Joint Center for Satellite Data Assimilation

Presentation to the
2nd National Operational Processing Centers
Observational Data Workshop

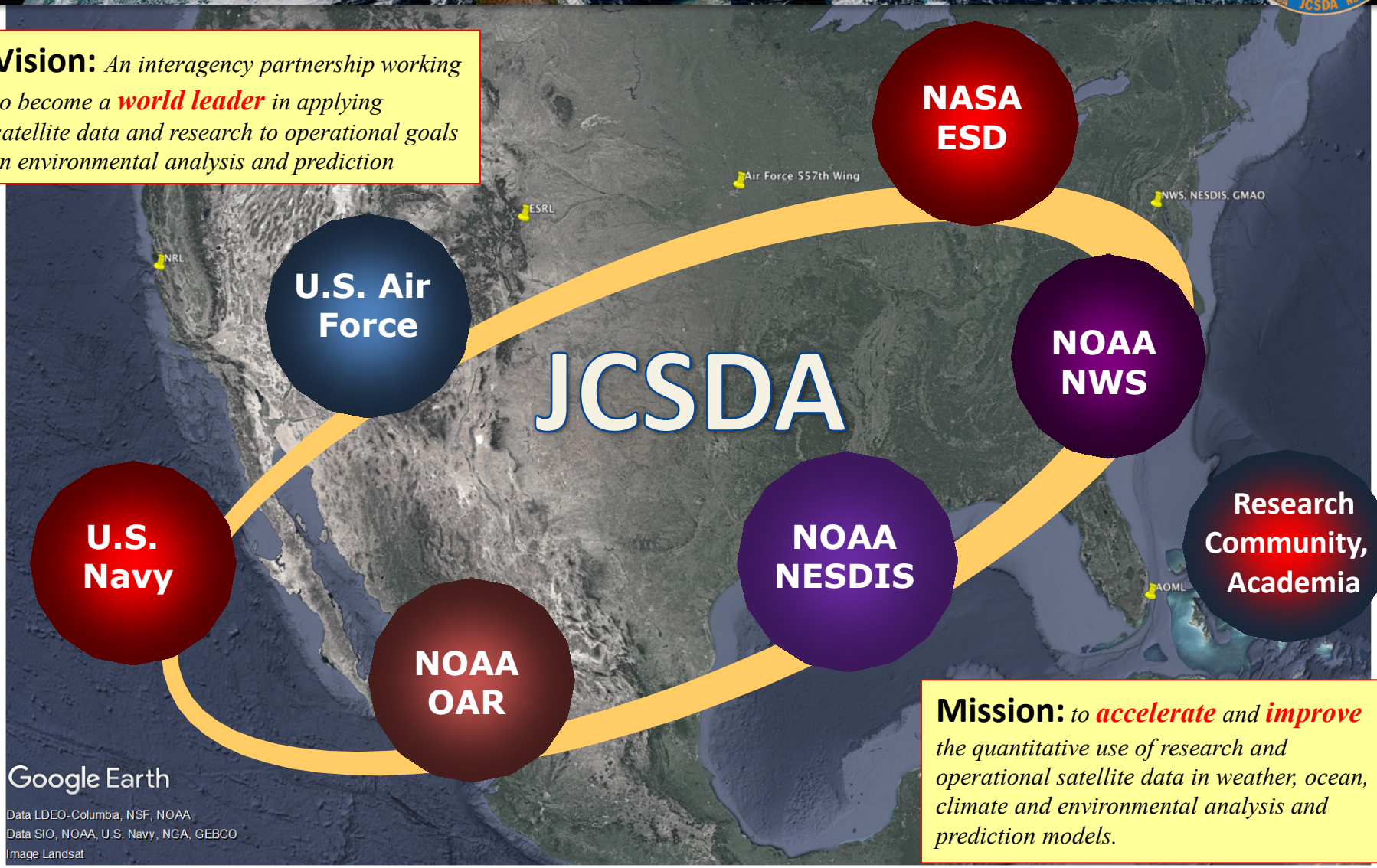
Jim Yoe, Chief Administrative Officer for the Joint Center for Satellite Data Assimilation (JCSDA)

Overview

- Partners, Mission, and Vision
- Management Structure & CONOPS
- Project and Project Management
  - Joint Effort for Data assimilation Integration (JEDI)
- Top Three Satellite Data Challenges

