

Challenges and Successes of Implementing NOAA/NESDIS/NCEI's Ingest of Archival Data in a Cloud Environment

Steven B. Rutz¹ (Steven.Rutz@noaa.gov), Kenneth S. Casey¹, Joseph T. Mangin², and David Fischman¹

NOAA/NESDIS ¹National Centers for Environmental Information and ²Office of the Assistant Chief Information Officer – Satellites

1. Abstract

Since its formation in 2015, the National Centers for Environmental Information (NCEI) has used disparate, legacy systems spread across several IT networks of the National Environmental Satellite, Data, and Information Service (NESDIS) to fulfill its data-stewardship functions. As part of its modernization and consolidation of these functions, NCEI implemented Common Ingest as the functional component that ingests approximately 200 data streams every month into its enterprise archival information system. In

parallel, NESDIS completed the Secure Ingest Gateway Project (SIGP), a pilot project to establish standard-enterprise secure methods for NESDIS and the rest of the National Oceanic and Atmospheric Administration (NOAA) to receive data in a cloud environment from their external partners. SIGP is now transitioning to operations as the Operational Secure Ingest Service (OSIS), which will be the on-ramp to NCEI's "Common Ingest" functionality when it too moves to the cloud. In addition, this ingest function will populate

and use a cloud-based metadata catalog, which will be the beating heart of the NESDIS and NCEI information systems in the cloud environment. The vision is to scale the ingest of environmental data to keep pace with its ever increasing volume, veracity, variety, and velocity. In this presentation to the ocean data community, the authors describe NESDIS and NCEI's challenges and successes with the implementation of the ingest function of their archival information system in a cloud environment.

2. Ingest in an Open Archival Information System

What does ingest have to do with data stewardship? The functional components of NCEI's archival information system, including the ingest component, all map to the standard for an Open Archival Information System (OAIS; International Organization for Standardization 14721:2012; Figure 1). In this standard, Ingest is defined as the functions and services that (1) accept Submission Information (or Data) Packages (SIPs) from Producers (Figure 2), (2) prepares Archival Information Packages (AIPs) for Archival Storage (Figure 3), and (3) ensures that the AIPs and their supporting descriptive information (metadata) become established within the OAIS, or what NCEI often refers to as The Archive. Consumers can then access the Dissemination Information Packages (DIPs) to use in their products (Figure 4).

Open Archival Information System (OAIS)

