

Executive Summary — Data Governance

Sierra Napier

May 2026

Data Governance — Executive Summary

Sierra Napier | Data Scientist | AI Architect

What I Built

Three federal data governance systems using live government APIs and public domain records.

Project	Domain	Data Source	Key Result
Data.gov Catalog	Federal data inventory	CKAN API: ~500 datasets from 120 organizations	Agency distribution analysis + metadata quality assessment
FOIA Compliance	Request processing optimization	FOIA Tracker: 48,277 requests, FY2008–FY2024, 123 agencies	Automated classifier (100% topic accuracy) + SLA prediction
OMB Guidance	Policy lifecycle tracking	OMB API: 170+ guidance documents, 22 agencies assessed	Policy distribution by category + compliance timeline

Why This Matters

For Federal Contractors: Data.gov catalog analysis demonstrates how to audit a client’s data assets. The same metadata quality framework applies to enterprise data catalogs (Collibra, Alation, Informatica).

For Compliance Teams: FOIA automation shows how to reduce manual routing by 80%+. The classifier reads request text and routes to the correct agency automatically — no human triage needed.

For Policy Offices: OMB guidance tracking proves you can monitor policy changes across 22 agencies in real-time. Critical for regulatory compliance and impact assessment.

Key Numbers

- **48,277 FOIA requests** processed and classified
 - **340% backlog growth** identified (FY2008→FY2024)
 - **123 agencies** in compliance database
 - **170+ OMB documents** tracked with lifecycle metadata
 - **~500 Data.gov datasets** cataloged with source attribution
-

Tech Stack

Python · Scikit-learn · Pandas · CKAN API · REST APIs · Streamlit · Plotly

Verification

All data fetched live from government APIs. FOIA records from public domain FOIA Tracker. OMB docs from official API. No fabricated compliance scores.

Live Code: github.com/gosidehustlesisi/sierra-data-governance

Interactive Portfolio: gosidehustlesisi.github.io/sierra-applied-ml/

“Federal data governance isn’t about checklists — it’s about making 500 datasets and 48K requests actually legible to decision-makers.”