



# Project Alexandria – Managing and Enabling Discovery of DNN R&D Data

Jack Sarle<sup>1</sup>, Christopher Ritter (PI)<sup>2</sup>, Dana Grisham<sup>3</sup>, Kelly Rose<sup>1</sup>, Carlos Soto<sup>4</sup>, Andrew Nicholson<sup>5</sup>, Jonathan MacCarthy<sup>6</sup>, Kate Gibbs<sup>7</sup>, Tim Bender<sup>8</sup> & Lindsay Roy<sup>9</sup>

1. NETL 2. INL 3. SNL 4. BNL 5. ORNL 6. LANL 7. PNNL 8. LLNL 9. SRNL

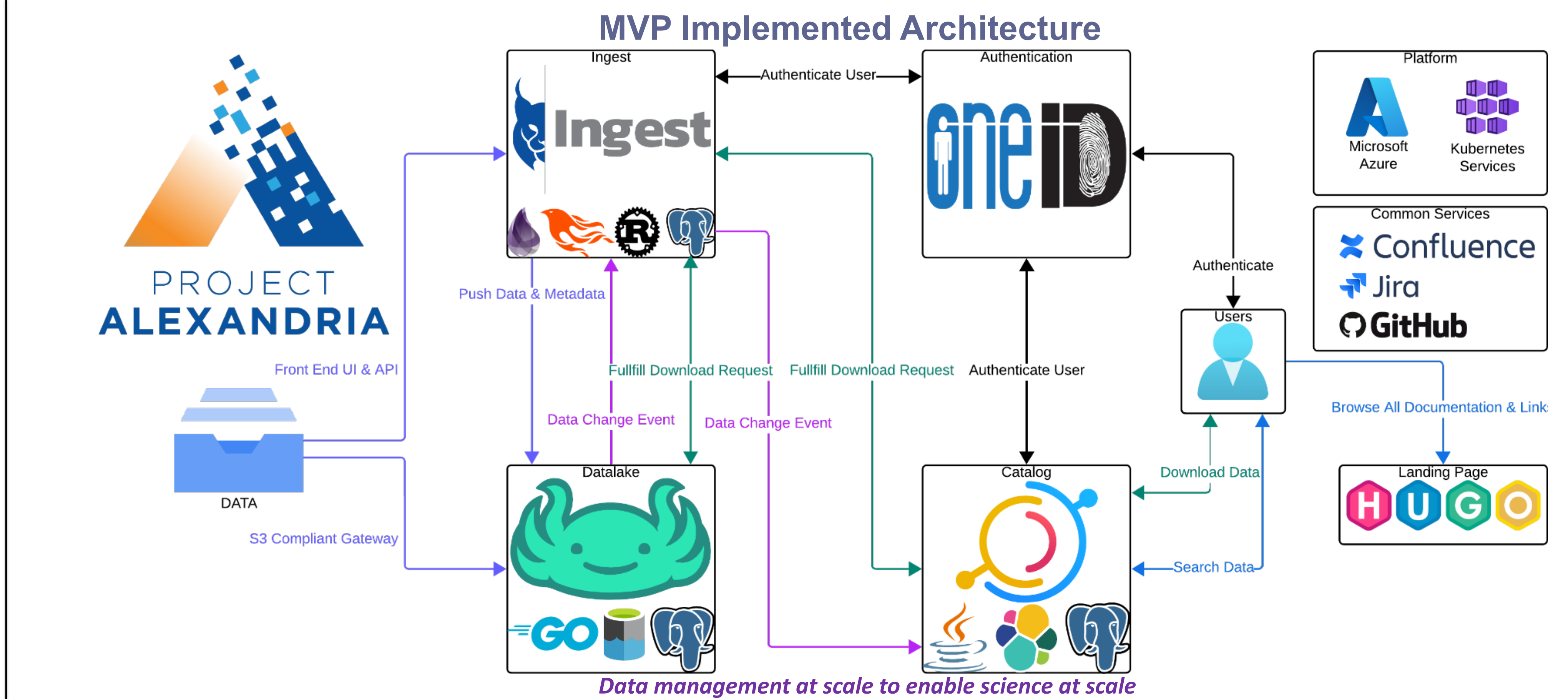
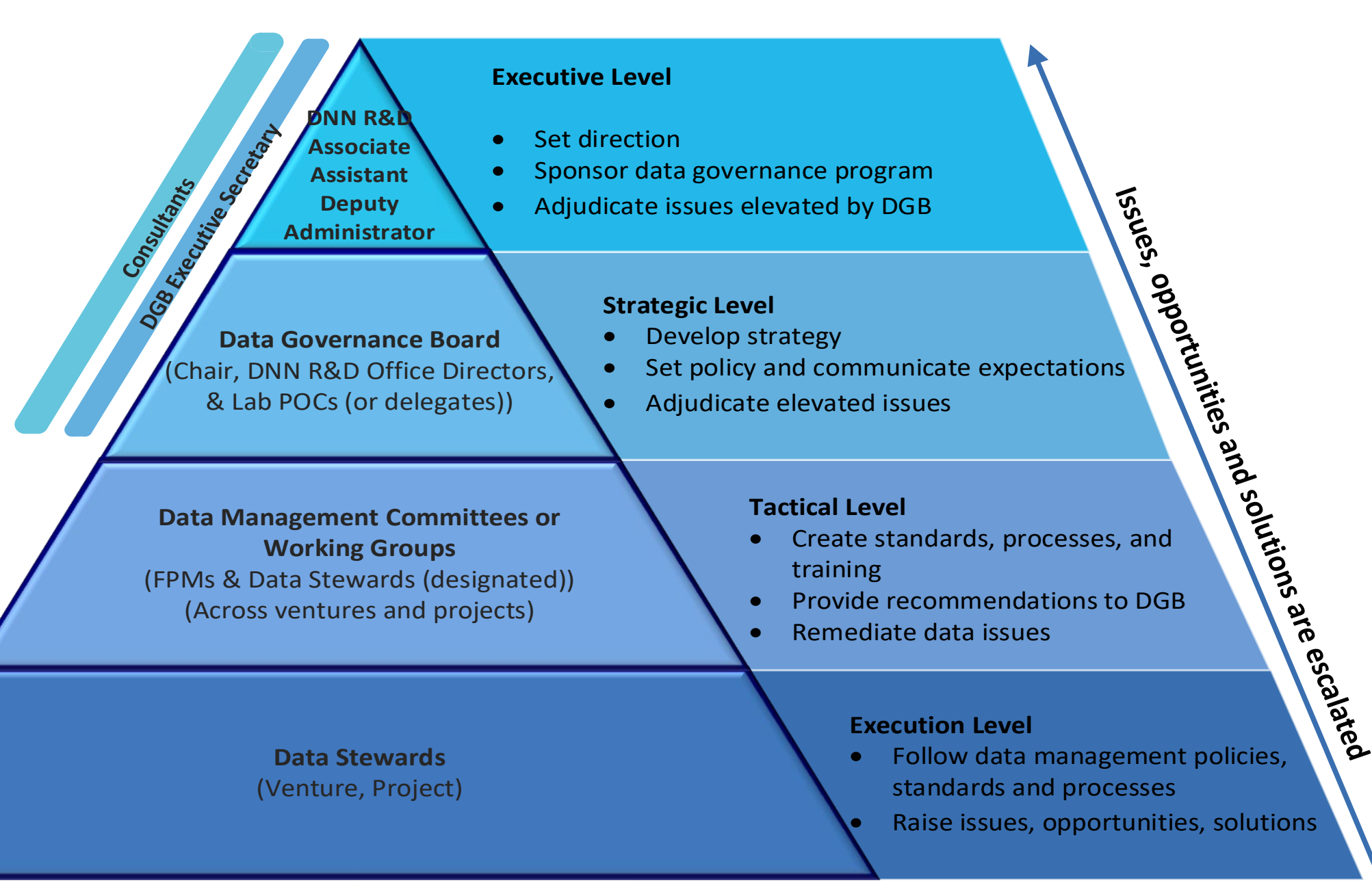
## Project Overview

