



**DIGITAL**  
EUROPE  
PROGRAMME

The illustration features four stylized human figures in a digital space. At the top left, a woman in a yellow shirt and dark skirt stands on a purple platform, holding a tablet. At the top right, a woman in a white lab coat and black pants stands on a purple platform, holding a yellow folder. At the bottom left, a man in a white shirt and blue pants stands on a purple platform, holding a smartphone. At the bottom right, a man in a yellow shirt and blue pants sits on a white folding chair, working on a laptop. In the center, five white icons are arranged horizontally: a server rack, a heart with a pulse line, a factory with a rising bar chart, a shield with a lock, and a server rack. The background is dark purple with wavy lines and a grid of white dots in the top right corner.

**#DigitalEUProgramme**

## Introduction to this document

This document provides **practical resources** such as questionnaires, checklists, decision trees and frameworks in the context of the [Big Data Test Infrastructure](#).

These materials, produced throughout the project, aim to **support public administrations to define, design and pilot their use cases** using the **BDTI service offering**.

# Tool: Questionnaire for problem assessment



## How do we assess the problem and identify the ideal outcome?

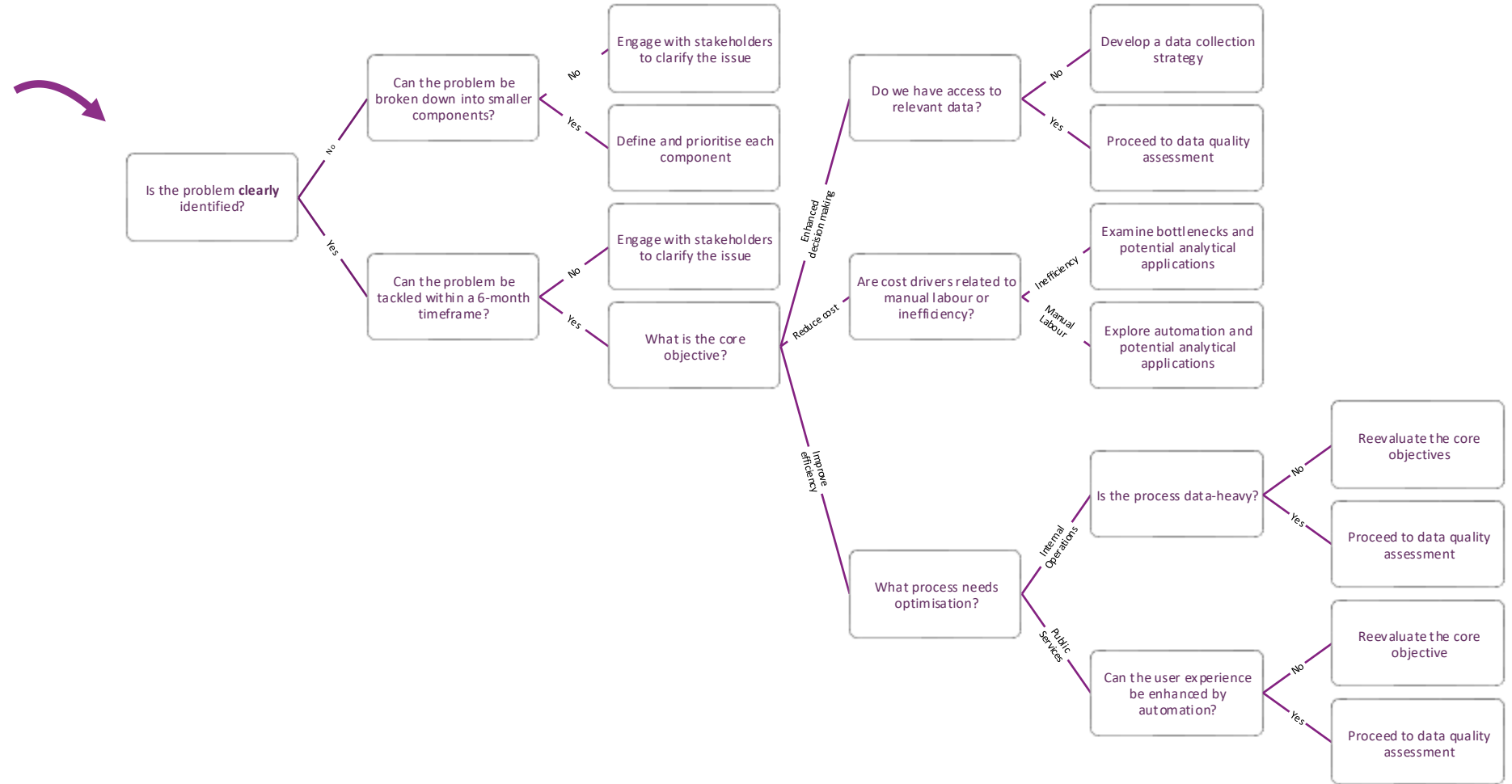
<b>Topic</b>	<b>Description</b>	<b>Questions</b>
<b>Problem Statement</b>	<p><u>Clarify the Problem Statement</u></p> <p>A problem statement identifies a specific issue or challenge that affects the efficiency, effectiveness, or equity of public services. It succinctly outlines the nature of the problem, its impact on the community or stakeholders, and the need for a solution to improve public governance and service delivery.</p>	<ul style="list-style-type: none"><li><input type="checkbox"/> What is the problem we are trying to address?</li><li><input type="checkbox"/> What exact problem are we aiming to solve with data analytics?</li><li><input type="checkbox"/> Who are the stakeholders or groups affected by this problem?</li><li><input type="checkbox"/> What is the current impact or consequences of this problem on the community or organisation?</li><li><input type="checkbox"/> What are the underlying causes or contributing factors to this problem?</li></ul>
<b>Objectives</b>	<p><u>Prioritise Objectives</u></p> <p>Objectives are clear, specific, and measurable targets set to achieve desired outcomes in public policies, programs, or initiatives. They provide direction and focus for actions, ensuring that efforts are aligned with broader goals of improving public welfare, efficiency, and service delivery. Objectives serve as benchmarks for evaluating progress and success in addressing public needs and challenges.</p>	<ul style="list-style-type: none"><li><input type="checkbox"/> Which objectives are most critical to achieving the desired outcome?</li><li><input type="checkbox"/> What are the specific outcomes or results we aim to achieve with this policy or program?</li><li><input type="checkbox"/> Who are the primary beneficiaries or target groups of these objectives?</li><li><input type="checkbox"/> What are the key actions or steps required to achieve these outcomes?</li><li><input type="checkbox"/> How will we measure and evaluate the success of these objectives?</li><li><input type="checkbox"/> What is the timeline for achieving these objectives, and what are the critical milestones along the way?</li></ul>
<b>Success Criteria</b>	<p><u>Define Success Criteria</u></p> <p>Success criteria are the specific, measurable standards or benchmarks used to evaluate the effectiveness and impact of policies, programs, or initiatives. They define the desired outcomes and performance indicators that signify the achievement of objectives, ensuring accountability and guiding decision-making processes to enhance public service delivery.</p>	<ul style="list-style-type: none"><li><input type="checkbox"/> What are the specific goals or objectives that need to be achieved?</li><li><input type="checkbox"/> How will we measure progress and determine if these goals are being met?</li><li><input type="checkbox"/> What are the key performance indicators (KPIs) or metrics to evaluate success?</li><li><input type="checkbox"/> What are the expected outcomes or benefits for the stakeholders or community?</li><li><input type="checkbox"/> What is the timeline for achieving these goals, and what milestones should be reached along the way?</li></ul>

# Tool: Use case readiness - Determine your next step



\*Clearly\* signifies that you should be capable of explaining this issue to a colleague in just one sentence.

❑ Use this tool to assess your use case readiness and determine your next step



# Tool: Questionnaire on Strategic Fit



## How do we verify if the use case fits into a broader strategy?

<b>Topic</b>	<b>Description</b>	<b>Questions</b>
<b>Policy Adherence</b>	<p>Policy adherence refers to the alignment between public administration’s resources, capabilities, and policy objectives, laid out in relevant documents, to effectively address public needs and challenges. It ensures that public sector strategies are coherent with the internal strengths of government agencies and the external socio-political environment. Achieving strategic fit in public administration enhances the efficiency and effectiveness of public service delivery and policy implementation.</p>	<ul style="list-style-type: none"> <li><input type="checkbox"/> <i>Is the project's result aligned with our entity's/EU's strategy?</i></li> <li><input type="checkbox"/> <i>What official documents can we use to support our reasoning?</i></li> <li><input type="checkbox"/> <i>What pillars of what EU strategy are we supporting with this project?</i></li> <li><input type="checkbox"/> <i>Are we promoting data accessibility and reusability?</i></li> <li><input type="checkbox"/> <i>Are we facilitating data sharing/interoperability?</i></li> <li><input type="checkbox"/> <i>Are we adhering to the ethical practices/regulations?</i></li> </ul>
<b>Value Identification</b>	<p>Value Identification involves pinpointing the specific benefits that the product delivers to its users and stakeholders. This includes identifying how the data product enhances decision-making, improves operational efficiency, and supports transparency and accountability within public sector organisations. By clearly articulating these benefits, public administrators can justify investments in data products and ensure they align with the broader goals of serving the public interest.</p>	<ul style="list-style-type: none"> <li><input type="checkbox"/> <i>Has the future value of the project's product been assessed?</i></li> <li><input type="checkbox"/> <i>Do we know what benefits it would bring? To whom? Citizens? Administration?</i></li> <li><input type="checkbox"/> <i>What local/EU policies would the end product and its benefits support?</i></li> </ul>

Use this table to understand what strategic fit may mean for your use case

Each topic includes a non-exhaustive sample of questions to help you prepare your own list

# Tool: Questionnaire on Collaboration (1/3)



## How do we want to efficiently and effectively collaborate?

<b>Topic</b>	<b>Description</b>	<b>Question</b>
<b>External Collaboration</b>	External collaboration in the context of public administration involves partnerships and cooperative efforts between government agencies and external entities such as non-governmental organisations (NGOs), private sector companies, academia and community groups. This collaboration aims to leverage the unique strengths and resources of each partner to address public issues more effectively and efficiently. By working together, these entities can enhance service delivery, foster innovation, and improve outcomes for the community.	<ul style="list-style-type: none"><li><input type="checkbox"/> Who can the key external stakeholders involved in this collaboration be? Academia? NGOs? Startups?</li><li><input type="checkbox"/> What are we getting out of this cooperation? What are they getting? Would they be inclined to collaborate at all?</li><li><input type="checkbox"/> Can we join a network of cities or any other relevant network to collaborate/learn from similar cases?</li><li><input type="checkbox"/> How are these stakeholders identified and engaged in the project?</li><li><input type="checkbox"/> What communication channels are established for effective collaboration?</li><li><input type="checkbox"/> What potential risks have been identified in this collaboration? How are these risks being mitigated and managed? Are there contingency plans in place for unforeseen challenges?</li><li><input type="checkbox"/> How are ethical considerations and data privacy managed in this collaboration?</li><li><input type="checkbox"/> What agreements or contracts are in place to formalise the partnership? What agreements are missing?</li></ul>
<b>Inter-departmental Collaboration</b>	Interdepartmental collaboration refers to the cooperative efforts and coordination between different departments or divisions within an organisation. This type of collaboration aims to break down silos, enhance communication, and leverage diverse expertise to achieve common goals and improve overall organisational performance. By working together, departments can share resources, align strategies, and address complex challenges more effectively.	<ul style="list-style-type: none"><li><input type="checkbox"/> Which departments should be involved in this collaboration?</li><li><input type="checkbox"/> Who are the key stakeholders from each department?</li><li><input type="checkbox"/> How are these stakeholders identified and engaged in the project?</li><li><input type="checkbox"/> What are the primary objectives of this interdepartmental collaboration?</li><li><input type="checkbox"/> How do these objectives align with the overall goals of the organisation? How do they align with each department's objectives? What's in it for them? Will they be willing to help?</li><li><input type="checkbox"/> What resources (financial, human, technological) will be contributed by each department? How are these resources managed and allocated throughout the project?</li><li><input type="checkbox"/> What communication channels are established for effective collaboration?</li><li><input type="checkbox"/> What potential risks have been identified in this collaboration?</li><li><input type="checkbox"/> Are there any legal or regulatory requirements that need to be addressed (e.g. data sharing)? How are ethical considerations and data privacy managed in this collaboration?</li></ul>

# Tool: Questionnaire on Collaboration (2/3)



## How do we want to efficiently and effectively collaborate?

<i>Topic</i>	<i>Description</i>	<i>Question</i>
<b>Administration – IT Collaboration</b>	Public administration and IT collaboration in data projects involves a collaborative decision-making process where both administration and technology engage with diverse stakeholders to make informed data and technology decisions. This partnership ensures alignment on key metrics and goals, facilitating optimal decision-making and resource allocation. By integrating civil service objectives with IT capabilities, public administration can enhance service delivery, operational efficiency, and innovation.	<ul style="list-style-type: none"><li><input type="checkbox"/> What processes are in place to ensure that both administration and IT are engaged with diverse stakeholders?</li><li><input type="checkbox"/> How are civil service objectives integrated with IT capabilities?</li><li><input type="checkbox"/> What challenges have been encountered in the collaboration, and how have they been addressed?</li><li><input type="checkbox"/> Do we have the buy-in of the IT decision makers?</li></ul>
<b>Executive Commitment</b>	Leaders understand importance of data's impact on the future of the public administration and are active in investing in and holding the organisation accountable for change. They provide strategic direction, resources, and governance to ensure the project's alignment with broader goals. They actively engage stakeholders, monitor performance, and advocate for the product's value to ensure sustained support and adoption. Additionally, they manage risks and enforce policies to maintain data privacy, security, and compliance. The leadership's support at various levels is essential for a project to succeed	<ul style="list-style-type: none"><li><input type="checkbox"/> Who are my key stakeholders? Data officers, department heads, city council, mayor/president, etc.?</li><li><input type="checkbox"/> Do the key stakeholders support this initiative?</li><li><input type="checkbox"/> Are they actively participating in promoting it?</li><li><input type="checkbox"/> If not, should they? How do we onboard them?</li></ul>

# Tool: Questionnaire on Collaboration (3/3)



## How do we want to efficiently and effectively collaborate?

Topic	Description	Question
<b>Delivery Commitment</b>	A delivery commitment refers to the formal pledge by a government entity to allocate necessary resources, including personnel, budget, and technology, to develop and implement a data solution. This commitment ensures that the project has the required support and oversight to meet its objectives and deliverables within the specified timeframe. It also involves accountability measures to track progress and ensure that the data solution effectively addresses the public needs it was designed to solve.	<ul style="list-style-type: none"><li><input type="checkbox"/> Do we have the commitment of particular departments (e.g. IT, data, analytics, etc.) to provide help?</li><li><input type="checkbox"/> Did these departments agree to provide resources to complete the project?</li><li><input type="checkbox"/> Even if they did, did we get the buy-in of the stakeholders who could influence the withdrawal of the agreement?</li></ul>

