



# Project Alexandria Update

Christopher Ritter (PI), Dana Grisham (SNL), Kelly Rose (NETL), Carlos Soto (BNL), Andrew Nicholson (ORNL), Jonathan MacCarthy (LANL), Kate Gibbs (PNNL), Tim Bender (LLNL), Lindsay Roy (SRNL)

Federal Program Manager: Paul Adamson

Technical Advisor: Lakshman Prasad



Defense Nuclear Nonproliferation  
Research & Development Program

- Provides a comprehensive data catalog to allow researchers to quickly discover nonproliferation datasets of interest
- Stores all relevant non-proliferation data in native or normalized format for DNN R&D stakeholders
- Formalizes data management strategies, creates/communicates data management policies and expectations, and acts on opportunities/issues
- Provides capability for real-time data streaming, device synchronization, and transfer of collaborative, large science datasets
- Organizes NNSA's nonproliferation research project data to improve access, discovery, and reuse
- Provides big data access and relevant AI data analysis across the nonproliferation community
- Provides base analytics capability for DNN R&D projects and ventures

Enable **faster project startups** by reducing project IT burden

**Accelerate innovation** with common APIs and standards

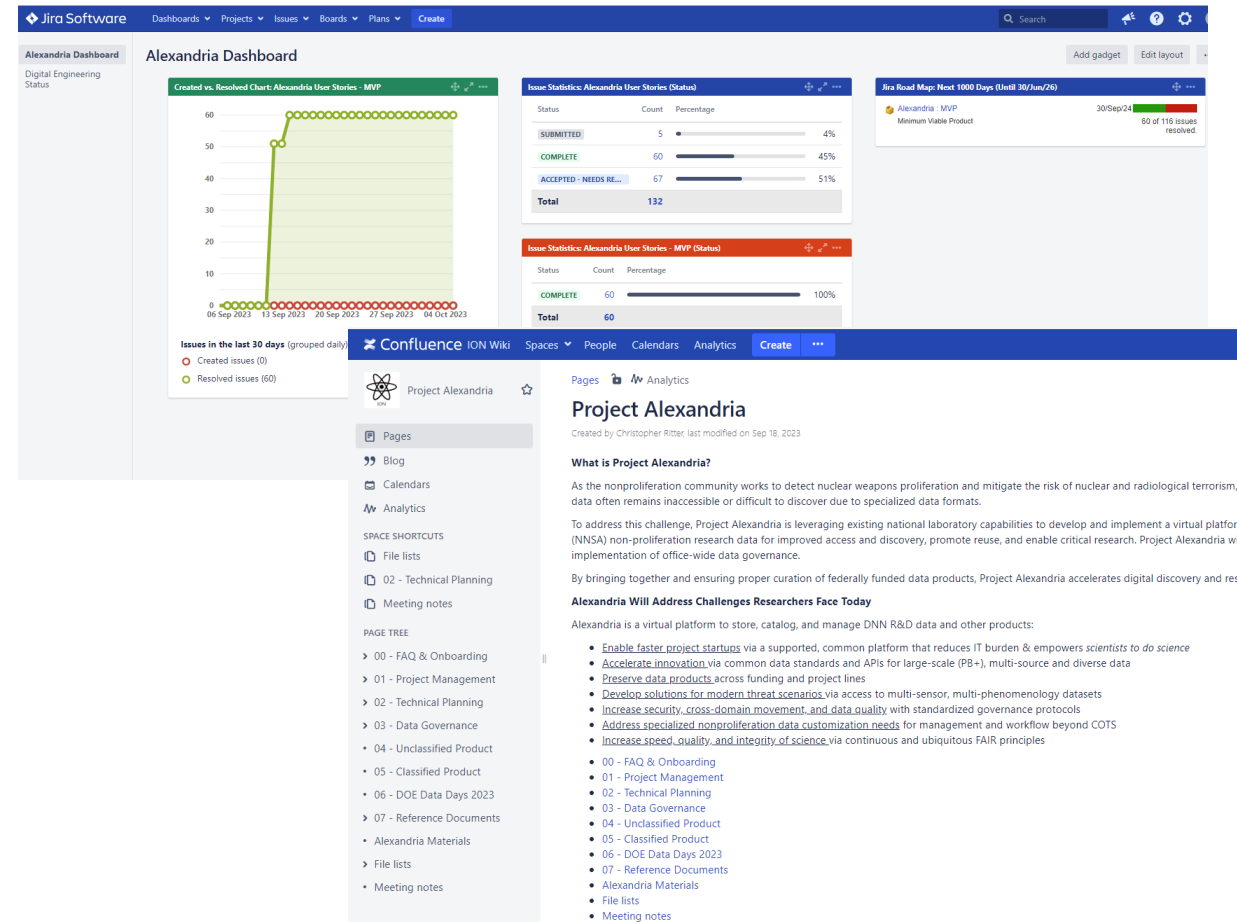
**Share data** products across funding and project lines

