CSP Closing Remarks

Bruno Merín
IVOA Committee on Science Priorities (CSP)
http://wiki.ivoa.net/twiki/bin/view/IVOA/IvoaSciencePriorities

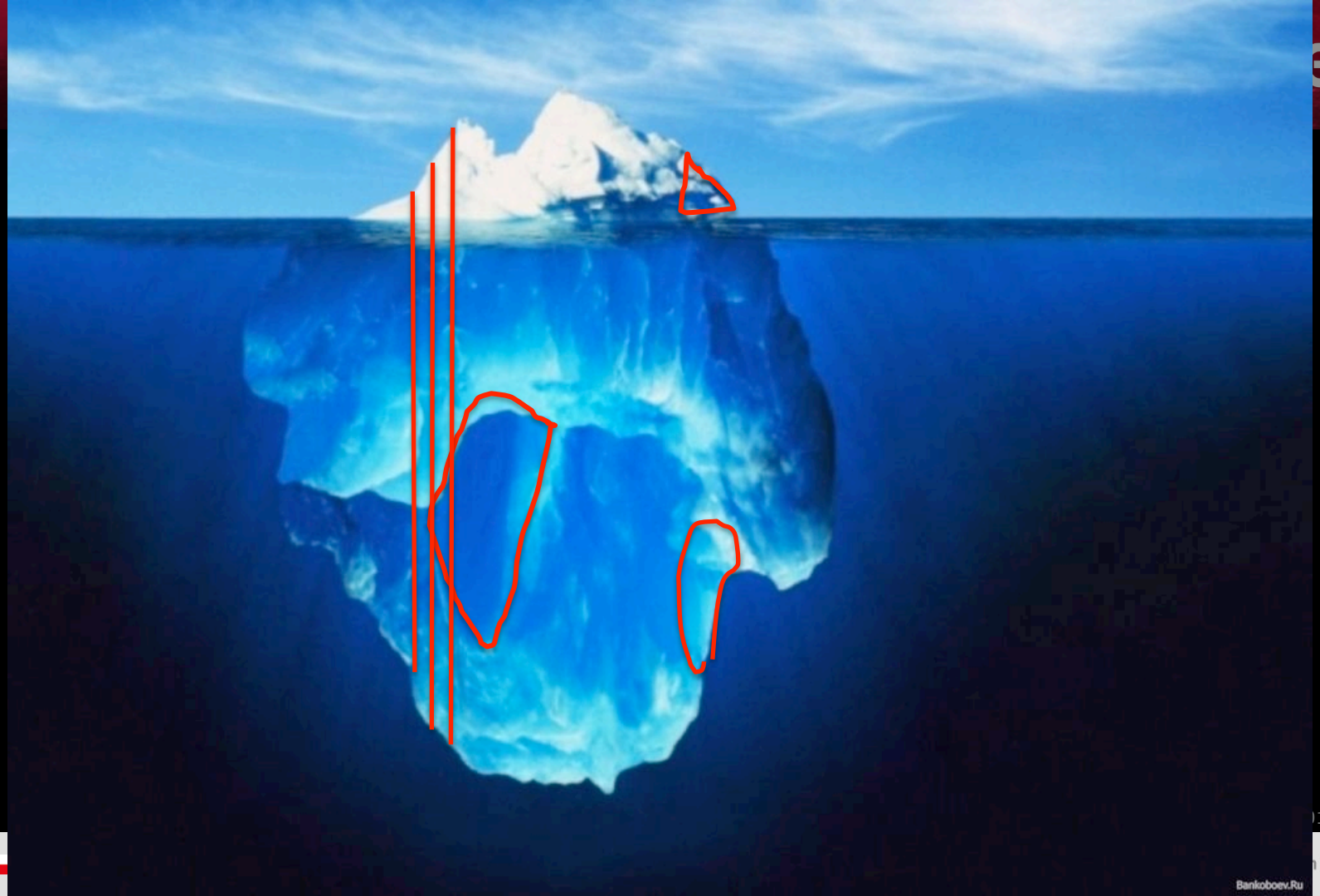
ESAC Science Data Centre (ESA), Madrid, Spain

