



GOT DATA? DISCOVERY IS THE KEY

Data Governance – a key component of Data Strategy

Purpose (why we exist)

- Data Governance supports **data-driven insights** and **access to data** to enable Sandia's **innovation acceleration and modern engineering goals**.

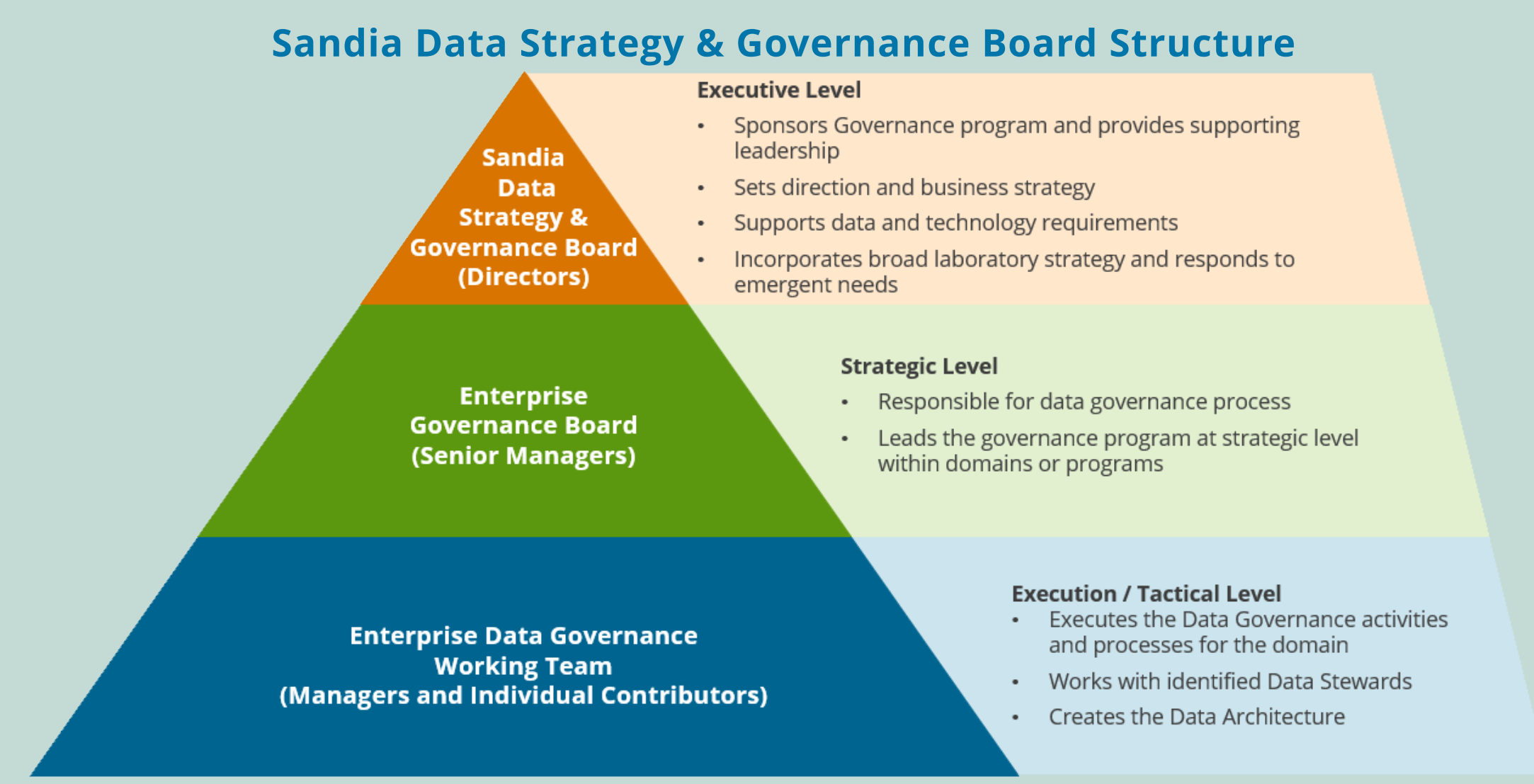
Vision (what we aspire to)

- Sandia data is securely accessible, understood, and trusted and is used in a consistent, responsible, and meaningful manner to **increase interoperability, ensure reliability, and inform decision making in our community**.

