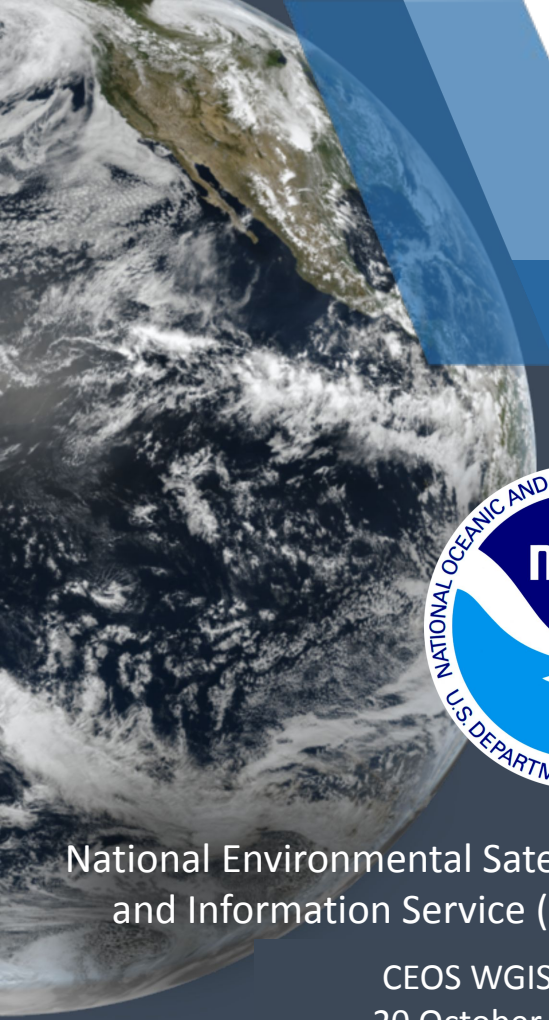


# The NOAA/NESDIS Cloud Archive Pilot (NCAP)

## Moving the NOAA Archive to the NESDIS Common Cloud Framework (NCCF)

Kenneth S. Casey, PhD
NOAA/NESDIS/NCEI

National Environmental Satellite, Data, and Information Service (NESDIS)

