Data Discovery in and Science Results by means of VOs

Masatoshi OHISHI　（大石雅寿）

Director, Astronomy Data Center,
National Astronomical Observatory of Japan
&
President of Commission 5, IAU

masatoshi.ohishi@nao.ac.jp

Structure of my Lecture

- Era of Data Intensive Sciences
  - toward "4th paradigm"
- Data Discovery in Astronomy
  - How to find necessary data for our research
- Towards Standardization
  - Differences can be overcome
- Examples of Science Results
  - Just a flavor
- Data Science in Other Fields
  - We share the same problems

Era of Data Intensive Sciences

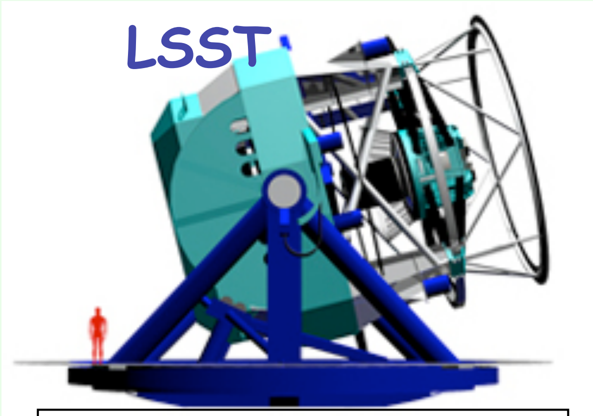
Accelerating Discoveries

- Issues, Planning
- Observation
- Data Reduction
  - Calib., Select, Combine, , ,
- Data Analysis
  - Physical Parameters
  - Thinking
  - Solution
- Publish

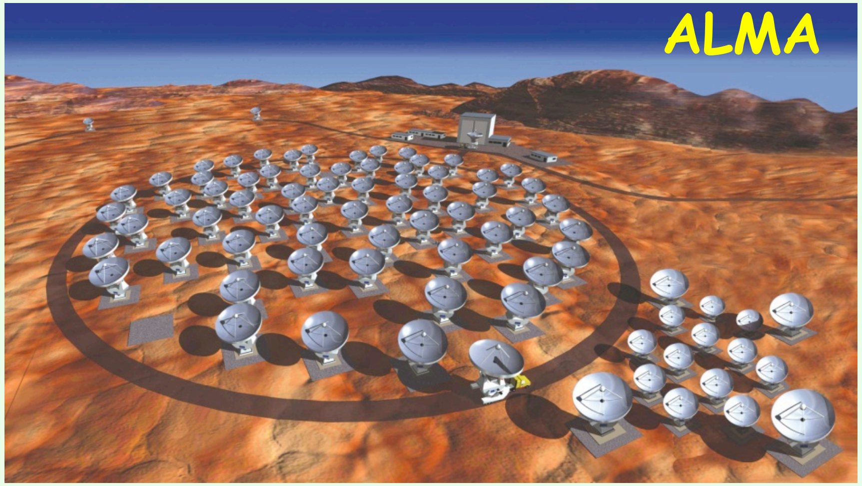
Data
↓
Information
↓
Knowledge
↓
Understanding
↓
Wisdom

Planned Future Astronomy Projects

- ALMA
- JWST
- LSST
- LOFAR
- SKA
- TMT
- Pan-STARRs
- E-ELT

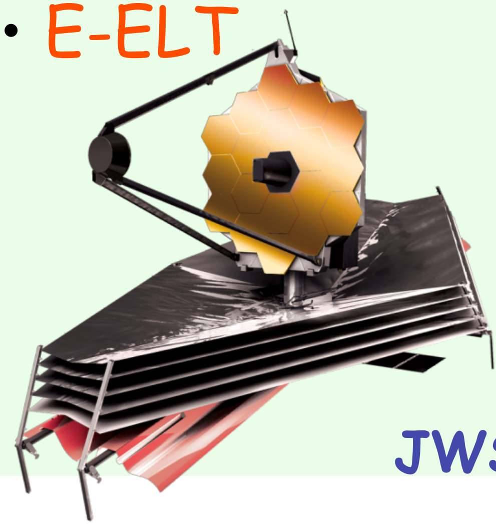


30 PB/yr x 6 yr ~ 200 PB

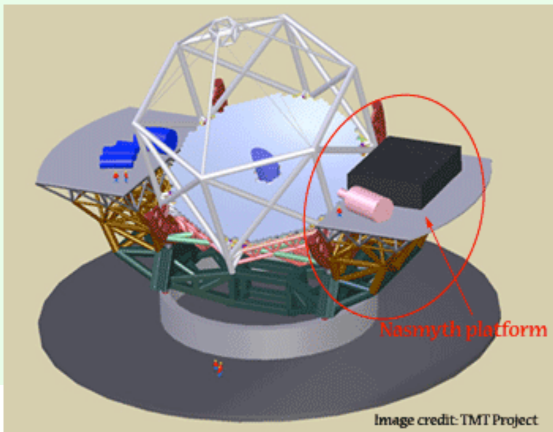


~ a few PB/yr

~ a few TB/night , only object params stored



JWST



TMT

Two Major Categories

Pointing Obs.

- ALMA
- JWST

Surveys

- LSST
- Pan-STARRs

cosmology, the large-scale structure of the Universe, formation of galaxies, star formation, variable stars, transient phenomena such as the Gamma-ray bursts, small bodies in the solar system, extrasolar planets, life in the Universe, dark matter and dark energy, and others

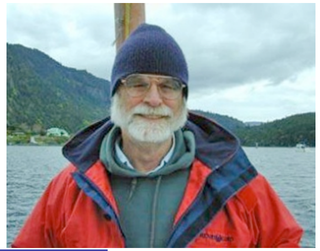
Large collecting area

High resolution

Whole sky

Time-domain astronomy

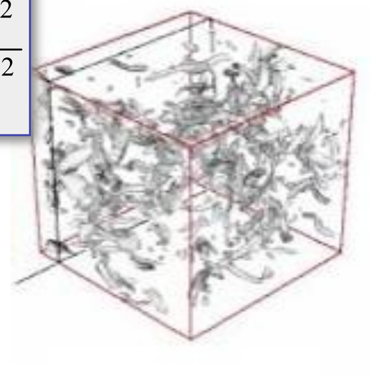
Science Paradigms



- Thousand years ago:
  science was **empirical**
  -- **observations / experiments**

- Last few hundred years:
  theoretical studies

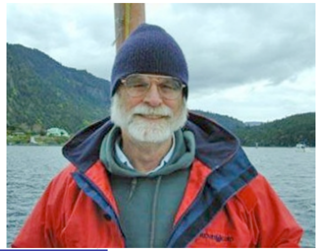
$$\left(\frac{\dot{a}}{a}\right)^2 = \frac{4\pi G\rho}{3} - K\frac{c^2}{a^2}$$



- Last few decades:
  simulations

- Today:
  data exploration (e-Science)
  unify theory, experiment, and simulation
  – High-speed network
  – Computers, storages, databases

Science Paradigms

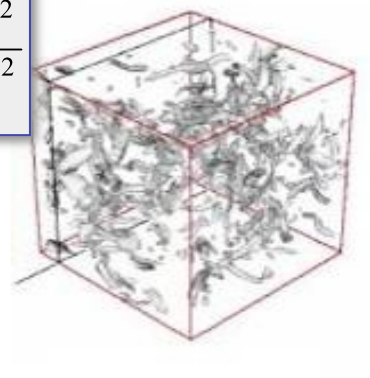


- Thousand years ago:
  science was **empirical**
  describing natural phenomena

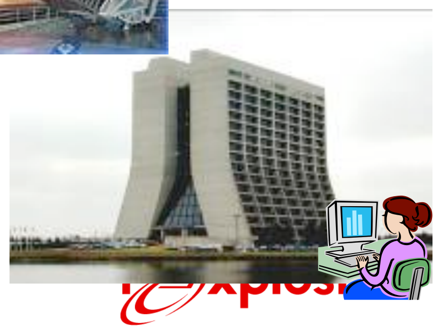
- Last few hundred years:
  theoretical branch
  using models, generalizations

$$\left(\frac{\dot{a}}{a}\right)^2 = \frac{4\pi G\rho}{3} - K\frac{c^2}{a^2}$$

- Last few decades:
  a **computational** branch
  simulating complex phenomena



- Today:
  data exploration (e-Science)
  unify theory, experiment, and simulation
  - Data captured by instruments
    Or generated by simulator
  - Processed by software
  - Information/Knowledge stored in computer
  - Scientist analyzes database / files
    using data management and statistics



Are we prepared for such a new era ?

Requirements in the Data Intensive Science Era

Data producer side

- Definition of data quality index, and establishment

Quality assurance of data (from obs. to data analyses)

Data center side

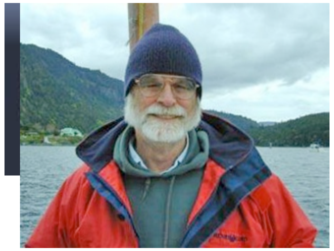
- Establishment of data handling environment

incl. data mining, knowledge discovery, statistics, event discovery

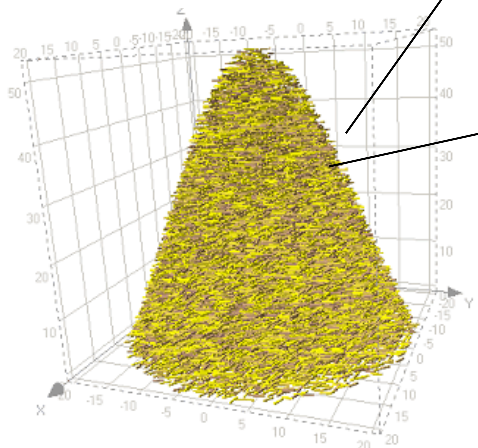
  – High-speed network

Data management / analysis cost will become a major issue

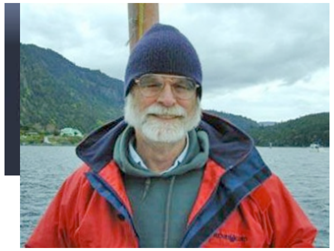
Data Analysis



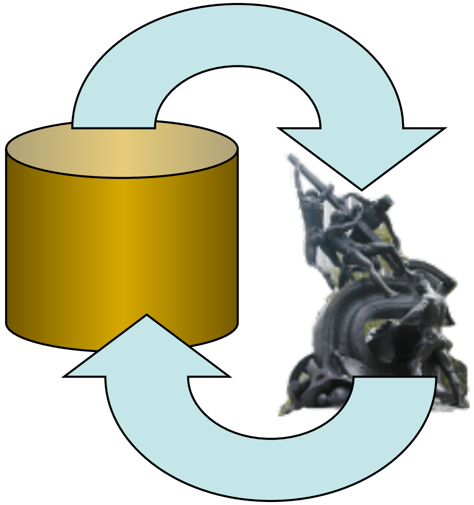
- Looking for
  - Needles in haystacks – the Higgs particle
  - Haystacks: Dark matter, Dark energy
- Needles are easier than haystacks
- Global statistics have poor scaling
  - Correlation functions are N^2, likelihood techniques N^3
- We can only do $N \log N$
- Must accept approximate answers
  New algorithms
- Requires combination of
  - statistics &
  - computer science



Accessing Data

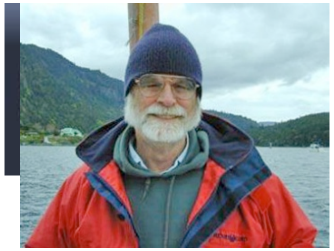


- If there is too much data to move around,
  take the analysis to the data!
- Do all data manipulations at database
  - Build custom procedures and functions in the database
- Automatic parallelism guaranteed
- Easy to build-in custom functionality
  - Databases & Procedures being unified
  - Example temporal and spatial indexing
  - Pixel processing
- Easy to reorganize the data
  - Multiple views, each optimal for certain analyses
  - Building hierarchical summaries are trivial
- Scalable to Petabyte datasets

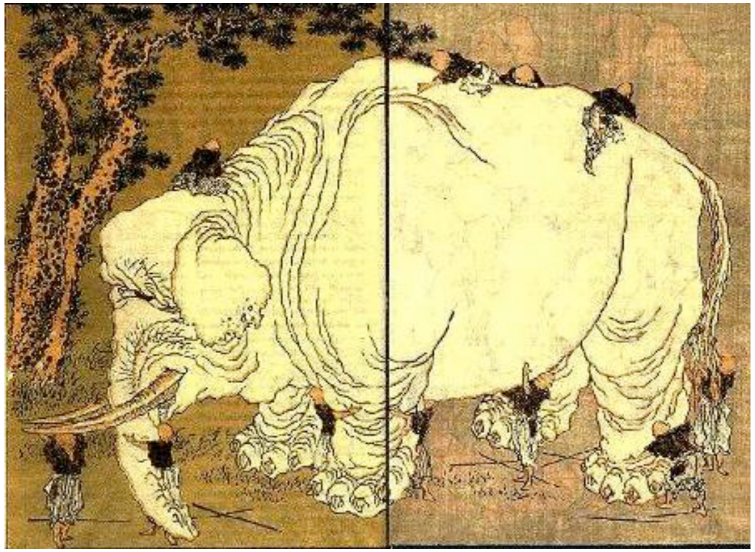


active databases!

