DATA GOVERNANCE INNOVATIONS

Emerging practices and trends across the data life cycle

Stefaan Verhulst, Begoña G. Otero

March 2026

GOVLAB

Table of Contents

Introduction

What this guide is.

This primer is a companion to the Data Governance Toolkit: Navigating Data in the Digital Age. That report, by the Broadband Commission Working Group on Data Governance, underscored the need for data governance frameworks that are adaptive, inclusive, and designed for the complexities of today's digital ecosystems. In light of that, it focuses on emerging governance practices that are becoming operationally important across the data life cycle. Rather than treating these practices as standalone trends, it maps them to the life cycle stages where they contribute most: planning, collecting, processing, sharing, analysing, and using data. The guide is intended as a living document; it will be revised on an as-needed basis and republished as new versions.

Practices vs. trends.

This guide distinguishes between emerging practices and structural trends. Practices are methods, tools, and governance approaches that organizations can actively adopt or embed in their operating models. Trends are broader forces that shape the environment in which governance decisions are made, but are not themselves implemented as practices. In this guide, the life cycle section focuses on practices, while the broader contextual discussion addresses the structural forces that make those practices increasingly necessary, including regulatory densification, data sovereignty pressures, and the growing enclosure of data.

Who it is for.

Data space operators, data governance leads and stewards, cross-functional programme managers, legal and policy designers, public sector officials, and public-interest data intermediaries. It is designed for readers who need a practical way to connect emerging governance developments to concrete stages of the data life cycle.

How to use it.

The guide can be read in sequence or used selectively. The opening sections explain the structural context in which new governance practices are emerging. The life cycle sections then show where particular practices matter most and what governance function they serve at each stage. A final section addresses cross-cutting practices that apply across the life cycle as a whole. Instead of a static references section, the authors encourage readers to consult our Living Library for the most up-to-date resources and examples.

SECTION I.

Why Are These Practices Important?

Emerging practices in data governance are not appearing in isolation. They are responses to a changing operating environment in which organizations must govern more data, of more types, under more demanding technical, legal, and societal conditions. Three developments are especially important.

First, data environments have become more complex. Organizations increasingly work with multimodal, real-time, and continuously generated data rather than small, static, and well-bounded datasets. This raises the governance uncertainties at every stage of the data life cycle: data must be more discoverable, better documented, easier to contextualise, and more consistently governed if it is to remain usable and trustworthy.

Second, artificial intelligence (AI) has moved from experimentation into routine operational use. As AI systems are integrated into search, classification, analysis, recommendation, and decision-support processes, data governance and AI governance increasingly overlap. This makes practices such as metadata and provenance standards, privacy-enhancing technologies, explainability mechanisms, and integrated AI-data governance more important than before. Data that is poorly documented, weakly governed, or difficult to trace is no longer simply inefficient; it becomes a direct governance risk in AI-enabled environments.

Third, governance expectations have hardened. Requirements relating to accountability, privacy, transparency, inclusion, sovereignty, and cross-border compliance are becoming more formalized across jurisdictions and sectors. At the same time, access to usable data is becoming more constrained. The growing enclosure of data—sometimes described as a "data winter"—means that governance must increasingly address not only how data is protected, but also how access, reuse, and public value can be sustained under conditions of legal, institutional, and commercial restriction.

Taken together, these changes mean that data governance can no longer be treated as a downstream compliance exercise. It must be built into the data life cycle, from planning and collection through to sharing, analysis, and use. The practices mapped in the section that follows should be read in that light: not as abstract innovations, but as operational responses to a more demanding governance landscape.

SECTION II.

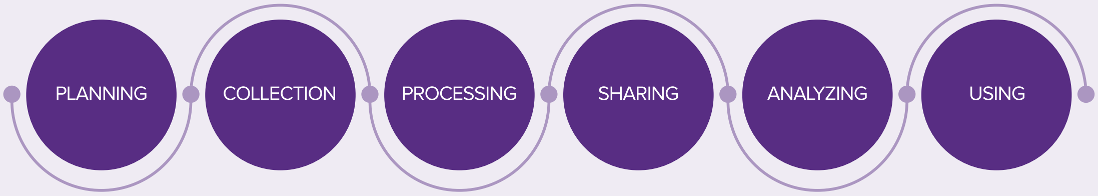
How To Read This Guide: The Data Life Cycle Approach

The emerging practices discussed in this guide are not separate silos. They operate across a common data life cycle: planning, collecting, processing, sharing, analysing, and using data. The purpose of the guide is therefore not to treat each practice as a standalone trend, but to show where in the life cycle it makes its strongest governance contribution.

