On-line information in astronomy

From networking
to
a virtual observatory

Why keeping data in astronomy?

- Long term observation of variable natural phenomena
- Major scientific objectives:
  - Variability, evolution, statistics
  - Observations at different wavelengths, with different techniques

  ⇨ *the physical phenomena at work*

Large projects

- Large ground- and space-based observatories
- Large surveys

⇨ Optimize the scientific return of ‘big science’ by increasing usage

in general 1year proprietary period

journals: 3 years but abstracts and TOC free

A few figures: catalogues (1)

- Hipparcos (2nd cent. BC), Almagest (2nd AD)
- mid-19th, early 20th: the first very large catalogues
  - Bonner Durchmusterung (1859-1862)

    325 000 objects
  - Southern Durchmusterung (1886,…)

    135 000 objects
  - Cape Photographic Durchmusterung (1895-1900)

    455 000 objects

A few figures: catalogues (2)

- HIPPARCOS

  100 000 stars, better positions

- Guide Star Catalogue (GSC)

  1992: 20 000 000 *; 2001: 456 000 000*

- USNO A1, A2

  500 000 000 stars

The very large surveys

- Sloan Digital Sky Survey
  100 000 000 objects, 5 colors, π sr, z~5
- DENIS, 2MASS infrared
  3 colors, 100 000 000 objects

Catalogues and surveys

http://vizier.u-strasbg.fr/viz-bin/VizieR

ISAAC

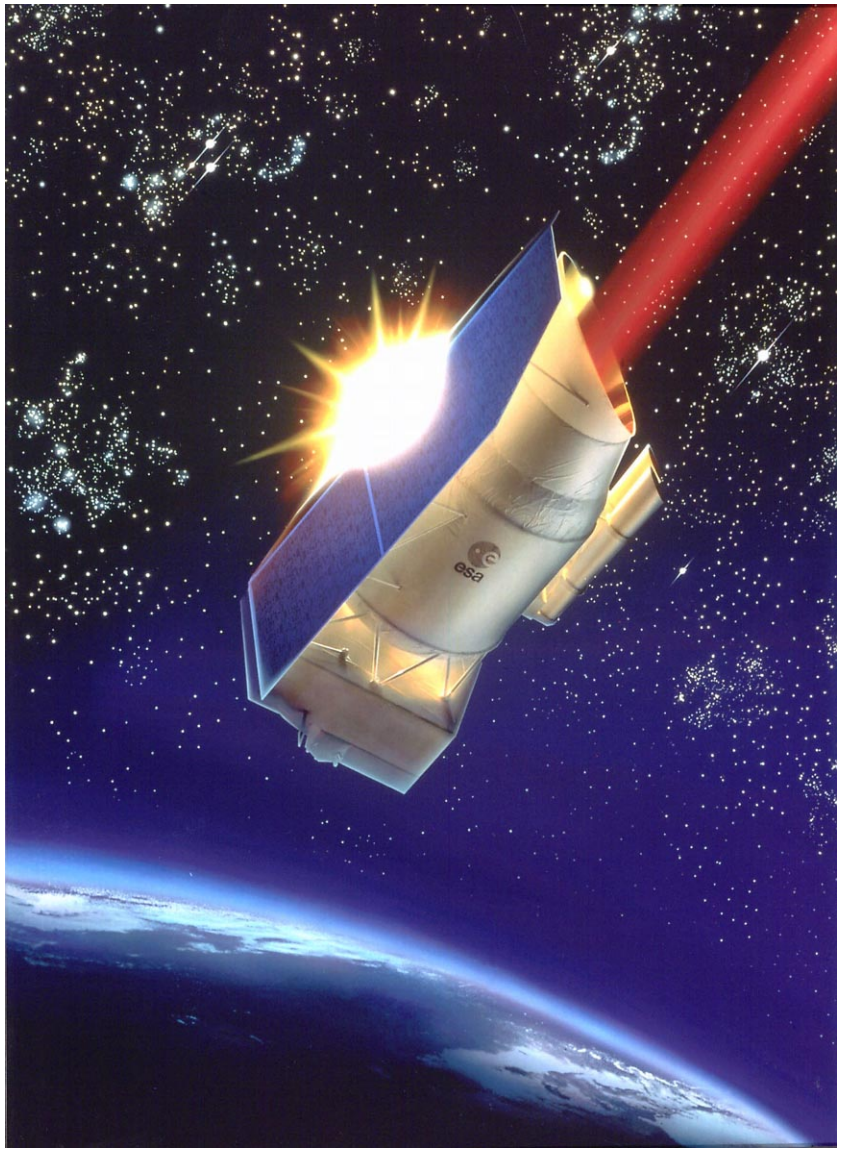
VLT

ISO

0.6m telescope

In a cryostat!

