NASA's EOSDIS Cumulus

:::

Ingesting, Archiving, Managing, and Distributing Earth Science Data from the Commercial Cloud

Katie Baynes - Civil Servant, NASA Goddard
Rahul Ramachandran - Civil Servant, NASA Marshall
Dan Pilone, Patrick Quinn, Jason Gilman - Element 84, Inc
Ian Schuler, Alireza Jazayeri - Development Seed

EARTHDATA
EOSDIS NASA'S EARTH OBSERVING SYSTEM DATA AND INFORMATION SYSTEM

EOSDIS in Context

Capture and Clean

Process

Archive

Transform*

Distribute

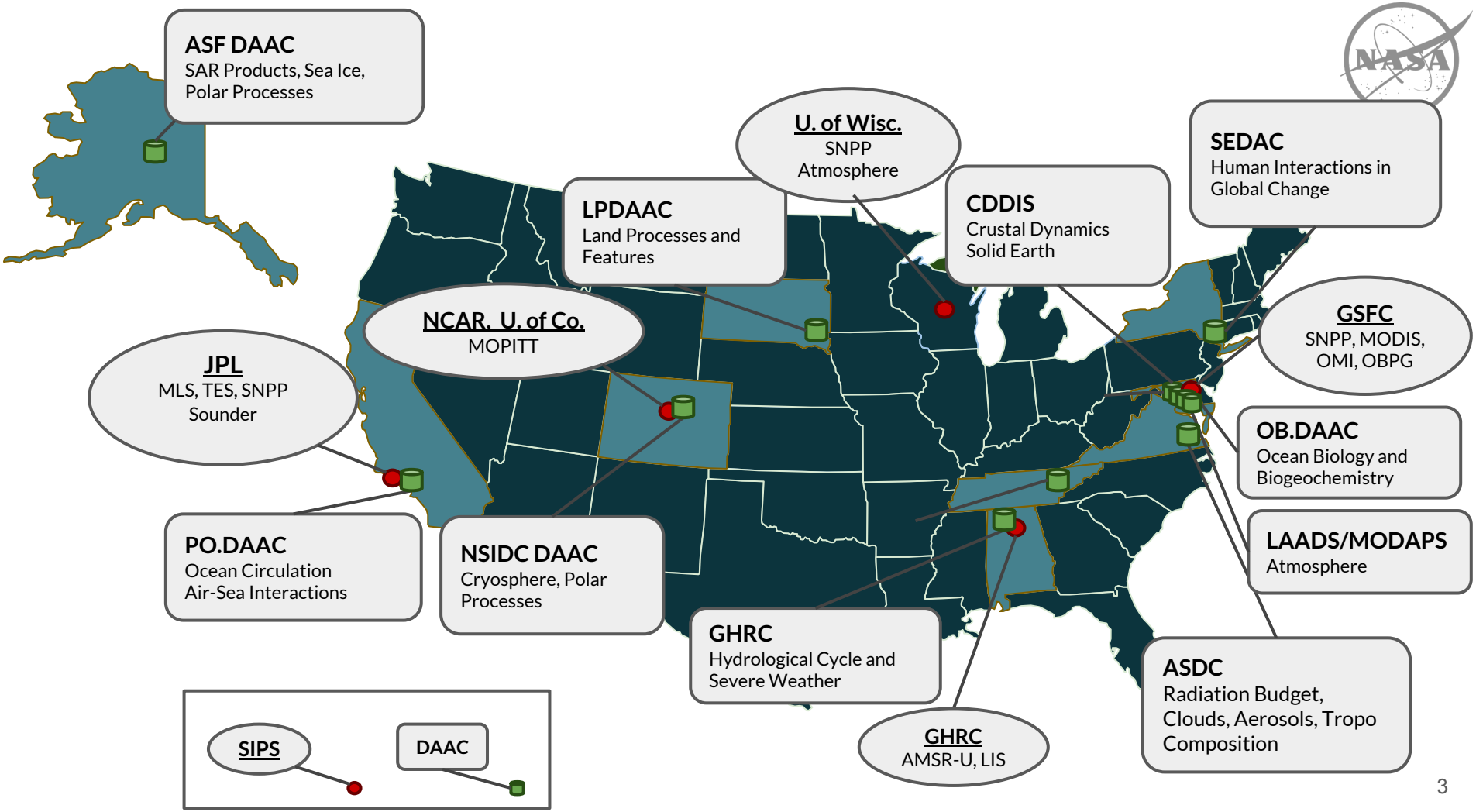
Research

Applications

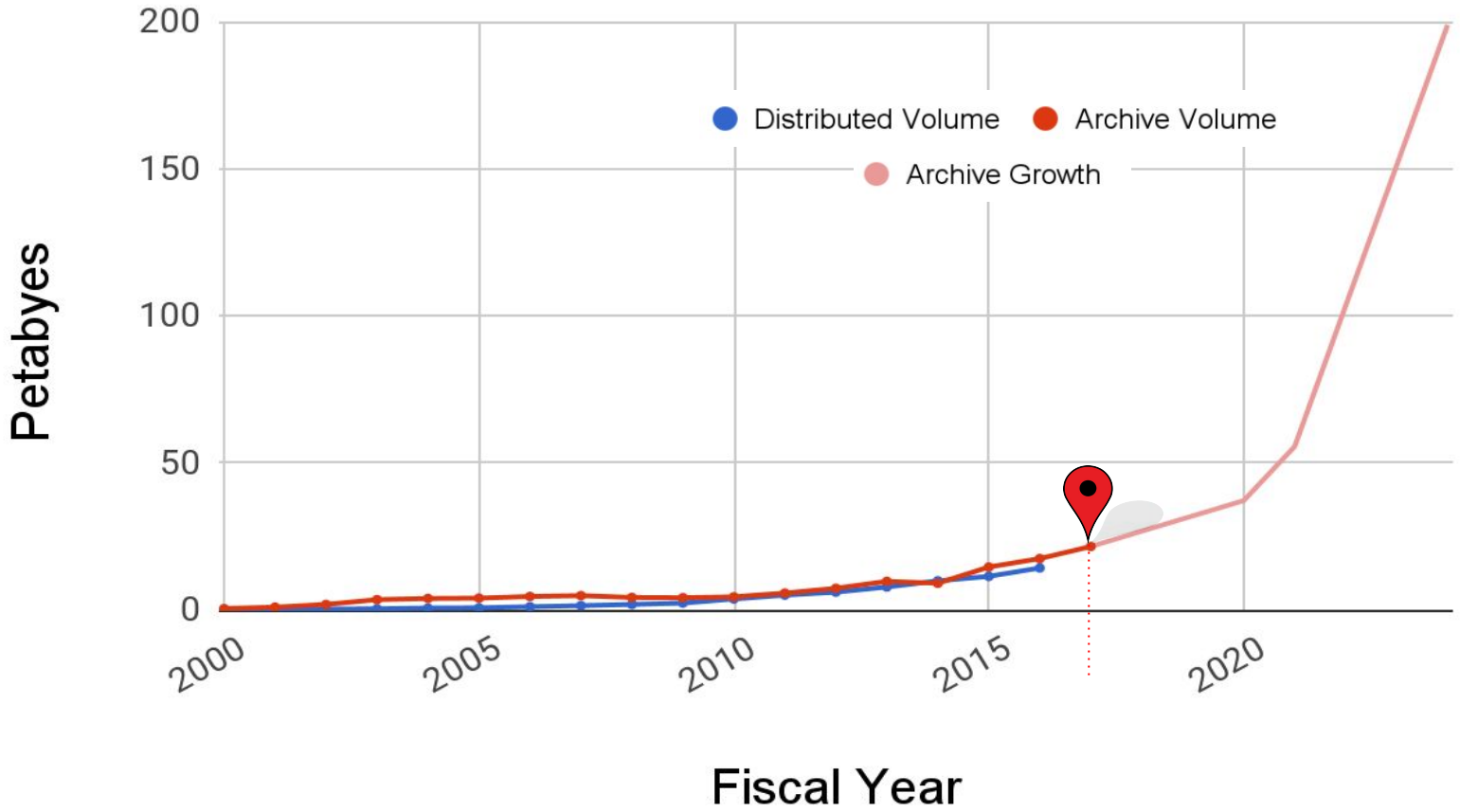
Education

*Subset, reformat, reproject

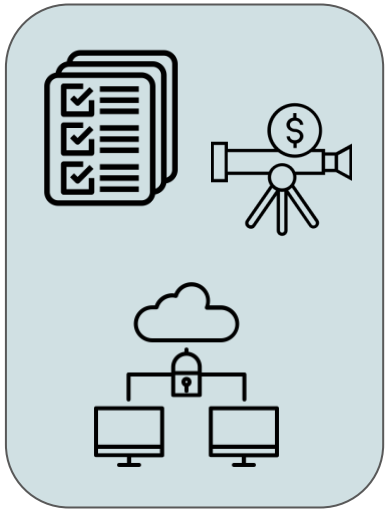
satellite, airplane, compass icons by Nook Fulloption,
database, transformation, decision, process, data mining, customer community by Becris
all from from thenounproject.com (CC 3.0)


