



Realizing FAIR Data and Open Science:

Improving the Findability, Accessibility,
Interoperability, and Reusability of NOAA Data

NSF CI4MF



National Oceanic and
Atmospheric Administration

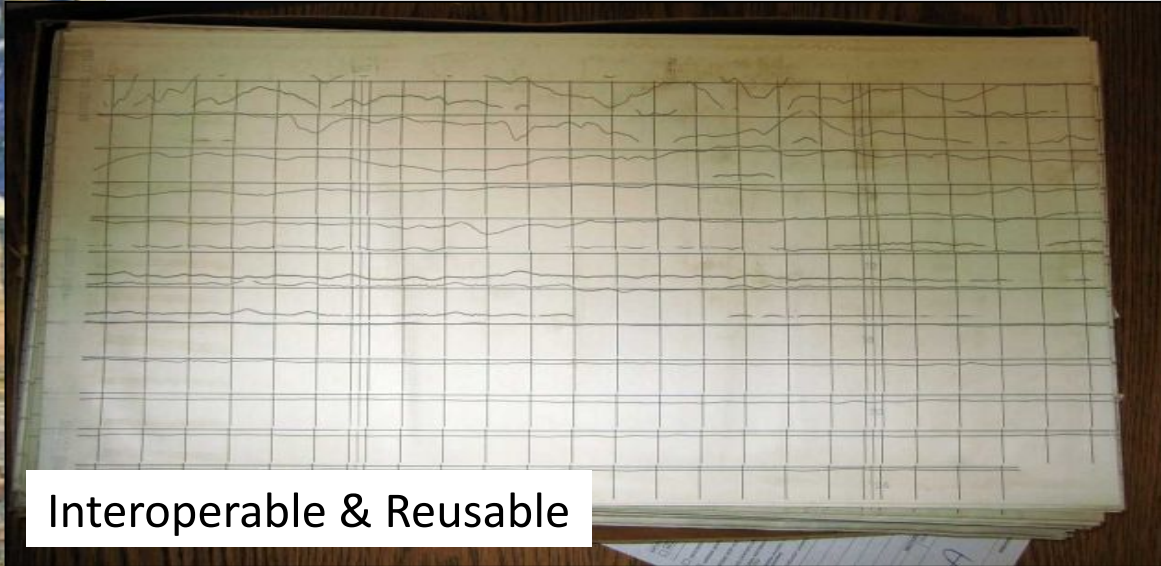
Monica Youngman, Ken Casey, Ryan Berkheimer
National Centers for Environmental Information

What does FAIR data and Open Science Mean?

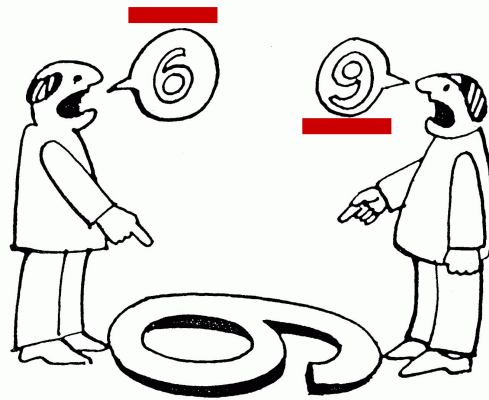
Findable & Accessible



Interoperable & Reusable



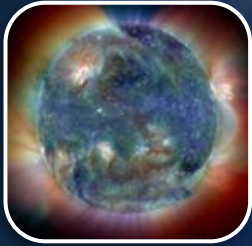
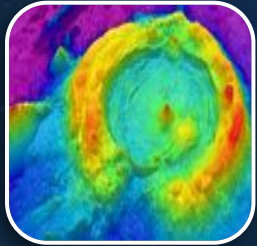
Transparent and accessible knowledge
Ensuring data and metadata are complete, documented, accessible, and understandable to a broad community



Outcome: Full value, impact, and credit of data and research to science, decision-making, problem solving and innovation.