CEOS WGISS 52
20 October 2021

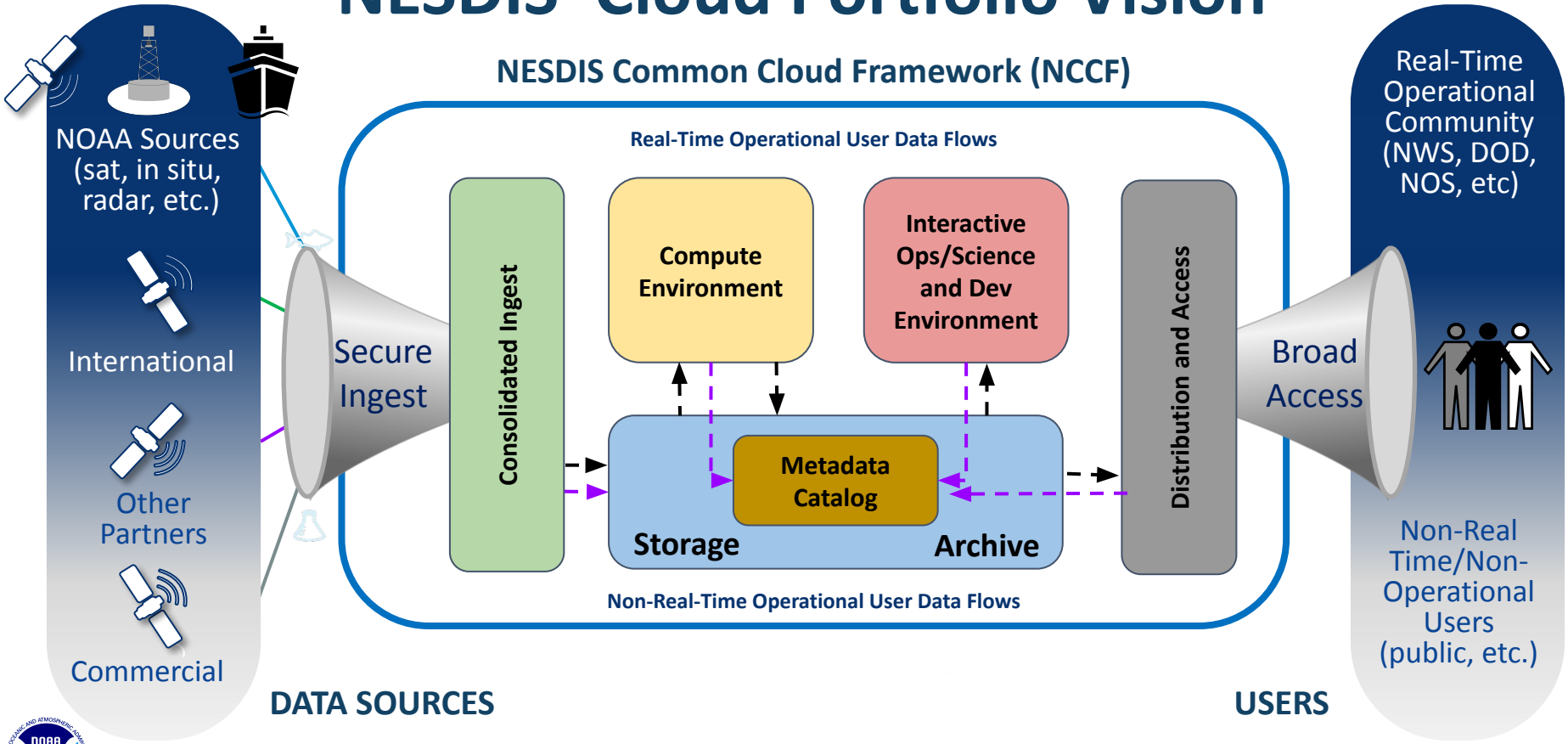
# One Slide Summary

- NOAA/NESDIS began working in earnest to migrate to the cloud in 2019 following release of its 2018 Cloud Computing Strategy
- After successful completion of two pilots, NESDIS began operationalizing the NESDIS Common Cloud Framework (NCCF) in late 2020, and in 2021 conducted its NESDIS Cloud Archive Pilot (NCAP) to finish the end-to-end cloud archive prototype. It is preparing to operationalize this capability, with the goal of achieving an IOC by October of 2022.
- Development of of the NCCF in general and the cloud archive is being managed using the Scaled Agile Framework (SAFe), allowing for the coordination of multiple Agile teams
- This slide deck provides both architectural and schedule details for reference, to emphasize significant progress having been achieved but also to stress the amount of work remaining
- The NCCF and the cloud archive component are being implemented within AWS, and the NASA CMR has been evaluated for possible incorporation into the NESDIS cloud solution - decision pending
-

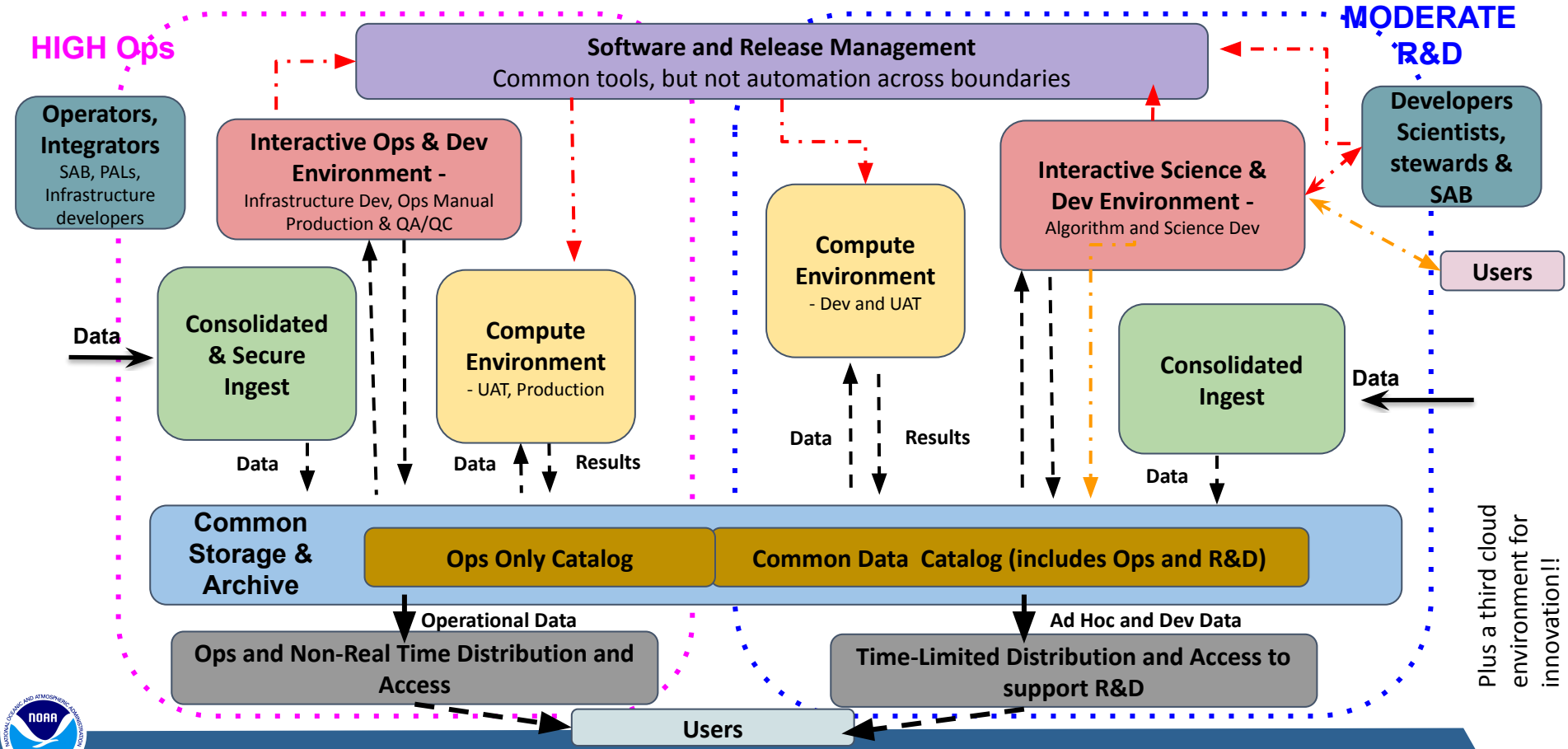
# Introduction and Background

- In the summer of 2018, NESDIS created its Cloud Computing Strategy and launched into a pilot effort to begin moving toward implementation
- The first phase of the pilots was successful, so NESDIS followed with simultaneous efforts to begin operationalizing what was demonstrated, and to continue piloting new capabilities
  - Maturing and iterating, NESDIS launched a Red Team to operationalize cloud functionality into the NCCF and recently wrapped up its second phase of piloting
  - While operationalization continues, NESDIS launched a final pilot phase known as the NESDIS Cloud Archive Pilot (NCAP) to demonstrate the cloud archive capabilities and also established a research and development environment known as the NESDIS Cloud-sandbox Infrastructure Service (NCIS)
- Throughout it all, NESDIS established a Cloud Program Management Office (CPMO) and enterprise Cloud Transformation Plan to manage the overall process, including not just technology aspects but the essential people and process transformations as well.
- These slides provide an overview of the NCCF, with a focus on NCAP, the cloud archive piece