ASF DAAC
SAR Products, Sea Ice, Polar Processes
U. of Wisc.
SNPP
Atmosphere
SEDAC
Human Interactions in Global Change
LPDAAC
Land Processes and Features
CDDIS
Crustal Dynamics
Solid Earth
NCAR, U. of Co.
MOPITT
GSFC
SNPP, MODIS, OMI, OBPG
JPL
MLS, TES, SNPP Sounder
OB.DAAC
Ocean Biology and Biogeochemistry
PO.DAAC
Ocean Circulation
Air-Sea Interactions
NSIDC DAAC
Cryosphere, Polar Processes
LAADS/MODAPS
Atmosphere
GHRC
Hydrological Cycle and Severe Weather
ASDC
Radiation Budget, Clouds, Aerosols, Tropo Composition
GHRC
AMSR-U, LIS
SIPS
DAAC

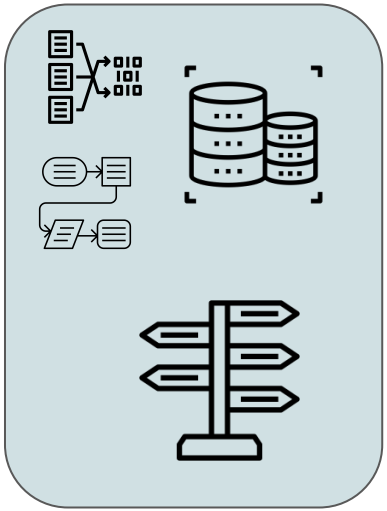
EOSDIS Holdings and Projected Growth



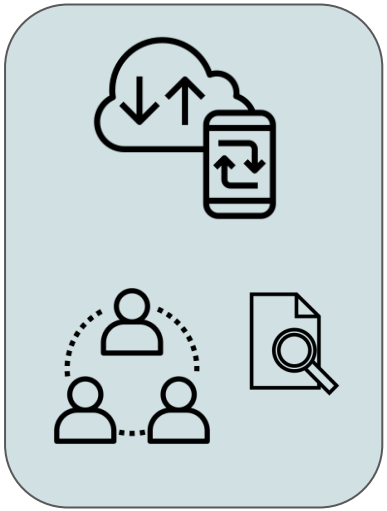
Past 24 Months: Focused on evaluation and planning for a cloud migration in 4 areas



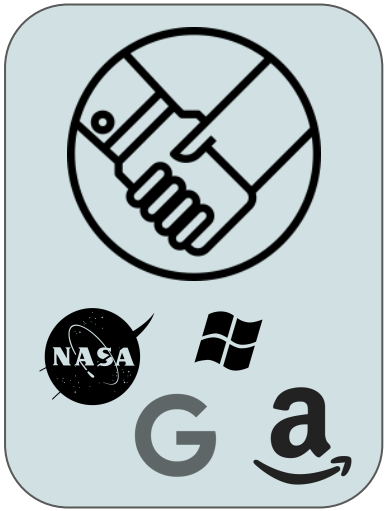
Compliance, Security, Cost Tracking



Core Archive Functionality and Processing



End-User Application Migration

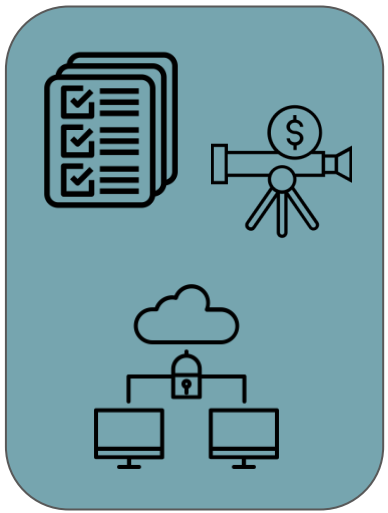


Pursuing Cloud Partnerships

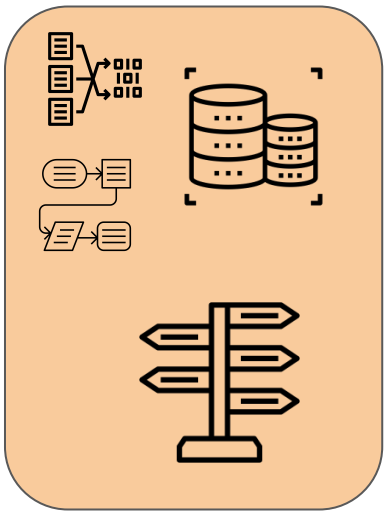
cloud network secured, file search icons by UNiCORN,
checklist, database, transformation, process, decision, customer community, handshake, cloud mobile icons by Becris
"Cost Per Impression" icon by Creative Stall
all from from thenounproject.com (CC 3.0)

Microsoft Windows, Google, Amazon logos all registered trademarks of their respective corporations

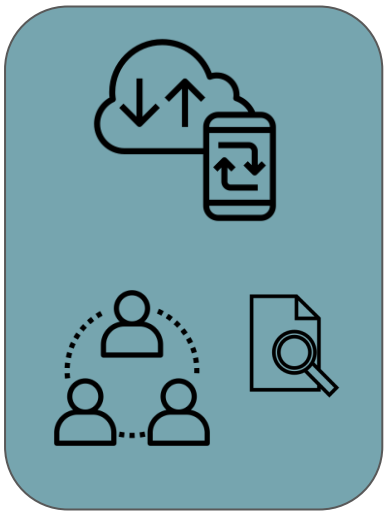
Past 24 Months: Focused on evaluation and planning for a cloud migration in 4 areas