<b>Approval Processes</b>	The approval process for conceiving and working on a data product in an entity. Takes into consideration the time needed for getting all the required approvals for starting a project but also making important decisions during its delivery. Includes the need for superiors' approvals, issuing public tenders and review by other public/government institutions whenever needed.	<ul style="list-style-type: none"><li><input type="checkbox"/> What are the steps involved in the approval process for starting a data product project?</li><li><input type="checkbox"/> How much time is typically required to obtain all necessary approvals?</li><li><input type="checkbox"/> What are the key milestones in the approval process?</li><li><input type="checkbox"/> Are there specific criteria that must be met for approval at each stage?</li><li><input type="checkbox"/> How are public tenders issued and managed as part of the approval process?</li><li><input type="checkbox"/> What role do other public or governmental institutions play in the approval process?</li><li><input type="checkbox"/> How are important decisions made and approved during the project's delivery? What documentation is required at each stage of the approval process?</li></ul>
---------------------------	--	---

Check how to identify and manage stakeholders on the next page

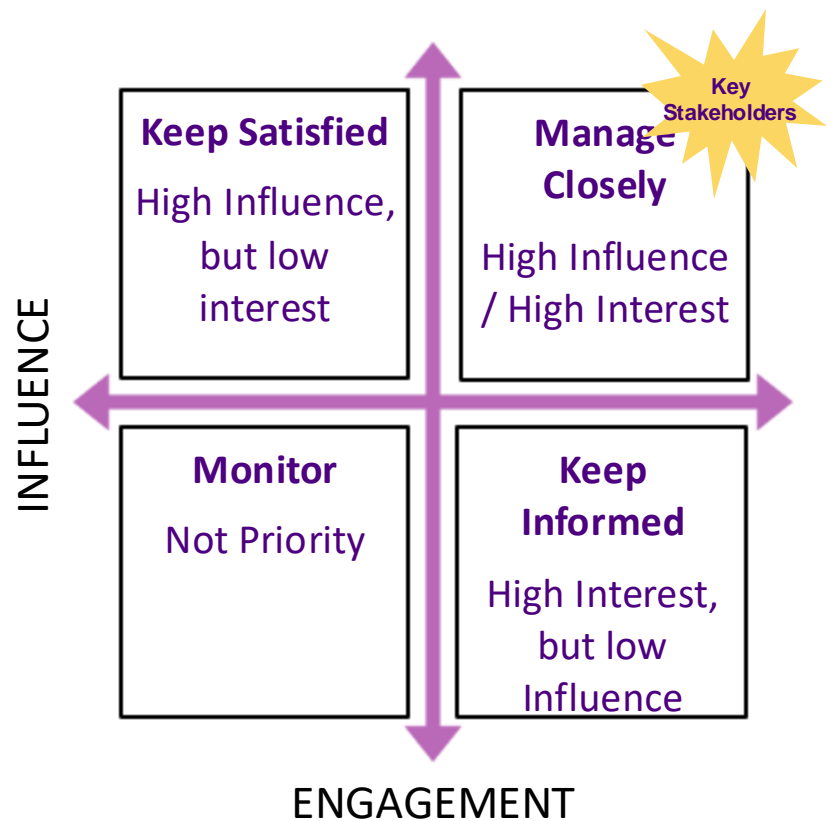




# Tool: Stakeholder mapping

How can we map our stakeholders to target our collaboration and communication efforts?

The Influence/Engagement grid is a tool that maps stakeholders, to help you identify the most important people to engage with, and to define how to collaborate with each.



## Stakeholder Mapping Tips and Tricks:

To effectively collaborate with your stakeholders, you can use the **Engagement vs. Influence Stakeholder Mapping Grid**. This tool helps you to categorise stakeholders based on their level of engagement and influence, allowing you to plan your collaboration based on the four quadrants.

**X-Axis (Engagement):** Measures the level of interest or engagement a stakeholder has in the project.

**Y-Axis (Influence):** Measures the level of influence or power a stakeholder has over the project.

Here are some examples of how you could define your interaction with the stakeholders per quadrant:

### *Keep Satisfied*

- Engage and consult
- Try to increase engagement and move into the box to the right

### *Monitor*

- Inform with basic communication: newsletters, emails, blogs etc.
- Enhance engagement by moving into the right-hand box

### *Manage Closely*

- Include in Governance and decision-making
- Engage directly and consult regularly

### *Keep informed*

- Use interest to get involved and consult in areas of expertise
- Use as supporters and ambassadors



# Tool: Technical skills assessment – Data Management

## Do we need those skills? Do we have them in the team?

<b>Role</b>	<b>Description</b>	<b>Skills Checklist</b>
<b>Data Engineers</b>	Data Engineers design, build, and maintain the infrastructure and systems that allow for the collection, storage, and analysis of large volumes of data. They ensure data are accessible, reliable, and efficiently processed for use by data scientists and analysts.	<ul style="list-style-type: none"><li><input type="checkbox"/> <i>SQL and NoSQL Databases: Proficiency in database management and querying.</i></li><li><input type="checkbox"/> <i>ETL (Extract, Transform, Load) Processes: Experience with data integration and transformation.</i></li><li><input type="checkbox"/> <i>Big Data Technologies: Knowledge of Hadoop, Spark, and other big data frameworks.</i></li><li><input type="checkbox"/> <i>Programming Languages: Proficiency in Python, Java, or Scala.</i></li><li><input type="checkbox"/> <i>Data Warehousing Solutions: Experience with tools like Amazon Redshift, Google BigQuery, or Snowflake.</i></li></ul>
<b>Data Architects</b>	Data Architects design and manage the data infrastructure of an organisation. They create blueprints for data management systems, ensuring data are organised, stored, and accessed efficiently and securely.	<ul style="list-style-type: none"><li><input type="checkbox"/> <i>Data Modelling: Proficiency in designing data models and schemas</i></li><li><input type="checkbox"/> <i>Database Management: Experience with SQL and NoSQL databases.</i></li><li><input type="checkbox"/> <i>Big Data Technologies: Knowledge of Hadoop, Spark, and other big data frameworks.</i></li><li><input type="checkbox"/> <i>Cloud Platforms: Experience with AWS, Azure, or Google Cloud.</i></li><li><input type="checkbox"/> <i>ETL Processes: Skills in data integration and transformation.</i></li></ul>

Use this table to understand what **DATA MANAGEMENT** may mean for your use case

# Tool: Technical skills assessment – Frontend/Visualisation



## Do we need those skills? Do we have them in the team?

<b>Role</b>	<b>Description</b>	<b>Skills Checklist</b>
<b>Data Visualisation Specialist</b>	Data Visualisation Specialists create visual representations of data to help stakeholders understand complex information. They use tools and techniques to design charts, graphs, and dashboards that make data insights accessible and actionable.	<ul style="list-style-type: none"><li><input type="checkbox"/> <i>Data Visualisation Tools: Proficiency in Tableau, Power BI, or D3.js.</i></li><li><input type="checkbox"/> <i>Design Principles: Understanding of visual design and user experience.</i></li><li><input type="checkbox"/> <i>Data Analysis: Ability to analyse and interpret data.</i></li><li><input type="checkbox"/> <i>Programming Languages: Skills in Python or R for data manipulation.</i></li><li><input type="checkbox"/> <i>Communication: Strong ability to present data insights effectively</i></li></ul>
<b>Frontend Developers</b>	Frontend Developers build the user interface and user experience of web applications. They use technologies like HTML, CSS, and JavaScript to create responsive and interactive websites that provide a seamless experience for users.	<ul style="list-style-type: none"><li><input type="checkbox"/> <i>HTML/CSS: Proficiency in web markup and styling.</i></li><li><input type="checkbox"/> <i>JavaScript: Strong skills in JavaScript and frameworks like React, Angular, or Vue.js.</i></li><li><input type="checkbox"/> <i>Responsive Design: Experience with creating responsive and mobile-friendly designs.</i></li><li><input type="checkbox"/> <i>Version Control: Proficiency in Git and GitHub.</i></li><li><input type="checkbox"/> <i>Cross-Browser Compatibility: Skills in ensuring compatibility across different browsers.</i></li></ul>

Use this table to understand what **FRONTEND AND VISUALISATION** may mean for your use case

# Tool: Technical skills assessment – Math and Statistics



## Do we need those skills? Do we have them in the team?

<b>Role</b>	<b>Description</b>	<b>Skills Checklist</b>
<b>Data Scientists</b>	Data Scientists analyse and interpret complex data to help organisations make informed decisions. They use statistical methods, machine learning, and data visualisation techniques to uncover patterns, trends, and insights from data.	<ul style="list-style-type: none"><li><input type="checkbox"/> <i>Statistical Analysis: Strong foundation in statistics and probability.</i></li><li><input type="checkbox"/> <i>Machine Learning: Proficiency in algorithms, model building, and evaluation.</i></li><li><input type="checkbox"/> <i>Programming Languages: Proficiency in Python or R.</i></li><li><input type="checkbox"/> <i>Data Visualisation: Skills in tools like Matplotlib, Seaborn, or Tableau.</i></li><li><input type="checkbox"/> <i>Data Wrangling: Ability to clean and preprocess data effectively.</i></li></ul>

Use this table to understand what **MATH AND STATISTICS** may mean for your use case

# Tool: Technical skills assessment – Programming



## Do we need those skills? Do we have them in the team?

<b>Role</b>	<b>Description</b>	<b>Skills Checklist</b>
<b>Backend Developers</b>	Backend Developers focus on the server-side logic, databases, and APIs that power web applications. They ensure that the backend systems are robust, scalable, and secure, enabling the frontend to function smoothly.	<ul style="list-style-type: none"><li><input type="checkbox"/> <i>Server-Side Languages: Proficiency in languages like Java, Python, Ruby, or Node.js.</i></li><li><input type="checkbox"/> <i>Database Management: Experience with SQL and NoSQL databases.</i></li><li><input type="checkbox"/> <i>API Development: Skills in creating and managing RESTful or GraphQL APIs.</i></li><li><input type="checkbox"/> <i>Security Best Practices: Knowledge of securing backend services.</i></li><li><input type="checkbox"/> <i>Version Control: Proficiency in Git and GitHub.</i></li></ul>

Use this table to understand what **PROGRAMMING** may mean for your use case

# Tool: Technical skills assessment – Operations



## Do we need those skills? Do we have them in the team?

<b>Role</b>	<b>Description</b>	<b>Skills Checklist</b>
<b>DevOps</b>	DevOps professionals bridge the gap between development and operations teams by automating and streamlining the software development and deployment process. They focus on improving collaboration, continuous integration, and delivery to enhance the efficiency and reliability of software systems.	<ul style="list-style-type: none"><li><input type="checkbox"/> <i>CI/CD Pipelines: Experience with continuous integration and continuous deployment tools.</i></li><li><input type="checkbox"/> <i>Infrastructure as Code (IaC): Proficiency in tools like Terraform or Ansible.</i></li><li><input type="checkbox"/> <i>Containerisation: Knowledge of Docker and Kubernetes.</i></li><li><input type="checkbox"/> <i>Monitoring and Logging: Experience with tools like Prometheus, Grafana, or ELK stack.</i></li><li><input type="checkbox"/> <i>Scripting Languages: Proficiency in Bash, Python, or PowerShell.</i></li></ul>
<b>MLOps</b>	MLOps (Machine Learning Operations) specialists integrate machine learning models into production environments, ensuring they are scalable, reliable, and maintainable. They work on automating the deployment, monitoring, and management of machine learning models.	<ul style="list-style-type: none"><li><input type="checkbox"/> <i>Model Deployment: Experience with deploying machine learning models in production.</i></li><li><input type="checkbox"/> <i>CI/CD for ML: Knowledge of continuous integration and deployment for machine learning.</i></li><li><input type="checkbox"/> <i>Monitoring and Maintenance: Skills in monitoring model performance and retraining.</i></li><li><input type="checkbox"/> <i>Data Versioning: Proficiency in tools like DVC or MLflow.</i></li><li><input type="checkbox"/> <i>Cloud Platforms: Experience with AWS, Azure, or Google Cloud for ML operations.</i></li></ul>
<b>Cloud Architects</b>	Cloud Architects design and oversee the cloud computing strategy of an organisation. They are responsible for planning, implementing, and managing cloud infrastructure and services to ensure scalability, security, and cost-efficiency.	<ul style="list-style-type: none"><li><input type="checkbox"/> <i>Cloud Platforms: Proficiency in AWS, Azure, or Google Cloud.</i></li><li><input type="checkbox"/> <i>Infrastructure as Code (IaC): Experience with Terraform, CloudFormation, or ARM templates.</i></li><li><input type="checkbox"/> <i>Security Best Practices: Knowledge of cloud security and compliance.</i></li><li><input type="checkbox"/> <i>Networking: Skills in cloud networking and connectivity.</i></li><li><input type="checkbox"/> <i>Cost Management: Ability to manage and optimise cloud costs.</i></li></ul>

Use this table to understand what OPERATIONS may mean for your use case

# Tool: Supporting skills assessment (1/2)



## Do we need those skills? Do we have them in the team?

**Role**      **Description**

**Project Manager**      A project manager in the public administration context is responsible for planning, executing, and overseeing projects that aim to improve public services and infrastructure. They ensure that projects are completed on time, within budget, and meet the required standards and regulations. This role involves coordinating with various stakeholders, managing resources, and mitigating risks to achieve the desired outcomes.

- Skills Checklist**
- Strong organisational and planning abilities*
  - Excellent communication and interpersonal skills*
  - Proficiency in project management software and tools*
  - Ability to manage budgets and financial resources*
  - Problem-solving and critical thinking skills*

**Delivery Lead**      A delivery lead in the public administration context is responsible for ensuring the successful delivery of projects and services that meet the needs of the community. They oversee the implementation process, manage teams, and ensure that deliverables are completed on time and within scope. This role involves close collaboration with stakeholders, continuous monitoring of progress, and addressing any issues that may arise to ensure smooth project execution.

- Strong leadership and team management skills*
- Excellent communication and stakeholder engagement abilities*
- Proficiency in project and service delivery methodologies*
- Ability to manage budgets and ensure timely delivery*
- Problem-solving and adaptability skills*

**Business Analyst**      A business analyst in the public administration context is responsible for analysing and improving processes, systems, and policies to enhance public services and operations. They gather and interpret data, identify areas for improvement, and provide recommendations to optimise efficiency and effectiveness. This role involves collaborating with various departments, understanding stakeholder needs, and ensuring that proposed solutions align with organisational goals.

- Strong analytical and critical thinking abilities*
- Excellent communication and interpersonal skills*
- Proficiency in data analysis and visualisation tools*
- Ability to understand and document business processes*
- Problem-solving and decision-making skills*

Use this table to understand what **SUPPORTING SKILLS** your use case may need

# Tool: Supporting skills assessment (2/2)



## Do we need those skills? Do we have them in the team?

<b>Role</b>	<b>Description</b>
<b>Subject Matter Expert</b>	A subject matter expert (SME) is an individual with extensive knowledge and expertise in specific areas of government operations, policies, and procedures. They provide valuable insights and guidance on best practices, regulatory compliance, and effective governance strategies. SMEs play a crucial role in advising public officials, developing policy frameworks, and ensuring the efficient and transparent functioning of public institutions.

