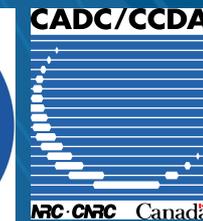




Thirty Years of Open Research Data

David Schade
Canadian Astronomy Data Centre
National Research Council Canada & University of
Victoria

Sept 22, National Data Service Summit, Montreal



A Canadian National Data Service must:

- **Be international in scope**
- **Support the work of professional researchers with adequate metadata and infrastructure**
- **Support the full spectrum of research communities**
- **Be integrated appropriately with data processing and other services that make the data usable for research**



University
of Victoria



University of
British Columbia

canarie



compute + calcul
CANADA



Why invest in Digital Data Management?

The essence of the value proposition:

- *maximize research output*
- produce more research for a given investment



University
of Victoria



University of
British Columbia

canarie



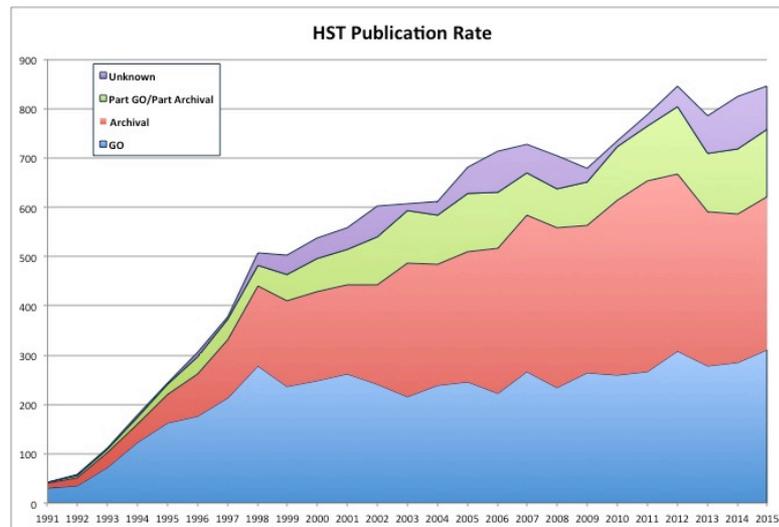
compute + calcul
CANADA



Astronomy digital data management

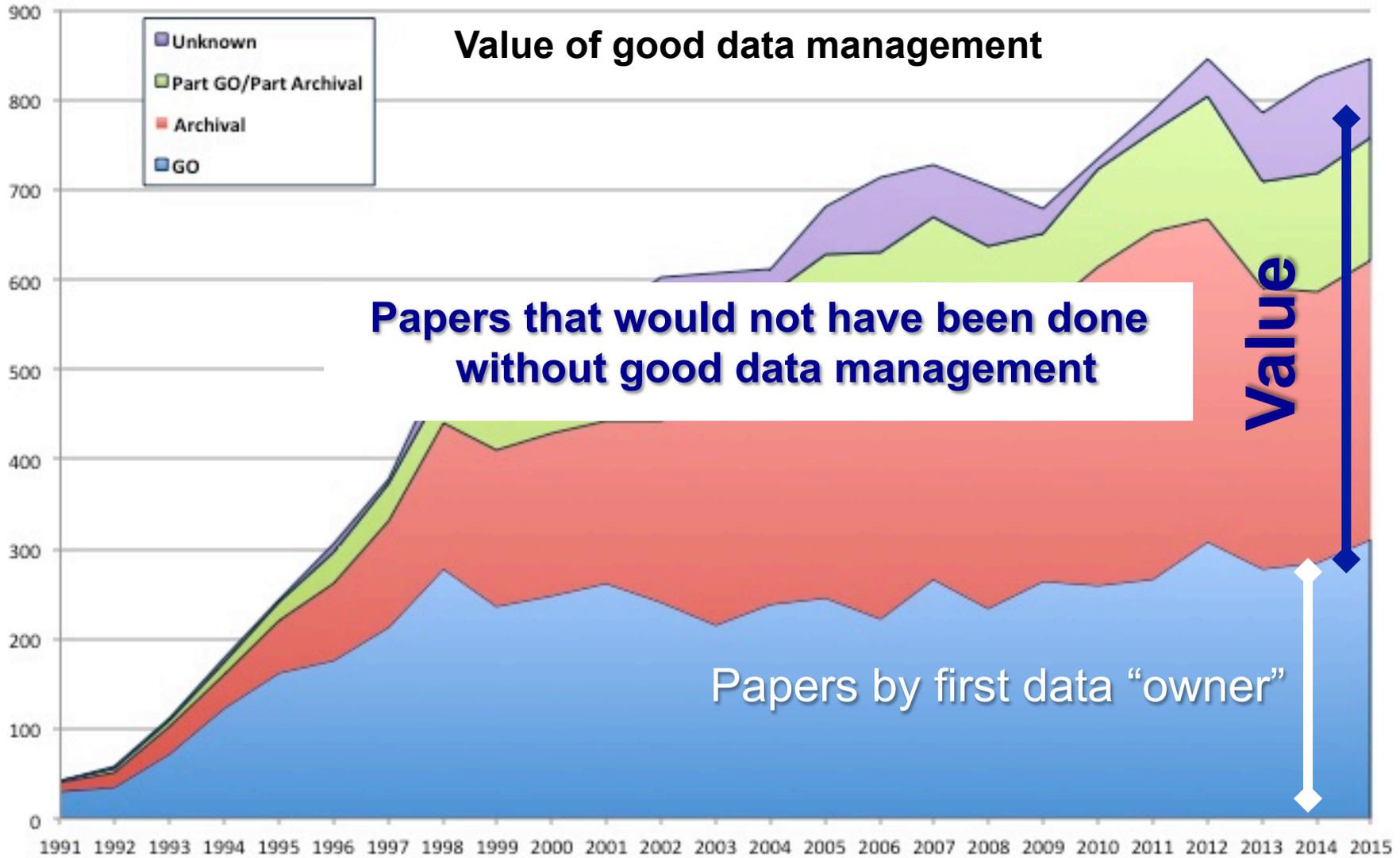
Astronomy is a leader among disciplines in the deployment of advanced data management services

And it shows in astronomy research outcomes



HST Publication Rate

Number of research papers



Value of good data management

Papers that would not have been done without good data management

Value

Papers by first data "owner"

Year of publication

Canadian Astronomy Data Centre

The National Research Council of Canada has operated a National Data Service *for astronomy* since 1986

Since 2010 we have operated a distributed data system using Compute Canada hardware

NRC has a Parliamentary mandate to support observatories and preserve and distribute data

In astronomy the GoC (NRC) has a special and close relationship with the university community