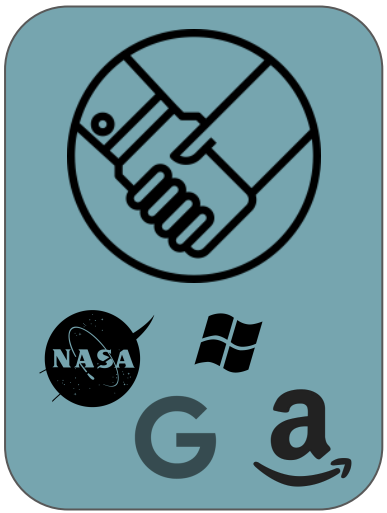
Compliance, Security, Cost Tracking



Core Archive Functionality and Processing



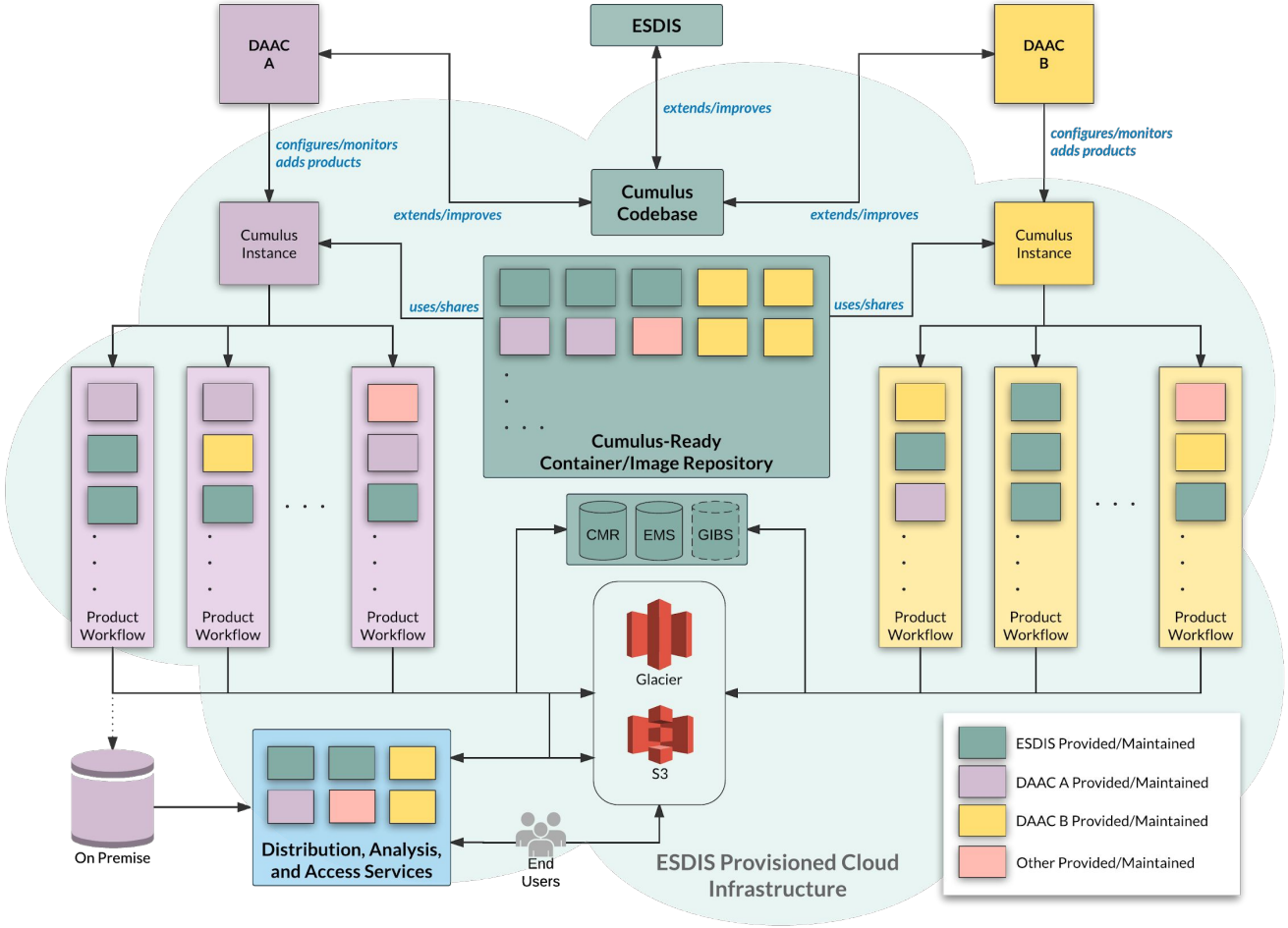
End-User Application Migration

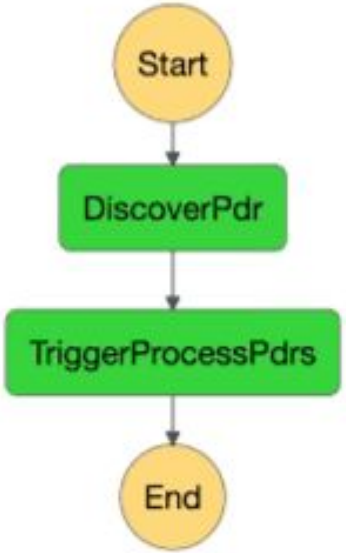


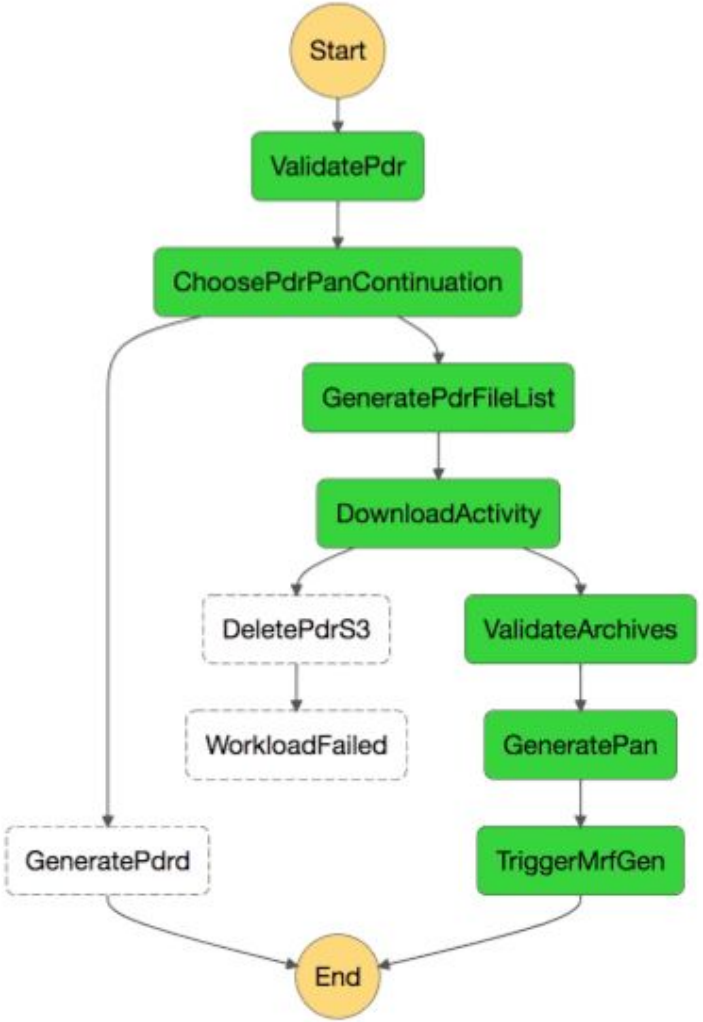
Pursuing Cloud Partnerships

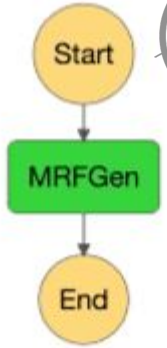
cloud network secured, file search icons by UNiCORN,
checklist, database, transformation, process, decision, customer community, handshake, cloud mobile icons by Becris
"Cost Per Impression" icon by Creative Stall
all from from thenounproject.com (CC 3.0)

Microsoft Windows, Google, Amazon logos all registered trademarks of their respective corporations


DAAC A
ESDIS
DAAC B
extends/improves
configures/monitors adds products
configures/monitors adds products
Cumulus Codebase
extends/improves
extends/improves
Cumulus Instance
Cumulus Instance
uses/shares
uses/shares
Cumulus-Ready Container/Image Repository
CMR
EMS
GIBS
Product Workflow
Product Workflow
Product Workflow
Product Workflow
Product Workflow
Product Workflow
Glacier
S3
On Premise
Distribution, Analysis, and Access Services
End Users
ESDIS Provisioned Cloud Infrastructure
ESDIS Provided/Maintained
DAAC A Provided/Maintained
DAAC B Provided/Maintained
Other Provided/Maintained