Analysis and Databases



- Much statistical analysis deals with
  - Creating uniform samples –
  - data filtering
  - Assembling relevant subsets
  - Estimating completeness
  - Censoring bad data
  - Counting and building histograms
  - Generating Monte-Carlo subsets
  - Likelihood calculations
  - Hypothesis testing

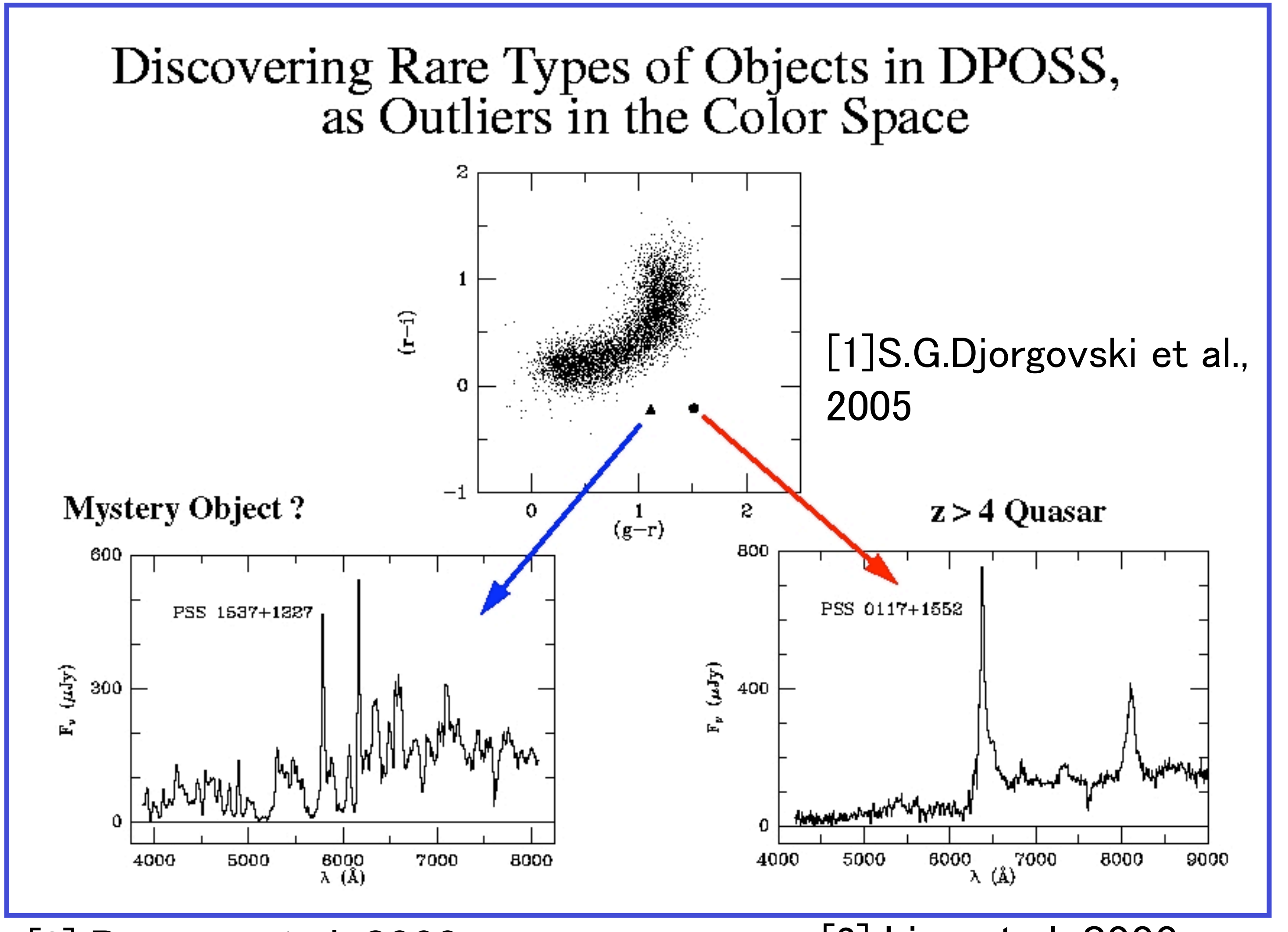


- Traditionally performed on files
- These tasks better done in structured store with
  - indexing,
  - aggregation,
  - parallelism
  - query, analysis,
  - visualization tools.

Getting Knowledge

- Approaches on Data analyses: mathematical statistics and/or taxonomy
- With scientific working hypothesis – what do we want to know from the deluge of data ?
  - We need to have a sensitive antenna
  - Serendipitous discoveries might be possible, but…
- Data publication as early as possible
- Challenging researchers in exploring the deluge of data

mystery outliers



[1]S.G.Djorgovski et al., 2005

[2] Brunner et al.,2003

[3] Lisa et al.,2000

graphics from US NVO project

Data Discovery in Astronomy

VO– New Research Infrastructure in the 21st Century

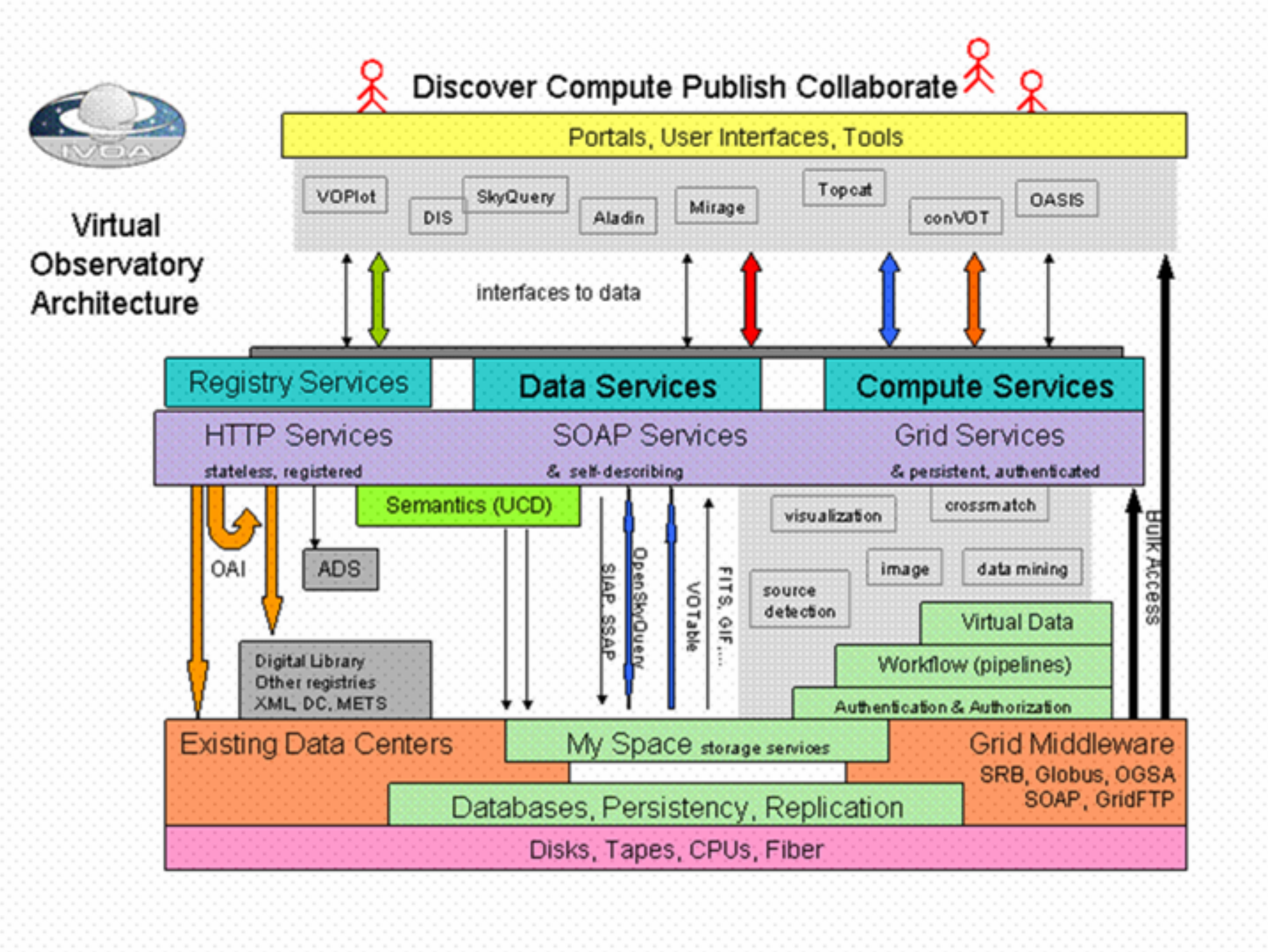
A collection of integrated astronomical data archives and software tools that utilize computer networks to create an environment in which research can be conducted.

http://www.encyclopedia.com/html/v1/virtobserv.asp

VO Projects in the world

- 18 members worldwide
- International Virtual Observatory Alliance (IVOA – http://www.ivoa.net/)
  → Standards to interoperate VOs

- No center (good-will), No shared project funding


Virtual Observatory Architecture
IVOA
Discover Compute Publish Collaborate
Portals, User Interfaces, Tools
VOPlot
DIS
SkyQuery
Aladin
Mirage
Topcat
conVOT
OASIS
interfaces to data
Registry Services
Data Services
Compute Services
HTTP Services
stateless, registered
SOAP Services
& self-describing
Grid Services
& persistent, authenticated
Semantics (UCD)
OAI
ADS
Digital Library
Other registries
XML, DC, METS
SIAP, SSAP
OpenSkyQuery
VOTable
FITS, GIF,...
visualization
crossmatch
image
data mining
source
detection
Virtual Data
Workflow (pipelines)
Authentication & Authorization
Bulk Access
Existing Data Centers
My Space storage services
Grid Middleware
SRB, Globus, OGSA
SOAP, GridFTP
Databases, Persistency, Replication
Disks, Tapes, CPUs, Fiber

Standardization in IVOA

- Meta-data
  - Contents & access protocol
- Access Images, Spectra, Catalogues
  - TAP, SIAP, SSAP, STC, etc.
- Query Language to Federated DBs (ADQL)
- Unified Attribute Names
  - UCD (Unified Contents Descriptions)
- Output format : VOTable (in XML)
  - FITS

