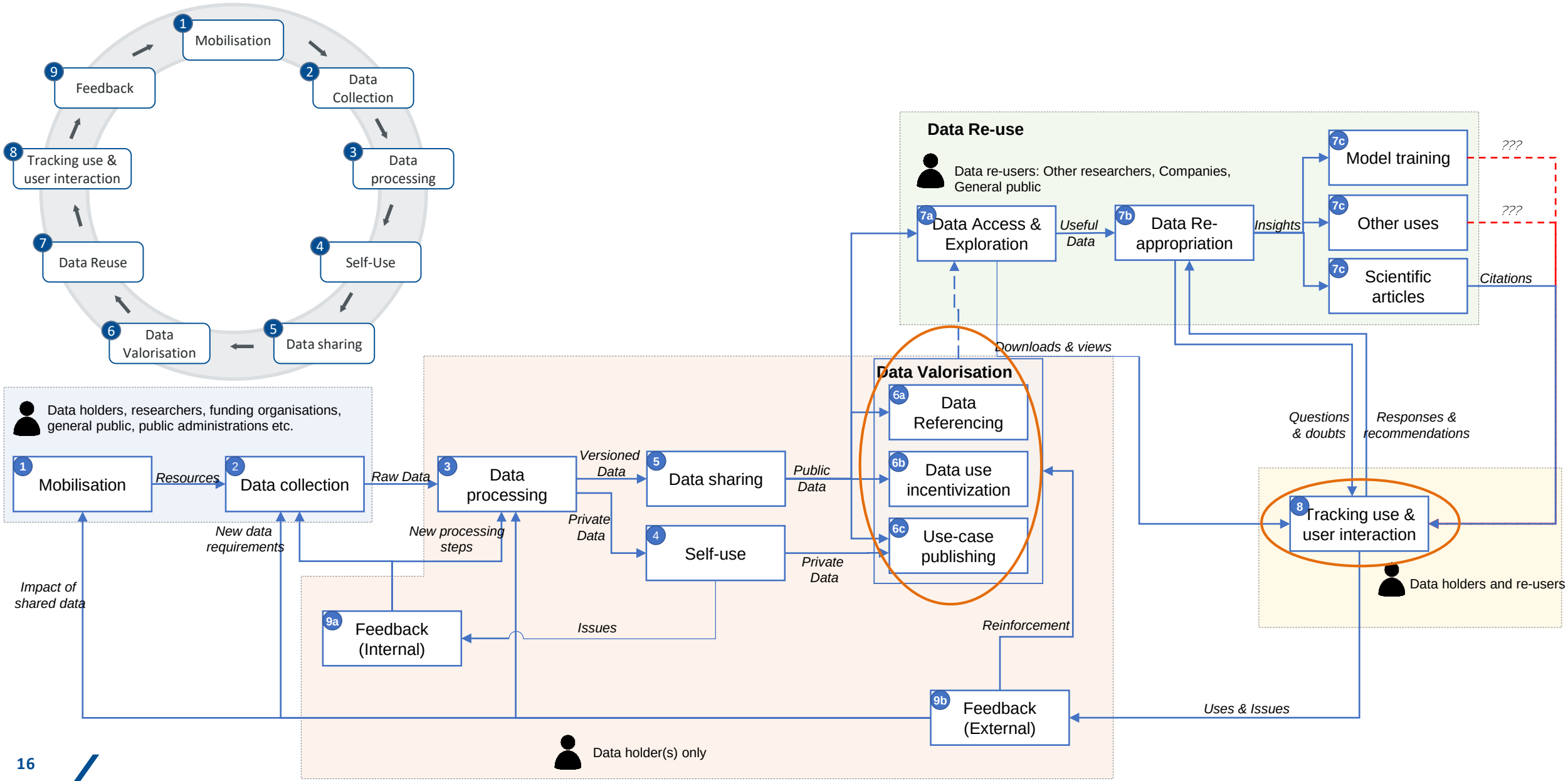
# Proposition of the thesis: Open Energy Data Lifecycle (OpEnDaLe)




## Changes with OpEnDaLe

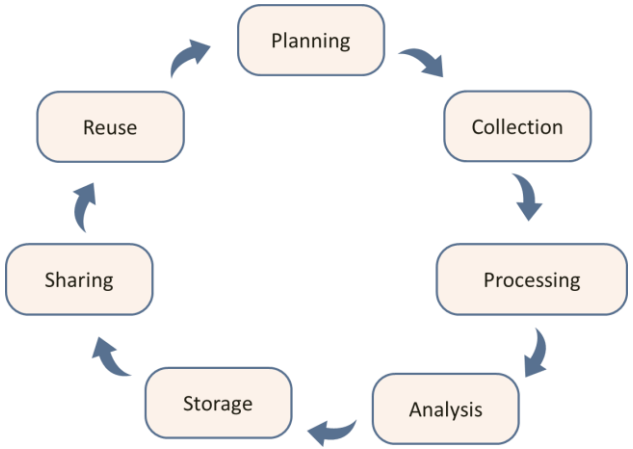
1. Extend "planning" to take into account the creation of new resources
4. Keep self-use separate from reuse
6. Encourage discovery of datasets through targeted activities
8. 
   - Gain more visibility into usage
   - Uncover implicit information between stakeholders

# Proposition: OpEnDaLe in full detail

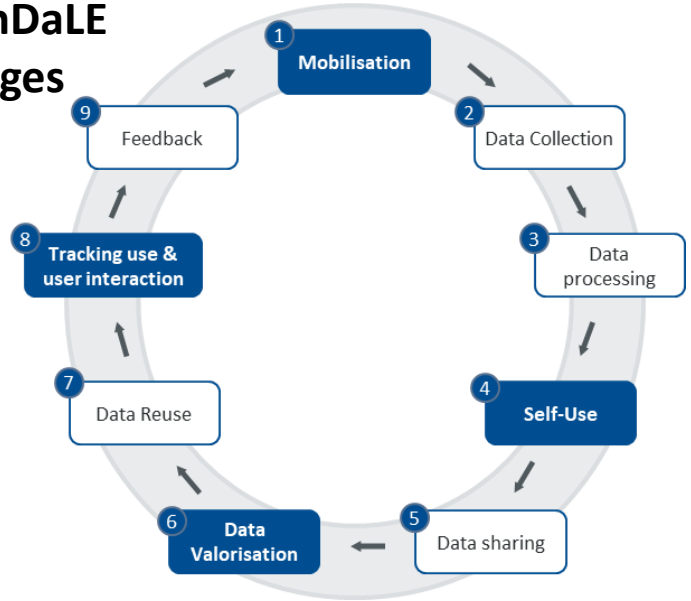


# On formalising a datalife cycle...

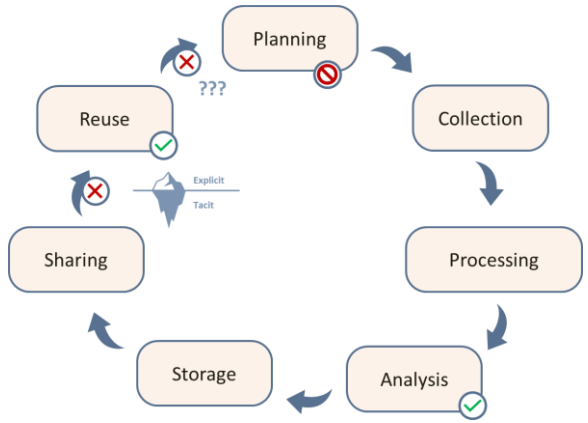
## Introduction to datalifecyles and a generalised model



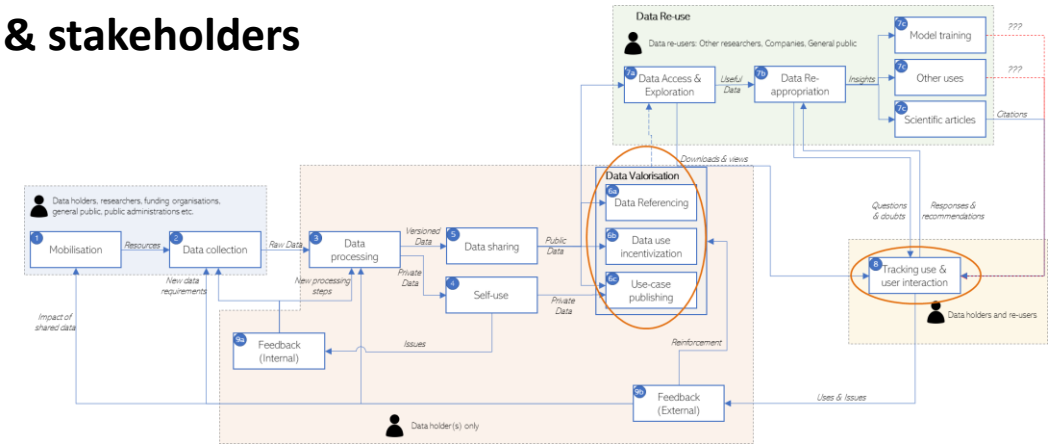

## My proposition: OpEnDaLE with the modified stages




## The shortcomings of this model for energy data




## OpEnDaLE in more detail flows & stakeholders



## Agenda

- Introduction
- Formalisation of the data lifecycle
- **Experimentation based on six real-life case studies**
  - ***Experimental setup***
  - *Impact of energy data typologies*
  - *Support for interactions between stakeholders*
- Conclusions, limitations and prospects