# NESDIS' Cloud Portfolio Vision



# Future NCCF Architecture Vision



# Enterprise Full Operational Capability Vision

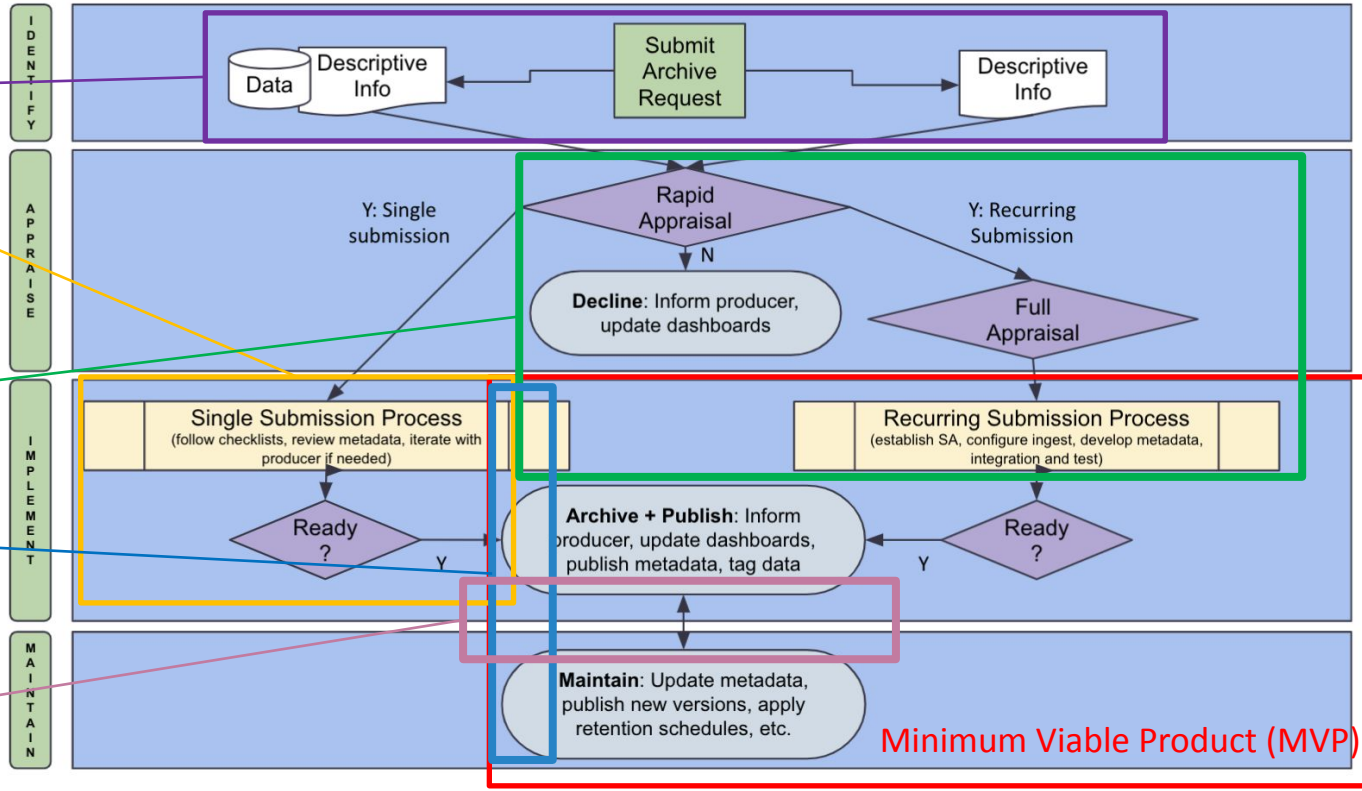
A) Single Request Portal

B) Single Data Submissions

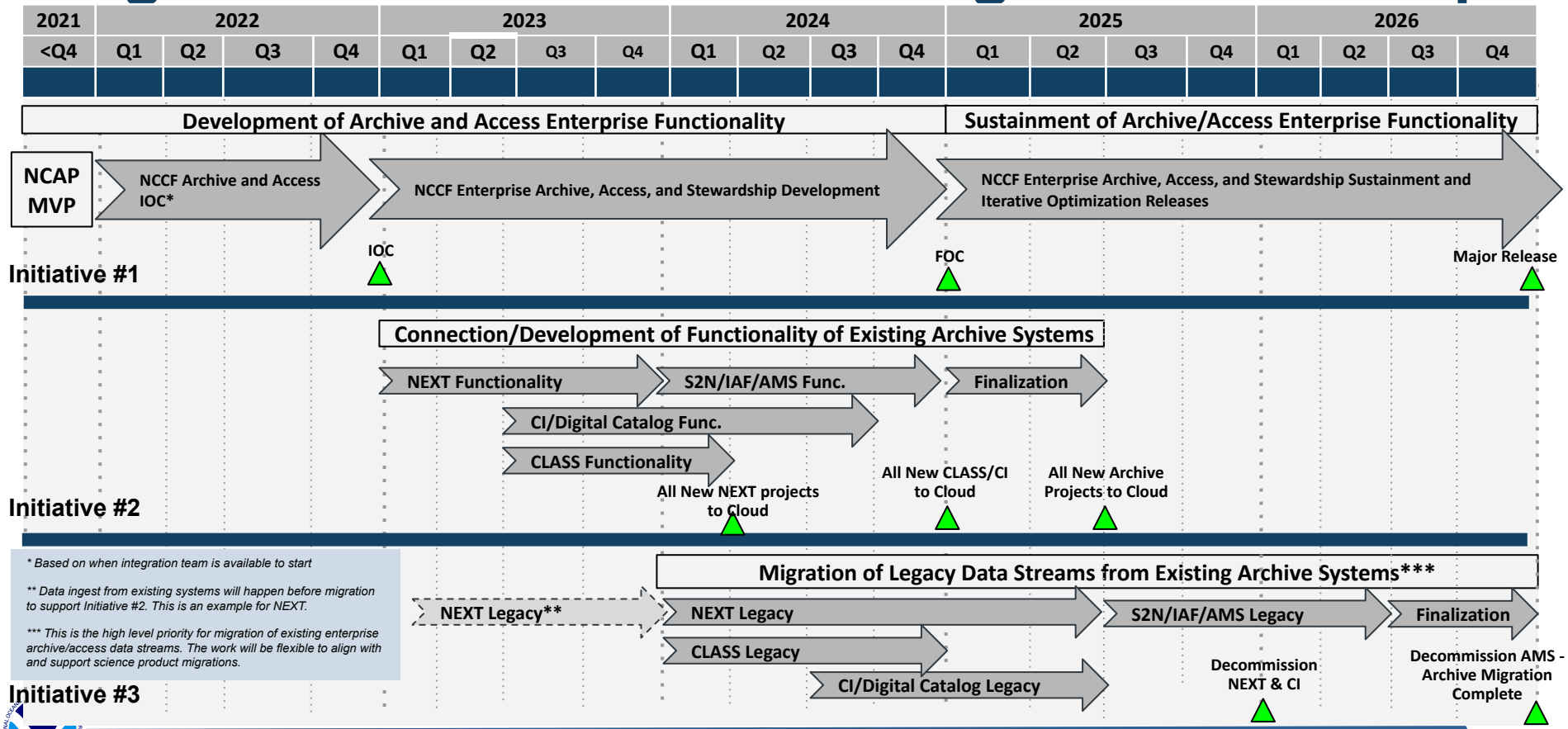
