Data for policy: data management and data access

DIGIT and Joint Research Centre

JRC.S3 – Science for Modelling and Evaluation

JRC.T4 – Data Governance and Services

The journey from data to evidence based policy

How to

… identify the data needed for policy making?

… find and acquire the data needed?

… improve decision making based on data?

… promote data sharing for further re-use?

European Commission

How to identify the data needed for policy making?

Start with the **problem at hand**. *Example, fight against VAT fraud*

Why	Frame desired outcomes	• Reduce VAT gap with a fair, transparent, and sustainable policy (*illustrative*)
What	Define objectives	• Resources allocation; Revenue collection; Compliance
Who	Identify stakeholders	• Policy makers; Inspectors; Administrators; Taxpayers; Tax consultants, …
How	Research past experience	• What has been already done? What works and does not work?
	Consider approaches	• Which policy or operational interventions to consider?
	Define success criteria	• How success looks like? Which criteria will be applied to qualify success?

What not to do:

- Focus on algorithms and ignore the policy question[1]
- Restrict to the data at hand and miss opportunities
- Avoid analytical approaches for lack of skills

[1]Everyone wants to do the model work, not the data work (Google Research)

How to identify the data needed for policy making?

Consider different **constraints** and…

Economic	Compliance
Purchase cost	Privacy
Storage cost	Legal
Scalability	Security
Maintenance	

requirements about potential data sources.

Quality	Usability	Richness
Accuracy	Interoperability	Granularity
Consistency	Timeliness	Coverage
Completeness	Accessibility	Frequency
Bias	Documentation	
	Durability	

drill down

- Compliance is about 3 components:
  - Privacy law: Are there personal data elements?
  - Legal: What is the legal basis ruling the data source?
  - Security: What is the confidentiality level attributed by the internal policy in your organisation?
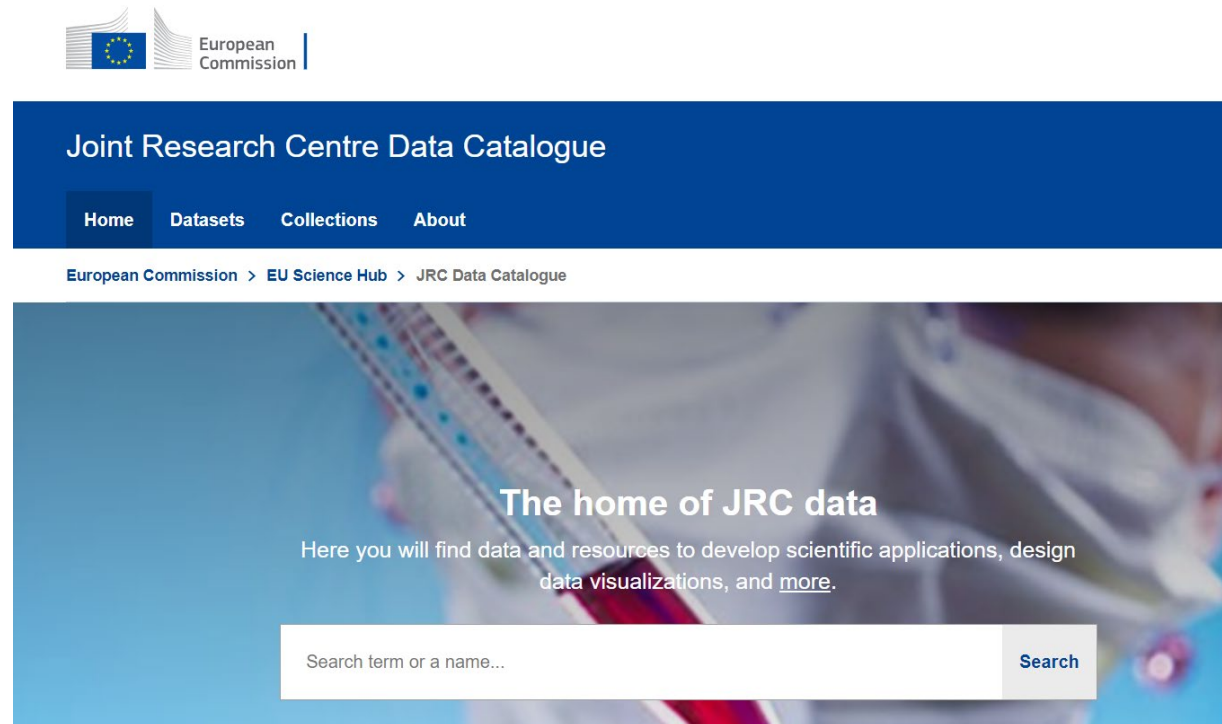
- Establish a **partnership with your Data Protection Officer**. This means that the intent is to maximise value from data remaining compliant.