DATA LIFE CYCLE STAGES

The data life cycle refers to the various stages data goes through—from its initial planning to its ultimate use in decision-making.

While different frameworks may highlight different stages or use varying terminology, the most commonly recognized phases include:



- **Planning –** Identifying data needs, intended uses, and governance requirements.
- **Collection –** Gathering data through surveys, sensors, transactions, or other means.
- **Processing –** Organizing data for use.
- **Sharing –** Making data accessible to others for re-use, whether through platforms, APIs, or data collaboratives.
- **Analyzing –** Interpreting the data to generate insights.
- **Using –** Applying those insights to inform decisions, policies, or services.

At each stage, data governance decisions—such as who has access, how data quality is maintained, and how privacy is protected—are made. These decisions are cumulative and can significantly shape what is possible in later stages. Poor governance early on (e.g., unclear purpose or unstructured collection) can generate negative consequences, or limit the value or usability of data downstream.

The section is organized by life cycle stage. Each stage identifies the emerging practices that are especially relevant there and explains the governance function they perform. Some practices appear more than once because their relevance spans multiple stages. Federated learning, for example, is relevant both when data is collected and when it is analysed; social licence and integrated AI-data governance apply across the full data life cycle.

This life cycle-first approach is intended to make emerging governance practices easier to operationalize. Rather than asking which topic is most fashionable, it asks a more practical question: what governance challenges arise at each stage of the data life cycle, and which emerging practices are most useful in addressing them?

1. PLANNING

- **AI for Systematic Review:** Automated tools to evaluate existing knowledge, reducing redundancy and improving topic mapping and problem definition, ensuring better traceability of data sources.
- **Semantic search:** AI-supported search methods that improve discovery of relevant data, documents, and related sources.
- **Knowledge Graphs and Data Discovery:** Advanced systems can be used to interlink data and metadata for enhanced discovery and contextualization.
- **Data Discovery and Tiering/Scouting:** Automated tiering systems classify data based on relevance and usability (and sensitivity).
- **Agentic AI for discovery horizon scanning and governance mapping:** AI agents that support multi-step discovery, monitoring, and mapping of data assets, sources, and governance gaps across organizational portfolios.

GOVERNANCE CONTRIBUTION

At the planning stage, these practices improve visibility over existing data, reduce unnecessary new collection, and strengthen early-stage decisions about scope, readiness, stewardship, and reuse. Their main governance contribution is to make upstream design more evidence-based and less ad hoc. They also help organizations distinguish between what data is available, what data is usable, and what data can responsibly support the intended purpose. Here, strategic data stewardship plays an important role.

The main governance risk at this stage is over-reliance on AI-assisted outputs without adequate verification, provenance checks, or human judgment.

2. COLLECTING

- **Decentralized Personal Data Management Systems:** Approaches that strengthen individual control over personal data and support more distributed stewardship models.
- **Support for Diverse Data Formats:** Increase in unstructured formats (e.g., images, videos, texts) that can be used for LLMS.
- **Synthetic Data:** Artificially generated data that can expand datasets and enable analysis while reducing exposure of real-world personal or sensitive data.
- **New Metadata and Data Provenance Standards:** Emerging standards that improve traceability, reliability, and documentation of data, particularly where data is used in AI systems.
- **PETs:** Technologies such as differential privacy, homomorphic encryption, and secure multi-party computation that reduce privacy risks during data collection and early-stage handling.
- **Social License:** Moving beyond individual consent to societal approval, addressing collective concerns.
- **Model Cards and Data Sheets:** Documentation tools that clarify intended use, limitations, quality considerations, and ethical implications of datasets and models.