University
of Victoria



University of
British Columbia

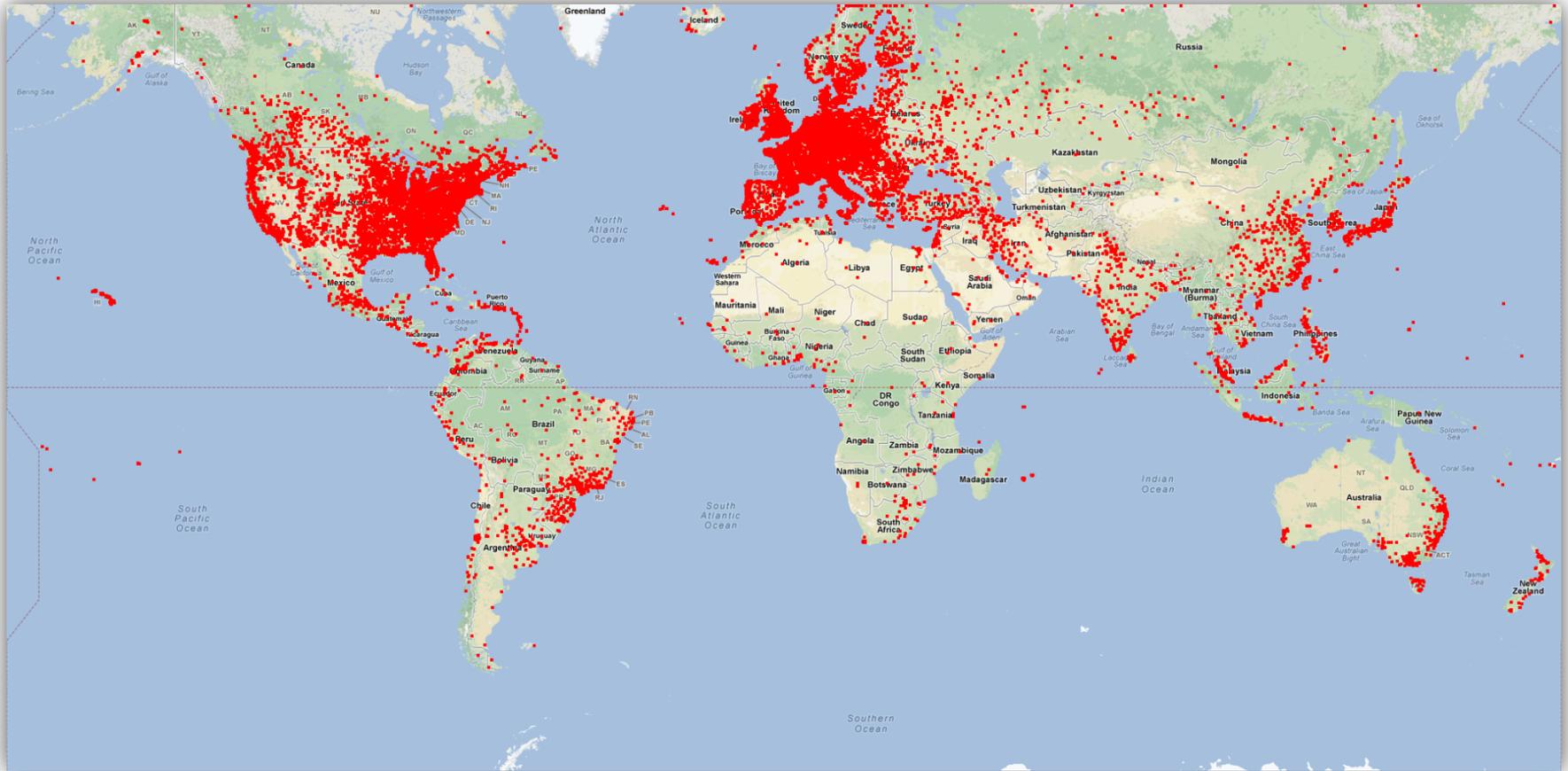
canarie



compute + calcul
CANADA



Delivery of Canadian astronomy data



National Data Service

The International Virtual Observatory Alliance has worked for 16 years to develop standards and protocols that support the federation of national and international data facilities

- National federation of NASA assets
- International federation of US, Canadian, and European, and other data services



University
of Victoria



University of
British Columbia

canarie



compute + calcul
CANADA



Disciplinary Groups doing data management

For example:

Health Data

Genomics

Cancer research

Neuroscience

Earth Observation

Snow and Ice

Agricultural Data

Chemistry

Digital Humanities

Astronomy

There is a wide range in the sophistication of the data management capabilities that exist now for these research fields

A scheme for a National Data Service must support existing capabilities and improve them

It would be wise for a National Data Service to digest the lessons learned from these fields



