

Data-Intensive Science in the Data Commons

The International Virtual Observatory experience

David Schade

Canadian Advanced Network for Astronomy Research (CANFAR)

Canadian Astronomy Data Centre

National Research Council Canada & University of Victoria



Subjects of this session

1. Data re-use
2. Data Preservation
3. Continued Value
4. Data integration

The first three items are essentially solved problems for astronomy

Data re-use and continued value : Hubble Space Telescope

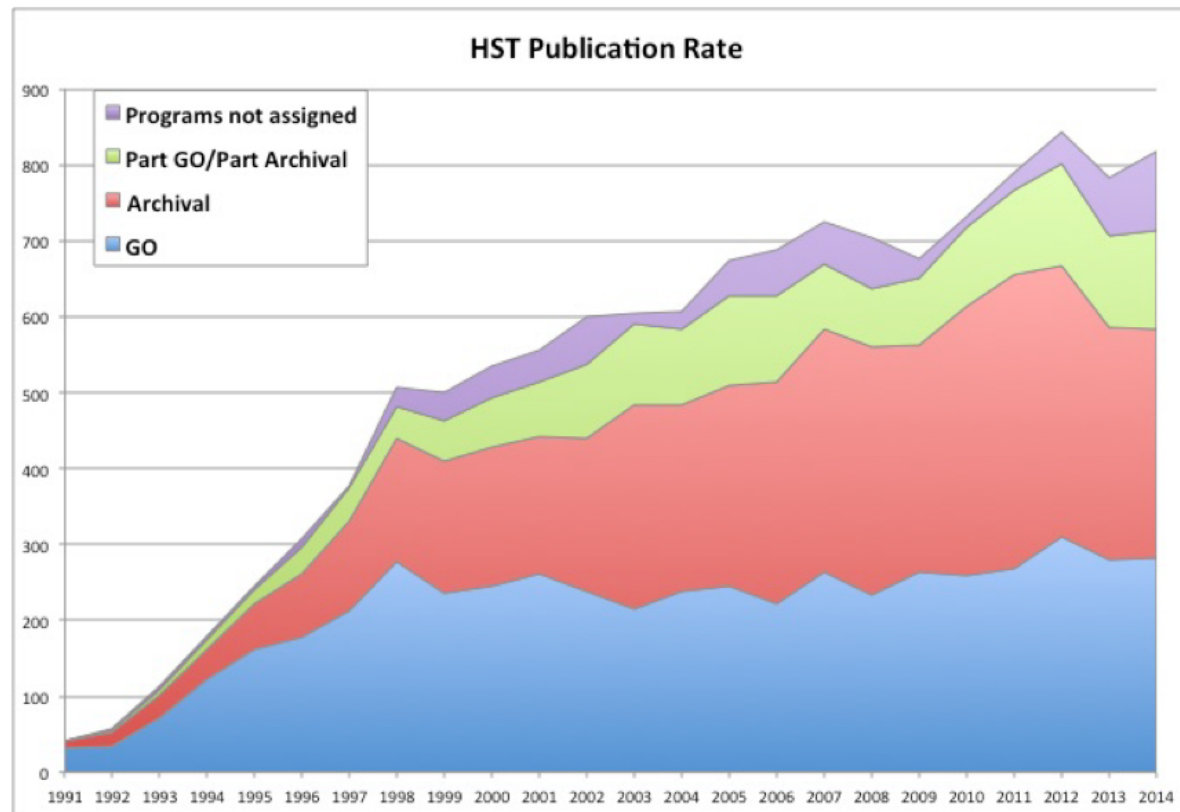


Figure 1: The number of Hubble Space Telescope papers per year to the end of 2014 showing that, since 2002, research based entirely or partly on archival data has produced more than 50% of the annual publications. The pioneering data management techniques developed by STScI, CADC, and ST-ECF deserve a

International Virtual Observatory Alliance (ivoa.net)

The starting point was a large set of centres for astronomy data & service distributed all over the world

- Diverse data collections
- Diverse services
- Diverse communities served

IVOA mission: *“to facilitate the international coordination and collaboration necessary for the development and deployment of the tools, systems and organizational structures necessary to enable the international utilization of astronomical archives as an integrated and interoperating virtual observatory.”*

