CERES Data Management Team Working Group Report

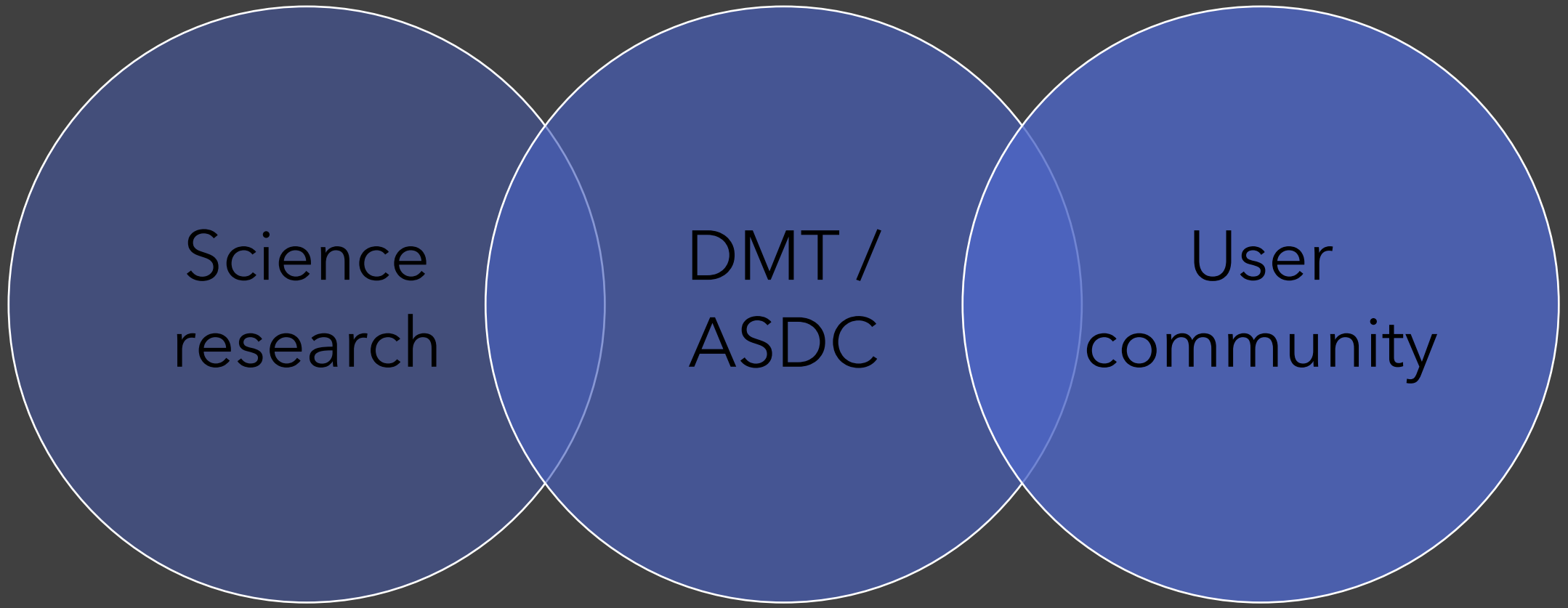
Katie Dejwakh
September 15, 2020

Outline

- DMT Overview
- Product Availability
- DMT Activities
- Code Re-architecture
- ASDC & Systems