4 instruments



An increasingly complex problem

- Volume and complexity increase
- Distributed, heterogeneous information

  Observations, results published in journals, compilations, software, models, …

- Re-use data

  ⇨ data ⊕ documentation

A major technical revolution

- Increased capacity to store and manage information
- Irruption of the WWW
  - Information easily accessible
  - Data/documentation integration
  - Navigation between distributed, heterogeneous information

Useful and appealing tools

But

- Careful work on contents and functionalities remains mandatory
- Information validation is critical

  producers and users!
- Maintenance of services and links

New constraints

- On Agencies

  Data conservation, diffusion versus instrumentation, operations

- On the scientific community

  Give a sufficient priority in projects, evaluation and prospective

  Resources needed (scientists, engineers)

- On projects

  Available and usable data

  selection, organisation, « project memory »

Networking of information in astronomy

- A small discipline
- Few commercial constraints
- A long term partnership to define exchange standards

Networking of astronomical information

Observations

Ground- and space-based observatories

Surveys

Results

Publications in electronic journals

The actors

- Data producers

  Know instruments and methods

  ⇨ preserve ‘usable’ data and project memory
- Disciplinary centers

  Data and information in a given domain
- Journals

  Results
- Data Centers

Centre de Données astronomiques de Strasbourg

- Created in 1972
  - Electronic data
  - Expertise about data
  - International role
  - Objective: tools for science

Collect, homogenize, preserve, distribute astronomical information for the usage of the whole astronomy community

Centre de Données astronomiques de Strasbourg

- Value-added, reference services
- Selected, validated, homogenized information

  from publications

  reference catalogues

  images + links
- Standards and tools

The CDS hub (1)