- Consider **expertise on GDPR and other legal regulations** essential to scale up data analytics work.

- Key topics for maximising data value in a compliant way:
  - Consent, data minimization, anonymization[1] and pseudoanonymisation, retention, transparency, privacy records

European Commission

[1]AEPD-EDPS joint paper on 10 misunderstandings related to anonymization

How to find and acquire the data needed?

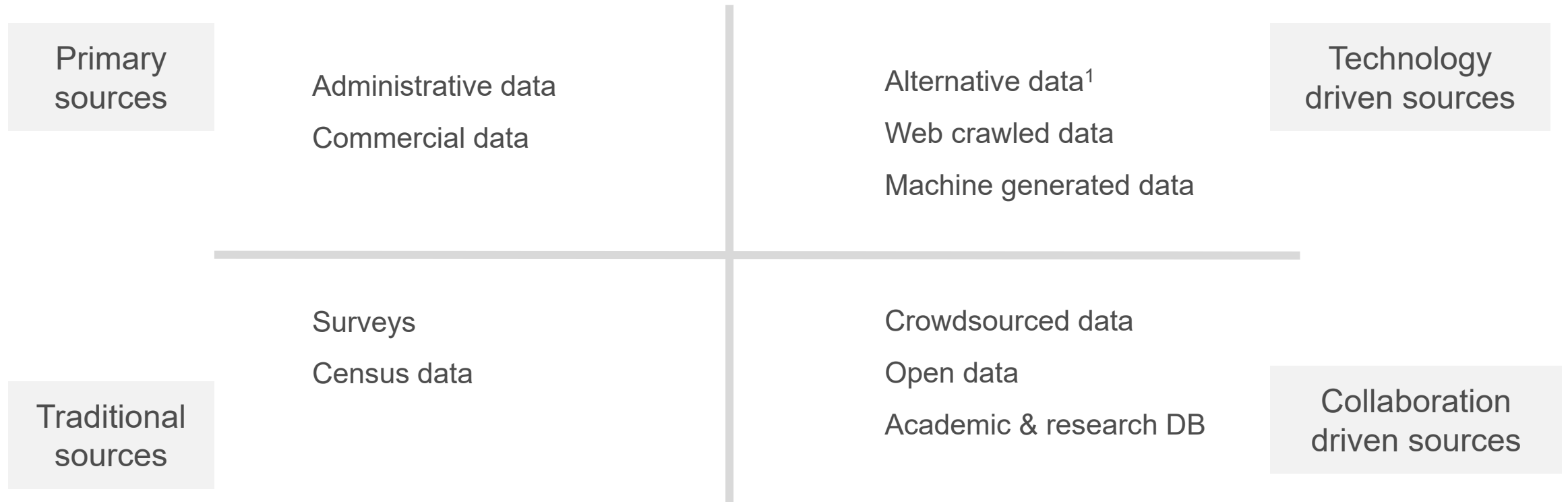**Data catalogues** are important to discover if the data is already available in-house.

How to find and acquire the data needed?