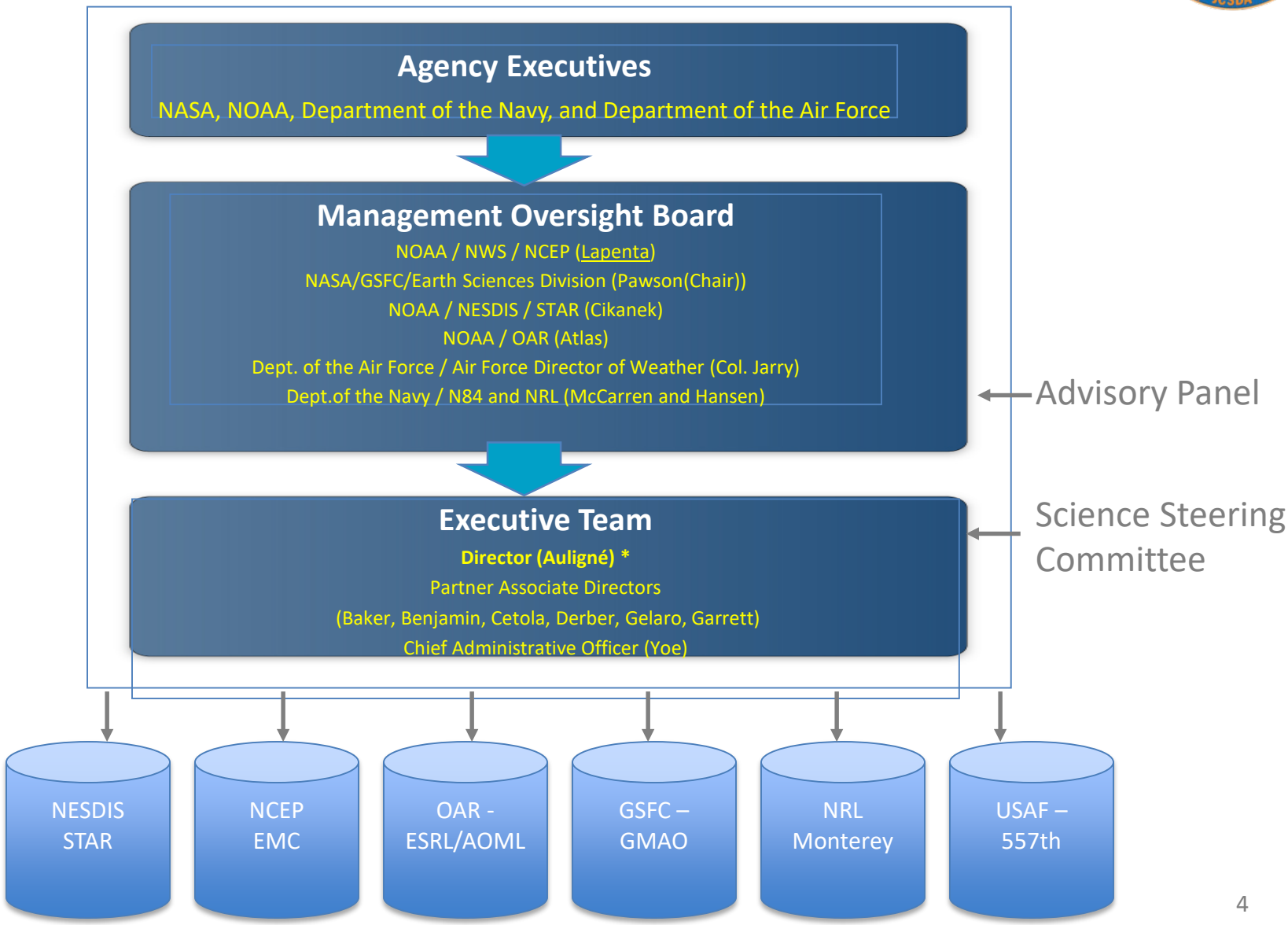
JCSDA Partners, Mission, and Vision

Vision: *An interagency partnership working to become a **world leader** in applying satellite data and research to operational goals in environmental analysis and prediction*



Mission: *to **accelerate** and **improve** the quantitative use of research and operational satellite data in weather, ocean, climate and environmental analysis and prediction models.*

Management Structure



Concept of Operations

• The reaffirmation of the **central role** of the **Executive Team** to guide science activities and ensure high level of collaboration, and of the **Management Oversight Board** to provide management-level oversight and strategic decisions.