### Skills Checklist

- Policy Analysis and Development*
- Domain-Specific Knowledge*
- Analytical and Critical Thinking*
- Communication Skills*
- Ethical Judgment and Integrity*

Use this table to understand what **SUPPORTING SKILLS** your use case may need

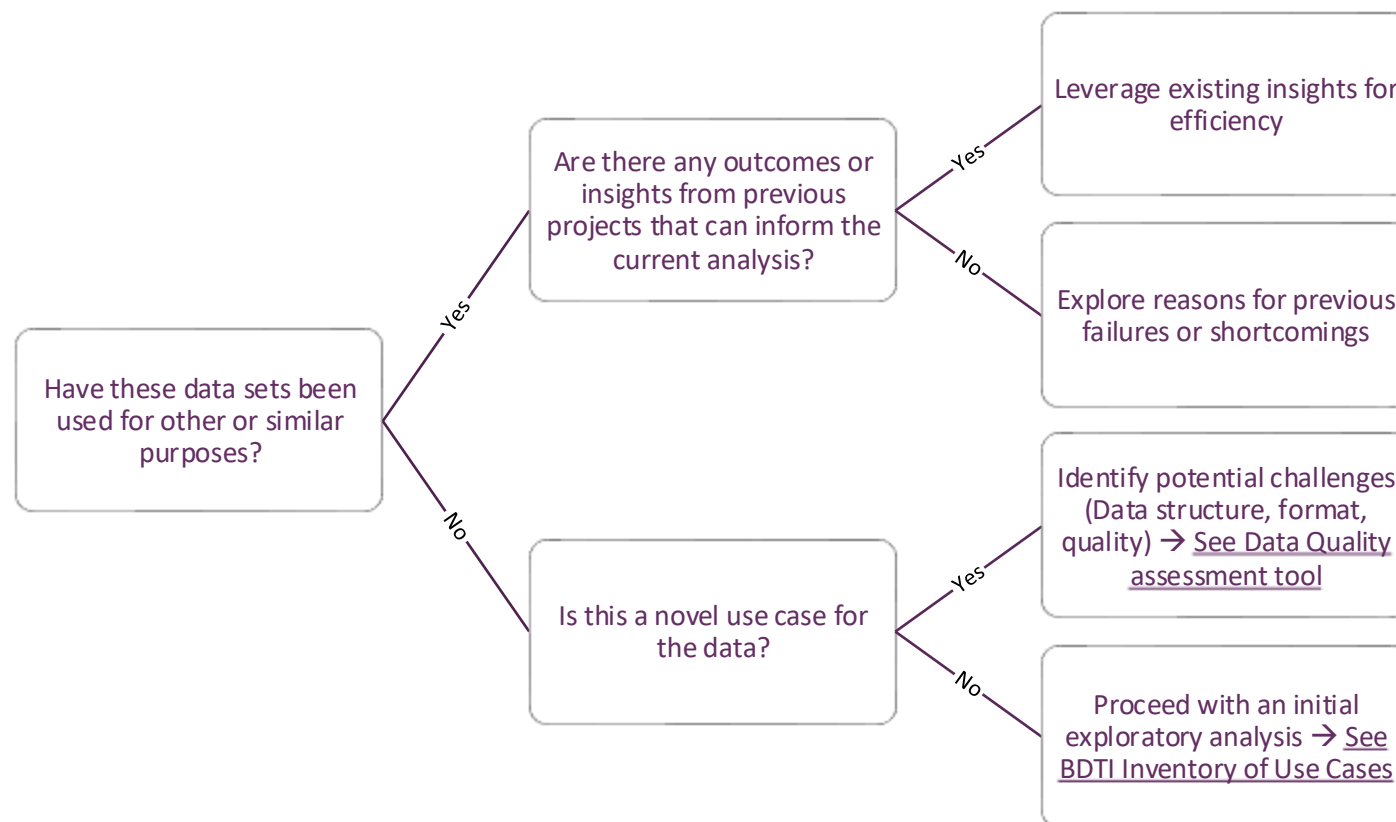




# Pre-Assessment: Have these data sets been used before?

Understanding the previous usage and outcomes of data sets can significantly enhance the efficiency and effectiveness of your data analysis.

- ❑ Use this tool to ensure your work is based on a more informed and strategic approach
- ❑ Explore the [European Data Portal](#), where almost 2M datasets are available for download through inter alia [SPARQL](#) or [REST API](#)



**IMPORTANT**

# Tool: Operational data assessment



## What operational data aspects should be considered?

<b>Role</b>	<b>Description</b>	<b>Question</b>
<b>Continuity</b>	Continuity in data refers to the seamless and uninterrupted flow of data over time, ensuring that data are consistently available and up-to-date for analysis and decision-making. Refers also to the fact of data being available over time (i.e. the data source being continuously maintained).	<ul style="list-style-type: none"><li><input type="checkbox"/> What is the certainty data will be available going forward?</li><li><input type="checkbox"/> How reliable is the data source in terms of SLAs (up-time, history, format backward compatibility, etc.)?</li><li><input type="checkbox"/> Are there any known issues with data latency or timeliness?</li><li><input type="checkbox"/> Is financial commitment required now/is it foreseen in the future?</li><li><input type="checkbox"/> What risks are associated with data discontinuity, and how are they mitigated?</li></ul>
<b>Sensitivity</b>	Data sensitivity refers to the classification of data based on its level of importance and the potential impact of its exposure. Sensitive data typically includes information that, if disclosed, could cause harm to individuals, organisations, or systems. This can include personal information, financial records, intellectual property, and other confidential or proprietary information.	<ul style="list-style-type: none"><li><input type="checkbox"/> Are there any compliance requirements (e.g., GDPR) that need to be considered?</li><li><input type="checkbox"/> Can the data be shared/displayed in the application?</li><li><input type="checkbox"/> Do we need to introduce masking/aggregation mechanisms?</li><li><input type="checkbox"/> How will we make sure we don't disclose inappropriate data, i.e. what technology and processes will we use?</li><li><input type="checkbox"/> What are the data retention policies for your organisation?</li><li><input type="checkbox"/> How long will the data need to be stored, and what are the implications for storage capacity?</li></ul>
<b>Quality</b>	Data quality refers to the condition of a set of values of qualitative or quantitative variables that are accurate, complete, reliable, and relevant to the intended use. High data quality ensures that the data are fit for its intended purposes in operations, decision-making, and planning. Poor data quality can lead to incorrect conclusions, inefficiencies, and potential risks in business processes.	<ul style="list-style-type: none"><li><input type="checkbox"/> What is the data quality at the source? How common are errors? Are the errors significant?</li><li><input type="checkbox"/> What validation processes are needed to ensure data conform to the expected format and structure?</li><li><input type="checkbox"/> How much effort would it be to remove errors from incoming data?</li></ul>

Use this table to assess the operational aspect of data for your use case

For a detailed data sensitivity and quality assessment, check out the next pages

# Tool: Technical data assessment



## What technical data aspects should be considered?

<b>Role</b>	<b>Description</b>	<b>Questions</b>
<b>Delivery</b>	Delivery pertains to the methods and processes used to transport data from one location to another, ensuring that it reaches its intended destination accurately and efficiently. The most common ways are automated batches (a bulk upload of the data), automated streaming (uploading data one by one, as they become available) or manual upload (a batch upload of a selected chunk of data).	<ul style="list-style-type: none"><li><input type="checkbox"/> What is the delivery method of the data? Batch? Streaming? Manual upload?</li><li><input type="checkbox"/> How frequently will the data need to be delivered (e.g., real-time, hourly, daily)?</li><li><input type="checkbox"/> How easy is it to integrate the data delivery method with the BDTI platform?</li><li><input type="checkbox"/> How many distinct sources will there be? What would be the workload to integrate all data sources?</li></ul>
<b>Volume</b>	Volume describes the amount of data being generated, stored, and processed, often measured in terms of bytes (and its multipliers), and is a key aspect of big data.	<ul style="list-style-type: none"><li><input type="checkbox"/> What is the expected volume of data to be transferred?</li><li><input type="checkbox"/> How will the volume of data impact processing times and performance?</li><li><input type="checkbox"/> What are the cost implications of storing and processing large volumes of data?</li><li><input type="checkbox"/> How will you manage and optimise costs related to data volume?</li></ul>
<b>Format</b>	Format refers to the structure and organisation of data, which can vary widely (e.g., CSV, JSON, XML, Avro, etc.) and affects how data are stored, processed, and interpreted.	<ul style="list-style-type: none"><li><input type="checkbox"/> What is the format of the data? Are there various formats?</li><li><input type="checkbox"/> Does BDTI natively support storing this format?</li><li><input type="checkbox"/> What (and how costly) transformations would need to be done on these data to store it?</li><li><input type="checkbox"/> Does BDTI natively support processing this format?</li><li><input type="checkbox"/> What (and how costly) transformations would need to be done on these data to use/display it?</li></ul>

Use this table to assess the technical aspect of data for your use case

For a detailed data sensitivity and quality assessment, check out the next pages

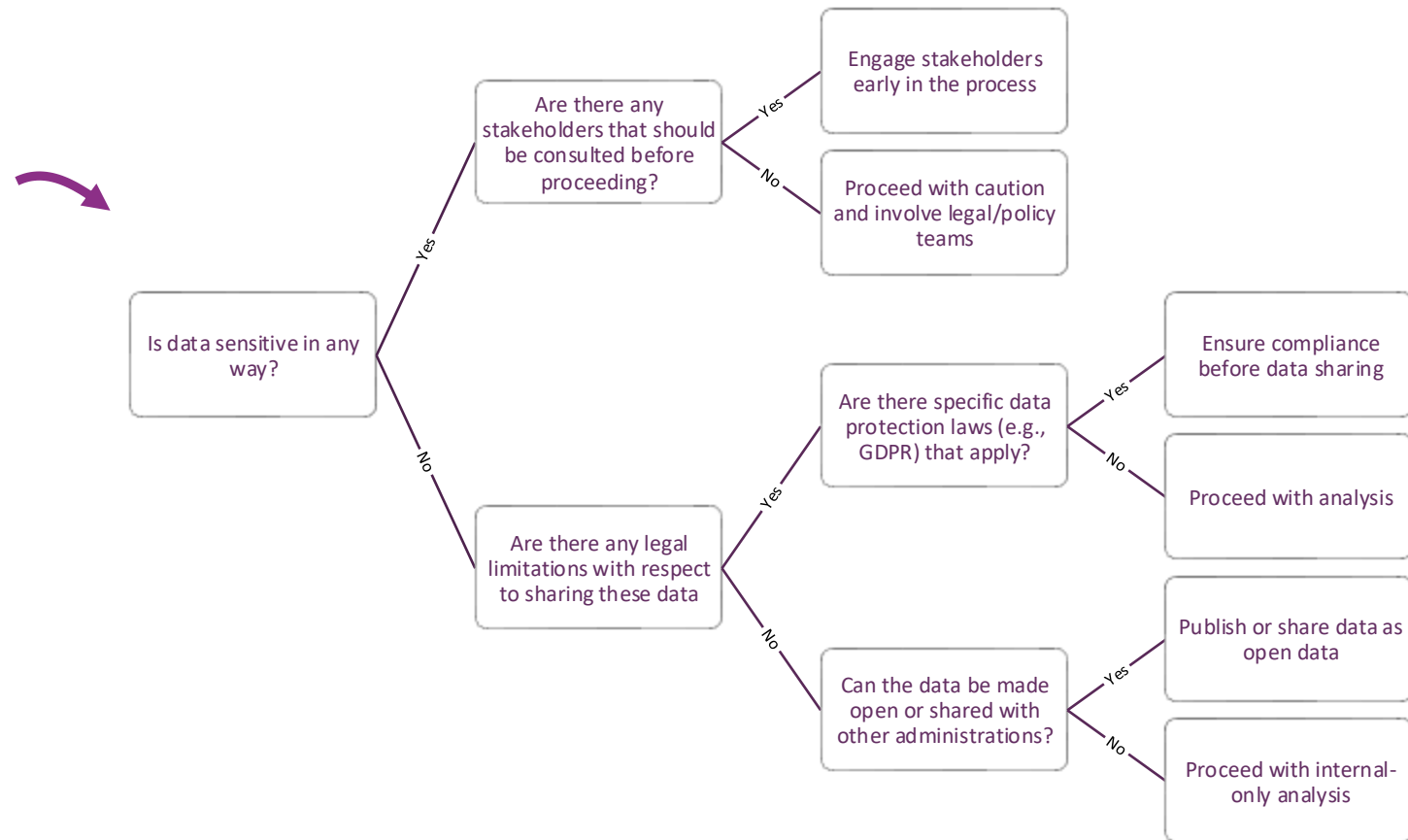
# Tool: Data Sensitivity assessment



It's important to consider data sensitivity when reusing data sets. Legal limitations, such as data protection laws, must be reviewed to ensure compliance.

Depending on these factors, data can be shared openly or used internally (with periodic legal reviews).

- ❑ Use this tool to ensure data sharing is both compliant and strategically sound
- ❑ If unsure, check the [EU data protection rules](#)



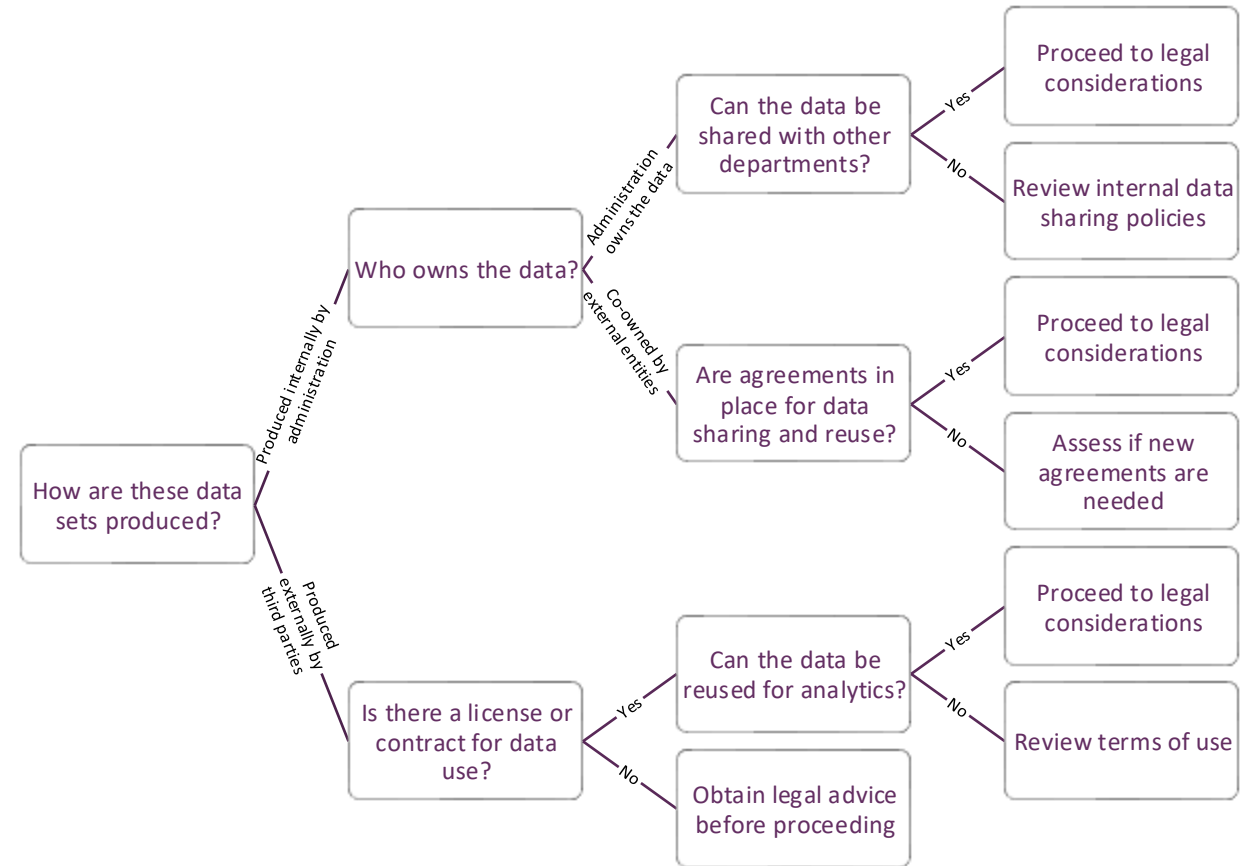
# Tool: Data Compliance assessment



Assessing how data sets are produced is crucial for public administrations because it helps ensure data integrity, ownership, and compliance with legal standards.

