

NASA Advancing Collaborative Connections for Earth System Science (ACCESS): **Data Access and the ECCO Ocean and Ice State Estimate**

Patrick Heimbach

*Institute for Computational Engineering and Sciences
The University of Texas at Austin*

Ian Fenty & Thomas Huang

*Jet Propulsion Laboratory
California Institute of Technology*

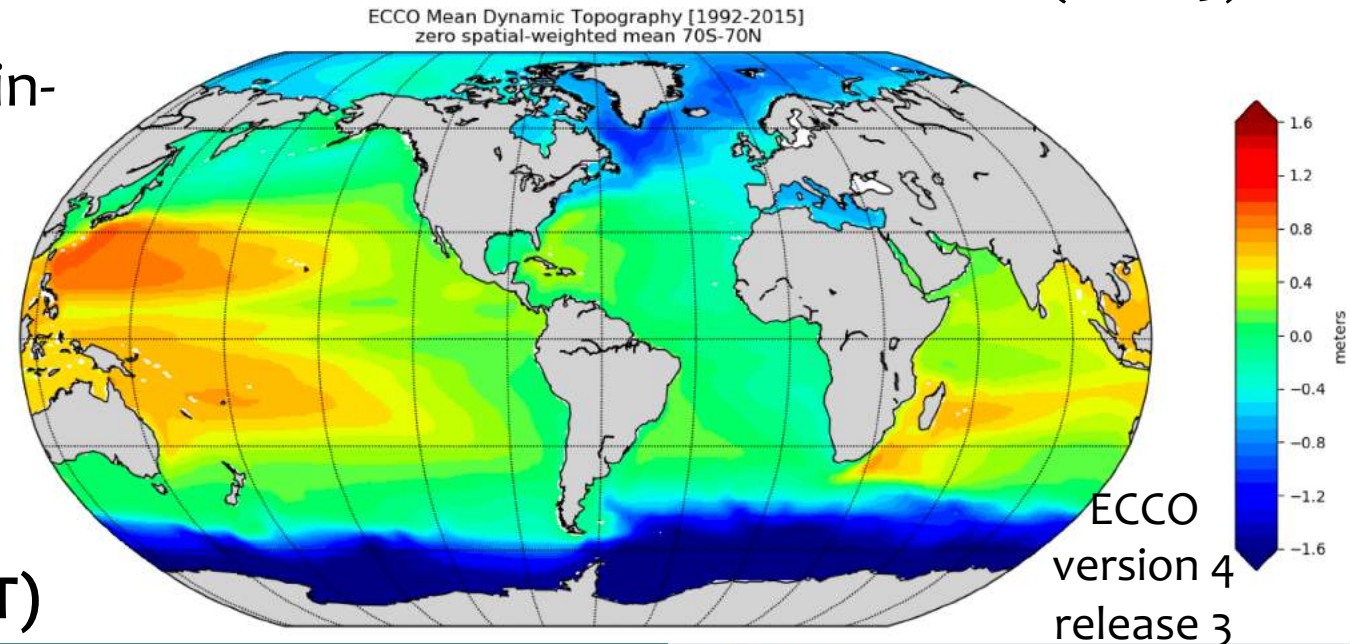
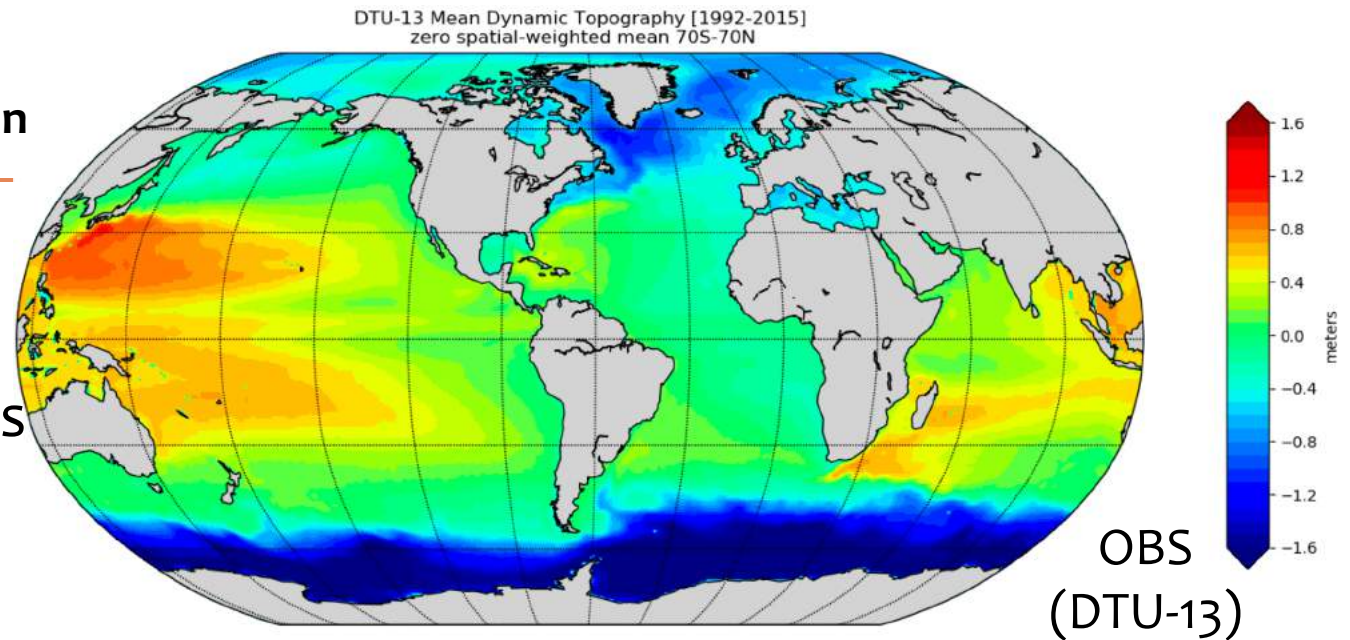
What is ECCO?