GOVERNANCE CONTRIBUTION

At the collecting stage, these practices strengthen traceability, documentation, privacy protection, and stewardship from the outset. Their governance contribution lies in making the collection more purposeful, better documented, and more compatible with later reuse, analysis, and accountability requirements. They also help ensure that data entering the life cycle is accompanied by sufficient contextual information to support lawful, explainable, and proportionate downstream use.

The main governance risk at this stage is to treat collection as a purely technical act, when in practice it establishes many of the constraints and possibilities that govern the rest of the data life cycle.

3. PROCESSING

- **Decentralized Storage Networks:** (Blockchain-based) systems decentralize data storage, to enhance resilience and security.
- **Data Mesh:** Decentralized data architecture to promote team autonomy and scalability.
- **Edge Computing:** Enables local data processing, reducing latency and ensuring real-time decision-making.
- **Data Products:** Pre-prepared, reusable, and modular datasets designed for specific use cases to streamline analysis and decision-making.
- **PETs in Processing:** Technologies like federated learning and secure multi-party computation to ensure secure collaborative processing.
- **Policy-as-code:** Machine-readable governance rules embedded into data pipelines and infrastructures to enable more consistent and automated policy enforcement.

GOVERNANCE CONTRIBUTION

At the processing stage, these practices enable more governable transformations, local control, secure collaboration, and automated rule enforcement. Their main contribution is to ensure that processing decisions remain visible, attributable, and aligned with the original purpose and governance conditions under which data entered the life cycle. They also help organizations move from informal or manual control to more systematic and enforceable governance arrangements.

The main governance risk at this stage is that technical optimisation outpaces governance visibility, leaving important transformations insufficiently documented, justified, or reviewable.

4. SHARING

- **Data Collaboratives/Trusts/Cooperatives/Commons:** Governance and operational models for data access and exchange while safeguarding individual and organizational interests.
- **Data Spaces:** Federated platforms for more secure, automated, and trusted data exchange across entities.
- **Data Sandboxes:** Controlled environments for testing data access, sharing, and reuse with lower regulatory and operational risk.
- **New Licensing Regimes:** Dynamic and use-case-specific licensing models that clarify rights, responsibilities, and conditions of reuse.
- **Advances in Identification Mechanisms:** Innovations in authentication, such as decentralized identifiers (DIDs) and biometric systems, provide more granular access control to data.
- **Zero-Knowledge Proofs:** Verifies data properties without exposing underlying data.
- **Data clean rooms:** Controlled environments that allow joint analysis across datasets while restricting direct access to raw or identifying data.

GOVERNANCE CONTRIBUTION

At the sharing stage, these practices support trusted exchange, access control, interoperability, and controlled experimentation without requiring unrestricted disclosure of data. Their main governance contribution is to make sharing more structured, more accountable, and more adaptable to different legal, institutional, and risk contexts. They also help organizations move beyond ad hoc bilateral agreements toward more repeatable and governable forms of exchange.

The main governance risk at this stage is to focus too narrowly on technical controls while underestimating the importance of institutional design, clear roles, and enforceable participation conditions.

5. ANALYZING

- **Federated Learning:** A decentralized machine-learning approach in which multiple parties train a shared model without pooling raw data; each party keeps data local and shares only model updates (e.g., gradients/weights) that are aggregated to form a global model.
- **Large Language Models (LLMs):** LLMs to enhance data analysis by uncovering patterns, summarizing findings, and generating natural language explanations for complex data.
- **Conversational data interfaces:** Natural-language interfaces that allow users to query and explore datasets more accessibly.
- **Agentic AI for multi-step analysis and orchestration:** AI agents that can coordinate analytical tasks, combine tools, and support more complex interpretive or workflow-based analysis.
- **Explainability and lineage mechanisms for AI-supported analysis:** Emerging practices that document analytical pathways, support interpretability, and strengthen accountability where AI shapes outputs or recommendations.

GOVERNANCE CONTRIBUTION

At the analyzing stage, these practices improve accessibility, expand analytical capacity, and support more collaborative forms of interpretation. Their main governance contribution is to make AI-supported analysis more reviewable, transparent, and accountable, especially when findings may influence consequential decisions. They also help distinguish between analysis as retrieval or summarization and analysis as judgment or decision support.