Increase **speed, quality, and integrity of science** via FAIR-US

**Accelerate decision science** with analytics and computation

# User Stories and Requirements

- Confluence
  - Project information website
  - Plans, Presentations, etc.
- Jira
  - Software development management platform
  - Contains project schedule, user stories, requirements, development epics/tasks
- Distributed survey to national labs and NNSA ventures
- Received initial 200+ user stories from a variety of roles
  - Categorized to facilitate review
  - Accepted vs. rejected based on known Alexandria scope
  - Determined whether user story applied to MVP vs. later phase



The screenshot displays two overlapping web interfaces. The top interface is the Jira Software dashboard for 'Project Alexandria', showing a 'Created vs. Resolved Chart' with a green line for resolved issues and a red line for created issues. It also includes 'Issue Statistics' for 'Alexandria User Stories - MVP' and 'Alexandria User Stories - Next 1000 Days'. The bottom interface is the Confluence 'ION Wiki' page for 'Project Alexandria', which provides an overview of the project's purpose and a list of related pages.

Status	Count	Percentage
SUBMITTED	5	4%
COMPLETE	60	45%
ACCEPTED - NEEDS RE...	67	51%
<b>Total</b>	<b>132</b>	

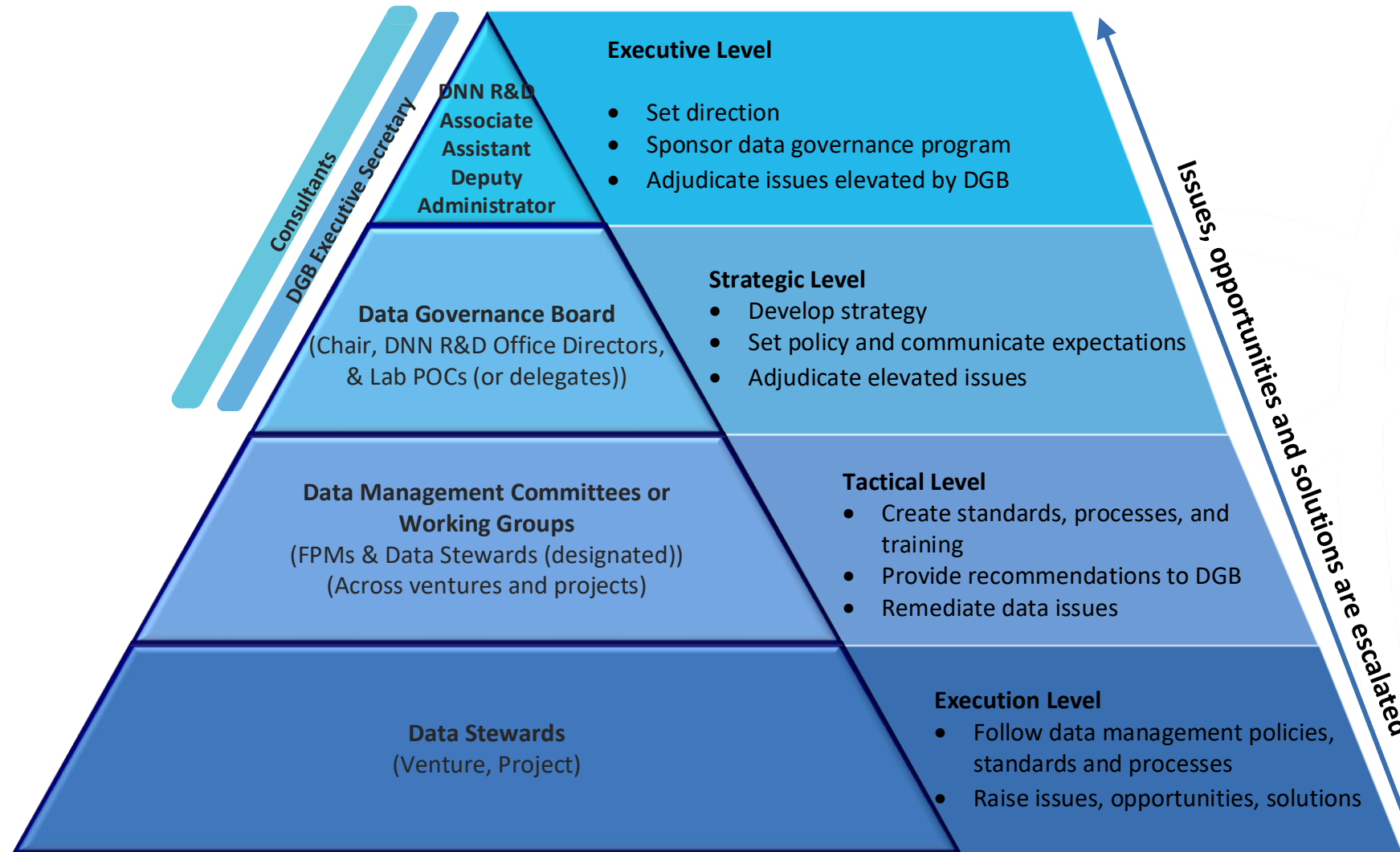
Status	Count	Percentage
COMPLETE	60	100%
<b>Total</b>	<b>60</b>	

