

Mapping the Institutional Data Landscape

A data inventory is a record of all assets that an institution defines as data. Data inventories focus on key components that aid in the management and governance of institutional data. This inventory is often structured towards a technical audience with an operational rather than descriptive focus.

Constructing a data inventory is necessary for understanding the scope and operational impact of institutional data assets. It can also be the first step in identifying the key audiences, stakeholders, and stewards that drive data culture in your institution.

Key Components of Data Inventory

- **Data identification** This involves recognizing and documenting all types of data that the organization handles. It ranges from biodemographic information, student performance, curriculum, compliance, employee records, finance data and intellectual property.
- **Data location** It's essential to document where each data set is stored. This can include various locations such as on-premises servers, cloud storage, or third-party databases.
- **Data ownership** This aspect identifies who is responsible for each data type within the organization. Assigning ownership ensures accountability and proper data management.
- **Data usage** Understanding how data is used in operations is crucial. This involves tracking the flow of data, its purpose, and how it contributes to business processes.
- **Data classification** Data is classified based on type, sensitivity, and importance. This classification helps determine the level of security and handling that each data type requires.
- **Data security and compliance** This involves documenting the security measures in place for each type of data and ensuring that these measures comply with relevant laws and regulations.

Once a data inventory is generated it can support the development of an institutional data map. By understanding the breadth of data, you can identify the right fit methodology and tools to support mapping your organizational data inventory in both technical and operational contexts.

Authors

Christen White, MS, PMP, Assistant Director, Data Services, UW School of Medicine

Institutional Data Mapping Methodology and Templates

Classification Models

1. Purpose

The purpose of Data Landscape Mapping is to establish a standardized approach to identifying, documenting, and visualizing the organization's data assets, their relationships, and flows across the enterprise. This ensures comprehensive visibility into data usage, ownership, and lineage, supporting governance, compliance, and strategic data management.

2. Scope

This method applies to all business units, systems, and stakeholders involved in the generation, storage, movement, or consumption of data within the organization. It covers structured, semi-structured, and unstructured data.

3. Methodology

3.1 Data Inventory and Classification

- Catalog all data assets, systems, applications, and repositories.
- Classify data based on sensitivity (e.g., public, confidential, sensitive).

3.2 Mapping Data Flows

- Identify data sources (internal and external).
- Trace data movement across systems including transformation, enrichment, and consumption.
- Document data integration mechanisms (e.g., APIs, ETL/ELT processes).

3.3 Metadata Collection

- Capture both business and technical metadata.
- Include data definitions, formats, lineage, and classification tags.

3.4 Roles and Responsibilities

- Assign Data Owners accountable for data content and quality.
- Identify Data Stewards responsible for maintaining metadata and resolving data issues.
- Define system and process owners involved in data processing.

3.5 Visualization Tools

- Use visualization tools to create a live and interactive map of the data landscape.
- Some off-the-shelf tools available are:
 - Microsoft Purview for data map, data catalog, data classification and data lineage
 - Collibra for Data Catalog and Governance
 - Alation for data catalog
 - Lucidchart for data flow diagrams.
 - Open source data catalog tools: Amundsen, Apache Atlas, DataHub and Metacat.
 - Glymr for data landscape mapping diagrams and consulting services

Institutional Data Mapping Methodology and Templates

Classification Models

4. Deliverables

- Data Asset Inventory
- Data Flow Diagrams
- Data Lineage Diagrams
- Ownership Matrix

5. Governance and Maintenance

- The Data Governance Office (DGO) will oversee the execution and updates of the data landscape map.
- Updates must be performed:
 - Quarterly for key systems.
 - Immediately following major changes (e.g., system implementation, migration).
- All changes are subject to review and approval by the Data Governance Council.
- Metadata management
 - Organizations should develop and maintain a Metadata Management Policy to ensure consistent practices that support data discoverability, quality, and lineage.
 - The policy defines the metadata life-cycle (capture, store, manage, and utilize across the organization).
 - Metadata should be stored in an approved enterprise metadata repository.
 - Metadata quality reviews must be conducted at least semi-annually and any discrepancies must be resolved promptly.
- Data Lineage management
 - Organizations should develop and maintain a Data Lineage Policy that establishes requirements for documenting and maintaining end-to-end data lineage across systems and processes.
 - Data Lineage supports traceability, compliance, impact analysis, and quality assurance.
 - Lineage must be visualized using standardized diagrams. Acceptable formats: automated tool outputs, annotated flowcharts, system diagrams.
 - Lineage diagrams must be reviewed quarterly or upon any major system change. Gaps in lineage documentation must be remediated with high priority.

6. Compliance and Audit Readiness

Data Landscape Mapping supports regulatory compliance efforts by:

- Ensuring traceability of personal and sensitive data
- Demonstrating control over data movement and usage
- Providing evidence of governance processes during audits

Institutional Data Mapping Methodology and Templates

Classification Models

7. Continuous Improvement

- Periodic reviews and stakeholder feedback will drive enhancements to the mapping process.
- Automation and integration with metadata management and data cataloging tools will be prioritized.

8. Resources

- DAMA-DMBOK (Data Management Body of Knowledge) Key chapters referenced: Data Governance, Metadata Management, Data Architecture, and Data Quality.
- Vendor-specific best practices for implementing metadata and lineage tracking:
 - <https://www.collibra.com>
 - <https://www.alation.com>
- Microsoft's official guidelines and use cases for data cataloging, lineage tracking, and governance:
 - <https://learn.microsoft.com/en-us/azure/purview/>
 - [Data lineage in classic Microsoft Purview Data Catalog | Microsoft Learn](#)
 - [What is Microsoft Purview? Data Driven Daily](#)
- The Open Group Architecture Framework (TOGAF): Concepts around enterprise architecture and information systems mapping align with the landscape methodology used.
- Others
 - [What is Data Governance? | Microsoft Azure](#)
 - [What is a Data Catalog? Uses, Benefits and Key Features | TechTarget](#)
 - <https://thedataecosystem.substack.com/p/issue-20-mapping-the-data-technology>
 - [Data Landscape Mapping™ from Glymr](#)

Authors

Prasad Chodavarapu, MS, Director, Data Management, The American Board of Family Medicine

References

- 1.Data Science Council of America. (2024, November 29). *What is data inventory and how to create it effectively?* Retrieved from dasca.org
- 2.AWS. (n.d.). *What is a data catalog?* Retrieved from aws.amazon.com
- 3.IBM. (n.d.). *What is a data catalog?* Retrieved from ibm.com
- 4.Herzberg, B. (2022, January 19). *Metadata repositories: Data dictionary vs. data inventory vs. data catalog*. Satori Cyber. Retrieved from Satori Cyber
- 5.Secoda. (2024, September 23). *Understanding the differences between a data inventory and a data catalog*.
.
- 6.Wikipedia. (n.d.). *Data steward*. Retrieved from Wikipedia
- 7.Knight, M. (2024, November 5). *What is data stewardship?* DATAVERSITY. Retrieved from DATAVERSITY
- 8.EDUCAUSE. *Understanding and Developing a Data-Informed Culture*. EDUCAUSE, May 2022,<https://library.educause.edu/-/media/files/library/2022/5/understandinganddevelopingdatainformedculture.pdf>

Authors

Christen White, MS, PMP, Assistant Director, Data Services, UW School of Medicine