Mission (what we do)

- We design and implement policy and standards to enable data management and data sharing in order to support **regulatory compliance, internal reporting & analytics, operational analytics for our business, data integration, and Artificial Intelligence (AI) and Machine Learning (ML) innovation and operations**.

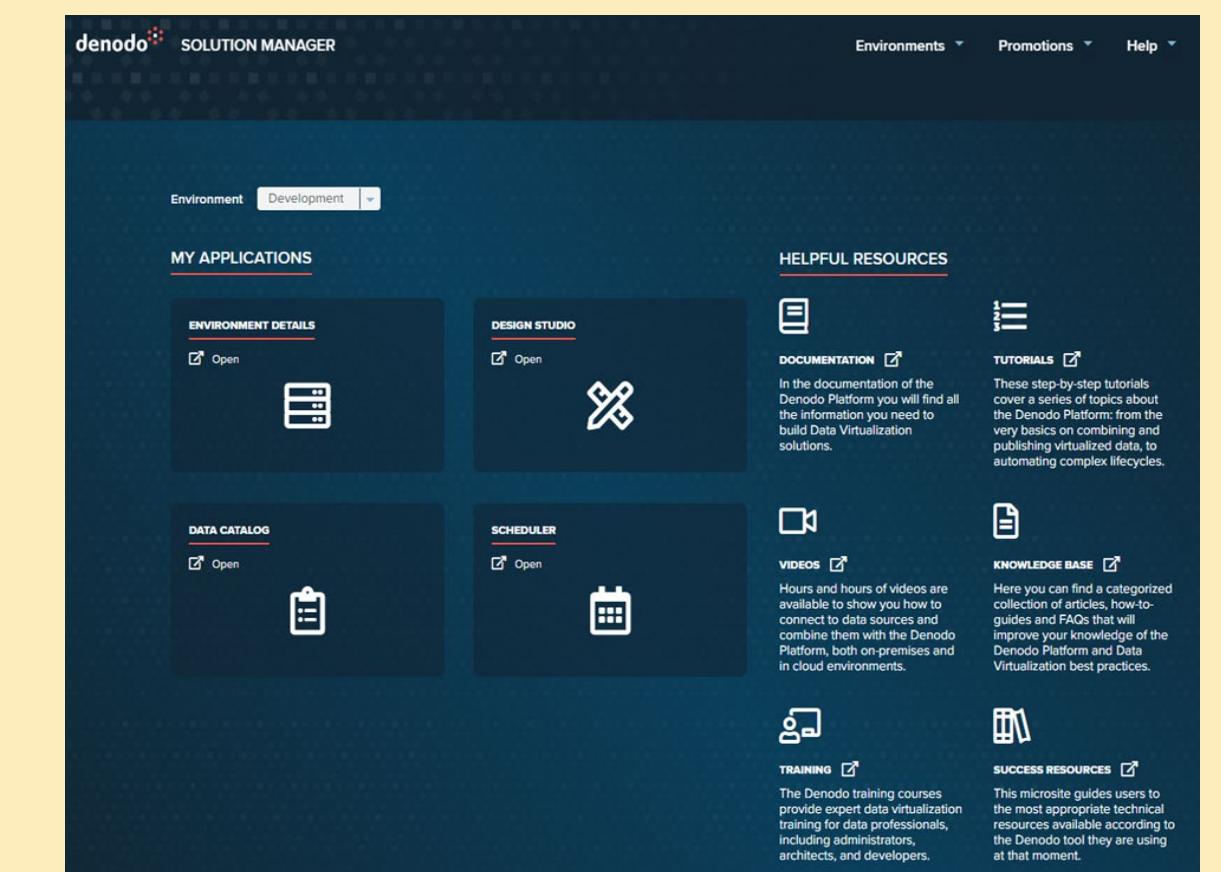
Who do we talk to about the data?