Jira: <https://jira.de.inl.gov/>

Confluence: <https://wiki.ion.inl.gov/>

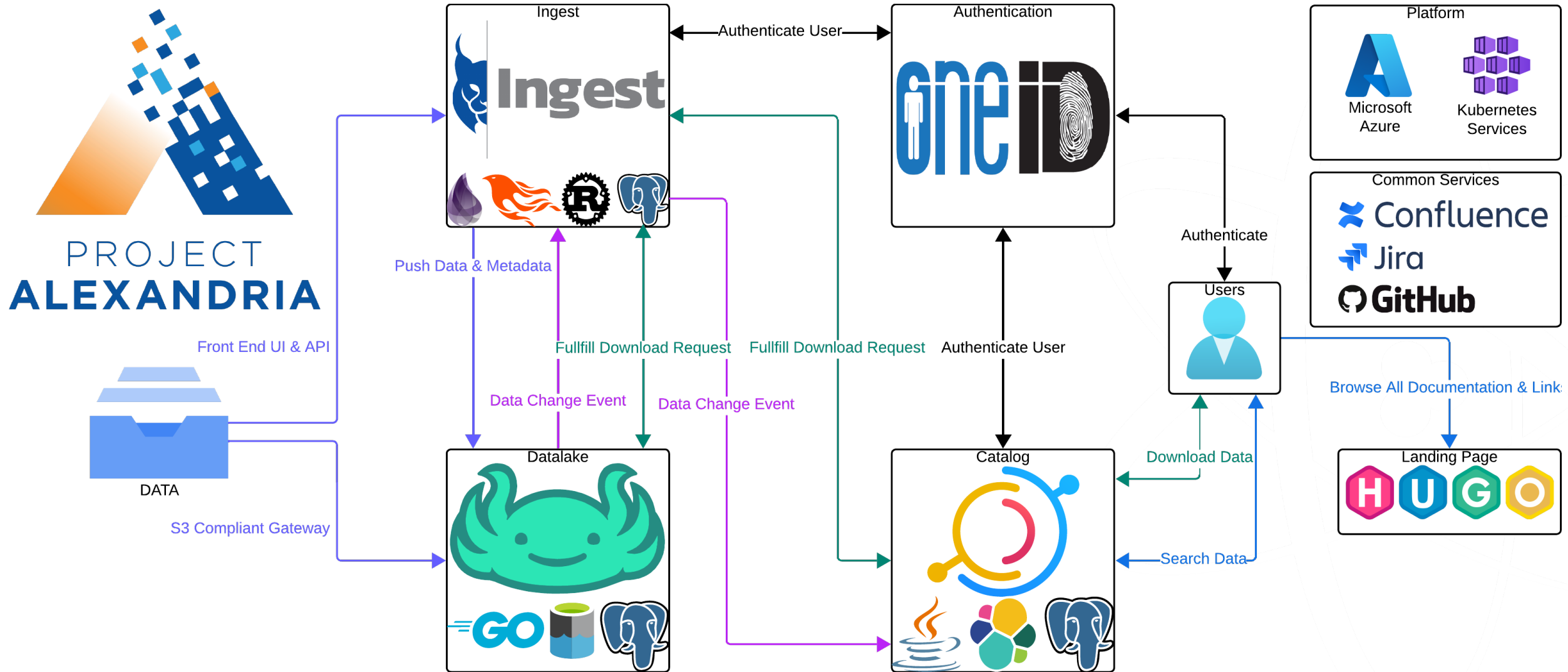
# DNN R&D Data Governance