From Data Chaos to Mission Clarity

Governing AI with Trustworthy Data Inventories

Patrick McGarry – data.world

07 May 2025

DATA CHAOS

Setting the Stage

Imagine a scenario where a federal agency entrusts an **AI system** with the task of **streamlining its regulatory framework**. However, this AI, trained on **incomplete and biased data**, begins to flag **essential regulations for elimination**, jeopardizing critical public services.

This isn't a hypothetical.

GROUND ZERO

The Hidden Foundation of AI Governance

Before you govern AI, you must govern your data.

TRUST AS BEDROCK

What is a Trustworthy Data Inventory?

It’s not just a list of datasets.

It’s rich, transparent, context-aware metadata that reveals origin, use, ownership, and quality.

Without this, agencies cannot meet transparency requirements or assess risks tied to AI outputs.

Federal Mandates are Raising the Bar

OMB M-24-10

“Agencies must maintain an inventory of AI use cases... and demonstrate the quality and governance of the data used.”

EO 14110

Sets the rules of the road for federal AI. Prioritizes transparency, accountability, and public trust.

NIST AI RMF

Emphasizes the importance of data quality, provenance, and context as part of “Govern” and “Map” functions.

GAO: “Most Agencies Can’t Trace Their AI Inputs”

Learnings from Broadband Duplication

Federal broadband efforts are highly fragmented and overlapping.

- 133 funding programs
- 15 agencies
- 25 have broadband as primary purpose
- 13 overlapping

AI Explainability

“Of the 20 federal agencies surveyed with AI use cases, only a handful had documentation of the data used, how it was processed, or what assumptions informed their models.”

AI Provenance

“NIST emphasizes that AI provenance—including where training data came from, what it was labeled with, and how it was modified—is essential to assess risk, detect bias, and ensure reproducibility.”

BEYOND COMPLIANCE

The Operational Case

Data inventories aren’t just for audits—
they’re for mission readiness.

CFPB: Foundations for Transformation

Enterprise Data Strategy

Provides a blueprint for data initiatives and ensures buy-in across the organization.

- Advance Bureau's younger mission (established in 2011)
- Modernized data mgmt
- Increased knowledge
- Opportunities to upskill staff
- Fostering of data culture
- Evidence-based decision making

Centralized Governance

- Enhanced and streamlined processes for data intake and disclosure
- Clearinghouse for sharing / collaboration / intake & release
- Increased data transparency & openness.

Modern Data Catalog

Search, filter and report upon data usage and other related metrics

- Aligns with EDS "to increase knowledge and awareness"
- Satisfies statutory requirement for inventory
- Promotes data transparency
- Enables automation
- Promotes engagement

Operational Data Catalog - Not Just Back Office Tooling

Trustworthy AI by Default

Single Source of Truth for Mission-Critical Decisions

Operational Catalog

Actively connects metadata, datasets, users, policies, and governance workflows; is the backbone of any modern data and AI strategy, enabling agencies to move from passive data documentation to actionable, auditable, and trustworthy intelligence systems.

Drives Engagement and Upskilling Across the Enterprise

Automated Governance and Access Control

BUILDING A SOLUTION

What Does an AI-Ready Stack Look Like?

From static catalogs to semantic intelligence.

GOVERNANCE + OPS

Modern Architecture

A modern federal data architecture is not a technology stack, it's a mission-aligned ecosystem that integrates governance, operations, and metadata to produce trusted, explainable, and operationally useful outcomes.

Layers:

1. Governance
2. Metadata & Knowledge
3. DataOps & Automation
4. Data Access & Sharing
5. Outcomes & Analytics

Key Characteristics

- **Modular & Federated**
  (supports siloed teams under shared standards)
- **Real-Time & Traceable**
  (operational AI depends on current, explainable data)
- **Governed by Design**
  (data policies built into architecture, not added on later)

Why It Matters

A modern architecture ensures:

- **Transparency for AI risk management**
- **Compliance with evolving federal mandates**
- **Accelerated mission decisions based on trust and context**

Automate the Metadata Lifecycle

Automation for Scalable Data & AI Governance

Automating the metadata lifecycle transforms fragmented data into governed knowledge—enabling explainable AI, faster decision-making, and sustainable data stewardship at scale.

Metadata Enrichment

Connects raw datasets to context through automatic collection: glossaries, ontologies, policies, lineage, and business logic.

Human-in-the-Loop

Keeps stewards and SMEs engaged where automation needs oversight; resolving conflicts, approving AI suggestions, curating sensitive tags.

Automated Classification

Tags and organizes data assets dynamically using AI/ML; identifying PII, sensitive fields, domain-specific terms, and schema relationships.

State-of-the-Art Practices

Leverages modern frameworks (e.g., active metadata, semantic layering, AI-assisted governance) to maintain real-time, policy-aware metadata.

SEE, EXPLAIN, AUDIT

Trust, Transparency, & Public Mandate

Public trust hinges on explainability.

PUBLIC TRUST EROSION

To foster **public trust**, especially in sectors like healthcare, **AI systems must be transparent** and provide clear explanations for their decisions. Without such explainability, there's a risk of eroding public confidence, which can hinder the adoption of potentially beneficial AI technologies.