IVOA Paris Interop, 17/05/2019

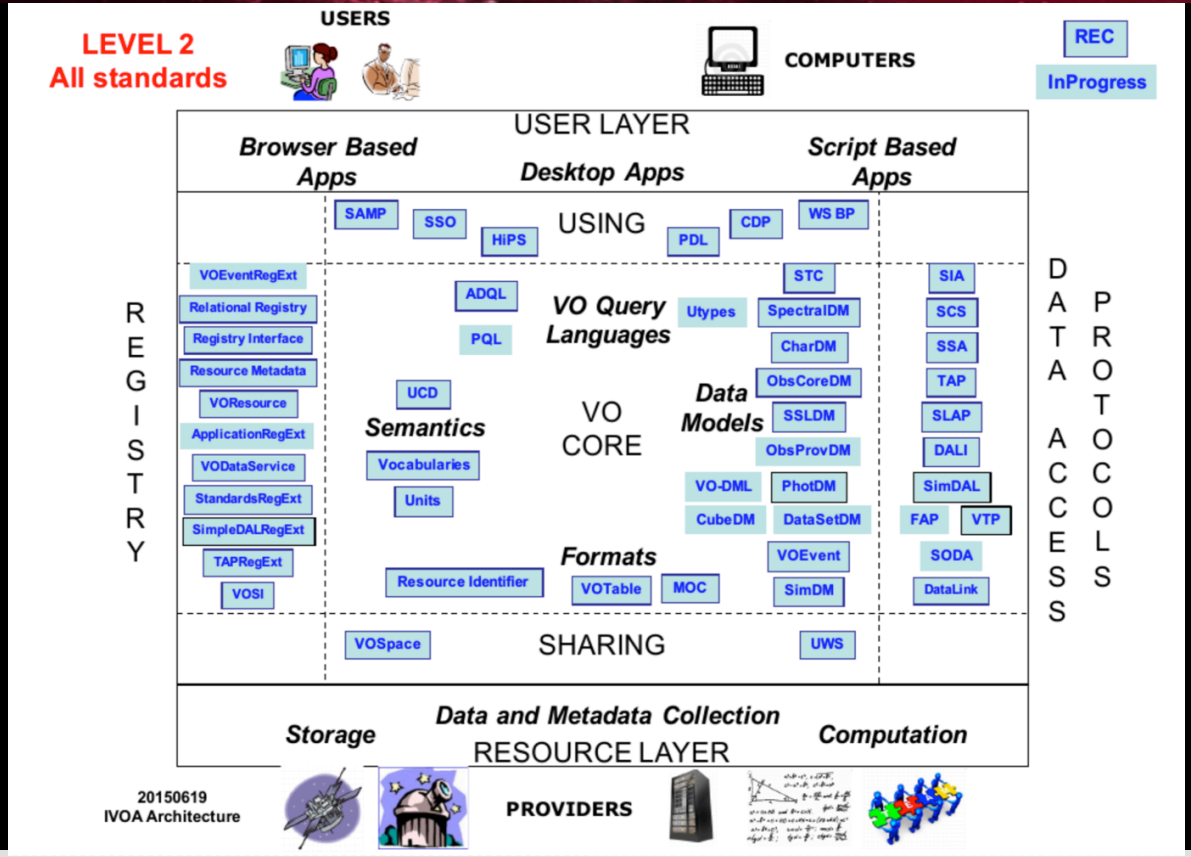
Outline

1. Motivation
2. Scientific priorities
    1. Currently identified
    2. Upcoming
3. Final recommendations

We do this because is has a lot of meaning !!


Bankoboev.Ru

What does the IVOA provide?



