SAND2018-3592PE

Open Data E

Enabling Mission Data M

Rick Moleres

Manager, Science and Engineeri

Cyber Security and Mission Co

April 11th, 2018

Sandia National Laboratories is a multi-missio

subsidiary of Honeywell International, Inc., fo

Agenda

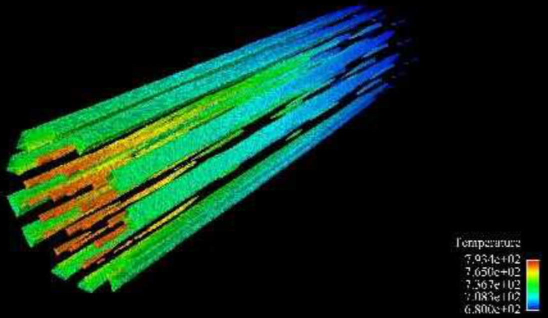
- Drivers for an Open Data Environment (ODE)
- Objectives of ODE
- Conceptual Architecture
- Key Challenges

Industry Metrics on Finding Data

Industry Metric	Impacts
Employees spend 20% or more of their time just searching for data	- Lost productivity - Lost opportunity - Cost of re-work when data is not found - Cost of using incomplete information

- "According to a McKinsey report, employees spend 1.8 hours every day searching and gathering information. Put another way, businesses hire 5 employees but only 4 show up to work; the fifth is off searching for answers, but not contributing any value." Source: Time Searching for Information.
- "19.8 per cent of business time – the equivalent of one day per working week – is wasted by employees searching for information to do their job effectively," according to Interact. Source: A Fifth of Business Time is Wasted Searching for Information, says Interact
- IDC data shows that "the knowledge worker spends about 2.5 hours per day, or roughly 30% of the workday, searching for information…" Source: Information: The Lifeblood of the Enterprise.

Sandia's Need for Mission Data Management



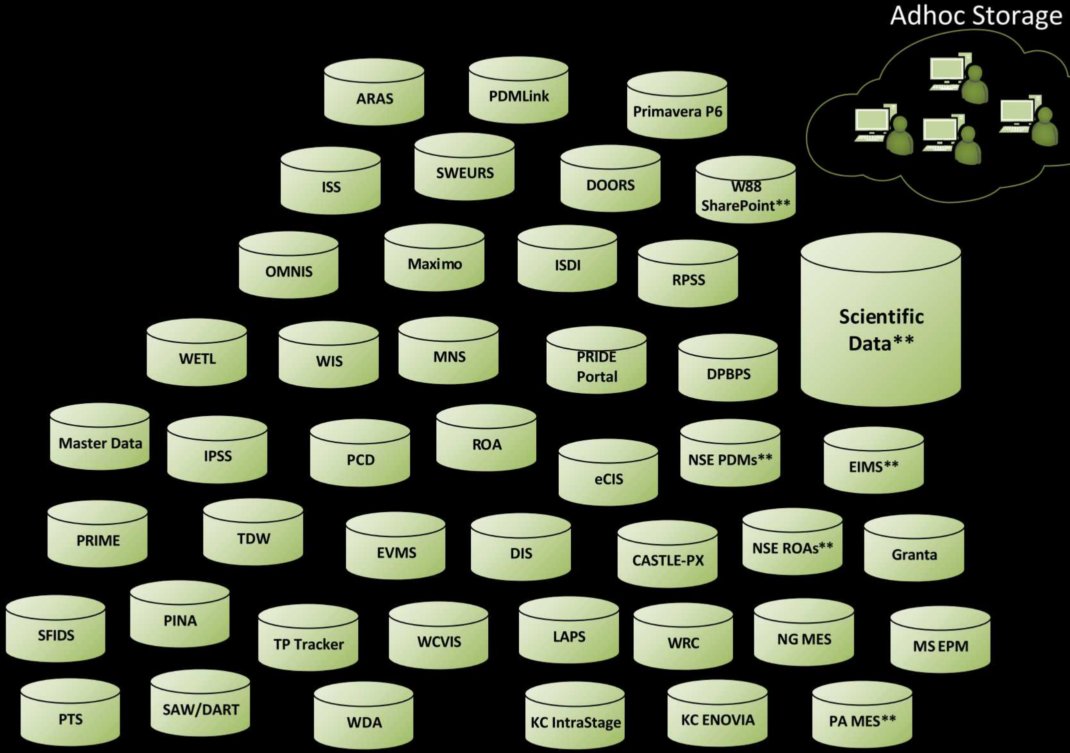
- Nuclear Deterrence (ND) Data Strategy (4/2017)
  - Formalized Chief Data Officer (CDO) and Data Governance in 2016
  - "ND must transform into a data-centric organization…"
  - "Strategic data management will be enabled by an architecture that truly connects our information collection, storage, retrieval and analysis systems …"
- Data Science Consortium Needs Assessment (9/2017)
  - "… centralized and electronically searchable storage of data and better data organization"
- Survey of Scientific Data Management and Archiving Needs (8/2017)
  - Driven by Engineering Sciences Research Foundation (ESRF)
  - "Broad need for data archiving and storage across the laboratory"
  - "Existing practices offer little potential for collaboration and/or data mining strategies"