- OTE-UGA as an experimental plaform
- Domains of case-studies used to study OpEnDaLe

G2Elab
Grenoble Génie Électrique
Grenoble Electrical Engineering

# What is the Observatoire de la Transition Énergétique?

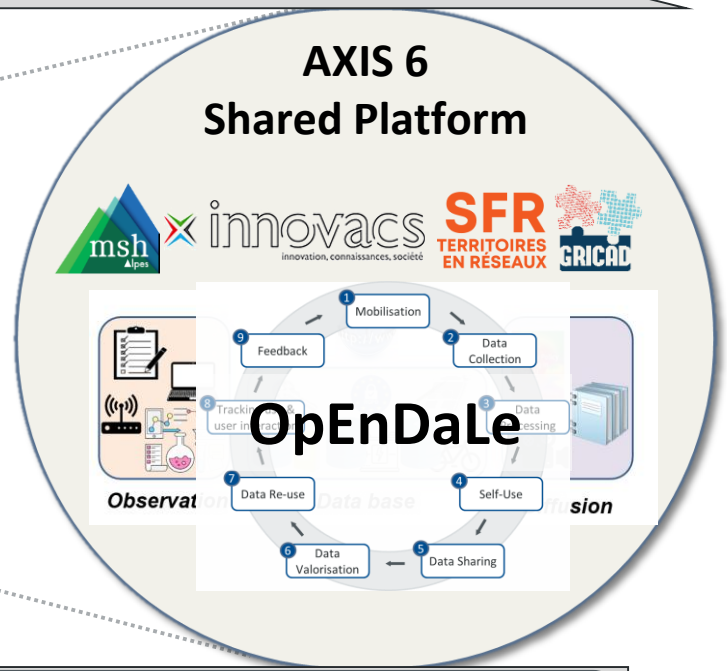
**OTE-UGA**

**Main advantages**

- Experimental structure I had easy access to
- Access to UGA resources, including a Data Protection Officer (DPO)
- Attractive non-profit status for panel
- Enables interaction with end users of energy

UGA

**AXIS 6**
**Shared Platform**




**Moderation "Sobriete" Research**
- Chaire Sobriete-Resilience
- ANR Satiable

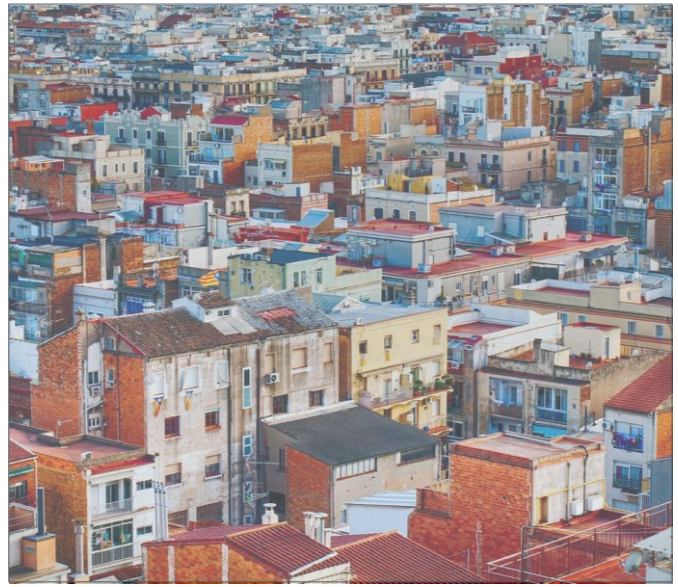
**Flexibility Research**
- FlexTASE project

# The datasets used as case studies covered different domains

## Electric Mobility


- EVE

## Residential buildings



- EtudELEC
- Etude xKY

## Tertiary buildings

- Predis-MHI Thermal
- GreEn-ER weather

In experimenting on the case studies, I worked on multiple aspects of the data lifecycles, including mobilisation, study design, data processing and sharing, and data valorisation, amongst others



Sources: Electric vehcles - Image from Pixabay; Residential buildings - Image from Pixabay; GreEn-ER buiding: Architecte du bâtiment:Groupe-6, Bruno Hallé — Milky2, CC BY-SA 3.0, https://fr.wikipedia.org/w/index.php?curid=9386438

# Agenda

- Introduction
- Formalisation of the data lifecycle
- **Experimentation based on six real-life case studies**
  - *Experimental setup*
  - ***Impact of data typologies***
  - *Support for interactions between stakeholders*
- Conclusions, limitations and prospects

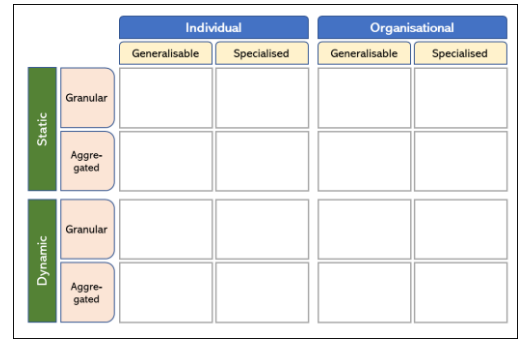
- Definition of data typology dimensions
- Mapping case studies to typology
- Comparison of case-study lifecycles
- Impact of typology dimensions on lifecycles

G2Elab
Grenoble Génie Électrique
Grenoble Electrical Engineering

# Proposed data typology dimensions and their attributes

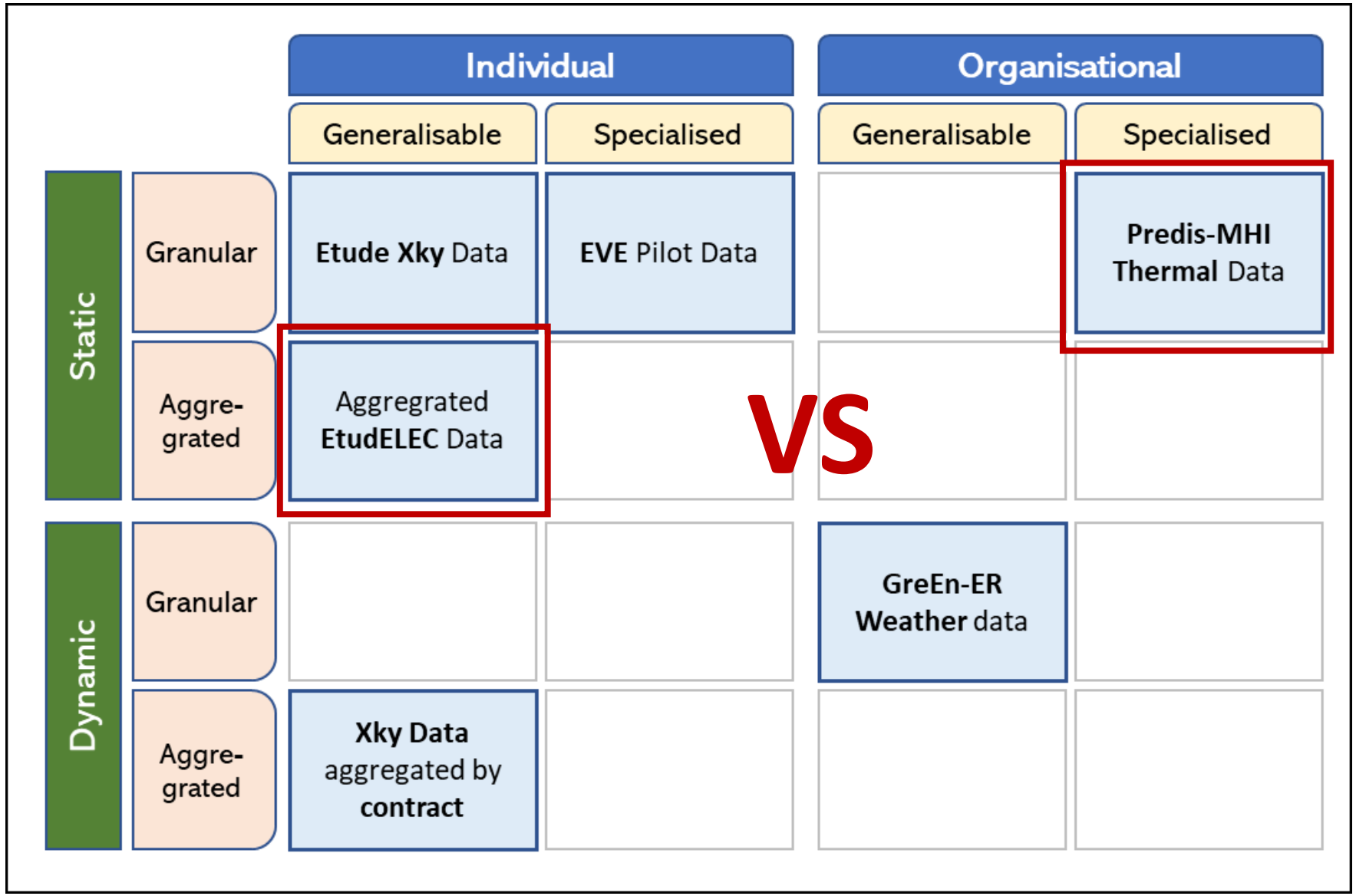
| Typology Dimensions | Dimension asks… | Typology attributes |
|---|---|---|
| Data subject | What does the data describe? | ▪ Individual<br>▪ Organisation |
| Temporality | Is the data fixed in time or does it change? | ▪ Static<br>▪ Dynamic/live |
| Granularity | How attributable is the data to a single data subject? | ▪ Granular (Meter/Entity)<br>▪ Aggregated |
| Prevalence of the concept | Is the data about a common concept or subject? | ▪ Generalisable<br>▪ Specialised |