Estimating the Circulation and Climate of the Ocean

A mathematically rigorous, advanced data tool that:

- synthesizes diverse Earth observations into a complete, physically consistent description of the 3-D time-varying global ocean and sea-ice state;
- combines more than 10^8 satellite and in-situ obs. with a state-of-the-art ocean and sea-ice general circulation model;
- enables accurate tracking of regional heat, salt, freshwater and sea level anomalies through the Earth system

E.g.: Mean Dynamic Topography (MDT)

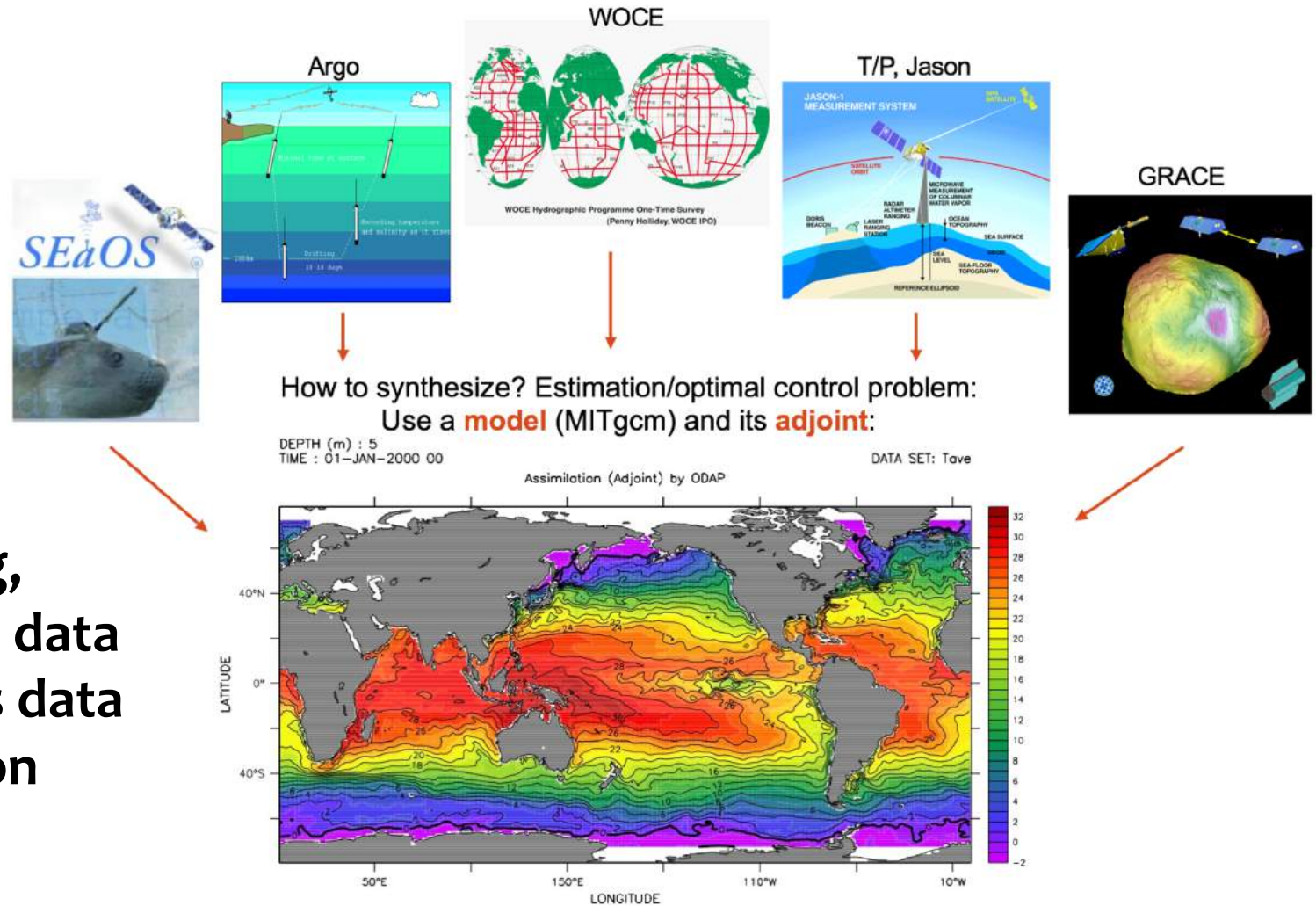


What is ECCO?

Estimating the Circulation and Climate of the Ocean

By integrating all measurements simultaneously into a common dynamically-consistent ocean model framework, ECCO makes optimal use of these valuable, diverse, and heterogeneous data streams

Achieving, maintaining, updating, improving and automating these data streams represents an enormous data preprocessing and transformation cyberinfrastructure challenge.



High-level goals

- I. Expand and accelerate the integration of NASA Earth system data into NASA's premier ocean and sea-ice reanalysis product, the Estimating the Circulation and Climate of the Ocean (ECCO) ocean & sea state estimate, in a sustainable and scalable manner

- II. Facilitate and expand the scientific utilization of NASA remote sensing data integrated in the ECCO state estimate by the growing community of interdisciplinary researchers in the oceanographic, sea-ice, sea level rise, and climate science fields.

GOAL I:

Expand and accelerate integration of NASA Earth system data into ECCO

Major activities for ingesting data into ECCO are

- preprocessing:
 - evaluating the data using statistical methods, quality control flags, etc., to infer robust uncertainty estimates
- transformation:
 - reprojecting and regridding data from their original spatial configs. (e.g., along-track, swath, or gridded) to the ocean model grid
 - converting from their native formats (e.g., NetCDF) to the flat unstructured binary files

GOAL I:

Expand and accelerate integration of NASA Earth system data into ECCO

A new ECCO cloud data provisioning system (DPS) is needed to address the sustainability, scalability, and time-lag issues related to integrating NASA (and other) data sets into ECCO

We will develop a new cloud-native storage and software system to automatically update, preprocess, and transform NASA ocean and sea-ice observations provided by EOSDIS close to the source of the data rather than at the multiple institutions.