International Virtual Observatory Alliance (ivoa.net)

InterOperability

IVOA lessons learned

Must define a “core mission” with tangible goals

For IVOA the core mission:

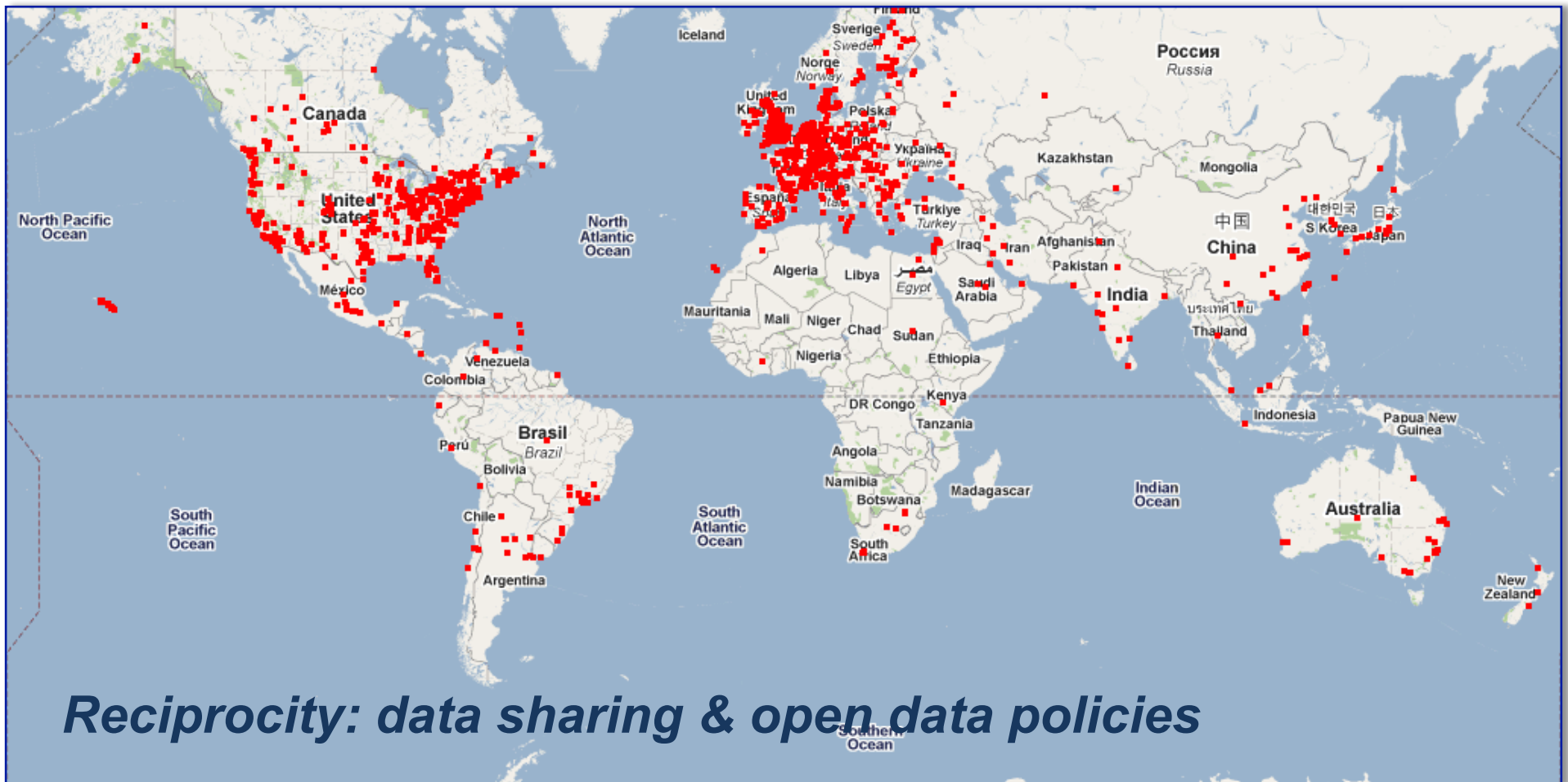
data discovery and access

Data: Canadian Astronomy Data Centre (CADDC) 2014

- Data Scale:
 - 932 Million files
 - 2.3 Petabytes
- Users:
 - Authenticated users: 762
 - Anonymous users: 7,544
 - Registered users: 7,018
- Data moved in the last year:
 - 1,106 Terabytes
 - 91 Million files



Why does Canada fund 100% of a data centre when less than 10% of its users are Canadian ?



