

NOAA's Next-Generation Cloud Archive

Kenneth Casey¹, Monica Youngman¹, Nancy Ritchey¹, John LaRocque², Rich Baldwin¹, Ryan Berkheimer¹, Jeff Arnfield¹, Richard Smith¹, Jeremy Hall¹, Drew Saunders¹, and Steven Rutz¹

¹NOAA National Centers for Environmental Information

²National Centers for Environmental Information

November 21, 2022

Abstract

In the two years since the Fall 2019 meeting, NOAA/NESDIS completed two significant projects whose goal was to prototype an enterprise data management system in the cloud. The two pilot efforts, run by the NESDIS Cloud Integrated Project Team, were completed in March of 2020 and March of 2021. A significant portion of those pilot efforts has been operationalized into what is known as the NESDIS Common Cloud Framework (NCCF). The NCCF is now up and running and capable of securely ingesting data and generating operational products. Another effort, the NESDIS Cloud Archive Project, picked up the unfinished work of the NESDIS cloud pilots and is focused on delivering an end-to-end cloud archive prototype in late 2021, with the goal of operationalizing that cloud archive capability into the NCCF in 2022. The current status of these efforts will be presented, including a technical focus on how NESDIS plans to archive data using commercial cloud services.

NOAA's Next-Generation Cloud Archive

Kenneth S. Casey, Monica Youngman, Nancy Ritchey, John LaRocque, Rich Baldwin, Ryan Berkheimer, Jeff Arnfield, Richard A. Smith, Jeremy Hall, Drew Saunders, and Steven Rutz, NOAA National Centers for Environmental Information (NCEI)

Abstract

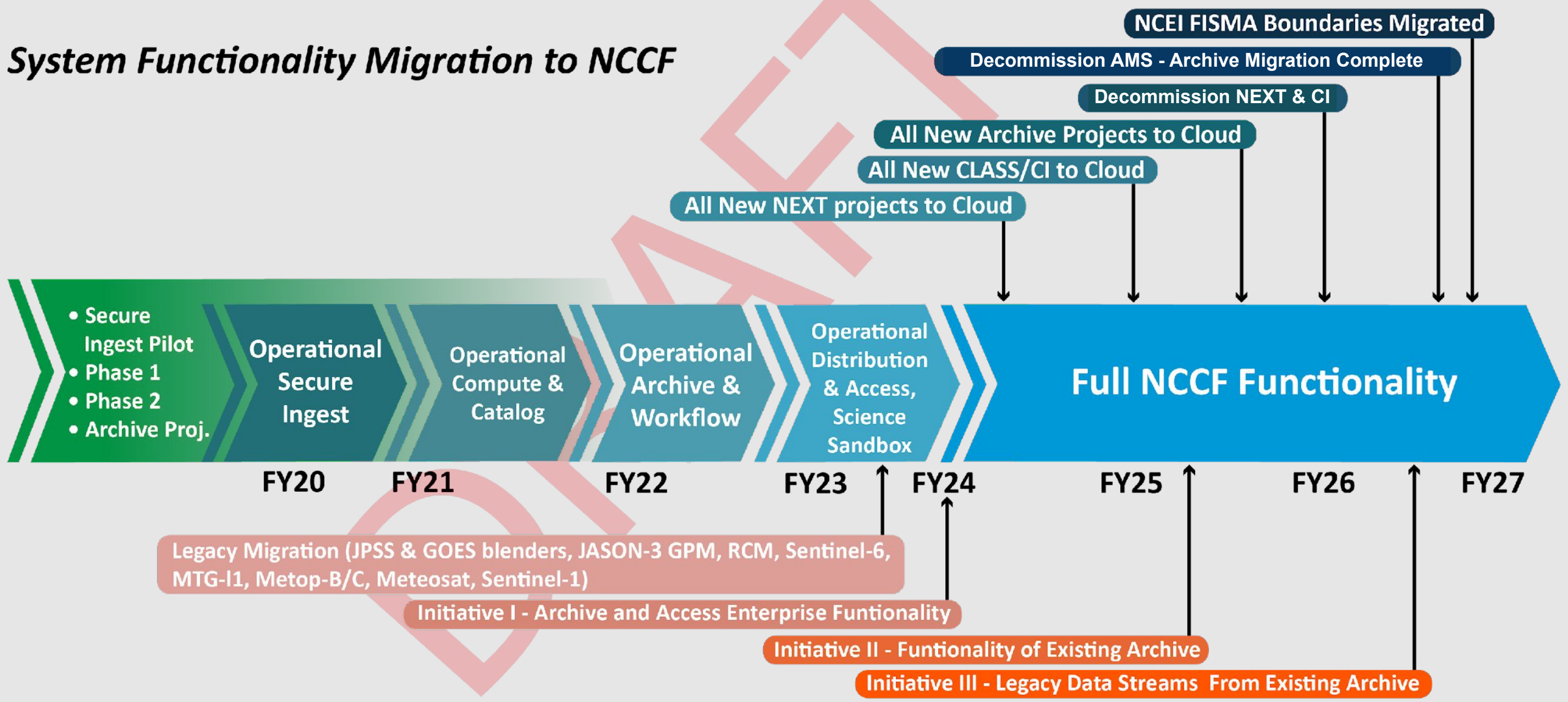
In the two years since the Fall 2019 meeting, NOAA/NESDIS completed two significant projects whose goal was to prototype an enterprise data management system in the cloud. The two pilot efforts, run by the NESDIS Cloud Integrated Project Team, were completed in March of 2020 and March of 2021. A significant portion of those pilot efforts has been operationalized into what is known as the NESDIS Common Cloud Framework (NCCF). The NCCF is now up and running and capable of securely ingesting data and generating operational products. Another effort, the NESDIS Cloud Archive Project (NCAP), picked up the unfinished work of the NESDIS cloud pilots and delivered an end-to-end cloud archive prototype in September 2021, with the goal of operationalizing that cloud archive capability into the NCCF in 2022. The current status of these efforts will be presented, including a technical focus on how NESDIS plans to archive data using commercial cloud services.