Astronomical objects

identification, bibliography, data, measurements

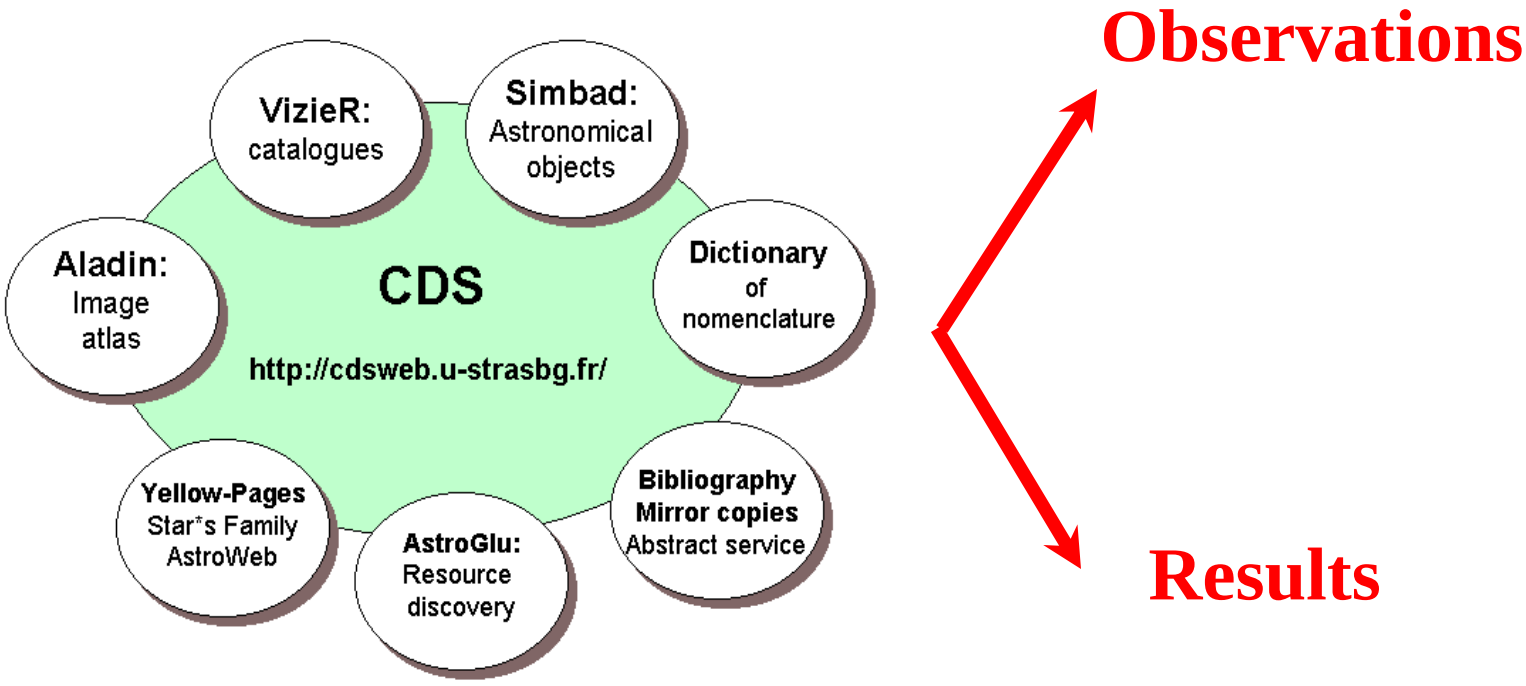
Information Federator

catalogues, published tables observation logs, surveys

Information Integrator

images, databases, catalogues, surveys, archives, *user data*

The CDS hub (2)


VizieR:
catalogues
Simbad:
Astronomical
objects
Aladin:
Image
atlas
CDS
http://cdsweb.u-strasbg.fr/
Dictionary
of
nomenclature
Yellow-Pages
Star*s Family
AstroWeb
AstroGlu:
Resource
discovery
Bibliography
Mirror copies
Abstract service
Observations
Results

Query types

- Search by position

  What is known in a region of the sky

  with links to distributed information

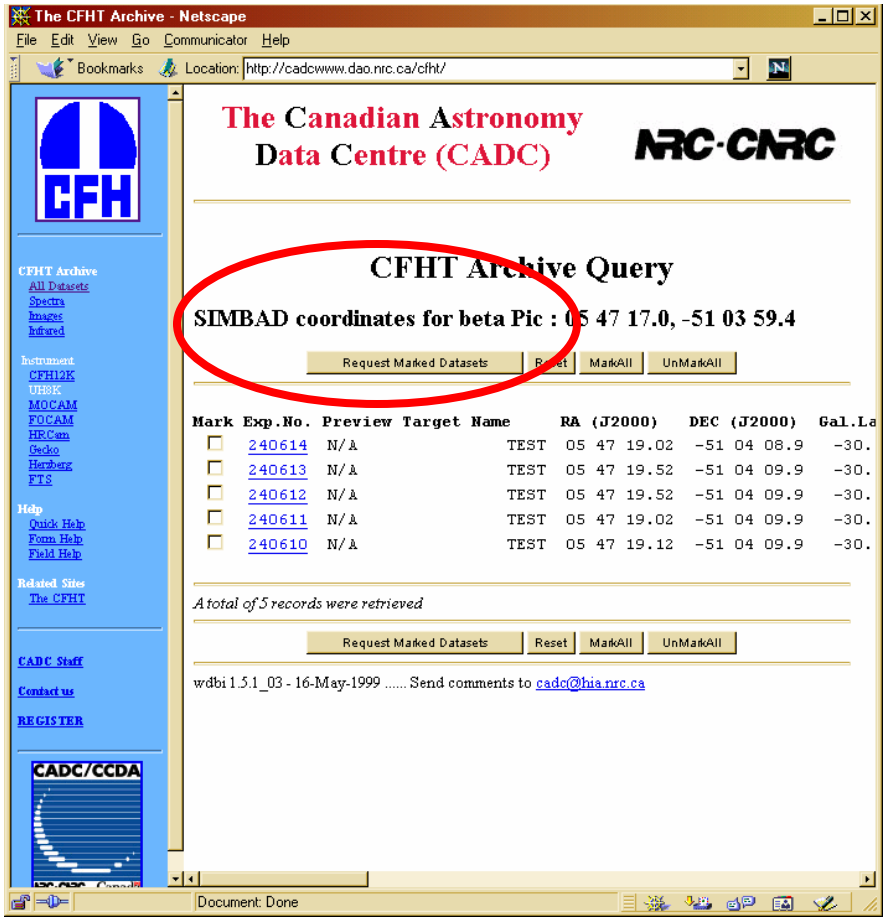
- Search by criteria, e.g.
  - gamma ray burst seen in 1999
  - published lists of objects from Chandra