C) Appraisal and Ingest Configuration

D) Mature Metadata Management Tools

E) Implement Enterprise Public Discovery and Access Capabilities

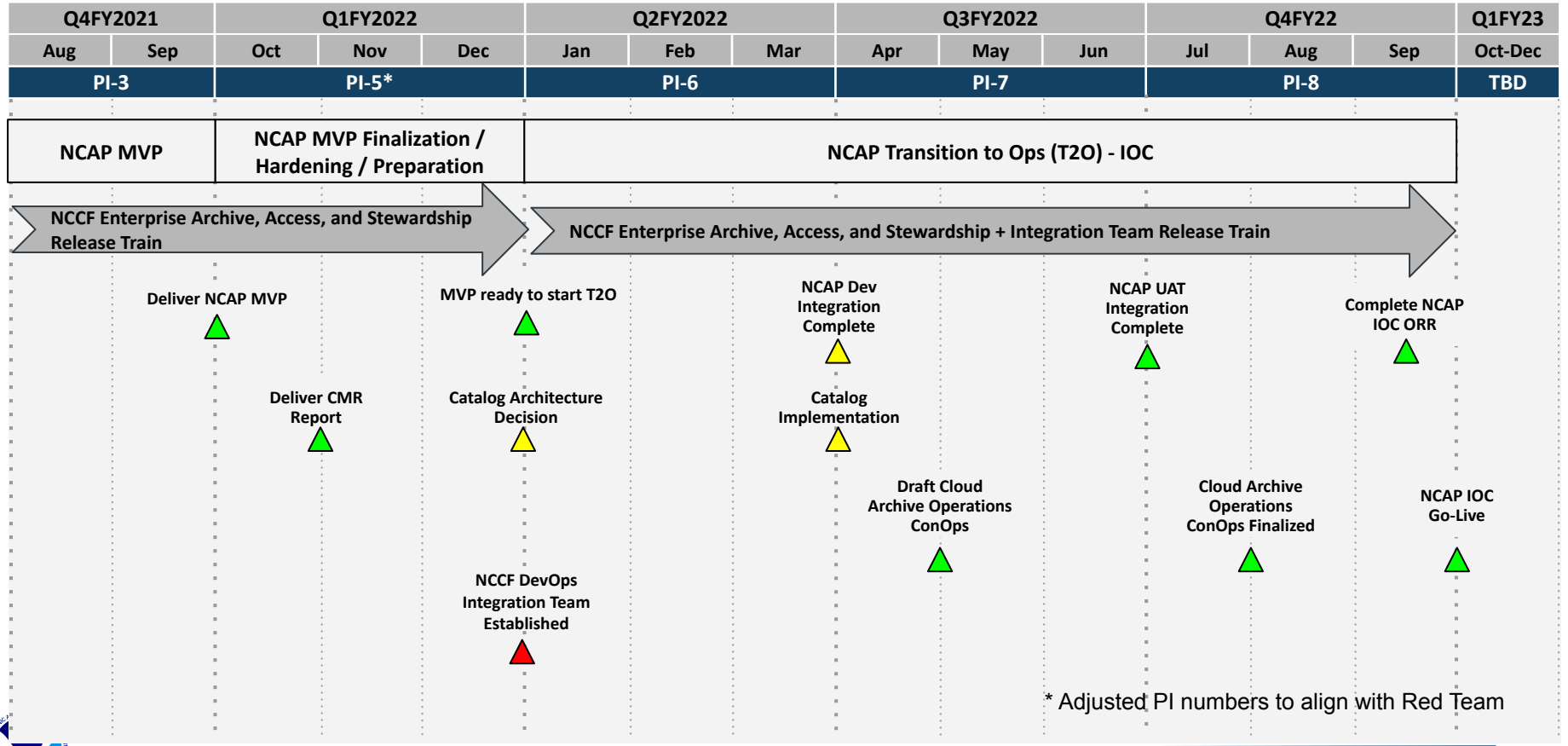


# High Level Archive Dev and Migration Roadmap



# Short-Term Schedule Details

Complete
On Track
Within Tolerance