**A 16-space typology grid**

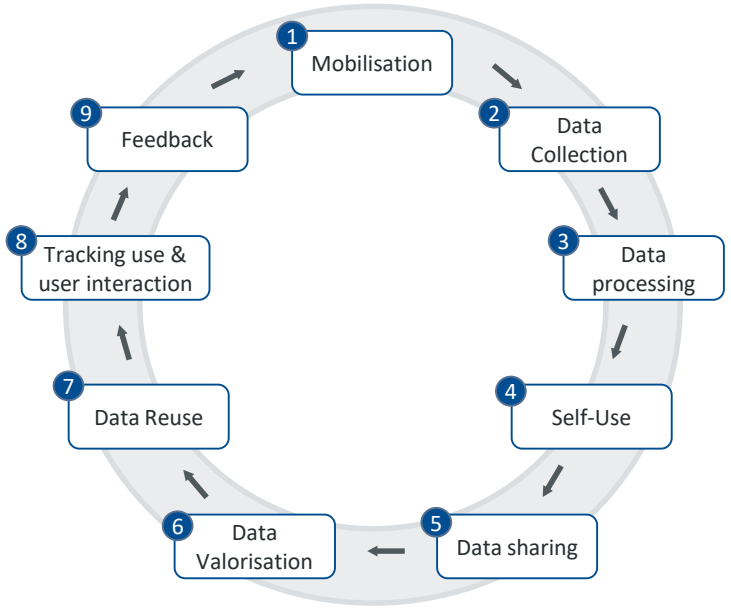


# The case studies fell into six different data typologies

- Data subject
- Temporality
- Granularity
- Prevalence of the concept

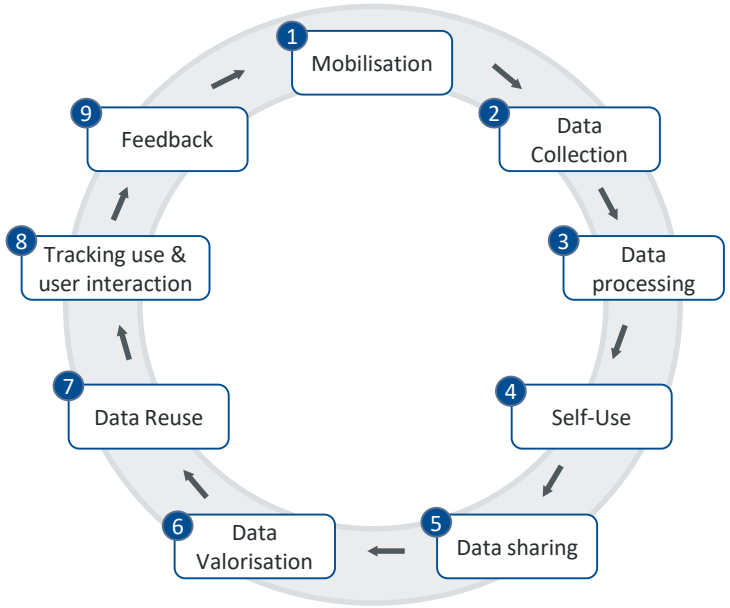


# Comparison of the data lifecycle for two cases




| | EtudELEC | Predis-MHI Thermal Data |
|---|---|---|
| | ▪ Metering infrastructure provided by 3rd party (DSO)<br>▪ Authorisations with all participants individually and with DSO (ENEDIS) | ▪ Metering infrastructure provided by the organisation<br>▪ Authorisations with one decision maker in the organisation (G2ELab) |
| | ▪ Personal data removal<br>▪ Aggregation to safeguard privacy | ▪ Limited personal data (check that none of the offices were single-occupied) |
| | ▪ Deposited on in a data repository with a DOI<br>▪ No data paper | ▪ Deposited on in a data repository with a DOI<br>▪ Data paper+ needed to provide more context |

# Comparison of the data lifecycle for two cases:

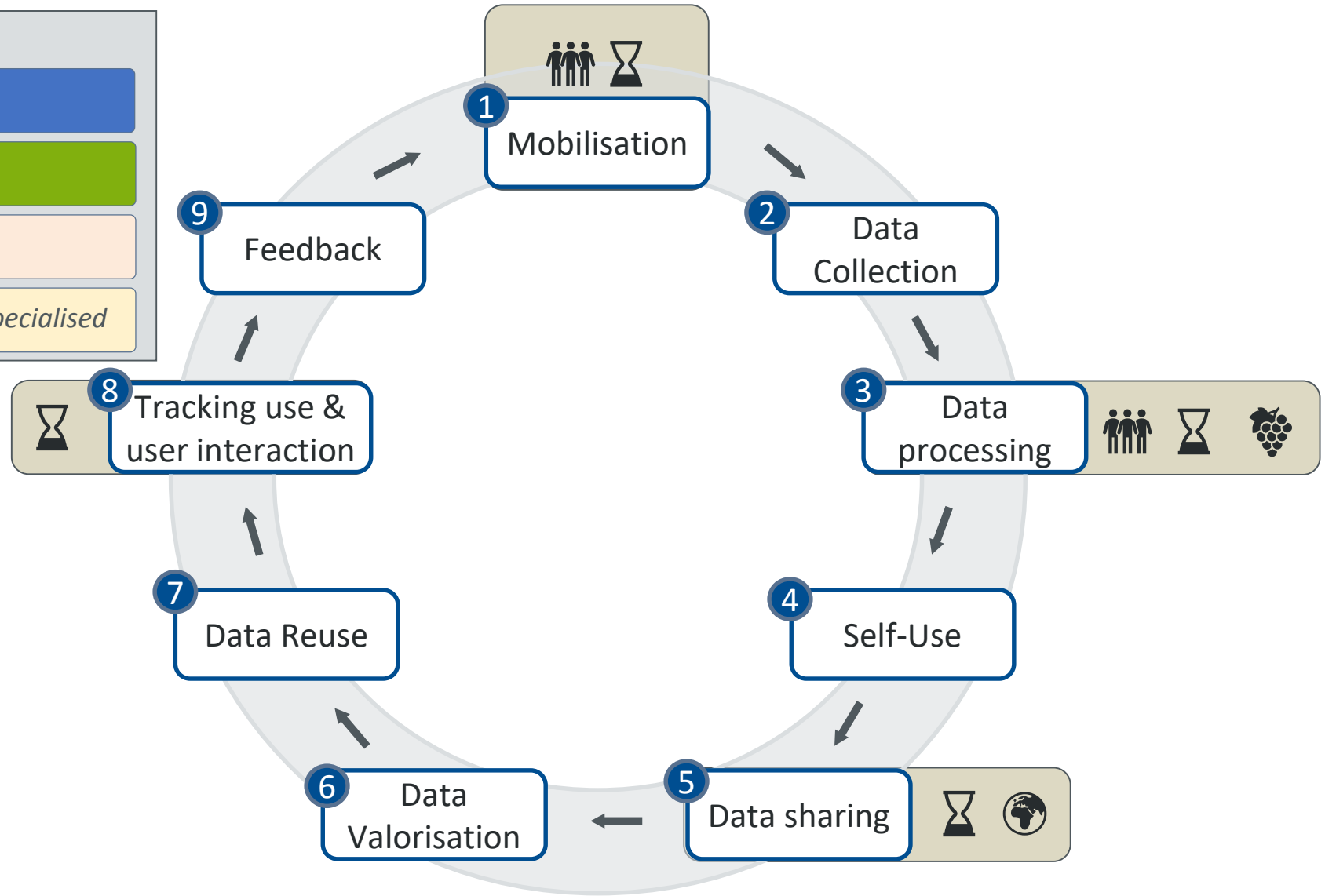



| | EVE | GreEn-ER Live Weather Data |
|---|---|---|
| | ▪ Metering infrastructure provided by the participants<br>▪ Authorisations with all participants individually | ▪ Metering infrastructure provided by the organisation<br>▪ Authorisations with one decision maker in the organisation (G2ELab) |
| | ▪ Personal data removal (VIN and personal data from questionnaires) | ▪ No personal data |
| | ▪ Deposited in a data repository with DOI<br>▪ Data paper+ needed to provide more context | ▪ Live data provided on a dashboard-style website<br>▪ Link and documentation deposited in a data repository with DOI |
| | ▪ Repository metrics (views, downloads, citations)<br>▪ Forum for discussion | ▪ Website metrics<br>▪ Repository metrics (views, downloads, citations)<br>▪ Forum for discussion |

