Ostfalia
University of
Applied Sciences

Structuring Data Sovereignty and Governance in Data Ecosystems - *A Literature Review and Model Proposal*

Michael Pleger, Torben Röhrs, Felix Schmidt and Ina Schiering

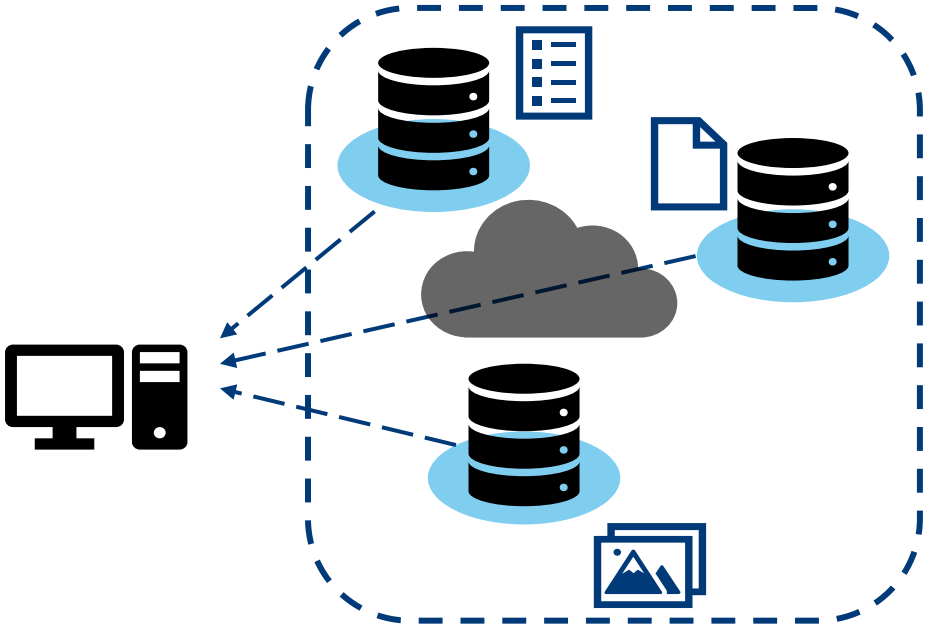
Ostfalia University of Applied Sciences
Salzdahlumer Str. 46/48
38302 Wolfenbüttel
Faculty of Computer Science

Motivation Context

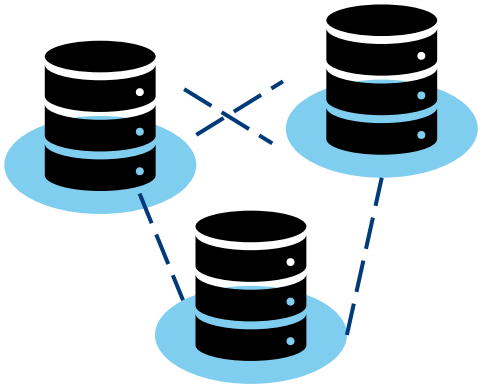
Data Ecosystems rely on shared infrastructure as their foundation for data spaces, but face structural challenges that limit their potential

Benefits

- Cross-institutional data sharing & reuse
- Decentralized storage and access
- Enable industrial collaborations
- Support public-sector digital infrastructure



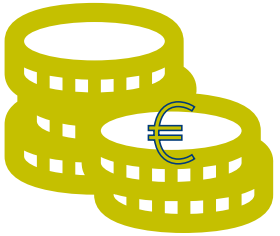
What are the facing Challenges for Data Ecosystems?



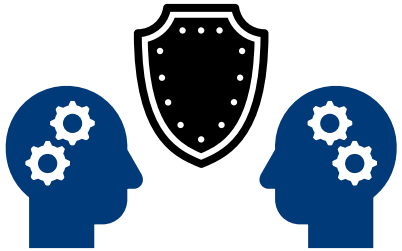
Fragmentation:
Data silos make collaboration and integration of data (sets) difficult

Interoperability:
Different formats & standards cannot be linked

Legal issues:
compliance, copyrights, terms of use



Economic value:
Data is kept exclusive



Lack of trust between stakeholders, fear of data misuse

Complexity:
Collecting data is easy but Data governance/management requires significant effort (strategy)

Technical challenges:
security, infrastructure, scalability

Illustration: Ostfalia

Who are the **Stakeholders** in a Data Ecosystem?