- ▶ The DNN R&D portfolio generates petabytes of data that often persist in specialized data formats, is inaccessible, and is undiscoverable to other internal or external projects. Additionally, there are unique workflow and data challenges to DNN R&D.
- ▶ Project Alexandria will leverage existing laboratory capabilities to architect and deliver a platform to store, catalog, and organize DNN R&D data assets with the goal to improve data access and discovery, promote the frequency of reuse, and enable transformative research.

## Data Governance FY24 Accomplishments

- ▶ Developed a Data Governance Charter
- ▶ Developed data quality standards, specialized DNN R&D metadata standards, policies, procedures, and processes
- ▶ Developed a data management plan template specific to DNN R&D needs

### DNN R&D Data Governance Structure



## Alexandria Will Address Current Research Challenges

- ▶ Enable faster project startups via a supported, common platform that reduces project-level IT burden & empowers *scientists to do science*
- ▶ Accelerate innovation via common data standards and APIs for large-scale (PB+), multi-source and diverse data
- ▶ Preserve data products across funding and project lines for future research needs
- ▶ Develop solutions for modern threat scenarios via common access to multi-sensor, multi-phenomenology datasets
- ▶ Increase security, cross-domain movement, and data quality with standardized governance protocols
- ▶ Address specialized nonproliferation data customization needs for management and workflow beyond COTS tools capability
- ▶ Increase speed, quality, and integrity of science via continuous and ubiquitous FAIR-US principles

- Ingest**
  - Purpose: Acts as the primary entry point for data into Alexandria.
  - Technology: Ingest is a completely open-source application that is written in a combination of Rust and Elixir with the Phoenix web framework
- Catalog**
  - Purpose: Facilitates easy discovery and retrieval of stored data.
  - Technology: Customized version of DataHub extended for comprehensive data management and accessibility.
- Datalake**
  - Purpose: Provides secure and scalable storage solutions.
  - Technology: Enterprise version of LakeFS



**CONTACT**  
 Christopher Ritter (PI)  
 301-910-1818  
 Christopher.Ritter@inl.gov