The goal is for ECCO to always incorporate the most recent data streams, even as the number and volume of data streams continue to increase

GOAL II:

Facilitate and expand the utilization of ECCO

Facilitate and expand the utilization of NASA remote sensing data integrated in ECCO by the growing community of interdisciplinary researchers

- Current system for distributing the ECCO products is limited and neither sustainable nor scalable.
- ECCO distribution extends only to FTP site (or related WebDAV)
- offers none of the advanced distribution capabilities now deployed across EOSDIS (e.g., common access interface)

Data harmonization:

- translate ECCO product metadata according to standards like ISO-19115, Global Climate Change Directory (GCMD), ECCHO-10, and the latest Unified Metadata Model for Collections (UMM-C) (ESDIS,2018)

GOAL II:

Facilitate and expand the utilization of ECCO

Radically streamline the distribution of updated ECCO products to EOSDIS, in particular to the Physical Oceanography Distributed Active Archive Center (PO.DAAC) via new cloud-native storage and software systems

Share common, open source, data analytic and visualization architecture that is being used by the NASA Sea Level Change Portal and PO.DAAC's new cloud-based data analytic solution

Provide new cloud-native tools to provide users the ability to access and conduct quantitative analyses on ECCO state estimates via a Data Analysis System (DAS)

GOAL II:

Facilitate and expand the utilization of ECCO

Leverage NASA's common access soution through the Earthdata enterprise Common Metadata Repository (CMR) and the Earthdata search capability

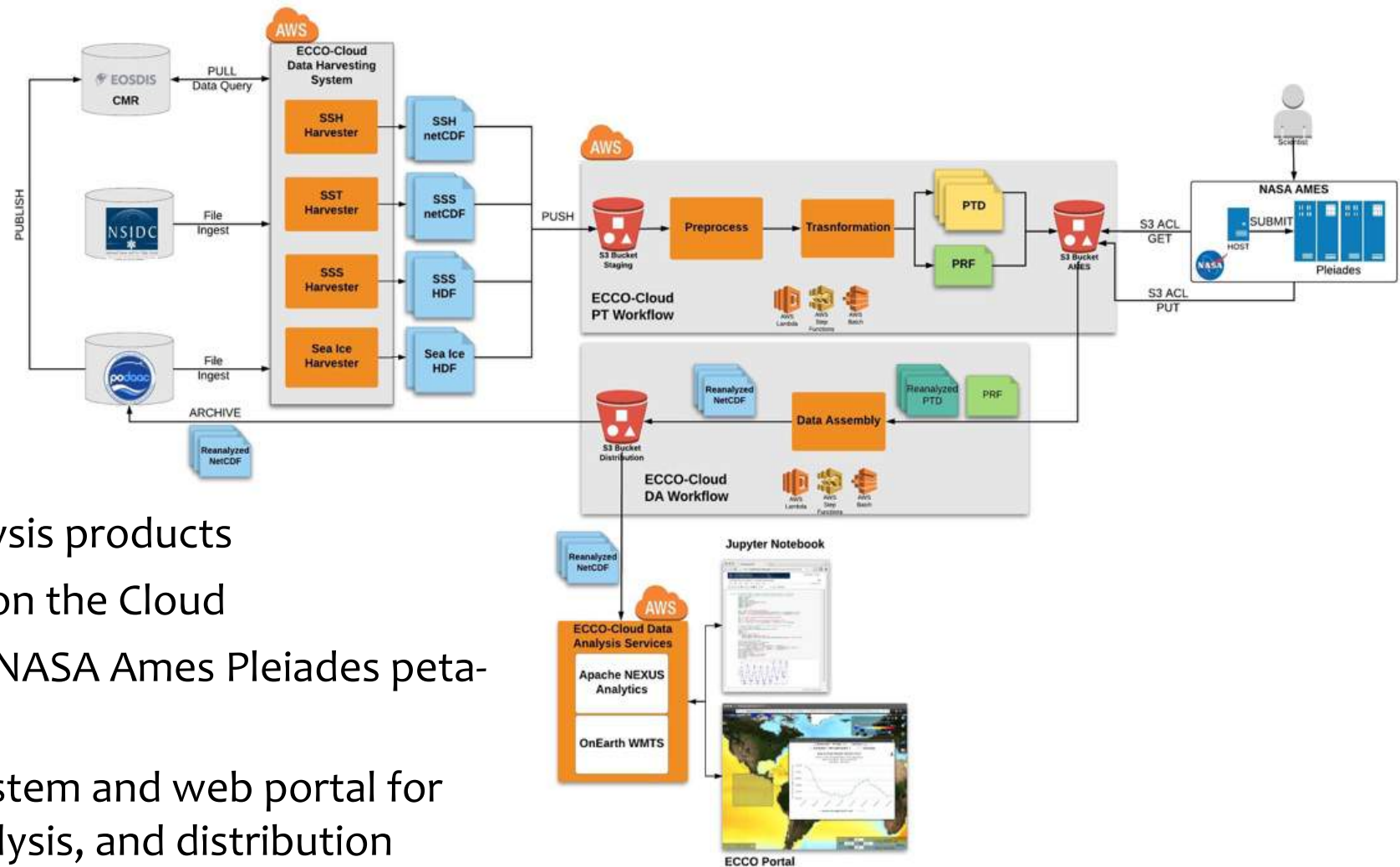
Provide web-portal and analytic platform on the Amazon Web Service (AWS) environment for users so that they can (a) visualize and analyze the full state estimate, and (b) formulate and conduct their own experimental simulations, based upon the ECCO configuration