DMT Overview

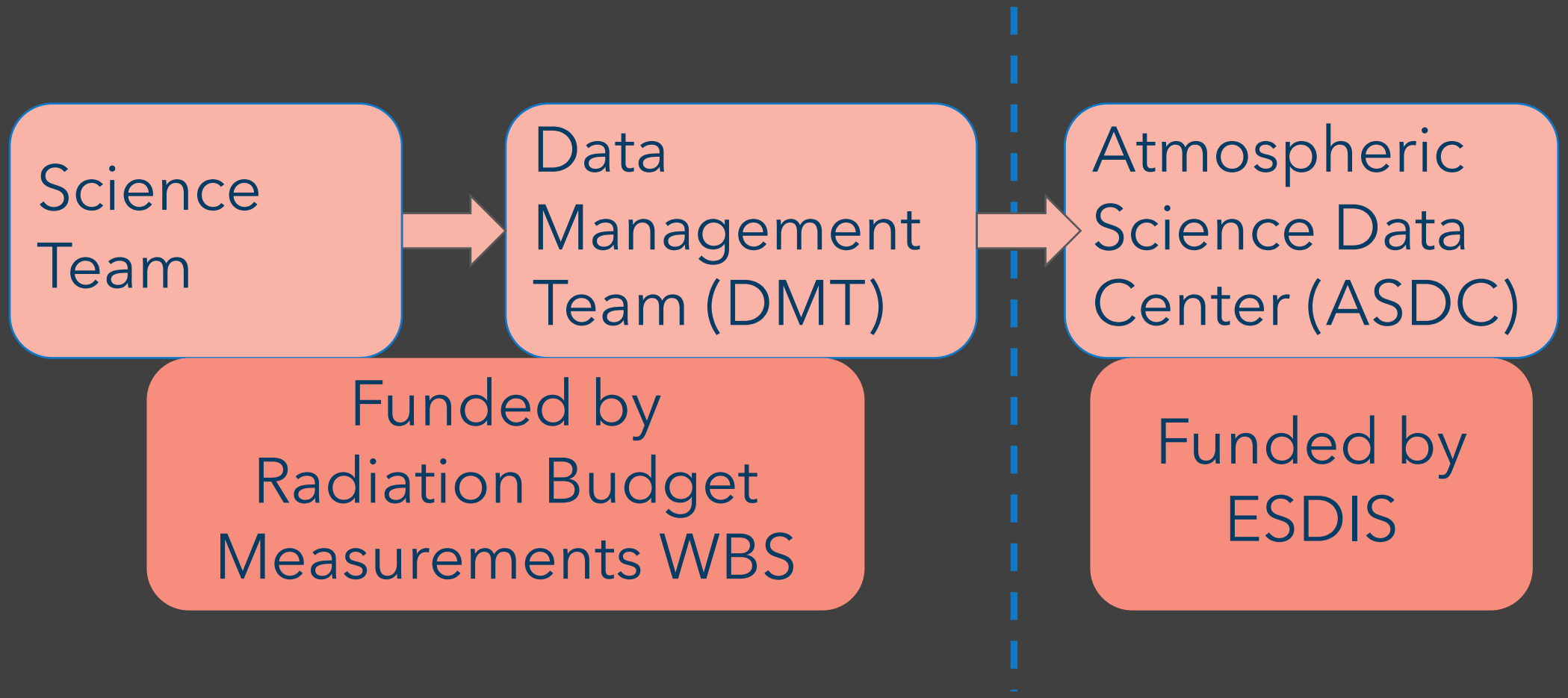
Data Management Team (DMT)



Data Management Team (DMT)

Instrument	ERBE-like	Clouds
Inversion	SARB	TISA
	Systems & CM	

DMT-ASDC Interface



Product Availability

Edition 4 Terra & Aqua

Product	Platform	Processed Thru	Publicly Available?
BDS	Terra, Aqua	May ‘20	Yes
SSF			
SSF1deg-Hour		Apr. ‘20	
SSF1deg-Day/-Month			
SYN1deg-Hour/3Hour/MHour	Terra+Aqua		
SYN1deg-Day/-Month			

Edition 4 Terra & Aqua

Product	Platform	Processed Thru	Publicly Available?
CldTypHist	Terra+Aqua	Apr. ‘20	Yes
FluxByCldTyp			
EBAF		Mar. ‘20	
EBAF ToA		May ‘20	