# Impact of energy data typology dimensions on the life cycle

**Typology dimensions**

- Data subject: *Individual/Organisational*
- Temporality: *Static/Dynamic*
- Granularity: *Granular/Aggregated*
- Prevalence of concept: *Generalisable/Specialised*




**Tested with datasets**

- EVE
- EtudELEC
- GreEn-ER live weather data
- xKy data
- Live aggregated xKy data
- Predis-MHI Thermal Data

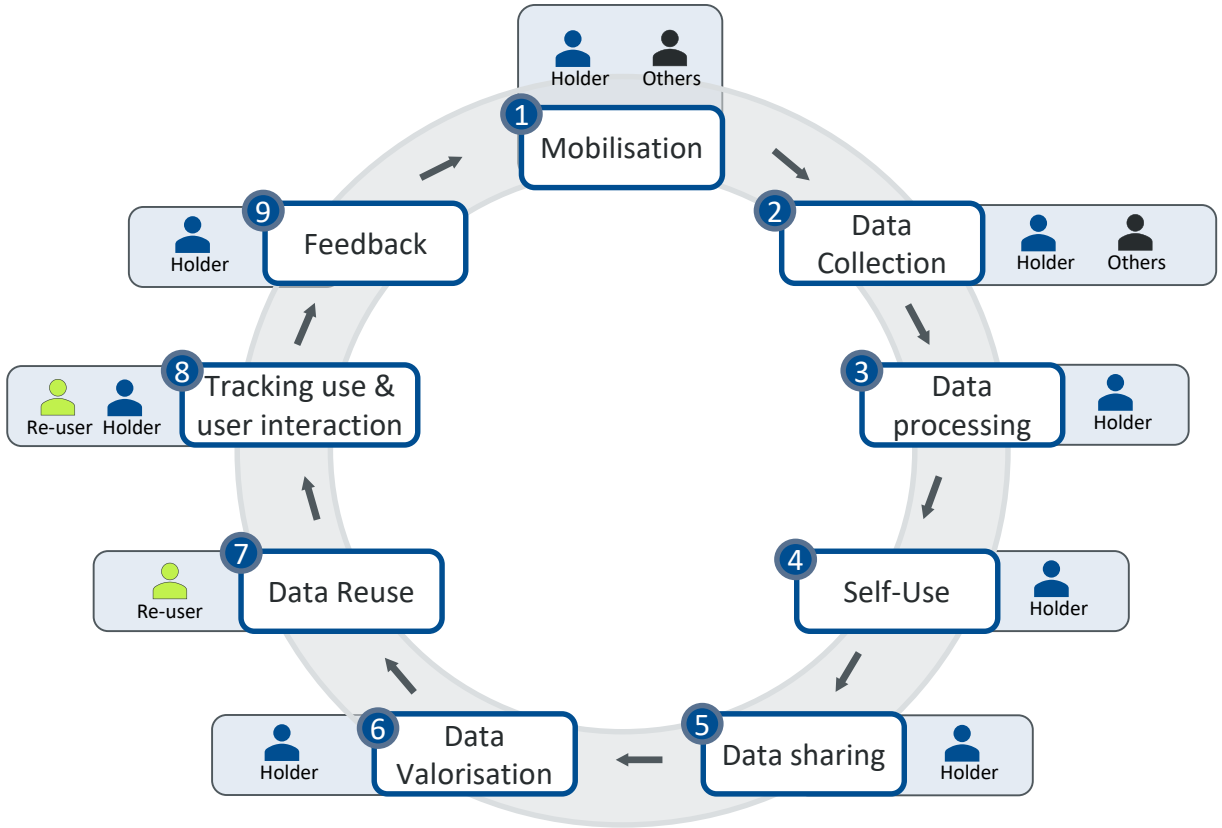
# Agenda

- Introduction
- Formalisation of the data lifecycle
- **Experimentation based on six real-life case studies**
  - *Experimental setup*
  - *Impact of data typologies*
  - ***Support for interactions between stakeholders***
- Conclusions, limitations and prospects

- Tools used for producer-user interactions
- OTE Forum and its interactions
- Co-creation of data using the forum
- Measuring certain metrics with a dashboard
- Quantifying valorisation impact: VALENS metrics

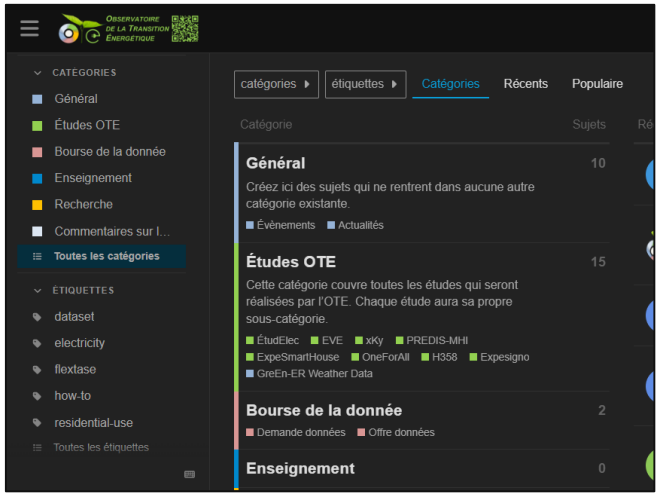
G2Elab
Grenoble Génie Électrique
Grenoble Electrical Engineering

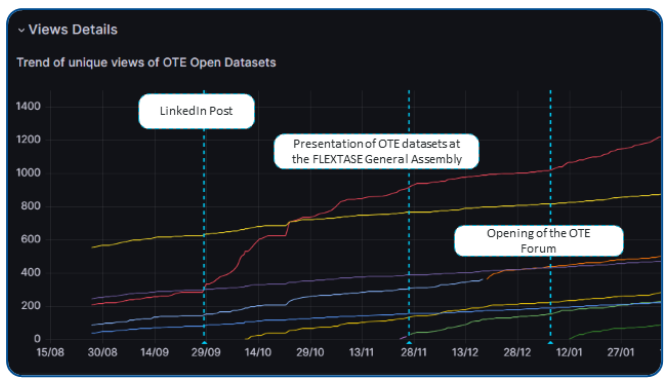
# There are different stakeholders for the stages of the lifecycle