Surfacing Data with Denodo

Denodo Design Studio

- Create Virtual Database (VDB) access
- Policy-based access
 - Role-based access controls (RBAC)
 - Attribute-based access controls (ABAC)



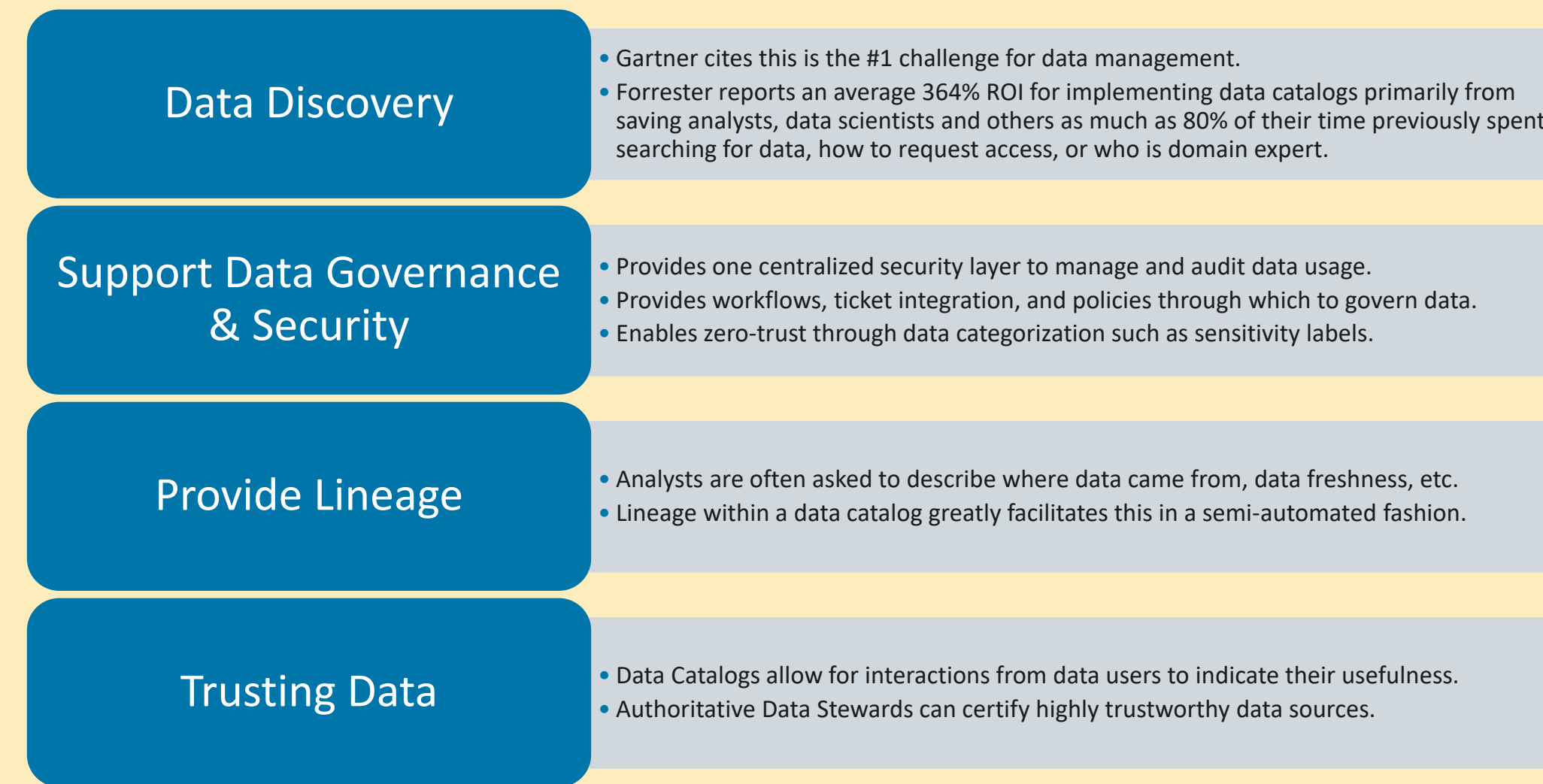
Denodo Data Catalog

- Enrich with metadata
- View lineage
- Search by keyword
- Filter fully curated data sets
- Surface data (once access is granted)