ND Data Strategy

- Drivers for change
  - Sustained Deterrence through better understanding of data
    - Better understanding of threats
    - Accelerated weapon program lifecycles
    - Increase in innovation and opportunities
  - Stockpile Evaluation and Assessment
    - Drive forward-looking assessment of the stockpile’s safety, security, and effectiveness (e.g., predictive analytics)
- Requires an agile, responsive data environment
  - Data is an asset, data is governed, data is shared
  - Data must be accessible, understandable, traceable, and secure
  - Integrate data from disparate systems
  - Enable discovery, access, and analysis

Example ND Data Haystack

- Difficult to find, access, and use information across data sources
  - Weeks to months, a manual hunt
- Difficult to get access to many data sources
- Sources may exist on the unclassified and classified networks
- This diagram represents only a subset – many more repos exist
  - Including ad-hoc on desktops, in cabinets, etc.
- New data are continually created



Adhoc Storage

ARAS
PDMLink
Primavera P6
ISS
SWEURS
DOORS
W88 SharePoint**
OMNIS
Maximo
ISDI
RPSS
Scientific Data**
WETL
WIS
MNS
PRIDE Portal
DPBPS
Master Data
IPSS
PCD
ROA
eCIS
NSE PDMs**
EIMS**
PRIME
TDW
EVMS
DIS
CASTLE-PX
NSE ROAs**
Granta
SFIDS
PINA
TP Tracker
WCVIS
LAPS
WRC
NG MES
MS EPM
PTS
SAW/DART
WDA
KC IntraStage
KC ENOVIA
PA MES**

