

# NOAA Update

# Data Preservation and Stewardship

**Nancy Ritchey, Archive Branch Chief, NCEI Data Stewardship Division**  
**Kenneth S. Casey, PhD, Deputy Director, NCEI Data Stewardship Division**

Thursday, 22 Mar 2021  
WGISS-51  
Virtual



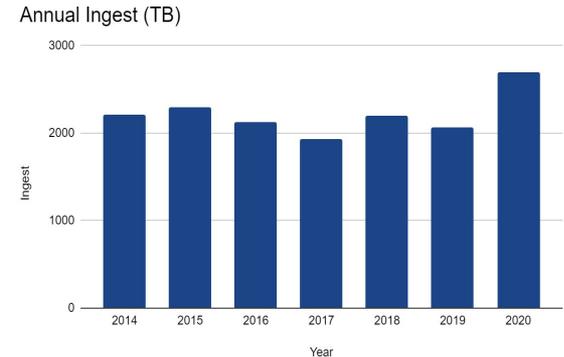
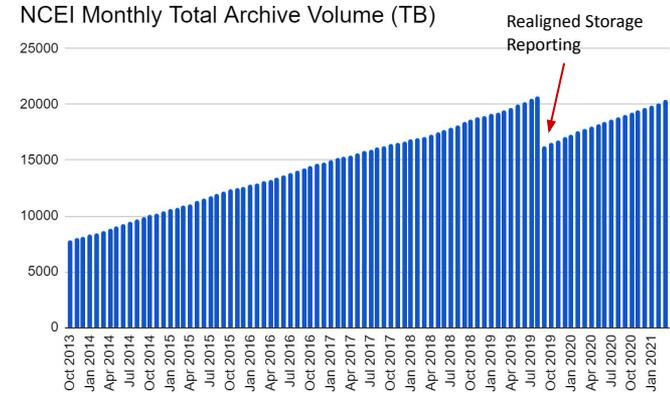
# Outline

- Archive Holdings
- Current Archiving Framework
- Future Cloud Workflows
- Challenges



# NCEI - By the Numbers

- NCEI formed in 2015
- Data currently stored across 5 systems
- Realigned storage reporting in 2019 for consistency
- Annual ingest of 2-2.5PB



# NCEI - Data Preservation

- Heterogeneous data types supporting the NOAA mission ([Archive Collecting Policy](#))
  - Large volume satellite, model and bathymetry, etc. data
  - Multi-Instrument event-related collections (cruises, field studies, hurricane flights)
  - Majority are small to med. collections consisting of 1 - 100s of records
  - Operational data streams and ad hoc, single submissions
  - L0 - L4, observations, ancillary, auxiliary, pre- and post- launch testing
- Heterogeneous spatial and temporal boundaries
  - Paleoclimate to current observations
  - Space weather to seafloor mapping
- Heterogeneous file formats ([Preferred File Formats](#))
  - MPEG, TIFF, JPG, BUFR, netCDF, HDF, ASCII, PDF, MSWord/Excel
  - Paper, film, microfiche, 9- and 7-track tapes



# NCEI - Information Preservation

- Climate Data Records require archival of software with workflow diagrams and detailed algorithm documentation
- Physical records are preserved until they are digitized
- Technical reports related to data and programs are preserved by the NOAA library
  - Collaborate to cross-link data and publication DOIs
- Draft NOAA Software Governance and Public Release Policy currently in review
  - Requires software to be publicly available and have a DOI



# NCEI - Data Packages

- Data packaging options currently supported:
  - Zip, Gzip, TAR, Bagit
- Package sizes related to:
  - User community needs
    - Spinning disc for small files with quick access needs
    - Daily 'tarballs' of most recent satellite data on spinning disc
  - Data Production system capabilities - not currently aligned with best practices
  - Storage system configuration is optimized for fast access
    - Spinning disc for collections to meet user needs
    - Tape for larger collections and for long-term storage
- Looking into Cloud- and AI-ready formats and packaging



