The NCI National Environmental Research Data Interoperability Platform (NERDIP) to support High Performance access to Oceans and Marine-related Interdisciplinary Research

NCI — NATIONAL COMPUTATIONAL INFRASTRUCTURE

Ben Evans[1,*], Tim Pugh[2], Edward King[3,4], Jonathan Hodge[3], Lesley Wyborn[1].

[1] National Computational Infrastructure (NCI), Australian National University, Canberra, Australia
[2] Bureau of Meteorology, Melbourne, Australia
[3] CSIRO Marine and Atmospheric Research, Australia
[4] Integrated Marine Observing System, Hobart, Tasmania
*Contact: Ben.Evans@anu.edu.au

As data volumes from ocean observation and modelling activities continue to increase exponentially, access to and the analysis of, long-term data, archives becomes increasingly challenging. The oceans community is not alone: all members of the Earth Systems and Environmental communities are facing the same challenge. We need a solution that will enable the Oceans and Marine community to not only manage its own growing data assets, but at the same time, facilitate seamless integration of oceans data sets with data from other communities (e.g., climate, atmospheric, near shore terrestrial and bioinformatics) to empower the next generation of large scale, high resolution, Data-intensive interdisciplinary research.

To progress towards this goal, the National Computational Infrastructure (NCI) at the Australian National University (ANU) has organised a priority set of more than 10 PBytes of national environmental, earth systems science and other data assets (Table 1, Figure 1) on a High Performance Data (HPD) Node within a High Performance Computing (HPC) facility. The node was developed under the Research Data Storage Infrastructure (RDSI) program, a component of the Australian National Collaborative Research Infrastructure Strategy (NCRIS).

Data Collections	Approx. Capacity
CMIP5, CORDEX	2.6 Pbytes
ACCESS products	2.6 Pbytes
LANDSAT, MODIS, VIIRS, AVHRR, INSAR, MERIS	2.0 Pbytes
Digital Elevation, Bathymetry, Onshore Geophysics	350 Tbytes
Seasonal Climate	600 Tbytes
Bureau of Meteorology Observations	400 Tbytes
Bureau of Meteorology Ocean-Marine	380 Tbytes
Terrestrial Ecosystem	400 Tbytes
Reanalysis products	160 Tbytes