The main governance risk at this stage is that increasingly capable analytical systems may produce outputs that appear authoritative without making their limitations, assumptions, or pathways sufficiently visible.

6. USING

- **Digital Twins:** Virtual replicas of physical systems to allow predictive modeling and scenario testing.
- **Simulations:** Advanced simulations to provide insights into potential outcomes and decision impacts.
- **Conversational Data Interfaces (Chatbots):** AI-driven chatbots to enable users to interact conversationally with datasets, making data more accessible and user-friendly.
- **Social license processes:** Practices that move beyond individual consent to address broader societal expectations, legitimacy, and collective concerns around data use.
- **Benefit-sharing mechanisms:** Approaches that seek to distribute the value generated from data use more fairly among affected actors or communities.
- **Integrated AI-data governance:** Governance approaches that connect data quality, data rights, model oversight, deployment controls, and monitoring within actual operational decision environments.

GOVERNANCE CONTRIBUTION

At the using stage, these practices support decision-making, legitimacy, benefit distribution, and closer alignment between data governance and AI system governance. Their main governance contribution is to ensure that outputs are not only operationally useful but also contestable, explainable, and socially legitimate in practice. They also help organizations move from a narrow focus on access and analysis toward the broader question of whether data use produces accountable and equitable outcomes.

The main governance risk at this stage is that technically sophisticated outputs may be used too confidently, too opaquely, or too far beyond the purposes and conditions under which the underlying data was originally governed.

SECTION III.

Cross-Cutting Practices

Some emerging practices are best understood not as belonging to a single stage of the data life cycle, but also as shaping how data governance is designed and exercised across the data life cycle. They provide coherence across planning, collecting, processing, sharing, analyzing, and using data, and are most effective when embedded **throughout** rather than treated as separate workstreams.

Social license and ethical innovation

Social license is the practice of building and maintaining collective, contextual legitimacy for a data initiative beyond formal consent or legal compliance. Its significance is cross-cutting because legitimacy is shaped not only by how data is ultimately used but also by how decisions are made, who is involved, whether benefits are distributed fairly, and whether concerns can be heard and addressed throughout the data life cycle. In this sense, social license gives operational form to broader governance commitments to equity, participation, and digital self-determination. In contexts involving Indigenous or community-governed data, this also requires attention to collective rights and responsibilities that cannot be reduced to individual consent alone.

In practice, social license requires more than consultation. It requires governance structures that are accessible to affected communities, visible responsiveness when concerns are raised, and mechanisms for adaptation or course correction when a project becomes contested. It also has an important equity dimension. Governance practices that demand high levels of technical, legal, or institutional capacity may be workable for large, well-resourced organizations but exclusionary for others. A socially legitimate governance framework must therefore be designed not only for rigor, but also for accessibility and proportionality.

Integrated AI-data governance

Integrated AI-data governance is the practice of treating AI governance and data governance as a single connected framework rather than as parallel domains. This has become increasingly necessary because questions of data quality, provenance, documentation, consent, bias, lineage, explainability, and oversight now arise across the full chain from training data and model development to deployment, monitoring, and downstream use. Organizations that maintain separate AI ethics policies and data governance frameworks often create accountability gaps precisely where the two converge most strongly.

In practice, integration means at least three things. First, data quality, documentation, and bias-related controls must apply to training, fine-tuning, and operational datasets used in AI systems. Second, data life cycle governance concepts such as provenance, lineage, stewardship, and purpose limitation must extend to model development, deployment, updating, and decommissioning. Third, agentic and automated AI systems should be treated as governed actors within the data environment, with explicit authorization boundaries, oversight requirements, and review mechanisms. Integrated governance is therefore not simply a matter of compliance; it is a design choice that reduces fragmentation across the broader data and AI stack.

Digital public infrastructure alignment

Digital public infrastructure alignment is the practice of designing and auditing data governance arrangements to ensure they are compatible with, and, where appropriate, contribute to, shared digital systems such as identity, payments, and data exchange layers. Its relevance is cross-cutting because governance choices made in these contexts do not remain local to a single dataset or project. They can become default conditions for access, participation, and accountability at the population scale.