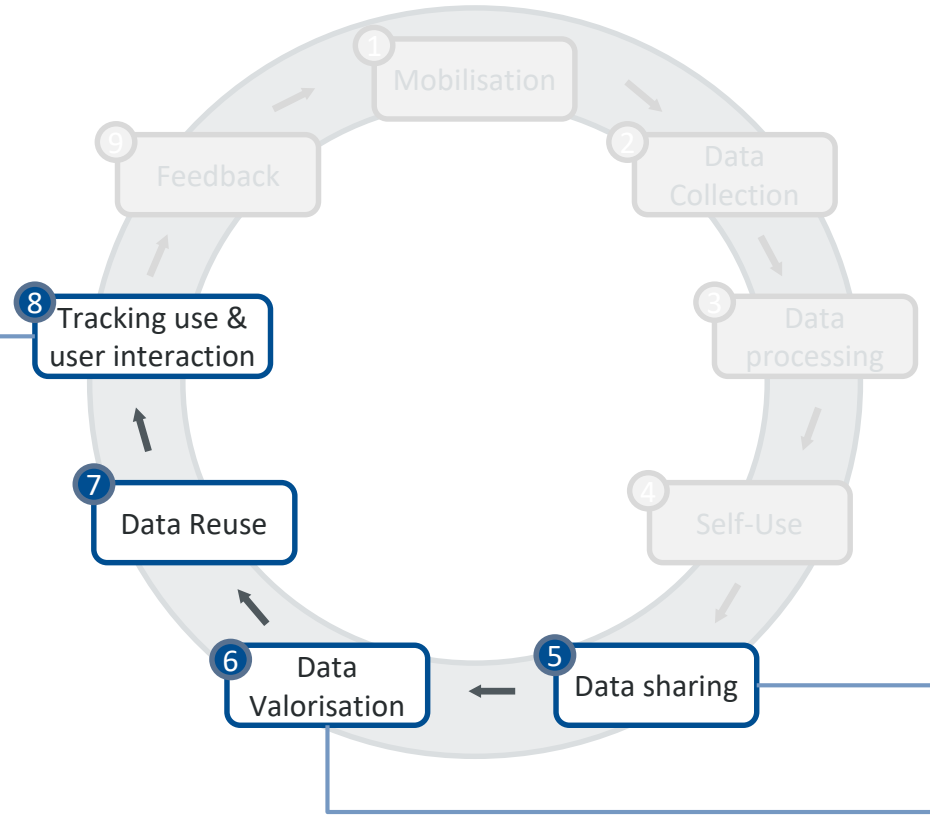


# Let’s focus on the tools that enable producer-user interactions

## OTE Forum & Metrics dashboard



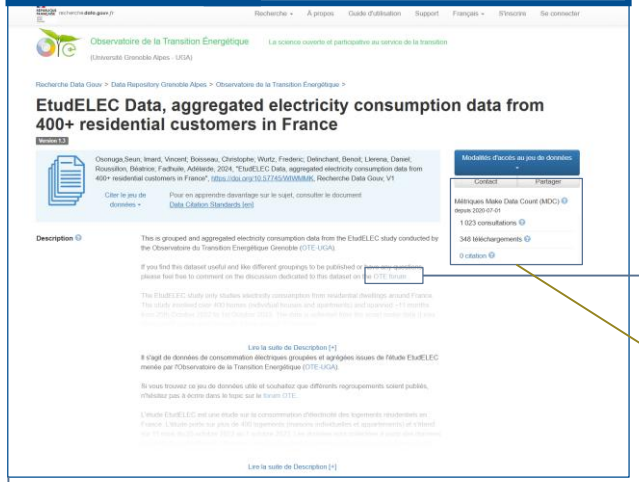




## Recherche.data.gouv

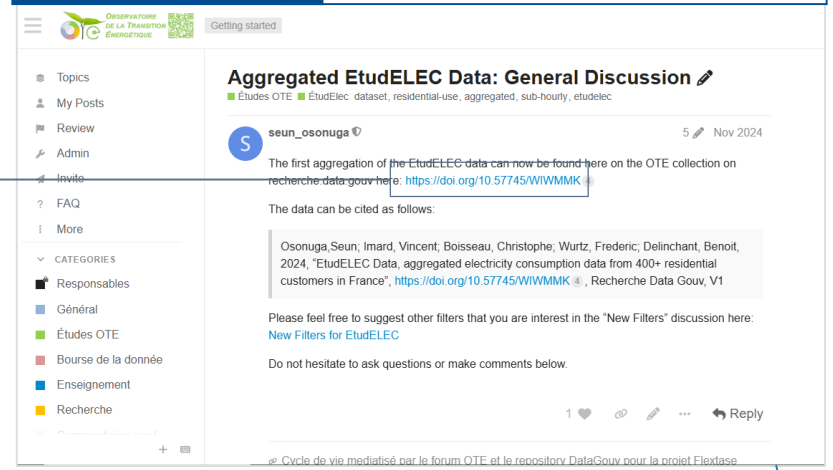

## LinkedIn amongst other

# Information exchange and references between these tools

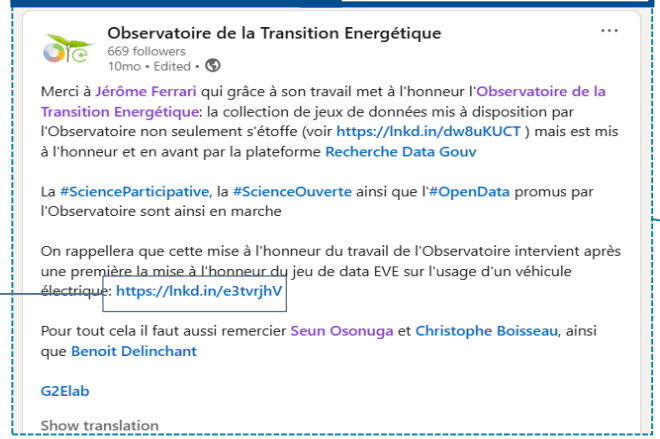
**Data sharing** Recherche.data.gouv



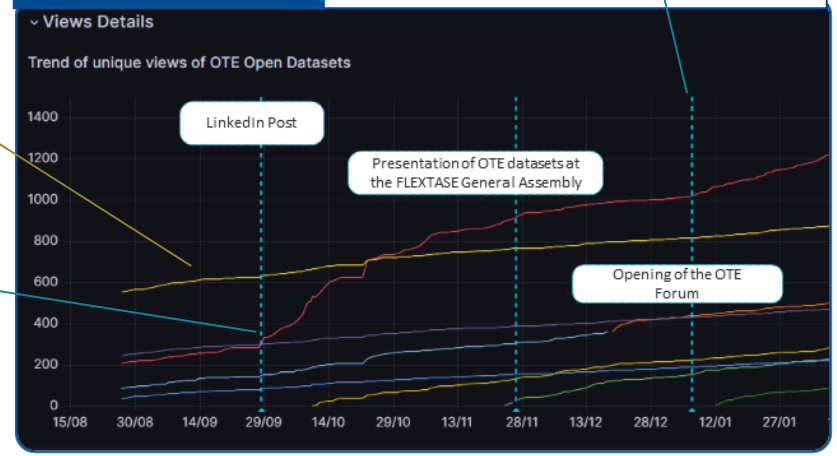

**User interaction** OTE Forum



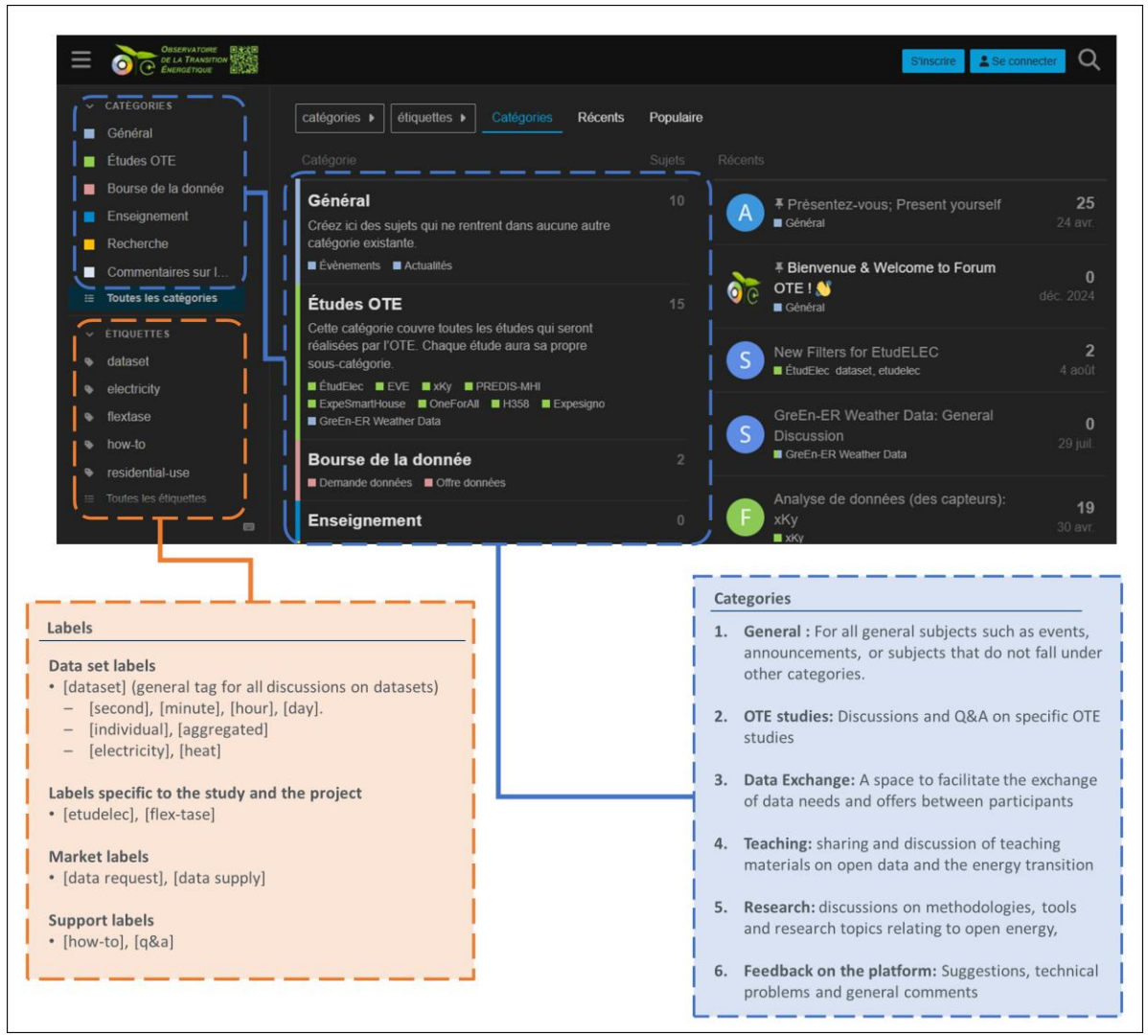

**Data Valorisation** LinkedIn




**Tracking use** Metrics dashboard



# What is the OTE FORUM?




## Key features

- 51 users (to date)
- Based on Discourse (open source software) with the following features:
  - Discussions
  - Customised notifications and periodic summaries
  - Direct messaging