Technology

P. Dowler
TCG Report

Current scientific priorities at IVOA

- **Time-domain astronomy**: light-curves -> see new proposed TIMESYS and T-MOC

- **Multi-dimensional data**: spectral or time cubes (sky + wavelength/frequency or sky + time)

  - **Python reference implementations** prioritized for major services?
  - Ways for accessing **large amounts of data** from future surveys?
  - Other growing areas/priorities?

Mapping of IVOA standards to science cases

WG	Standard	Task Driver
DAL	TAP - Table Access Protocol (Rel)	Querying catalogues remotely
DAL	SLAP - Simple Line Access Protocol	VAMDC Accessing lines
DM	SSLDM v2.0	VAMDC Accessing lines
DAL	ObsLocTAP - Observation Locator Table Access Protocol	XMM-Newton, Integral, Chandra Coordinating observations between facilities
DAL	ObjVisSAP - Object Visibility Simple Access Protocol	XMM-Newton, Integral, Chandra Coordinating observations between facilities
DAL	ProvTAP - Provenance Table Access Protocol	CTA, Pollux, SVOM, MuseWise, CDS DB, triple Store Accessing provenance DM through TAP. Also useful for ML
DM	Provenance	CTA, Pollux, SVOM, MuseWise, CDS DB, triple Store Accessing provenance DM through TAP. Also useful for ML
DAL	ProvSAP	CTA, Pollux, SVOM, MuseWise, CDS DB, triple Store Accessing provenance DM through TAP. Also useful for ML
DAL	Datalink	Data discovery Links to further data
DAL	DALI - Data Access Layer Interface	Core standard to support other DAL standards. Allow DAL standards to do things consistently IVOA infrastructure
DAL	ADQL - Astronomical Data Query Language (Rel)	Community Input

Mapping of IVOA standards to science cases

WG	Standard	Task Driver
DM	Coordinates model	CSP priority - Multi Dimension use case - Needed for the Cubes - Search for space, time and Coordinates Partial replacement of STC
DM	Transform model	CSP priority - Multi Dimension use case - Needed for the Cubes - Search for space, time and Coordinates Partial replacement of STC
DM	Measurement model	CSP priority - Multi Dimension use case - Needed for the Cubes - Search for space, time and Coordinates Partial replacement of STC
DM	Cube v1.0	CSP priority - Multi Dimension use case - Needed for the Cubes - Search for space, time and Coordinates Partial replacement of STC
DM	VODML-Mapping	IVOA Infrastructure Needed to serialize IVOA DMs
DM	DatasetMetadata v1.0	Generalized DM for datasets Link with CAOM ? Multi Dimension phase 2
DM	Translation of DMs to VODML/ VODML modeling utilities	IVOA Infrastructure
DM	SourceDM	Gaia, Chandra DM for catalogues - Data discovery for sources
DM	TimeSeries	Comparing light curves
GWS	Group Membership Service 1.0	User authorization - Data sharing between groups of users
GWS	VOSI Best Practice	IVOA Infrastructure

Mapping of IVOA standards to science cases

WG	Standard	Task Driver
GWS	VOSI	IVOA Infrastructure
GWS	Science Platforms	Bring the code to the data Share data and code among users Collaborative virtual environment
Reg	RegTAP	IVOA Infrastructure TAP query to the IVOA Registry
Reg	STC in Registry	Identify the coverage of the VOResources in the Registry
Semantics	Maintenance of the list of UCD words	IVOA infrastructure
Semantics	UCD1+ controlled vocabulary	IVOA infrastructure

High-level view of current priorities

- Visualization tools
- Data access
  - Discovery
  - Access
  - Interoperability
- Data context
- Computing resources
- Data analysis
- Follow-up time-domain event

A possible way to prioritize the WG and IG activities

WSJF = Weighted Shortest Job First

https://www.scaledagileframework.com/wsjf/

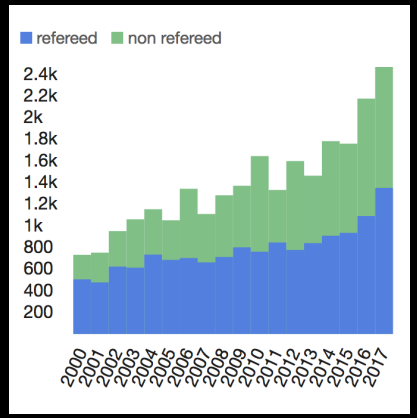
Possible prioritization of activities within IVOA