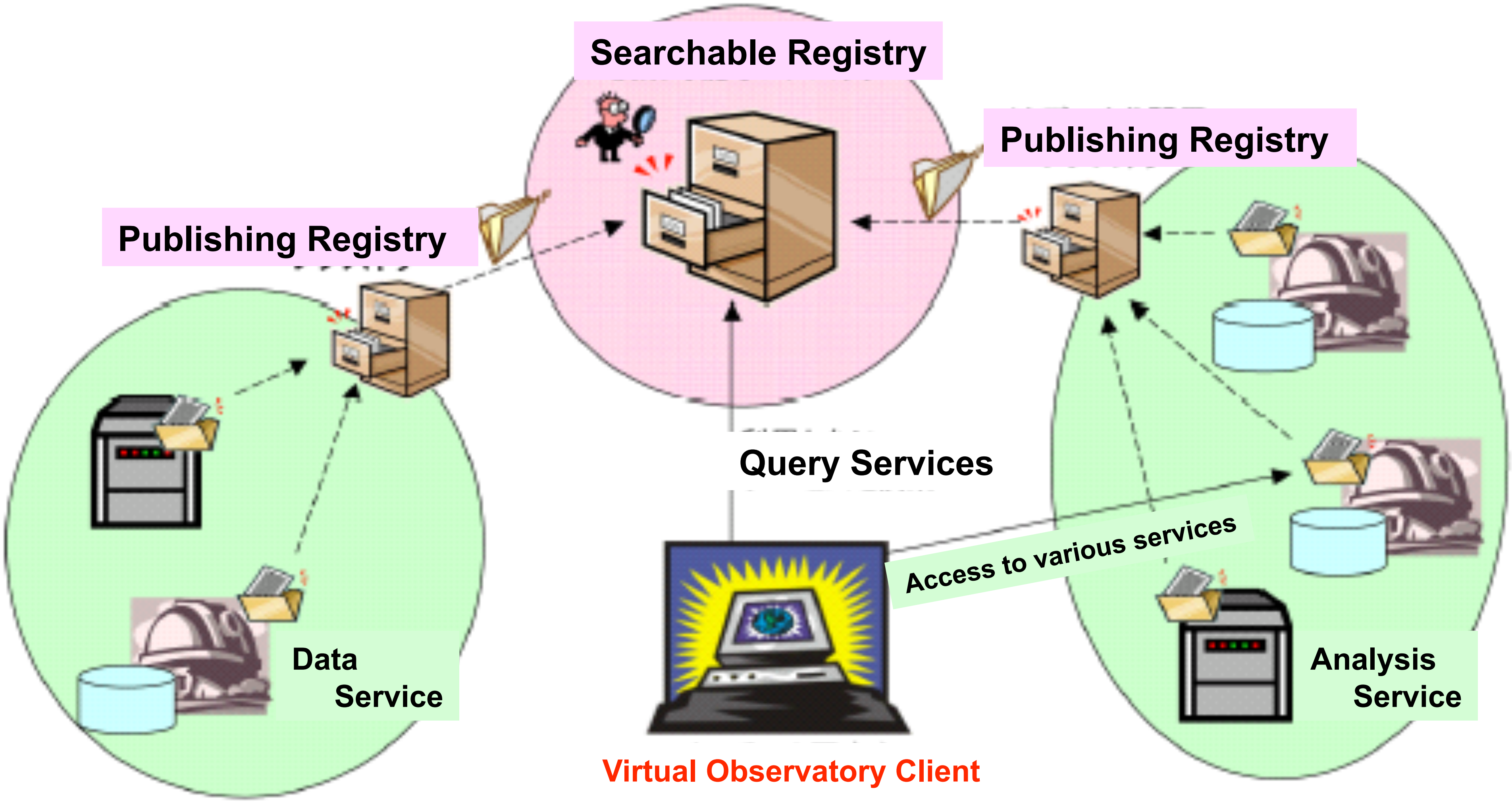
Resource Metadata

2011 Sep 28

- Resource Identification:
  - Title, ShortName, Identifier
- Curation:
  - Publisher, PublisherID, Creator, Creator.Logo, Contributor, Date, Version, Contact.Name, Contact.Email
- General content:
  - Subject, Description, Source, ReferenceURL, Type, ContentLevel, Relationship, RelationshipID

Exchange of Meta Data: OAI-PMH

2011 Sep 28



Data Access Protocols

- Parameter query in terms of the HTTP

```
http://jvo.nao.ac.jp/imageData?Pos=24,5&Size=0.2&format=VOTable
```

  - Simple Image Access Protocol (SIAP)
  - Simple Spectrum Access Protocol (SSAP)
  - Table Access Protocol (TAP)

  etc.

- Unified query language (JVOQL) for both the catalog and observation data such as image data, spectrum, 3D-cube, photon list …

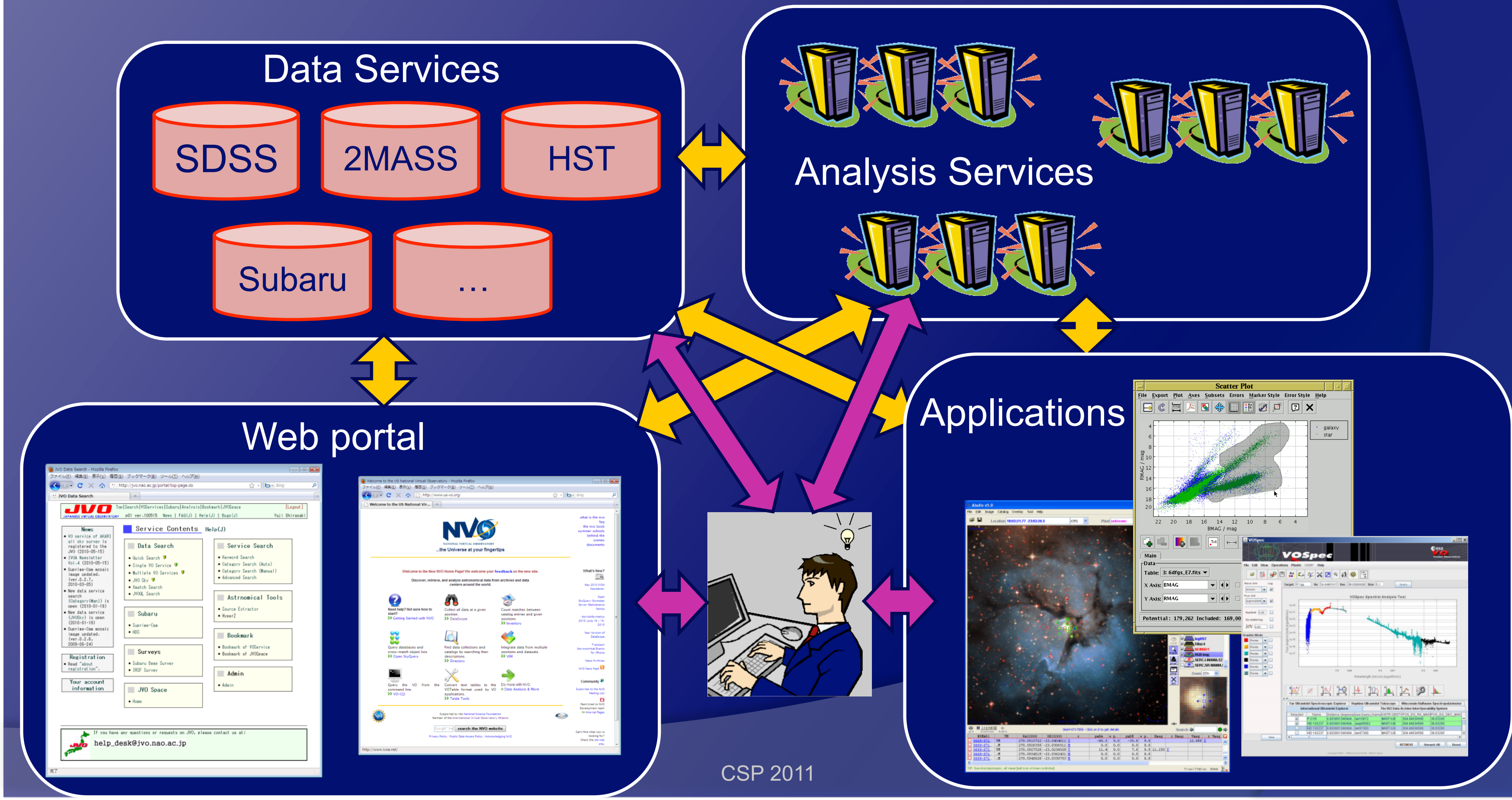
```
Select    imageURL, …
From      naoj:imageData
Where     pos=Point(24,5) and size=0.2 and format='VOTable'
```

File Formats

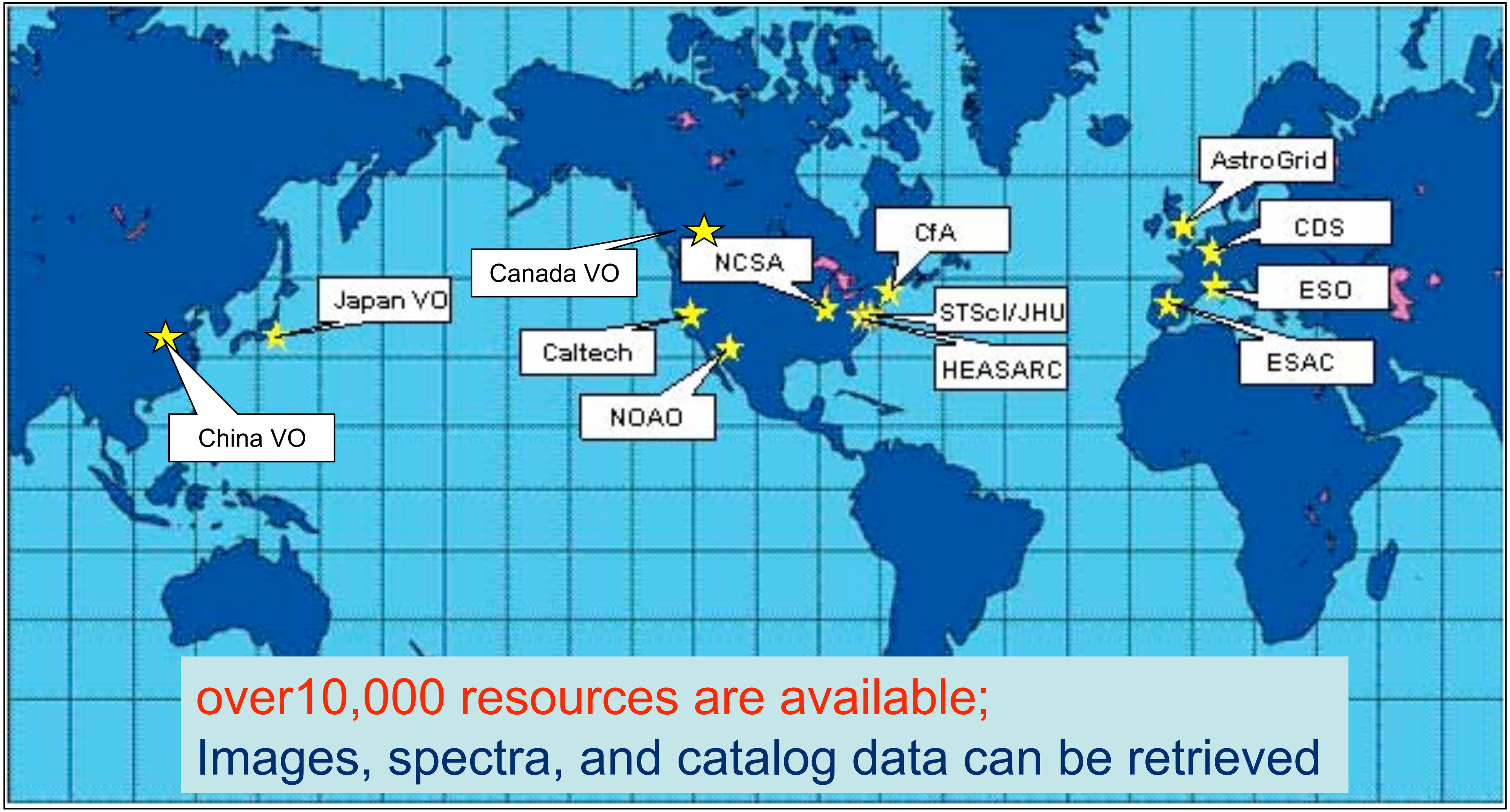
- Flexible Image Transfer System (FITS)
  - standardized in early 80's to exchange observed data
  - 1 record = 2880 bytes
  - (Header, Data)(Header, Data)･･･
  - IAU has the FITS WG to maintain its specification
- VOTable
  - used in Virtual Observatories as an output format
  - described in XML, and standardized in IVOA
  - can inline FITS files / contain a link to FITS files

Virtual Observatory

- ✓ Infrastructure for **efficient** research environment
- ✓ **International standards** for data publication & access
- ✓ **Sharing** data worldwide, **Maximize** scientific return