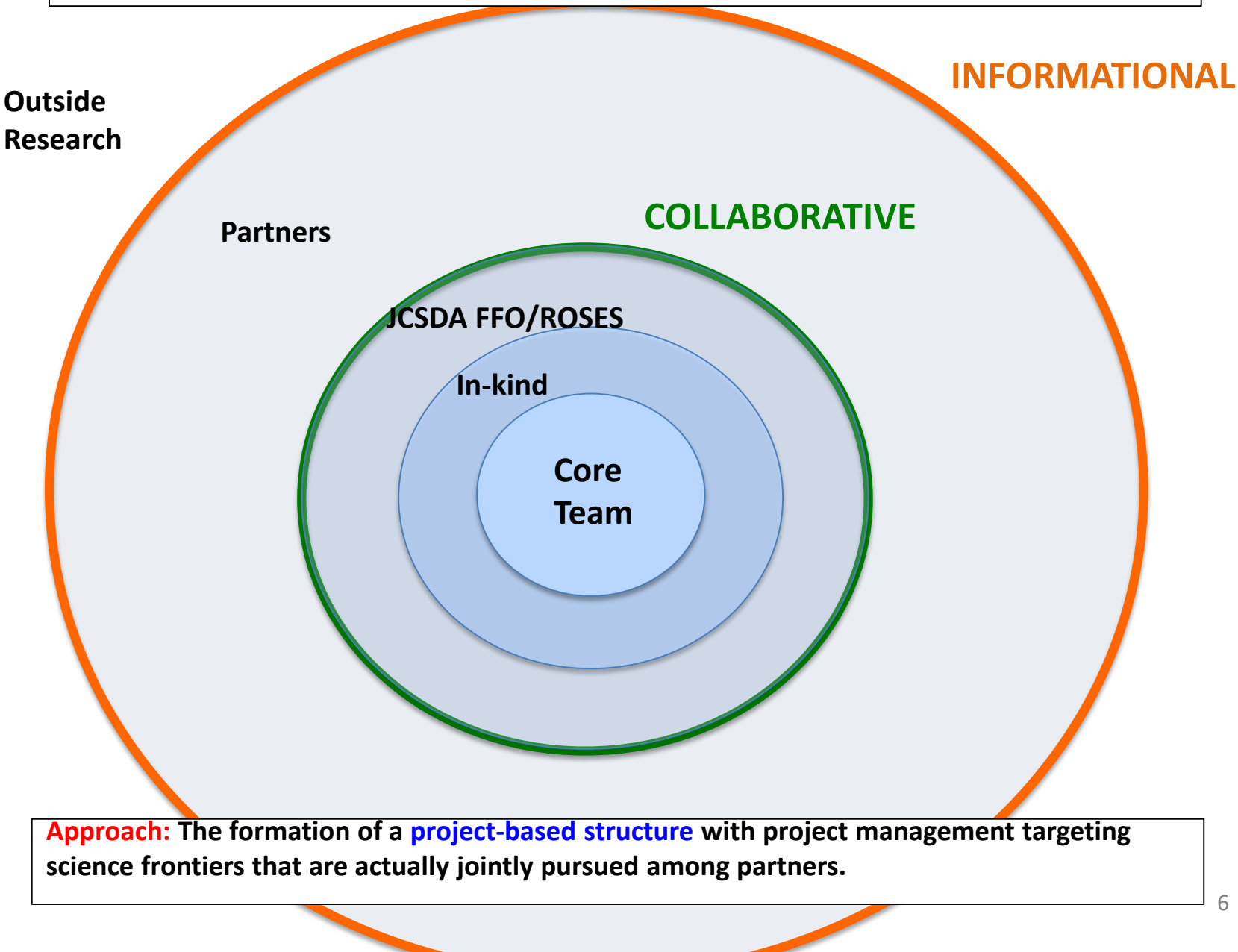
• The **transition of programmatic, administrative, and operational management** to a Non-Government Research Organization (NGRO), which will increase accountability to the JCSDA Director while maintaining close interaction with and oversight from the partner federal agencies.

• The clarification of the **scope of activities** and the associated decision process to determine what constitutes the purview of the JCSDA.

• The formation of a **project-based structure** with project management targeting science frontiers that are actually jointly pursued among partners.

• The establishment of a **formalized annual cycle** to coordinate the planning, budgeting, execution and reporting of JCSDA activities.

Scope of activities of JCSDA: Collaborative, inter-dependent activities inside AOP
Metric of success = *added* value of doing work *jointly* via the JCSDA



Approach: The formation of a project-based structure with project management targeting science frontiers that are actually jointly pursued among partners.

JCDSA Projects

- Director's Office (DOF)
- **New and Improved Observations (NIO)**
  - Prepare for the assimilation of JPSS, GOES-16, COSMIC-2, evaluate Satellite Commercial data
  - Assimilation of radiances over land and sea-ice with improved estimation of surface emissivity
  - Improved use of all-sky radiances
- **Impact of Observing Systems (IOS)**
  - OSEs, FSOI, etc.
- **Community Radiative Transfer Model (CRTM)**
  - Acceleration via software optimization
  - Improved scattering tables for clouds and precipitation
- **Sea-Ice, Ocean, Coupled Analysis (SOCA)**
  - Build Sea-ice DA components following JEDI standardized observation access
  - Initial integration into unified forward operator
- **Joint Effort for Data assimilation Integration (JEDI)**
  - Unified Data Assimilation Planning Workshop
  - Prototype of Unified Forward Operator
  - Requirements and initial prototype of standardized observation access