InterOperability

IVOA decided early to focus on developing standards

IVOA is an unfunded alliance of 20 national Virtual Observatory Projects

IVOA is not an alliance of:

- data centres
- universities
- scientists

IVOA Structure: Working Groups

Working Group Page	Previous Messages	Subscribe	Send Mail	Chair	Vice Chair
Applications	archive	options	apps@ivoa.net	Pierre Fernique	Tom Donaldson
Data Access Layer	archive	options	dal@ivoa.net	François Bonnarel	Marco Molinaro
Data Model	archive	options	dm@ivoa.net	Mark Cresitello-Dittmar	Laurent Michel
Grid & Web Services	archive	options	grid@ivoa.net	Brian Major	Giuliano Taffoni
Registry	archive	options	registry@ivoa.net	Markus Demleitner	Theresa Dower
Semantics	archive	options	semantics@ivoa.net	Mireille Louys	Alberto Accomazzi

IVOA Structure: Interest Groups

Interest Group Page	Previous Messages	Subscribe	Mailing List	Chair	Vice Chair
Data Curation & Preservation	archives	options	datacp@ivoa.net	Francoise Genova	
Education	archives	options	edu@ivoa.net	Massimo Ramella	Sudhanshu Barway
Knowledge Discovery in Databases	archives	options	kdd@ivoa.net	George Djorgovski	N/A
Operations	archives	options	ops@ivoa.net	Tom McGlynn	Mark Taylor
Theory	archives	options	theory@ivoa.net	Franck Le Petit	Carlos Rodrigo
Time Domain	archives	options	voevent@ivoa.net	John Swinbank	Mike Fitzpatrick

IVOA Structure: Other Groups

Group/Committee Page	Previous Messages	Subscribe	Mailing List	Chair	Vice Chair
Interop	archives	options	interop@ivoa.net	N/A	N/A
Exec			exec@ivoa.net	Christophe Arviset	Enrique Solano
Technical Coordination Group			tcg@ivoa.net	Matthew Graham	Patrick Dowler
Liaison Committee				Masatoshi Ohishi	
Standing Committee on Science Priorities				Mark Allen	N/A
Standing Committee on Standards & Processes	archives	options	stdproc@ivoa.net	Francoise Genova	N/A
IVOA Document Coordinator			ivoadoc@ivoa.net	Giulia lafrate	N/A