Astronomical Virtual Observatories
~ Data Grid ~



Towards Standardization

Establishing Standards

- Standards are quite effective
  - Access protocols, data format, etc.
  - Interoperability → wider dissemination and application
  - Endorsement by the IAU (VO WG)

- Painful process
  - Philosophy, intention, life time of project,,,
  - Compromise, patience
  - Establishment of relationship：respect to each other
  - Coffee/tea breaks and lunch/dinner talks are crucial

IVOA Interoperability meetings

- Twice a year, since 2003
- Discussions toward standardization
- Human network as a basis for cyber network (Layer 0)

Nara, 2010 December

International Endorsements

- IAU XXVth GA Res. (2003 Jul.)
- OECD Rec. (‘04 Aug)
  - place archives that may be accessible via internet
  - provide adequate funding as long-term issues
- IAU DivXII (union-wide activity)
  - VO WG to endorse IVOA Recs since 2006

Examples of Science Results

VO Science

VO enables researchers

1) to find a small particular data subset from a large collection of catalog and observation data

2) to retrieve and use large amount data in an automated way

VO science papers

- ✓ http://www.ivoa.net/newsletter/ 150 refereed papers
- ✓ Over 1600 related publications

✓More and more VO science papers are coming

✓However, most of them are type-1 science case

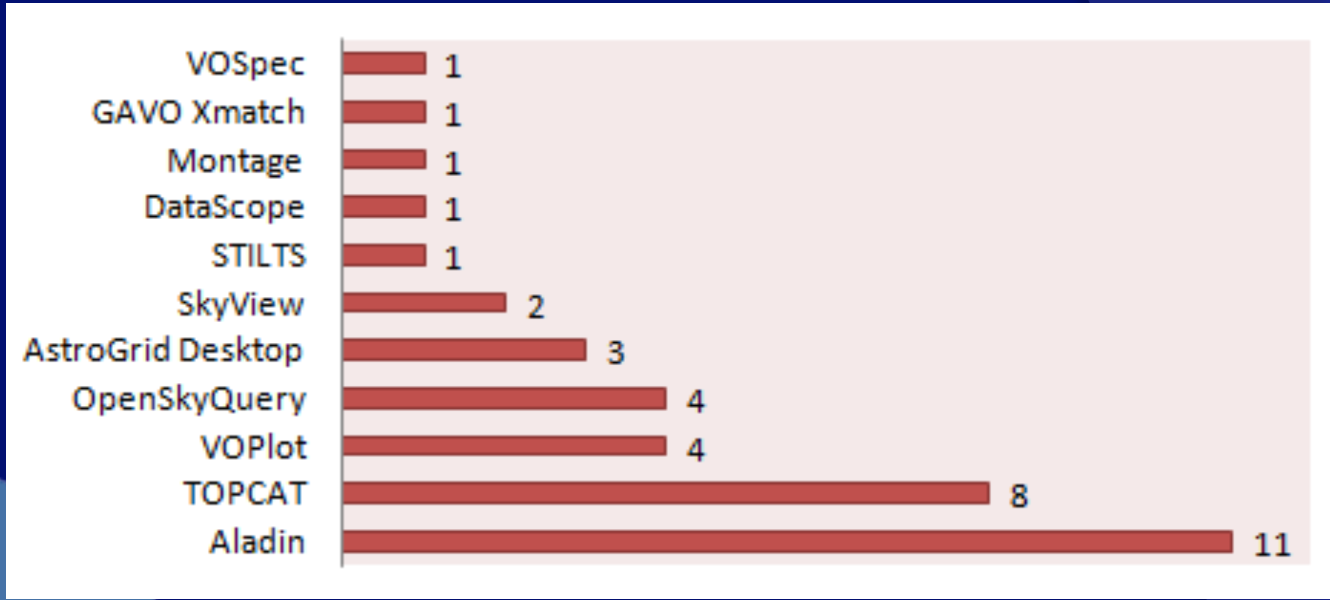
✓type-2 science case ?

Refereed Papers by Euro-VO

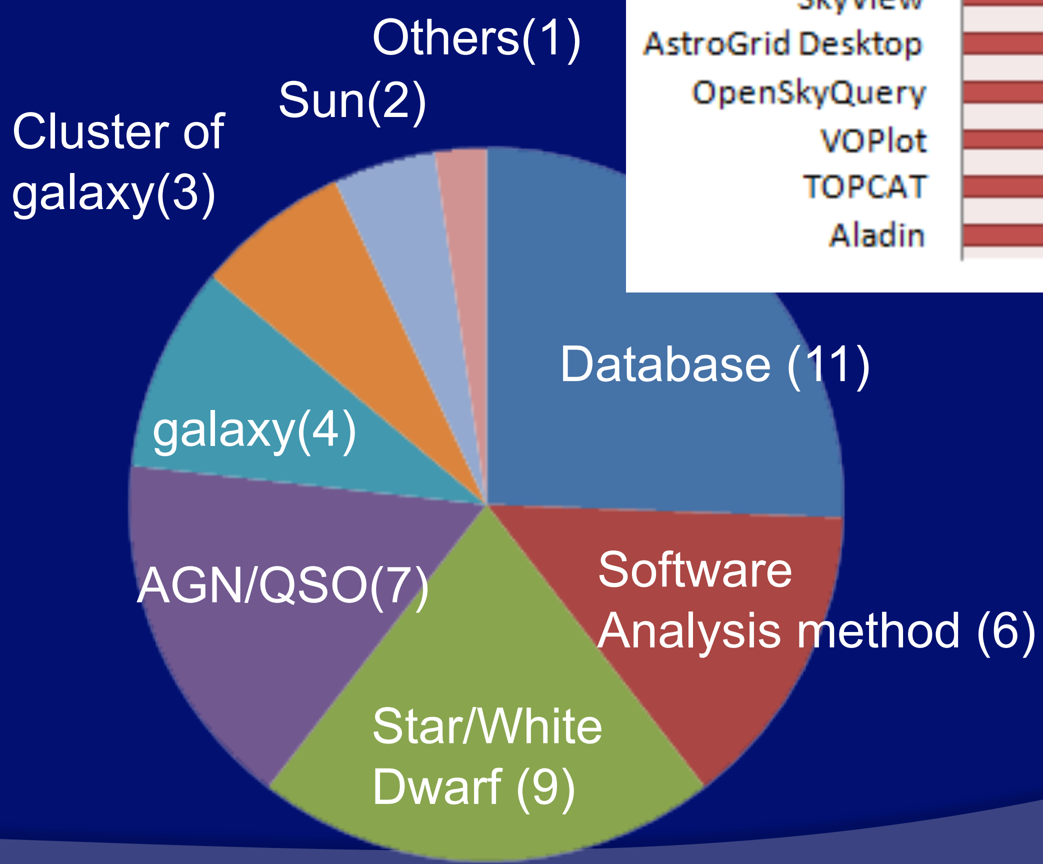
http://www.euro-vo.org/pub/fc/papers.html 43 papers

- The GalMer database: Galaxy Mergers in the Virtual Observatory Chilingarian I., Di Matteo P., Combes F., Melchior A.-L., Semelin B., A&A, in press
- Scale Lengths of Disk Galaxies Fathi K., Allen M., Boch Th., Hatziminaoglou E., Peletier R., MNRAS, in press
- SDSSJ150634.27+013331.6: the second compact elliptical galaxy in the NGC5846 group Chilingarian I & Bergond G., MNRAS Letters, in press
- VisIVO-Integrated Tools and Services for Large-Scale Astrophysical Visualization Becciani et al., 2010, PASP, 122, 119
- The SPECFIND V2.0 catalogue of radio cross-identifications and spectra. SPECFIND meets the Virtual Observatory Vollmer et al., 2010, A&A, 511, 53
- Montage: a grid portal and software toolkit for science-grade astronomical image mosaicking Jacob J.C. et al., Int. J. Computational Science and Engineering, 2009, vol 4, No. 2
- A Population of Compact Elliptical Galaxies Detected with the Virtual Observatory Chilingarian I. et al., 2009, Science, 326, 1379
- Properties of dusty tori in active galactic nuclei - II. Type 2 AGN Hatziminaoglou E., Fritz J., Jarrett T., 2009, MNRAS, 399, 1206
- The LAEX and NASA portals for CoRoT public data Solano et al., 2009, A&A, 506, 455
- The chemical abundance analysis of normal early A- and late B-type stars Fossati et al., 2009, A&A, 503, 945
- Exo-Dat: An Information System in Support of the CoRoT/Exoplanet Science Deleuil et al., 2009, AJ, 138, 649
- GALEX-SDSS Catalogs for Statistical Studies Budavari et al., 2009, ApJ, 694, 1281
- Data Mining and Machine Learning in Astronomy Ball M., Brunner R.J., 2009, International Journal of Modern Physics
- Quasar candidates selection in the Virtual Observatory era D'Abrusco R., Longo G., Walton N.A., 2009, MNRAS, 396, 223
- X-Ray Variability of sigma Orionis Young Stars as Observed with ROSAT Caballero J.A. et al., 2009, AJ, 137, 5012
- SDSS J125637-022452: A High Proper Motion L Subdwarf Sivarani T., Lépine S., Kembhavi A.K., Gupchup J., 2009, ApJ, 694, 140
- An IPHAS-based search for accreting very low-mass objects using VO tools Valdivielso L. et al., 2009, A&A, 497, 973
- The construction of the large quasar astrometric catalogue (LQAC) Souchay J. et al., 2009, A&A, 494, 799
- Stellar populations in a standard ISOGAL field in the Galactic disc Ganesh et al., 2009, A&A, 493, 785
- The Hubble Legacy Archive NICMOS grism data Freudling W. et al., 2008, A&A, 490, 1165
- Infrared Imaging of Sloan Digital Sky Survey Quasars: Implications for the Quasar K Correction Kennefick & Bursick, 2008, AJ, 136, 1799
- VOSA: Virtual Observatory SED Analyzer. An application to the Collinder 69 open cluster Bayo A. et al., 2008, A&A, 492, 277
- The TVO Archive for Cosmological Simulations: Web Services and Architecture Costa A. et al., 2008, PASP, 120, 933
- Initial data release from the INT Photometric Hα Survey of the Northern Galactic Plane (IPHAS) Gonzalez-Solares E. et al., 2008, MNRAS, 388, 89
- Young stars and brown dwarfs surrounding Alnilam (eps Ori) and Mintaka (del Ori) Caballero J.A. & Solano E., 2008, A&A, 485, 931
- SDSSJ124155.33+114003.7 -- a Missing Link Between Compact Elliptical and Ultracompact Dwarf Galaxies Chilingarian I.V. & Mamon G.A., 2008, MNRAS, 385, 83
- Invisible sunspots and rate of solar magnetic flux emergence Dalla S., Fletcher L., Walton, N. A., 2008, A&A, 479, L1
- Stars and brown dwarfs in the σ Orionis cluster: the Mayrit catalogue Caballero J.A., 2008, A&A, 478, 667
- Fossil Groups in the Sloan Digital Sky Survey Santos W.A., Mendes de Oliveira C., Sodre L. Jr, 2007, AJ, 134, 1551
- Albus 1: A Very Bright White Dwarf Candidate Caballero J.A., Solano E., 2007, ApJ, 665, L151
- Visualization, Exploration, and Data Analysis of Complex Astrophysical Data Comparato et al., 2007, PASP, 119, 898
- Using VO tools to investigate distant radio starbursts hosting obscured AGN in the HDF(N) region Richards A.M.S., Muxlow T.W.B., Beswick,R., et al., 2007, A&A, 472, 805
- Flare productivity of newly-emerged paired and isolated solar active regions Dalla S., Fletcher L., Walton N.A., 2007, A&A, 468, 1103
- eSDO algorithms, data centre and visualization tools Auden E., Toutain T., Zharkov S., 2007, AN, 328, 356
- The DRaGONS Survey: A Search for High-Redshift Radio Galaxies and Heavily Obscured Active Galactic Nuclei Schmidt S., Connolly A., Hopkins A., 2006, ApJ, 649, 63
- Cluster Merger Variance and the Luminosity Gap Statistic Milosavljevic M., Miller C.J., Furlanetto S.R., Cooray A., 2006, ApJ, 637, L9
- Spatial orientation of galaxies in the core of the Shapley concentration - the cluster Abell 3558 Aryal B., Kandel S.M., Saurer W., 2006, A&A, 458, 357
- SearchCal: a virtual observatory tool for searching calibrators in optical long baseline interferometry. I. The bright object case Bonneau D., Clausse J.-M., Delfosse X. et al., 2006, A&A, 456, 789
- The shaping of planetary nebula Sh2-188 through interaction with the interstellar medium Wareing C.J., O'Brien T.J., Zijlstra A.A. et al., 2006, MNRAS, 366, 387
- Radio-loud Narrow-Line Type 1 Quasars Komossa S., Voges W., Xu D., Mathur S. et al., 2006, ApJ, 132, 531
- Luminous AGB stars in nearby galaxies. A study using virtual observatory tools Tsalmantza P., Kontizas E., Cambrésy L., Genova F., Dapergolas A., Kontizas M., 2006, A&A, 447, 89
- Using Virtual Observatory Tools for Astronomical Research Kim S.C., Taylor J.D., Panter B., Sohn S.T., Heavens A.F., Mann R.G., 2005, Journal of the Korean Astronomical Society, 38, 85
- Discovery of optically faint obscured quasars with Virtual Observatory tools Padovani R., Allen M.G., Rosati P., Walton N.A., 2004, A&A, 424, 545