In practice, this means that governance arrangements connected to digital public infrastructure must support interoperability, privacy, accountability, inclusion, and secure processing at a systemic rather than purely project level. As data spaces, federated learning, and decentralized identifiers are increasingly incorporated into broader public and quasi-public digital systems, governance choices around access, documentation, oversight, and redress take on infrastructural significance. The standard applied in these contexts must therefore be higher, because the consequences of poor design are broader, harder to reverse, and more widely distributed.

SECTION IV.

Structural Forces Shaping Practice

The practices described in this guide do not emerge in a vacuum. They are responses to broader structural forces that shape the conditions under which data governance now operates. These forces are not themselves practices: organizations do not implement regulatory convergence, data sovereignty, or a data winter. They navigate them. The purpose of this section is therefore not to add another layer of operational detail, but to clarify the environment to which the practices in this guide are responding.

AI deployment and the mediated governance environment

AI is increasingly embedded in routine organizational processes, public administration, service delivery and data-intensive decision environments. This shift changes the governance environment in at least two ways. First, it reinforces a broader systems approach to governing with AI: trustworthy AI depends not only on models, but on the full set of enablers, guardrails, oversight arrangements, and data conditions that make AI deployment governable in practice. AI raises the baseline requirements for data quality, provenance, documentation, interoperability and reuse. Data that is poorly described, weakly governed or difficult to trace cannot reliably support the training, validation, deployment or monitoring of AI systems. Thus, AI does not simply increase demand for data; it increases demand for governable data.

Second, AI is no longer used as a tool for prediction or summarization. The emergence of agentic AI landscape introduces systems that can coordinate multi-step tasks, retrieve and combine information from multiple sources, call tools, sequence actions and generate outputs with reduced human intervention and oversight. This expands the governance challenge from datasets and models to workflows, permissions, oversight boundaries and non-human action within the data environment. Governance must then cover both the quality and legality of the underlying data and the authority, scope, traceability and explainability of AI-enabled processes acting upon that data.

Thus, AI amplifies the consequences of weak data governance. Poor metadata, unclear provenance, fragmented stewardship or inadequate oversight are no longer only efficiency problems. In AI-enabled environments they become direct sources of opacity, bias, error propagation, overreach and reduced accountability. This is why emerging practices such as metadata and provenance standards, explainability and lineage mechanisms, policy-as-code, agentic oversight, and integrated AI-data governance now matter operationally rather than aspirationally.

Regulatory convergence and enforcement

Data governance and AI governance are increasingly shaped by binding legal, regulatory, and institutional frameworks rather than by soft law or voluntary principles alone. This does not take a single form across jurisdictions, but the broader direction is clear: expectations around documentation, accountability, transparency, and oversight are becoming more formalized and more enforceable. In practical terms, this creates a governance floor below which organizations cannot safely operate. The significance of this force is not limited to compliance. It changes what counts as adequate governance design in the first place, especially where data and AI systems interact.

Data sovereignty and cross-border governance

Data sovereignty has moved from a largely geopolitical idea to an operational governance constraint. Questions of where data originates, where it is stored, which legal regime applies, and who has authority over access and transfer now shape governance decisions across the full data life cycle. This is especially important in cross-border environments, where data flows may be technically feasible but legally or politically constrained. The practical implication is that governance frameworks increasingly need to record jurisdictional origin, clarify legal bases for transfer and access, and design for varying sovereignty requirements across sectors, territories, and communities.

Data winter and the reuse gap

A further force is the growing enclosure of data, described as a “data winter”: a condition in which data becomes simultaneously more valuable and less accessible for public-interest use, science, and governance. This force is driven not by a single cause but by a combination of institutional risk aversion, retrenchment in open data commitments, growing sensitivity around training data, geopolitical restrictions, and the closure of once-accessible private and research datasets. At the same time, the problem is not only access but also reuse. The reuse gap describes the widening distance between making data available and creating the stewardship, incentives, and operational capacity needed to turn it into meaningful public value. A governance framework designed only for data abundance will underperform under these conditions.

