EODDIS System of Systems Architecture circa 2018

- Distributed Active Archive Center System Functions
  - Ingest and Archive – Acquire and validate data products from multiple data providers, store multiple versions, ensure archive integrity.
  - Distribution – Enable users to learn about data products, format, theoretical basis. Enable users to search for and access data products. Provide subsetting, reprojection and format conversion.
  - Data Providers – Satellite Mission Science Data Systems, Aircraft and Ground Validation Campaigns, Principle Investigators collect and deliver data products to DAACs for archive and distribution.
  - Enterprise Functions – Data Search and Discovery

EOSS Component Systems and External Interfaces

- Data Provider: may be internal to EODDIS such as SIPS or DAAC, or external such as a NASA Principle Investigator; may be from on-premise or in-cloud. DAAC on-premise facility serves as proxy science data source.
- Data Consumer: may be external to EODDIS such as SIPS or DAAC, or external such as public user service; may be on-premise or in-cloud. The interface is agreed to in advance and can represent a one-time or an ongoing active data transfer.
- Common Metadata Repository (CMR): catalogs all DAAC science data and services; metadata records are registered, modified, and access via tools such as the MMT or via standard APIs.
- Metadata Management Tool (MMT): allows metadata authors (data owners) to create, update, publish, view, delete and manage their metadata records in CMR.
- Earthdata Search Client: Web application allowing users to search, discover, visualize, and access NASA Earth science data products using CMR and GIBS.
- Earthdata Login: single sign-on user registration and user management management system for public users to get Earth science data from any of the DAACs.
- Global Imagery Browse Services: provides standard image services from DAAC science data.
- Worldview: public Web tool for interactively browsing global imagery from GIBS.
- Data Preservation Archive: Secure copy of unique and irreplaceable science data and information at a separate location.
- Cumulus: a framework for data ingest and archive management configurable to perform acquisition, ingest, validation, processing, metadata harvesting & creation, publication to CMR, data distribution and metrics reporting.

Major Functional Concepts and Scenarios

- Data Management:
  - Setup and retirement of ingest, archive and distribution workflows & services for data collections.
  - Metadata curation, product guide development, Website landing page development.
  - Setup test environments, build/checkout metadata and data service workflows.
  - Cumulus Dashboard used to add/delete new dataset collections or individual granules.
  - MMT or discipline-specific metadata tool used to add collection metadata to the CMR.
  - Worldview used to test/check product imagery generation and delivery to GIBS.
  - Testcheck dataset and documentation backup and restore configuration.
  - Control data access based on guidance from the Principle Investigator or science team.
  - Monitoring, troubleshooting workflow interrupts and checking quality of data services.
  - Auditing product granule availability, checking for data gaps, removing bad/duplicate data.
  -Augmenting collection and granule metadata as necessary

- Distribution:
  - Cloud transfer; may be on-premise facility.
  - The user is authenticated and authorize via Earthdata login.
  - The user requests and downloads data products.
  - Network throttling or circuit breakers may implemented to control egress costs.
  - Transfer rates are managed similar to DAAC on-premise bandwidth to the internet.
  - Direct access to DAAC AWS archive via external AWS account within same AWS region.
  - The user is authenticated and authorize via Earthdata login.
  - Temporary access is granted with no egress costs in the cloud-to-cloud transfer.
  - Utilization is monitor & controlled by data collection or by registered user.

- Maintenance:
  - DAAC Application Owners run their own instance of Cumulus on their own NGAP account.
  - Deployments to operations accounts occur without downtime.
  - User Acceptance Test environment is in separate isolated NGAP account.
  - Used by DAAC operators and selected Data Providers and Consumers.
  - Acceptance test new releases on realistic workflows.
  - Development teams will have NGAP sandbox accounts to work on new code.
  - DAAC Application Owners ensure integration of the latest releases of NGAP and Cumulus.

A Concept of Operations for Earth Science Data Archive and Distribution in the Cloud

John F Moses1, Kathleen Baynes1, Jeanne Behnke1
1 NASA Goddard Space Flight Center, Greenbelt, MD, United States.