**Source**: https://forum-ote.univ-grenoble-alpes.fr/

# Interactions between stakeholders on the forum

## Roles/characters

Data users

Data producers

Other interested parties

Forum moderators and administrators

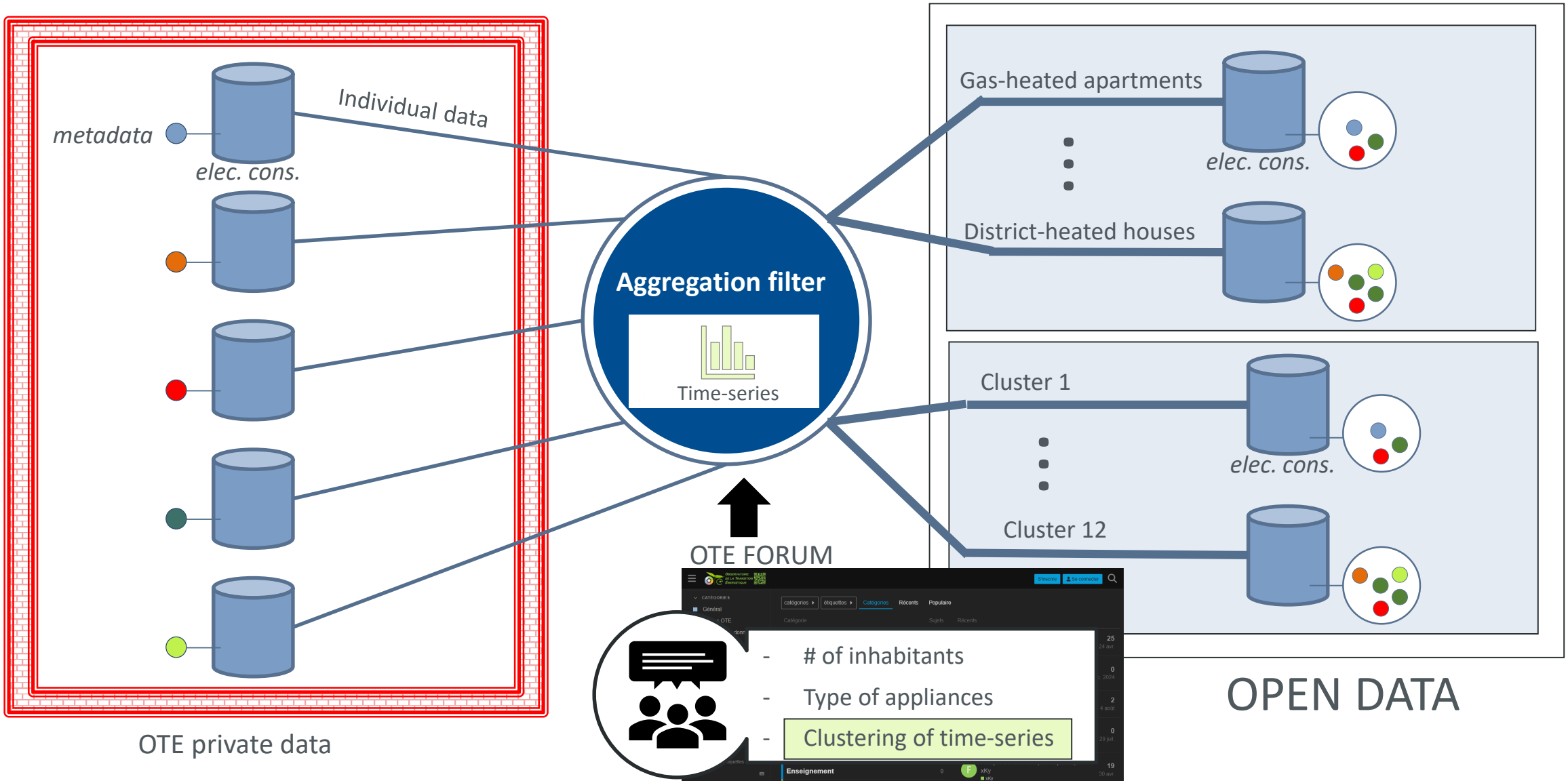
## Witnesses to interactions

- Clarifications between producers and users
- Producer-user corrections
- Collaboration between users
- Collaborations between interested parties
- Co-creation of data by producers and users