Edition 1 S-NPP

Product	Platform	Processed Thru	Publicly Available?
BDS	S-NPP	Jul. '20	Yes
SSF		Jul. '20	
SSF1deg-Hour		Sep. '19*	
SSF1deg-Day/-Month		Sep. '19*	
SYN1deg-Hour/3Hour/MHour	Terra+S-NPP	Nov. '17*	
SYN1deg-Day/-Month		Nov. '17*	

* L3 processing is paused. Instrument in RAPS mode

DMT Activities

Forthcoming Data Product Editions

Edition 2 S-NPP

- Processing thru: Instrument & ERBE-like
- Available early Winter 2020
- Need C2 VIIRS data
- No L3 products from 9/2019 onward, while S-NPP in biaxial mode

Edition 1 NOAA-20

- Processing thru: Instrument, SSF, and Level 3
- Available late Fall 2020 – improved Clouds!

Infrastructure

- Automation support for L1-3 processing of:
  - Ed 1 NOAA-20
  - Ed 2 S-NPP

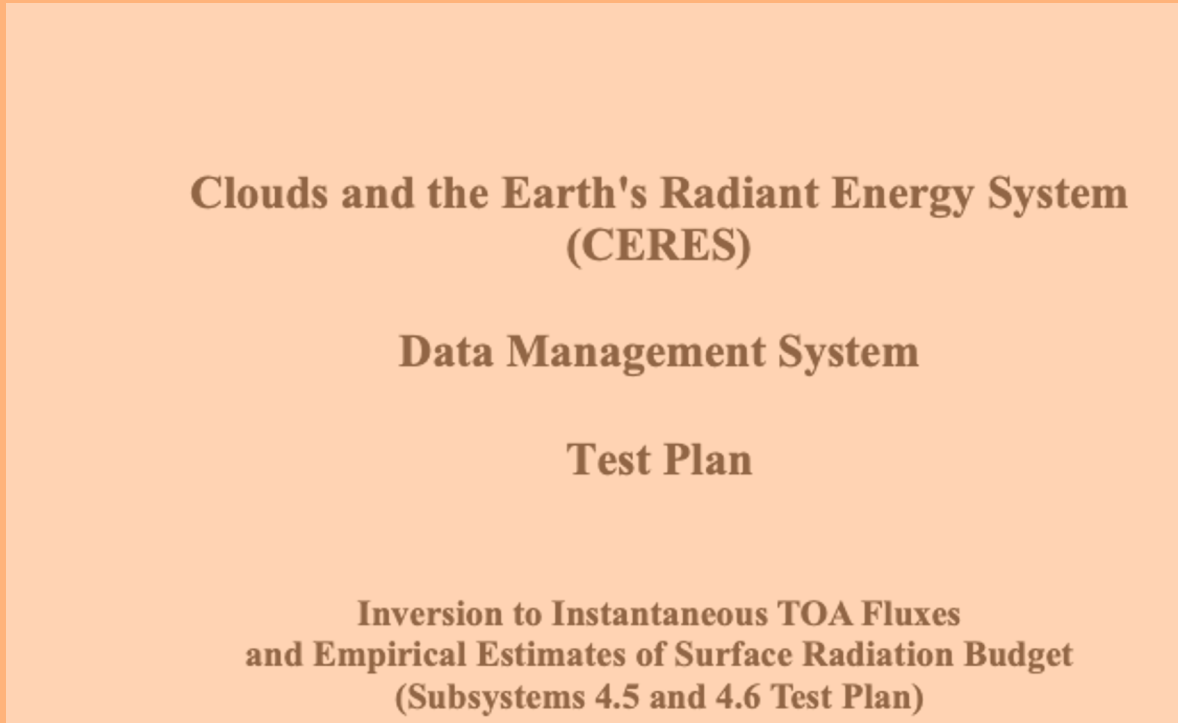
- Enhanced archiver streams:
  - Parallelizing streams – interleaving forward- and re-processing
  - Freezing streams

- Enable capability to use gains from previous month to temporarily apply to the first day of next month