At Risk

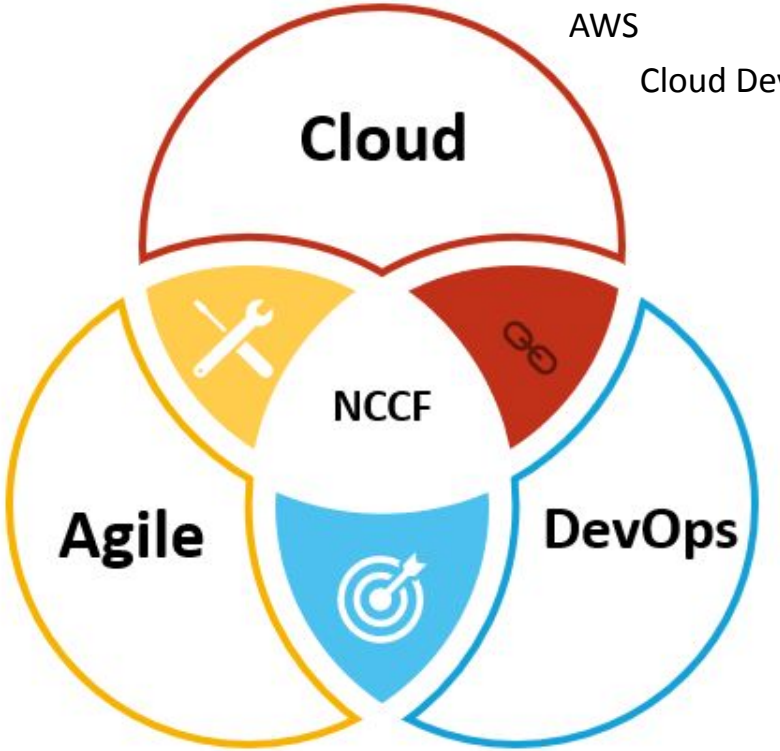



* Adjusted PI numbers to align with Red Team

# Current Enterprise Solution Enablers

AWS

Cloud Development Platform




Methodology : Agile,

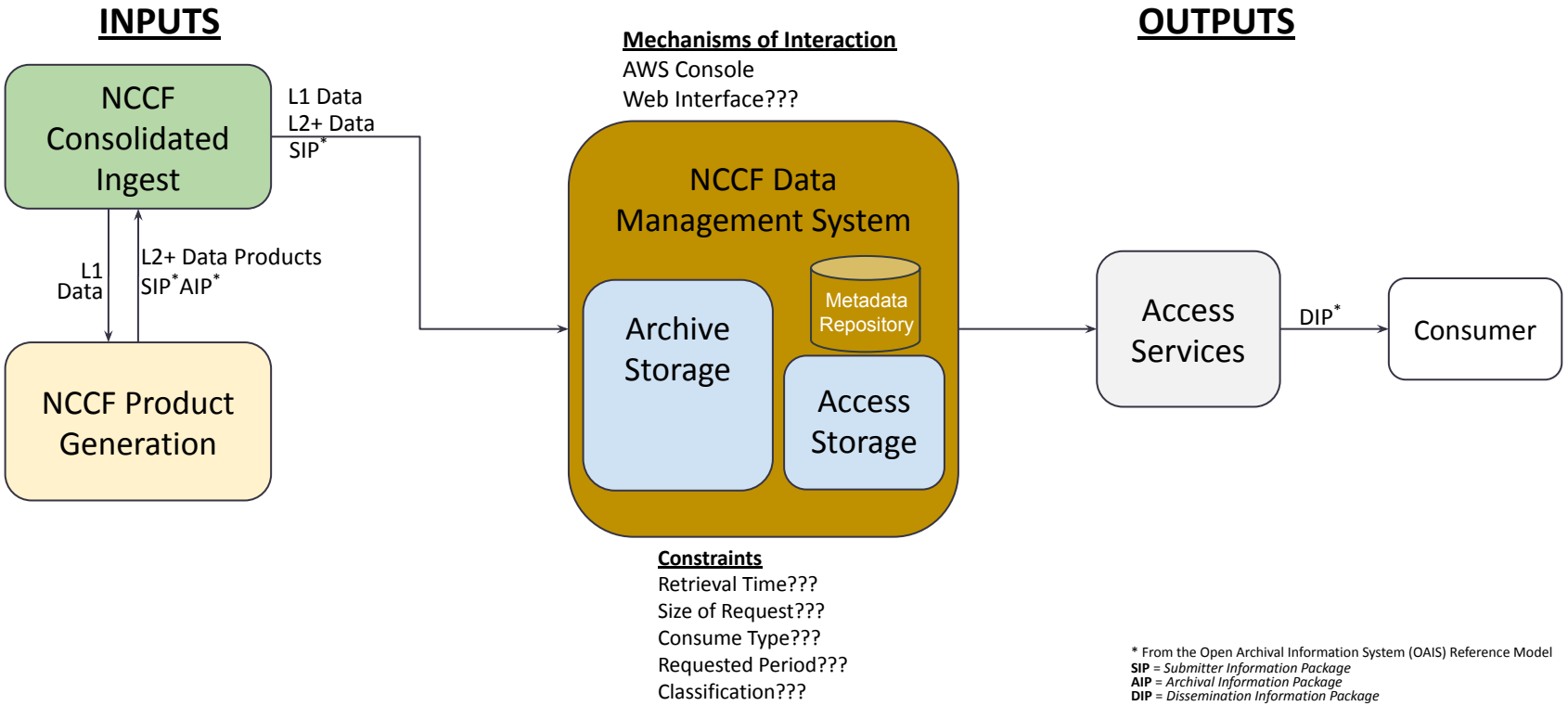
Scaled Agile Framework (SAFe)