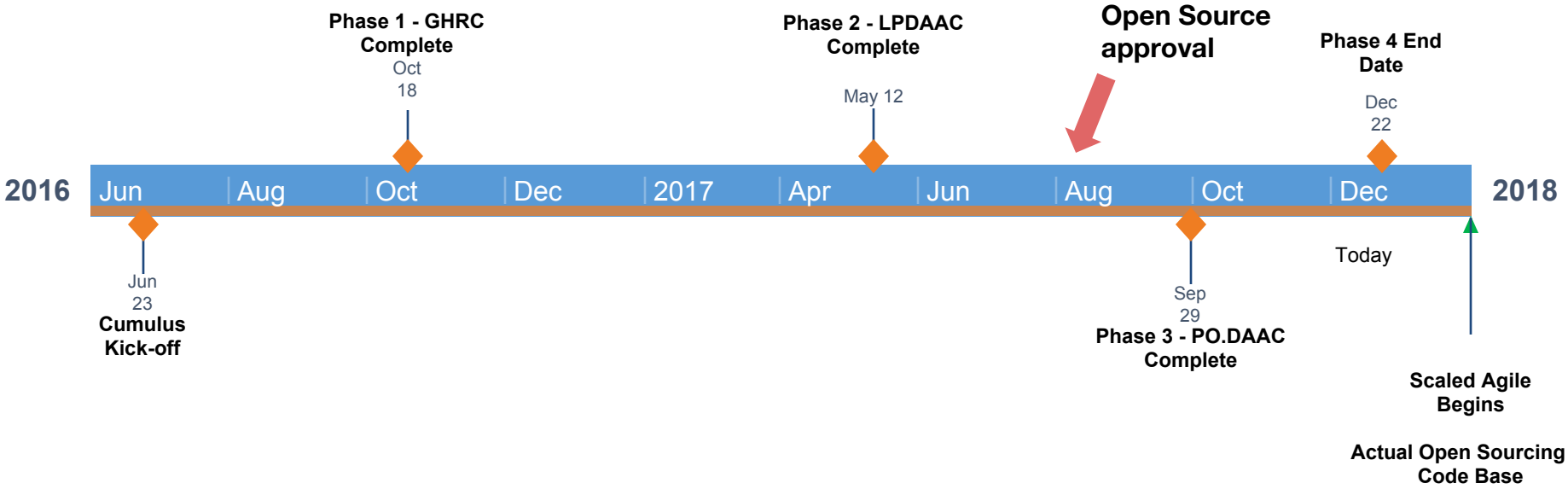


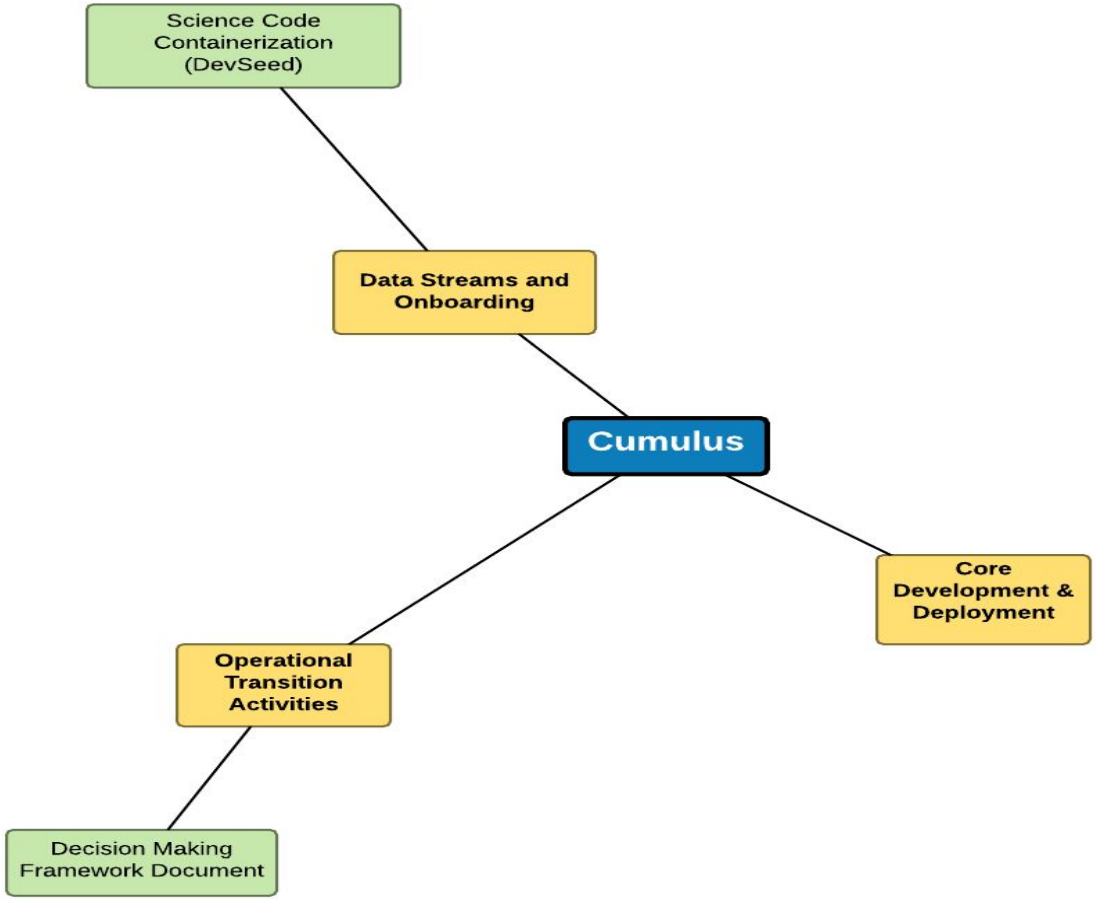


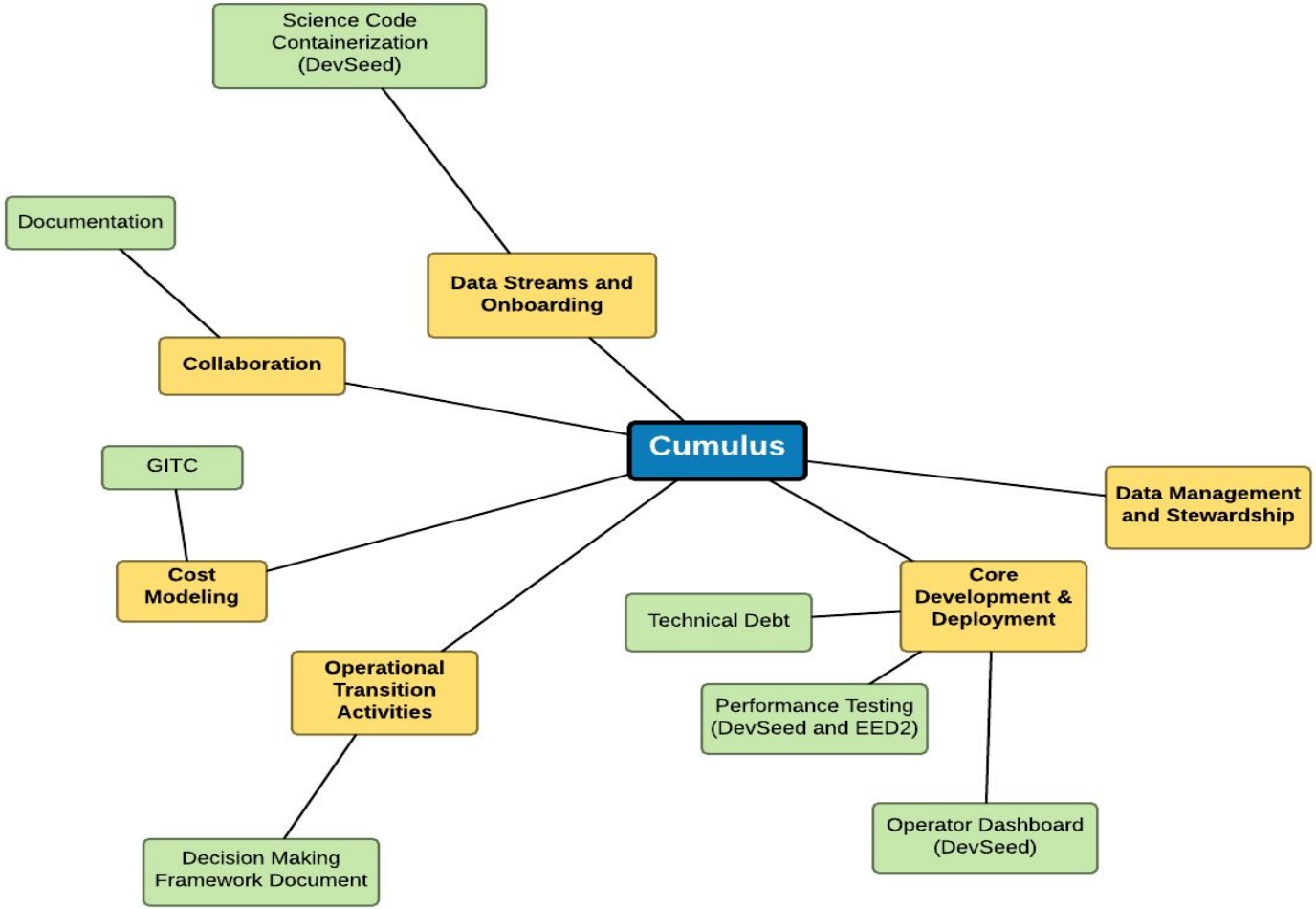


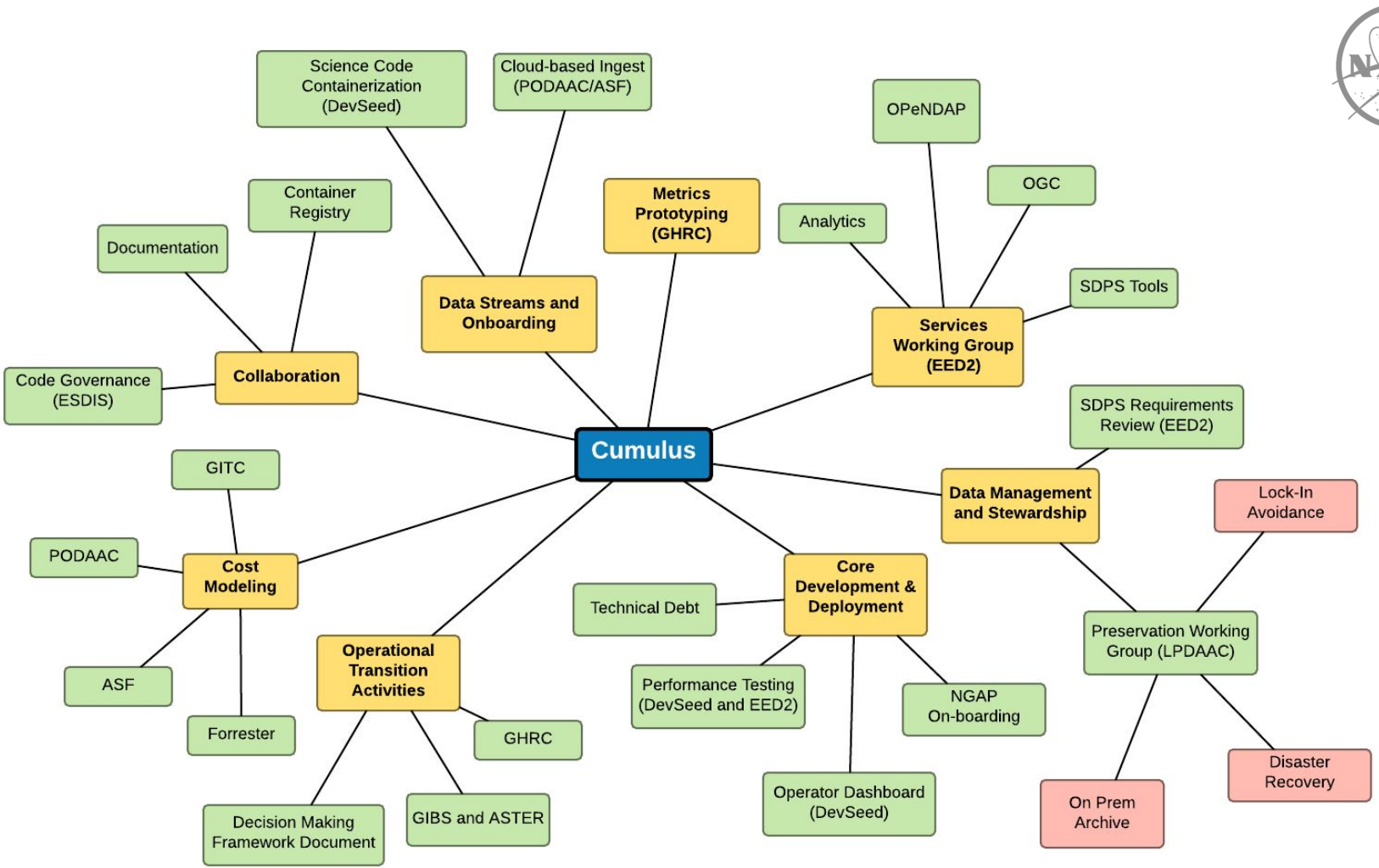
Example Step Function for GIBS in the Cloud