AWS Technologies in Use

AWS Service	Service Category	AWS Service	Service Category
OpenSearch Service	Analytics	Budgets	Financial Management
EventBridge	App. Integration	CloudTrail	Mgmt & Governance
Simple Notification Service	App. Integration	CloudWatch	Mgmt & Governance
Simple Queue Service	App. Integration	Config	Mgmt & Governance
Step Functions	App. Integration	Systems Manager	Mgmt & Governance
Simple Workflow Service	App. Integration	EC2-ELB	Network/Content Del.
EC2-Instances	Compute	VPC	Network/Content Del.
Lambda	Compute	Identity & Access Management	Security, Identity, Compl.
EC2 Container Registry	Containers	Key Management Service	Security, Identity, Compl.
ECS for Kubernetes	Containers	Secrets Manager	Security, Identity, Compl.
Elastic Container Service	Containers	Backup	Storage
DynamoDB	Database	Elastic File System	Storage
Relational Database Service	Database	Glacier	Storage
SimpleDB	Database	Simple Storage Service	Storage
CodeBuild	Developer Tools	Neptune	Knowledge Graph
CodeCommit	Developer Tools		

NOAA is Moving its Archive to the Cloud

NCEI's Proposed Cloud Archive Implementation Roadmap



Consolidated Cloud Archive Workflow

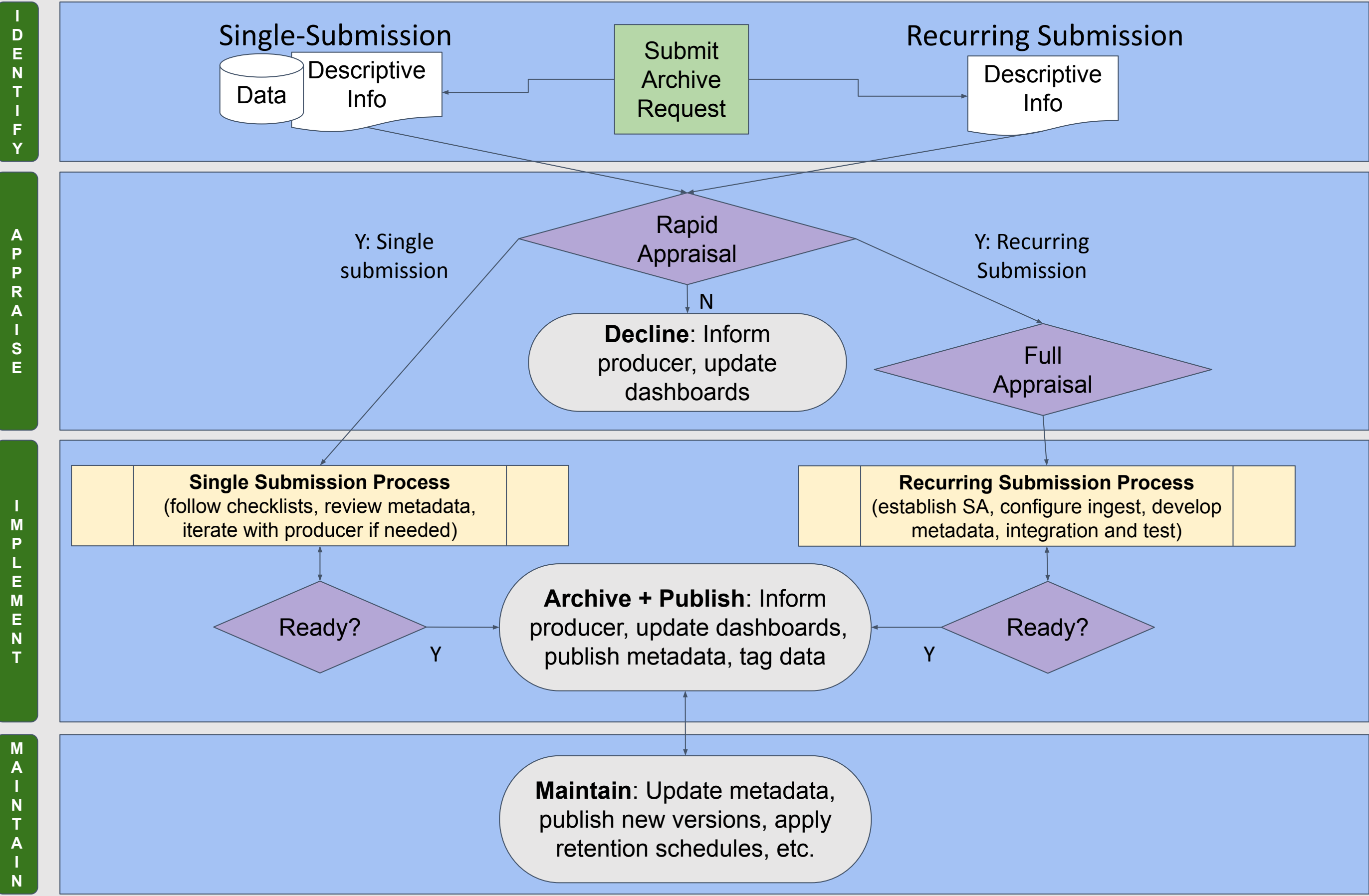


Figure 1: Consolidated Cloud Archive - This figure illustrates the high-level view of the consolidated cloud archive workflow, combining the two primary archive paradigms (single submissions and recurring submissions) into a unified workflow. This unified workflow will be managed using the enterprise cloud archive system to be deployed in the NCCF.