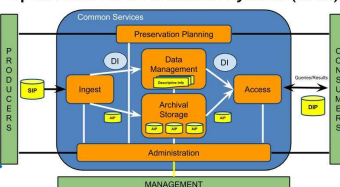


Figure 1

AIP in Archival Storage

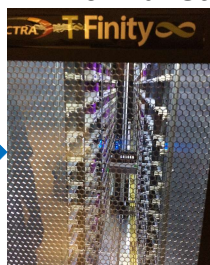


Figure 3

SIP

DIP

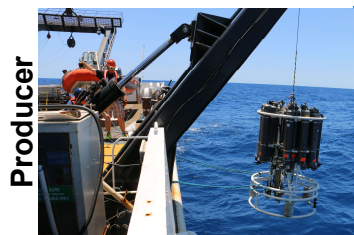


Figure 2

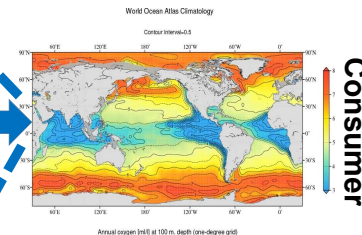


Figure 4

NCEI Open Data Framework

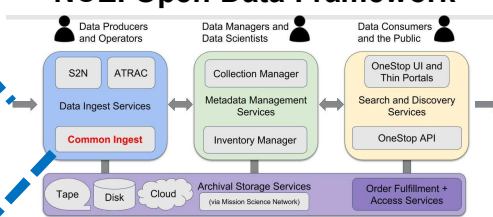


Figure 5

Cloud Framework

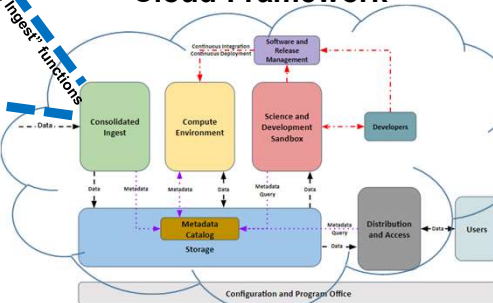


Figure 7

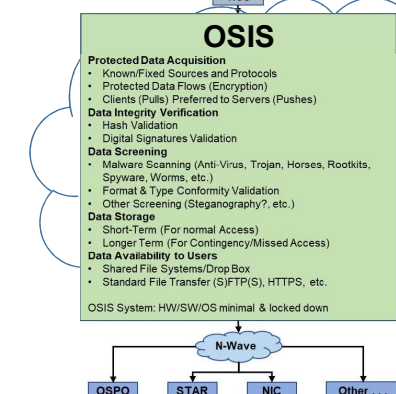


Figure 6

5. Operational Secure Ingest Service (OSIS)

In parallel to NCEI's Common Ingest Migration Project, NESDIS completed the Secure Ingest Gateway Project (SIGP), a pilot project to establish standard enterprise-secure methods for NESDIS and the rest of NOAA to receive data in a cloud environment from their external partners such as from the China Meteorological Administration and the U.S. Geological Survey. SIGP is now transitioning to operations as the Operational Secure Ingest Service (OSIS; Figure 6).

The OSIS mission is to provide a secure cloud-based, enterprise-level, 24/7-operational service that will:

- Ingest large data flows safely from multiple external data providers (international partners, commercial, academia, etc.),
- Securely screen the data, and
- Provide secure access for retrieval of data to internal users in a timely manner.

6. Cloud Framework: Consolidated Ingest

In another effort, NESDIS initiated the Cloud Integrated Project Team and pilot project to establish agile and scalable services that improve the efficiency of service deliverables and data ingest from all sources. Phase I of the pilot project, which is in progress through April 2020, includes providing a framework for operating securely in a cloud environment (Figure 7). Consolidated Ingest, a component of the Cloud Framework, will move environmental data to cloud-based Storage and populate and use a cloud-based Metadata Catalog – the beating heart of the future NESDIS information system.

In this future Consolidated Ingest, OSIS will be the on-ramp to "Common Ingest" functionality when NCEI's archival information system moves to the cloud. NESDIS and NCEI's vision is to scale their Ingest functions and services to keep pace with the ever-increasing volume, veracity, variety, and velocity of environmental data.

3. Introducing NCEI's Common Ingest

Flying up to a 10,000-foot view of the Open Data Framework (NCEI's implementation of the OAIS; Figure 5), Ingest, and NCEI's Common Ingest in particular, maps to Data Ingest Services. Common Ingest brings SIPs containing environmental data into NCEI and then prepares and passes each SIP's data and associated metadata on to the other components of the system, including Metadata Management and Archival Storage Services.

Why is NCEI using Common Ingest? Early in 2015, NOAA merged their three National Data Centers to form NCEI. Each of the three had their own systems to ingest environmental data into the Archive, which made it difficult to have a cohesive process. In 2016, NCEI evaluated the existing ingest systems and recommended the former National Geophysical Data Center's relatively new Common Ingest system as NCEI's enterprise ingest system.

4. NCEI Common Ingest Migration Project

In FY2016, NCEI embarked on a project to:

- Port the Common Ingest system to the NCEI-mission network and add functions necessary for ingesting unique archival data streams such as NOAA Next Generation Radar (NEXRAD) Level 3 Products from the National Weather Service (<https://data.nodc.noaa.gov/cgi-bin/iso?id=gov.noaa.ncdc:C00708>), and
- Migrate all the archival data streams from a legacy ingest system to the Common Ingest system.

In August 2017, Common Ingest passed an Operational Readiness Review and began ingesting data into archival storage. By July 2019, NCEI had migrated all of the 141 active archival data streams from the legacy system on the NCEI-mission network to the Common Ingest system. In addition, during this period, they began ingesting 52 new data streams (i.e., data streams that NCEI never ingested with the legacy system). Common Ingest was a success!

However, during the migration, numerous file system problems arose that limited Common Ingest's performance. To address these performance problems and to better prepare NCEI for the next step in the evolution of its archival information system – using cloud-based services – NCEI is implementing for Common Ingest data the use of object storage – an architecture used by most commercial cloud services.

7. Acknowledgements

NESDIS and NCEI have presented these individuals with awards for their work on the NCEI Common Ingest Migration Project: Michael Allen, Jeff Bliss, David Bowman, Heather Brown, Jon Burroughs, Linda Copley, Ryan Cox, John Fauerbach, Evan McQuinn, Dave Neufeld, Pichai Polprasert, Ron Ray, Ken Roberts, Steven Rutz, Ken Schmidt, Richard Smith, Jason Symonds, Mike Urzen, and Lou Vasquez. The U.S. Department of Commerce presented these individuals with a Gold Medal Award in 2019 for their work on NESDIS SIGP: Kate Becker, Michelle Detommaso, Matthew Jochum, Chi Kang, Joseph Mangin, Irene Parker, Robert Sears, Cameron Shelton, Doug Whiteley, and James Yoe.