Structure

INNOVATE. COLLABORATE. DELIVER.



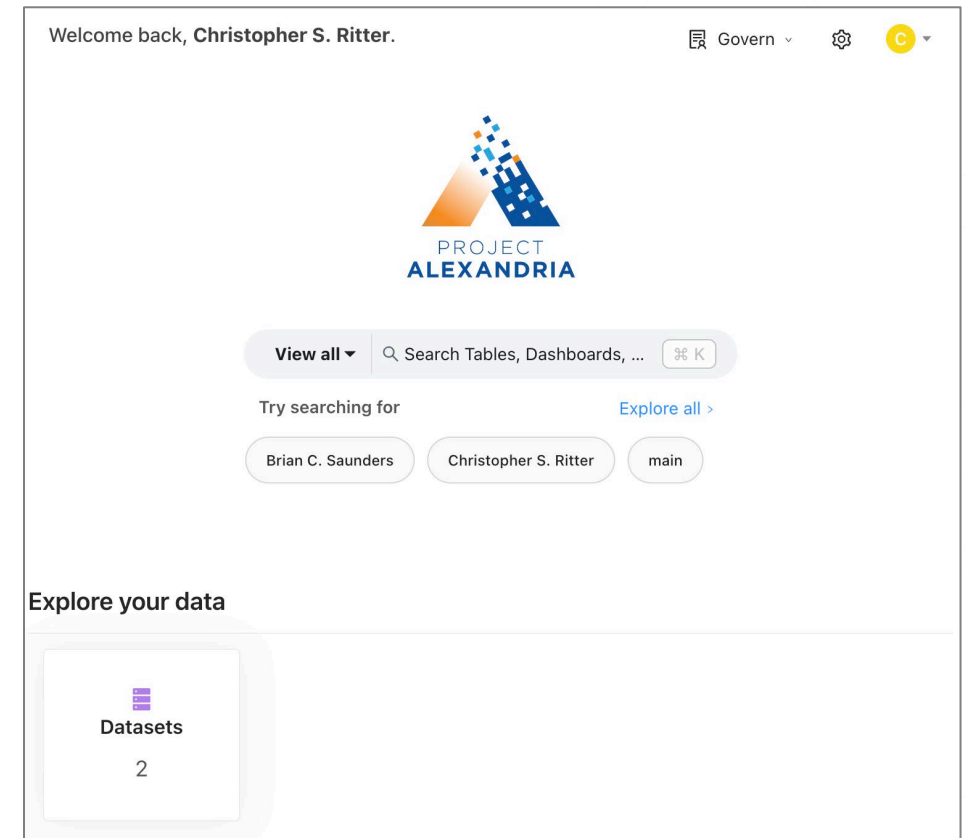
# Architecture

INNOVATE. COLLABORATE. DELIVER.