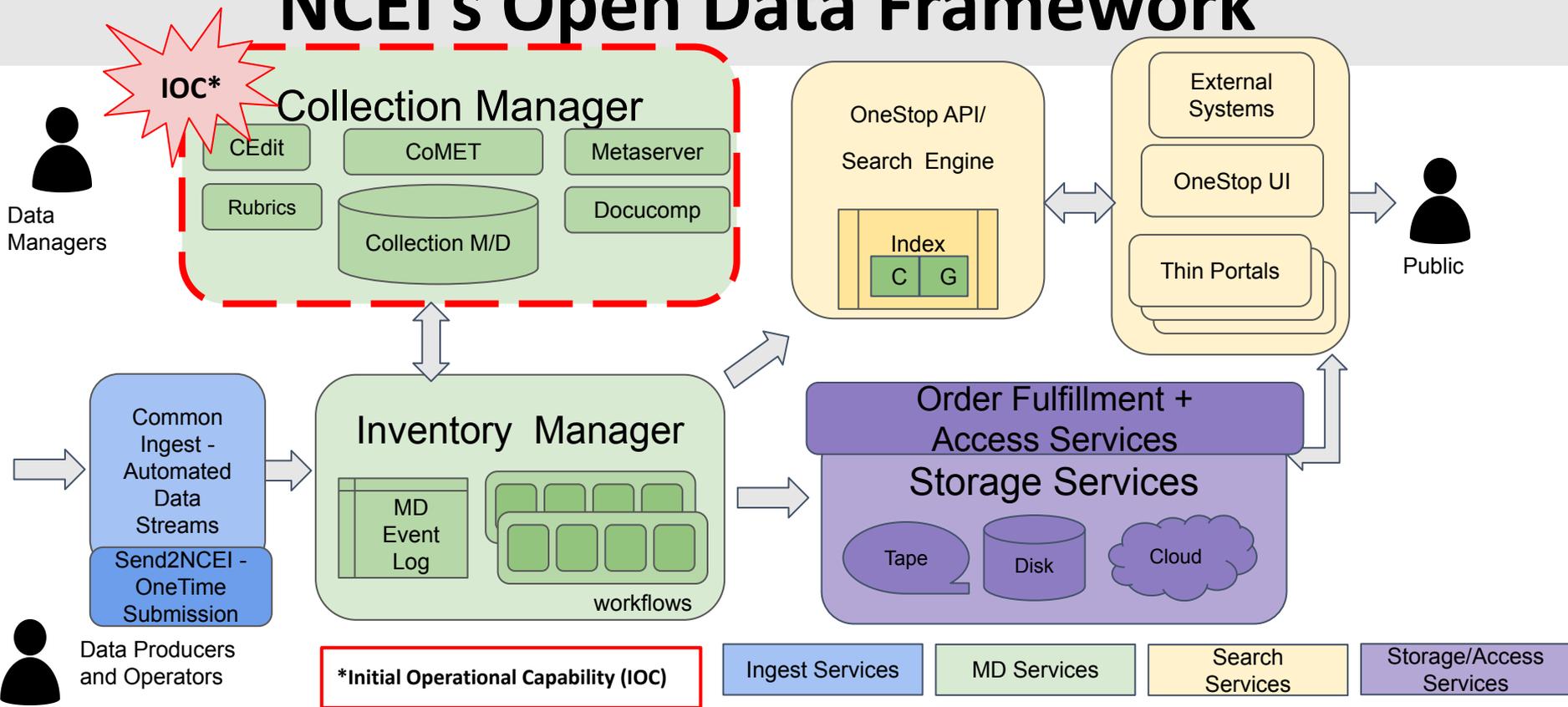
# NCEI - Today

A wide-angle photograph of a lush green wetland or marsh area. In the foreground, there are tall green reeds and lily pads in a shallow pond. The middle ground shows a winding waterway through dense vegetation. In the background, there are some buildings and utility poles under a clear sky.