Configuration Management

- Improved documentation – PGE-specific:
  - Operator's Manuals
  - Test Plans

- Quicker access to distilled information



Clouds and the Earth's Radiant Energy System
(CERES)

Data Management System

Test Plan

Inversion to Instantaneous TOA Fluxes
and Empirical Estimates of Surface Radiation Budget
(Subsystems 4.5 and 4.6 Test Plan)

Production Request Tool

- OpenShift

- Opportunities to leverage virtual layer to
  - Create
  - Upgrade
  - Roll back
  - Destroy

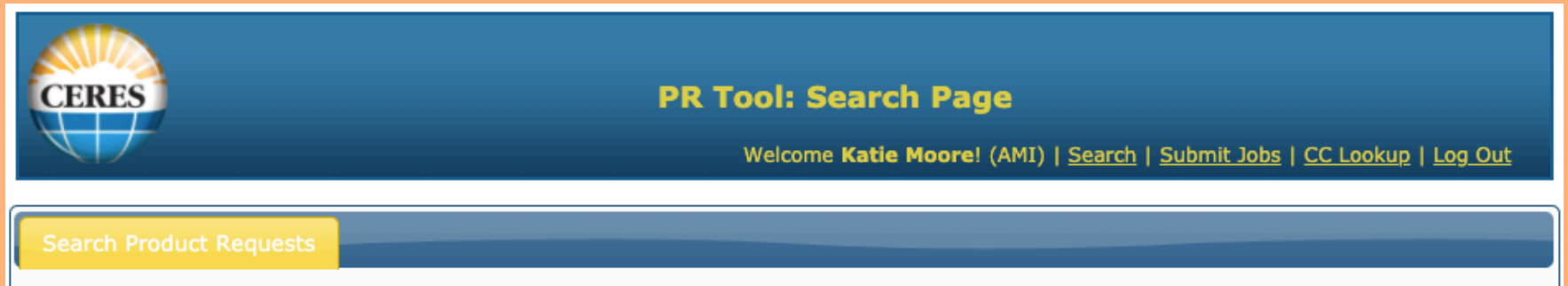
  processes more easily, as necessary

- Flexibility front- and back-end

Production Request Tool

- Multi-"pod" application deployment
  - Storing session data outside of pod
  - Backup pod retrieves data if primary pod down

- Automated database updates (and rollbacks)
  - Database changes now in version control (Phinx)
  - PRTool web code changes tied to database state