Making Data Useful

Earth Observing Systems



National
Centers for
Environmental
Information

Data and Science products for decision-making

Harmful
Algal
Blooms
Observing

Coastal
Digital
Elevation
Models

Heating &
Cooling
Degree Days

Hurricane
Tracks

Solar
Activity/
Sunspots

Tsunami
Warning

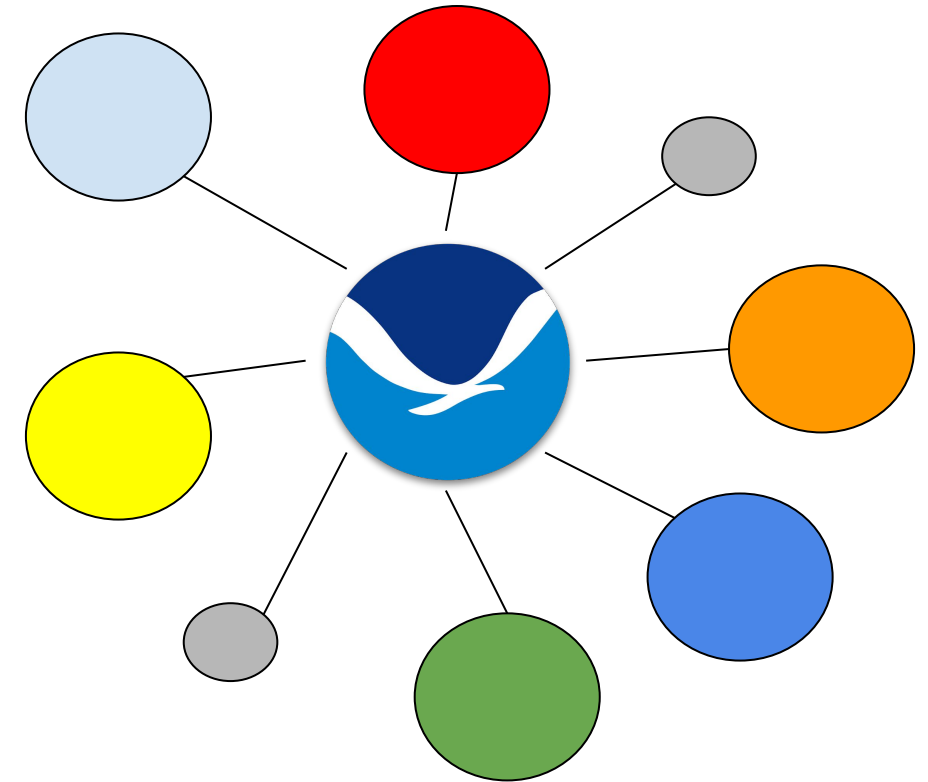
Imagine a Future...

Imagine: Google/Amazon/Netflix-like tool for searching recommending, and accessing government data

- “Users like you also looked at...”
- “Papers and presentations you might be interested in...”
- “Authors you might know...”

Imagine Further: Government data connected across agencies from environmental data to Census data, business data, land management, and more

- “The percent of the population in your county that has high fire risk... with the elderly at risk population being...”
- “Water resources in your county have increased/declined by... over the past decade while the population increased/decreased by... and land development changed by ...”