Early interoperability tools

- **FITS** - description of images, spectra
  - common tools to deal with observations from any telescope
  - maintained by an IAU WG
- NED and SIMBAD name resolvers

Acces to observatory archives: by position

Object name > position: Using SIMBAD



Networking
The astronomy bibliographic network

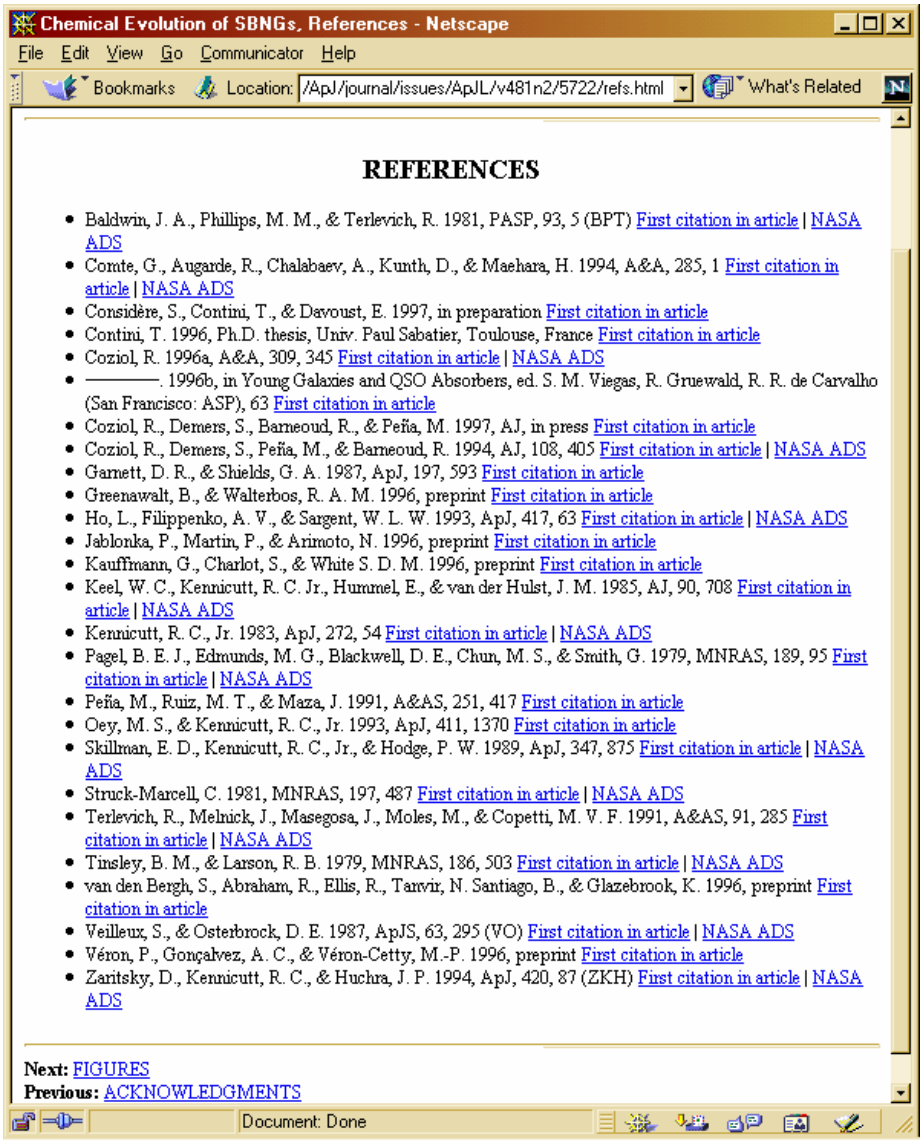
Links between

- electronic journals (common keyword schema)
- the ADS bibliographic database
- on-line services (SIMBAD, NED)
- archival data