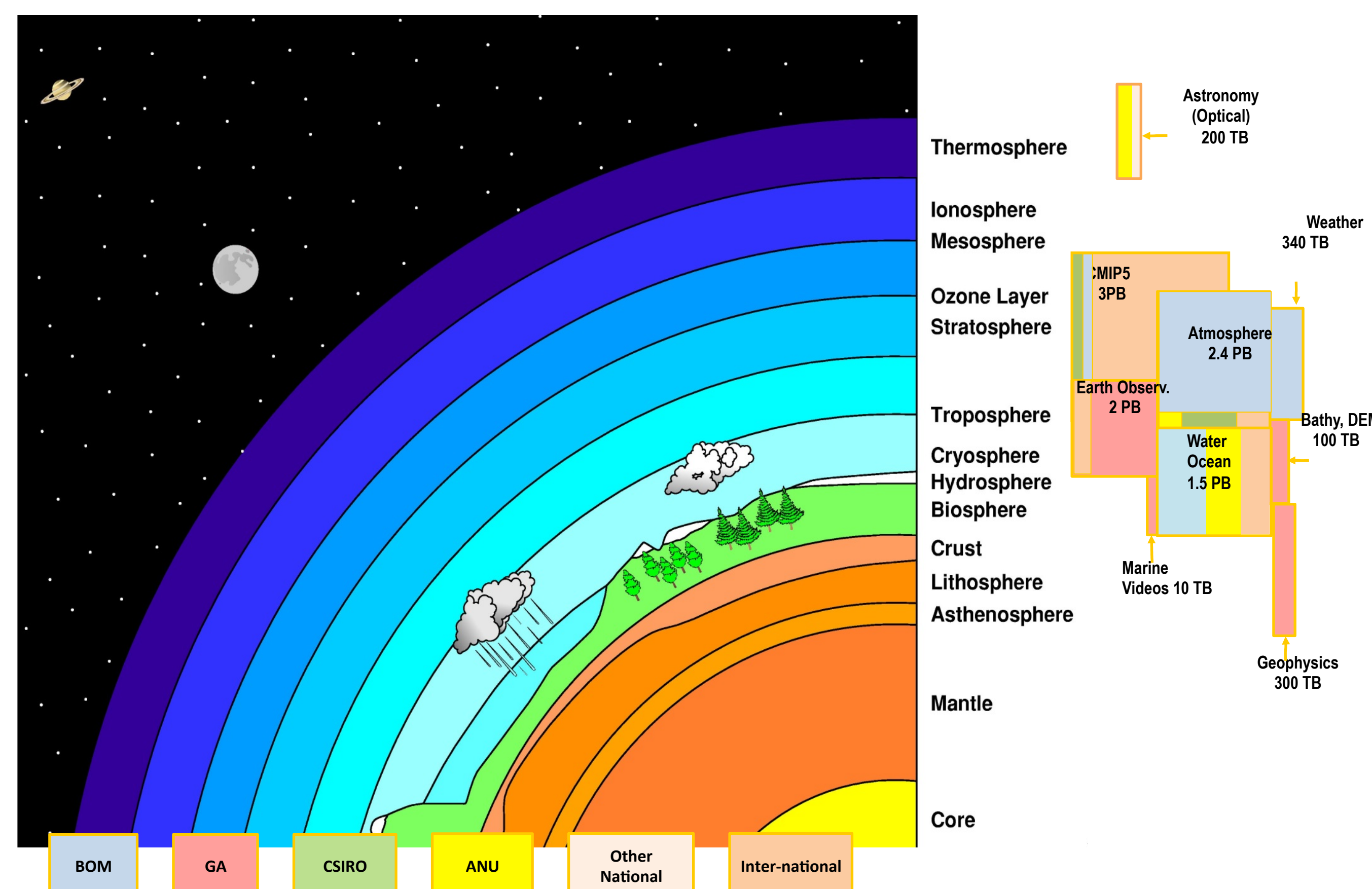Table 1: Major National Environmental Research Data Collections at NCI

Figure 1: The 10 Pbytes+ of National Data Collections are from Government Science Agencies, Universities and other sources. They are categorised into:

1. Earth system sciences, climate and weather model data assets;
2. Earth and marine observations and products;
3. Geosciences;
4. Terrestrial ecosystems;
5. Water management and hydrology; and
6. Astronomy, social science and biosciences.