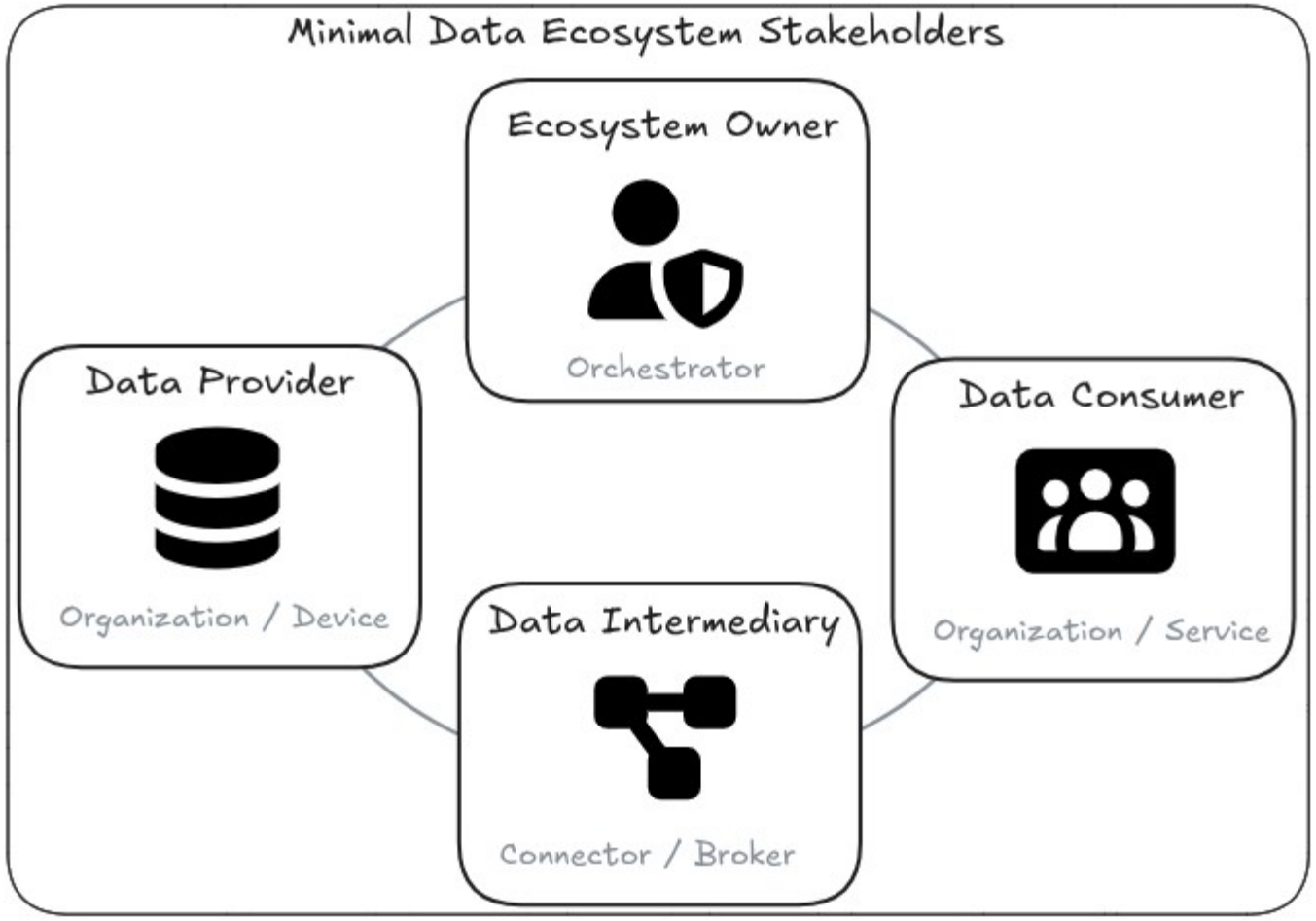


Illustration: Ostfalia

What are the key factors?

How are responsibilities **distributed**?

How can data usage be **governed**?