	Business value	Time criticality	Risk reduction - Opportunity enablement	Cost of Delay	Job size	Weighted Shortest Job First
Data access - access	20	8	2	30	2	15
Data Access - discovery	13	13	3	29	2	14,5
follow-up time-domain events	2	20	13	35	3	11,66666667
Data access - interoperability	8	5	8	21	5	4,2
Computing resources	5	8	8	21	8	2,625
Visualization tools	8	3	5	16	8	2
Data context (provenance, quality)	1	1	20	22	13	1,692307692
Data analysis tools	3	8	1	12	20	0,6

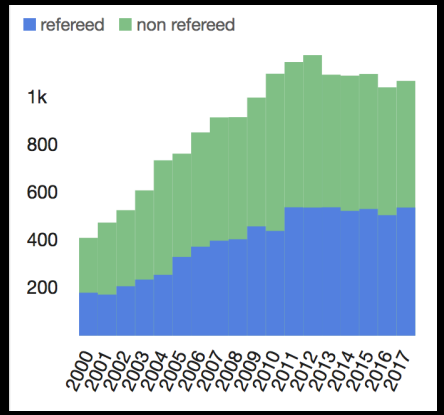
Draft!

Upcoming scientific priorities at IVOA

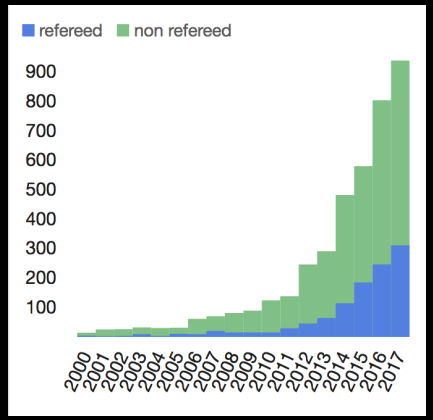
Gravitational waves



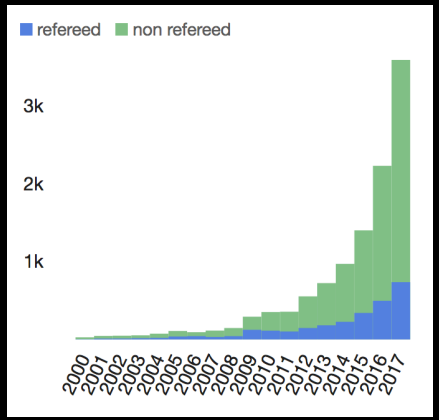
Multi-wavelength



Python



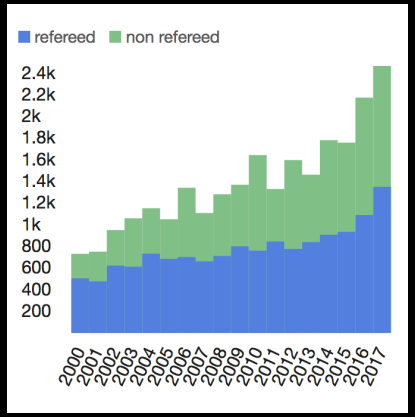
Machine learning



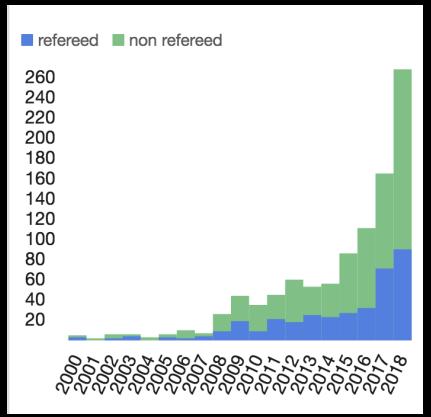
ADS-listed articles containing those key-words as a function of time