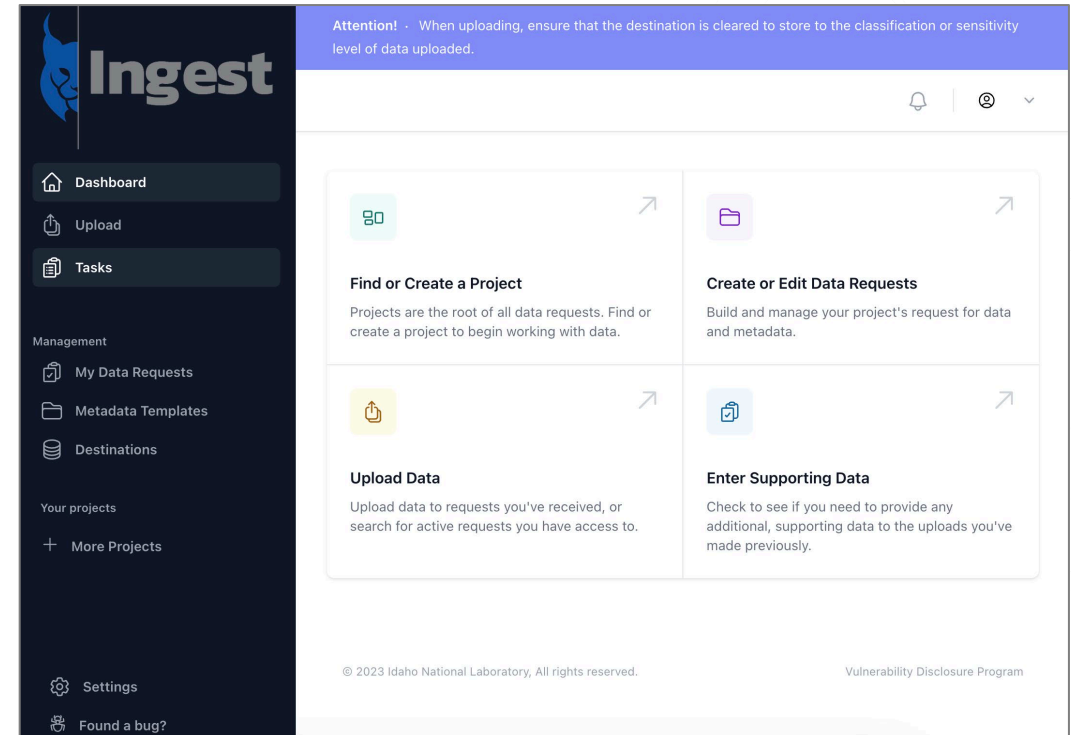
# Catalog Platform (DataHub)

- Purpose: Facilitates easy discovery and retrieval of stored data.
  - Functionality:
    - Serves as a centralized metadata catalog for all data stored in Alexandria.
    - Enables users to search across multiple catalogs for seamless data discovery.
    - Supports governance features such as access control, ownership management, and export controls.
- Technology: Customized version of DataHub extended for comprehensive data management and accessibility.