How can stakeholders **retain control** in shared infrastructures?

What are the goals of **Data Governance**?

> *Data governance is a system of rules, roles, processes and standards that ensures data is accurate, secure, accessible and used appropriately throughout its lifecycle.*

Transparency & Trust:
Traceability of data usage, building stakeholder confidence, accountability

Ensuring high quality data to be: accurate, complete, relevant, up-to-date

Compliance:
Meeting legal and regulatory requirements (e.g. GDPR, industry standards)

Illustration: Ostfalia

Example of **Data Governance** in Forestry

“The Landesforsten deploy bioacoustic sensors across state forests to monitor forest biodiversity.”

Data Ownership	Privacy & Compliance	Access Control	Purpose Limitation
• all audio recordings, • all derived species detections and • all biodiversity indicators are **owned by the state** and managed under public-sector data rules	• automatic voice filtering before storage • deletion of raw audio after a short retention period • GDPR-compliant processing on EU-based servers	• **Full access** Internal biodiversity and GIS teams • **Limited access** Researchers (species-level data only) • **Public access** Aggregated indicators (e.g. forest bird index), no raw audio	• biodiversity monitoring • species trend analysis • reporting to environmental authorities

What is **Data Sovereignty**?

- Data sovereignty refers to the control, access and usage of data by individuals, organizations or states
- Key legislative frameworks include:
  - EU Data Governance Act
  - Data Act
  - GDPR
- It is a **collective term** summarizing all frameworks, processes, and policies governing how data is managed and utilized within an organization

- > *Data sovereignty ensures that data remains subject to the laws and governance structures within the nation, region or organization where it is collected*

- Essential for responsible data management, compliance, and trust in digital ecosystems
- Deciding who gets to:

Illustration: Ostfalia, DGA Icon: generated by AI

Example of **Data Sovereignty** in Forestry

“A private forest owner in allows a service provider to run LiDAR scans and AI-based stress detection over their forest.”

Without Data Sovereignty

- The provider could store the data indefinitely
- Sell aggregated forest health data to insurers or timber buyers
- Use the data to train commercial AI models
- Share data with third parties without explicit consent

With Data Sovereignty

- The owner defines a contract stating:
  - data may only be used for the specific monitoring purpose
  - no resale or secondary use
  - no training of external AI models
  - deletion after a defined period
- The owner retains full rights to the raw and processed data
- The provider acts only as a technical processor

How was the structured literature reviewed?

Search strategy

- Tag based search
- Source: *Web of Science, IEEE Xplore, Springer*
- Year: 2020+

Selection

- Limited on meta studies (surveys & literature reviews)
- Ordered by citations

Screening

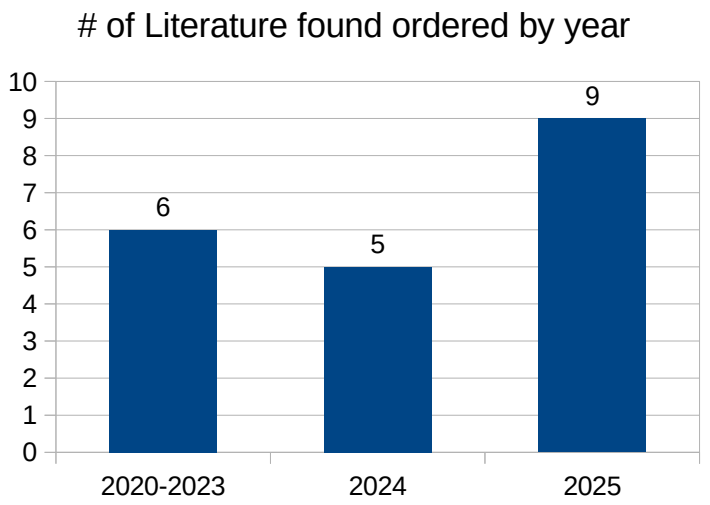
- Filtering through manual abstract & full paper screening
- Extension through snowballing

Definition alignment

- Addresses multiple stakeholders
- Data through federated / shared infrastructure

Analysis

- Power asymmetries
- Federated identity management
- Usage control
- Trust mechanisms



Findings - Platforms

Three platforms appear repeatedly for **federated sovereignty-preserving** data sharing