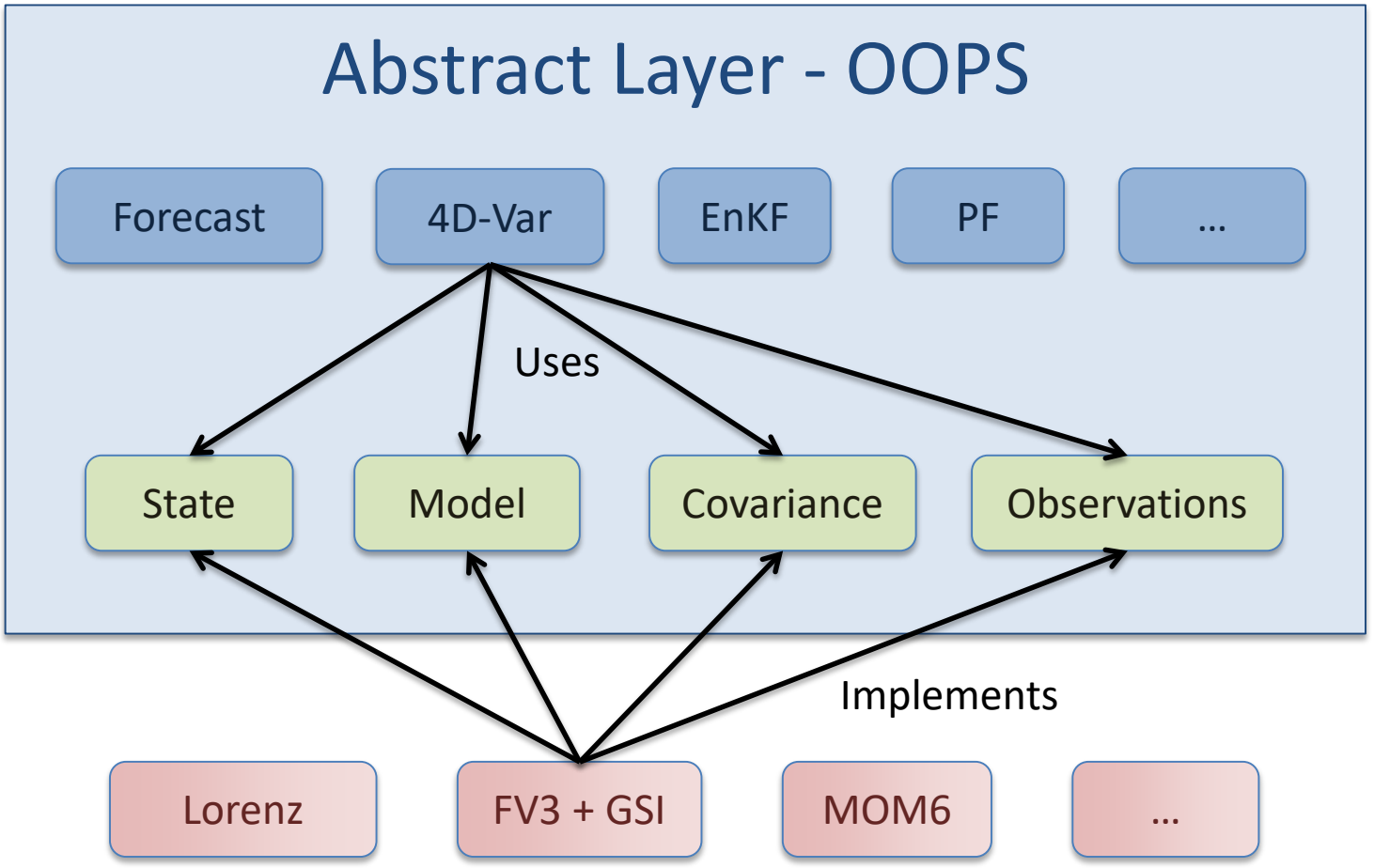
JEDI: Motivations and Objectives

The Joint Effort for Data assimilation Integration (JEDI) is a collaborative development between JCSDA partners.

Develop a unified data assimilation system:

- From toy models to Earth system coupled models
- Unified observation (forward) operators (UFO)