**NCEI's set of stewardship systems, products, services, and people functioning together in a symbiotic way can be described as an ecosystem**

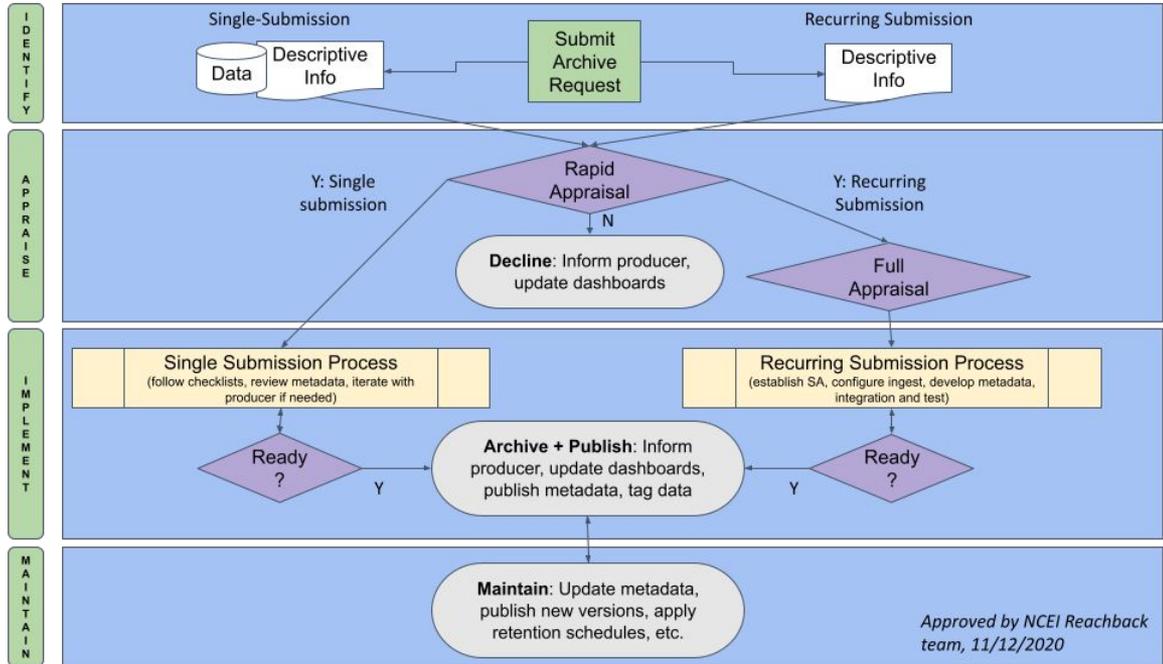


# NCEI's Open Data Framework



# NCEI - Cloud Archive Workflow + Tech

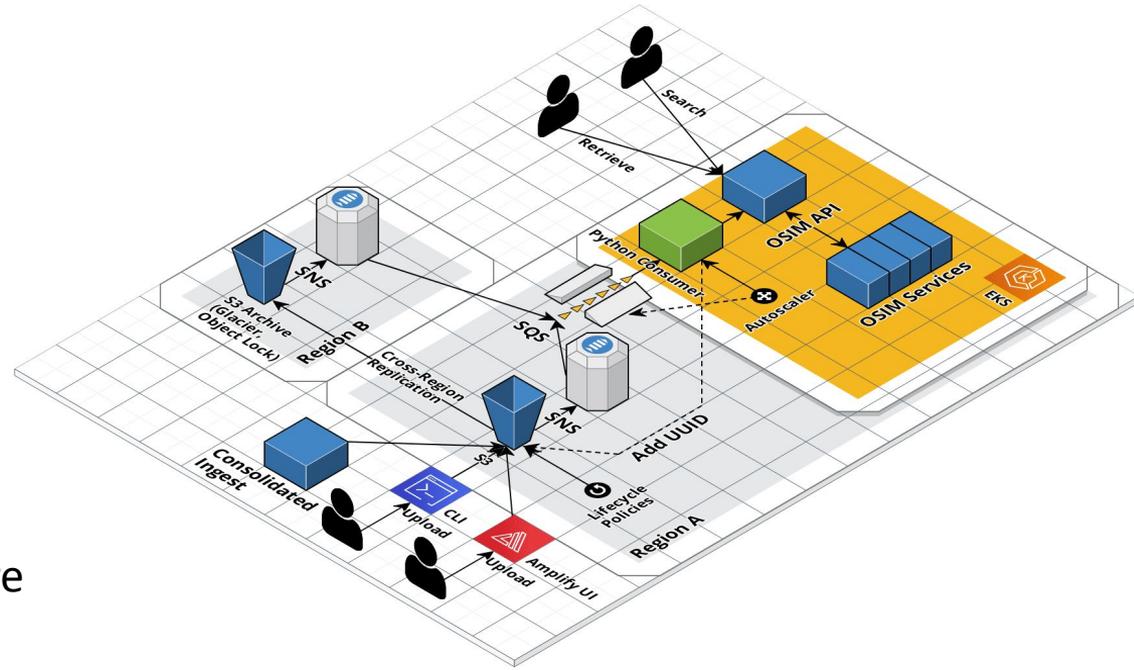