GAIA-X

- European federated secure data infrastructure initiative
- Provides governance mechanisms and interoperability standards
- Enables sovereignty enforcement through policy-based approaches

IDS (International Data Spaces)

- Reference architecture for sovereignty-preserving data exchange
- Defines usage control frameworks, data connectors and certified components
- Key challenge: post-transfer usage control remains technically unresolved

CEDS (Common European Data Spaces)

- EU policy framework for domain-specific data spaces (health, energy, mobility…)
- Promotes interoperability through shared standards and legal frameworks

Findings - Governance

Trust is still governed more by **legal agreements** than by verifiable technical guarantees

- Many frameworks remain conceptual - insufficient validation in real-world ecosystems
- Ambiguous responsibilities and conflicting incentives
- Misalignment between organizational and ecosystem-level objectives
- **Power asymmetries**
  - dominant organizations may impose rules disadvantaging smaller stakeholders
- Post-transfer usage control are technically unresolved
  - rely on contractual mechanisms

Findings - Sovereignty

Mechanisms for **enforcing sovereignty** across different stages of data access and use are listed

- Access Control Models
  - Role-based access control: permissions based on assigned roles
  - Attribute-based: fine-grained access by data and user attributes
  - Policy-based: rule-driven enforcement across organizational boundaries
- Usage Control
  - Regulates processing
  - Still a **open challenge** in production environments
- Monitoring & Contractual
  - Origin tracking: traceability of data sources
  - Consent management: explicit usage permissions per stakeholder
  - Contractual agreements: define rights, obligations and responsibilities

Differences in Governance and Sovereignty

Governance and Sovereignty provide complementary guidelines for data ecosystems but conflict with each other in fundamental ways

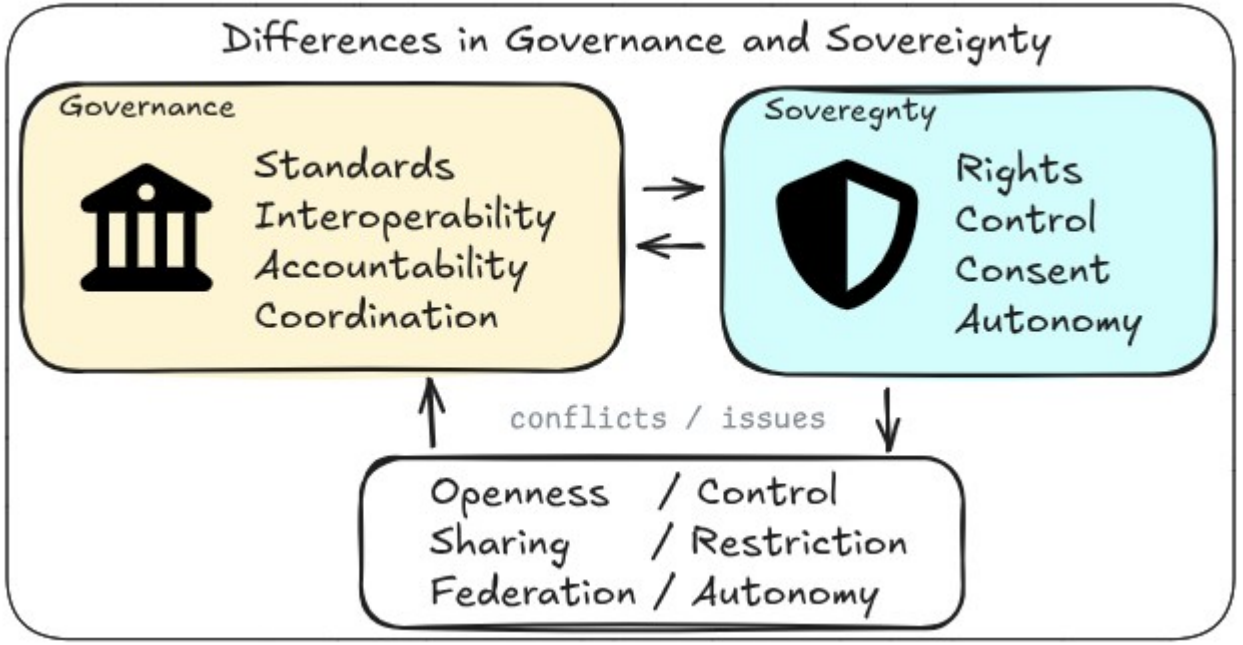


Illustration: Ostfalia

Layered **Conceptual Model**

Combination of organizational, technical and social factors

Governance Layer

- Organizational & procedural baseline
- Generates transparent, reproducible processes & task clarity for all stakeholders