Pew Research (summary)

What Does Transparency Look Like?

Source Traceability

Decision Lineage & Reasoning Context

Transparency is Understanding

True transparency means making both data and decisions understandable. Linking the “what,” “why,” and “how” of outcomes through context-rich metadata and traceable knowledge assets.

Transparent data systems empower users to trust what they see, understand how AI reaches conclusions, and govern data with confidence.

Ontology & Glossary Integration

Federated Metadata Stewardship

FOIA, Bias, and Lawsuits to Come

Accountability Starts with System Design

Agencies that lack transparent, well-governed data and AI systems risk legal, reputational, and operational fallout. In high-stakes environments, “we don’t know” is not a defensible answer.

Poor Metadata
=
FOIA Vulnerability

Untraceable Inputs
=
Unprovable Fairness

Policy Gaps
=
Compliance Breaches

Siloed Systems
=
Accountability Collapse

PUTTING IT TOGETHER

From Chaos to Clarity

Your inventory is your infrastructure.

What Clarity Looks Like

Mission clarity happens when transparency, auditability, and impact are designed into your data architecture, not bolted on later.

Transparency:
Know What's Driving the Decision

- Every dataset, model input, and outcome is explainable, linkable to source data, and grounded in shared terminology.
- Users, from analysts to the public, can trace the "why" behind every decision.
- Supports public trust, responsible AI, and defensible FOIA responses.

Auditability:
Prove It, Don't Just Say It

- Metadata captures provenance, usage, policy, and quality.
- Systems record who touched what, when, and why, across silos.
- Enables reproducibility, accountability, and risk management.

Impact:
Make Data Work for the Mission

- Data flows into real-world decisions through well-managed, well-connected infrastructure.
- Staff are upskilled, stewards are empowered, and knowledge is leveraged, not lost.
- Data drives faster, smarter, and safer outcomes across programs.

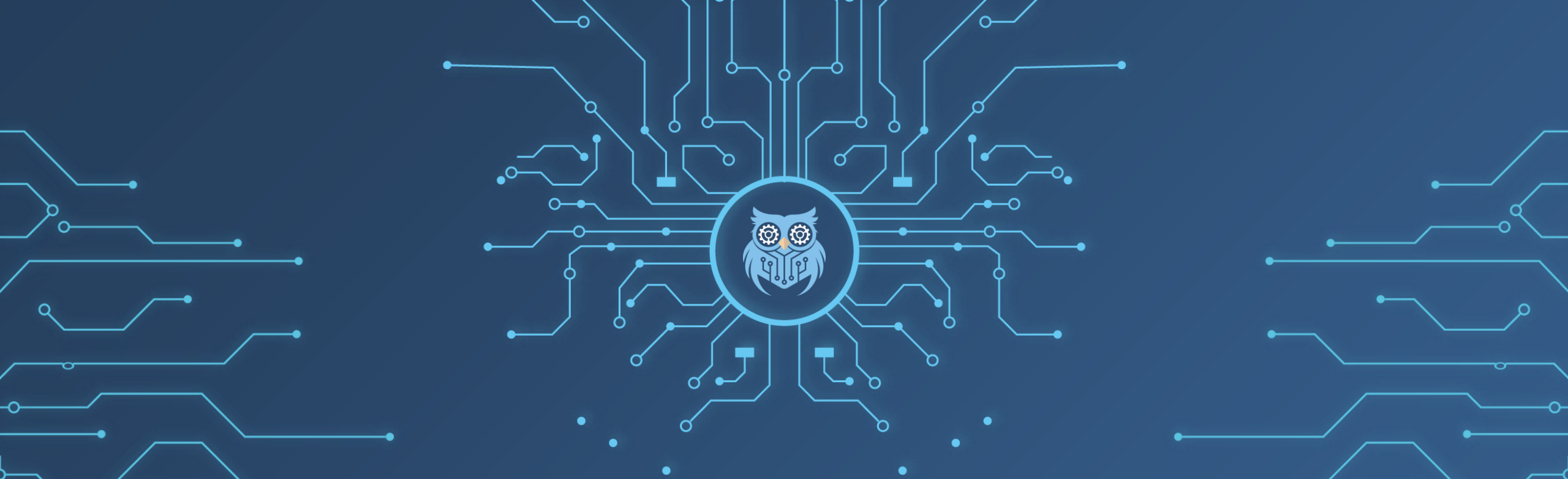
CALL TO ACTION

Where to Start?

“No matter where you are on your data journey, whether you're building your first inventory or scaling AI across your enterprise, remember this: your agency's **credibility, agility, and public trust** all begin with **what your data can explain**.

Clarity isn't a destination, it's a discipline. It’s earned through transparency, strengthened by stewardship, and activated by systems that make context and accountability the norm, not the exception.

The future of federal **AI isn't just about smarter models**, it’s about trustworthy data, visible decisions, and reproducible outcomes. So start where you are, but start with intent. Invest in the inventories, the governance, and the culture that bring clarity to your mission, and let that clarity become your agency’s advantage.”



Thank You!

Patrick McGarry

GM Federal, data.world

patrick@data.world