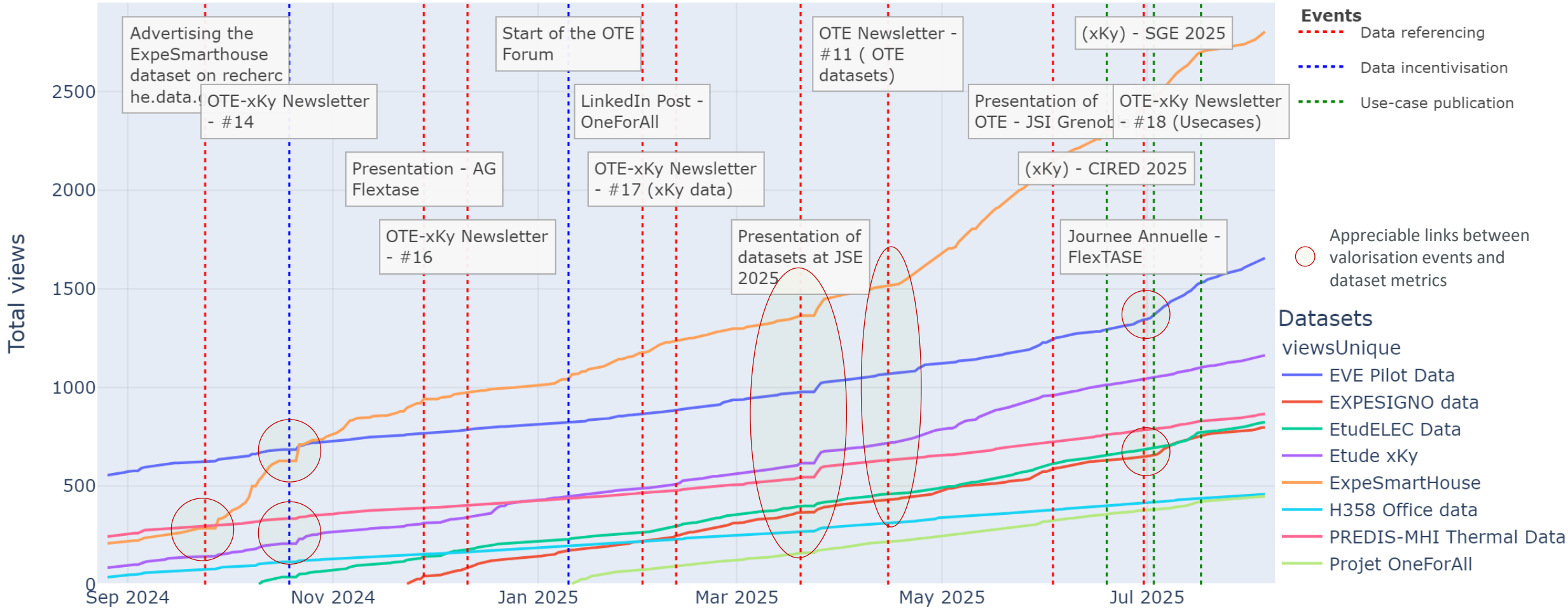


**Source**: https://forum-ote.univ-grenoble-alpes.fr/

# Co-creation of data with users: An EtudELEC example

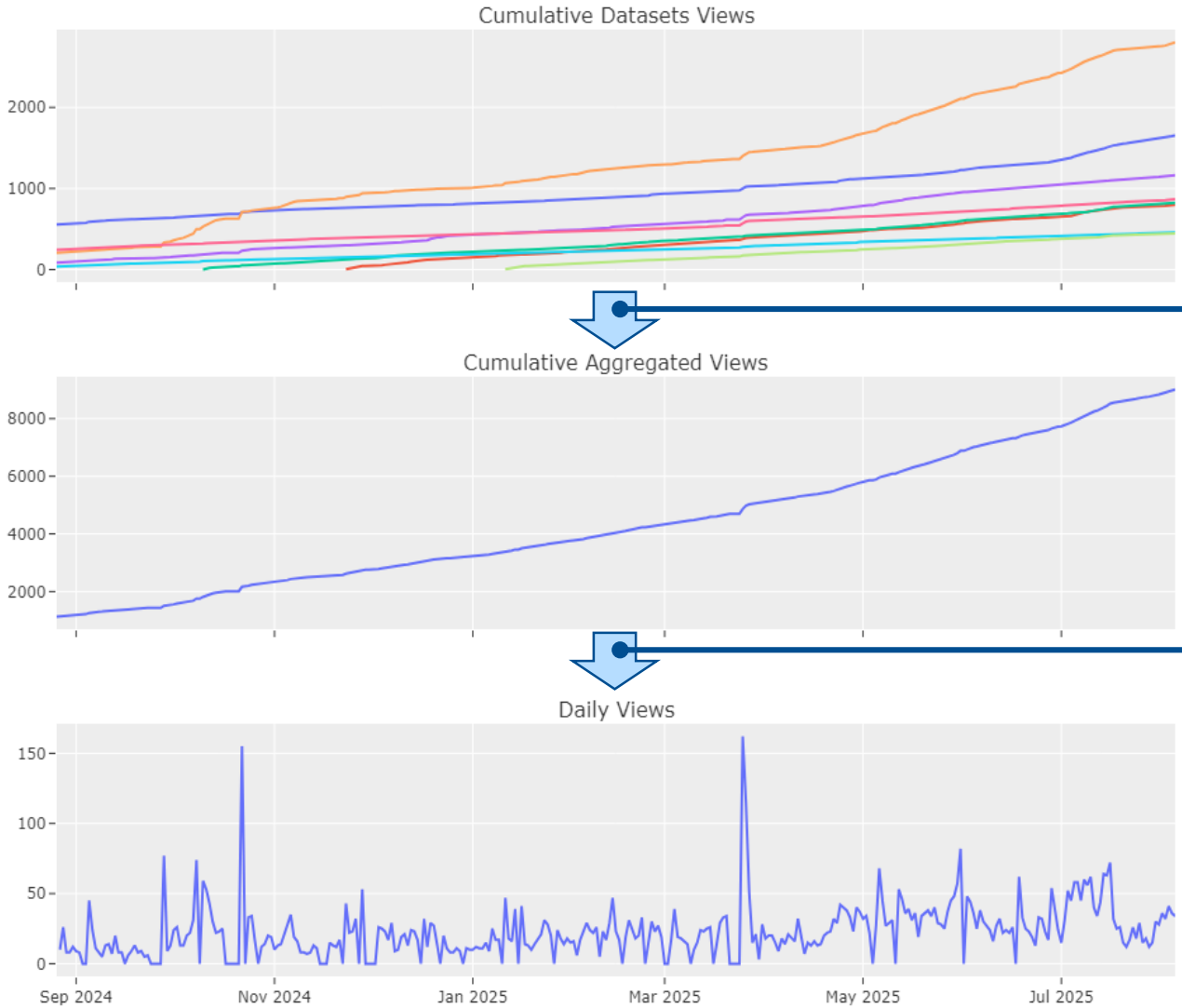


# The metrics dashboard




- Metrics for each dataset are collected daily via the recherche.data.gouv API, where valuation events are manually entered into a database.
- Similar dashboards are created for downloads and citations.

# VALENS indicators: data processing




Aggregated the metrics for all datasets as often valorisation activities affect multiple datasets