Distribution of VO Apps Usage



Classification on VO papers



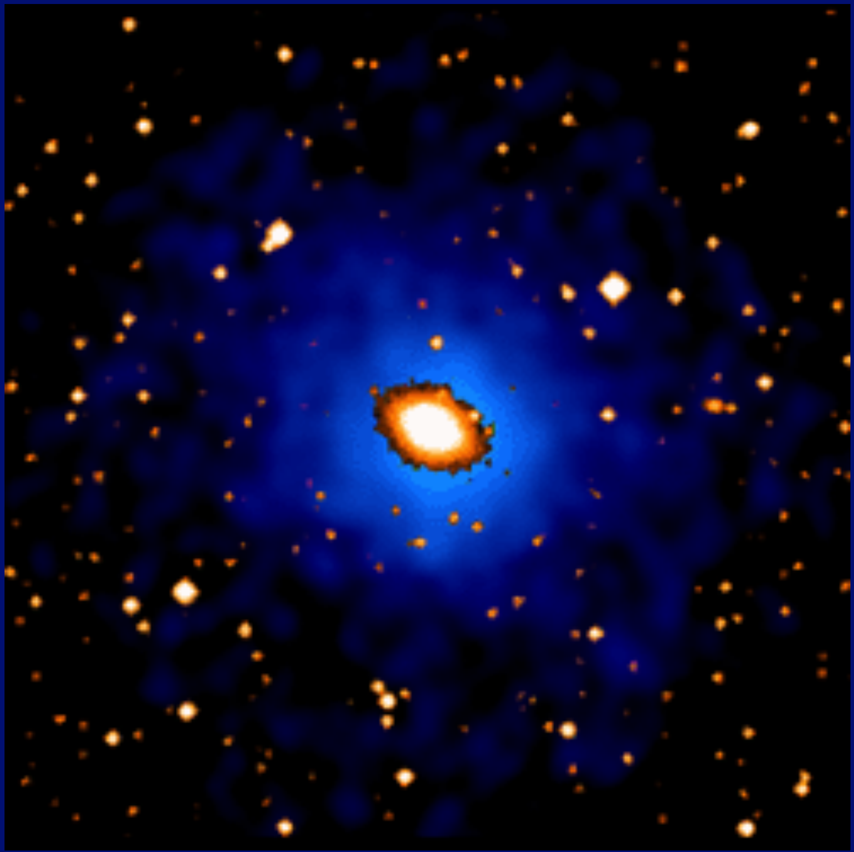
Use Cases in Using VOs (1)

- ✓ Specify data services; similar to the traditional ways
  - – Find suitable data services, and access them one by one
    - ➢ VOs provide user interfaces to access DBs
  - – Simple & homogeneous query IFs
    - ➢ Region queries, Get all data, etc.
    - ➢ Possible to use a single IF in accessing different data services
  - – Web portals vs applications
    - ➢ Web portals : no need to install IFs
    - ➢ Apps : provides with high-level GUIs
  - – Save results in the VOTable format
    - ➢ Readable in various VO apps

Ex. 1: "Fossil Groups in the Sloan Digital Sky Survey" W. A.Santos et al, 2007, ApJ, 134, 1551

- ✓ Exploration of Fossil groups by means of VOs
- ✓ Fossil group :
  - – A system with an isolated bright, giant elliptical galaxy at its center
  - – Mass & X-ray brightness ~ local group
  - – Possibly a final status after merges and/or coagulations
- ✓ Method
  - – Use OpenSkyQuery (NVO portal)
  - – Cross-matching SDSS LRG & Rosat All Sky catalogues
  - – Find elliptical galaxies with extended X-ray emissions
  - – Associated galaxies within 0.5 h^{-1}_{70} Mpc are dimmer by more than 2 mag
- ✓ 34 candidates (only 15 before)

Fossil Group



Chandra X-ray observations of the giant elliptical galaxy NGC 6482

DSS image of NGC 6482

Ex. 2: "ALBUS 1: A Very Bright White Dwarf Candidate"

J. A. Caballero & E. Solano, 2007, ApJ, 665, L151

- ✓ Discovery of (candidate of) a bright white dwarf (B_T = 11.8)
- ✓ Accidental discovery while investigating regions around Alnilam (ε Ori) and Mintaka (δ Ori)
- ✓ X-match Tycho-2 and 2MASS catalogues by Aladin
- ✓ Extremely blue star on a color-mag diagram
- ✓ 12th brightest white dwarf (isolated)

[5]Caballero et al.,2007

FIG. 1.—False-color composite image, 5.6 × 5.6 arcmin^2 wide, centered on Albus 1. North is up, east is left.

Color-Mag diagram

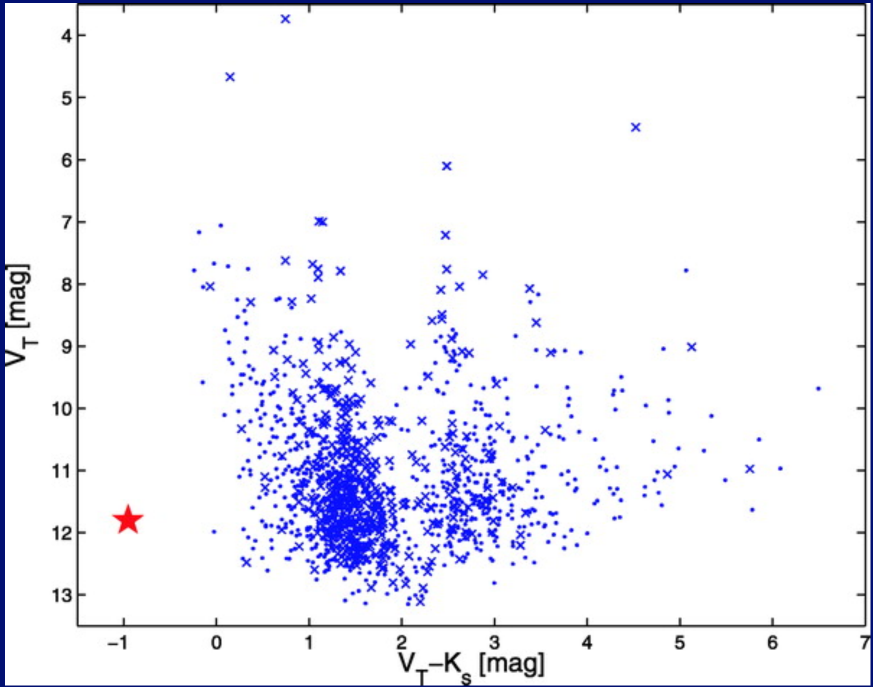


FIG. 2.—V_T vs. $V_T - K_s$ color-magnitude diagram from the data in J. A. Caballero & E. Solano (2007, in preparation). Tycho-2/2MASS sources with proper motions larger and smaller than 15 mas yr^{-1} are shown with crosses and dots, respectively. Albus 1 is highlighted with a big filled star.

SED of the candidate (Albus 1)

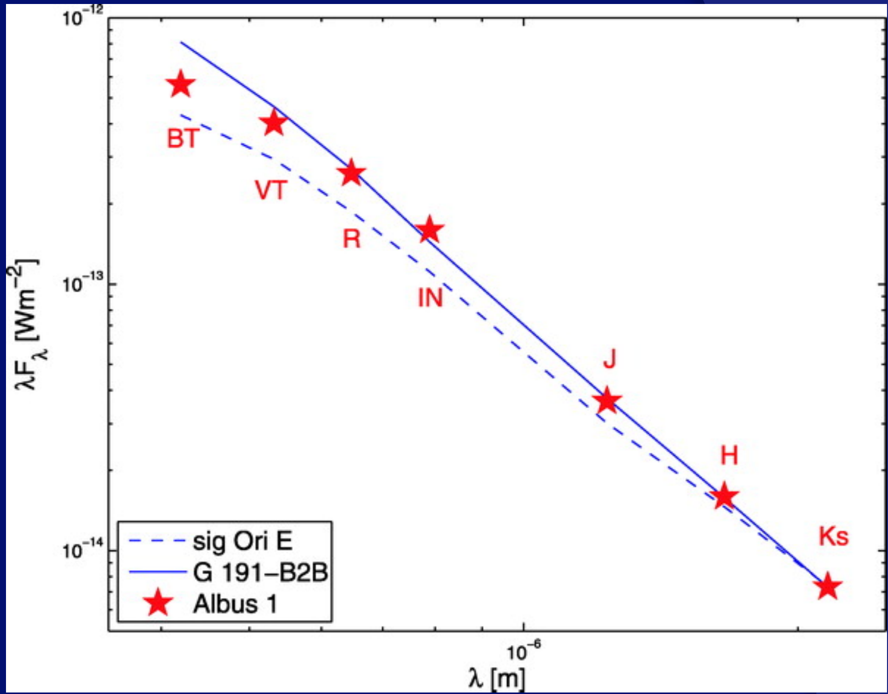


FIG. 4.—Spectral energy distributions of Albus 1, the DA1 white dwarf G191-B2B, and the B2 Vp star σ Ori E (shifted to a heliocentric distance of 0.5 kpc). The seven passbands ($B_T V_T R I_N J H K_s$) are indicated.