Expected significance

- **data provisioning system** will
 - decrease the amount of ECCO scientist's time spend manually updating data sets,
 - reduce the time lag between release of new data and their integration,
 - through (a) and (b) increase the value and utilization of ECCO
 - enable regular extensions to the state estimate time window
- **data distribution system** will
 - accelerate the distribution of ECCO products to PO.DAAC and other distribution channels
 - improve its accessibility to the research community
- **data analysis system** will
 - increase access, exploration, and use of ECCO products and NASA data through web-based tools and web-service interfaces

The ECCO Cloud Data Provisioning System (DPS) Architecture

- Leverage NASA-funded open source big data analytic solution, Apache SDAP (<https://sdap.apache.org>)
- Automate ingestion, processing and packaging of ECCO reanalysis products
- Automate delivery to PO.DAAC on the Cloud
- Integrating Amazon Cloud with NASA Ames Pleiades petascale supercomputer
- Establish ECCO Data Analysis System and web portal for interactive visualization and analysis, and distribution



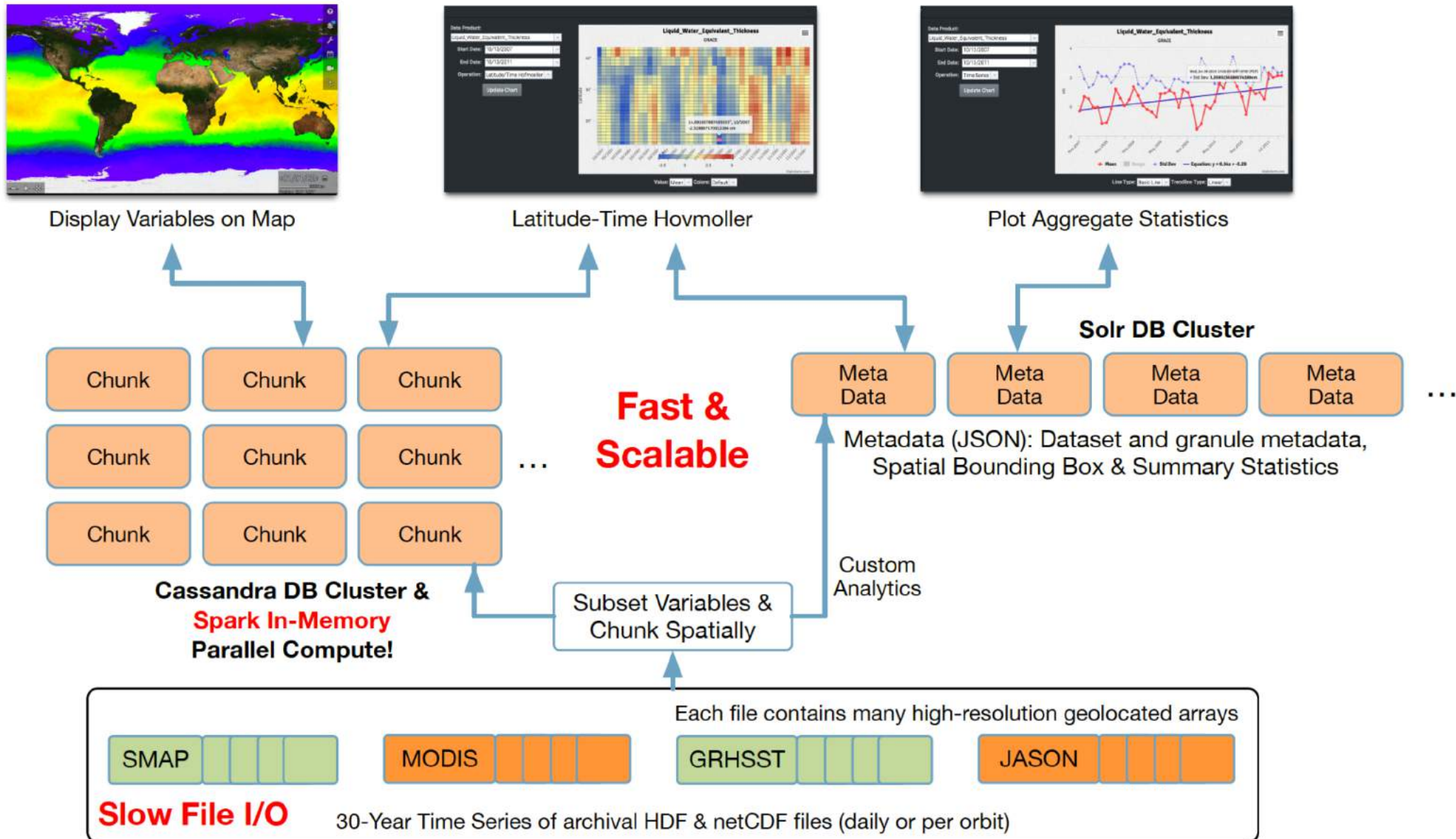
The ECCO Cloud Data Provisioning System (DPS) Architecture

Automation of data discovery and data provisioning in DPS requires the following inputs:

- TBD (p. 10/11)

NEXUS' tile-based architecture for science data processing and analysis on the cloud

• ...



ECCO-CLOUD: Data Distribution System (DDS)