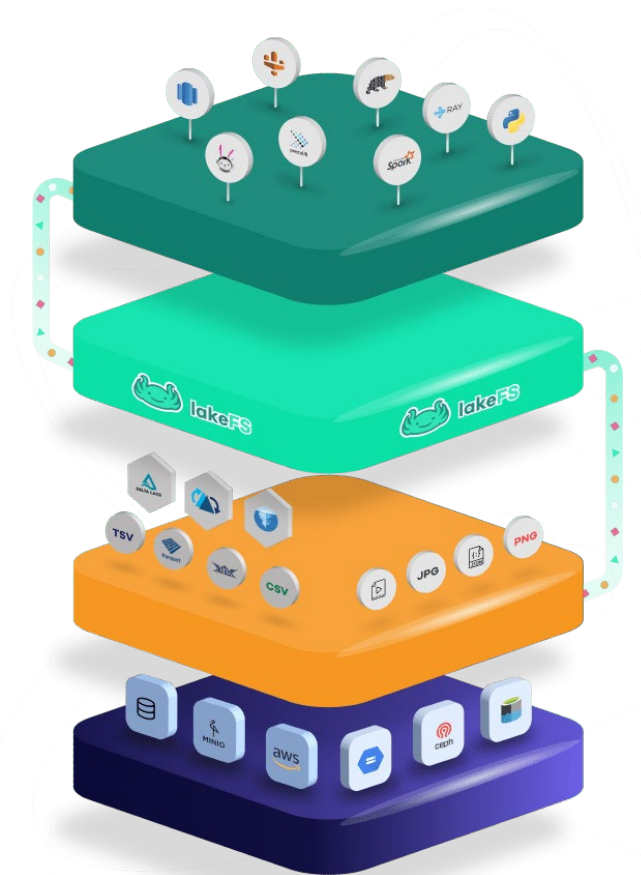
# Data Collection Platform (Ingest)

- **Purpose:** Acts as the primary entry point for data into Alexandria.
- **Functionality:**
  - Allows for the ingestion of data along with important metadata to facilitate organization and retrieval.
  - Supports uploads of files up to 40GB via HTTPS.
  - Ensures that all incoming data is appropriately tagged and cataloged for easy retrieval and management
- **Technology:** Ingest is a completely open-source application that is written in a combination of Rust and Elixir with the Phoenix web framework



# Storage Platform (LakeFS)

- Purpose: Provides secure and scalable storage solutions.
- Functionality:
  - Acts as a version control system (similar to GitHub for data), enabling branching, merging, and rollback functionalities without duplicating data.
  - Decoupled from ingestion and catalog layers to ensure flexibility and scalability.
  - Utilizes object storage for distributed and resilient data storage.
- Technology: Enterprise version of LakeFS