Sovereignty Layer

- Social mechanisms for data control
- Ensures data usage aligns with rights of data-providing stakeholders

Ecosystem Dynamics Layer

- Interpretation & application of previous layers via technical implementations

Outcome Layer

- Representation of the system effects

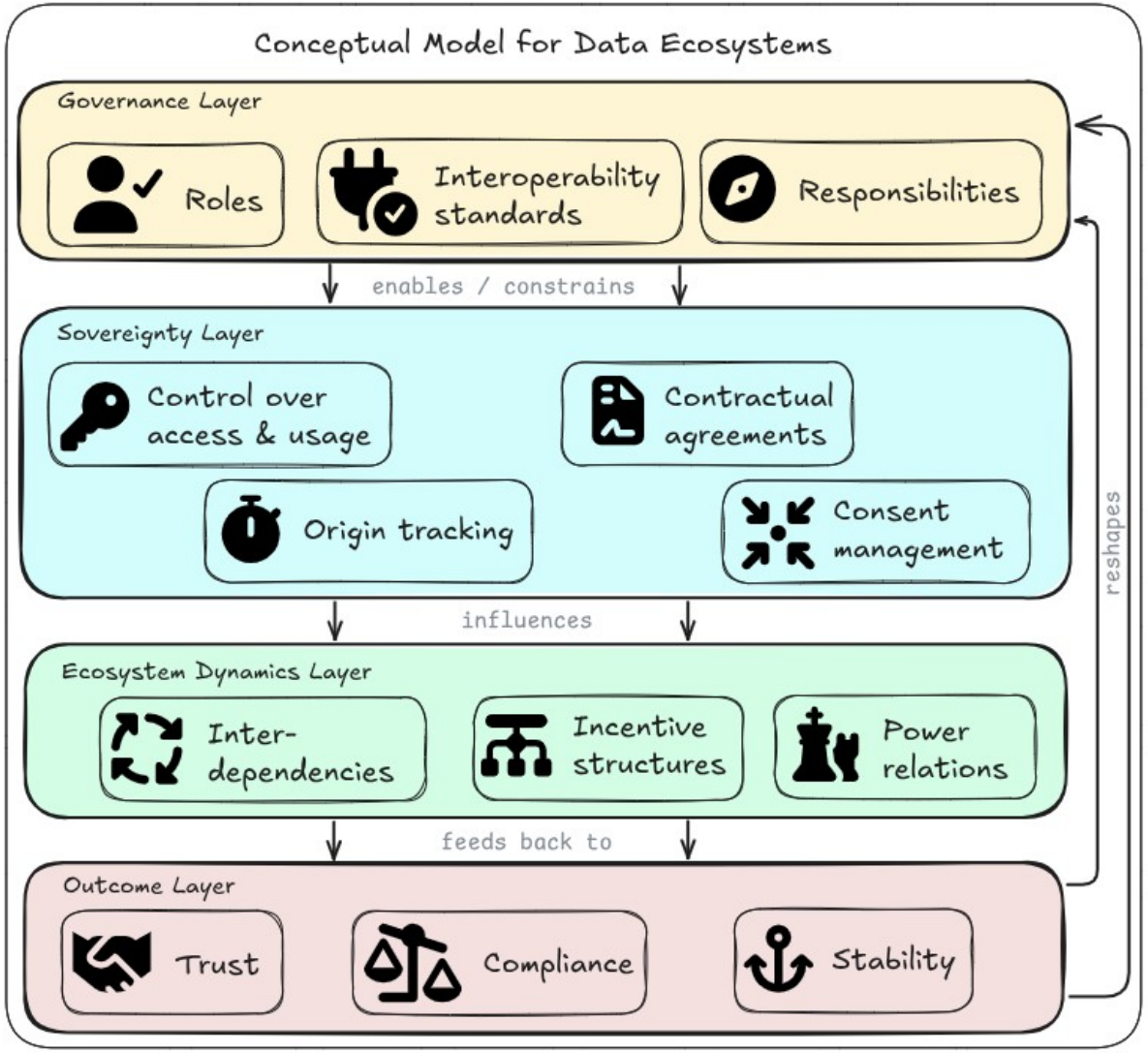


Illustration: Ostfalia

Problem **recap**

How are responsibilities **distributed**?