- For research and operations (including R2O/O2R)
- Share as much as possible without imposing one approach

Abstract Design: separation of coAncerns



Abstract interfaces are the most important aspect of the design

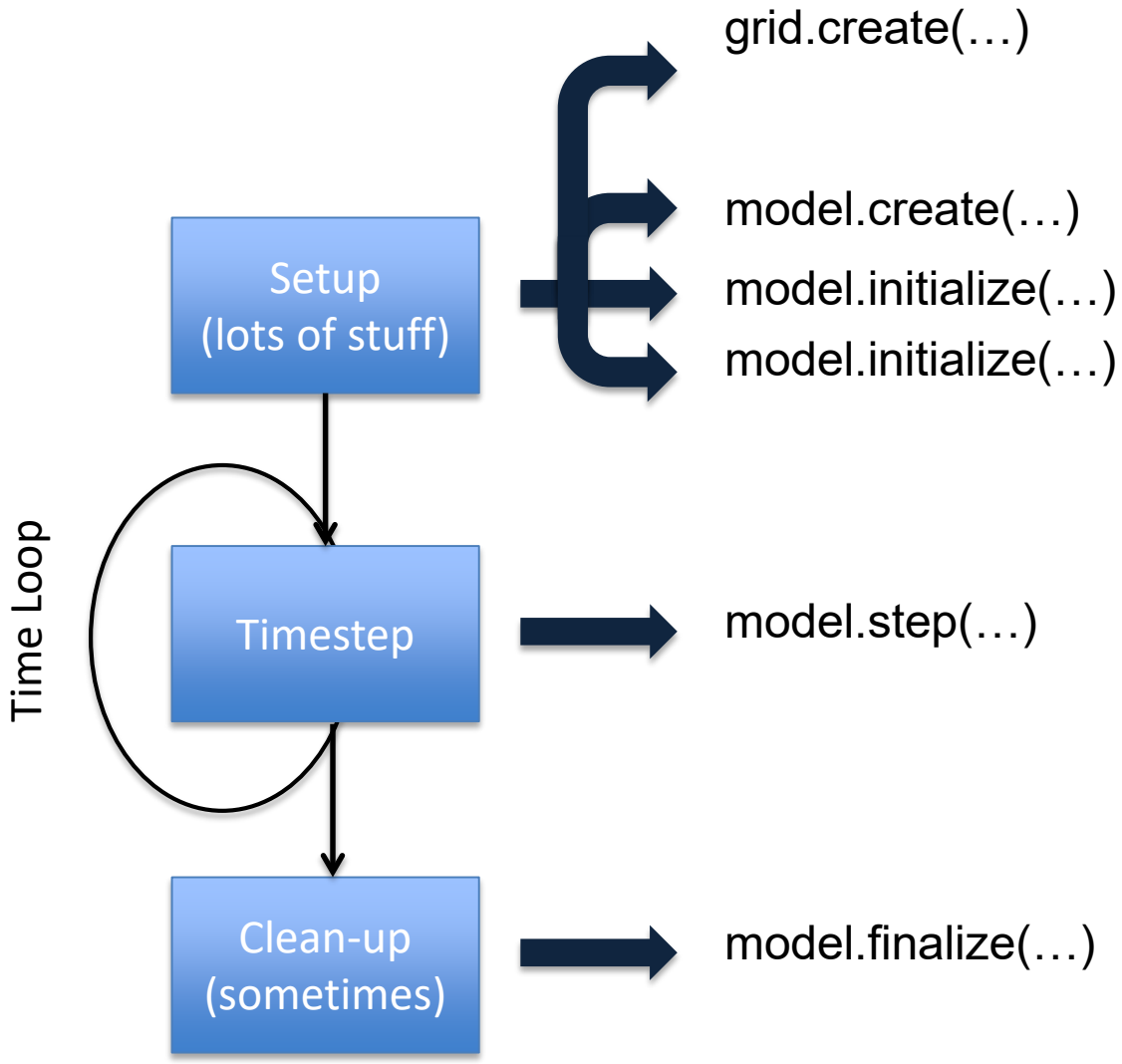
Joint Effort for Data assimilation Integration