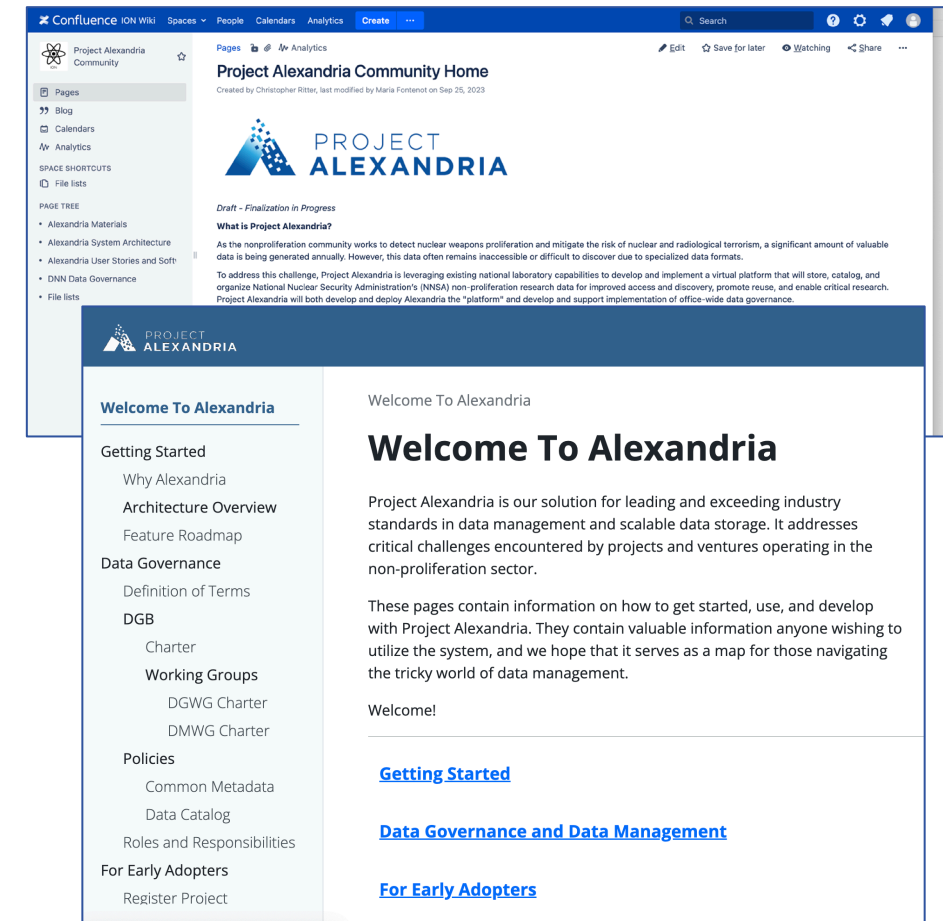
# Alexandria Multi-Lab Roles

INNOVATE. COLLABORATE. DELIVER.

Project Alexandria Laboratory	DNN R&D Data Science POC	Data Governance Team	Requirements Development Team	Architecture & Software Development Team
INL	Chris Ritter	Scott Thompson	Peter Suyderhoud (Co-lead)	John Darrington (Lead)
SNL	Dana Grisham	Dana Grisham	Gerardo Armenta, Brandon Klein	Gerardo Armenta
LANL	Norma Pawley	Richard Stead	Jonathan MacCarthy	Mark Myshatyn
LLNL	Eddy Banks	Rebecca Rodd (Lead)	Tim Bender	Tim Bender
ORNL	Philip Bingham	Andrew Nicholson	Andrew Nicholson	Kent Campbell
PNNL	Kate Gibb	Lynn Wood	Sean Story, Lynn Wood, Amelia Hough-Ross	Nate Tenney
NETL	Kelly Rose		Chad Rowan (Co-lead), Jack Sarle	Jack Sarle
BNL		Donny Hornback		Uma Ganapathy
SRNL		Tom Danielson	Justin Sims	Lindsay Roy
NNSS (new)	Ian McKenna (Tentative)	Ian McKenna (Tentative)		
ANL (new)	Alec Poczatek	Alec Poczatek		Alec Poczatek
LBNL (new)	Brian Quiter	Matthew Henderson		Brian Quiter