By understanding the origin of the data, public administrations can better manage data-sharing agreements, protect sensitive information, and ensure that the data can be used effectively for data analytics and decision-making.

Use this tool to evaluate the ownership of the data sets



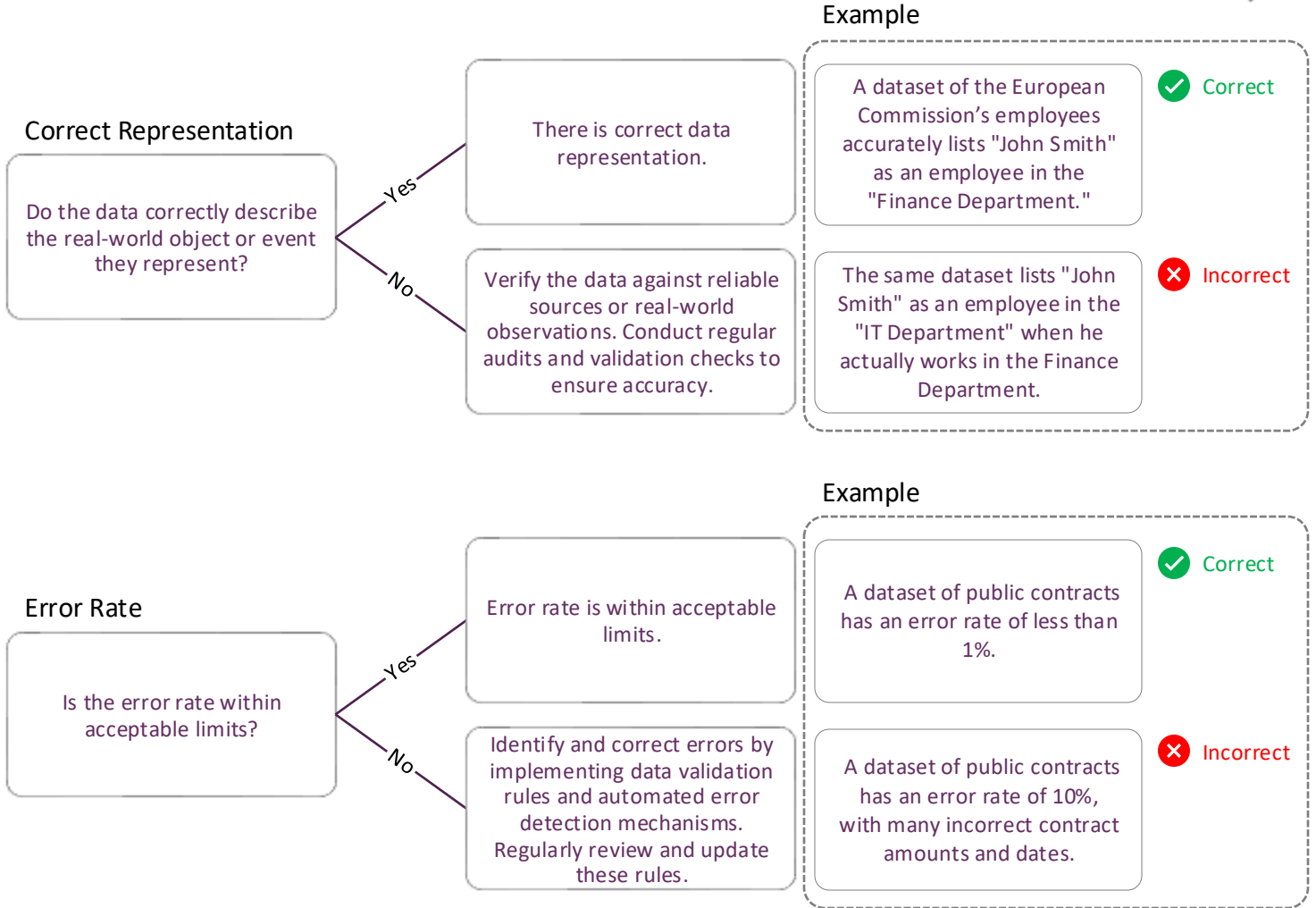
# Tool: Assessment on Accuracy



Accuracy of datasets is vital for public administrations as it ensures reliable decision-making and effective use of data.

Regular validation, audits, and error rate assessments help maintain data quality and identify inaccuracies, enabling better analytics and informed policy implementation.

- Use this tool to assess the accuracy of datasets and identifying error rates to maintain data quality.



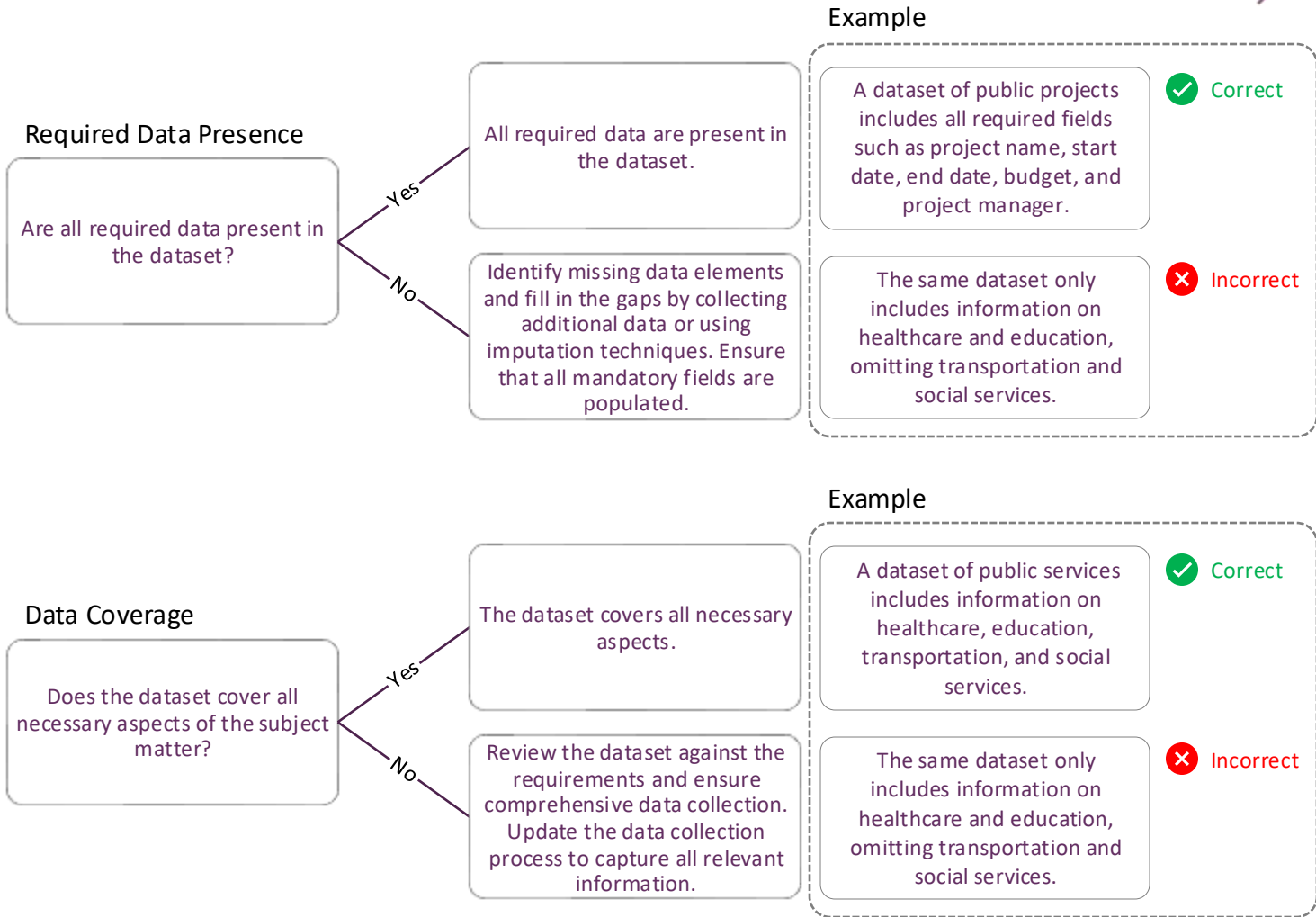
# Tool: Assessment on Completeness



Ensuring data presence and coverage is essential for effective dataset use. Address missing elements, such as compliance data in regulatory datasets, through collection or imputation.

Verify datasets cover key areas. Regular reviews ensure completeness and alignment with evolving needs.

- ❑ Use this tool to identify missing data elements, ensure mandatory fields are populated, and verify that datasets comprehensively cover all relevant aspects of the subject matter.



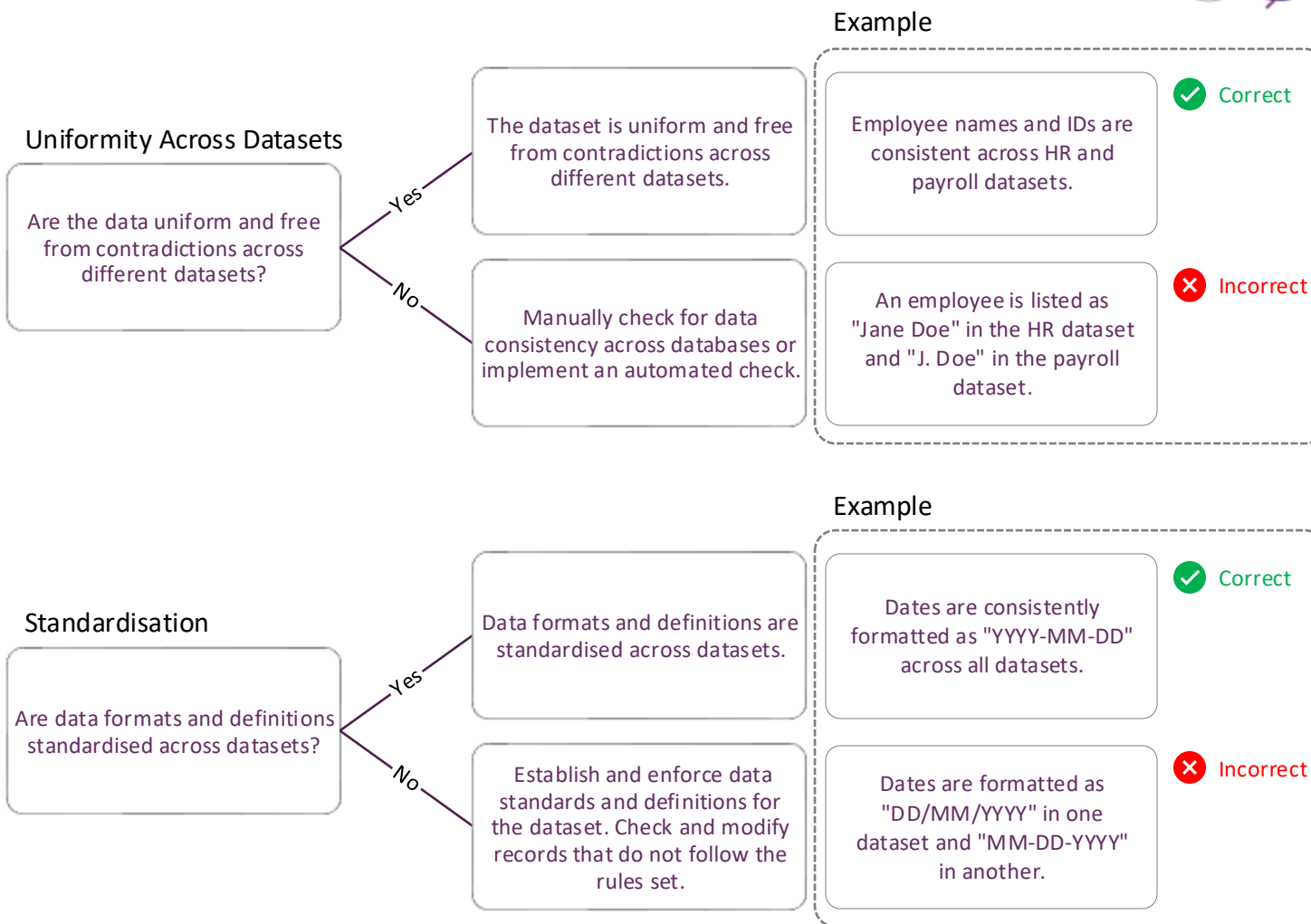


# Tool: Assessment on Consistency

Ensure data consistency by verifying uniformity across datasets and resolving contradictions through manual or automated checks.

Standardise formats and definitions, correcting non-compliant records to maintain reliability and compatibility.

- Use this tool to ensure data uniformity by reconciling inconsistencies and enforcing standard formats and definitions across datasets for improved consistency.