Focus Areas for Sandia Data Strategy & Governance

- Enterprise Assets** – Advocate the treatment of Sandia data as a corporate and government asset
- Use Ready** – Enable access and use of data for timely data-driven decisions
- Ease of Use** – Support and develop mutually beneficial standards and practices to provide efficient and effective access to data
- Support Community** – Develop and socialize the responsibilities of Data Stewards
- Taxonomy** – Align the organizational strategies, digital transformation and data strategy, and data governance objectives between Enterprise and Mission data governance efforts
- Funding** – Secure support and funding for agreed upon data strategy and data governance efforts
- Communication** – Provide organization-wide visibility to and flow-up / flow-down communication of data strategy and data governance efforts
- Data Management Standards** – Federate governance across program and mission areas

Data Integration Benefits



Denodo Advantages

- Web UI is relatively easy to navigate
- Unlimited catalog editors, licensing per core not per user
- Query data from within the catalog
- Denodo lineage shows provenance and dependencies between tables
- Open data into PowerBI or Tableau with a single click
- Singular integration point and access controls enable easy data governance
- Column descriptions defined once at "base view" and flow to dependent views
- Good customer engagement with vendor

Denodo Challenges

- Encountered many bugs and breaking features
- Backend server challenging to maintain
- Substantial amount of time testing and working with the vendor
- Attribute-based access control (ABAC) policies are difficult to debug

FAIR* Data Principles — GO FAIR is a bottom-up, stakeholder-driven and self-governed initiative

Findability

- Data catalog for democratization of data
 - Metadata standards of common information
 - Data searchable

Accessibility

- Data catalog with virtualization capabilities
 - Denodo with role-based access control (RBAC) and attribute-based access control (ABAC)