Process Archetypes

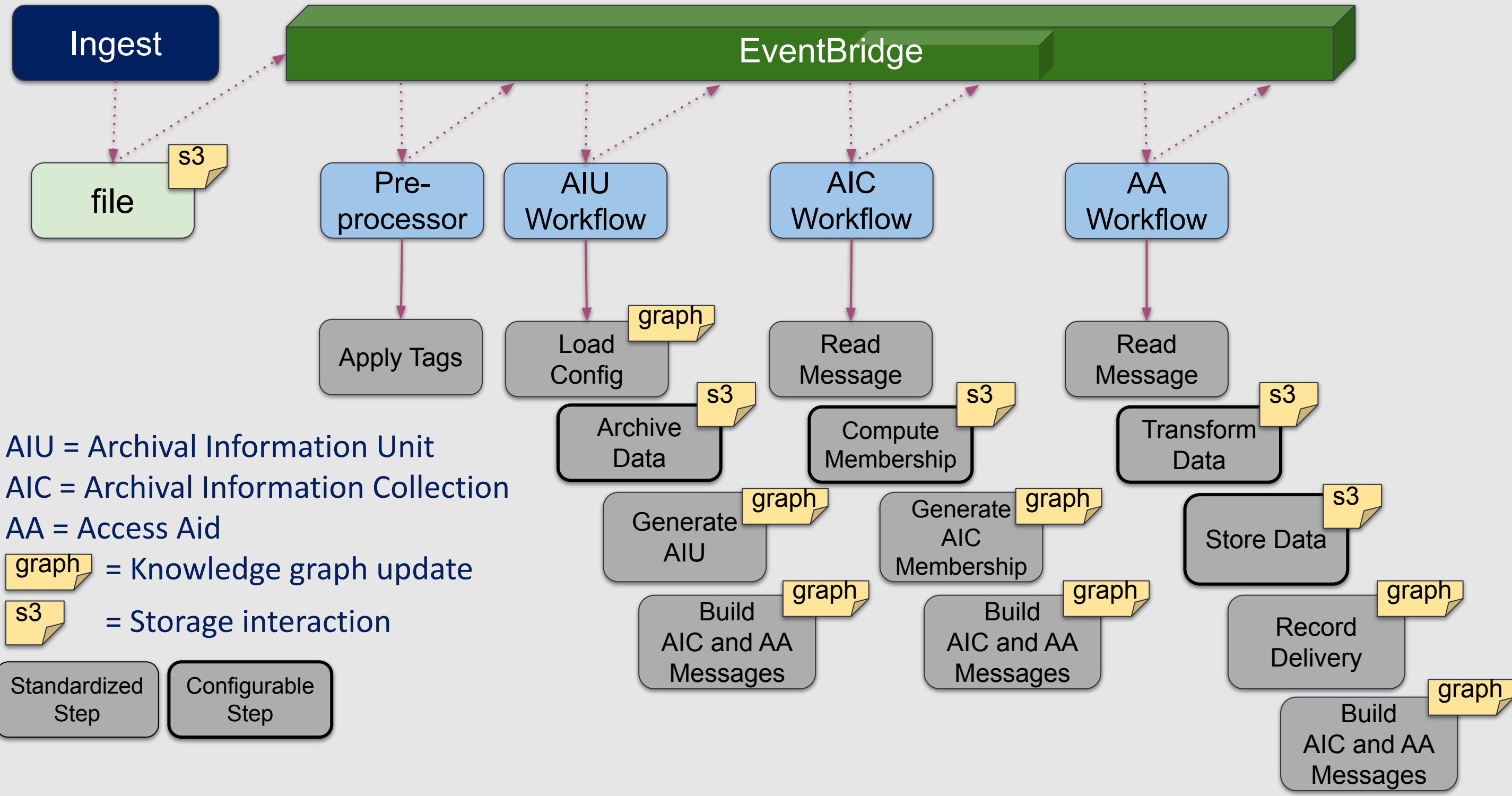


Figure 2: Process Archetypes - This figure illustrates how process archetypes or templates are used to coordinate specific workflows using AWS Eventbridge. This approach simplifies, adds rigor, and ensures consistency between different instances of the workflows and ensures archival requirements are being met for all data ingested into the NCEI cloud archive. Shown are example templates for Archival Information Unit (AIU) processing, Archival Information Collection (AIC) process steps, and Access Aid (AA) workflows. Storage interactions with