Taken together, these forces explain why emerging governance practices now matter operationally rather than aspirationally. They also reinforce a central principle of the Toolkit: practices should be selected in response to purpose and context, not adopted simply because they are new. Regulatory densification, sovereignty pressures, and the enclosure of reusable data do not point to a single governance model. They do, however, make clear that governance frameworks must now be more adaptive, more explicit, and better able to function under conditions of legal pluralism, constrained access, and uneven institutional capacity.

SECTION V.

Glossary

A. Cross-cutting and foundational concepts

Data Governance: The set of processes, people, policies, practices, and technology that seek to govern the data life cycle toward meeting the purpose of increasing trust, value and equity, while minimizing risk and harm in alignment with a set of core principles.

AI-Ready Data: Data that is prepared, structured, documented, and governed in ways that support its effective and responsible use in artificial intelligence applications.

Social License for Data Reuse: Societal approval or legitimacy for a data reuse practice, established through participatory processes that go beyond formal consent and legal compliance to include ethical, cultural, and community expectations.

Data Life Cycle: The full journey of data from planning and collection to processing, sharing, analysis, and use.

Integrated AI-Data Governance: A governance approach that treats data governance and AI governance as a connected framework, linking data quality, provenance, bias, documentation, oversight, deployment, and monitoring across the full life cycle.

Data Sovereignty: The capacity of the relevant actor - whether states, communities, organizations, or individuals- to exercise meaningful control and governance authority over data, infrastructure, and data conditions of access, use, and sharing according to their own legal, political, or governance frameworks.

Data Winter: An emerging term describing a period in which data that could support public-interest use, research, or innovation becomes increasingly inaccessible, restricted, or immobilized despite rising demand for data-driven systems.

Digital Public Infrastructure (DPI): Shared, secure, and interoperable digital systems and governance arrangements that enable delivery of public and private services at societal scale, often including digital identity, payment, and data exchange layers.

B. Planning stage

Systematic Review (AI-Enabled): The use of AI and automation to support the identification, screening, synthesis, and mapping of existing knowledge, datasets, and evidence relevant to a policy, governance, or research question, used wihtin a structured review methodology.

Semantic Search: Search methods that use meaning, context, and relationships rather than simple keyword matching to improve discovery of relevant data and documents.

Knowledge Graphs: Structured networks that link entities, concepts, datasets, and their relationships, enhancing contextualization, discovery, and interoperability.

Automated Data Tiering: The rule-based or AI-assisted classification of data according to factors such as sensitivity, criticality, quality, access needs, relevance, or anticipated use, in order to support life cycle governance.

Agentic AI for Discovery, Horizon Scanning, and Governance Mapping: AI agents that support multi-step discovery, monitoring, and mapping of data assets, sources, and governance gaps across organizational portfolios.

C. Collection stage

Decentralized Personal Data Management: Systems that enable individuals to control access to and sharing of their own data, often using decentralized identity or storage models.

Synthetic Data: Artificially generated data that reproduces relevant features of real-world data while reducing direct exposure of personal or sensitive records.

Metadata Standards: Formal schemas, vocabularies, or specifications used to describe data consistently, including attributes such as provenance, format, purpose, quality, and access conditions, in order to support traceability and interoperability.

Provenance: Information about the origin, context, method of collection, and transformation history of a dataset.

Privacy-Enhancing Technologies (PETs): Technical methods that reduce privacy risks in data collection, processing, analysis, or sharing, including approaches such as differential privacy, homomorphic encryption, secure multi-party computation, and federated techniques.

Model Cards / Data Sheets: Structured documentation artifacts describing datasets or models, including intended use, limitations, ethical considerations, and relevant performance or quality information.

D. Processing stage

Decentralized Storage Networks: Distributed storage infrastructures in which data is stored across multiple nodes rather than a single centralized repository, sometimes using blockchain-linked coordination mechanisms to enhance resilience, redundancy, or integrity.

Data Mesh: A decentralized data architecture in which domain teams manage their own data as a product using shared standards and interoperable interfaces.

Edge Computing: Localized data processing at or near the point of generation in order to reduce latency, improve responsiveness, and support privacy or residency requirements.

Data Products: Curated, reusable, and documented datasets or data assets designed for specific analytical or operational use cases.