Looking at the Past 18 Months

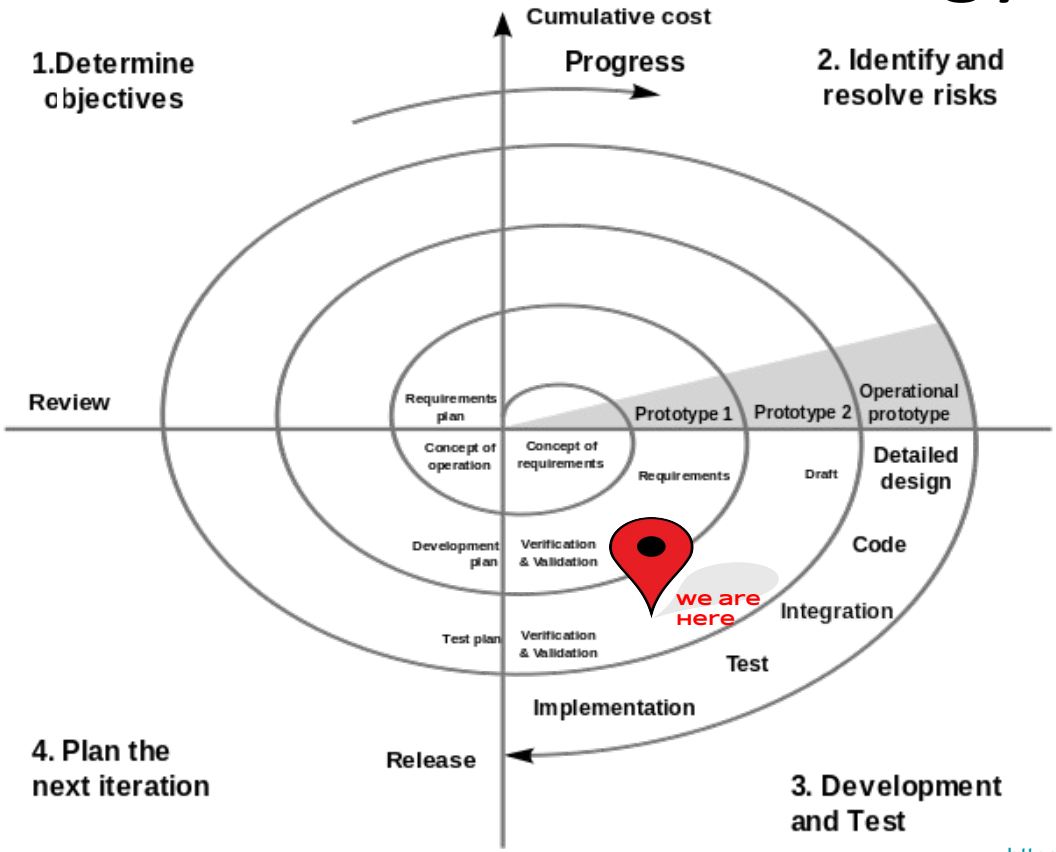



Science Code Containerization (DevSeed)
Data Streams and Onboarding
Cumulus
Core Development & Deployment
Operational Transition Activities
Decision Making Framework Document


Science Code Containerization (DevSeed)
Documentation
Data Streams and Onboarding
Collaboration
Cumulus
GITC
Data Management and Stewardship
Cost Modeling
Core Development & Deployment
Technical Debt
Operational Transition Activities
Performance Testing (DevSeed and EED2)
Operator Dashboard (DevSeed)
Decision Making Framework Document

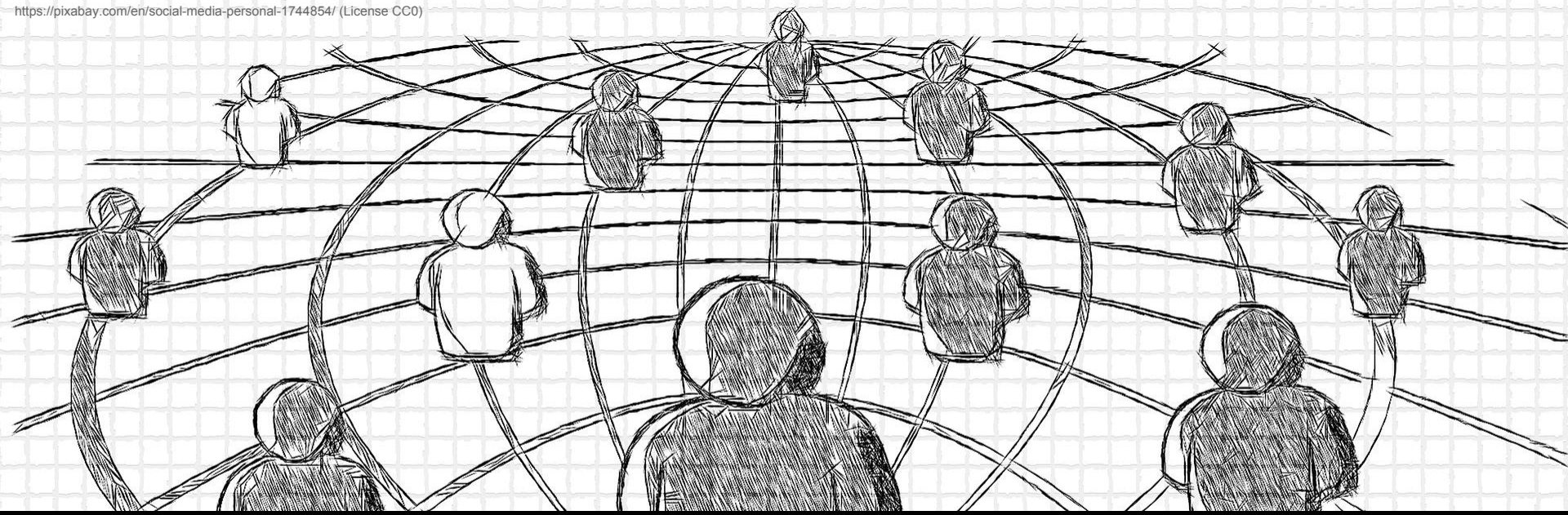

Cumulus
Data Streams and Onboarding
Science Code Containerization (DevSeed)
Cloud-based Ingest (PODAAC/ASF)
Collaboration
Container Registry
Documentation
Code Governance (ESDIS)
Metrics Prototyping (GHRC)
Services Working Group (EED2)
OPeNDAP
OGC
Analytics
SDPS Tools
Data Management and Stewardship
SDPS Requirements Review (EED2)
Lock-In Avoidance
Preservation Working Group (LPDAAC)
Disaster Recovery
On Prem Archive
Core Development & Deployment
Technical Debt
Performance Testing (DevSeed and EED2)
NGAP On-boarding
Operator Dashboard (DevSeed)
Operational Transition Activities
GHRC
GIBS and ASTER
Decision Making Framework Document
Cost Modeling
GITC
PODAAC
ASF
Forrester