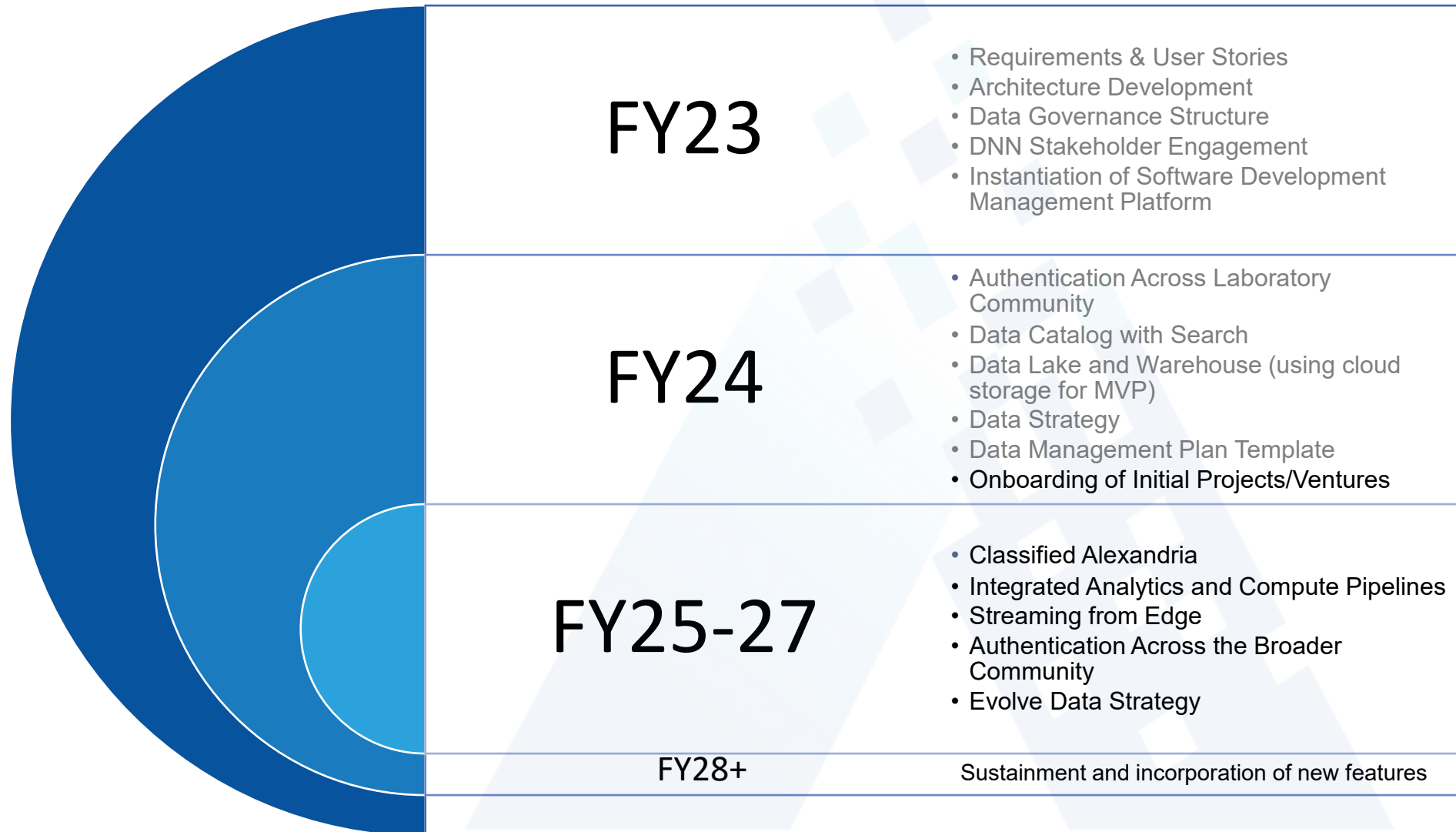
# Communication

- Engagements:
  - Events:
    - NSARD: April 2023, 2024 (poster + presentation)
    - SWRD: May 2023, 2024 (presentation + poster)
    - Monterey Data Conference: 2023 August 2023 (poster)
    - AFTAC Summit: September 2023+January 2024 (presentation)
    - OSD FedDEF: October 2023 (presentation)
    - DOE Data Days: October 2023 (poster + presentation)
    - Federal CIO Council: March 2024 (panel)
    - DOE CyberCon: July 2024 (presentation)
  - National Laboratories: All nine current Alexandria laboratories plus ANL (DarkNet), LBNL (BDC), and NNS (SET/ADventure) on separate projects
  - Government: DOE OCIO, DOE-IN CDO, IC Data Services, DoD NIWC Pacific (Navy Fusion platform)
  - Industry: Microsoft, Amazon, DataBricks, Collibra
- Digital:
  - National Laboratory Community Website
  - National Laboratory Community Jira
  - Team Email Address (alexandria@INL.gov)
  - DNN Stakeholder Survey



# Alexandria Overall Schedule

INNOVATE. COLLABORATE. DELIVER.



# Project Alexandria Summary

- Over **200 initial user stories** captured and refined into requirements for minimal viable product
- Delivery of **DNN Data Governance Board Charter** to community and formal structure identified
- Consistent **external communication** across DNN events, focus groups, and online
- Risk mitigation through **maximum reuse** of COTS, GOTS, existing ATOs, and open-source technology
- Minimal viable product (**MVP**) **now available**
- Testing and initial **onboarding underway**





# Questions?

[Alexandria@inl.gov](mailto:Alexandria@inl.gov)



Defense Nuclear Nonproliferation  
Research & Development Program