Go from cumulative values to daily values to better quantify impact

# Quantifying the impact of valorisation events: VALENS

## VALorisation Events Normalised Score

**Lift percentage**

Comparison of the daily metrics in between pre and post windows

$$Lift\% = \frac{Avg_{post\_window} - Avg_{pre\_window}}{Avg_{pre\_window}}$$

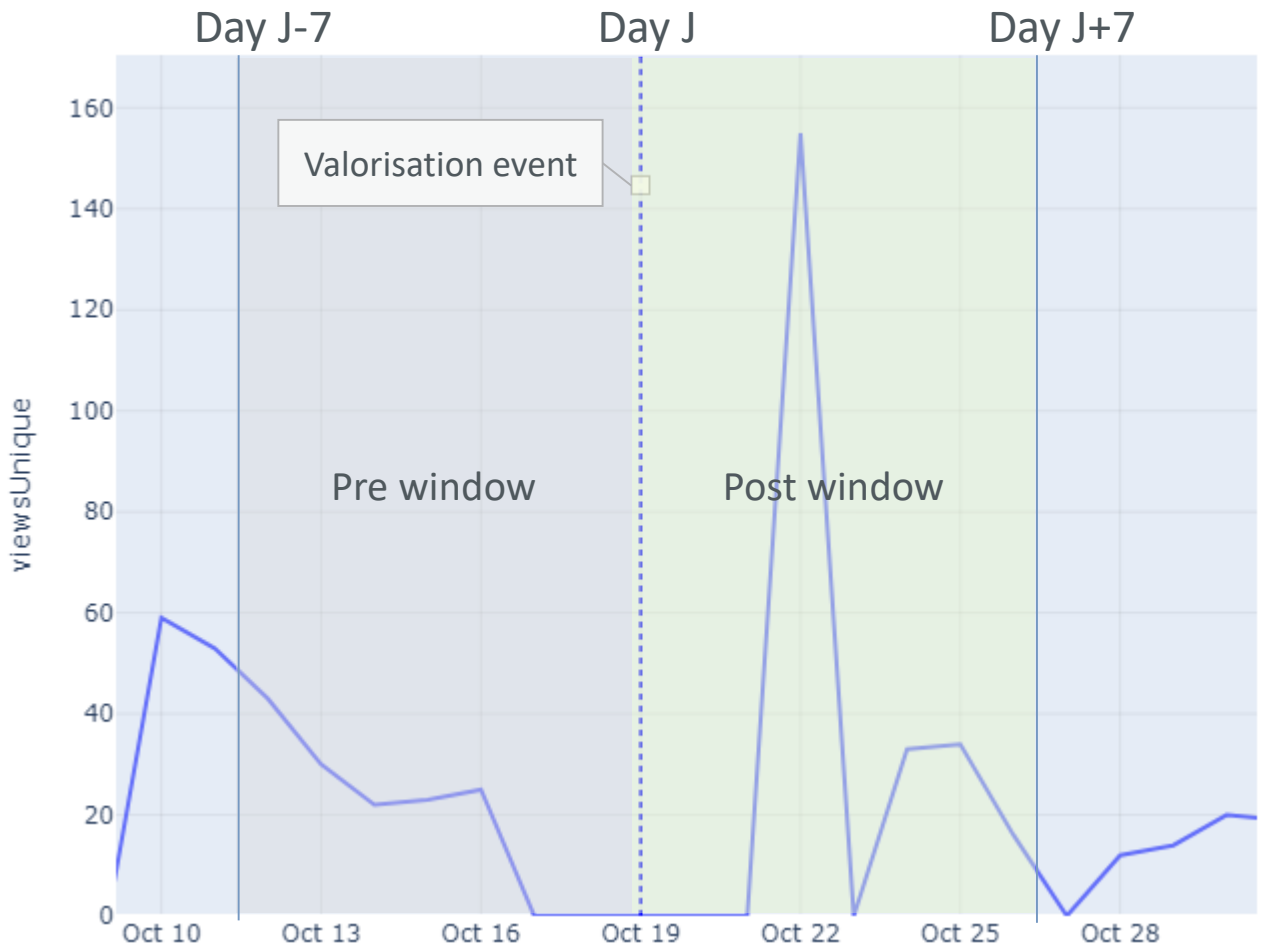
**Absolute lift score**

Difference in residuals between post and pre-windows

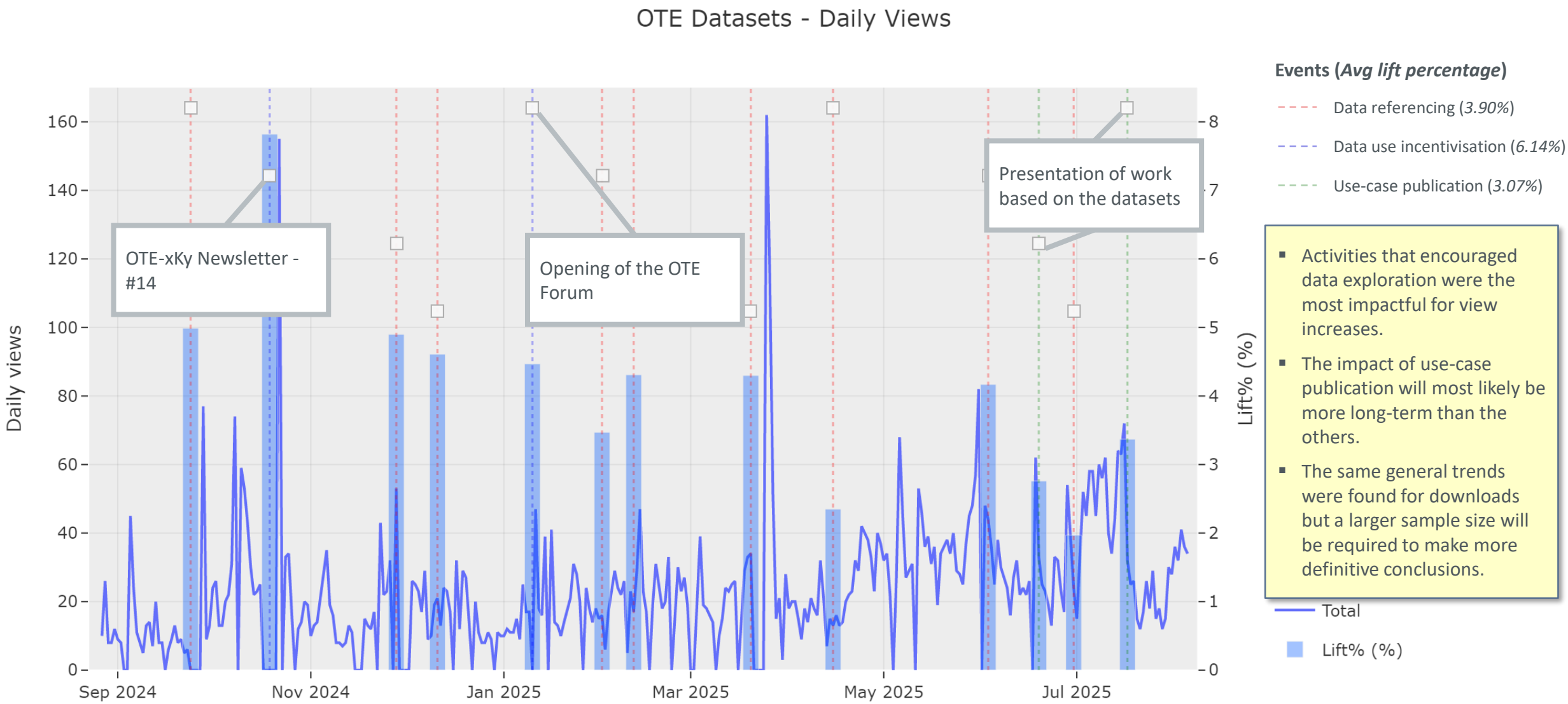
$$res_t = y - \tilde{y}_{DoW\_median}$$

$$Abs\_lift = \sum_{post} res_t - a.\sum_{pre} res_t$$

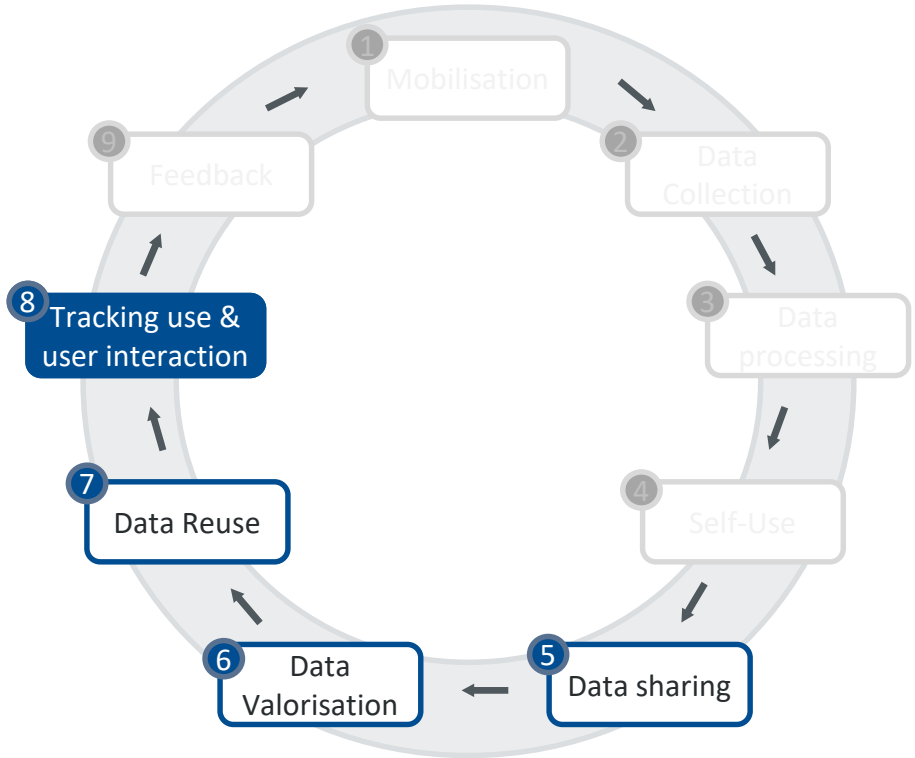
$$a = k + 1/k$$



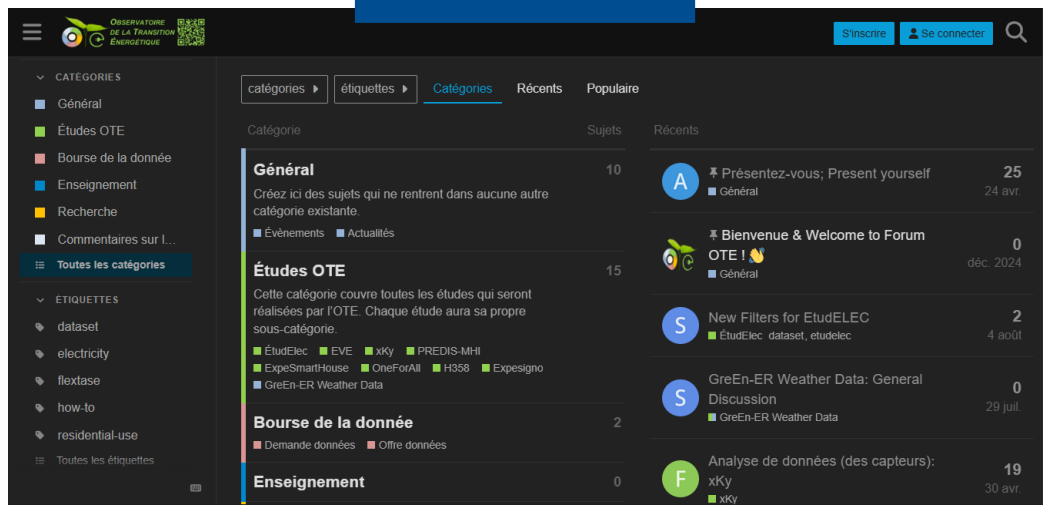
# VALENS results



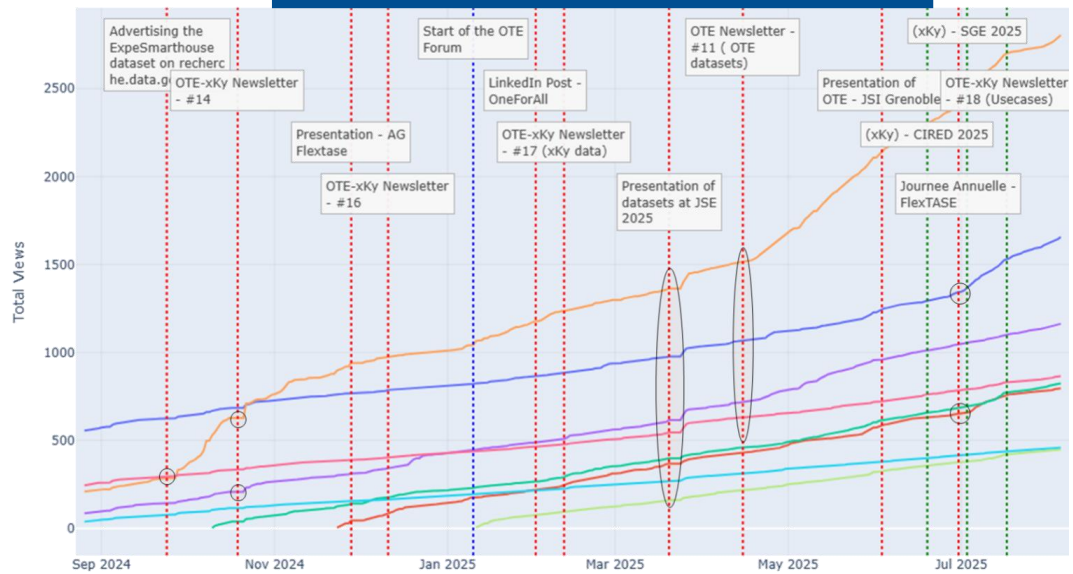
# On the tools to support interactions.




## OTE Forum




## Metrics dashboard & VALENS




- The OTE forum and the interactions around the forum
- An example of how the forum was used for data co-creation
- Measuring data metrics and valorisation events with a dashboard
- Quantifying the impact of valorisation events with VALENS

# Agenda

- Introduction
- Formalisation of the data lifecycle
- Experimentation based on six real-life case studies
  - *Experimental setup*
  - *Impact of data typologies*
  - *Support for interactions between stakeholders*
- **Conclusions, limitations and prospects**

G2Elab
Grenoble Génie Électrique
Grenoble Electrical Engineering

# Overall, we discussed...

## Our questions for today...

1. **HOW** can energy data be made **ACCESSIBLE**?
2. Will this be **VALID** for the **DIVERSITY** of energy data?
3. How will this **FACILITATE INTERACTIONS** between stakeholders?

- OpEnDaLe, a data lifecycle for the energy sector that addresses some of these barriers through additional steps and modifications
- The experimental setup used in my thesis: OTE-UGA, and the various published datasets
- Relevant data typologies and their impact on the data lifecycle, demonstrating the robustness of this cycle.
- The tools and indicators we use in the data lifecycle to facilitate interactions between stakeholders

# But the work is neither perfect nor final…

## Limitations

- Centered around only individuals and research communities
- Lack of metrics to study the effects adequately (esp. citations)
- Insufficient control study on the individual lifecycle additions

## Outlook

- Continue to use OpEnDaLe and the tools at OTE-UGA on current projects (FlexTASE, Chaire Sobriete Resilience, ANR Satiable)
- Extension of the implementation and testing to other organisations
- Battery of tests for various phases (interaction & valorisation) with improved metrics pipelines
- Deeper experimentation on more technical privacy-preserving techniques
- Exploring semantic data enrichment and leveraging AI in the data lifecycle process

# Thank you very much

# Question time!