- **Transformation rules:**

- describe the conversion of model output from its native 13-tile Cartesian grid to user-friendly geographical projections

- **Data assembly rules:**

- describe how to repackage the model's flat binary output files into NetCDF

- **Announcement rules:**

- describe how to alert various distribution systems to the presence of new ECCO products

ECCO-CLOUD: Data Analysis System (DAS)

- p.11/12

EOSDIS data products from active instruments ingested by ECCO

Variable	Data Product	Level	Daily mb
Sea surface height	Jason-3 GPS based orbit and SSHA OGDR	2	25
	SARAL Near-Real-Time Value-added Oper. Geophys. DR Sea Surface Height Anom.	2	25
Sea surface temperature	MODIS Aqua & Terra Level 3 SST Thermal IR Daily 4km Day & Night v2014.0	3	40
	MODIS Aqua & Terra Level 3 SST Mid-IR Daily 4km Day & Night v2014.0	3	20
	Reynolds NCEP Level 4 Optimally Interpolated SST Weekly Version 2	4	1
Sea surface salinity	JPL SMAP Level 2B CAP Sea Surface Salinity and Extreme Winds V3.0 Dataset	2B	200
	JPL SMAP Level 3 CAP Sea Surface Salinity Std Mapped Img 8-Day Run Mean V3.0	3	200
Sea-ice concentration	Near-Real-Time NOAA/NSIDC Climate Data Record of Pas. Mic. Sea-ice Conc, V1.	4	4
	NOAA/NSIDC Climate Data Record of Pas. Mic. Sea-ice Conc., V3	4	4
	MASAM2: Daily 4 km Arctic Sea-ice Concentration, Version 1	4	2
Ocean bottom pressure	JPL GRACE Mascon Ocean, Ice, and Hydrology Equivalent Water Height JPL RL05M.1 Version 2	2	2

Amazon Web Services Cost Management

- p. 13