If not, consider different **options to acquire data**:

Primary sources		Technology driven sources
Administrative data	Alternative data[1]	
Commercial data	Web crawled data	
	Machine generated data	
	Crowdsourced data	
Surveys	Open data	
Census data	Academic & research DB	
Traditional sources		Collaboration driven sources

[1][Uses of Alternative Data Sources for Public Health Statistics and Policymaking: Challenges and Opportunities](#)

European Commission

How to find and acquire the data needed?

Alternatively, do not acquire the data. Provide incentives for the data holders to perform the required analysis.

A community of practice consists of a range of actors (a community) including

(A) scientists,
(B) policy officials and possibly
(C) officers of data-holding institutions

to discuss what works in terms of project design and methodology, and to compare problems and solutions across teams.

- What worked in one Country may or may not work for another one.

- However, if one team listens to the experience of another team in another Member State, this may inspire the design of new policy implementation or of the evaluation of similar policies.

- This is the peer effect in learning to learn.

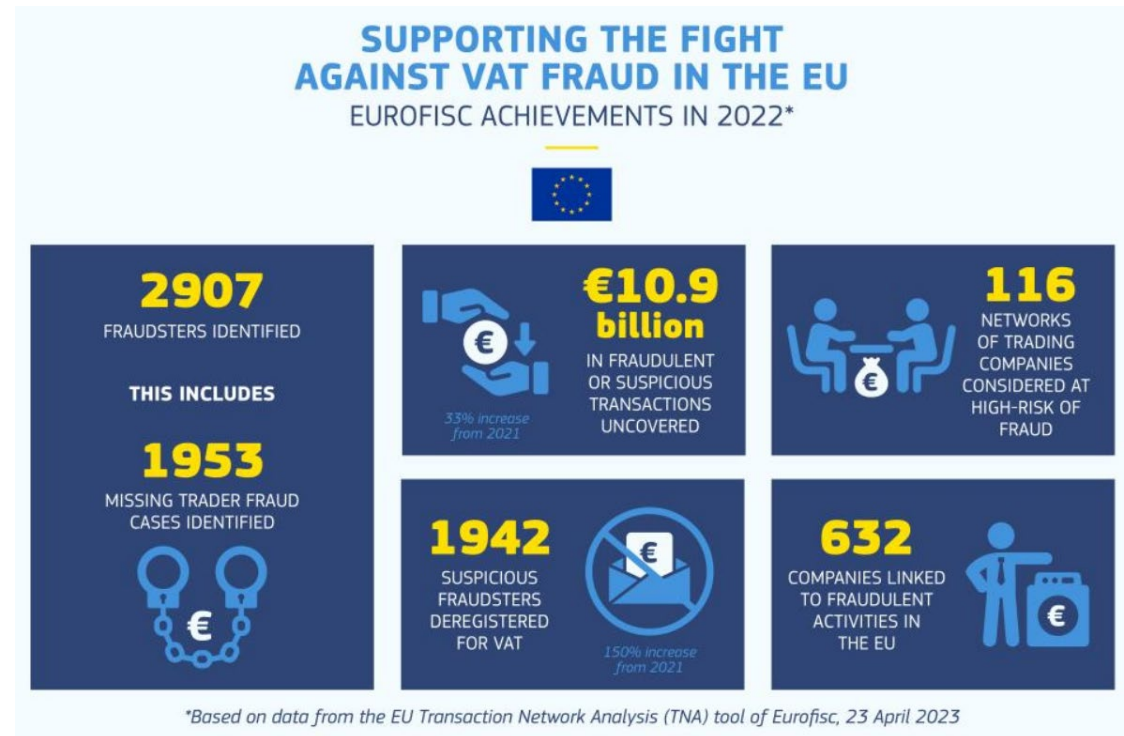European Commission

How to find and acquire the data needed?

Communities of Practice with policy-makers, data holders and users for (country-specific) best practices.