Cumulus Phased Methodology

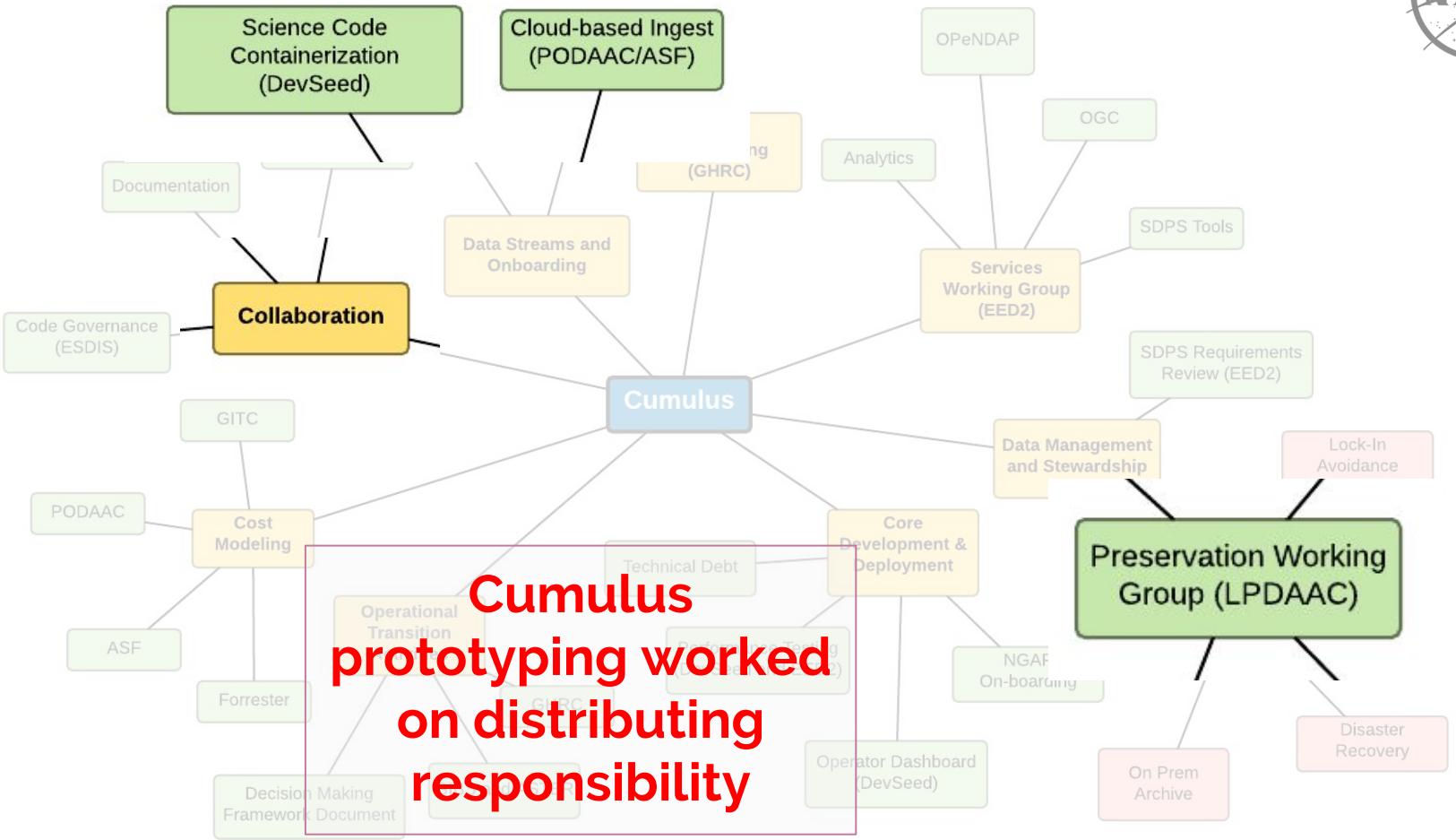


https://en.wikipedia.org/wiki/Spiral_model

https://pixabay.com/en/social-media-personal-1744854/ (License CC0)

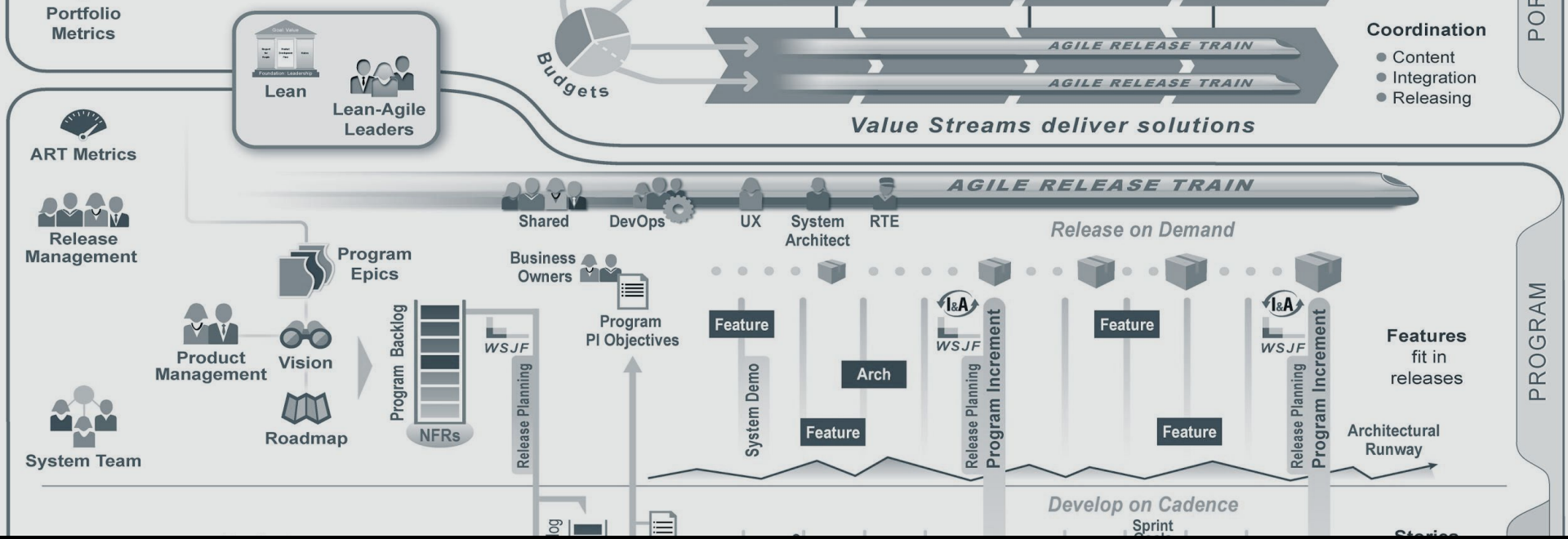


So, what does this mean in practical terms for EOSDIS?


Science Code Containerization (DevSeed)
Cloud-based Ingest (PODAAC/ASF)
OPeNDAP
OGC
Analytics
SDPS Tools
Documentation
Data Streams and Onboarding
(GHRC)
Services Working Group (EED2)
Collaboration
Code Governance (ESDIS)
SDPS Requirements Review (EED2)
Cumulus
GITC
Data Management and Stewardship
Lock-In Avoidance
PODAAC
Cost Modeling
Core Development & Deployment
Preservation Working Group (LPDAAC)
Technical Debt
Operational Transition
ASF
Forrester
NGAP On-boarding
Disaster Recovery
Operator Dashboard (DevSeed)
On Prem Archive
Decision Making Framework Document
Cumulus prototyping worked on distributing responsibility

Earthdata Cloud 2021 Strategy for DAAC Migration

https://pixabay.com/en/birds-formation-flying-south-north-1835510/ (License CC0)

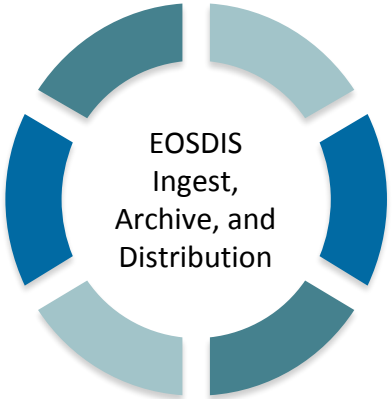


Using SAFe®* to align Earthdata Cloud 2021 Migrations

*Yes, I am aware of the huge number of buzzwords on image above. And it is less “agile”. But this project involves 100s of people and this system has proven extremely useful.

http://www.scaledagileframework.com/program-increment/

DAAC Cloud Migration Organization January



- Product Manager
- System Architect
- Release Train Engineer

Cumulus Core Development

GHRC Migration Team

LPDAAC Migration Team

GIBS Migration Team

NSIDC Prototyping Team

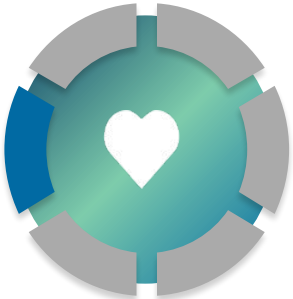
ASDC Prototyping Team

PO.DAAC and ASF Prototyping Teams

DAAC Team Organization

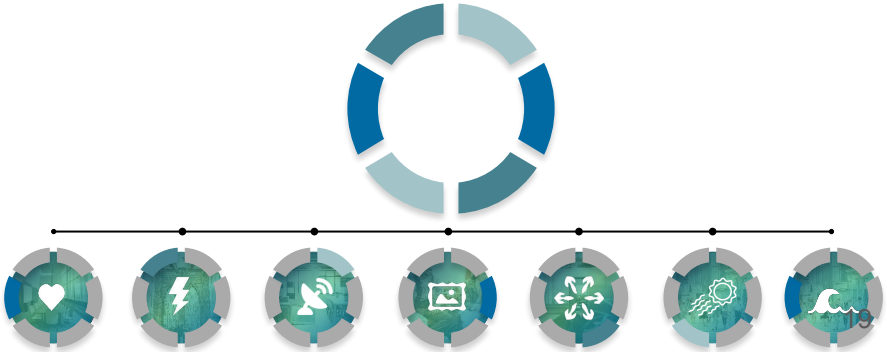
DAAC-owned Teams

- Led from internal DAAC resources
- Development, operators, testing
- Requests external support from Core team for specific tasks

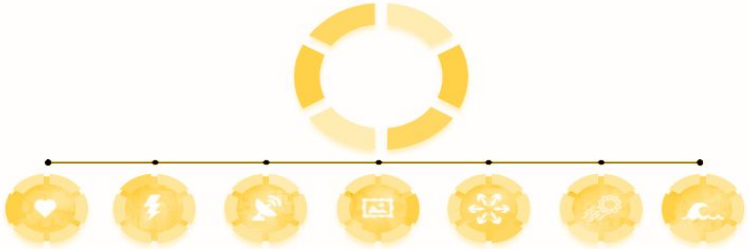


Cumulus Platform Core

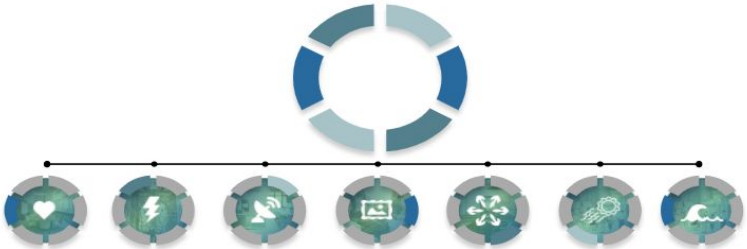
- Mainly ESDIS contracted development/design resources
- Working on smoothing out the delivery, versioning, on-boarding
- Aiding DAAC teams at specific intervals (development help, face to face meetings, etc) “We need help integrating xyz, etc”