Model Space Interfaces

Observation Space Interfaces

Infrastructure and working practices

Model Design: $x_t=M(x_0)$



Model Design: Post-processors

- Between model “steps” OOPS calls post-processors
  - OOPS manages when post-processing should be called
  - Post-processing moved away from model code (**separation of concerns**)
  - Adding a post-processor is just adding it to a list

- A post-processor can be anything that
  - Is called (regularly) during model integration
  - Does not modify the State

- OOPS relies on post-processors for isolating data assimilation from the model (**separation of concerns**)
  - Computing simulated observations H(x)
  - Jc-DFI, …

Joint Effort for Data assimilation Integration

Model Space Interfaces

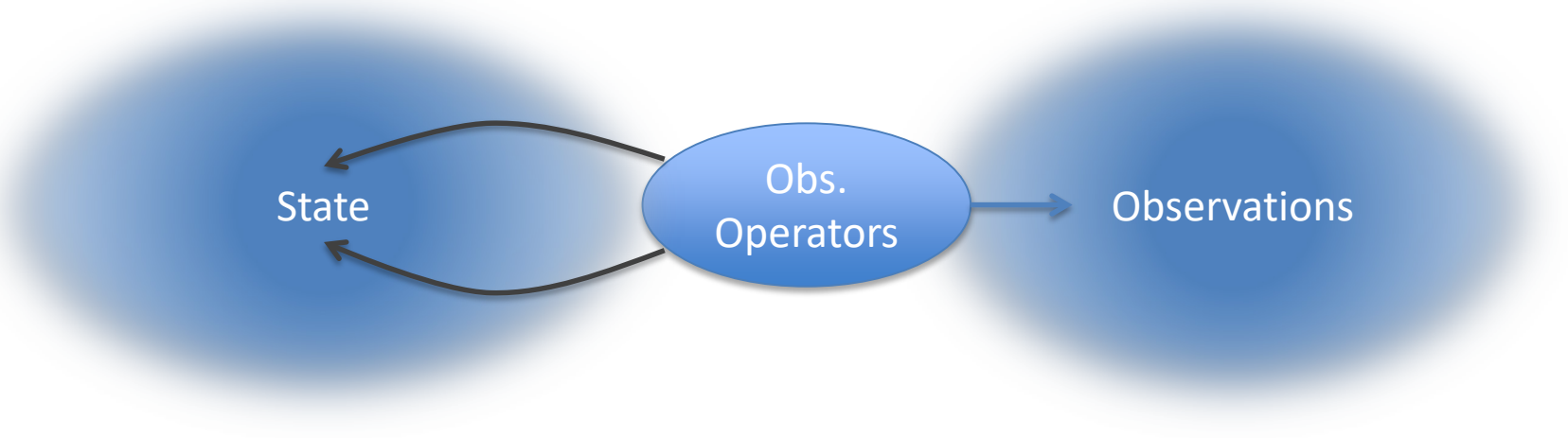
Observation Space Interfaces

Infrastructure and working practices

Observation Space Objectives

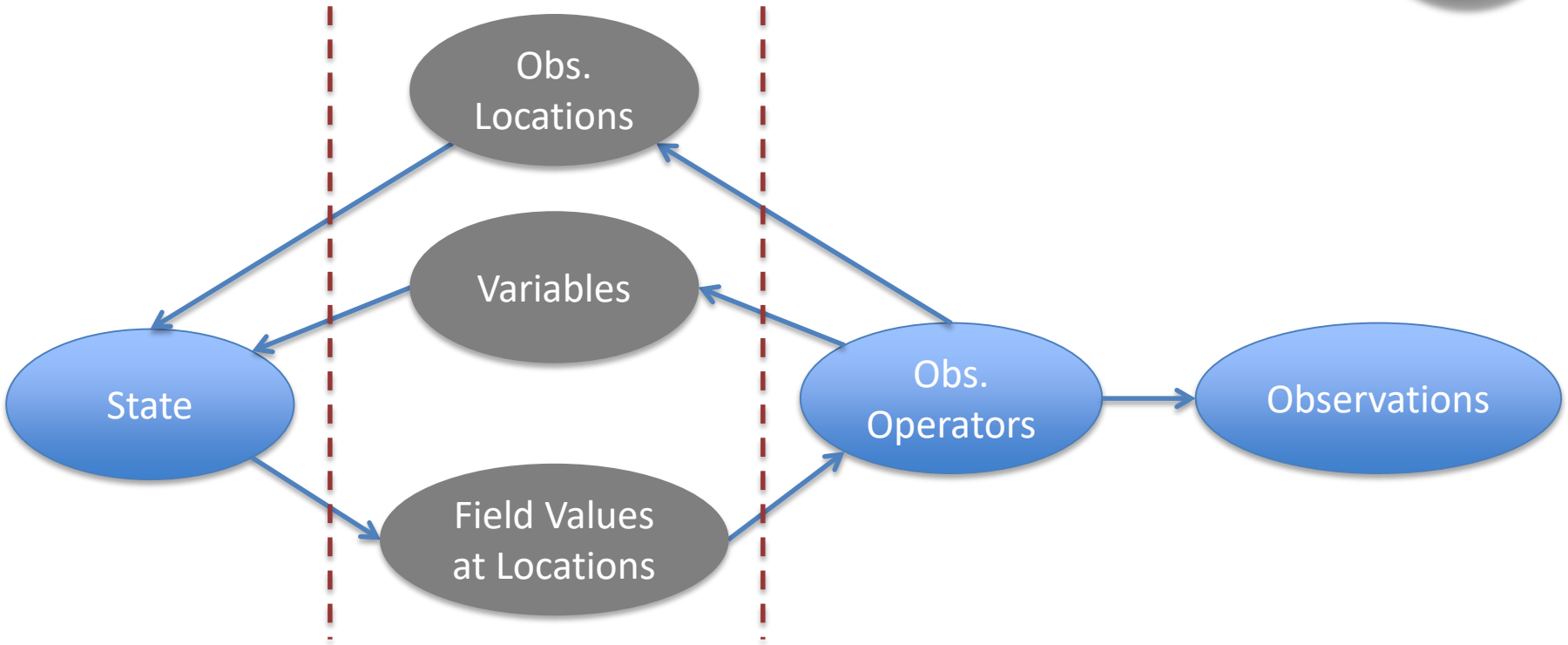
- Share observation operators between JCSDA partners and reduce duplication
  - Include instruments science teams
- Faster use of new observing platforms
  - Regular satellite missions are expensive
  - Cube-sat have short expected life time
- Unified Forward Operator (UFO)
  - Build a community app-store of observation operators ("op-store")

Observation Operators



- In most existing systems, observation operators directly access state/model data
- Observation operators, and as a result DA systems, are very model specific

UFO: the interface advantage



- JEDI/UFO introduces standard interfaces between the model and observation worlds
- Observation operators are independent of the model and can easily be shared, exchanged, compared

UFO Observation "filters"

- JEDI/UFO calls abstract "observation filters" before and after the actual observation operator
- Filters can be written once and used with many observation types
- Observation filters are generic and have access to
  - Observation values and metadata
  - Simulated observation value (post-filter)
  - Their own private data
- Examples:
  - Quality control (background check, buddy check, cloud detection…)
  - Thinning
  - Saving linearization trajectory or Jacobians

Interface for Observation Data Access (IODA)

Interface to isolate science code from data storage

Three levels:

- Long term storage (historic database)
- Files on disk (one DA cycle)
- In memory handling of observations (hardware specific?)

Two environments:

- Plotting, analyzing, verifying on workstation
- DA and other HPC applications (MPI, threads, GPUs…)

Goal: one interface, possibly several implementations?