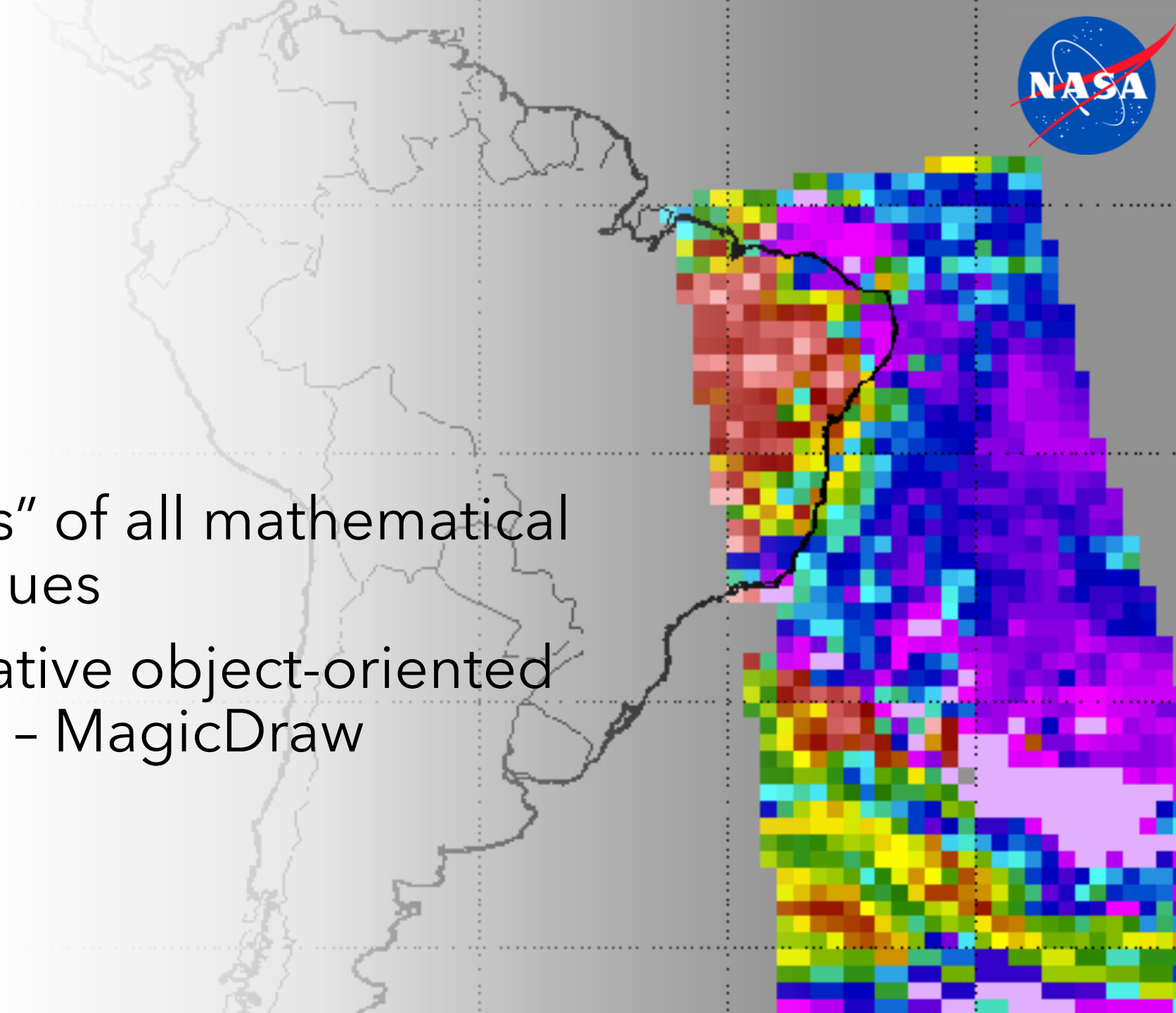


Code Improvement Activities

Clouds Status

Cloud mask

- Taking “census” of all mathematical operations/values
- Mapping tentative object-oriented class structure – MagicDraw