Tools: Jira and Confluence

(DevSecOps) Cross functional teams

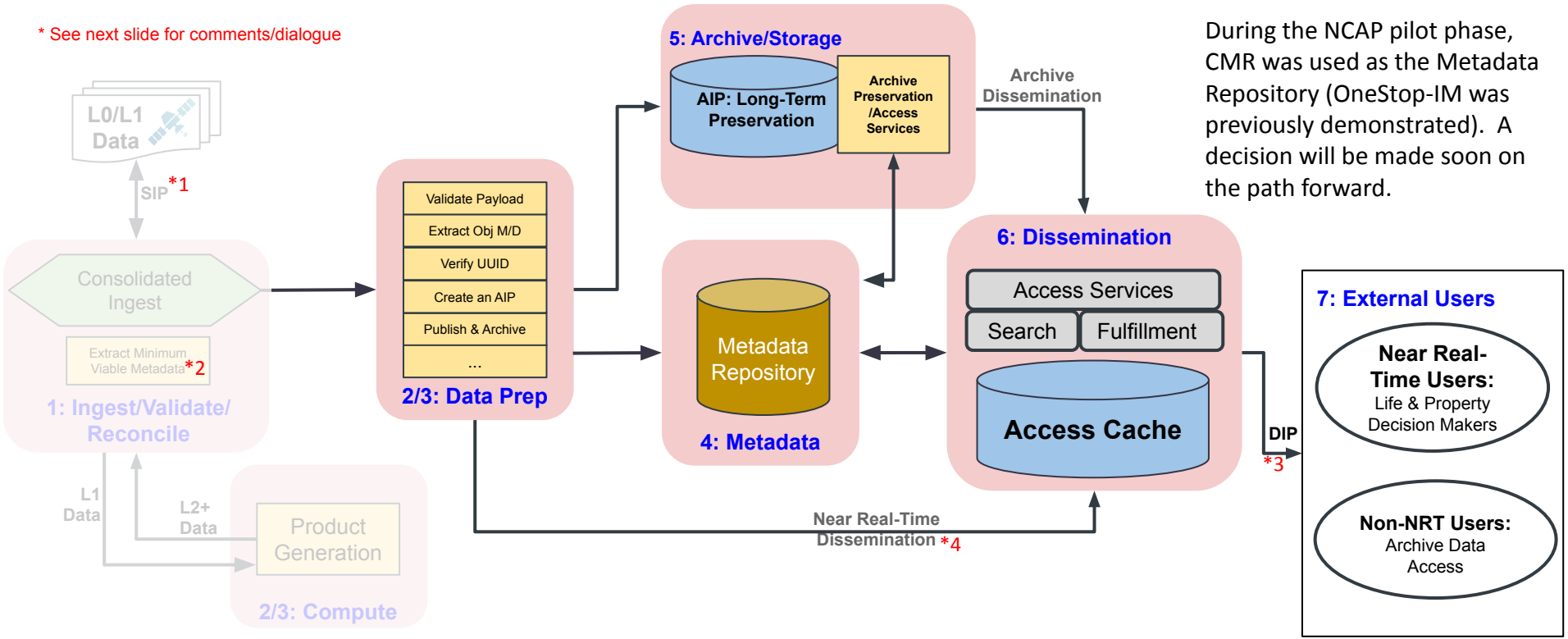
# Details of the Cloud Archive Workflow

# NCCF Archive Context Diagram




* From the Open Archival Information System (OAIS) Reference Model
**SIP** = *Submitter Information Package*
**AIP** = *Archival Information Package*
**DIP** = *Dissemination Information Package*