Open Data + Technology



Step Changes to an Interoperable Open OneNOAA

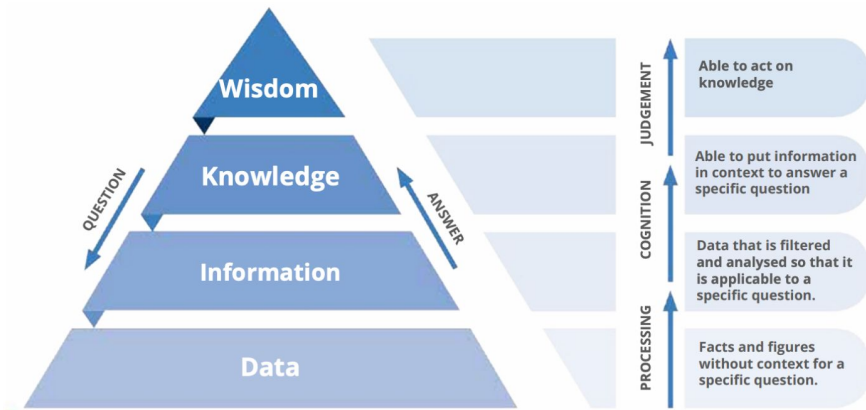
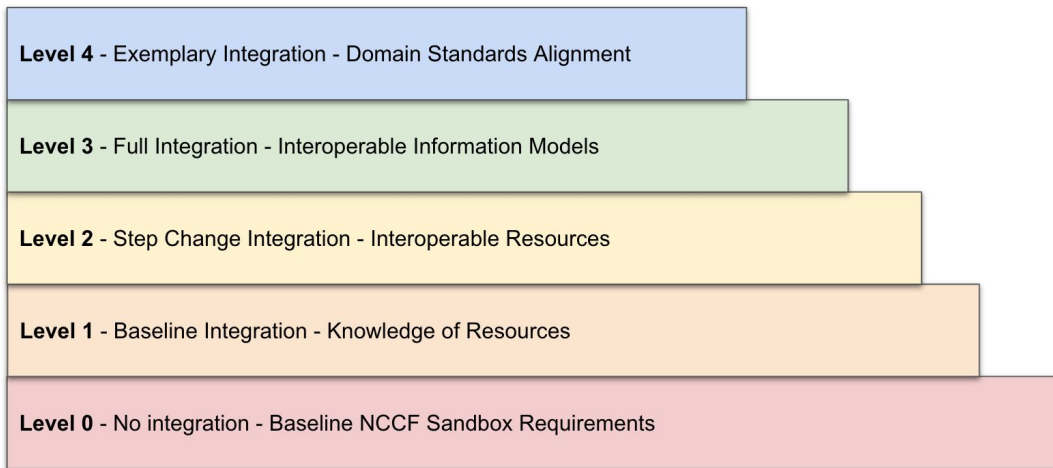
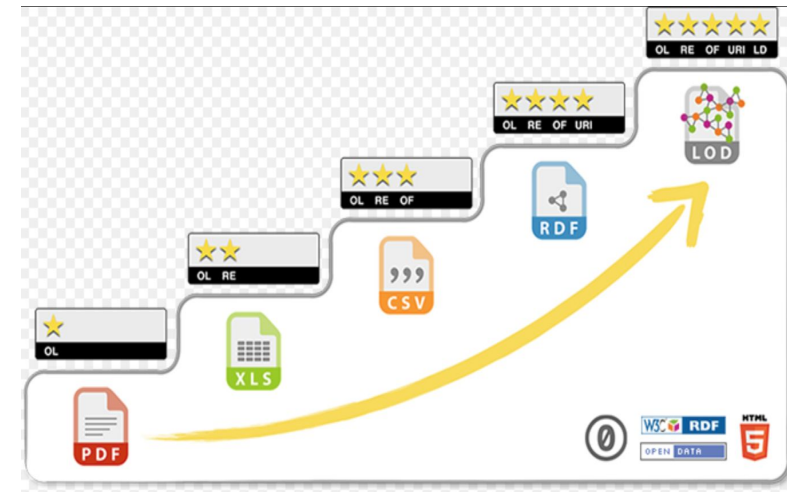
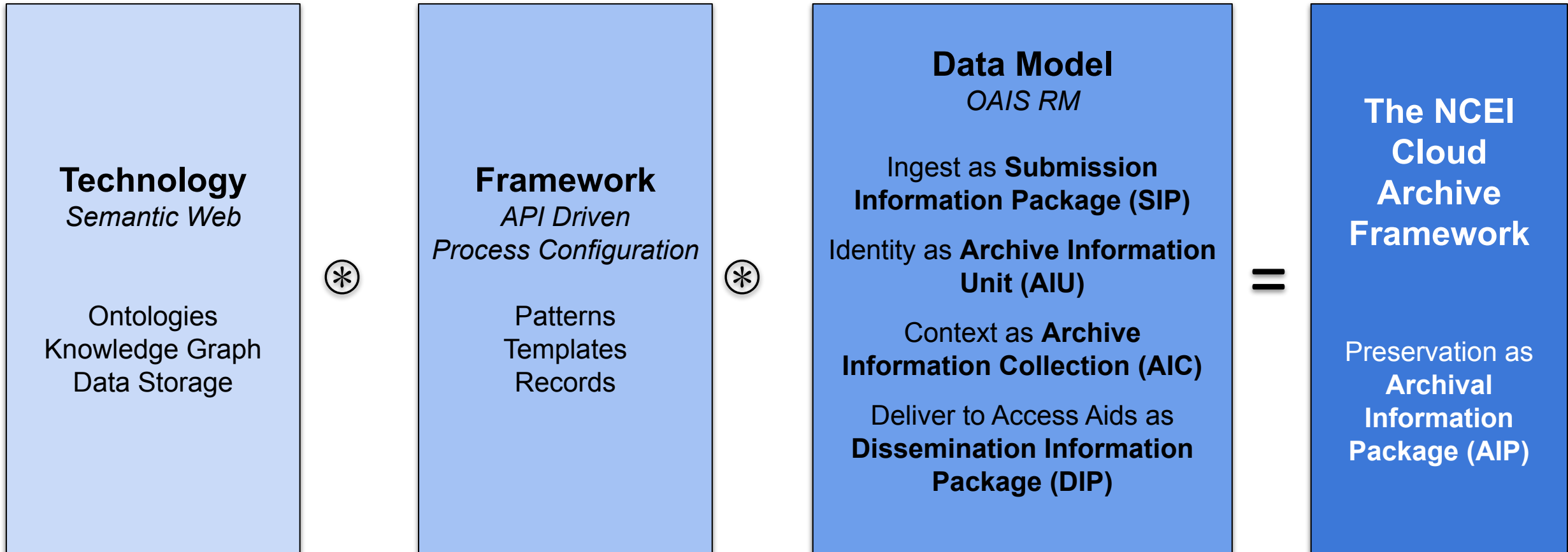


Figure 4. From data to information, knowledge and wisdom. Adapted from DIKW Model for knowledge management and data value extraction.