**multitude of data sources

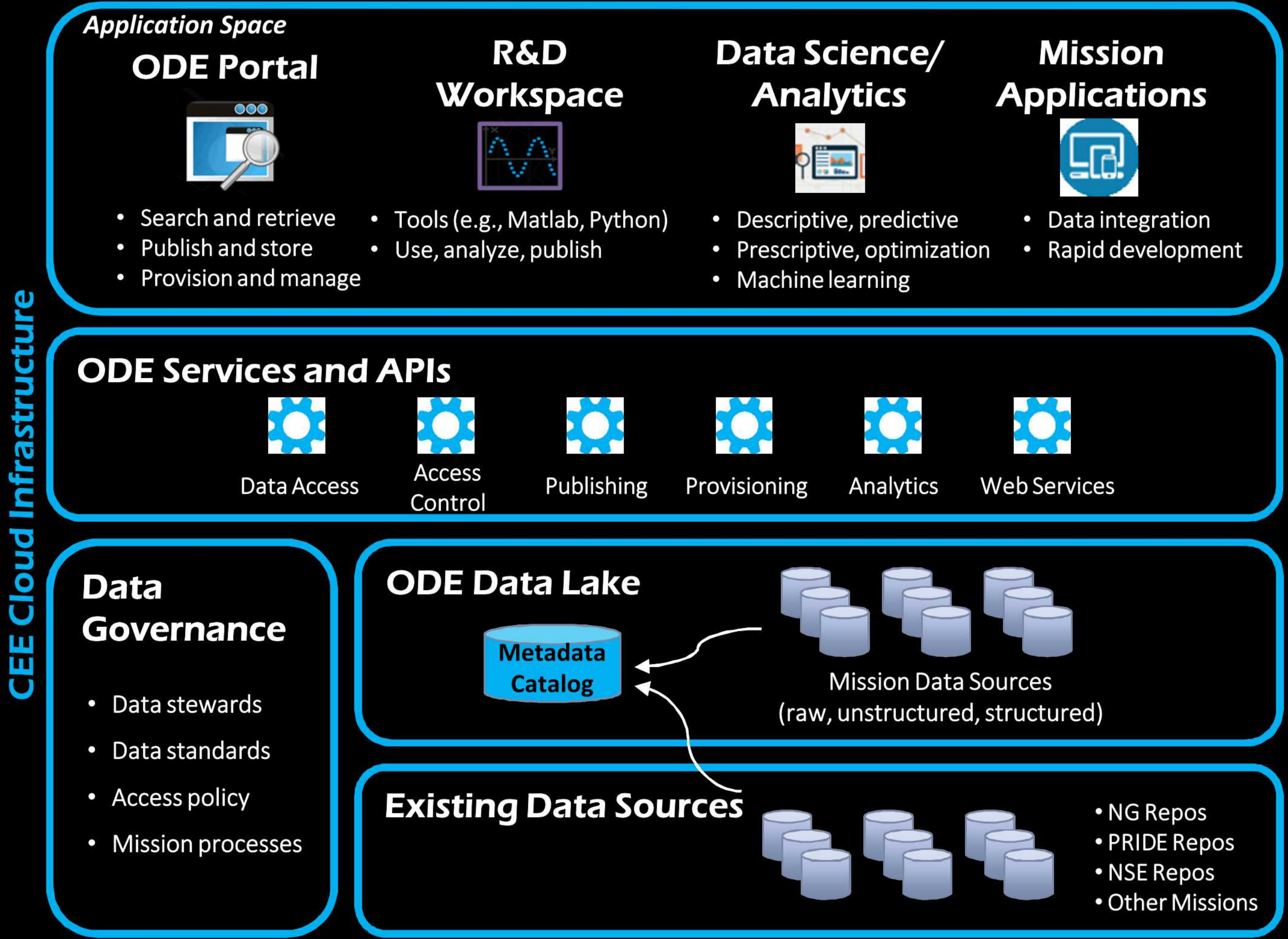
Data Management Issues in General

- Stovepipe storage and archiving solutions exist because there is no comprehensive, corporate data management solution
  - Data science is achieved in pockets, no enablement strategy or platform, and little to no automation
  - Raw data is often thrown away or lost after analyses are complete or derived data sets are created
- If not involved in the data creation, it is difficult to find existing data
  - Data will often get re-created because there is no way to search or find existing data
- Inconsistent, crude, or lack of metadata across mission data repositories, makes searching and correlation difficult to impossible
- Access controls and processes are inconsistent and complex across mission programs and IT solutions

Primary Objectives of ODE

- Make mission data easy to find and retrieve across a multitude of data sources and data owners
  - intelligent search and faster, secure access to data
- Provide a centralized, common storage location for mission data
  - secure and persistent, high storage and network capacities, data management tools
- Enable end users, software applications, and data science and analytics engines to easily use and store data
  - using a rich and extensible set of services and APIs
- Align with mission data strategies like the ND Data Strategy to support data governance and access control policies and processes

Currently driven by ND, but applies generally across Sandia mission and scientific data domains