* Confirmed to be a B subdwarf by spectrum observations

S. Vennes et al (2007) ApJ 668 L59
"CPD -20 1123 (Albus 1) Is a Bright He-B Subdwarf"

Use Cases in Using VOs (2)

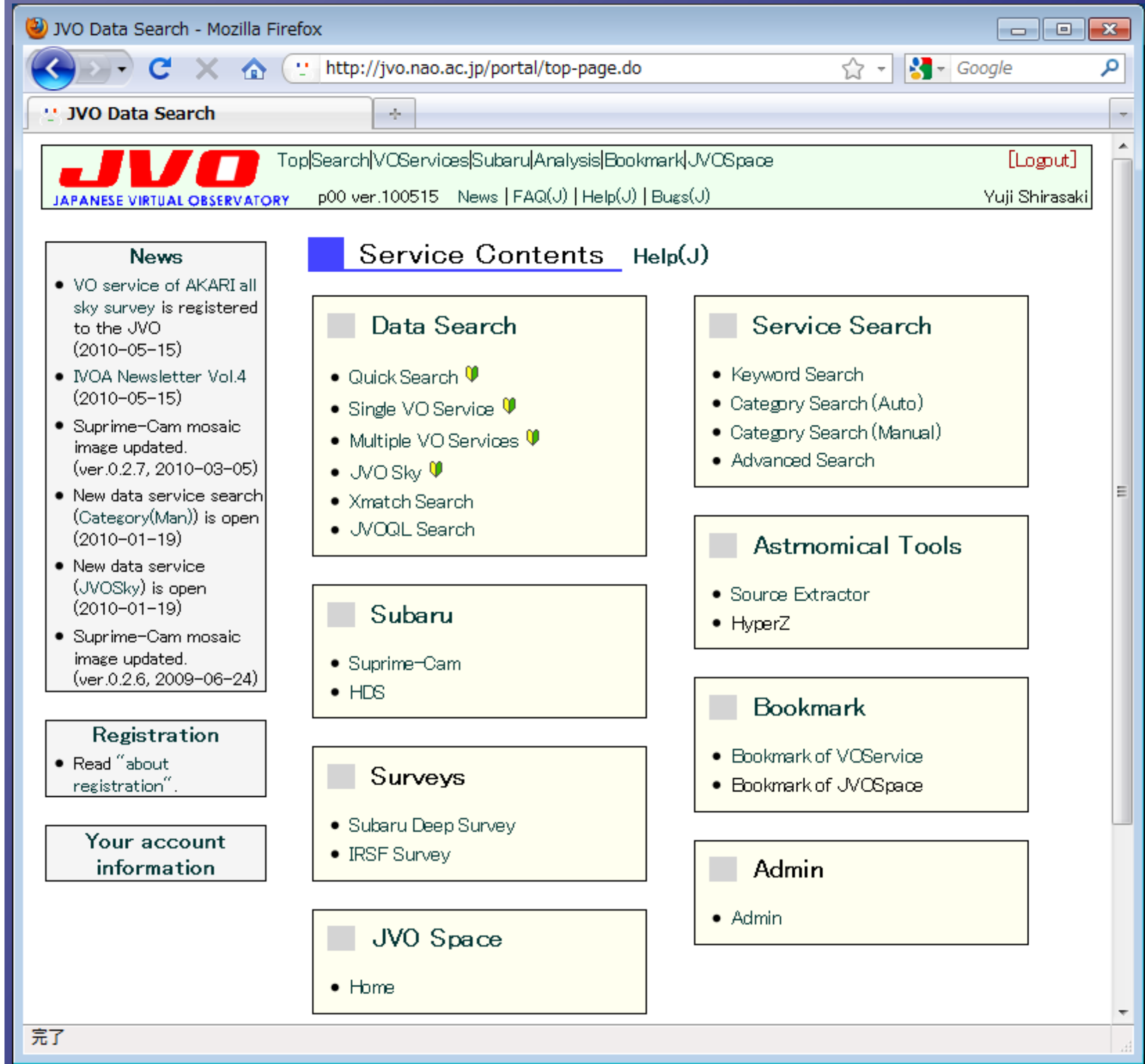
- ✓ Cross-Query to multiple datasets
  - Query multi-wavelength data in a single query within (an) interested sky region(s)
  - Easier data discoveries
  - Easier to utilize multiple DBs

Possible to conduct automated data reductions of huge data

→ makes it possible to conduct (effectively) impossible research in the past

JVO portal

http://jvo.nao.ac.jp/portal

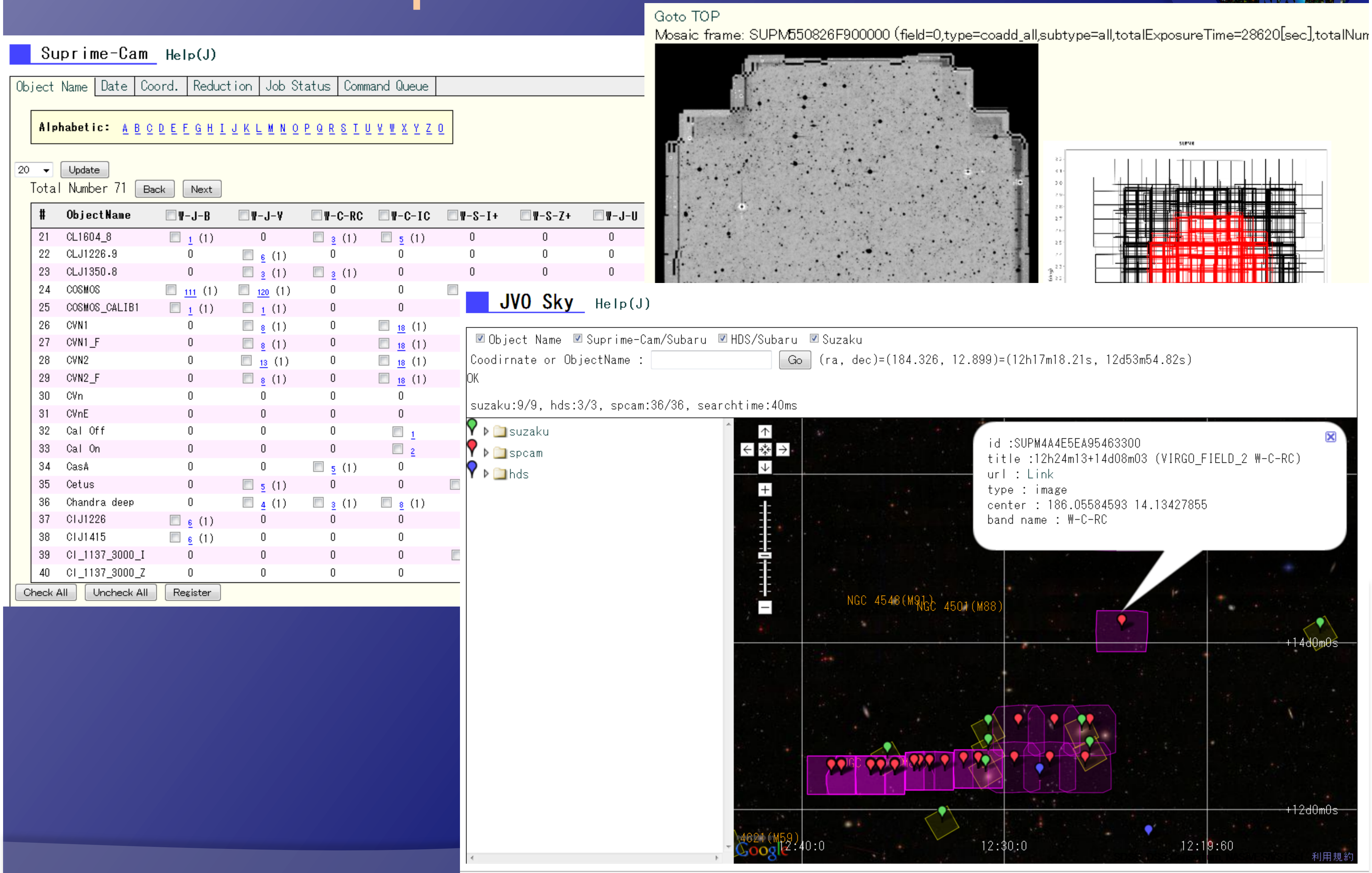


- ✓ 10,551 Data Resources
  - 7,397 Catalogs
  - 208 Image Services
  - 84 Spectrum Services
  - …
- ✓ Reduced Subaru Data
  - Suprime-Cam
  - HDS

JVO Subaru archive

- ✓ Suprime-Cam data reduction system
  - Data archive and parallel computing system are connected with a dedicated network (128Gbps)
  - The whole data can be processed in two weeks (using 48 CPU cores)
- ✓ VO access as well as a dedicated GUI
  - Data retrieval is “programmable”
  - Possible to retrieve cutout image for specified region
  - No need to download all the data (~10TB), data can be retrieved on demand

GUI for Suprime-Cam archive



#	ObjectName	W-J-B	W-J-V	W-C-RC	W-C-IC	W-S-I+	W-S-Z+	W-J-U
21	CL1604_8	1 (1)	0	3 (1)	5 (1)	0	0	0
22	CLJ1226.9	0	6 (1)	0	0	0	0	0
23	CLJ1350.8	0	3 (1)	3 (1)	0	0	0	0
24	COSMOS	111 (1)	120 (1)	0	0			
25	COSMOS_CALIB1	1 (1)	1 (1)	0	0			
26	CVN1	0	8 (1)	0	18 (1)			
27	CVN1_F	0	8 (1)	0	18 (1)			
28	CVN2	0	13 (1)	0	18 (1)			
29	CVN2_F	0	8 (1)	0	18 (1)			
30	CVn	0	0	0	0			
31	CVnE	0	0	0	0			
32	Cal Off	0	0	0	1			
33	Cal On	0	0	0	2			
34	CasA	0	0	5 (1)	0			
35	Cetus	0	5 (1)	0	0			
36	Chandra deep	0	4 (1)	3 (1)	8 (1)			
37	ClJ1226	6 (1)	0	0	0			
38	ClJ1415	6 (1)	0	0	0			
39	Cl_1137_3000_I	0	0	0	0			
40	Cl_1137_3000_Z	0	0	0	0			

"Early Science Result from the Japanese Virtual Observatory: AGN and Galaxy Clustering at z = 0.3 to 3.0"

Y.Shirasaki et al., PASJ, Vol.63, No.SP2, 469–491.

- ✓ Measurement of AGN-Galaxy cross-correlation
  - Fueling mechanism of AGN
  - Co-evolution of galaxy and black hole
  - Use all the data of Suprime-Cam archive (nobody have done !)
- ✓ Previous works
  - Redshift measurement → 3D cross-correlation
  - $z < 0.6$: Good statistics using SDSS data
  - $z > 0.6$: Relatively poor stat. (several tens).
    - Hard to improve statistics:
    - Statistics at small scale (~1Mpc) is extremely poor
    - Affected by Cosmic Variance ← small number of samples
    - Biased to red galaxies in spectroscopic target selection

Dataset

AGN samples

- ✓Veron QSO/AGN Catalog (12th ed)
- ✓SDSS DR-5 QSO Catalog (4th ed)

Galaxy samples

- ✓JVO Suprime-Cam Archive (B, V, R, I, i', z' bands)
- ✓UKIDSS DR-2 Catalog (K band)
- ✓Deepest observation data was used for each AGN

Suprime-Cam	484
UKIDSS	1325
Total	1809

Search Suprime-Cam image around AGN

JAPANESE VIRTUAL OBSERVATORY

Top|Search|VOServices|Subaru|Analysis|Bookmark|JVOSpace [Logout]

p01 ver.100925 News | FAQ(J) | Help(J) | Bugs(J) Yuji Shirasaki

=> Location: Top Page > Search > JVOQL Search

Input JVOQL

```
SELECT  qso.*, img.*
FROM    ivo://jvo/vizier/VII/235:qso_veron_2006 AS qso,
        ivo://jvo/subaru/spcam:image_cutout AS img
WHERE   qso.z >= 1.0 and qso.z < 1.1
        AND img.region = Circle(qso.raj2000, qso.dej2000, 0.14)
```

Similar to SQL (Structured Query Language)

Submit | Genarate JVOQL | Clear

Service | Table | Region | Criteria | Samples

Search Result

JAPANESE VIRTUAL OBSERVATORY

Top | Search | VOServices | Subaru | Analysis | Bookmark | JVOSpace [Logout]

p01 ver.100925 News | FAQ(J) | Help(J) | Bugs(J) Yuji Shirasaki

=> Location: Top Page > VOTable Viewer

Save/Download | Filter | Metadata | Graphic | Add Column | Appearance

Total 5390 records page: 1 go << < > >>

Alias Name		C42	C31	C43	C23	C30
check	download	QSO.NAME sort	QSO.RAJ2000 sort	QSO.DEJ2000 sort	IMG.IMAGE_TITLE sort	IMG.ACCESS_REF sort
☐	Download	Q J02399-0134	02 39 56.6	-01 34 27	A370 (W-C-RC)	Link
☐	Download	Q J02399-0134	02 39 56.6	-01 34 27	A370-new (W-S-Z+)	Link
☐	Download	Q J02399-0134	02 39 56.6	-01 34 27	A370-wide (W-S-Z+)	Link
☐	Download	Q J02399-0134	02 39 56.6	-01 34 27	A370 (W-S-Z+)	Link
☐	Download	TEX 2152+172	21 54 39.9	+17 27 39	A2390 (W-S-I+)	Link
☐	Download	SDSS J17110+6400	17 11 05.3	+64 00 14	A2255 (W-C-RC)	Link
☐	Download	SDSS J14022+0308	14 02 14.4	+03 08 12	A1835 (W-S-I+)	Link
☐	Download	SDSS J09570+0238	09 57 01.6	+02 38 57	COSMOS (W-J-B)	Link
☐	Download	SDSS J09589+0213	09 58 57.3	+02 13 14	COSMOS (W-J-B)	Link
☐	Download	SDSS J09597+0247	09 59 46.0	+02 47 43	COSMOS (W-J-B)	Link
check	download	QSO.NAME	QSO.RAJ2000	QSO.DEJ2000	IMG.IMAGE_TITLE	IMG.ACCESS_REF
☐	Download	SDSS J09567+0205	09 56 42.3	+02 05 53	COSMOS (W-S-Z+)	Link
☐	Download	SDSS J09589+0213	09 58 57.3	+02 13 14	COSMOS (W-S-Z+)	Link
☐	Download	2QZ J095958+0108	09 59 58.2	+01 08 47	COSMOS (W-S-Z+)	Link
☐	Download	SDSS J09589+0213	09 58 57.3	+02 13 14	COSMOS (W-S-Z+)	Link
☐	Download	SDSS J09589+0213	09 58 57.3	+02 13 14	COSMOS (W-J-V)	Link