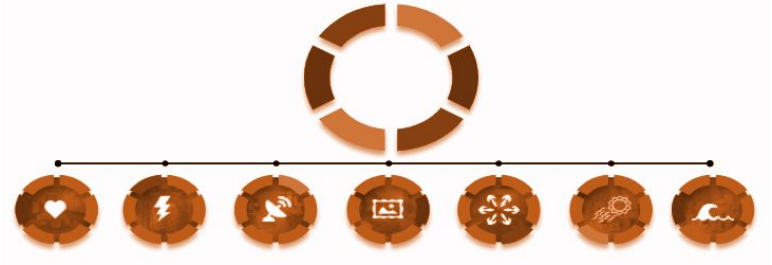
Ingest, Archive, and Distribution is only one of 3 three potential scaled agile teams



Data Usage (Clients and Services)



Ingest, Archive, Distribution



Cloud Hosting Platform

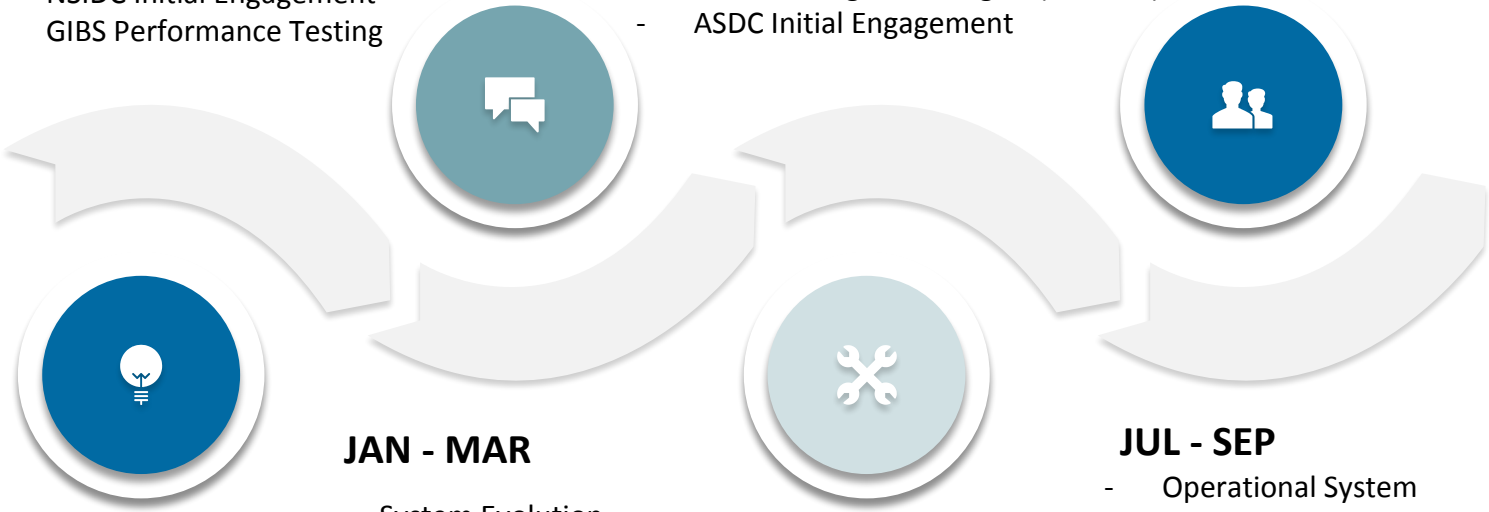
DAAC Cloud Migration PI Roadmap

OCT - DEC

- Planning Activities
- Process alignment
- NSIDC Initial Engagement
- GIBS Performance Testing

APR - JUN

- Operational Data Onboarding Begins (LPDAAC, GHRC, GIBS)
- Service Integration Begins (LPDAAC)*
- ASDC Initial Engagement

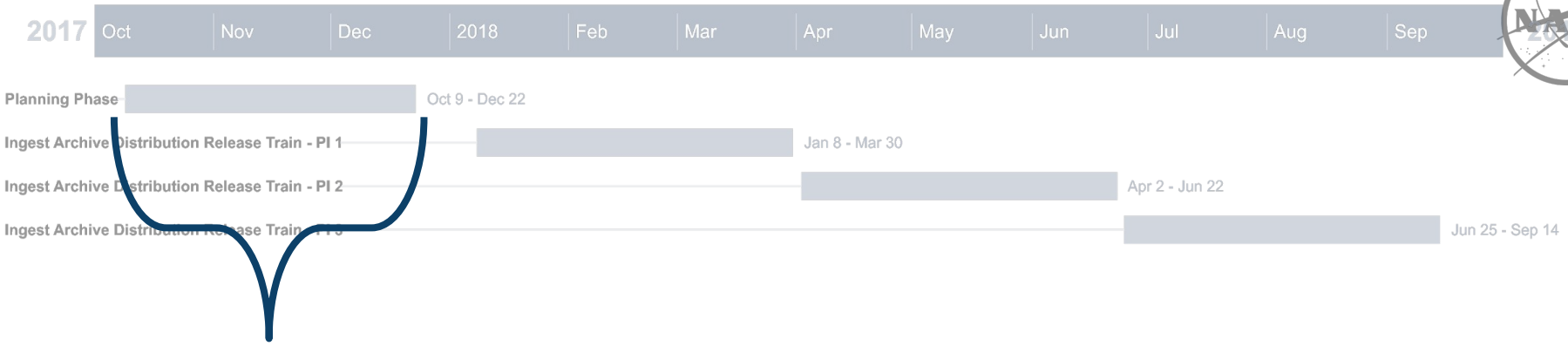


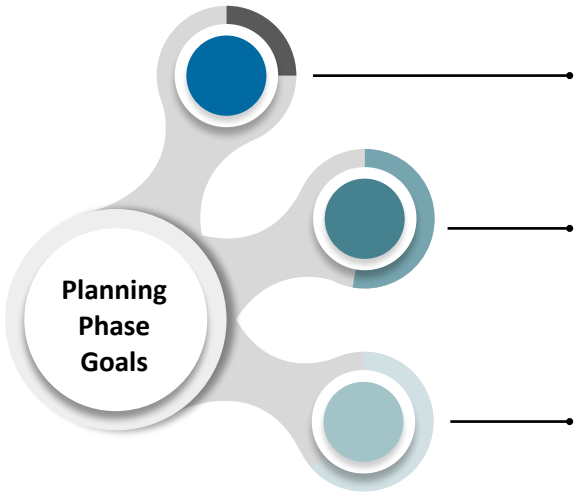
JAN - MAR

- System Evolution
- Collaboration Coordination
- NSIDC New Data Streams

JUL - SEP

- Operational System Testing (LPDAAC, GHRC, GIBS)
- Continued Service Integration
- ASDC New Data Streams





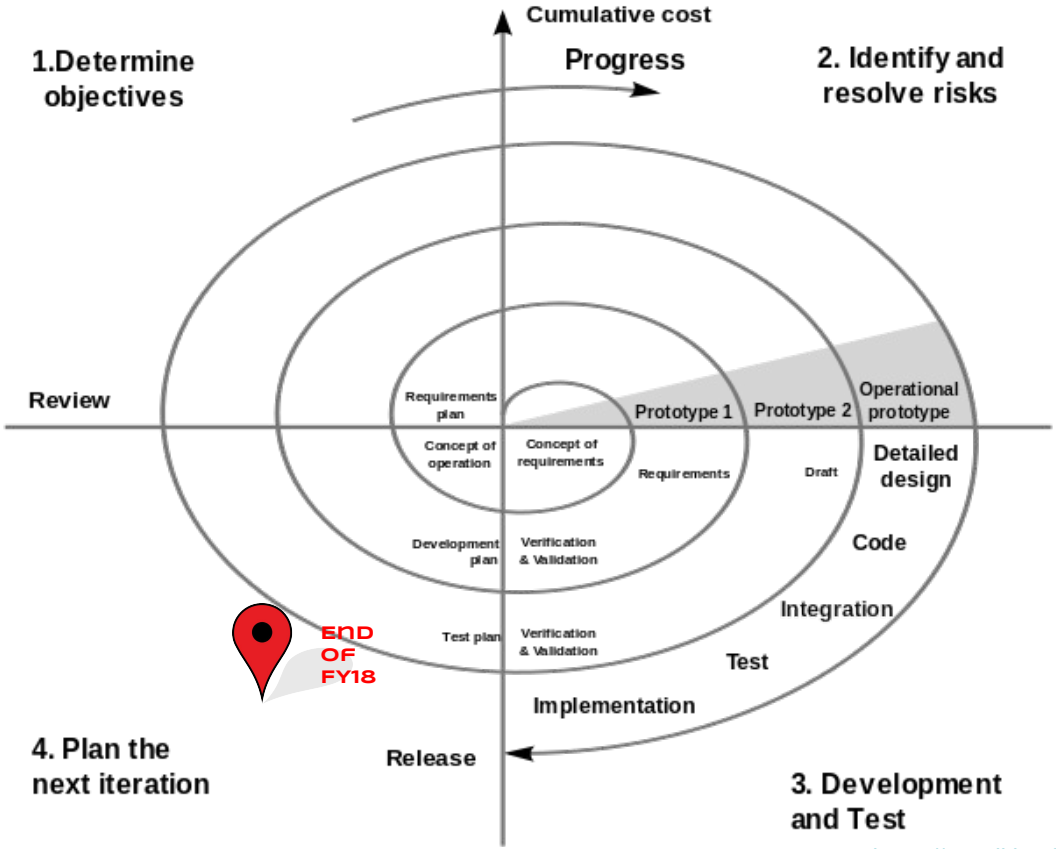
All migration teams are able to start onboarding data and metadata **no later than April 2018** in order to meet Earthdata Cloud 2021 milestones