AWS S3 are noted for specific steps in the workflow, as are updates to the archive knowledge graph, currently implemented using AWS Neptune.

Virtual Archival Information Package (vAIP)

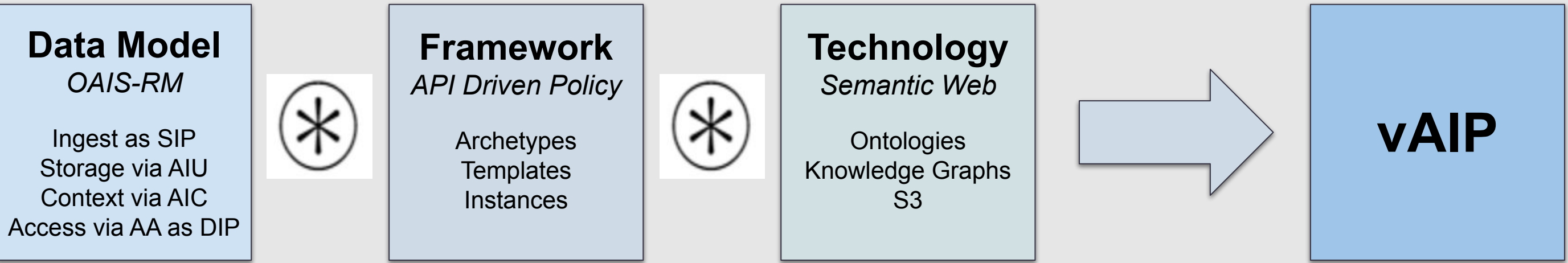


Figure 3: Virtual Archival Information Packages (vAIP) - Supporting the NCEI cloud archive is the vAIP, which is the result of implementing specific technology choices like AWS S3 for Storage and Neptune for knowledge graphs, within an API driven policy framework of archetype templates aligned to the data model of the Open Archival Information System Reference Model (OAIS-RM).