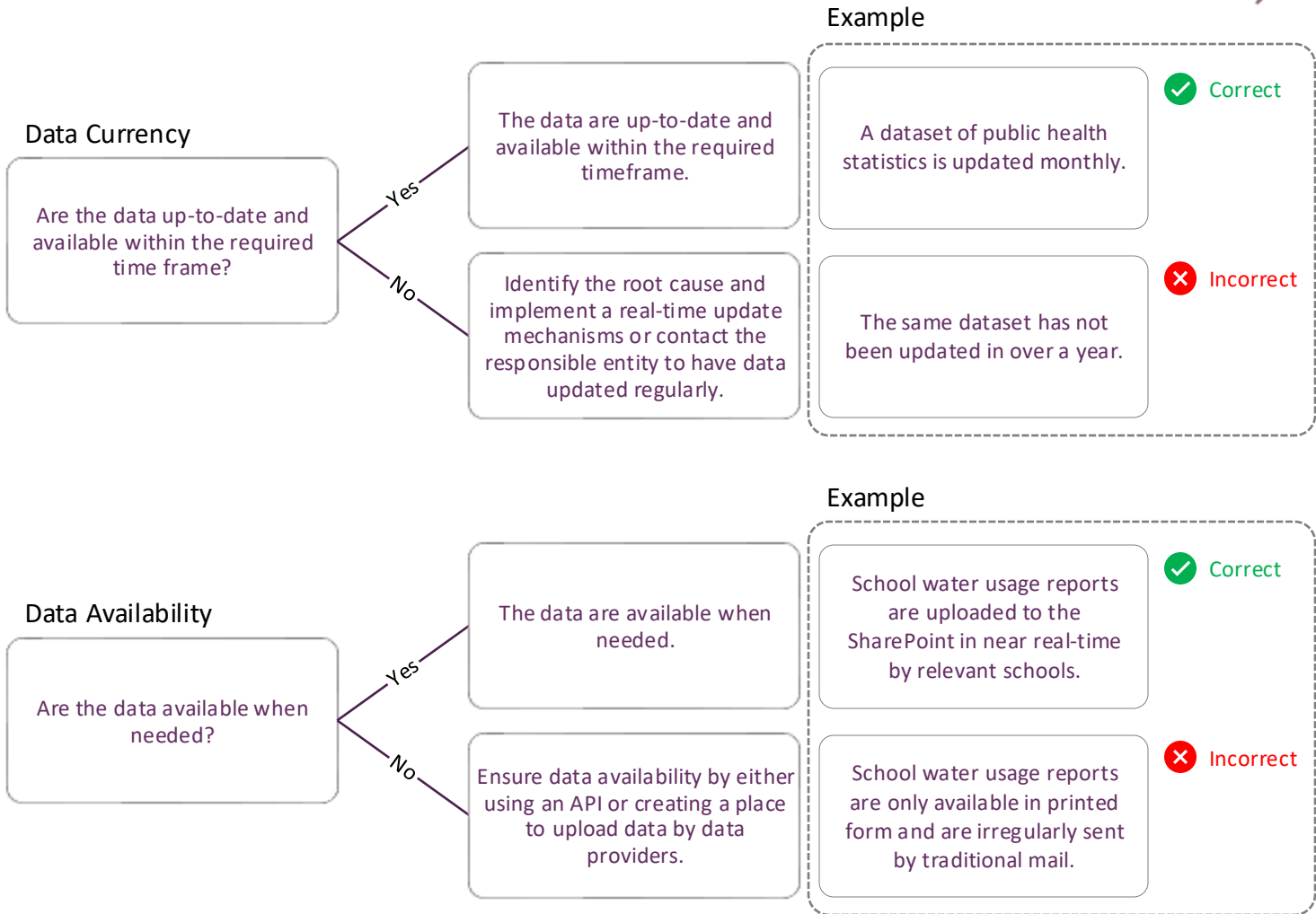
# Tool: Assessment on Timeliness



Currency of data refers to the extent to which information is current, accurate, and updated within the required timeframe to meet its intended purpose effectively.

Data availability ensures that information is consistently accessible and retrievable whenever needed, supported by reliable systems or mechanisms to prevent delays or disruptions.

- ❑ Use this tool to ensure data are up-to-date with automated refresh schedules and real-time updates, and to optimise accessibility through efficient storage and retrieval processes.

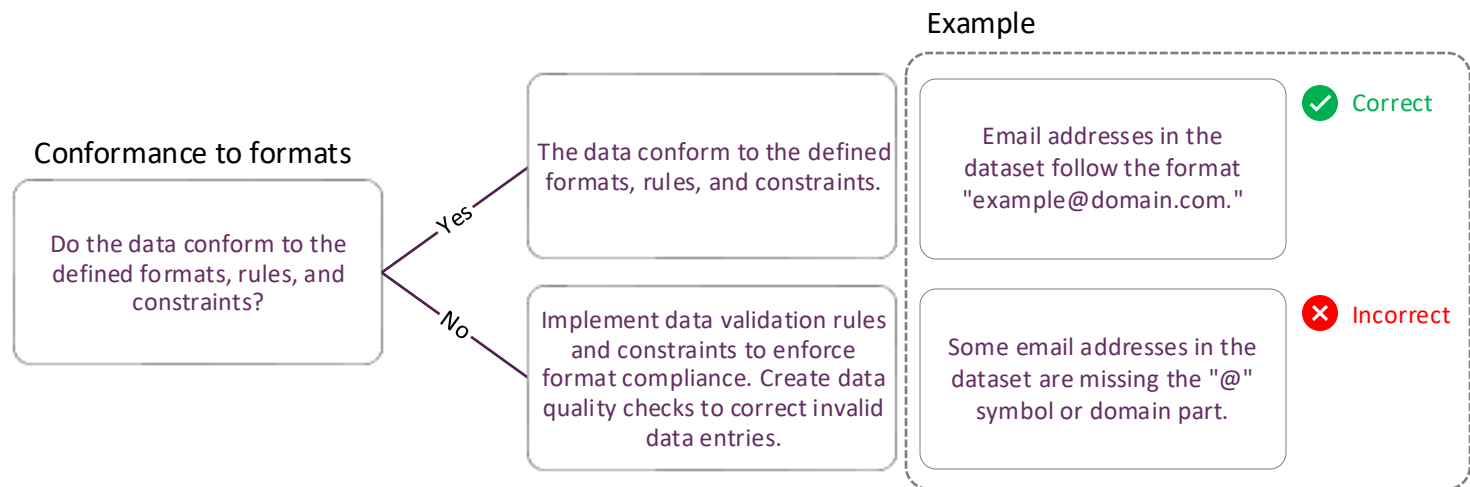


# Tool: Assessment on Validity



Data validation is the process of ensuring that data comply with predefined formats, rules, and constraints, thereby maintaining its accuracy, consistency, and reliability. This process involves checking data for errors, discrepancies, or invalid entries and applying corrective measures to uphold data quality standards.

- ❑ Use this tool to enforce data format compliance with validation rules, detect and correct errors using data quality tools, and ensure adherence to business rules through governance policies and rule-based engines.



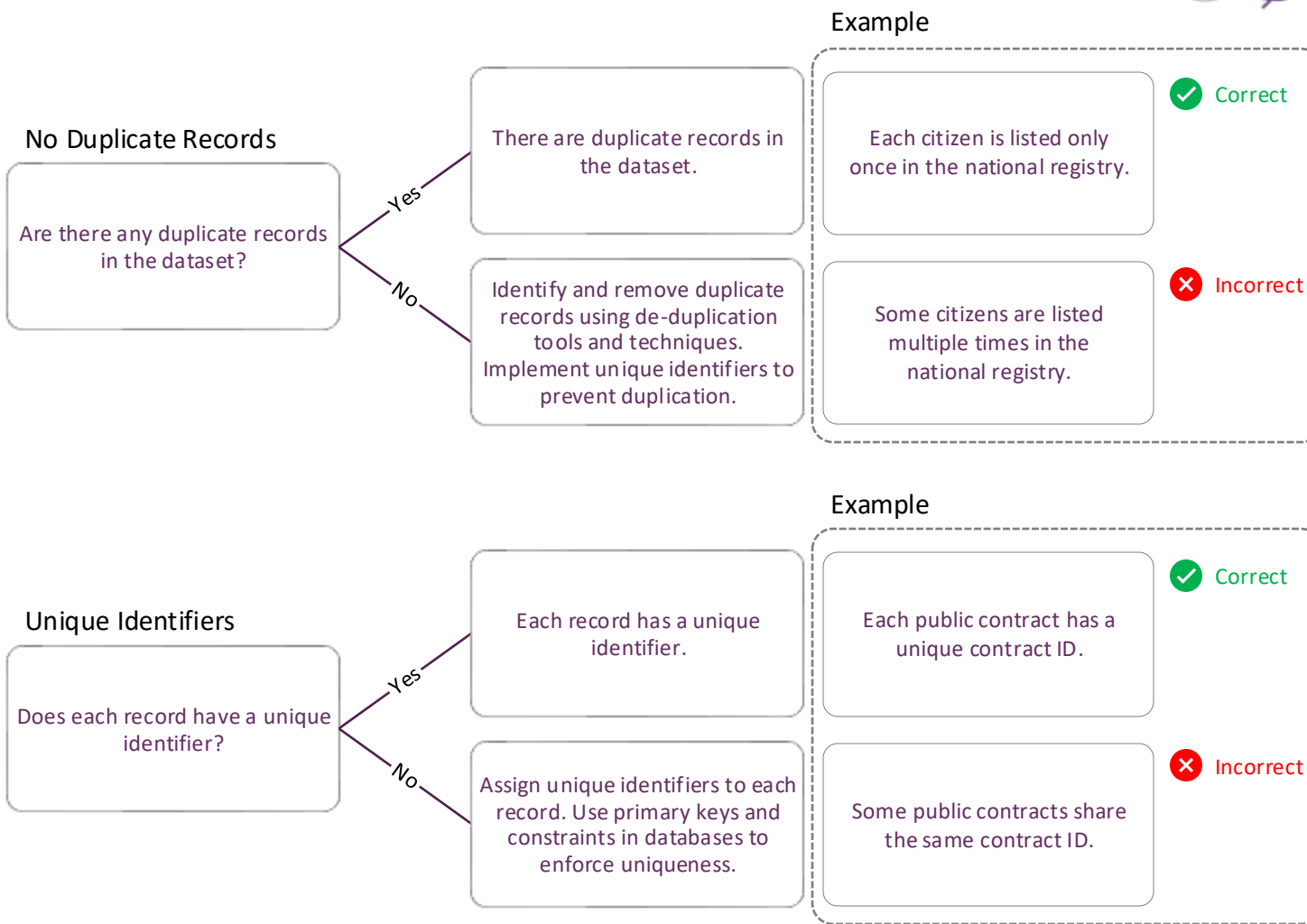


# Tool: Assessment on Uniqueness

Ensure datasets are free from duplicates and each record has a unique identifier.

Use de-duplication techniques to identify and remove duplicates, and implement unique identifiers, such as primary keys, to enforce record uniqueness and maintain data integrity.

- Use this tool to evaluate the presence of duplicate records and ensure unique identifiers are applied to each record and guiding necessary corrective actions.



# Tool: Assessment on Integrity



Data integrity refers to the accuracy, consistency, and reliability of data, ensuring it remains logically coherent and maintains valid relationships between data elements throughout its lifecycle.

- ❑ Use this tool to ensure logical coherence by enforcing referential integrity, maintaining relationships between data elements, and validating data through foreign key constraints and regular audits.



## Integrity

Are the data logically coherent and maintain relationships between different data elements?

Yes

The data are logically coherent and maintain relationships between different data elements.

No

Ensure logical coherence by checking data relationships. If broken, check at the source or employ techniques to fix or remove affected entities.

## Example

Each employee record is linked to a valid department record.

✔ Correct

Some employee records reference non-existent departments.

✘ Incorrect



# Tool: Weighting by value and effort (part 1)

- To determine which use case to prioritise, a relative estimation between use cases is crucial.
- Estimating the relative value/size of a use case is done by assigning the appropriate value following the Fibonacci scale (1, 2, 3, 5, 8, 13, 21 and so on).
- Higher numbers are rarely used

We will revisit this later and estimate Duration and the final WSJF score

Example using WSJF

Use cases	Business Value 	+	Time Criticality 	+	Risk Reduction 	=	Overall Value
"Data insight through visuals"	5		5		1		11
"Timeseries forecasting"	8		8		8		24
...	...		...		...		...
...	...		...		...		...

# Tool: Go/No Go Chart at end of Phase 1: Define



Category	Topic	Go/No-Go Criteria	Status
<b>Strategic Alignment</b>	Understanding of Strategic Fit	Does the use case align with strategic objectives?	Ready / Not Ready
<b>Collaboration</b>	Stakeholder Mapping and Engagement Skills	Are key stakeholders aligned and engaged?	Ready / Not Ready
	Communication Skills	Is communication clear across all teams?	Ready / Not Ready
<b>Technical Skills</b>	Data Management (SQL, ETL, Big Data tools)	Are data engineers and architects available?	Ready / Not Ready
	Programming (Python, R, JavaScript)	Are programmers skilled in required tools?	Ready / Not Ready
	Machine Learning and Statistical Analysis	Are data scientists capable of building ML models?	Ready / Not Ready
	Frontend Visualisation (Power BI, Tableau)	Are visualisation specialists available?	Ready / Not Ready
	DevOps and MLOps	Are DevOps and MLOps specialists in place?	Ready / Not Ready
	Cloud Architects	Are there Cloud Architects in the team?	Ready / Not Ready
<b>Supporting Skills</b>	Leadership and Project Management	Are strong leaders managing timelines and resources?	Ready / Not Ready
	Domain Expertise	Are SMEs engaged in the project?	Ready / Not Ready
	Policy and Compliance Understanding	Are data and policies compliant with regulations?	Ready / Not Ready
	Ethical Decision-Making	Are ethical standards being considered?	Ready / Not Ready
<b>Data Quality</b>	Data Quality Dimensions	Do data pass the quality checks in all dimensions?	Ready / Not Ready
<b>Planning &amp; Evaluation</b>	Prioritisation (WSJF Scoring)	Have use cases been scored and prioritised?	Ready / Not Ready
	Metrics for Success	Are KPIs well-defined for evaluation?	Ready / Not Ready