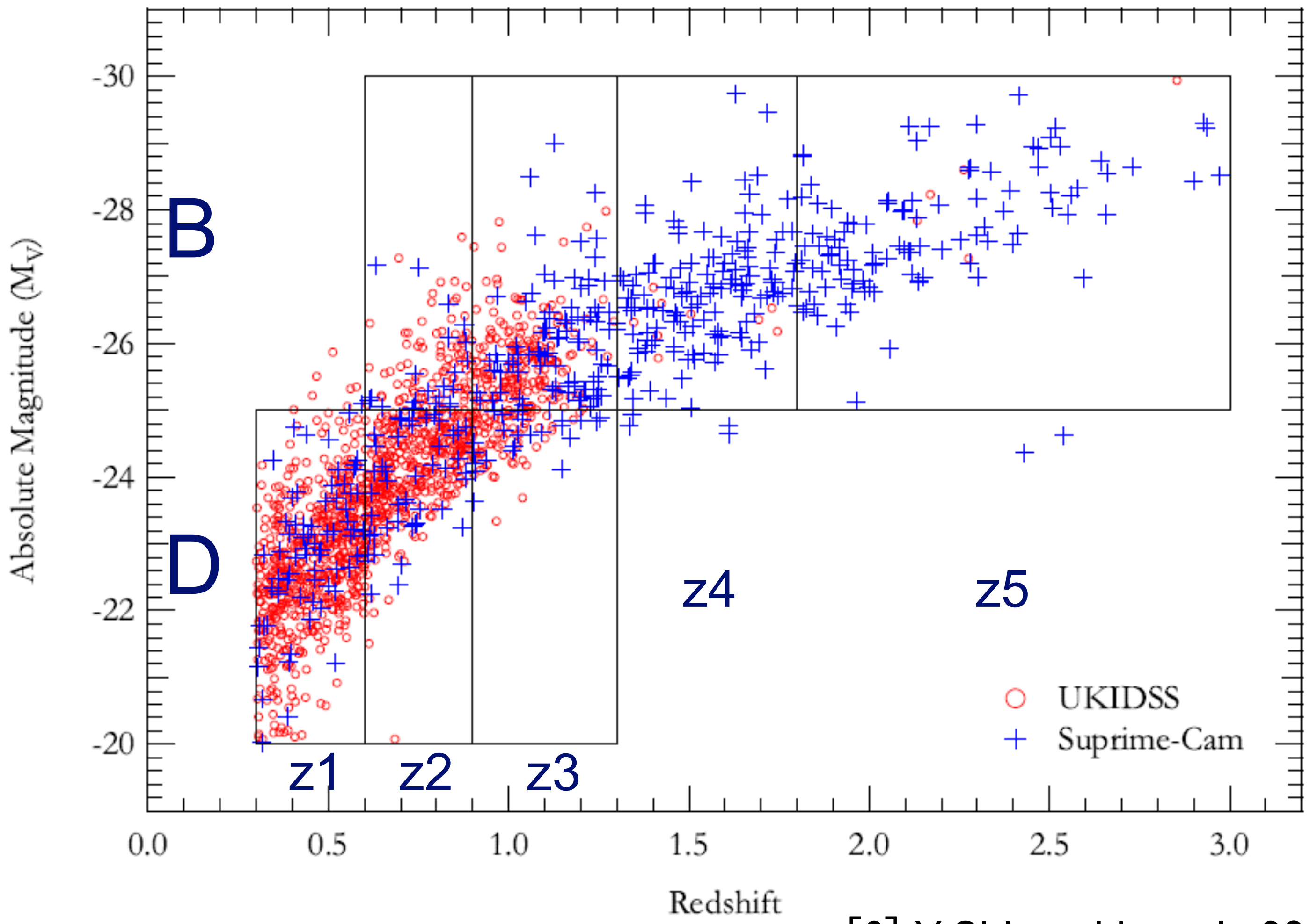
- ✓ Only a part of data is displayed
- ✓ No way to download all the images at once
- ✓ Download coordinates in CSV format

- ✓ Create a list of AGNs observed with Suprime-Cam

Automate using a script

- ✓ Download and analysis for ~12,000 AGNs
  - – hard or impossible to do manually
- ✓ Make a script (e.g. shell script)
  - – Script to create a catalog from retrieved image data ...
  - – Execute this script for each AGN
  - – 12,000 AGNs → 40 parallels. Completed in one day.
- ✓ Access to VO
  - – Use command line access tool.
  - – Useful for repeating the same query by changing a query condition.

AGN redshift & absolute mag



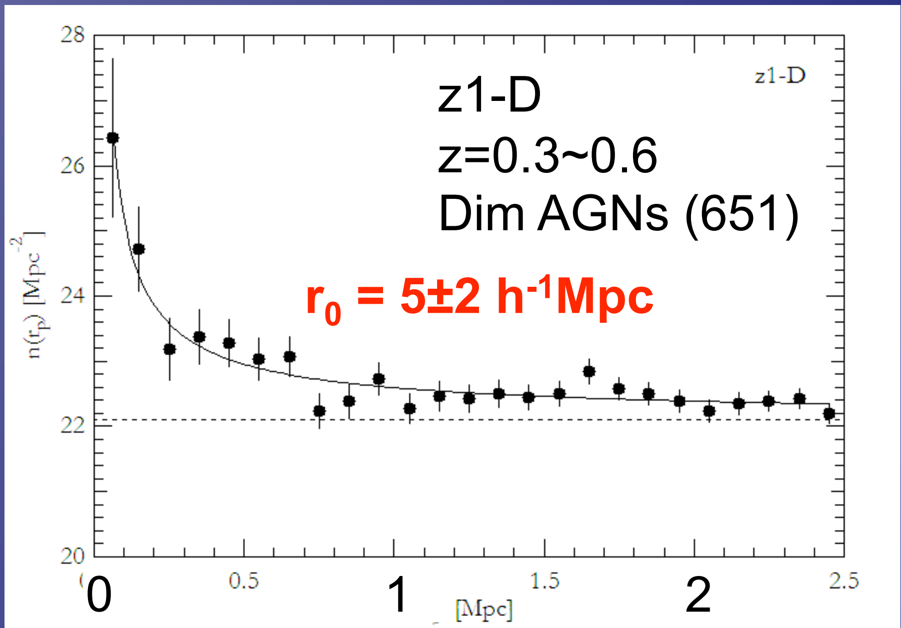
[6] Y.Shirasaki et al., 2011

Fig. 5. K-corrected V band absolute magnitude vs redshift of the AGNs used in this work. Open circles represent AGN samples for which the galaxy sample is derived from the UKIDSS data, and the crosses represent AGN samples for which the galaxy sample is derived from the Suprime-Cam data.

Analysis

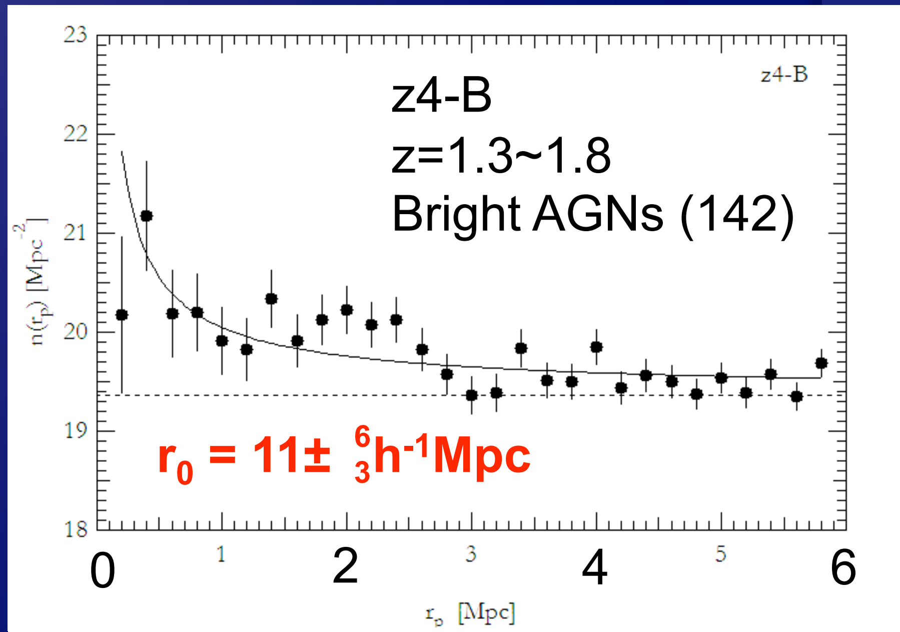
✓ Density of galaxies within r_p : $n(r_p)$; γ=1.8

$$< n(r_p) >= r_p \left(\frac{r_0}{r_p}\right)^{\gamma} \frac{\Gamma(1/2)\Gamma((\gamma-1)/2)}{\Gamma(\gamma/2)} < \rho_0 > + < n_{bg} >$$



Transverse comoving distance (Mpc)

2011 Sep 28



Transverse comoving distance (Mpc)