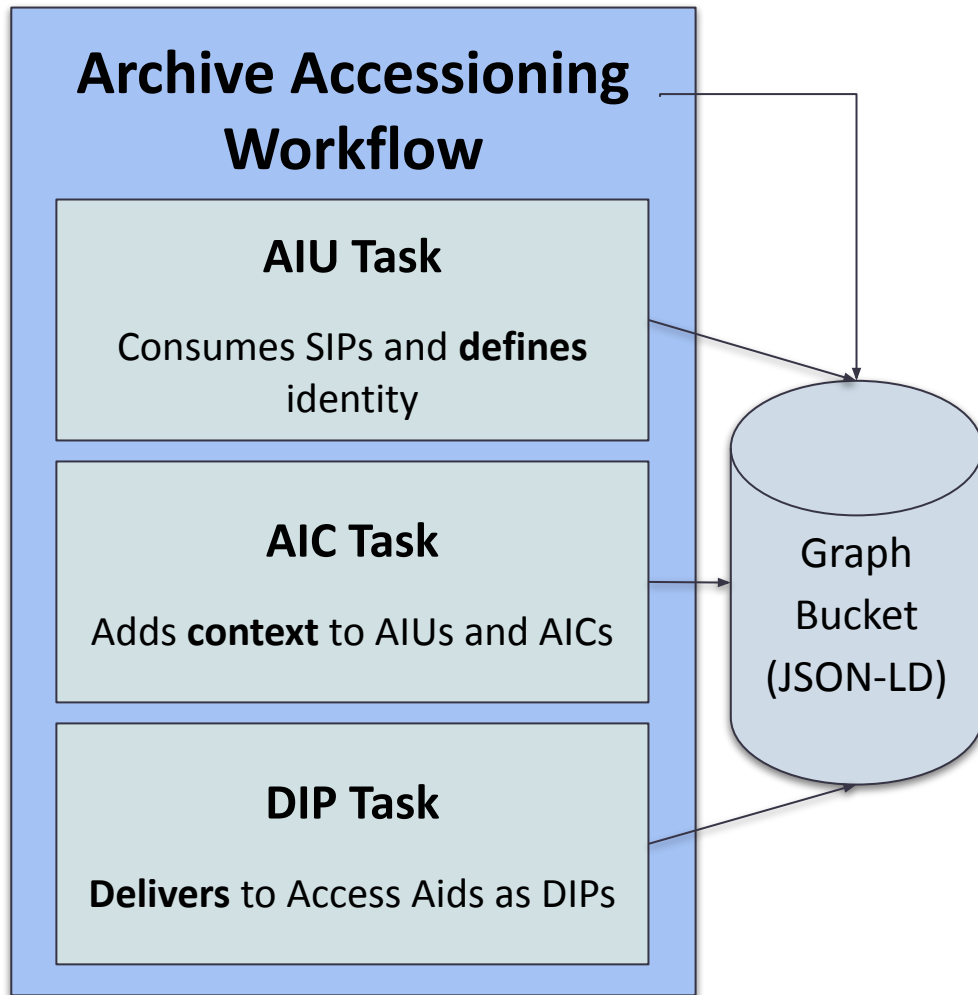


The Building Blocks



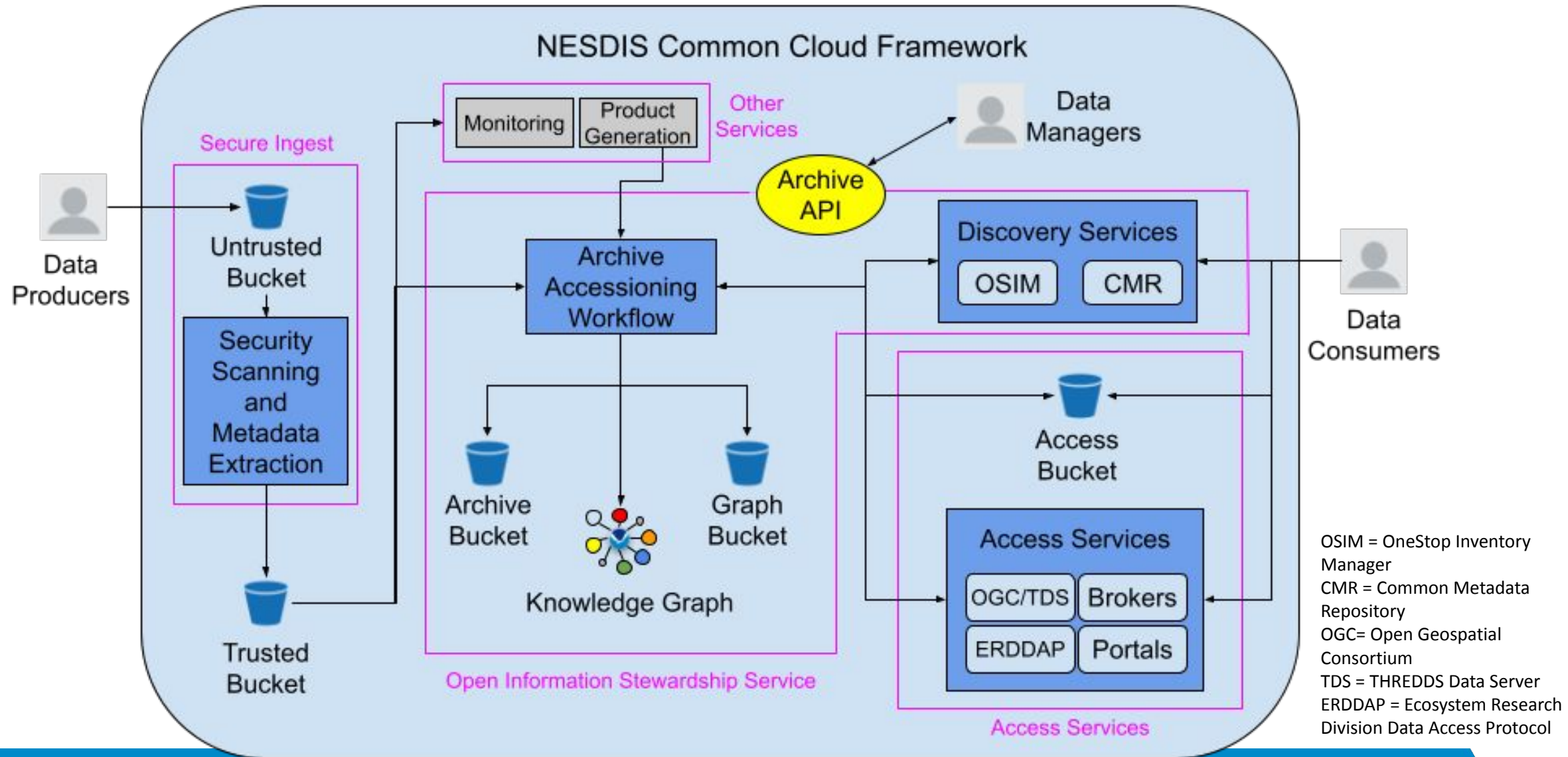
OAIS RM = Open Archival Information System Reference Model (ISO 14712)