University of Victoria



University of British Columbia

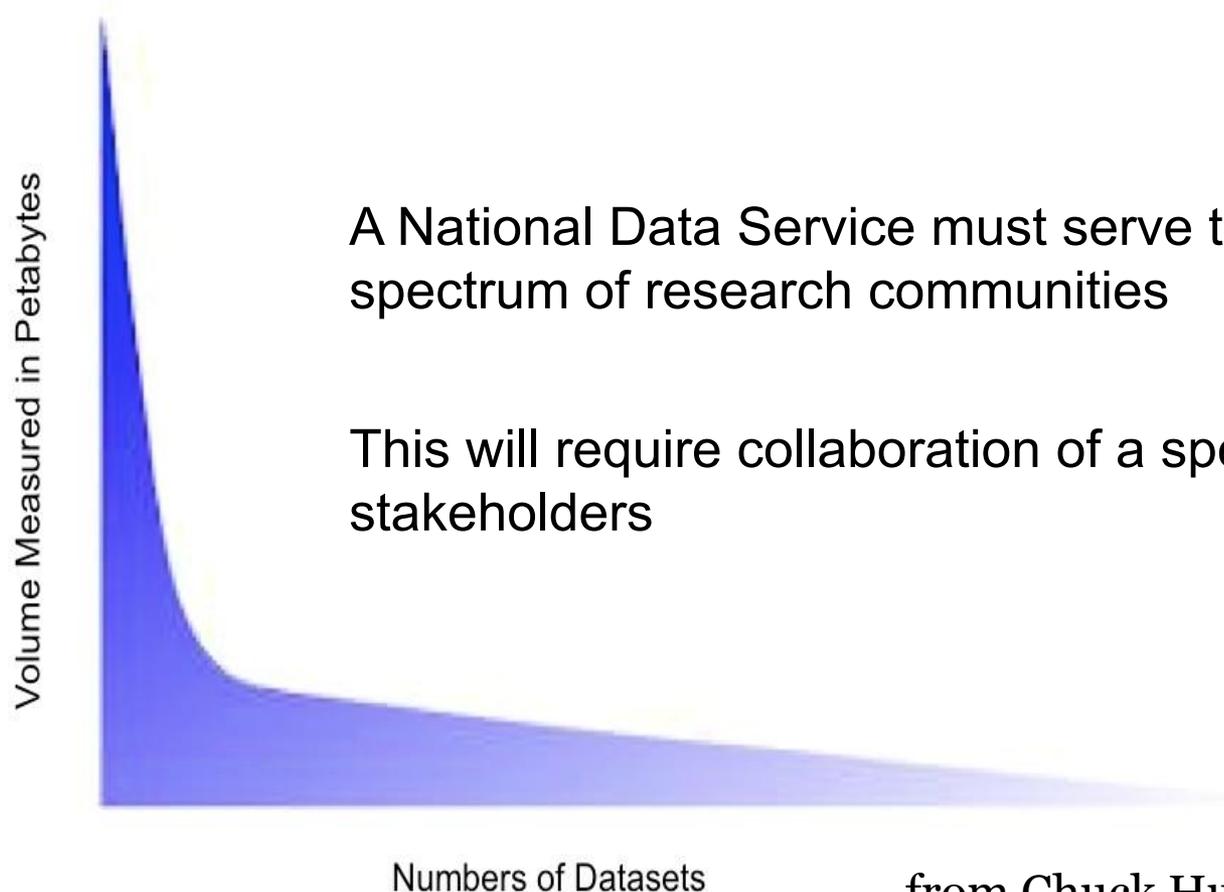
canarie



compute + calcul
CANADA



Long Tail of Research Data



A National Data Service must serve the full spectrum of research communities

This will require collaboration of a spectrum of stakeholders

from Chuck Humphrey

Disciplinary and Long-Tail data management

There is a continuum of data landscapes

The variable that changes *most* across the spectrum is not infrastructure or technology it is:

Metadata



University
of Victoria



University of
British Columbia

canarie



compute+calcul
CANADA



Disciplinary data groups

What can Astronomy contribute ?

Common Data Models

- integration of diverse data into a single metadata representation



University
of Victoria



University of
British Columbia

canarie



compute + calcul
CANADA



Domain Specific Organizations

There is a high level of generality of the outcomes of domain specific data management

- In infrastructure (with CANARIE and Compute Canada)
- In metadata management

Astronomy has taken a number of major steps toward generality by

- IVOA and international standards
- Common Data models for diverse data collections
- Shared infrastructure (Compute Canada)