Example of **Eurofisc**, launched in 2010 to combat cross-border VAT fraud, is comprised of liaison officials from the 27 Member States and Norway.

SUPPORTING THE FIGHT AGAINST VAT FRAUD IN THE EU
EUROFISC ACHIEVEMENTS IN 2022*

2907 FRAUDSTERS IDENTIFIED
THIS INCLUDES
1953 MISSING TRADER FRAUD CASES IDENTIFIED

€10.9 billion IN FRAUDULENT OR SUSPICIOUS TRANSACTIONS UNCOVERED
33% increase from 2021

116 NETWORKS OF TRADING COMPANIES CONSIDERED AT HIGH-RISK OF FRAUD

1942 SUSPICIOUS FRAUDSTERS DEREGISTERED FOR VAT
150% increase from 2021

632 COMPANIES LINKED TO FRAUDULENT ACTIVITIES IN THE EU

*Based on data from the EU Transaction Network Analysis (TNA) tool of Eurofisc, 23 April 2023

European Commission

How to improve decision making based on data?

Practices to **trust data** and the **analytical process**.

- ## Guidelines for FAIR data
  - Findable, accessible, interoperable, reusable

- ## Publishing data alongside results
  - reproducibility (data + analytics governance)

- ## Training scientists
  - ethical principals
  - data skills
  - communicating with data

European Commission

Joint Research Center (JRC)

Modelling Inventory and Knowledge Management System of the European Commission (MIDAS)

Home   Explore ∨   **Search**   API   **Publications**   Help ∨

Home > Search

Search inside MIDAS   Search

European Commission

How to improve decision making based on data?

Good decisions require **good context** and **good data**. This is possible through collaboration between Policy and Data people.

	Policy Officer	*collaboration*	**Data Analyst**

Data Process

Policy Understanding	Defines context and policy questions/goals.	Translates the business questions/goals into analytical approaches and identifies the data elements required for the analysis.
Data Understanding	Provides business knowledge to understand the meaning of the data and validates its consistency.	Locates and collects the data required, assesses the quality of the data, and profiles the data and explores the key variables.
Data Preparation	Clarifies the business meaning of data and may request specific rules for naming variables.	Prepares the data for analysis (e.g. puts together different files, transforms variables, selects a certain scope of rows/columns).
Analytics	Validates intermediate results and provide feedback to drive the analysis adaption	Executes the analysis iteratively, involving the policy maker for feedback on a need basis.
Evaluation	Performs policy evaluation of analytics outcomes.	Performs quantitative evaluation of analytics outcomes (it means different things, depending on the type of outcomes)

European Commission

How to promote data sharing for further re-use?

There are **multiple barriers to data exchange**, both inside and between organizations. Most are non-technical.

Economic	Cultural	Organizational	Governance	Compliance	Technical
Resources not shared by default	Control and trust	Change	Roles and responsibilities missing	Ownership	Accessibility
Need for incentives in place	Secondary usages	Budget to perform work	Poor metadata	Regulations	Interoperability
	Silos and collaboration	Bureaucracy for approvals	Misuse	Legislation (incl. privacy law)	Volume
	Politics			Reputation	Formats
				Security	
				Intellectual property	

"The processing of personal data should be designed to serve mankind. The right to the protection of personal data is not an absolute right; it must be considered in relation to its function in society and be balanced against other fundamental rights, in accordance with the principle of proportionality."[1]

A timely quote.

[1]European Parliament, Council of the European Union. Regulation EU 2016/679 of the European Parliament and of the Council of 27 April 2016 (General Data Protection Regulation). Official Journal of the European Union (2016).

European Commission

How to promote data sharing for further re-use?