- Working across NESDIS, developed a unified cloud archive workflow
- Unites multiple on-prem systems for managing both single and recurring archive submissions in the cloud
- Prototyping underway, with significant progress on backend functions in AWS



# NCEI - Cloud Archive Workflow + Tech

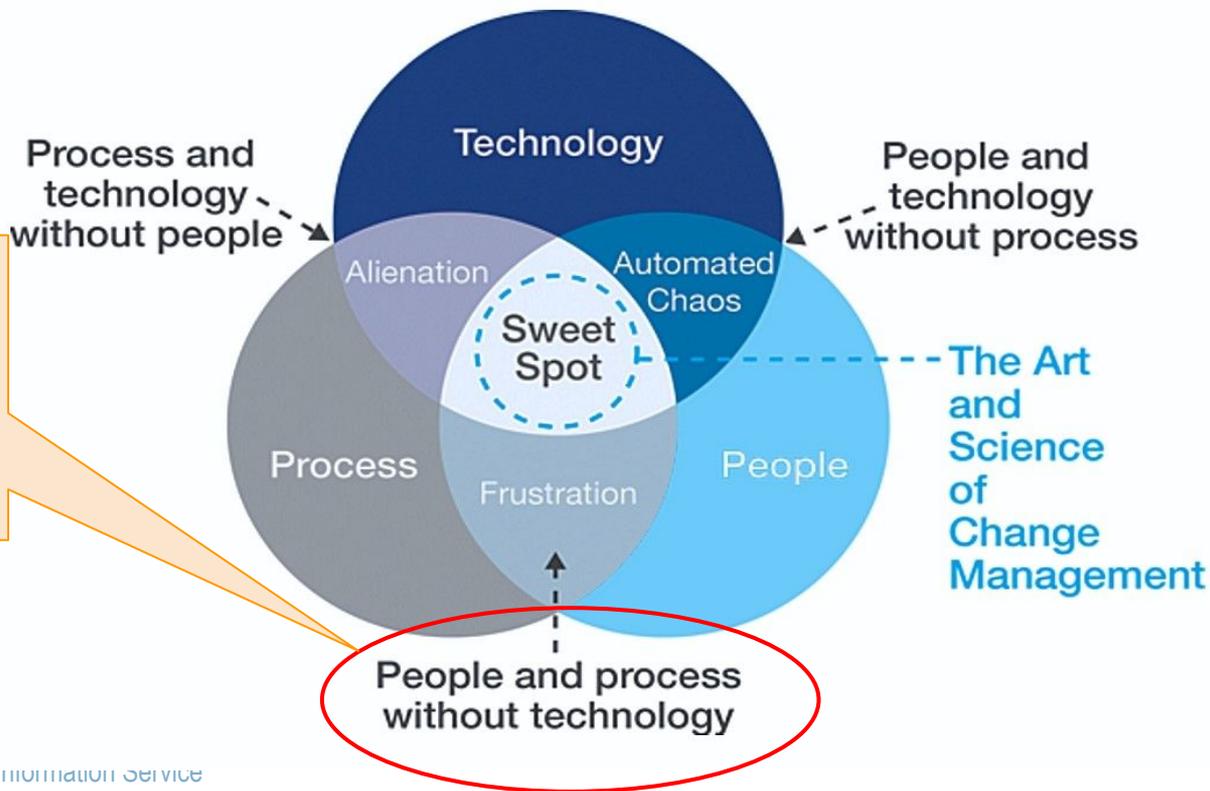
Backend functions ensure that:

- Producers can be notified of updates to their submissions
- Appropriate records retention schedules can be applied to the data
- Data are inventoried rapidly, assigned a unique identifier, and copied to redundant storage (cross-region replication) to ensure resilience to regional scale disasters



(Diagram courtesy of Evan McQuinn)

# Business Readiness: People, Process, and Technology

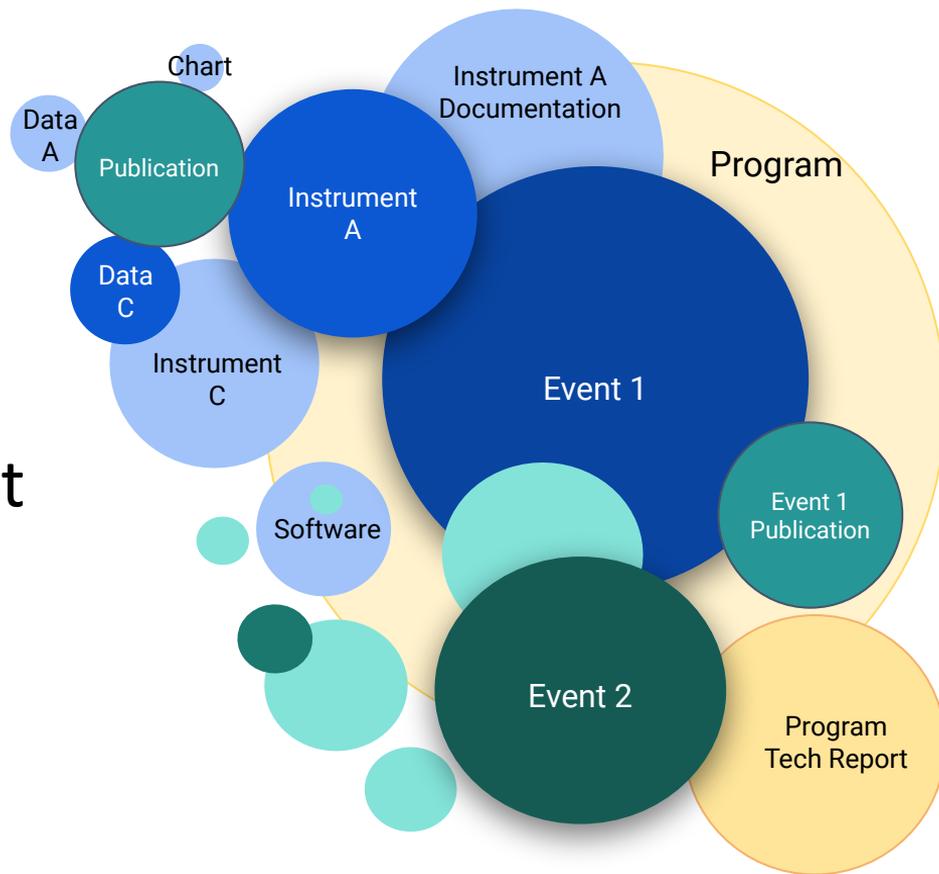


The challenge of simultaneously developing a new system while maintaining old systems

# Challenges

How do we ensure sufficient citation and linking of assets?

Thoughtful community best practices need to be developed and adopted





# Thank You!