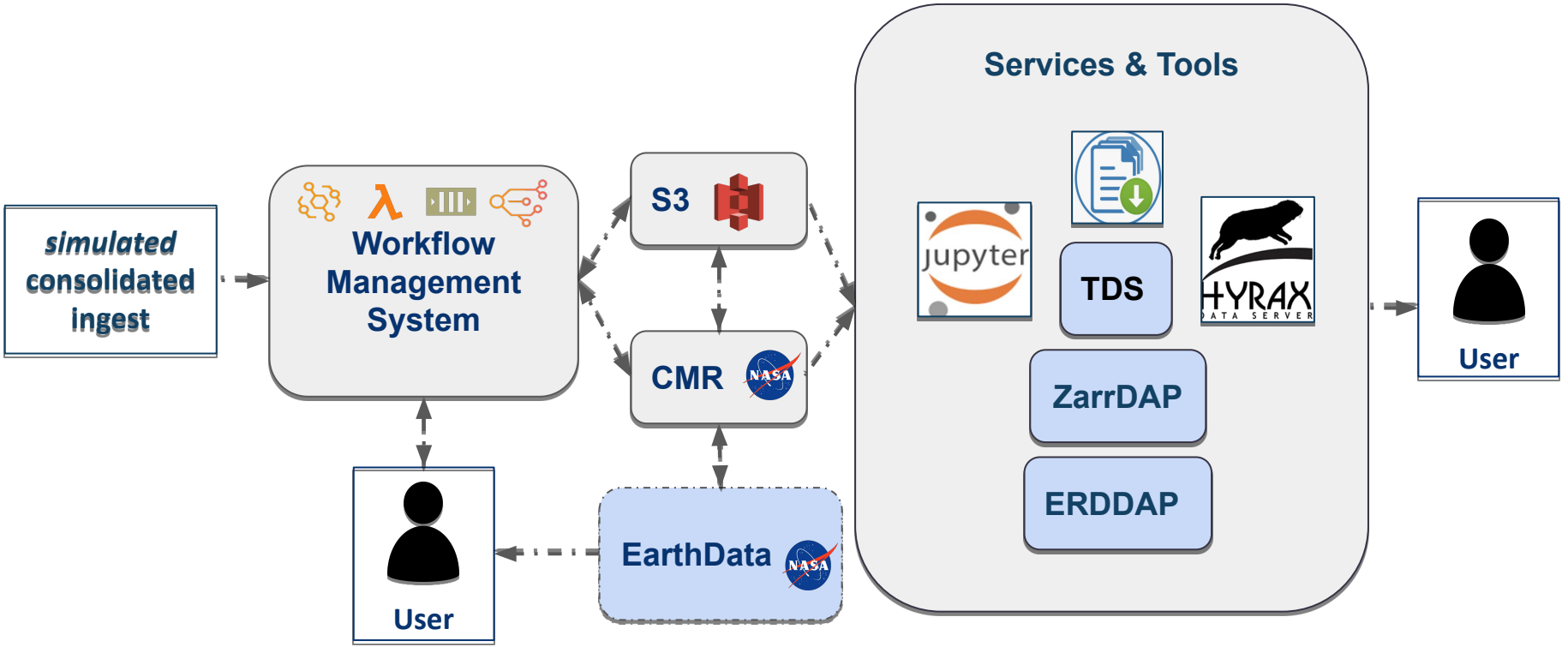
# Satellite/Archive Workflow Diagram




During the NCAP pilot phase, CMR was used as the Metadata Repository (OneStop-IM was previously demonstrated). A decision will be made soon on the path forward.

From the OAIS Reference Model
**SIP** = *Submitter Information Package*
**AIP** = *Archival Information Package*
**DIP** = *Dissemination Information Package*

# Access Demonstration



# Looking Ahead:

- Next program increments will focus on hardening of the demonstrated cloud archive features, and completing the remaining handful of features needed for the MVP
- Resources are being aligned to support the cloud archive operationalization effort as early in 2022 as possible

# Backup Slides

# Concept: The Data Manager's Toolkit

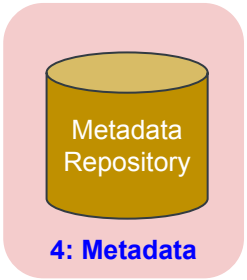
- The tools available to the **Archivists**, **Data Managers**, **Metadata Experts**, or **Operators**.
- These are tools which are available in each step of the workflow.
- Metadata Tools
  - *Cross-Role Tools (Tools Available to the Archist, Data Manager, Metadata Archivist, or Operator)*
    - Reporting tools - generate custom reports (velocity, successes/failures, etc.)
  - *Archivist Tools*
    - Verification reports that analyze metadata state and ensure the information represents the latest known state of metadata.
  - *Data Manager Tools*
    - Editing tools to backfill missing, incomplete or invalid Metadata.
  - *Metadata Archivist Tools*
    - Metadata Editing for standardizing snippets or components of metadata records that require updating many metadata records simultaneously.
  - *Operator Tools*
    - UI-based monitoring tools (Dashboard) - how many, velocity, success/failure, etc. of collection, granule,AIP, DIP metadata
    - Capacity Management Assessment - what is the total byte/memory/etc footprint of the holdings, what is the cost, etc.

# Step 2 and 3: Satellite Data Prep

1. Validate the payload.
2. Extract the object metadata (S3 metadata).
3. Verify UUID of the object.
   a. If no UUID exists, then throw an error and exit.
4. Create an AIP for the data object (granule)
   a. Extract full metadata (for AIP Package Description)
   b. Create AIP Packaging Information
   c. Create AIP Representation Information
   d. Create AIP Preservation Description Information
5. Publish data object metadata.
6. Archive the data object.
7. Notify the operator.