Three essential factors to **address barriers to data sharing**:

Data Strategy

…for **clarity of purpose** on data usage:

- Increase citizen service level in terms of speed, personalization, and accuracy.
- Reduce costs by automating tasks.
- Improve compliance by reducing costs and fraud (which leads to increase of revenue, like tax revenue)

Data Governance

…to foster data sharing, possibly steered by a data department:[1]

- Clarify **Ownership** inside the organization for fast and transparent decision making
- **Knowledge** about meaning of data
- Keep **quality** in check
- **FAIR** principles promote sharing
- Address **compliance** proactively to maximize sharing

Data Architecture

…to ease data sharing by design:

- Cloud based technologies
- Smart contracts
- Common European data spaces[2]

Common European data spaces

Data governance and data policies
at the European Commission

European Commission
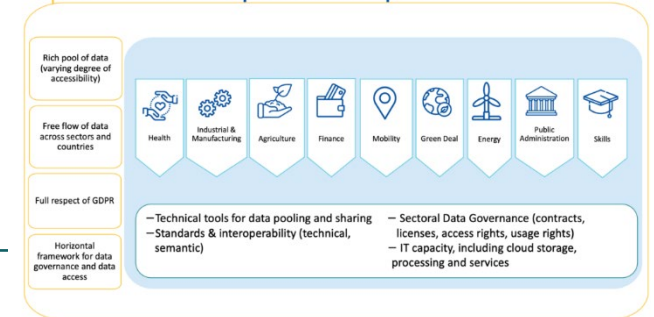
[1]summary-data-governance-data-policies_en.pdf (europa.eu)
[2]Common European data space for cultural heritage

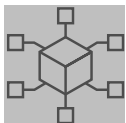# Guiding questions: What would it take to accelerate the use of **data** at your organization?

|	**Strategy**	How can we better **link** our data strategy to specific milestones, use cases and initiatives that help us get there?
	Governance and Operating Model	How can we **embed** the importance of data owners and data stewards in our operating model?
	Infrastructure and Architecture	What steps can we take to **simplify** our data architecture and provide an working environment?
	Talent and Culture	How do we best **equip** our talent with data and analytics skills and embed data-use in the day-to-day work of our teams?

European Commission

Time to collaborate!

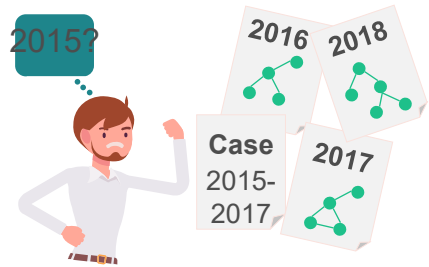# What is the downside of not getting data governance right?
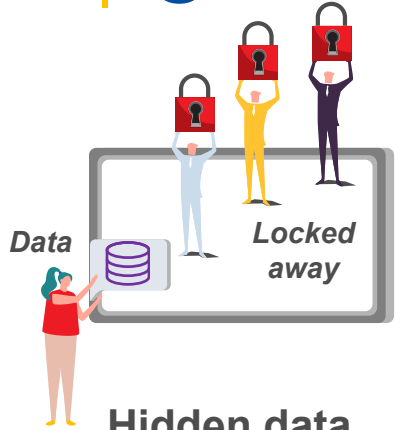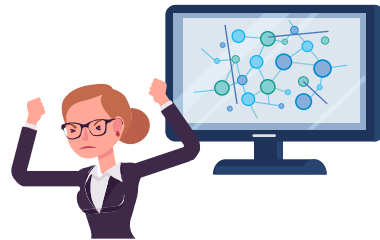
Hidden data siloes

Multiple versions of the truth
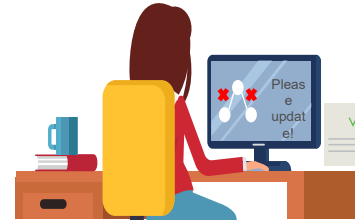
Highly manual, time-consuming processes

Limited historical views

Low confidence/ high uncertainty

Risk of regulatory violations

~30% of total enterprise time is spent on non-value-added tasks because of poor data quality and availability

>50% of an analytics team time is spend on data processing and cleanup in organizations without good data governance

Source: McKinsey

European Commission