Upcoming scientific priorities at IVOA

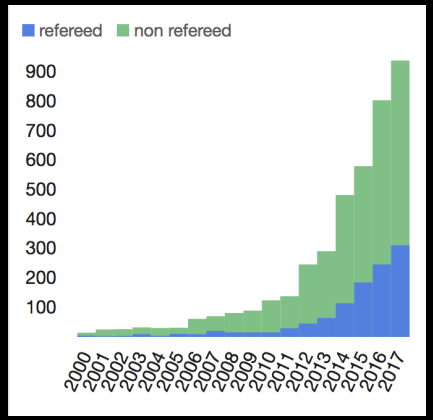
Gravitational waves



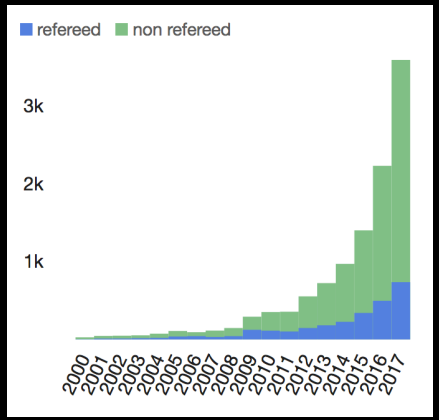
Multi-messenger



Python