Archive Accessioning Workflow - “Process as Data”



- AIU tasks define the identify of each object, consuming SIPs as they are ingested
- AIC tasks provide the context for AIUs, associating them with collections to enable long-term **preservation of knowledge** (context is key!!)
- DIP tasks deliver the content to access aids (AA), enabling data access and reuse
- Each step in the workflow writes its results as a record to the KG, where the **workflow and patterns themselves are also documented**

Archive Accessioning Workflow in the Broader Context

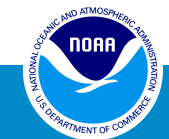


Slide on Steps Major Facilities Can Take/Opportunities

- Consider specific pathways laid out in the [UN Steps to Geoverse](#) transition tables
- Coordinate step change activities via cross-agency working groups to ensure interoperability runway
- Leverage existing knowledge and lessons learned from other large inter-agency efforts, like [NSF OKN](#), to identify ‘north star’ implementations
- Achieve early buy-in from leadership to ensure harmony and confidence in holistic agency change
- Align workforce development strategy with the idealized ‘open science culture’ end state
- Implement an iterative migration framework that exemplifies the Open Science characteristics desired by the solutions it facilitates



Backup



NCEI Tiers of Data Stewardship



Data Stewardship Services

Stewardship Planning

Coordinate with data providers early to set expectations, identify resources, and avoid rework

- Archive Recommendation Package
- Submission Agreements
- Cost Estimates
- General Education and Outreach
- Participation in NOAA and external forums to engage with the community

Archiving Services

Services providing during the course of archiving.

- NOAA Metadata Catalog
- Digital Object Identifiers (DOIs)
- Collection Manager Metadata Tools
- Data Stewardship Maturity Matrix (DSMM)
- Stewardship Support Desk