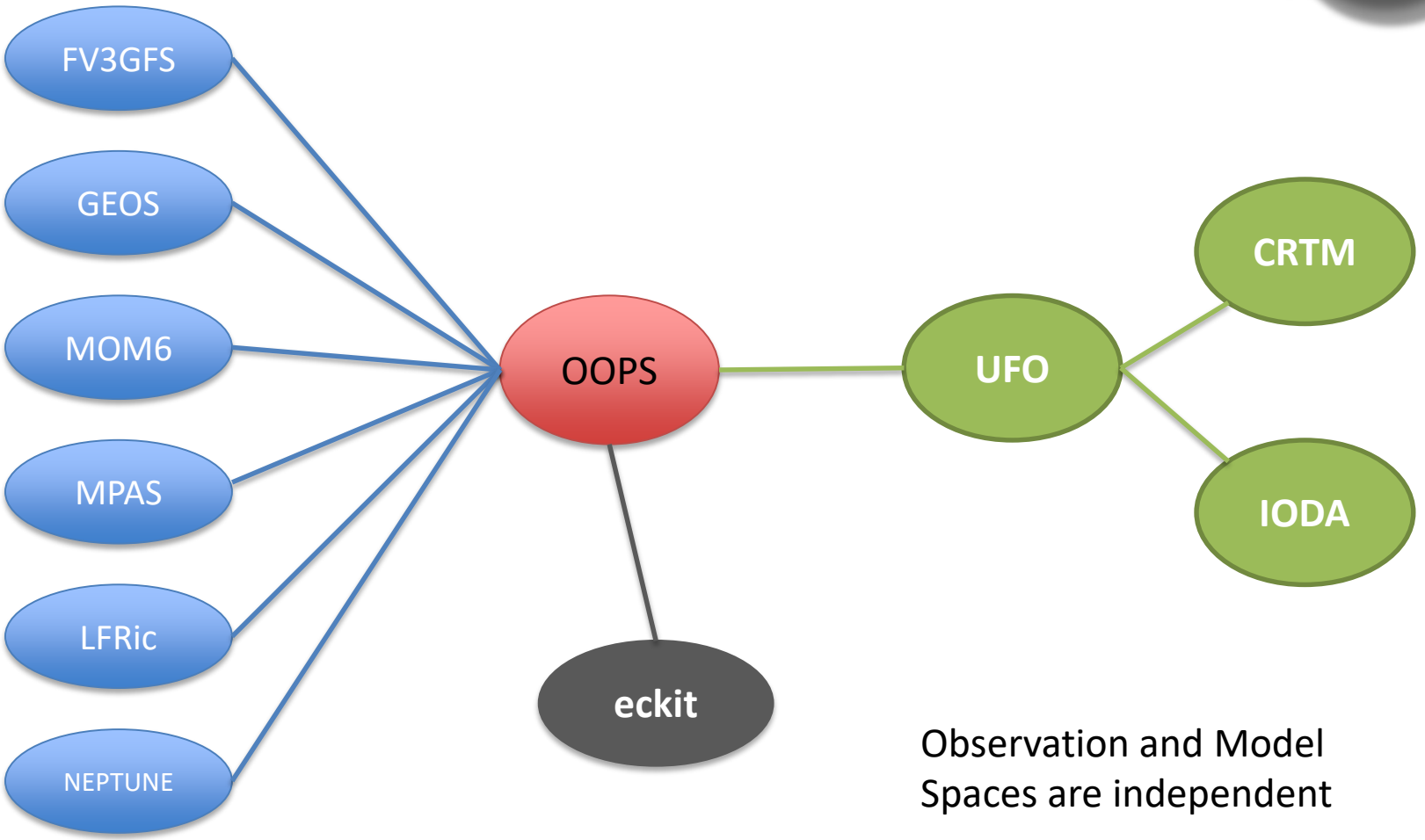
Joint Effort for Data assimilation Integration