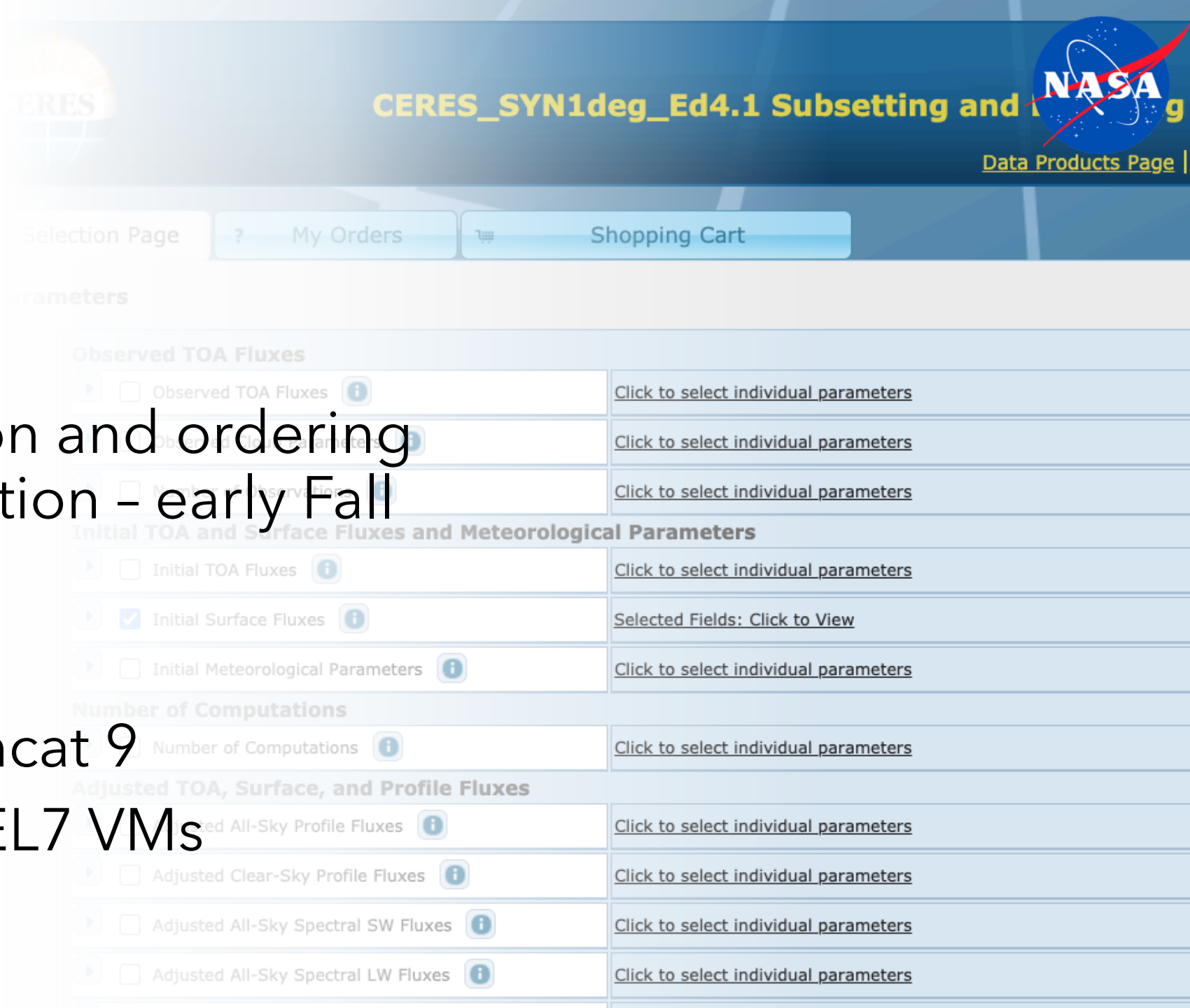


TISA Status

- Fortran (current codebase)
  - Cloud property container
    - Compatibility with JSON format
    - Testing averaging routines
- C++ (future codebase)
  - Gridding implementation
  - JSON I/O for properties