CEE Cloud Infrastructure

Application Space

ODE Portal

- Search and retrieve
- Publish and store
- Provision and manage

R&D Workspace

- Tools (e.g., Matlab, Python)
- Use, analyze, publish

Data Science/ Analytics

- Descriptive, predictive
- Prescriptive, optimization
- Machine learning

Mission Applications

- Data integration
- Rapid development

ODE Services and APIs

Data Access

Access Control

Publishing

Provisioning

Analytics

Web Services

Data Governance

- Data stewards
- Data standards
- Access policy
- Mission processes

ODE Data Lake

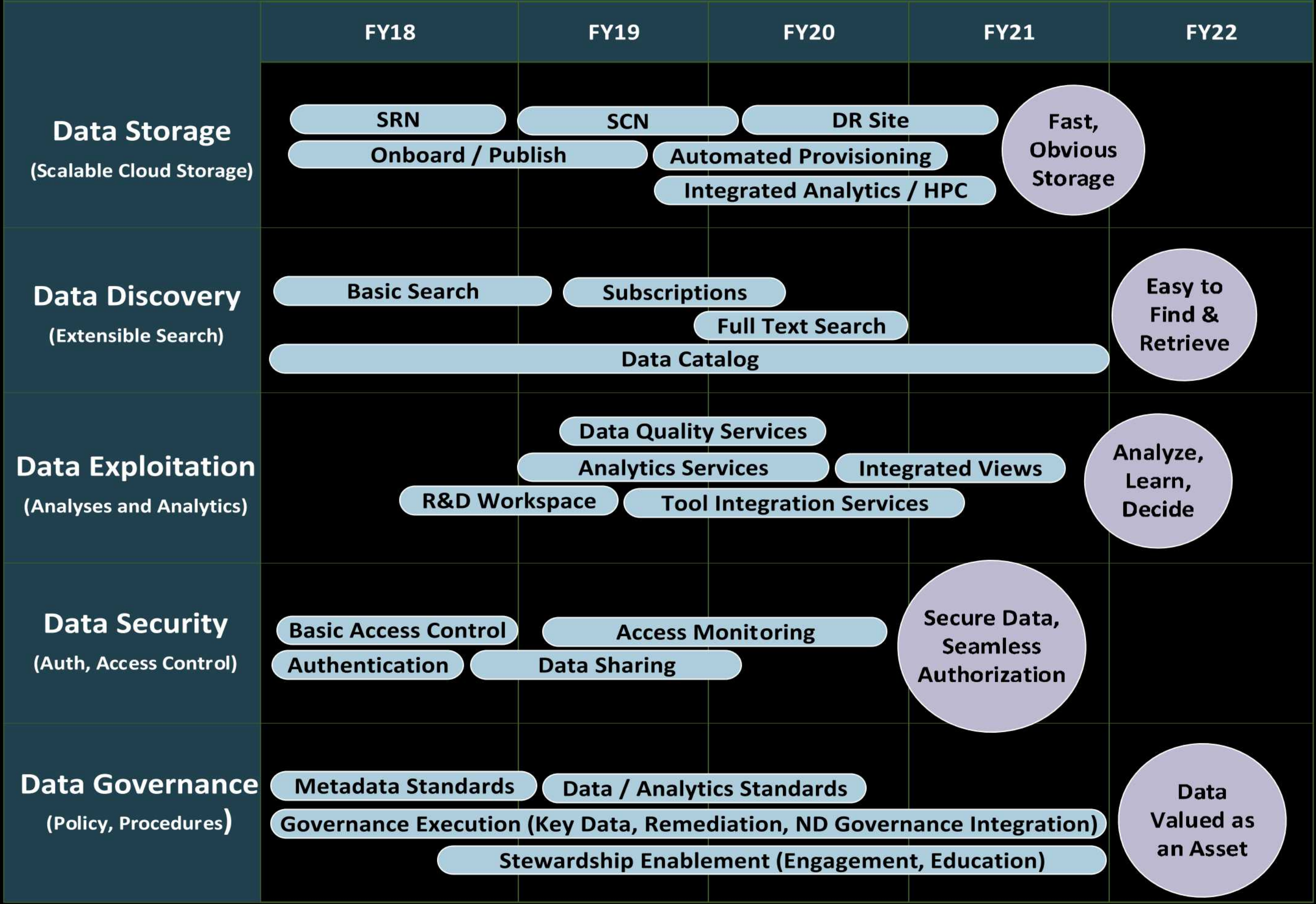
Metadata Catalog

Mission Data Sources (raw, unstructured, structured)

Existing Data Sources

- NG Repos
- PRIDE Repos
- NSE Repos
- Other Missions

ODE Capability Roadmap



FY18 | FY19 | FY20 | FY21 | FY22

Data Storage
(Scalable Cloud Storage)

- SRN
- SCN
- DR Site
- Onboard / Publish
- Automated Provisioning
- Integrated Analytics / HPC
- Fast, Obvious Storage

Data Discovery
(Extensible Search)

- Basic Search
- Subscriptions
- Full Text Search
- Data Catalog
- Easy to Find & Retrieve

Data Exploitation
(Analyses and Analytics)

- Data Quality Services
- Analytics Services
- Integrated Views
- R&D Workspace
- Tool Integration Services
- Analyze, Learn, Decide

Data Security
(Auth, Access Control)

- Basic Access Control
- Access Monitoring
- Authentication
- Data Sharing
- Secure Data, Seamless Authorization

Data Governance
(Policy, Procedures)

- Metadata Standards
- Data / Analytics Standards
- Governance Execution (Key Data, Remediation, ND Governance Integration)
- Stewardship Enablement (Engagement, Education)
- Data Valued as an Asset

Key Challenges

- Data governance is critical to engage stakeholder communities in data sharing and data storage policy and procedures
  - Culture change required both for data sharing and data storage
  - Must define requirements, standards, and use cases for IT systems
- ODE must provide an appropriate level of data security to satisfy concerns that easier access to data for Sandians can be a safe practice
- ODE cloud infrastructure – storage, network, compute – is maturing at Sandia but there are new technologies and infrastructure components (e.g., object storage) that still need to be proven
- Ensure a sustained capability and funding model
  - Governance, CIO/CISO/CDO investment and alignment, mission value

Summary

- ODE is a mission data access and management capability
  - A critical element to broadly enable data sciences at Sandia
  - Currently driven by ND, but intended to solve a laboratory problem
- There are key challenges ahead to overcome