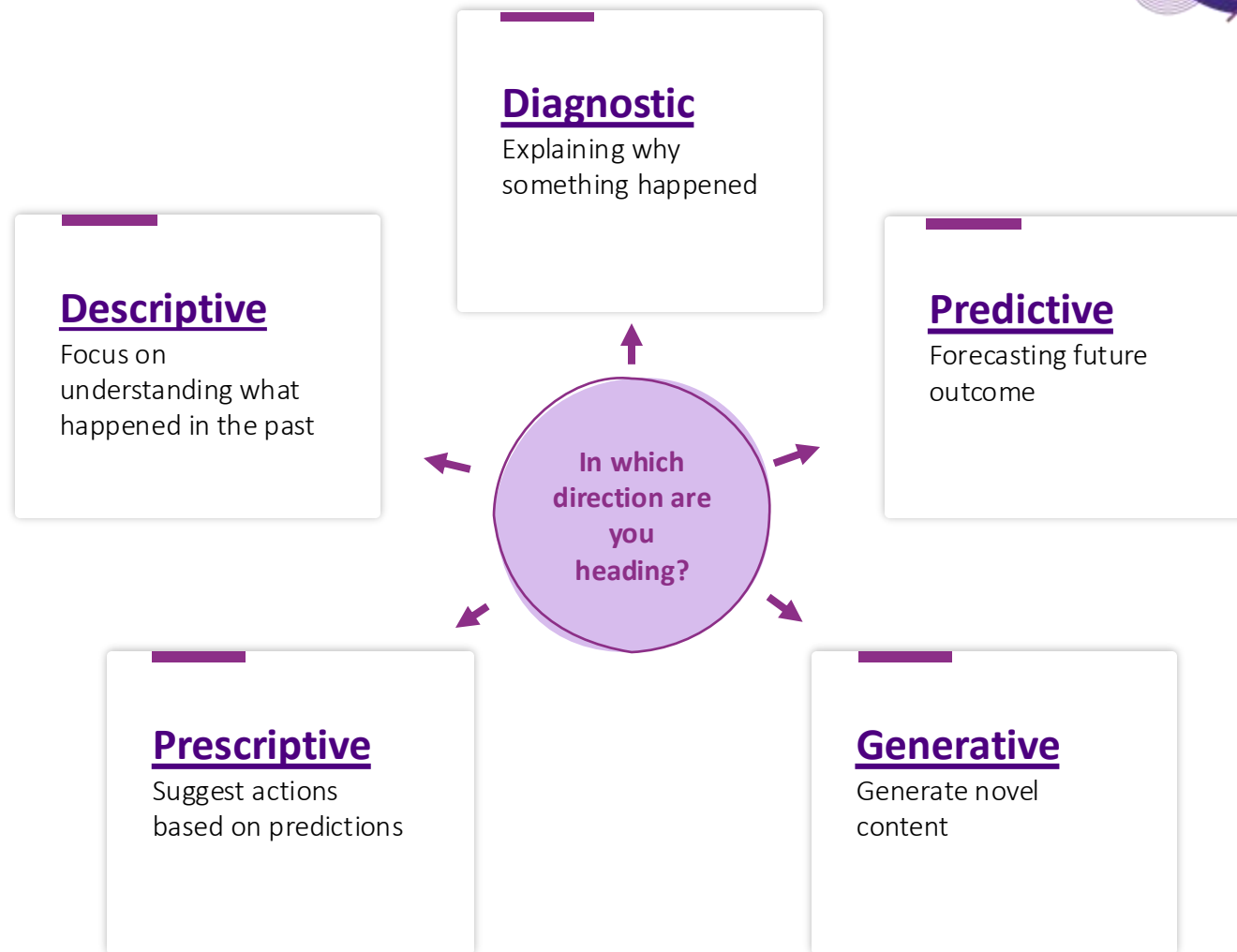
For each criterion, mark as "Ready" or "Not Ready". All should ideally be marked "Ready" before moving to the next phase or there should be a valid reason for any not being ready



# Tool: Data analytics types – Which direction?

When dealing with data analytics, numerous options are available to us. Selecting the appropriate method will aid in accomplishing your goal.

- Use the upcoming slides to delve into each of these disciplines



# Tool: Data analytics types – Descriptive and Diagnostic



## Descriptive Analytics

Focus on understanding what happened in the past. It provides insights into historical data to help understand trends, patterns, and behaviours

### Types of analysis:

- Data aggregation and summarisation
- Reporting and dashboard creation
- Trend analysis over time

### Use Case Suitability:

- Gain insight into past trends
- Understanding historical data to inform future decisions



Lisbon uses descriptive analytics for crowd monitoring.

Check the inventory of use cases to learn more about this and other applications.

## Diagnostic Analytics

Explaining why something happened and identify the causes or factors that led to a specific outcome

### Types of analysis:

- Root cause analysis
- Correlation and causality analysis
- Drill-down and data mining techniques

### Use Case Suitability:

- Understanding patterns and identifying root causes
- Help uncover hidden patterns or relationships in data



Galicia uses diagnostic analytics to assess environmental impact in the mining sector.

Check the inventory of use cases to learn more about this and other applications.



# Tool: Data analytics types – Predictive and Prescriptive



## Predictive Analytics

Forecasting future outcomes based on historical data. Use statistical models and machine learning to predict trends and behaviours

### Types of analysis:

- Predictive modelling and forecasting
- Risk assessment and scenario analysis
- Proactive decision making

### Use Case Suitability:

- Predicting demand for services, forecasting budget requirements, or assessing the likelihood of policy success



Ljubljana uses predictive analytics for controlling stormwater flows.

Check the inventory of use cases to learn more about this and other applications.

## Prescriptive Analytics

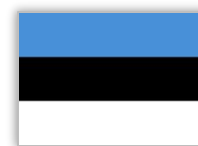
Not only predict outcomes but also suggests actions to achieve desired results. Provide recommendation based on predictive models with constraints/objectives

### Types of analysis:

- Optimisation models and simulations
- Decision trees and recommendation systems
- Scenario planning with actionable insights

### Use Case Suitability:

- Complex decision-making scenarios: Optimising resource allocation, planning emergency response strategies, improving delivery process



Tallin uses prescriptive analytics to improve urban transport planning.

Check the inventory of use cases to learn more about this and other applications.

# Tool: Data analytics types – Generative



## Generative Analytics

Focuses on creating novel content or scenarios based on existing data. It can be used to generate new text, images, simulations by learning from patterns in data

### Types of analysis:

- Text and content generation
- Synthetic data creating for training AI models
- Scenario generation for planning and innovation

### Use Case Suitability:

- Generating new policy drafts, creating training data for AI models, or simulate outcomes in urban planning

Please note that advanced generative AI (e.g. LLM) is currently not supported on the BDTI platform



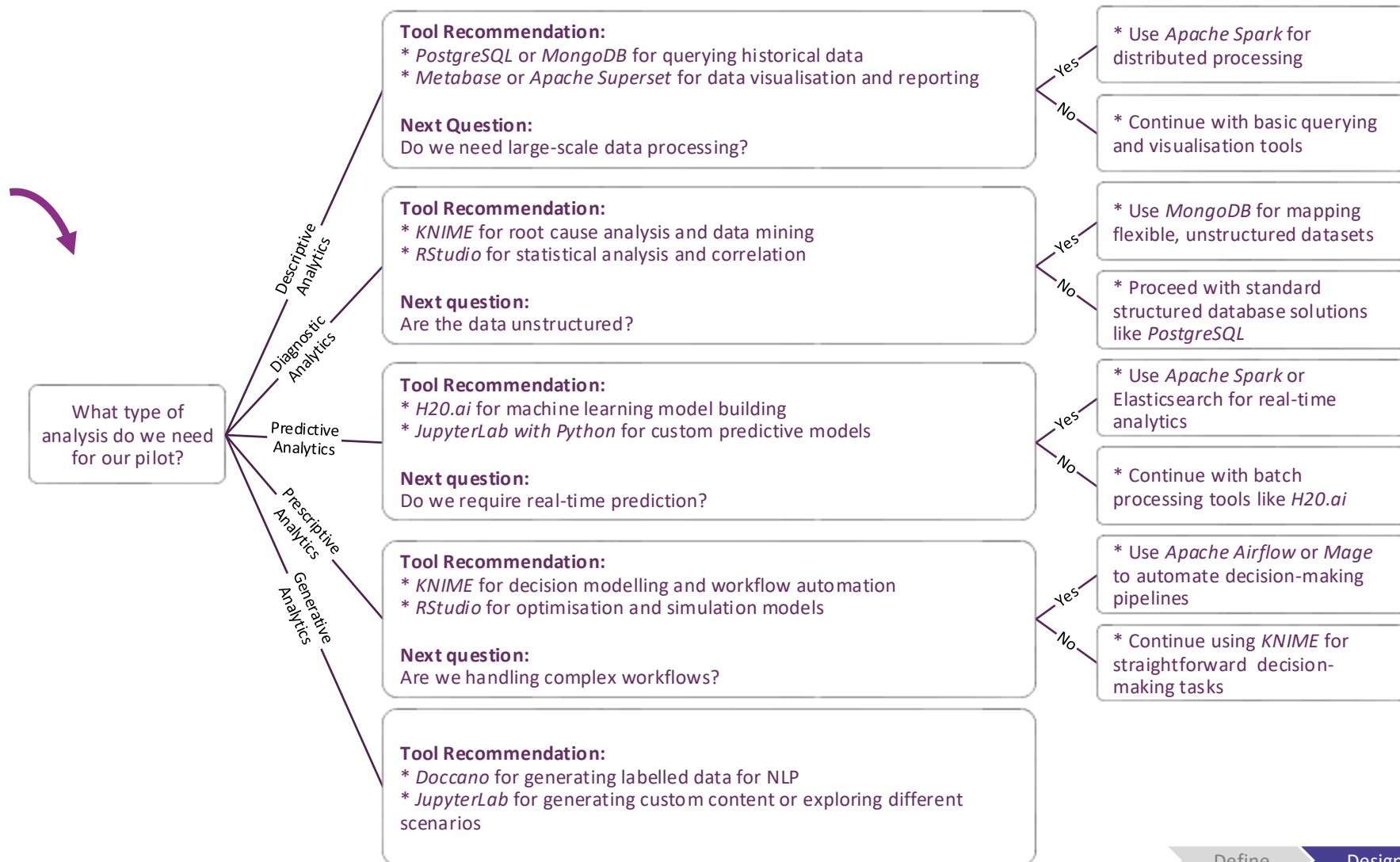
# Tool: Linking to BDTI Service offering

BDTI has many service offerings. Selecting the appropriate service is crucial for accomplishing your feature/task.

*Disclaimer: Listed tools are chosen from an ease-of-getting-started perspective. More advanced users might choose differently.*


Use this tool to quickly get started on the BDTI platform with the most appropriate tool

**Alternative:** Use the “Linking analytic tools on BDTI” Excel Spreadsheet






# Tool: BDTI software stack vs. Skills – Data Lake



<i>Tool Name</i>	<i>Description</i>	<i>Difficulty to Learn</i>	<i>Difficulty Explanation</i>	<i>Learn More</i>
<b>MinIO</b> 	MinIO is a high-performance, distributed object storage system designed for large-scale data infrastructure. It is compatible with Amazon S3 cloud storage service.	Moderate	MinIO requires understanding of storage concepts and setting up the server. While the initial setup is straightforward, configuring it for high availability, security, and performance optimisation can be challenging. Advanced features like erasure coding and object locking add to the complexity.	<a href="#">MinIO on BDTI's website</a> <a href="#">MinIO's website</a> <a href="#">MinIO's documentation</a>





# Tool: BDTI software stack vs. Skills – Databases

<i>Tool Name</i>	<i>Description</i>	<i>Difficulty to Learn</i>	<i>Difficulty Explanation</i>	<i>Learn More</i>
<b>PostgreSQL</b> 	PostgreSQL is a powerful, open-source relational database management system (RDBMS) known for its robustness, extensibility, and SQL compliance.	Moderate	PostgreSQL involves learning SQL and relational database concepts. Basic CRUD operations are easy, but mastering advanced features like indexing, query optimisation, and stored procedures requires significant effort. Understanding its extensive configuration options and extensions also adds to the learning curve.	<ul style="list-style-type: none"> <li>• <a href="#">PostgreSQL on BDTI's website</a></li> <li>• <a href="#">PostgreSQL's website</a></li> <li>• <a href="#">PostgreSQL's documentation</a></li> </ul>
<b>MongoDB</b> 	MongoDB is a popular NoSQL database that uses a flexible, JSON-like document structure, making it ideal for handling unstructured data and scaling horizontally.	Moderate	MongoDB requires understanding of document-oriented databases and basic CRUD operations. While it is easier to start with compared to relational databases, advanced features like aggregation pipelines, indexing strategies, and sharding for distributed data management add complexity.	<ul style="list-style-type: none"> <li>• <a href="#">Mongo DB on BDTI's website</a></li> <li>• <a href="#">MongoDB's website</a></li> <li>• <a href="#">MongoDB's documentation</a></li> </ul>
<b>Virtuoso</b> 	Virtuoso is a multi-model database management system that supports relational, graph, and RDF data models, often used for linked data and semantic web applications.	Hard	Virtuoso supports multiple data models and query languages, including SQL, RDF, and SPARQL. This requires a deep understanding of both relational and semantic web technologies. The complexity is further increased by the need to manage and optimise performance across these different models.	<ul style="list-style-type: none"> <li>• <a href="#">Virtuoso on BDTI's website</a></li> <li>• <a href="#">Virtuoso's website</a></li> <li>• <a href="#">Virtuoso's documentation</a></li> </ul>




# Tool: BDTI software stack vs. Skills - Orchestration



<i>Tool Name</i>	<i>Description</i>	<i>Difficulty to Learn</i>	<i>Difficulty Explanation</i>	<i>Learn More</i>
<b>Airflow</b> 	Apache Airflow is an open-source platform to programmatically author, schedule, and monitor workflows, making it ideal for managing complex data pipelines.	Moderate	Airflow requires understanding of Python and workflow orchestration concepts. While setting up basic workflows is straightforward, managing complex dependencies, handling failures, and scaling the system for large workflows can be challenging. Advanced features like custom operators and sensors add to the complexity.	<ul style="list-style-type: none"><li>• <a href="#">Airflow on BDTI's website</a></li><li>• <a href="#">Airflow's website</a></li><li>• <a href="#">Airflow's documentation</a></li></ul>
<b>Mage</b> 	Mage is an open-source data pipeline tool that simplifies the process of building, running, and monitoring data workflows, focusing on ease of use and scalability.	Moderate	Mage involves understanding data engineering concepts and programming skills. Setting up basic data pipelines is easy, but handling complex data transformations, integrations with various data sources, and ensuring data quality and reliability require more effort.	<ul style="list-style-type: none"><li>• <a href="#">Mage on BDTI's website</a></li><li>• <a href="#">Mage's website</a></li><li>• <a href="#">Mage's documentation</a></li></ul>





# Tool: BDTI software stack vs. Skills – Environments (1/2)



<i>Tool Name</i>	<i>Description</i>	<i>Difficulty to Learn</i>	<i>Difficulty Explanation</i>	<i>Learn More</i>
 <b>JupyterLab</b>	JupyterLab is an open-source web-based interactive development environment for Jupyter notebooks, code, and data, widely used in data science and research.	Easy	JupyterLab is easy to learn if you are familiar with Python and data analysis. Basic usage is straightforward, but advanced features like custom extensions, interactive widgets, and integrating with other tools and libraries require more effort. Managing large notebooks and optimising performance can also be challenging.	<ul style="list-style-type: none"><li>• <a href="#">JupyterLab on BDTI's website</a></li><li>• <a href="#">JupyterLab's website</a></li><li>• <a href="#">JupyterLab's documentation</a></li></ul>
 <b>RStudio</b>	RStudio is an integrated development environment (IDE) for R, a programming language for statistical computing and graphics, providing tools for data analysis and visualisation.	Easy	RStudio requires knowledge of the R programming language and statistical analysis. Basic usage is easy, but advanced features like package development, Shiny applications, and integrating with other tools and libraries add complexity. Understanding R's extensive ecosystem and optimising performance also require significant effort.	<ul style="list-style-type: none"><li>• <a href="#">RStudio on BDTI's website</a></li><li>• <a href="#">RStudio's website</a></li><li>• <a href="#">RStudio's documentation</a></li></ul>
 <b>VS Code</b>	Visual Studio Code (VS Code) is a free, open-source code editor developed by Microsoft, known for its versatility, extensions, and support for various programming languages.	Easy	VS Code is user-friendly and easy to learn for basic coding and development tasks. Basic usage is straightforward, but advanced features like custom extensions, debugging, and integrating with other tools and services require more effort. Managing large projects and optimising performance can also be challenging.	<ul style="list-style-type: none"><li>• <a href="#">VS Code on BDTI's website</a></li><li>• <a href="#">VS Code's website</a></li><li>• <a href="#">VS Code's documentation</a></li></ul>