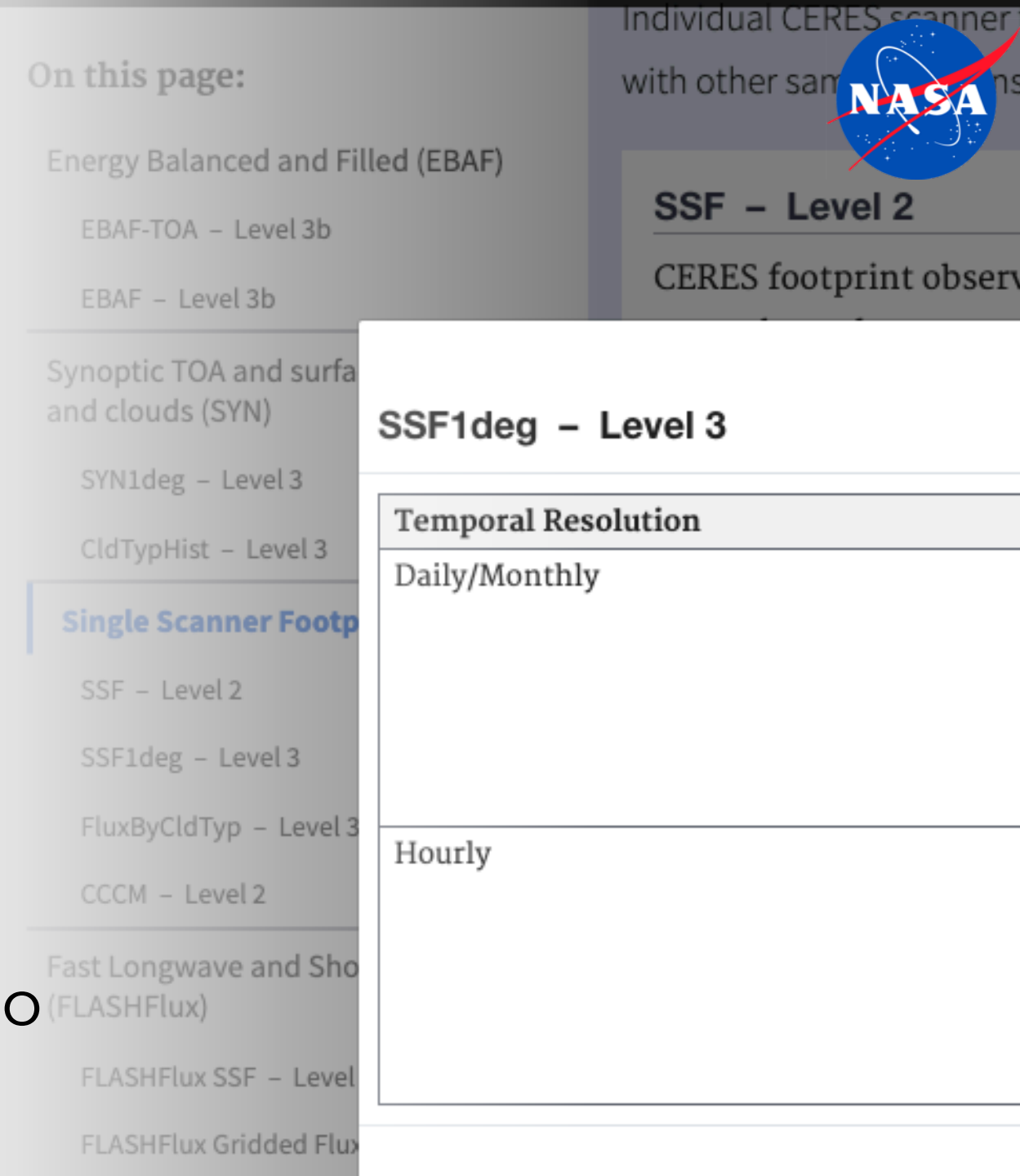
Subsetter/ Ordering Tool

- Data visualization and ordering hardware migration – early Fall
- Upgrades:
  - Python 3
  - Java 11 / Tomcat 9
- Running on RHEL7 VMs



Subsetter/ Ordering Tool

- Tailoring new CERES website data product interface
- Hooking to Ordering Tool
- Continuing to update Tool to accept upcoming Editions



ASDC & Systems

ASDC: Targeted Engagement – Processing

- Long-term: clarifying DAAC procedures, especially with respect to data flow

- Investigating future data staging design
  - Mount DPO
  - Separate staging area
  - Hybrid?