University
of Victoria



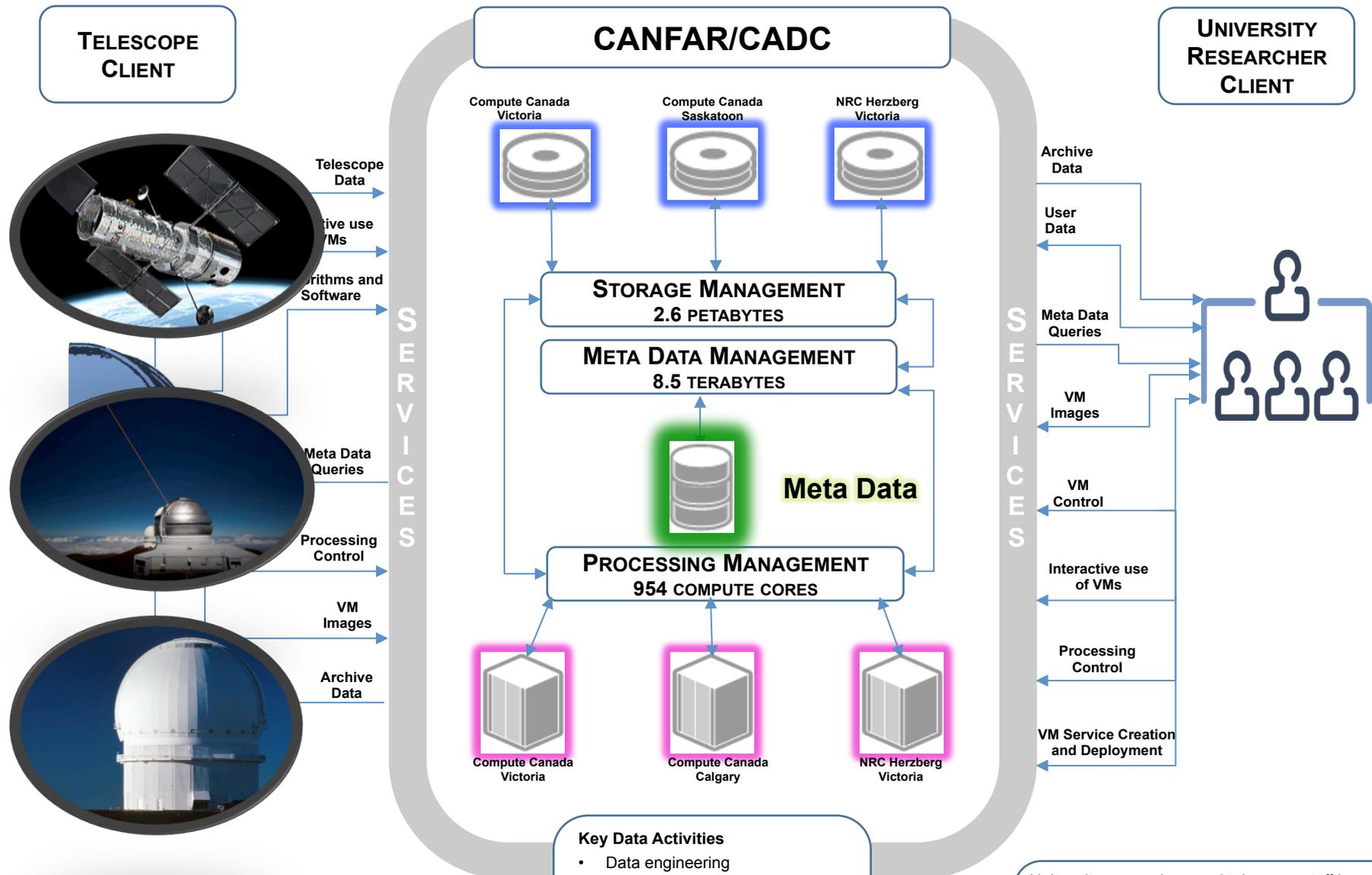
University of
British Columbia

canarie



compute + calcul
CANADA





	Data In		Data Out	
	# of files	Terabytes	# of files	Terabytes
Peak per day	2,169,190	8.0	648,093	16.8
Avg per day	130,952	0.4	99,253	2.6

- Key Data Activities**
- Data engineering
 - Operations and user support
 - Software development
 - Software integration
 - Data processing
 - Data management
 - User web services
 - User web interfaces

University researchers and telescope staff have privileges to upload data, create VMs and install science applications, run interactive VM sessions, submit batch processing jobs to VMs, share their VMs, control the life-cycle for their VMs, offer software-as-a-service applications in their VMs.

Definition: VM – Virtual Machine

A Canadian National Data Service must:

- **Be international in scope**
- **Support the work of professional researchers with adequate metadata and infrastructure**
- **Support the full spectrum of research communities**
- **Be integrated appropriately with data processing and other services that make the data usable for research**



University
of Victoria



University of
British Columbia

canarie



compute + calcul
CANADA