Data Access Services

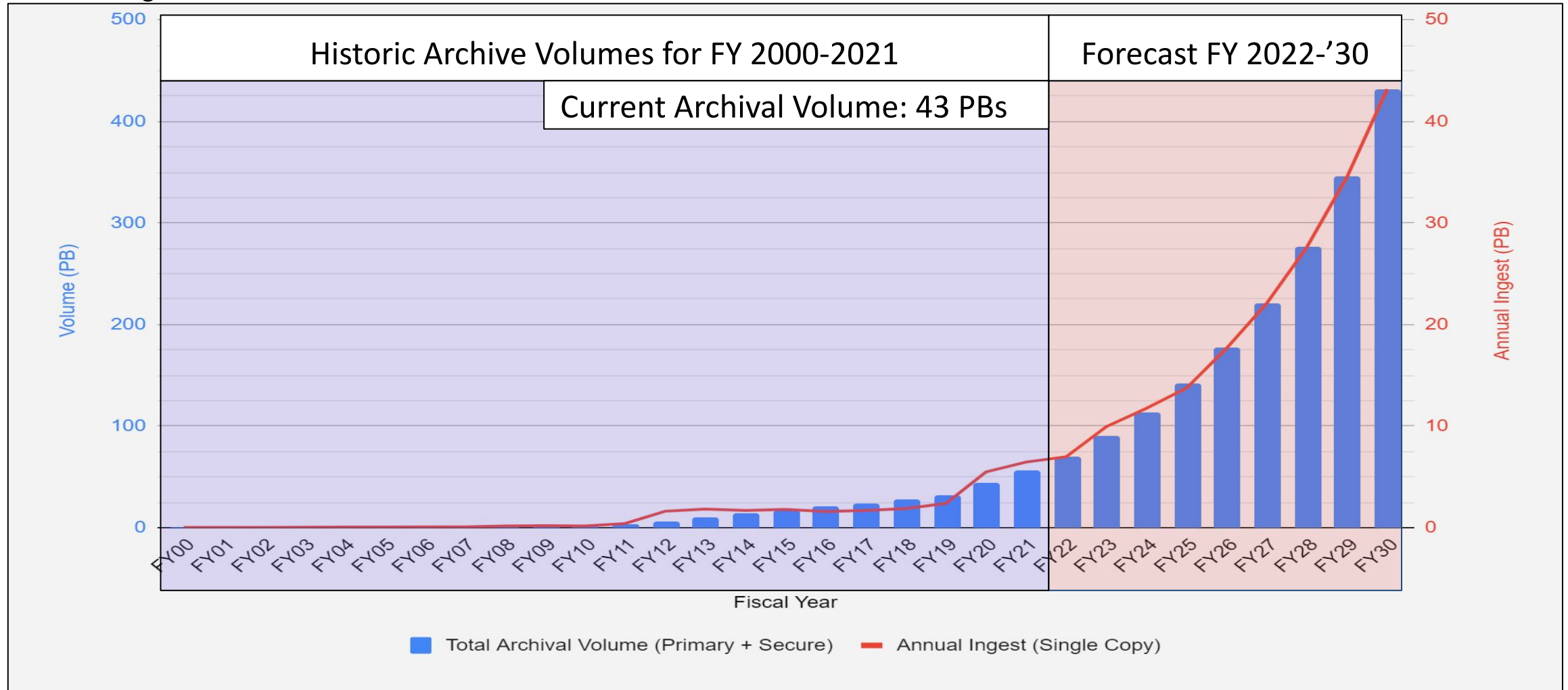
Data discovery and access services, including data interoperability and analysis-ready data

- Search/discovery, download (Tier 1)
- Delivery to external catalog and Schema.org
- Data and metadata APIs
- OPeNDAP and OGC services
- GIS map viewers
- Customized order output and certification