**Assumptions:**

1. The data collection for the satellite data exists.
2. The data object (granule) is considered complete.

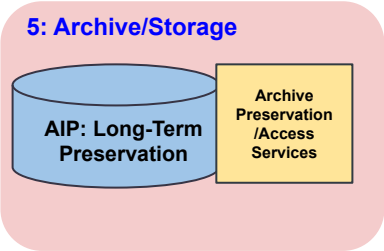



# Step 4: Metadata

For NCAP and the NCIS, the "Metadata Repository" is the NASA CMR application managed by the NCIS Infrastructure Team.

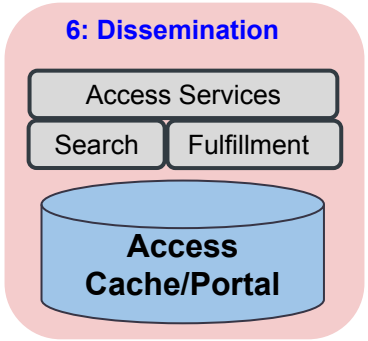
Other Metadata Tools for the Satellite Archive Workflow include

1. CMR is part of the prototype and used as part of MVP
2. OneStop-Inventory Manager (OSIM) were previously tested in the NCCF pilots
3. A decision is pending on what catalog solution to utilize in the IOC (CMR, OSIM, both)




# Step 5: Archive & Storage

- Connected with Metadata, Data Prep, and Dissemination is the **Archive/Storage** piece of the workflow.
- The Archive/Storage contains the preservation copy (the “archive copy” or “gold copy”) of the object.
- From Data Prep, the AIP (which includes the data object) are sent to two storage services:
  - **Copy #1** is sent to the Archive/Storage service for preservation.
    - The AIP will be written **directly** to S3 Glacier or S3 Glacier Deep Archive.
  - **Copy #2** is sent to the Access Cache in the Dissemination service for immediate search, discovery, and access.
    - The Access Cache S3 bucket will keep objects in S3 Standard for up to a certain age (e.g., 5 days and objects older than 5 days are deleted).



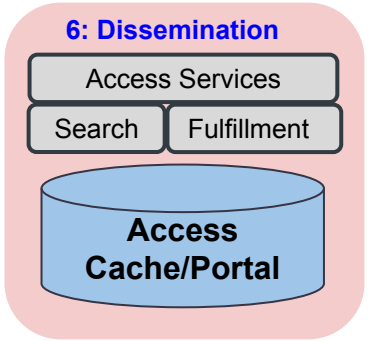

# Step 6: Dissemination

**Dissemination** is defined as the services to support distribution of data, search and discovery, and order fulfillment.

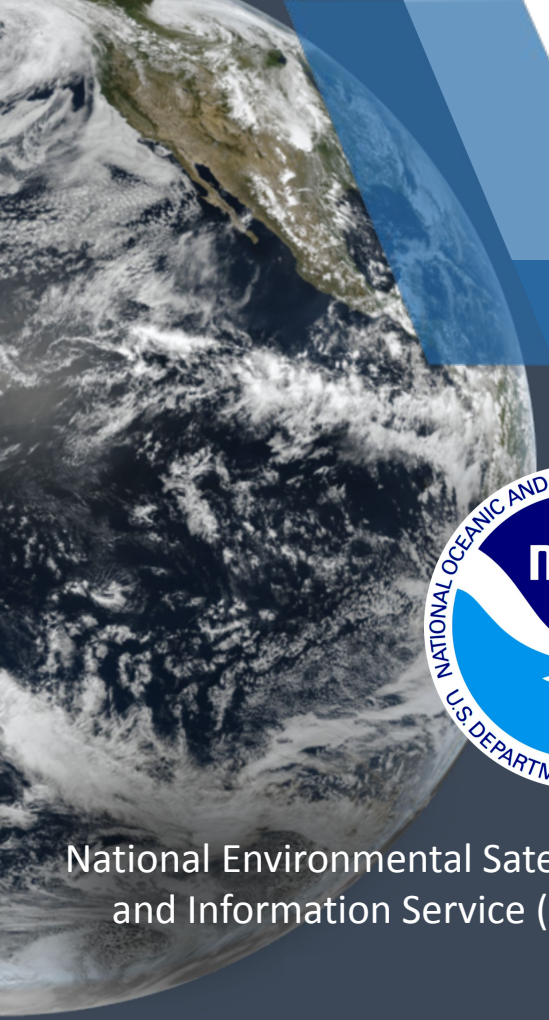
What are the major components of Dissemination aka Access Services?

- Search
  - Metadata search/discovery UI -> data
  - **Complete, accurate metadata is key to successful implementation of Service Oriented Architecture**
- Fulfillment
  - Synchronous and Asynchronous requests processed
  - Data Management
    - Real-time to NRT -> PDA/CLASS (satellite product focused)
    - NRT (next day) to historical -> NCEI/CLASS (all products)
- Core Services - OAIS Features

# Step 6: Dissemination - Core Services




- Search/Discovery - metadata and data APIs
- Download data
- Visualize structured data
- Subsetting of structured data
- Data Transformation
- Ad Hoc orders/requests (Alt: Ad Hoc Data Requests)
- Subscription orders/requests (Alt: Recurring Data Requests)
- Bulk orders/requests (Alt: Large Data Requests)
- Event-based orders/requests
- Certification orders/requests
- Data Use Monitoring/Administration/Reporting
- Order fulfillment Monitoring/Administration/Reporting
- Subscription Monitoring/Administration/Reporting
- A/A for restricted data access



# Questions?

National Environmental Satellite, Data, and Information Service (NESDIS)