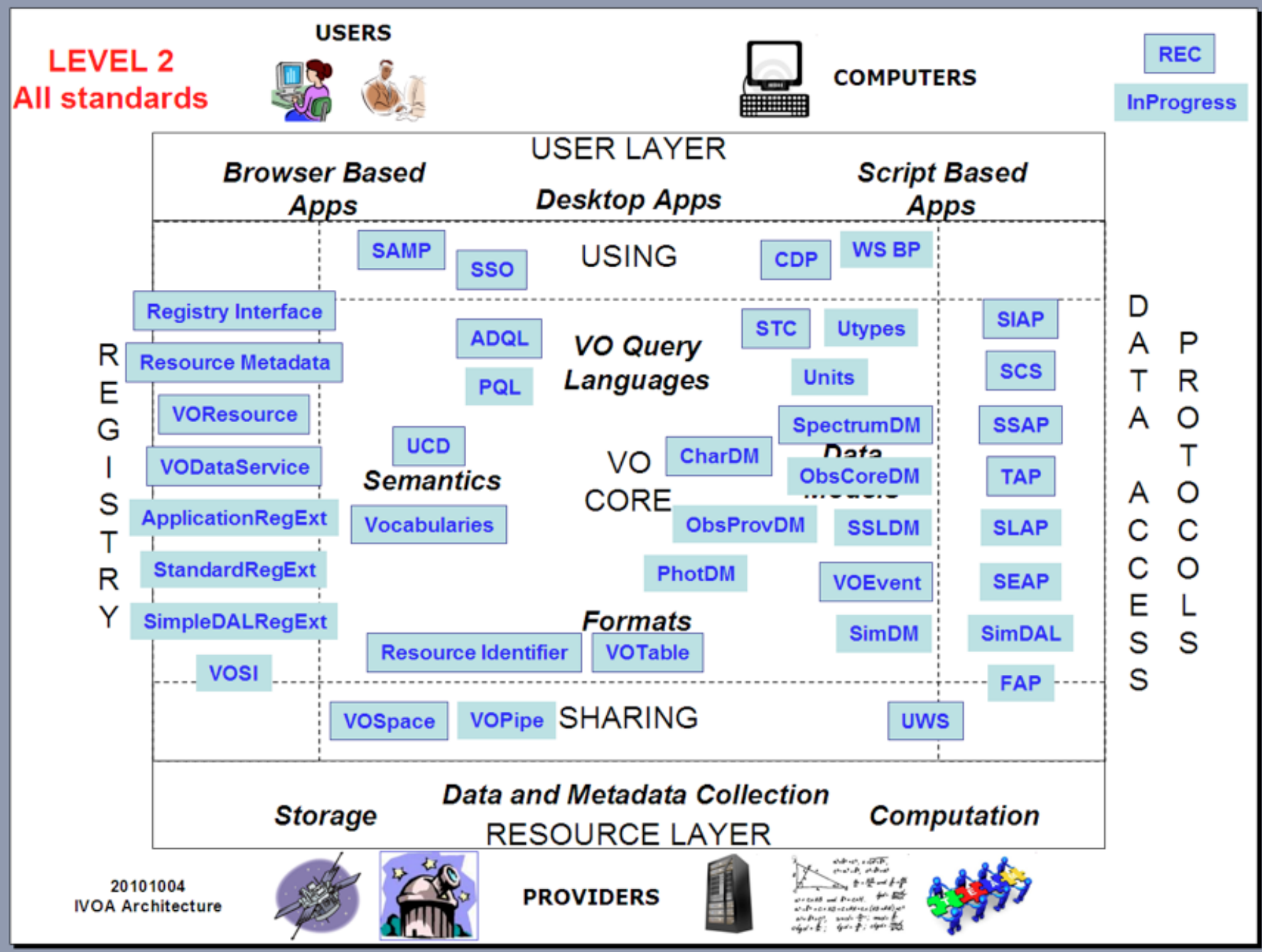


Figure 3: IVOA Architecture Level 2

IVOA Standards cover a wide scope of issues

From:

Data Models (standard metadata and descriptors)

To:

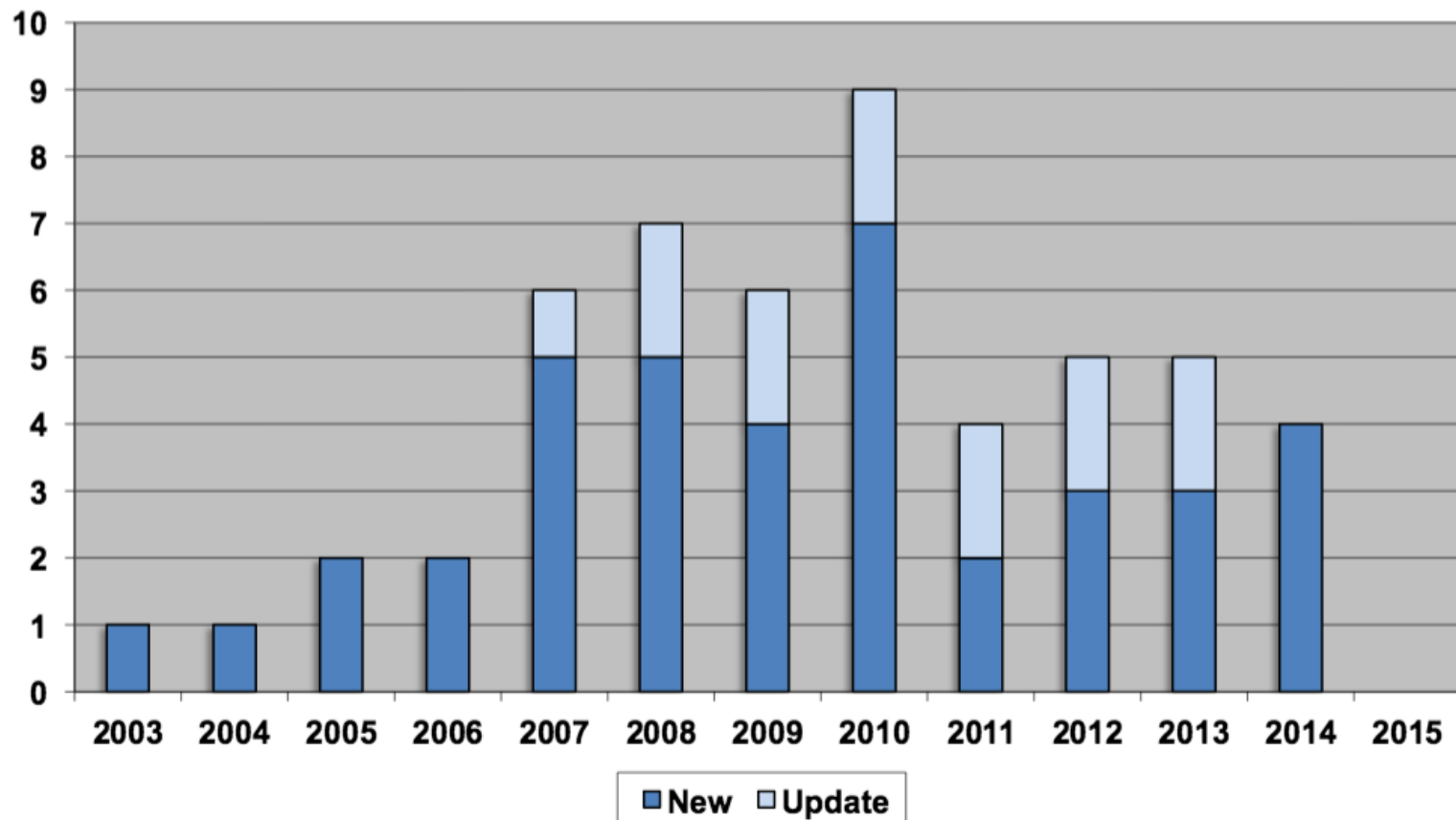
low level job control infrastructure

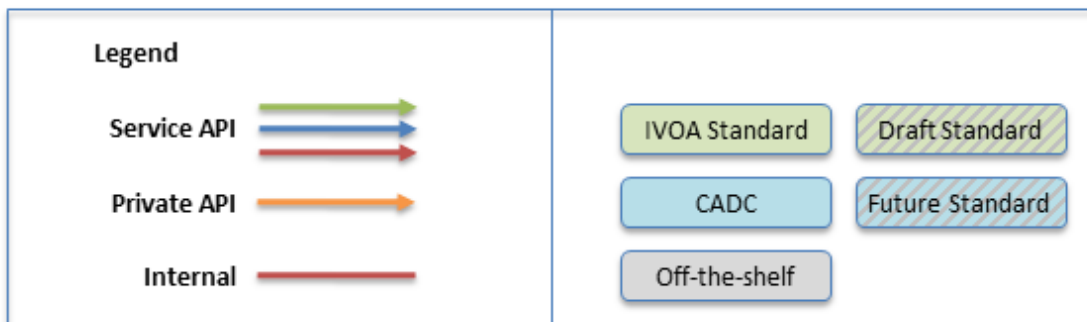
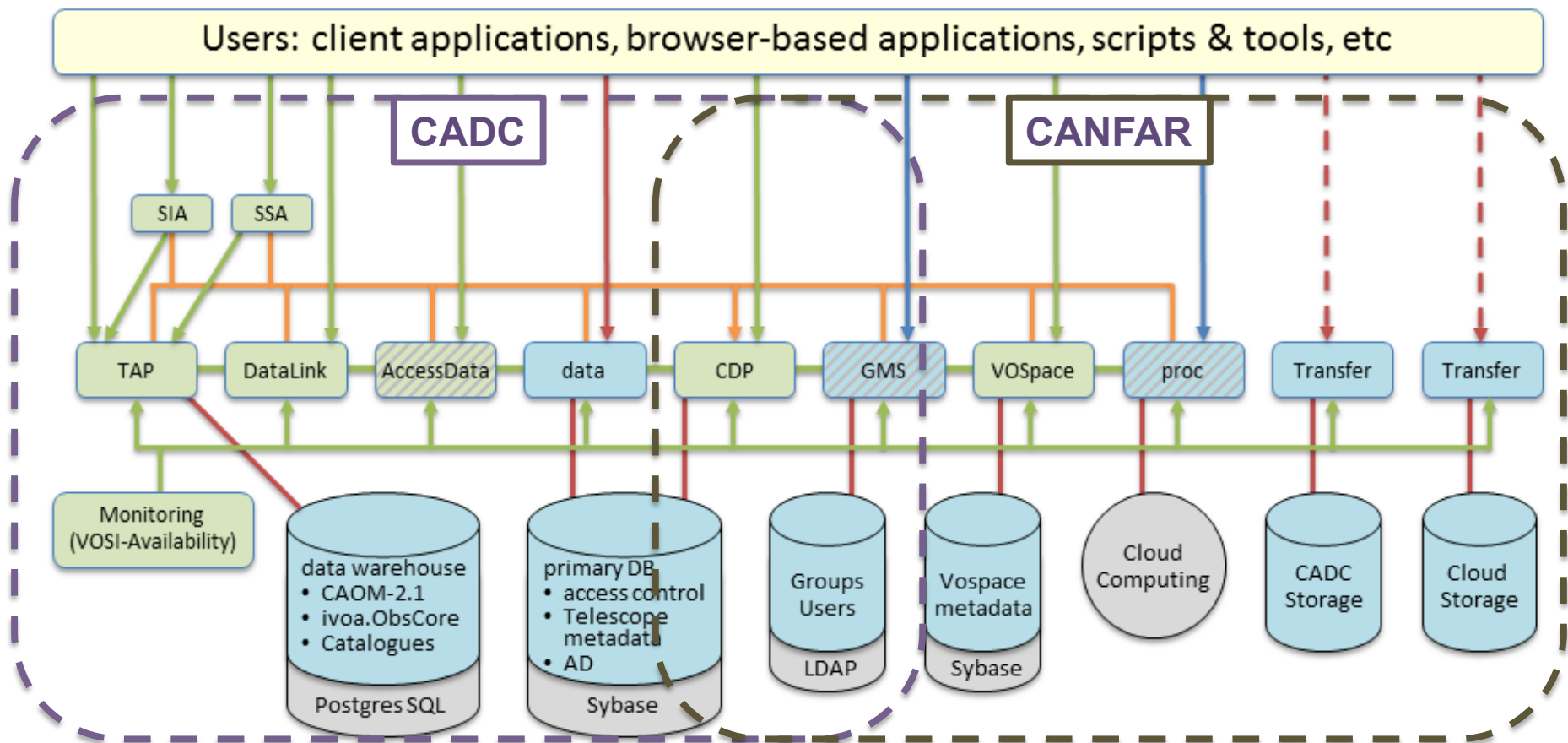


IVOA Standards per year

Revised with help from Françoise Genova and Christophe Arviset

IVOA Standards Recommended per Year





IVOA lessons learned

IVOA defined very good governance and management structure

IVOA developed very good process for working on and approving standards formally

IVOA developed an architecture that showed how its standards were linked together

Gap between standards and implementation

Writing standards is not enough:

Standards must be implemented on major data collections

There was a gap in our thinking that left a gap between the standards and their implementation

This problem is being addressed successfully

Bottom line

- IVOA has produced a valuable degree of interoperability
- It took 13 years to do this
- It took 5 years to get good at this
- Now that we know how to do it:
 - We could do it much better the second time
- The experience we have gained is of greater value than the standards that we produce
- The experience that we gained could be of great value to others

IVOA lessons learned

Nearly all IVOA standards have nothing to do with astronomy

They are domain-agnostic and can be re-used by other domains

They are available for re-use

IVOA lessons learned

We should work very hard to share the lessons of IVOA with those working in other domains and in cross-domain data sharing