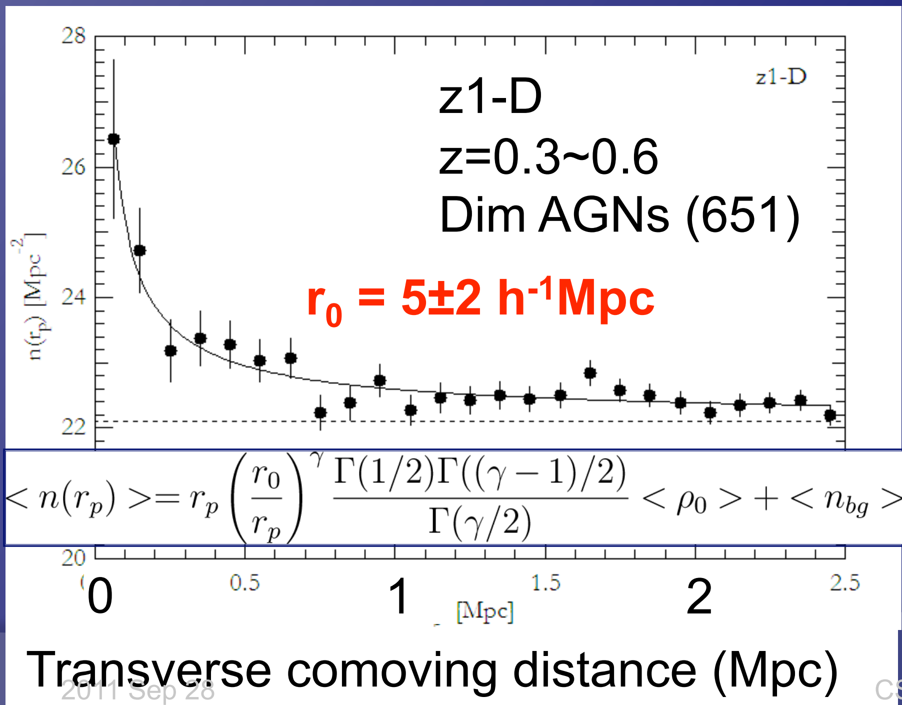
[6] Y.Shirasaki et al., 2011

Analysis

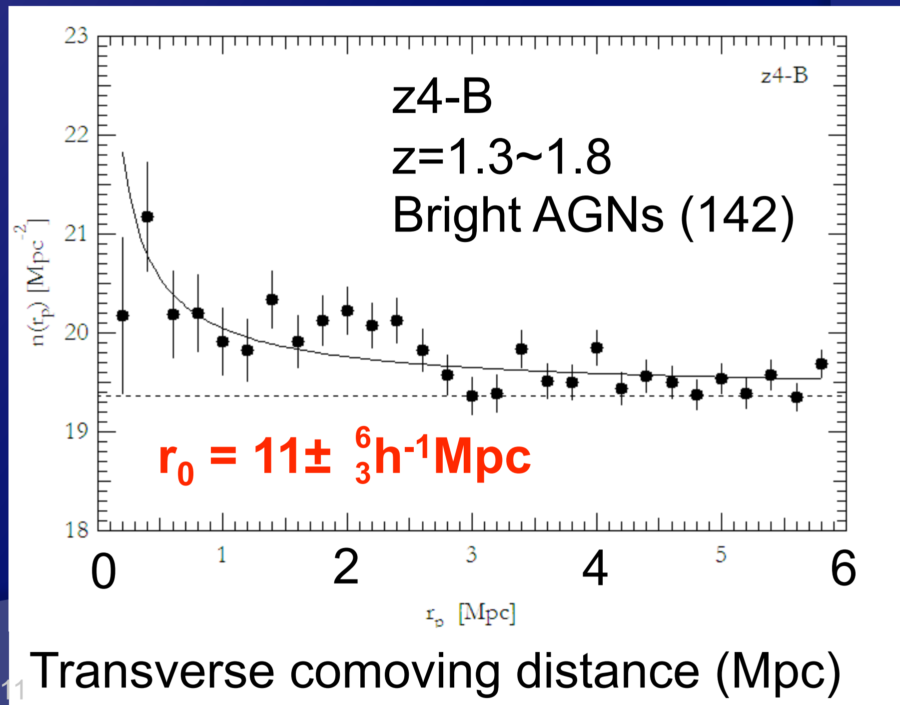
✓ Projected correlation function : $\omega(r_p)$

$$\omega(r_p) = 2\int_0^\infty \xi(r_p,\pi)d\pi = r_p\left(\frac{r_0}{r_p}\right)^\gamma \frac{\Gamma(1/2)\Gamma((\gamma-1)/2)}{\Gamma(\gamma/2)}$$

$$\boxed{\xi(r) = (r_0/r)^\gamma.} \qquad = \frac{1}{\rho_0}\int_{-\infty}^{\infty}(\rho(r)-\rho_0)d\pi = \frac{n(r_p)-n_{\rm bg}}{\rho_0}$$

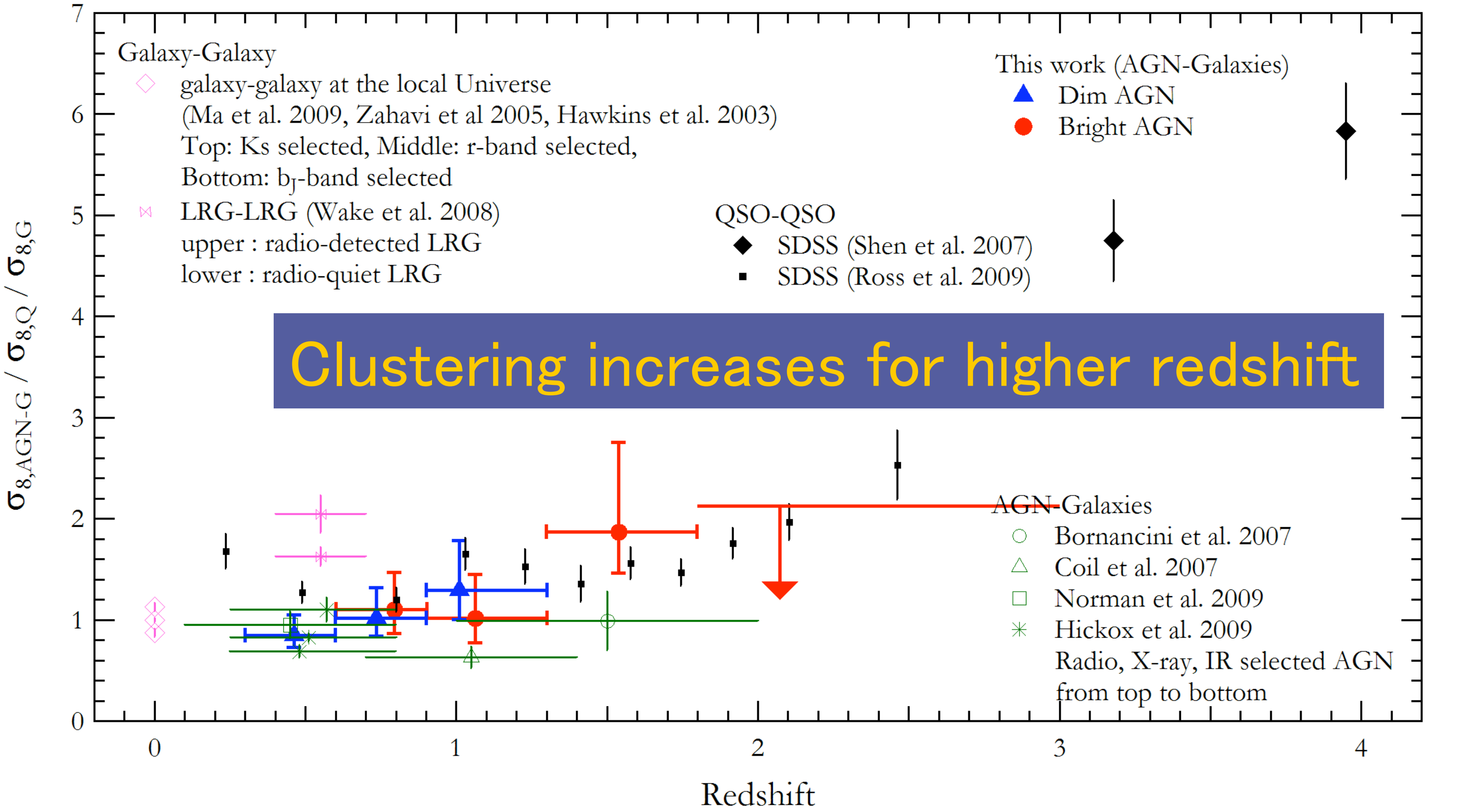


$$< n(r_p) >= r_p\left(\frac{r_0}{r_p}\right)^\gamma \frac{\Gamma(1/2)\Gamma((\gamma-1)/2)}{\Gamma(\gamma/2)} < \rho_0 > + < n_{bg} >$$



Result

σ_8 : rms of correlation function at < 8 Mpc



Data Science in Other Fields

Virtual Observatories in Planetary Sciences

- ease discovery, access and use of planetary data
- NASA, ESA, JAXA and others
- Refers to the IVOA standard protocols, w/ some modifications

http://planetarydata.org/

Successful Models on Data Sharing

1. Protein Data Banks (PDB)
2. OneGeology/CGI model
3. Intergovernmental Panel of Climate Change (IPCC)
4. International Virtual Observatory (IVOA)

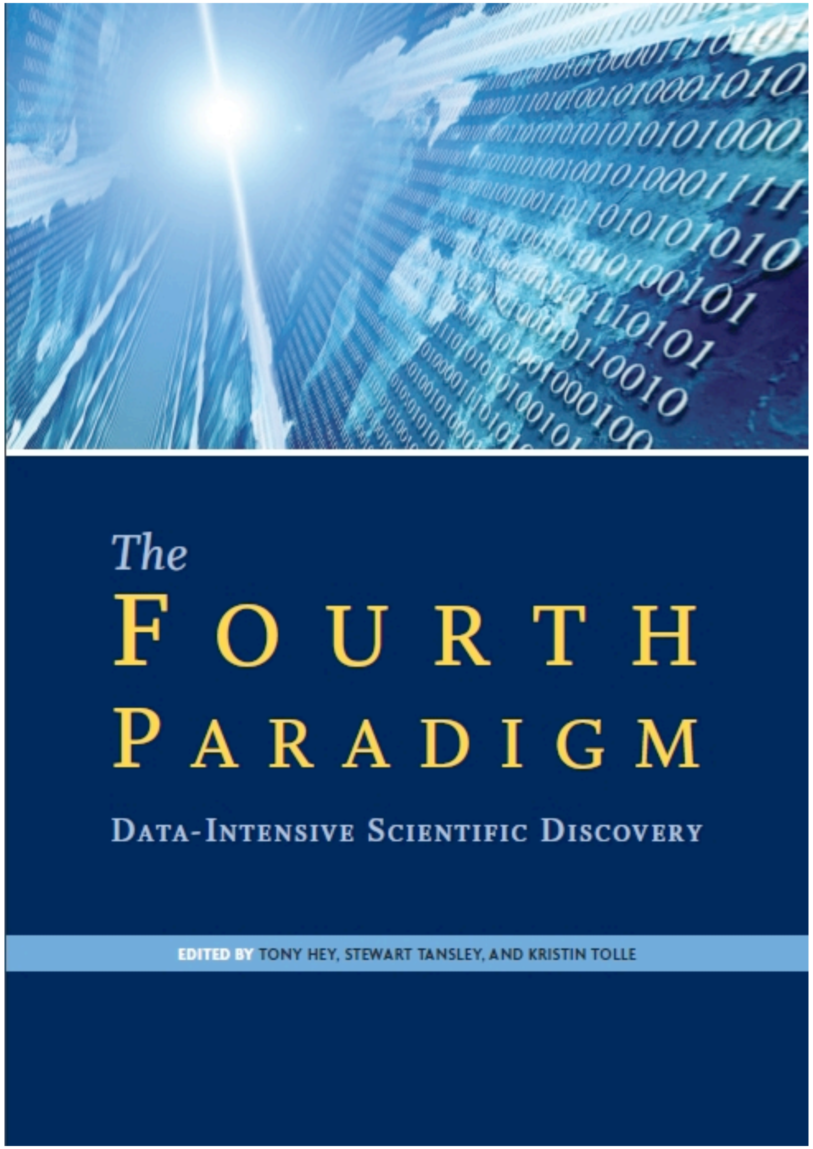
World Data System

- “Virtual Observatories” in a variety of science fields
  = advanced interconnections between data management components for disciplinary and multidisciplinary applications

- Organized under the ICSU (Int’l Council for Sciences)

Data Intensive Science

- Data deluge
  - Huge data size
  - Wide variety
  - Transient data
  - time-domain
- New paradigm in scientific research by introducing data management and advanced data analysis



References

[1] S.G.Djorgovski,2005:Virtual Astronomy, Information Technology, and the New Scientific Methodology,eprint arXiv:astro-ph/0504651,7pp

[2] Brunner, Robert J., Hall, Patrick B., Djorgovski, S. George, Gal, R. R., Mahabal, A. A., Lopes, P. A. A., de Carvalho, R. R., Odewahn, S. C., Castro, S., Thompson, D., Chaffee, F., Darling, J., Desai, V.,2003:Peculiar Broad Absorption Line Quasars Found in The Digitized Palomar Observatory Sky Survey, The Astronomical Journal,126,53-62

[3] Lisa J. Storrie-Lombardi, Arthur M. Wolfe.,2000:Surveys for z > 3 Damped Lyman-alpha Absorption Systems, Astrophys.J,543,552-576

[4] W. A.Santos et al, 2007 : Fossil Groups in the Sloan Digital Sky Survey, ApJ, 134,1551–1559

[5] J. A. Caballero & E. Solano, 2007: ALBUS 1: A Very Bright White Dwarf Candidate, ApJ, 665, L151-L154

[6] Y.Shirasaki et al., 2011: Early Science Result from the Japanese Virtual Observatory: AGN and Galaxy Clustering at z = 0.3 to 3.0,63, PASJ, No.SP2, 469–491

References

JVO(Japan Vatial Observatory)
    http://jvo.nao.ac.jp/portal

IVOA(International Virtual Observatory Alliance)
    http://www.ivoa.net/

IPDA(Interknational Planetary DATA Alliance)
    http://planetarydata.org/

The fourth Paradigm: Data-Intensive Scientific Discovery
    http://research.microsoft.com/en-us/collaboration/fourthparadigm/