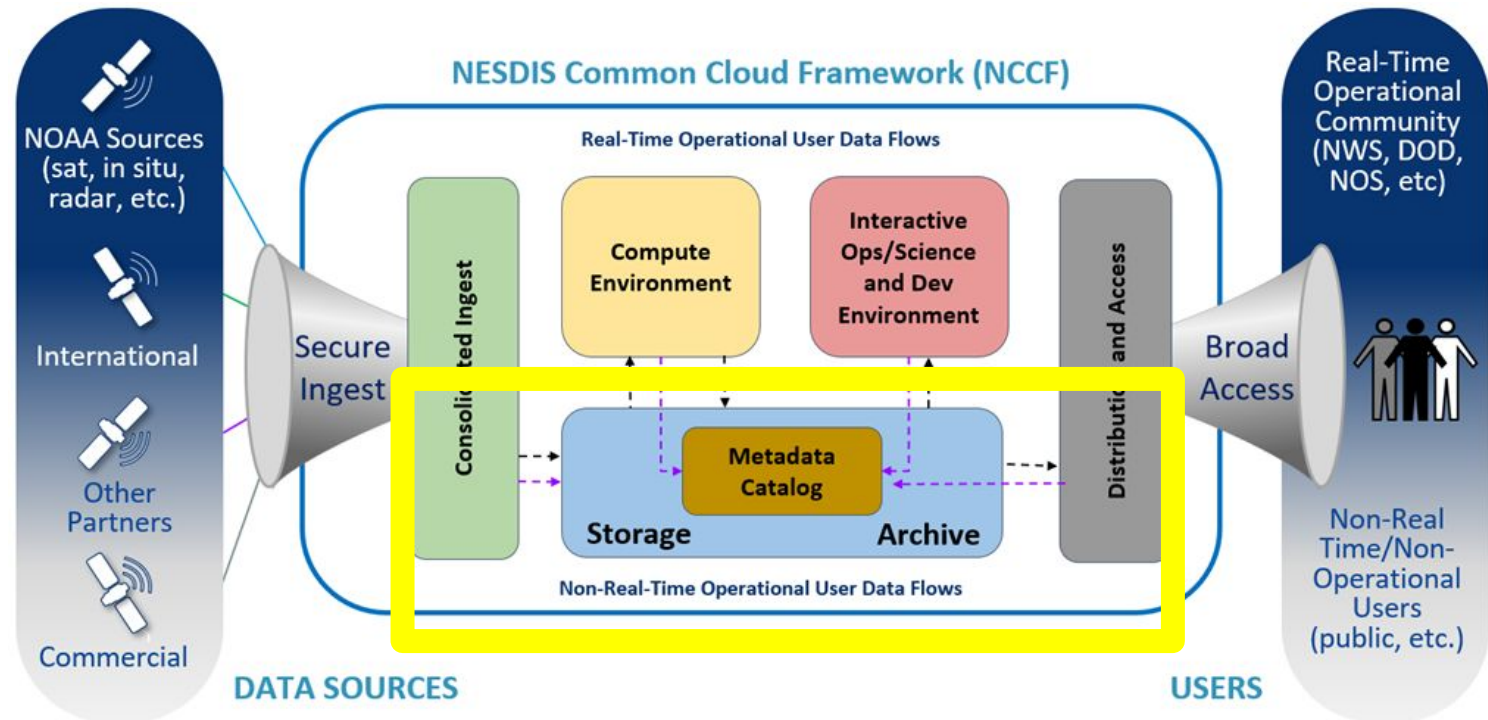
Archival Volume History and Forecast

Increasing Data Volumes from Station, Model, Radar, UxS, Acoustics, 'Omics, and Satellite Sources



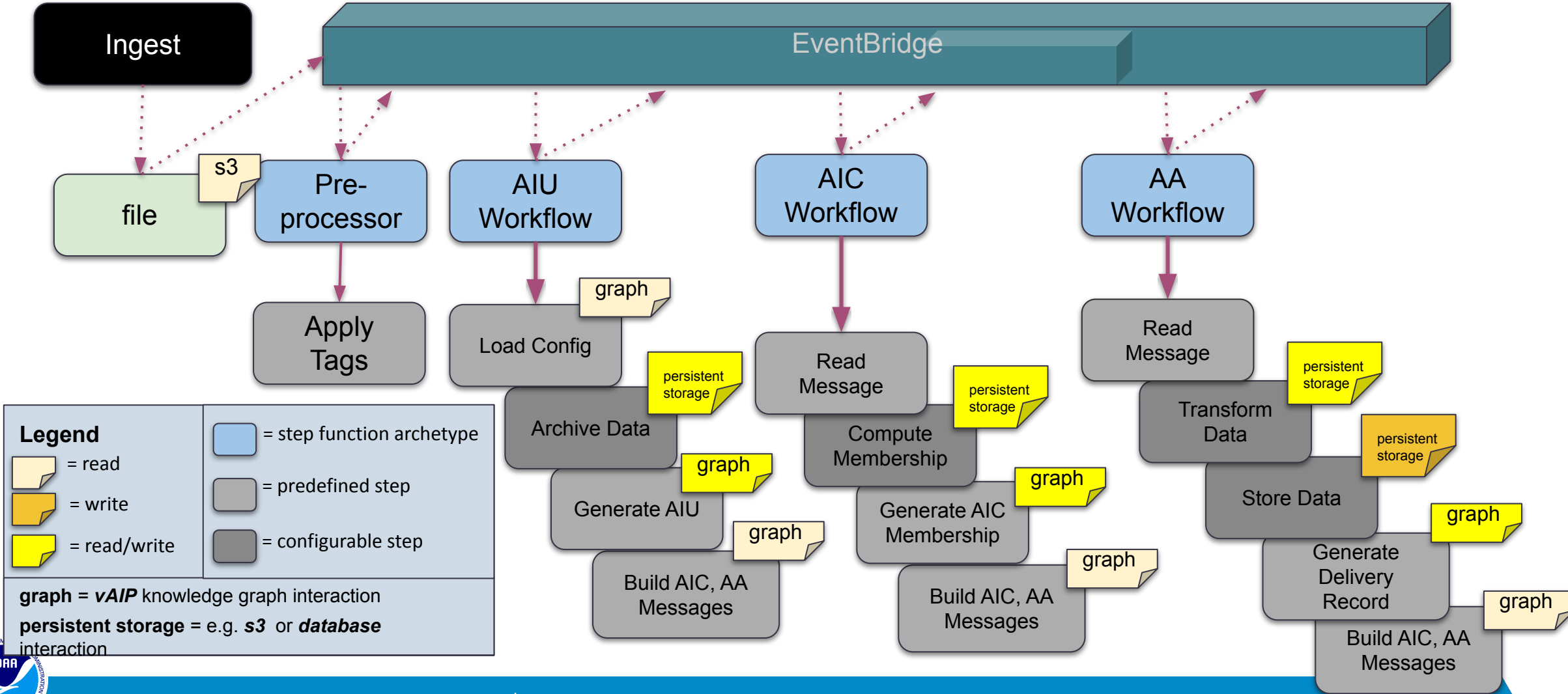
NCCF Architecture

- Cloud enterprise architecture with a multi-account and multi-Virtual Private Clouds (VPC) deployment
- Secure, fault tolerant, highly available, scalable architecture
- Loosely coupled, independently scalable services
- Implemented using Infrastructure as Code (IaC)
 - Reduced failover risk
 - Allows to minimize vendor lock in