Interoperability

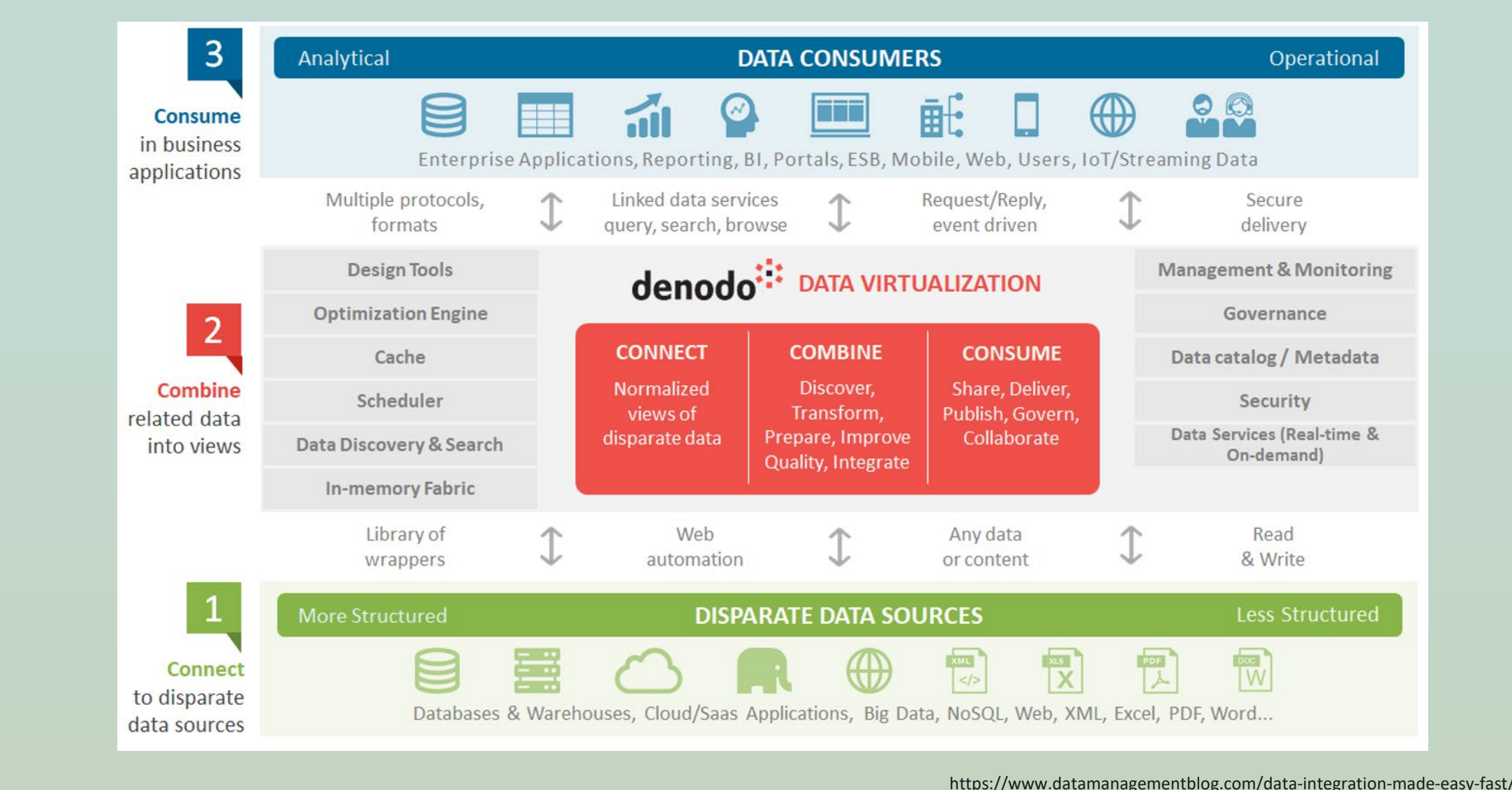
- Use a formal method for metadata that can work with other products

Reuse of Digital Assets

- Describe the data so others can reuse the data

<https://www.go-fair.org/fair-principles/>

Denodo Data Virtualization



<https://www.datamanagementblog.com/data-integration-made-easy-fast/>

Surfacing Data with Denodo

- Federated data access, cross-site data connectivity
- Cross-site alignment on metadata, governance, tagging, etc.
- AI and model governance and cataloging
 - Ties into software governance
 - Inventory of deployed models just as important as data!
- Generative AI governance
 - Models will need access to large amounts of data
 - National Institute of Standards and Technology (NIST) AI Risk Management Framework (RMF)
 - Securing data, data pipelines into indexing, and model delivery all still require access controls and need-to-know (NTK)
- Mature data governance, improve culture, decrease time to review and deliver data



GOT DATA? DISCOVERY IS THE KEY

Abstract

Data is growing at an exponential rate. Powerful tooling exists that can leverage this data to glean insights and strategic value critical to keeping our competitive edge.

- Data Governance – a key component of Data Strategy
- Focus Areas for Sandia Data Strategy & Governance
- Degrees of Information Usefulness
- How to follow the FAIR* data principles of Findability, Accessibility, Interoperability, and Reuse of digital assets
- Data Consumer Challenges: Where is the data? Who grants access? Who understands the data?
- What metadata is critical to capture?
- How to monitor data and insure high quality and accuracy over time
- Use of Denodo Virtualization Tool and Data Catalog in data discovery
- Future Methods: Incorporating Generative AI to aid in data search within Denodo

Sandia's Enterprise Data Governance team would like to share our experiences with using the Denodo Data Catalog to manage data, share pros and cons.

* <https://www.go-fair.org/fair-principles/>