How can stakeholders **retain control** in shared infrastructures?

How can data usage be **governed**?

Contact:
M. Sc. Michael Pleger
mic.pleger@ostfalia.de

https://www.ostfalia.de/hochschule/fakultaeten/fakultaet-informatik/fakultaetsteam/m-sc-michael-pleger

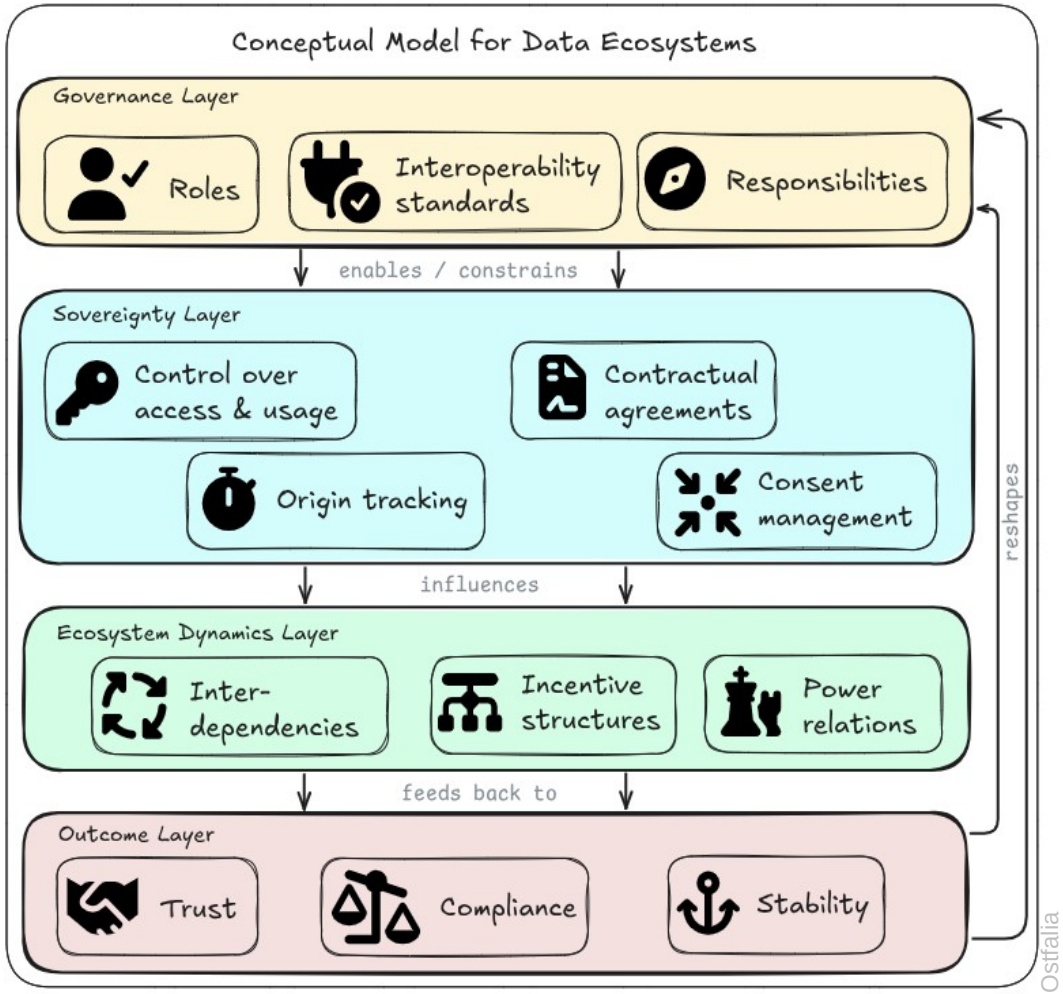


Illustration: Ostfalia