- CERES processing data volume is significant

ASDC: Targeted Engagement – Ingest

- New ingest system in development
- Operational by June 2022
- Ensuring smooth CERES data delivery:
  - Product Delivery Record (PDR) format update
  - Ingest priority levels
  - Tarred-up Process Control (PCF) and Log Files

Updated Metadata

Upcoming Editions

- Conventions:
  - Unified Metadata Model (UMM)
  - Climate-Forecasting (CF)
  - Attribute Convention for Data Discovery (ACDD)
  - CERES-specific

- Attributes in HDF5/netCDF4 files

- Separate "*.met" file until no longer necessary

Updated Data Formats

Upcoming Editions:

HDF5/netCDF4

- Many iterations with ASDC – ingest/metadata check
- First products: Ed 1 NOAA-20 and Ed 2 S-NPP **SSF**

ASDC: New Systems

Store… (high-performance, DMZ)

Process…

EMC Isilon Storage

- Computational Research Facility (CRF)
- Configuration in progress

HPE DL360 Servers

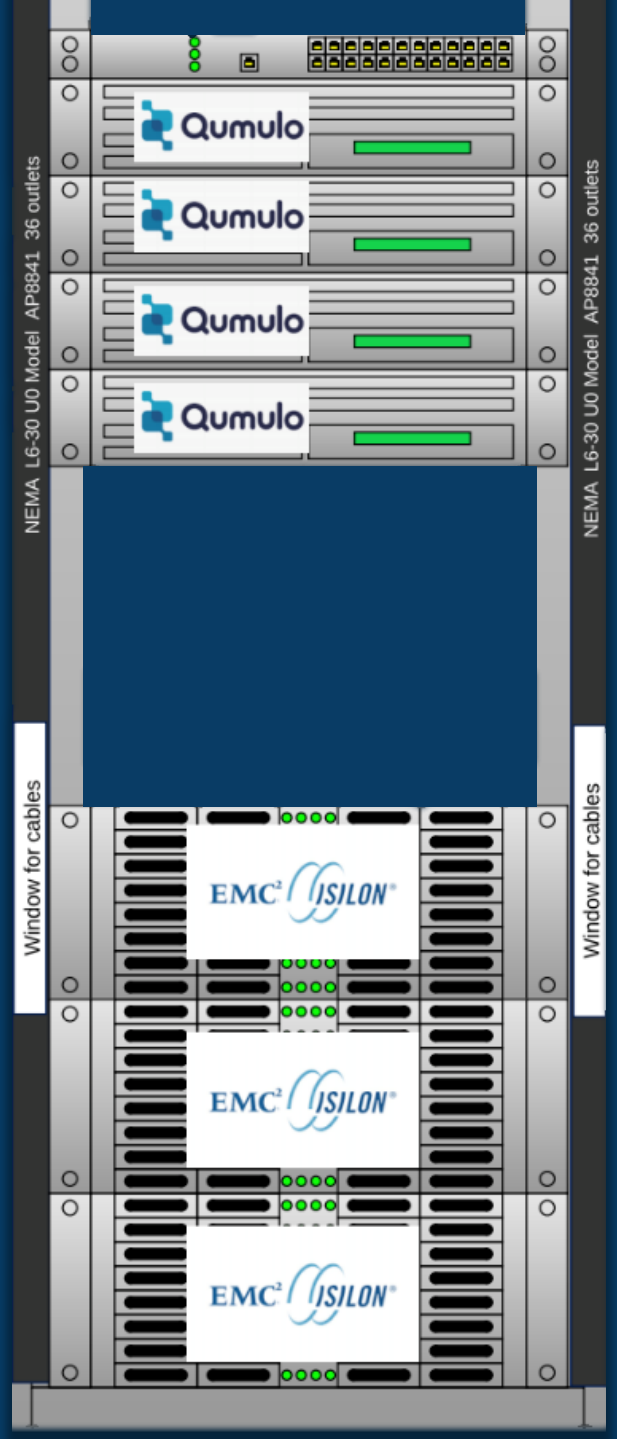
- ASDC then CRF
- Virtualized (RHEL7)
- Batch processing

ASDC: New Systems

Store… (internal)

HPE Scale-Out NAS Storage with Qumulo

- CRF
- Fall 2020
- Storage by-the-yard



Summary

- New editions on the way!
- Continuing to virtualize
- Engaging in cloud preparation
- Refreshing file formats, metadata
- Improving user experience