All migration teams are capable of **standing up and operating Cumulus independently** in order to scale development capabilities

All migration teams have **detailed migration roadmap** for reaching identified milestones for FY 2018

- All migration teams are **capable of contributing and testing code/tasks/services to Cumulus** core in order to scale development capabilities **(March 2018)**
- **Cumulus Core Evolution and Cross-DAAC System Engineering Activities**
  - Updated, cloud-native **requirements** documentation
  - Cloud-native **metrics** system
  - **Backup and recovery** procedures for archived data, workflows, and tasks
  - **Distribution API and Egress Shaping**

Ingest/Archive/Distribution in FY18



https://en.wikipedia.org/wiki/Spiral_model

Overview

Code Governance Document: http://bit.ly/2h88Nm2

Cumulus Docs https://cumulus-nasa.github.io/

Cumulus Code Base: https://github.com/cumulus-nasa

DAAC Concept of Operations: Coming Soon

What about Code Governance?

Because this as an **organizational** shift,
not just a technology shift

Drafting a contribution guide

1. Provide working definitions of the high-level components of the Cumulus system, including specifying which of those components are governed by this document.
2. Establish roles and responsibilities for contributions to Cumulus NASA's EOSDIS
3. Identify key communication flows, as well as information on documentation, testing and deployment paradigms
4. Outline high-level process expectations for Cumulus contributions and provide example process flows for these contributions

Project Structure

This diagram lays out the relationships between the various roles in the team, showing how each team interacts at a high-level.

Project Management Board

sets the vision for

Project Leadership Team

provides the "what" for

Technical Leadership Team

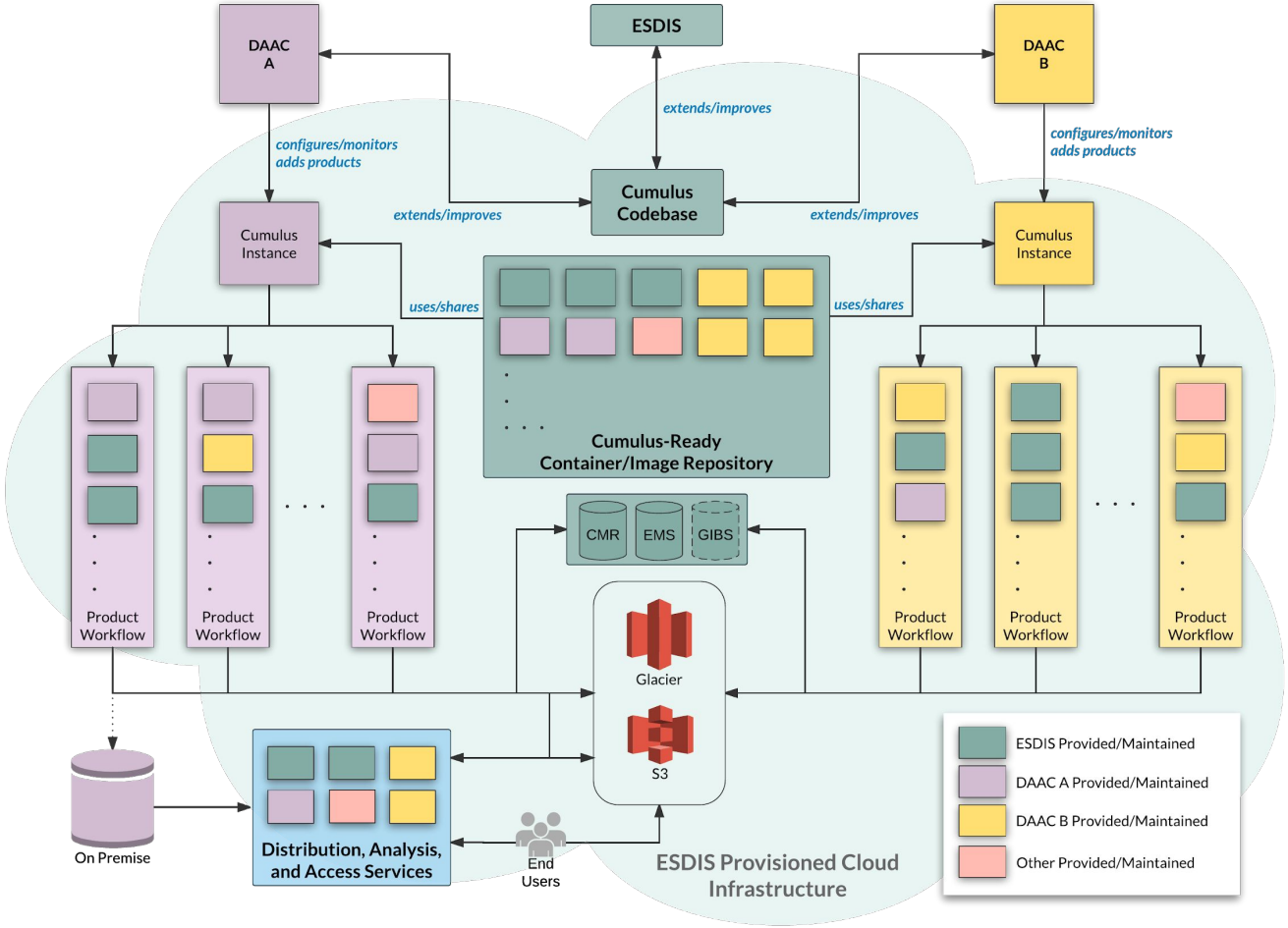
guides the "how" for

Committers

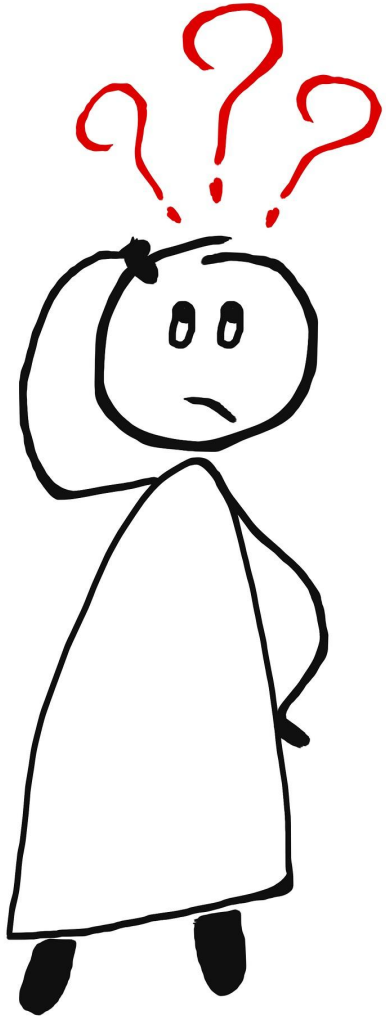
Contributors

As we scale (up or down) we can adjust/combine/tailor these roles. We can adapt to other projects/systems.

People group, leader, developer, teamwork, group icons by Oksana Latysheva from thenounproject.com (CC 3.0)


DAAC A
ESDIS
DAAC B
extends/improves
configures/monitors adds products
configures/monitors adds products
Cumulus Codebase
extends/improves
extends/improves
Cumulus Instance
Cumulus Instance
uses/shares
uses/shares
Cumulus-Ready Container/Image Repository
CMR
EMS
GIBS
Product Workflow
Product Workflow
Product Workflow
Product Workflow
Product Workflow
Product Workflow
Glacier
S3
On Premise
Distribution, Analysis, and Access Services
End Users
ESDIS Provisioned Cloud Infrastructure
ESDIS Provided/Maintained
DAAC A Provided/Maintained
DAAC B Provided/Maintained
Other Provided/Maintained

https://pixabay.com/en/questions-demand-doubts-psychology-1922476/ (License CC0)



Questions?

katie.baynes@nasa.gov

DAAC	Distributed Active Archive Center
SIPS	Science Investigator-led Processing System
EOSDIS	Earth Observation System Data and Information System
CMR	Common Metadata Repository
EMS	EOSDIS Metrics System
NGAP	NASA Compliant General Application Platform
ESDIS	Earth Science Data and Information System
GIBS	Global Imagery Browse Services
GHRC	Global Hydrology Resource Center
LP DAAC	Land Processes DAAC
PO DAAC	Physical Oceanography DAAC
ASDC	Atmospheric Science Data Center
NSIDC	National Snow and Ice Data Center
SAFe®	Scaled Agile Framework®
ASF	Alaska Satellite Facility

Acronym List!