The co-location of these large volume collections with a high performance infrastructure is designed to support Data-Intensive Science, whereby HPC analytics can be directly undertaken across all the data content for interdisciplinary analysis. A unified data platform, the NCI National Environmental Research Data Interoperability Platform (NERDIP), is being built to enable the same data to be used for multiple use cases both within, and beyond the Oceans and Marine community. To achieve this, formats need to be self-describing and all attributes need to conform to international standards for vocabularies and ontologies (Figure 2). High Performance access to data is facilitated through OpenDAP, OGC and other services, and fast programmatically-searchable catalogues: direct access to NCI's supercomputer (Raijin) and High Performance Cloud (Tenjin) is also available (Figure 3).
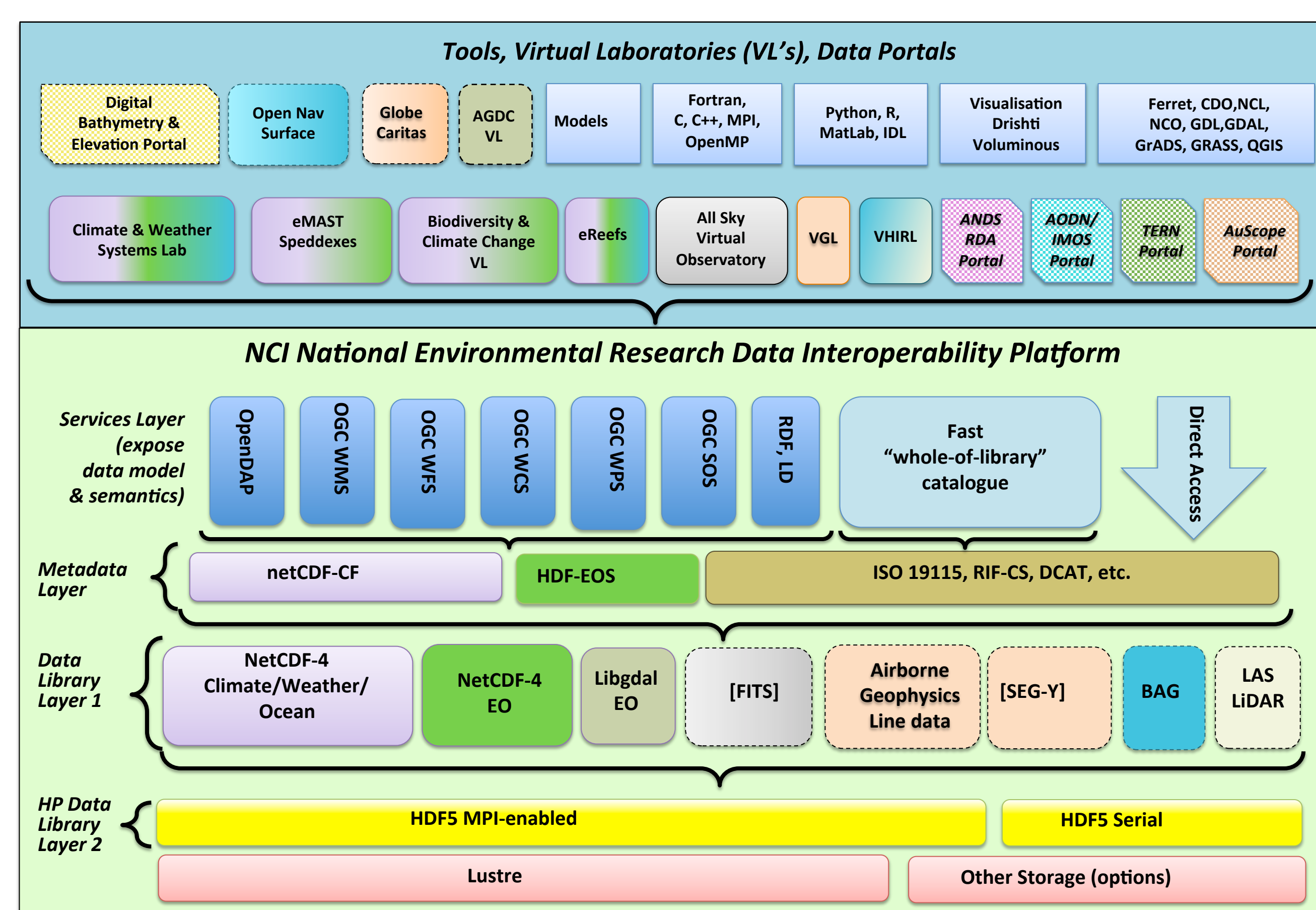