# Tool: BDTI software stack vs. Skills – Environments (2/2)



<b>Tool Name</b>	<b>Description</b>	<b>Difficulty to Learn</b>	<b>Difficulty Explanation</b>	<b>Learn More</b>
 <b>KNIME</b>	KNIME (Konstanz Information Miner) is an open-source data analytics, reporting, and integration platform, enabling users to create data workflows through a visual interface.	Moderate	KNIME requires understanding of data analytics concepts and some configuration skills. Basic usage is easy, but advanced features like custom nodes, complex workflows, and integrating with various data sources add complexity. Managing and optimising performance also require significant effort.	<ul style="list-style-type: none"> <li>• <a href="#">KNIME on BDTI's website</a></li> <li>• <a href="#">KNIME's website</a></li> <li>• <a href="#">KNIME's documentation</a></li> </ul>
 <b>H2O.ai</b>	H2O.ai is an open-source platform for machine learning and artificial intelligence, offering tools for building and deploying predictive models.	Hard	H2O.ai requires knowledge of machine learning concepts, data preprocessing, and programming. The complexity increases with the need to optimise models, handle large-scale data, and integrate with other tools and libraries. Understanding H2O's extensive ecosystem and managing performance also require significant effort.	<ul style="list-style-type: none"> <li>• <a href="#">H2O.ai on BDTI's website</a></li> <li>• <a href="#">H2O.ai's website</a></li> <li>• <a href="#">H2O.ai's documentation</a></li> </ul>
 <b>QGIS</b>	QGIS (Quantum GIS) is an open-source geographic information system (GIS) application that supports viewing, editing, and analysing geospatial data.	Moderate	QGIS requires understanding of GIS concepts and some configuration skills. Basic usage is easy, but advanced features like custom plugins, spatial analysis, and integrating with various data sources add complexity. Managing and optimising performance also require significant effort.	<ul style="list-style-type: none"> <li>• <a href="#">QGIS on BDTI's website</a></li> <li>• <a href="#">QGIS's website</a></li> <li>• <a href="#">QGIS's documentation</a></li> </ul>
 <b>doccano</b>	doccano is an open-source annotation tool for text data, enabling users to create labelled datasets for natural language processing (NLP) tasks.	Easy	doccano is relatively easy to learn for basic text annotation tasks. Basic usage is straightforward, but advanced features like custom annotation schemes, integrating with other tools and services, and managing large annotation projects require more effort. Optimising performance and ensuring data quality can also be challenging.	<ul style="list-style-type: none"> <li>• <a href="#">doccano on BDTI's website</a></li> <li>• <a href="#">doccano's website</a></li> <li>• <a href="#">doccano's documentation</a></li> </ul>



# Important: Data Quality Assessment – FAIR Principles



## Findable

**Unique Identifiers:** Percentage of datasets with globally unique and persistent identifiers.

**Metadata Richness:** Average number of metadata elements per dataset.

**Searchability:** Number of datasets indexed in searchable repositories or registries.

**Metadata Completeness:** Percentage of datasets with complete metadata descriptions.

## Interoperable

**Standardised Formats:** Percentage of datasets using standardised, machine-readable formats (e.g., JSON, XML, CSV).

**Controlled Vocabularies:** Percentage of datasets using controlled vocabularies or ontologies (e.g., Data Catalogue vocabulary (DCAT)) that follow FAIR principles.

**Cross-Referencing:** Percentage of datasets with qualified references to other datasets or metadata.

**Semantic Annotation:** Percentage of datasets with semantic annotations to enhance machine readability and interoperability.

## Accessible

**Protocol Standardisation:** Percentage of datasets accessible via standardised communication protocols (e.g., HTTPS, FTP).

**Open Access:** Percentage of datasets that are openly accessible without restrictions.

**Authentication and Authorisation:** Percentage of datasets that support authentication and authorisation mechanisms where necessary.

**Metadata Availability:** Percentage of datasets with accessible metadata, even if the data themselves are restricted or no longer available.

## Reusable

**Licensing Clarity:** Percentage of datasets with clear and accessible data usage licenses.

**Provenance Information:** Percentage of datasets with detailed provenance information (e.g., data source, creation date, modifications).

**Community Standards Compliance:** Percentage of datasets that comply with domain-relevant community standards.

**Data Quality Indicators:** Number of datasets with documented data quality indicators (e.g., accuracy, completeness, consistency).



IMPORTANT

# Tool: Assessment on Data Quality – Findable



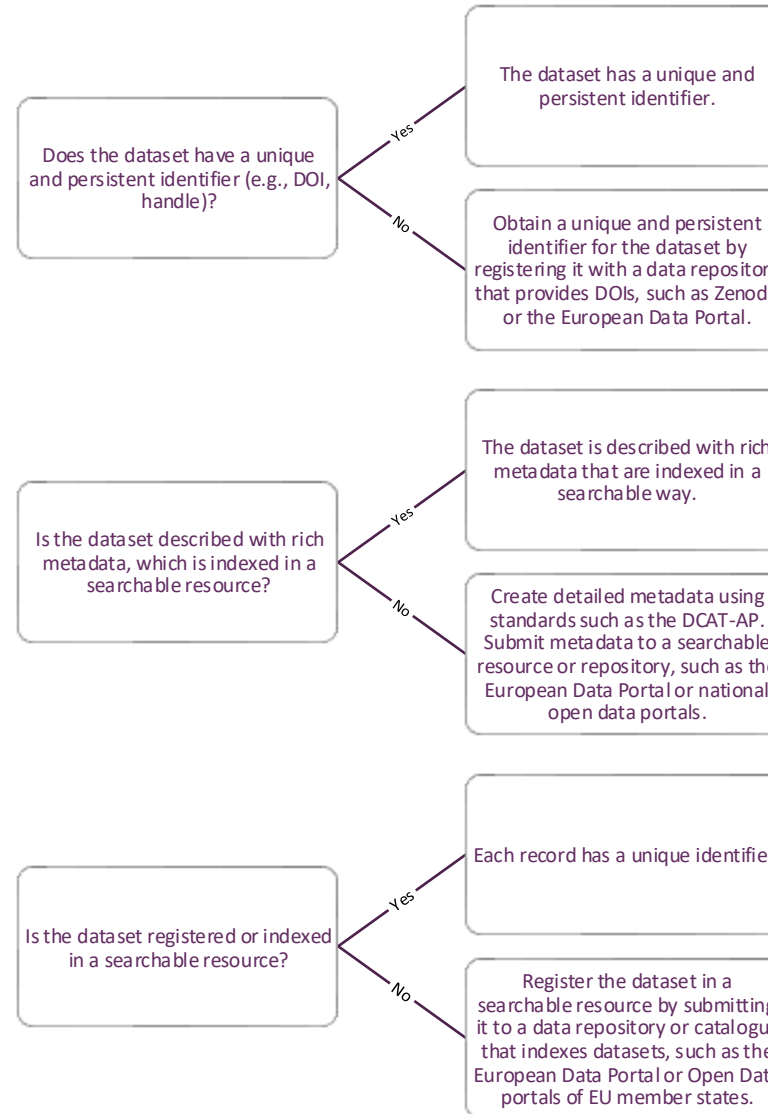
Ensure datasets are discoverable, accessible, and reusable by assigning unique identifiers (e.g., DOI) and creating rich metadata using standards like DCAT-AP.

Register datasets in searchable repositories such as the European Data Portal to enhance visibility and long-term access. Use this tool to verify compliance with metadata standards.

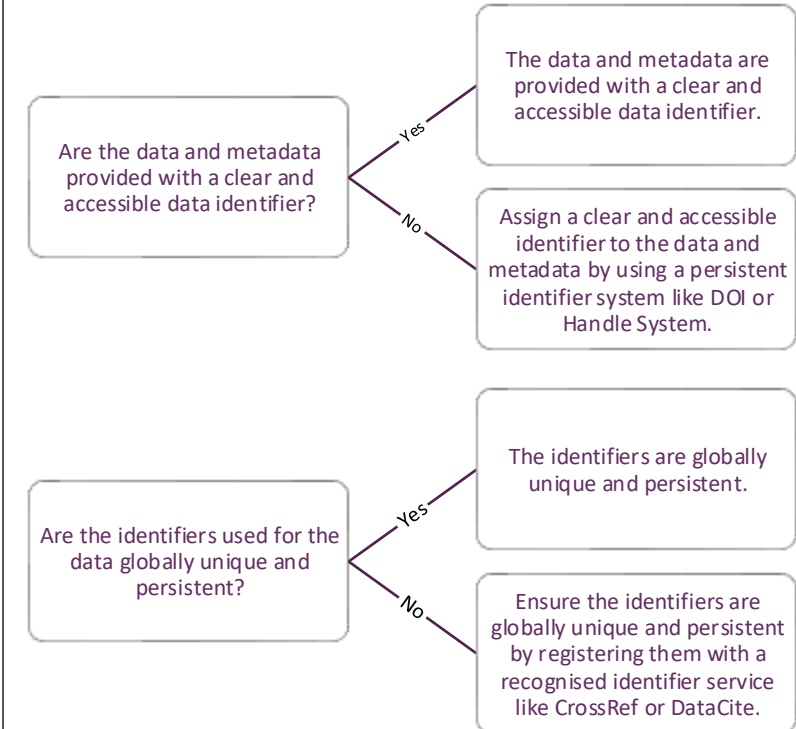
- ❑ Use this tool to ensure datasets have unique identifiers and rich metadata, registered in searchable repositories for accessibility and reuse.



## Metadata and Documentation



## Data Identification



# Tool: Assessment on Data Quality – Accessible



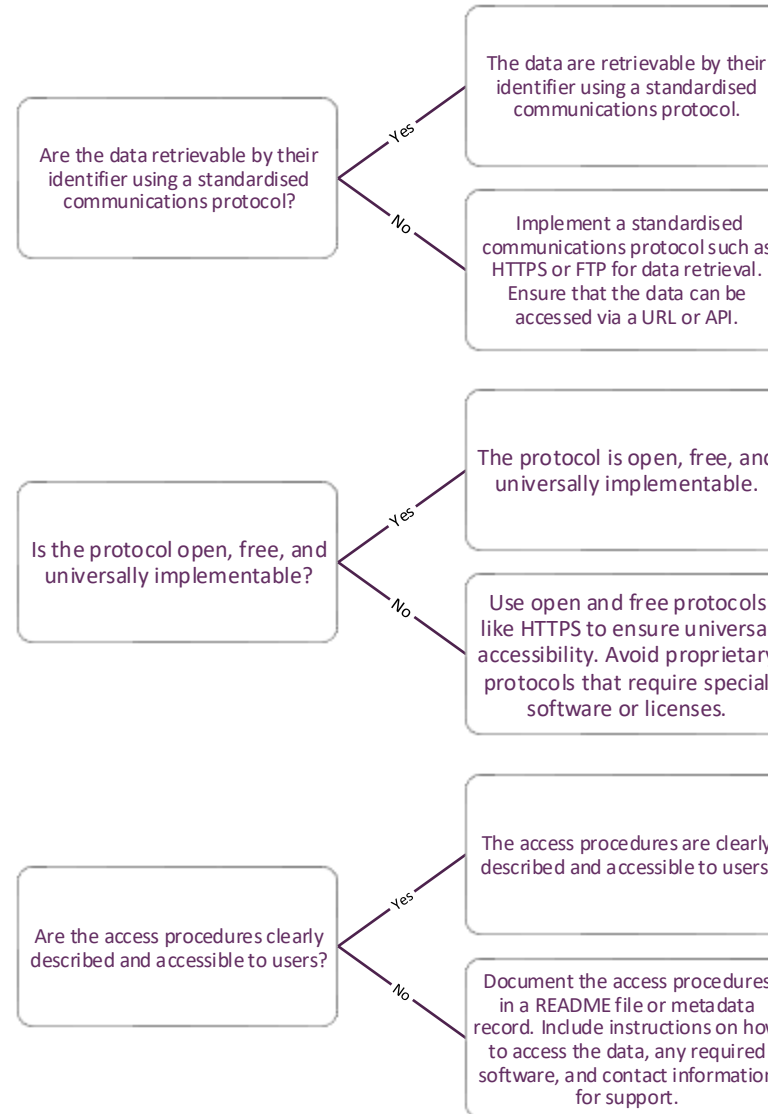
Ensuring data access involves using open and standardised protocols for data retrieval, while clearly documenting access procedures for users.

Metadata should remain accessible even if the data are unavailable. Clear conditions for data access and reuse should be defined and communicated through appropriate licensing to promote transparency and compliance.

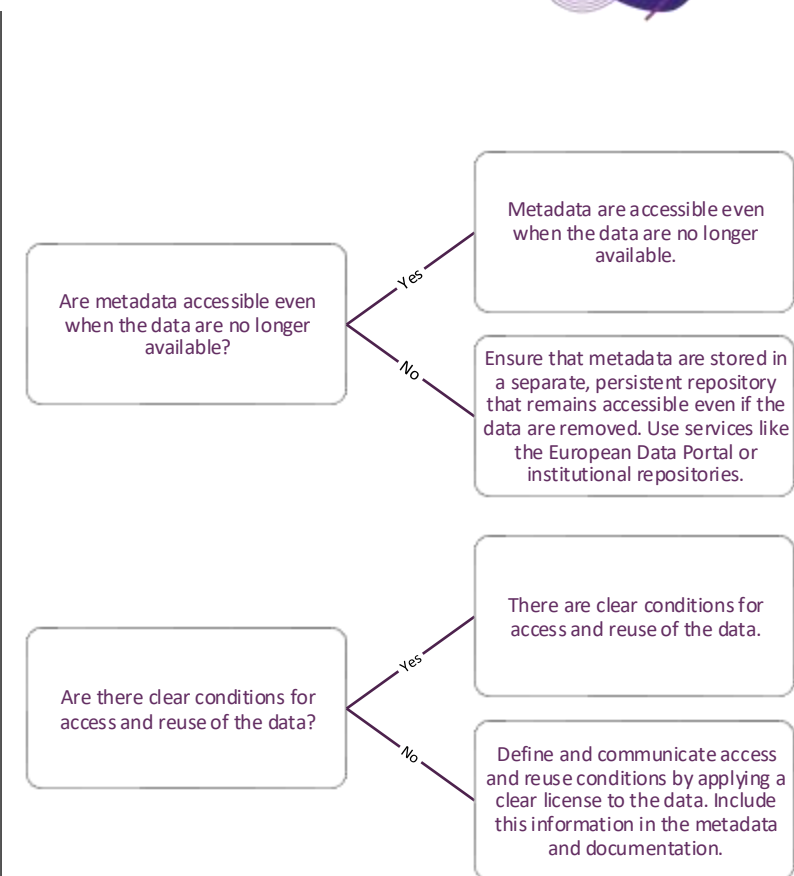
- ❑ Use this tool to ensure data are accessible via open protocols, retrieval procedures are documented, metadata remains available, and access conditions are clearly licensed.