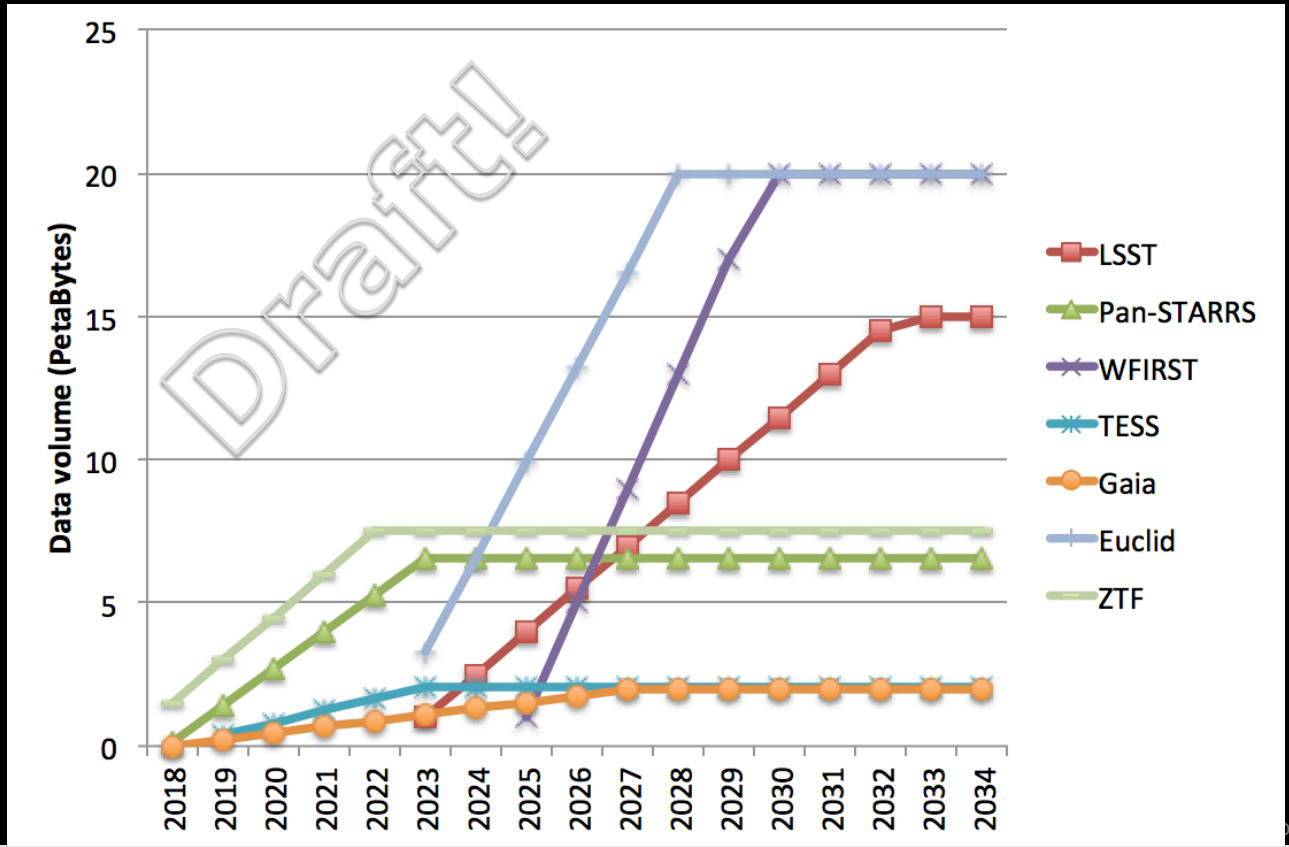
Machine learning



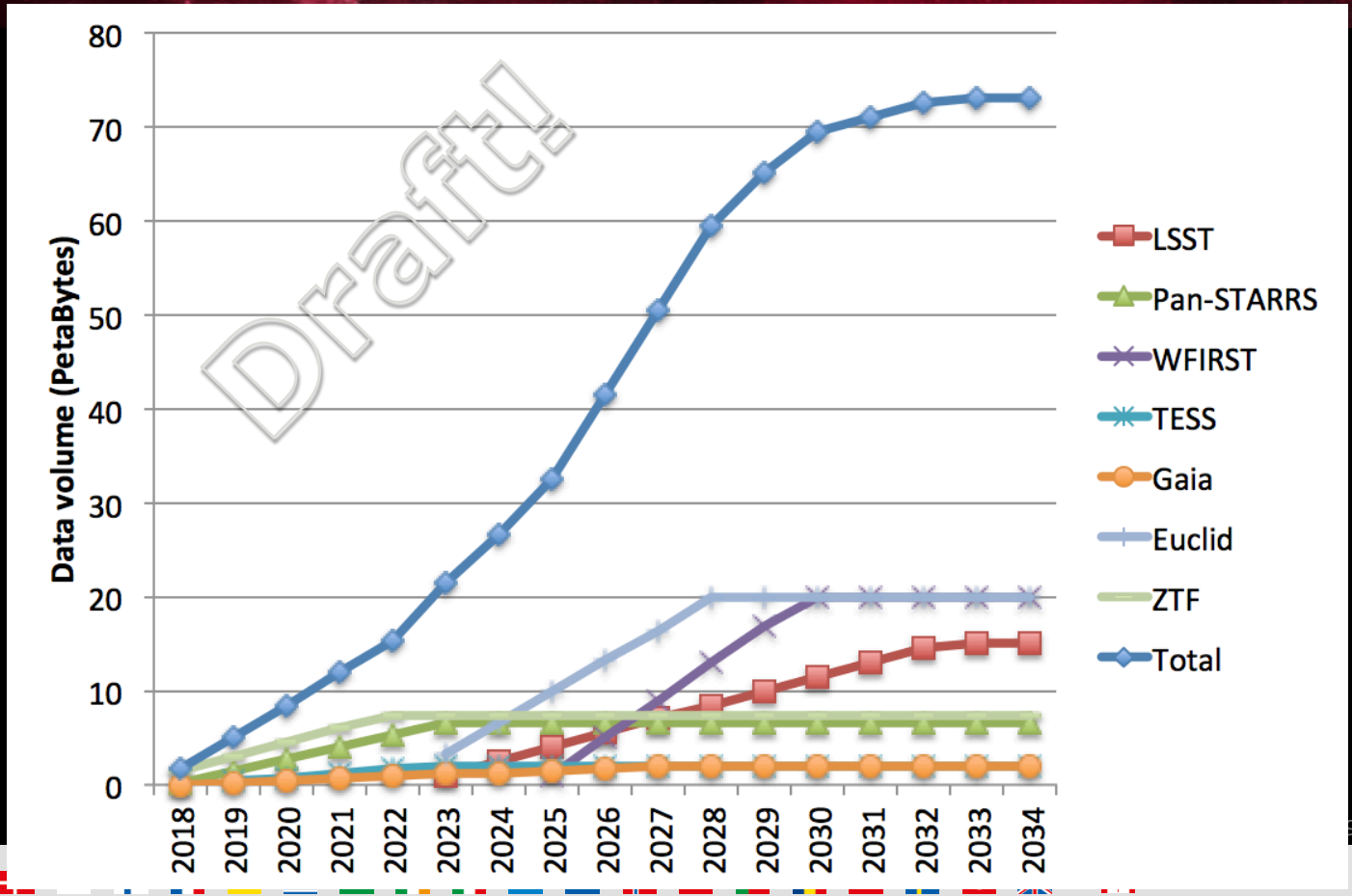
ADS-listed articles containing those key-words as a function of time