Model Space Interfaces

Observation Space Interfaces

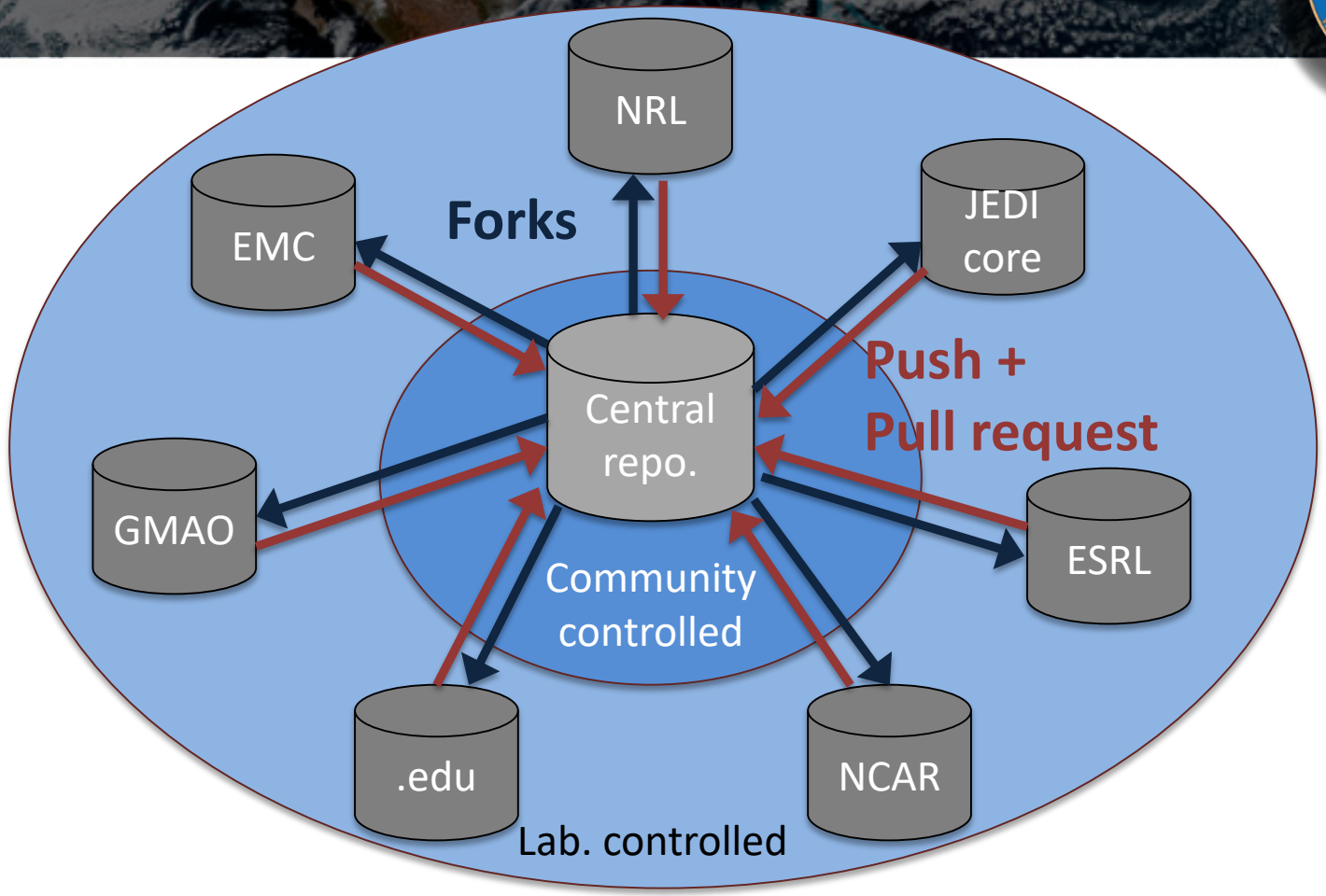
Infrastructure and working practices

Code and repositories



Observation and Model Spaces are independent

Collaborating: Repositories



Permission to fork repository are very easy to obtain

Contributing code is very controlled:

- Pushing a branch requires write permission on central repository
- Pull request triggers code review and approval for merging to higher level branch

Governance and code reviews

Governance is about management keeping in control and deciding what features should be in the system

Code reviews are about quality of the code

Two different levels of control

- Good code can stay outside of central repository (stability of interfaces is important)
- A desired feature that does not satisfy quality requirements cannot be accepted as is

Testing is a pre-requirement for code reviewing

Different people and different pace: **Separation of concerns**…

Infrastructure, working practices

- Project methodology inspired by Agile/SCRUM
  - Adapted to distributed teams and part time members
- Collaborative environment
  - Easy access to up-to-date source code (github)
  - Easy exchange of information (Confluence, zenhub)
- Flexible build system (ecbuild-based)
- Coding norms
- Documentation, tutorials, JEDI Academy

Infrastructure, working practices

- Continuous Integration, Testing framework
  - Toolbox for writing tests
  - Automated running of tests (on push to repositories)
- Effort on portability
  - Automatically run tests with several compilers
  - JEDI available in containers (Docker, singularity)
- Enforce software quality (correctness, coding norms, efficiency)
- Change in working practices take time…

Code Sprints

- Gather 8-10 people in a room for 2 weeks
  - NICAS B Matrix (Aug 2017)
  - Observation Operators (Nov. 2017)
- Efficient use of time, especially for part time contributors
- Involve people from all partner institutions in project
- Very motivating (before, during, after)

Summary

- JEDI is critical to next-generation DA development (hence to NGGPS)
- Provides scalability/reusability to support multiple applications, users, and contributors
- Builds off successful example (Object Oriented Prediction System – OOPS)
- Coding began in August, 2017
  - Successful components already exist
    - See SOCA example in JCSDA Round-Up Presentation

DA Science Grand Challenges

Observations

- Big Data paradigm (volume, variety, velocity): most of total error reduction comes from a large number of observations with **small or moderate individual impacts**
- **Diversity of Satellite Observation Sources (commercial, international, etc.)**

Models

- Better value for society: forecast model for more components of Earth system (Ocean, Waves, Cryosphere, Land, Hydrology, Aerosols, Atmospheric composition, Ionosphere, etc.)
- Models are getting coupled to better account for interactions

Data Assimilation Algorithms

- DA systems becoming increasingly complex as science progresses: comparing algorithms almost impossible. Optimum may be application/machine dependent

Workforce

- Individuals need years to attain interdisciplinary skills to contribute; harnessing the talent of academic, private sectors; substantial fraction of pool non-US

Closing Remarks

JCSDA improving its operations

- AOP improving up-front coordination and accountability
- Targeting inter-dependent activities with clear added value
- Project-based structure focusing on measureable deliverables
- JCSDA staff committed to collaboration
- **Enhancing satellite DA to support the OCs**

Upcoming Events

- 16th Annual JSCDA Science and Technical Workshop – 5/30 – 6/1 in Boulder, CO
- JCSDA Summer Colloquium - Bozeman, MT July 22 – Aug 3