## Data Access



## Authorisation and Authentication





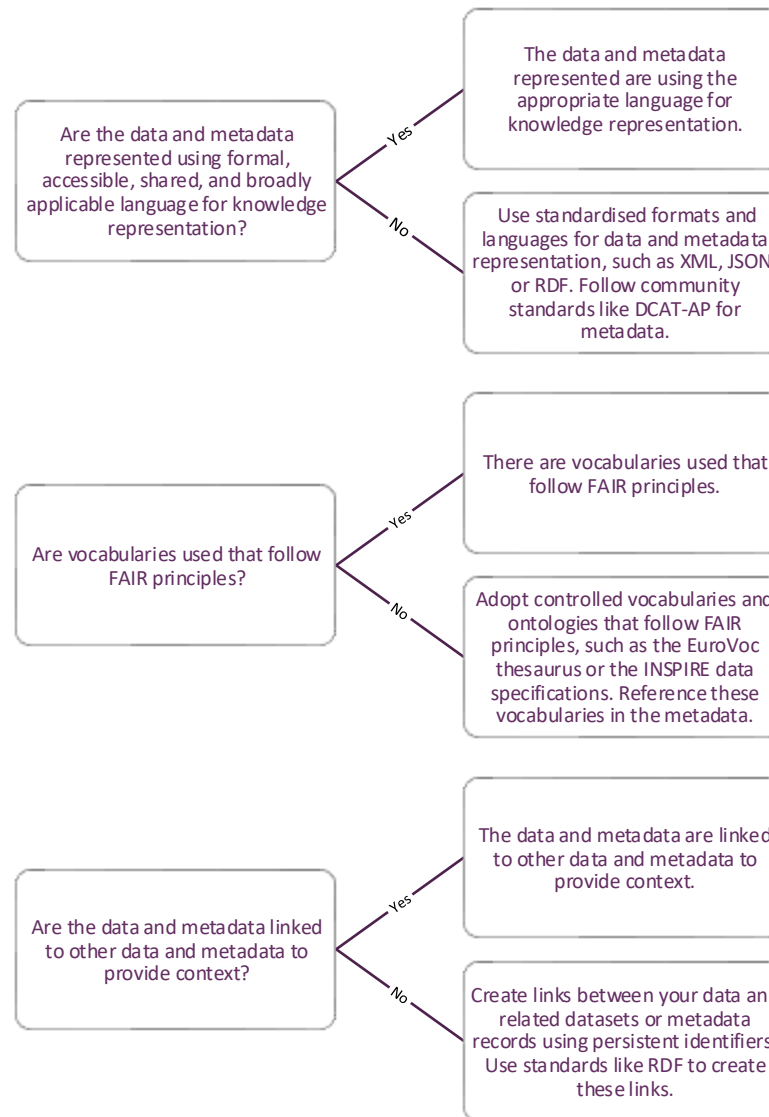
# Tool: Assessment on Data Quality – Interoperable

Ensure data and metadata use standardised formats like XML, JSON, or RDF and adopt FAIR-aligned vocabularies such as EuroVoc or INSPIRE. Link datasets using persistent identifiers for context and follow metadata standards like DCAT-AP. Use common data formats like CSV or JSON to ensure interoperability with workflows.

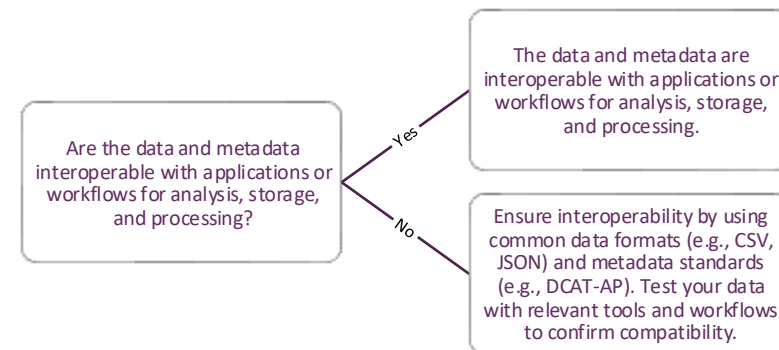
- ❑ Use this tool to ensure data and metadata follow standards, adopt FAIR vocabularies, link datasets, and use interoperable formats.



## Data and Metadata Standards



## Data Integration



# Tool: Assessment on Data Quality – Reusable



Ensure data and metadata are richly described with accurate attributes, detailed methodologies, and clear context using standards like DDI.

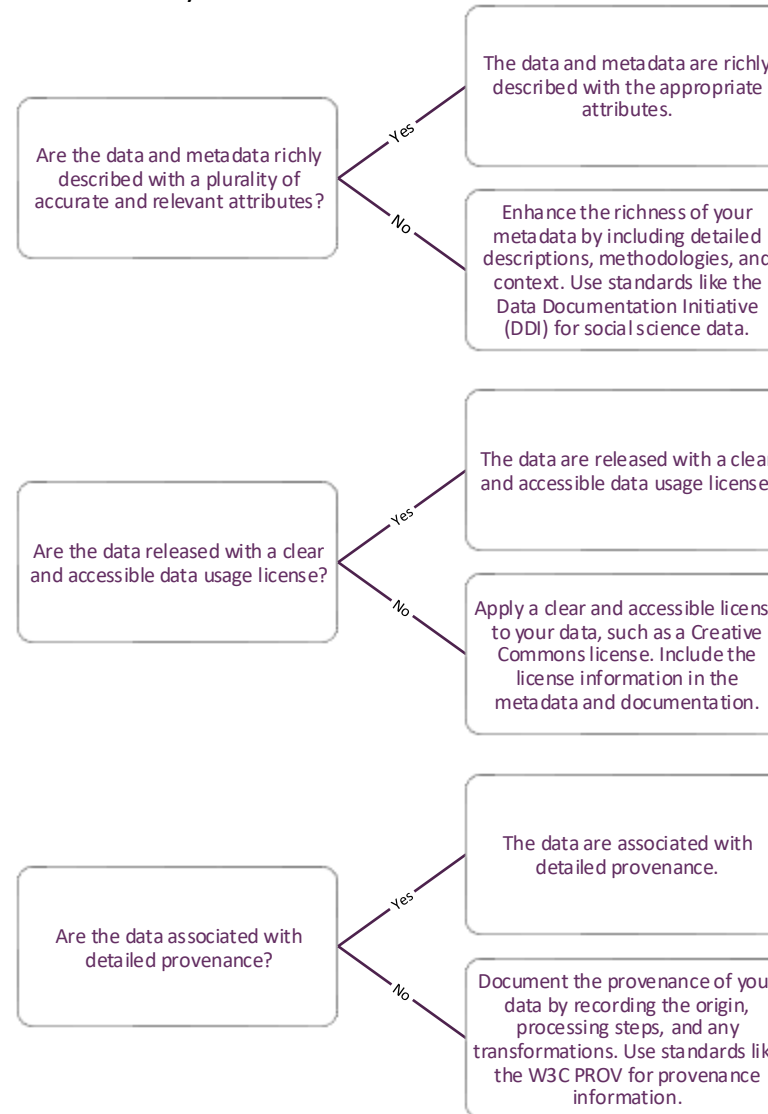
Apply an accessible license, and document provenance with origins, processing steps, and transformations such as W3C PROV.

Comply with community standards like the INSPIRE Directive or DCAT-AP to enhance interoperability and usability.

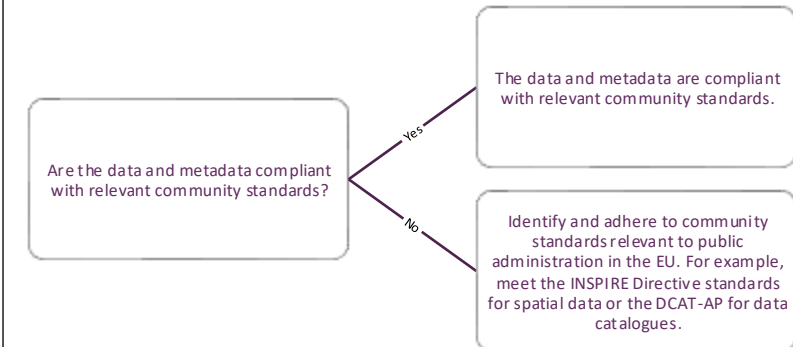
- Use this tool to ensure rich metadata, apply a clear license, document provenance with W3C PROV, and meet standards like INSPIRE or DCAT-AP for interoperability.



## Data Quality and Provenance



## Community Standards





# Tool: Questionnaire for Ethical Considerations (1/2)

## How do we avoid bias and ensure fairness in analytic decision-making?

Topic	Description	Question
<b>Bias</b>	<p>Identify Potential Sources of Bias.</p> <p>Bias refers to the systematic error introduced into data collection, analysis, or interpretation that skews results in a particular direction. This can occur due to various factors such as the selection of data sources, the design of algorithms, or the subjective decisions made by data scientists. Addressing bias is crucial to ensure that the data product provides fair, accurate, and representative insights for public policy and decision-making.</p>	<ul style="list-style-type: none"><li><input type="checkbox"/> <i>What biases might exist in the data being used?</i></li><li><input type="checkbox"/> <i>Are certain groups underrepresented or overrepresented?</i></li></ul>
<b>Fairness</b>	<p>Implement Fairness Mechanisms.</p> <p>Fairness mechanisms ensure that data collection, analysis, and usage are unbiased and equitable, promoting transparency and accountability. These mechanisms include implementing algorithms that prevent discrimination, ensuring diverse and representative data sets, and providing clear guidelines for ethical data usage. By doing so, public administration can make informed decisions that benefit all citizens fairly and justly.</p>	<ul style="list-style-type: none"><li><input type="checkbox"/> <i>What techniques can be used to ensure fairness, such as rebalancing datasets or applying fairness constraints in model training?</i></li></ul>
<b>Fairness</b>	<p>Engage Stakeholders in Fairness Discussions.</p> <p>Engaging stakeholders in fairness discussions involves actively involving diverse groups in conversations about equitable data practices. This includes organising forums and workshops to gather input from various community members, ensuring their perspectives are considered in decision-making processes. By fostering open dialogue, public administration can build trust and create data products that reflect the needs and values of all stakeholders.</p>	<ul style="list-style-type: none"><li><input type="checkbox"/> <i>How can we ensure that fairness considerations are transparent and understood by all stakeholders?</i></li></ul>

# Tool: Questionnaire for Ethical Considerations (2/2)



## How do we avoid bias and ensure fairness in analytic decision-making?

<b>Topic</b>	<b>Description</b>	<b>Question</b>
<b>Privacy</b>	<p>Assess Data Privacy Requirements.</p> <p>Assessing data privacy requirements involves identifying and understanding the legal and ethical standards for protecting personal information. This includes conducting thorough privacy impact assessments to evaluate potential risks and implementing robust data protection measures. By ensuring compliance with privacy regulations, public administration can safeguard citizens' data and maintain public trust.</p>	<input type="checkbox"/> <i>What are the legal and regulatory requirements regarding data privacy?</i>
<b>Privacy</b>	<p>Implement Data Anonymisation.</p> <p>Implementing data anonymisation involves transforming personal data into a format that prevents the identification of individuals. This process includes techniques such as data masking, aggregation, and encryption to ensure that sensitive information is protected. By anonymising data, public administration can use and share data responsibly while maintaining privacy and compliance with regulations.</p>	<input type="checkbox"/> <i>What methods can be used to anonymise or pseudonymise data to protect individual privacy?</i>

# Tool: Translation into actionable insights



## Interpreting Findings:

- What do the analytical results reveal about the problem or opportunity at hand?
- How do these findings align with the objectives defined at the beginning of the project?

Act



- Summarise the findings in a way that highlights the most relevant points
- Take inspiration from existing BDTI pilots : [Pilot & Success Stories](#)

## Translating Insights into Strategic Recommendations:

- Based on the analysis, what specific actions should the organisation take?
- How do these recommendations support or enhance existing initiatives and programs?

Act



- Provide clear, evidence-based recommendations that are directly linked to the insights derived from the data
- Discover the [Data in Publications Guide](#) available at data.europa.eu

## Communicating Insights to Stakeholders:

- How should the insights be communicated to different stakeholder groups, including decision-makers, operational teams, and the public?

Act



- Visualise data and insights where possible: [example on BDTI](#)
- Discover the [Data Visualisation Guide](#) available at data.europa.eu





# Tool: Weighting by value and effort (part 2)

Continued example using WSJF

Use cases	Business Value 	+	Time Criticality 	+	Risk Reduction 	=	Overall Value	/	Job Size 	=	WSJF
"Data insight through visuals"	5		5		1		11		1		11.0
"Timeseries forecasting"	8		8		8		24		5		4.8
...	...		...		...		...		..		..
...	...		...		...		...		..		..

Use this tool to focus effort on use cases with the highest WSJF score



# Tool: Go/No Go Chart at end of Phase 2: Design



Category	Topic	Go/No-Go Criteria	Status
<b>Analytics Needs</b>	Aligning objectives with analytics type (descriptive, predictive, etc.).	Have you identified the correct analytics type for your objective and ensured data quality supports it?	<span>Ready</span> / <span>Not Ready</span>
<b>Data Quality</b>	FAIR principles: Findable, Accessible, Interoperable, Reusable.	Are your datasets FAIR-compliant, with unique identifiers, rich metadata, and open access where possible?	<span>Ready</span> / <span>Not Ready</span>
<b>Tool Knowledge</b>	Familiarity with tools like KNIME, RStudio, PostgreSQL, JupyterLab, etc.	Do you or your team have the required expertise in the selected tools, or viable alternatives for quick upskilling?	<span>Ready</span> / <span>Not Ready</span>
<b>Ethical Considerations</b>	Addressing bias, privacy, data protection, and inclusivity.	Have you addressed potential biases, ensured fairness, and established compliance with data protection laws?	<span>Ready</span> / <span>Not Ready</span>
<b>Translation to Insights</b>	Interpreting findings, strategic recommendations, and stakeholder communication.	Can you interpret results clearly, provide actionable recommendations, and effectively communicate insights to stakeholders?	<span>Ready</span> / <span>Not Ready</span>
<b>Weighting by Effort</b>	Calculating WSJF scores (Value, Time, Risk Reduction, Effort).	Have you calculated WSJF scores to prioritise tasks effectively, focusing on the highest-value use cases?	<span>Ready</span> / <span>Not Ready</span>
<b>Implementation Planning</b>	Understanding dependencies, risks, and duration.	Have you evaluated technical risks, dependencies, and duration to ensure project feasibility within the timeline?	<span>Ready</span> / <span>Not Ready</span>

For each criterion, mark as "Ready" or "Not Ready". All should ideally be marked "Ready" before proceeding with the pilot