Survey	Q1: max. data volumes & dates	Q2: plans	Q3: code to data?	Q4: X-correlate?	Q5: IVOA?
LSST	30e12 sources 15 PBs (2023 to 2033)	Portal + Notebook + Web APIs	Yes	At least w/ Gaia	TAP+ADQL, SIAv2, SODA, VOSpace, WebDAV, PyVO
Pan-STARRS	11e9 sources 1.4 PBs (Jan 2019)	MASP portal + astroquery	AWS ?	?	MAST API, TAP, Cone, likely DataLink, SODA
WFIRST	20 PBs (~2025 TBC)	MASP portal + astroquery	AWS ?	?	MAST API, TAP, Cone, likely DataLink, SODA
TESS	260 TBs (2018 – 2020+)	MASP portal + astroquery	AWS ?	?	MAST API, TAP, Cone, likely DataLink, SODA
Gaia	2 PB (2018 - 2027)	Gaia archive + astroquery	SEPP	Several all-sky cats.	TAP+, datalink, SODA, VOSpace
Euclid	20 PB (2022 - 2028)	Euclid archive + astroquery	SEPP	Several all-sky cats.	TAP+, datalink, SODA, VOSpace
SKA	600 PB/year (2028 -)	TBD	TBD	TBD	TBD
ZTF	63e9 sources 1.5 PBs (2018 -)	IRSA APIs	?	?	IRSA APIs

Forecast of evolution in public astronomy data volumes



Forecast of evolution in public astronomy data volumes



Possible solutions to the big data challenges

- **To distribute TBs of data to users**
  - **Most users don't need TBs, but just MBs or GBs**
  - **They will run jobs in a cloud and download less data**
- **Allow people discovering datasets from new surveys**
  - **Add data to registry?**
- **Provide sufficient computing resources for users (who pays?)**
  - **Clone data to commercial clouds?**
- **What is missing in the IVOA set of standards?**
  - **SODA for cut-outs, rebinning, resampling?**
  - **Interoperable Notebooks?**
  - **Hierarchical data structures for discoverability?**
  - **Standard for code-to-the-data?**
  - **New data distribution standards like torrent?**

The IVOA needs you

- We need active and enthusiastic scientists at the Committee of Science Priorities!!
- Talk to us if you are interested!!

Final recommendations from the CSP

- You have the best judgement for what to work on !
- But always ask the question: how is the user going to use this?
- Try follow the user workflow to the paper and keep the big picture (is provenance clear? Can I explain/make a plot of this?)
- Connect to the future generation of users where they are: e.g. python, github, open source projects, social media, online open fora, connected to new big astronomy projects, using mobile devices and expecting quick answers

Thanks for all the great work!

Committee on Science Priorities : csp@ivoa.net