Federated Learning: A decentralized machine-learning approach in which a shared model is trained across multiple devices or organizations holding local data, without centralizing the underlying raw data; privacy protection may be strengthened further through complementary techniques such as secure aggregation or differential privacy.

Policy-as-Code: The formalization of governance and compliance rules in machine-readable form so they can be versioned, tested, and automatically enforced across data pipelines, infrastructures, services, and workflows.

E. Sharing stage

Data Collaboratives: Cross-sector partnerships in which participants exchange data to address public or shared problems while balancing privacy, trust, and mutual benefit.

Data Trusts: Legal or institutional arrangements for stewarding data on behalf of specified beneficiaries or purposes under defined governance, access, and accountability rules.

Data Cooperatives: Member-owned organizations that collectively manage and share data for the benefit of their members.

Data Commons: Shared governance models for data access and stewardship, often inspired by Elinor Ostrom's commons-based governance theory.

Data Spaces: Federated and interoperable environments that enable trusted data sharing and reuse across organizations, sectors, or borders under shared governance rules, standards, and services.

Data Sandboxes: Controlled environments in which data can be accessed, tested, or reused under specified conditions, often for innovation, experimentation, or regulatory learning.

Dynamic Licensing Regimes: Context-sensitive and, increasingly, machine-readable usage arrangements that specify permissions, obligations, and conditions for access, reuse, or onward sharing.

Decentralized Identifiers (DIDs): A type of identifier designed to enable verifiable, decentralized digital identity, controlled by the relevant DID controller and decoupled from centralized registries, identity providers, or certificate authorities.

Zero-Knowledge Proofs: Cryptographic methods that allow a verifier to confirm the truth of a statement without accessing the underlying data itself.

Data Clean Rooms: Controlled environments that allow joint analysis across datasets while restricting direct access to raw or identifying data.

F. Analysis stage

Large Language Models (LLMs): Large-scale AI models trained primarily on extensive text data (and sometimes multimodal data) to model and generate language, which can be applied to generation and analysis tasks—such as summarization, extraction, classification, and explanation—producing natural-language and structured outputs that support data interpretation and accessibility. Although commonly deployed as generative systems, they are also frequently used in non-generative analytical workflows.

Conversational Data Interfaces: Natural-language interfaces that allow users to query, explore, and interpret datasets through dialogue rather than specialized query languages.

Agentic AI for Multi-Step Analysis and Orchestration: AI agents that coordinate analytical tasks, combine tools, and support more complex interpretive or workflow-based analysis across multiple steps.

Explainability: The extent to which the workings, outputs, or implications of an AI-supported system can be understood and interpreted by relevant users or overseers.

Lineage: A documented record of how data, models, or analytical outputs were created, transformed, and used over time.

Explainability and Lineage Mechanisms for AI-Supported Analysis: Practices that document analytical pathways, support interpretability, and strengthen accountability where AI systems shape outputs, recommendations, or decisions.

G. Use stage

Digital Twins: Virtual representations of physical systems that use real-time or regularly updated data to simulate, monitor, and predict behavior or outcomes.

Simulations: Computational models that represent complex systems or processes, used to test interventions, evaluate alternatives, or explore scenarios.

Social License Processes: Participatory and trust-building governance practices that seek to establish legitimacy for data use beyond formal consent and legal compliance by addressing broader societal expectations and collective concerns.

Benefit-Sharing Mechanisms: Approaches that seek to distribute the value generated through data use more fairly among affected individuals, communities, or contributing actors.

H. Ethical and governance frameworks

Ethical Innovation: Approaches to technology and data use that proactively consider fairness, rights, inclusion, and long-term societal impacts.

Participatory Governance: Mechanisms that involve communities and stakeholders in the design, oversight, and evaluation of data governance practices.

Rights-Based Data Governance: Frameworks that prioritize human rights, such as privacy, autonomy, equality, and due process, in the collection, use, and sharing of data.

CARE Principles: Principles emphasizing Collective Benefit, Authority to Control, Responsibility, and Ethics, especially in relation to Indigenous and community-governed data.

FAIR Principles: Principles emphasizing that data should be Findable, Accessible, Interoperable, and Reusable.

GOVLAB

DATA GOVERNANCE INNOVATIONS

Emerging practices and trends across the data life cycle