Figure 2: NERDIP is loosely coupled to multiple Tools, VL's and Data Portals (Components in progress or proposed have dashed outlines).
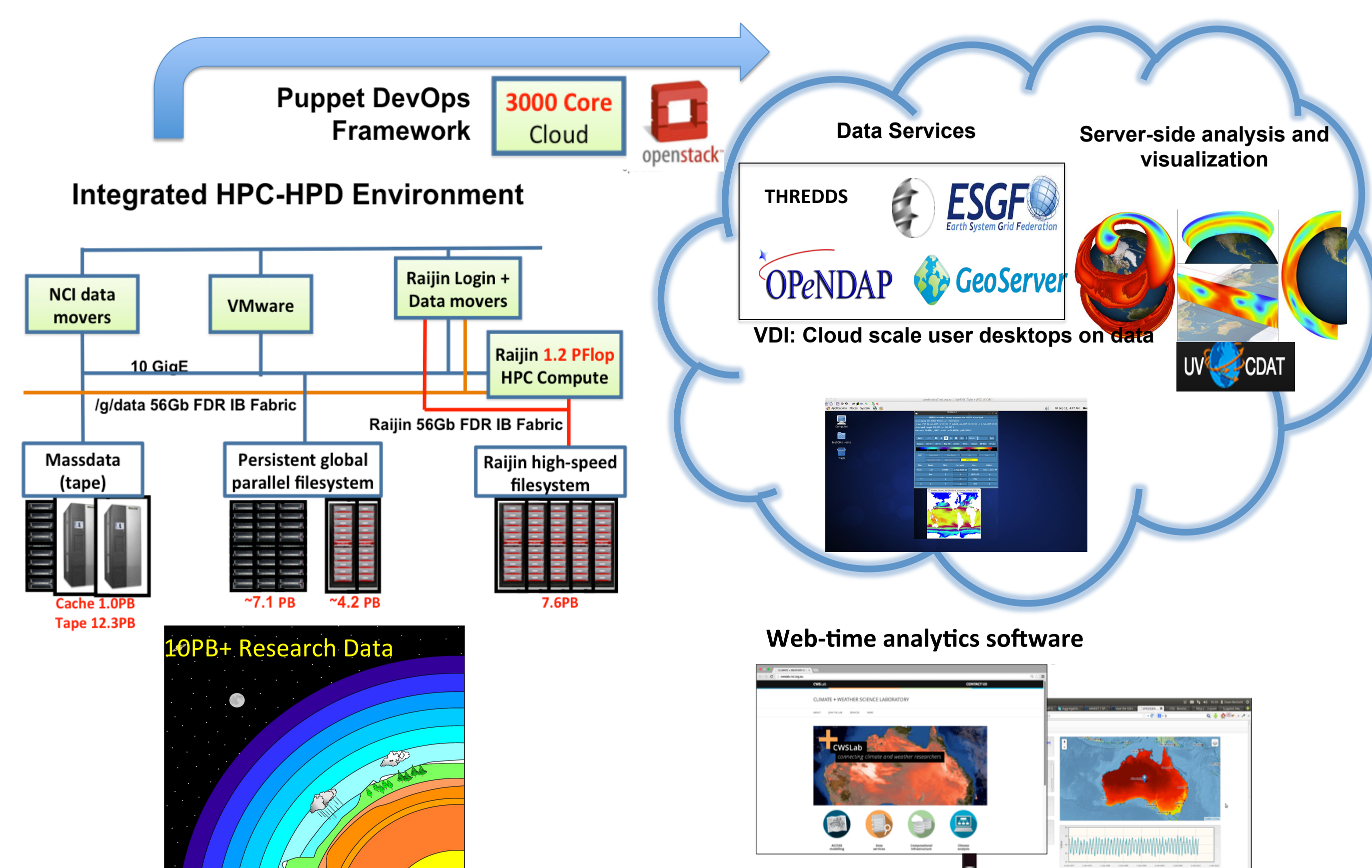

Figure 3: The Integrated High Performance infrastructure at NCI

Properly architected the National Environmental Research Data Interoperability Platform (NERDIP) will lead to:

- A dramatic improvement in the scale, resolution, reach and integration of Australian Oceans and Marine research;
- Seamless high performance access to nationally significant data collections using new capabilities to support cutting-edge research methods; and
- Realisation of synergies with related international research infrastructure programs, particularly those of the Oceans and Marine community.

nci.org.au