- Operationally deployed for high availability workflows in the Amazon Web Services (AWS) commercial cloud
- Inclusion of automation to ensure secure deployment of new software and service updates

Cloud Workflow Leveraging Knowledge Graphs



Dual Model Architectures

- Enables the builder through the use of:
 - “*Ontologies* -> **Models**” and “*Archetypes* -> **Templates**”
 - To represent multiple contexts that leverage the same core model for data persistence.
- **Core data model** which represents a specific ontology which in this case is based on UML and concepts described in the **Open Archive Information System**.
- On top of that model we can **implement other models** (called *archetypes*) that represent themselves **within that context** such as **dataset specifications**.
- **SME driven** archive and access through **user provided**, instanceable data models, inherently interoperable through a small, **never-changing class set**, **manipulatable** through an **API** and user-provided code.

Finding and Accessing Data Today

Physical Archives - ncei.info@noaa.gov

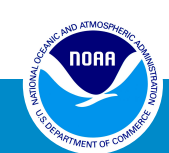
Data Discovery - OneStop

- Discovery of data across NOAA
- API Available
- <https://data.noaa.gov/onestop/>

Data Access - General Data Access

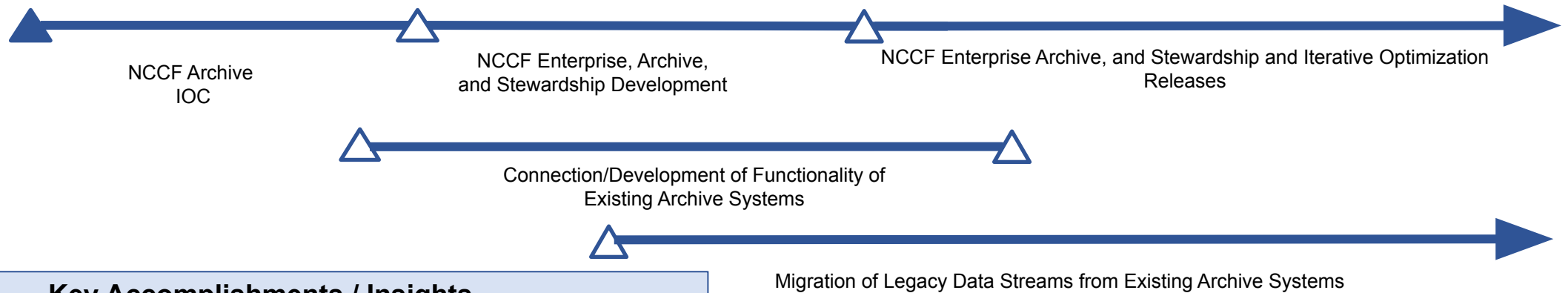
- Subsetting of select data sets, such as GOES-R Series Solar Ultraviolet Imager (SUVI) Level 2 Products
- API available
- <https://www.ncei.noaa.gov/access/search/index>

The image shows two screenshots of NOAA web portals. The top screenshot is the NOAA OneStop homepage, featuring a large image of Earth from space. The header includes the NOAA logo and the text 'OneStop NATIONAL OCEANIC AND ATMOSPHERIC ADMINISTRATION'. A search bar is visible with the text 'Search NOAA Data'. The bottom screenshot is the NCEI 'Data Access' page, featuring a background image of a forest. The header includes the NOAA logo and the text 'National Centers for Environmental Information NATIONAL OCEANIC AND ATMOSPHERIC ADMINISTRATION'. A search bar is visible with the text 'Search NCEI'. The page content includes the text 'Data Access' and a search bar with the text 'Ex: Global Summary of The Year' and a 'Search Datasets' button. There is also an 'Explore Datasets' button.



Cloud Archive Timeline

PI2	PI3	PI4	PI5	PI6	PI7	PI8	PI9	PI10	PI11	PI12	PI13	PI14	PI15	PI16	PI17	PI18	PI19	PI20	PI21	PI22	PI23	PI24	PI25	PI26	PI27	PI28	PI29
2021				2022				2023				2024				2025				2026				2027			



Key Accomplishments / Insights

- Transition from pilot to NCCF in progress
- NCAP - Piloted archive capability in NCCF within 12 months
- Initial Operational Capability for Archive end of Q2FY23



ConOps Vision

General Notes
 -Processes are user defined, API validates
 -Executable task business logic is either user provided, through CI/CD or is a remote API call to run in federation; may be long running and/or human-in-the-middle (HITM)
 -All Tasks publish to Metadata Bucket on Task resolution
 -Process Requests shown are inbound or outbound and made by passing a predefined message to API to trigger predefined process