bibcode e.g. 1999A&A...351.1003G

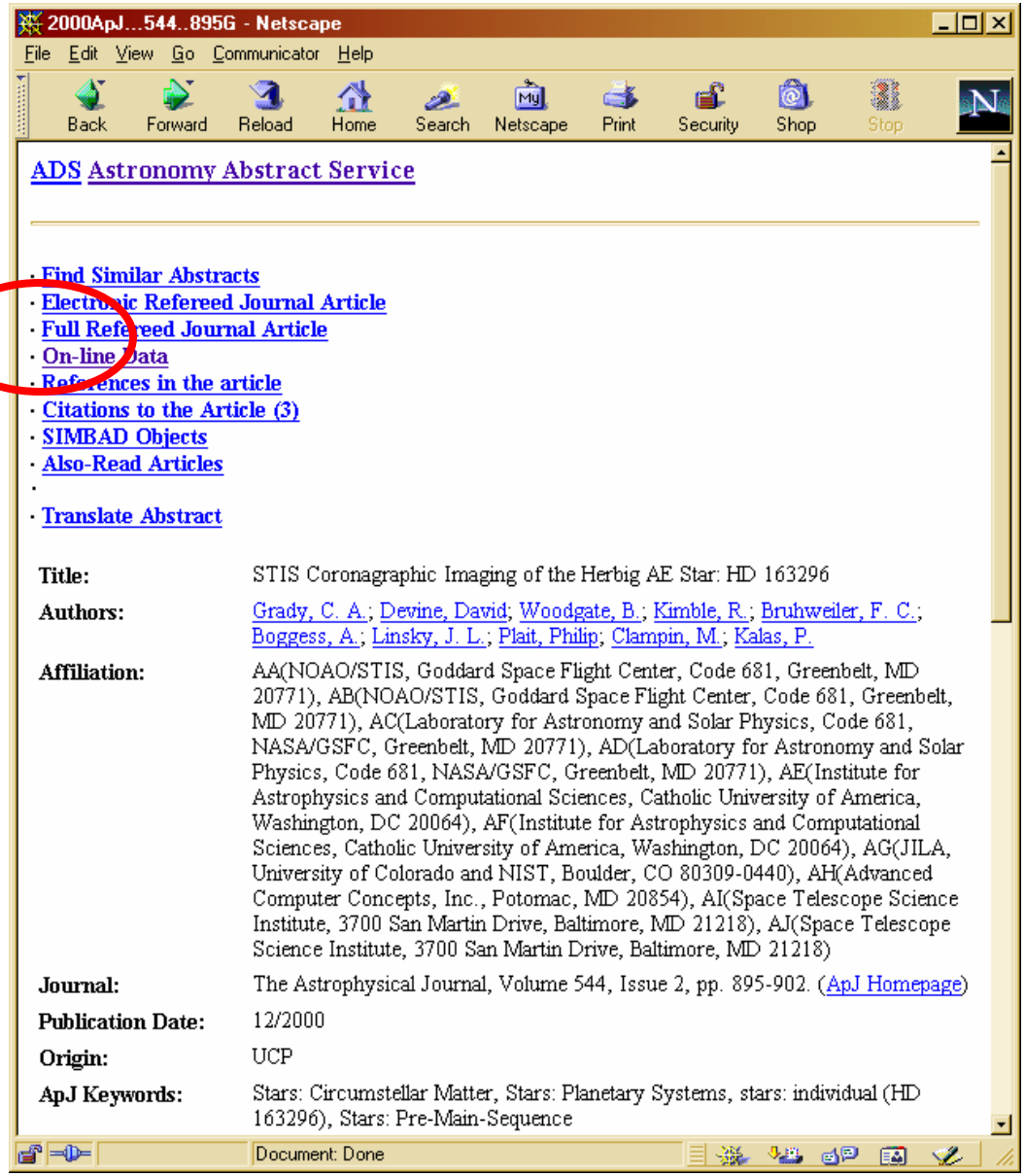
References in an on-line article

- links to ADS
- quality check

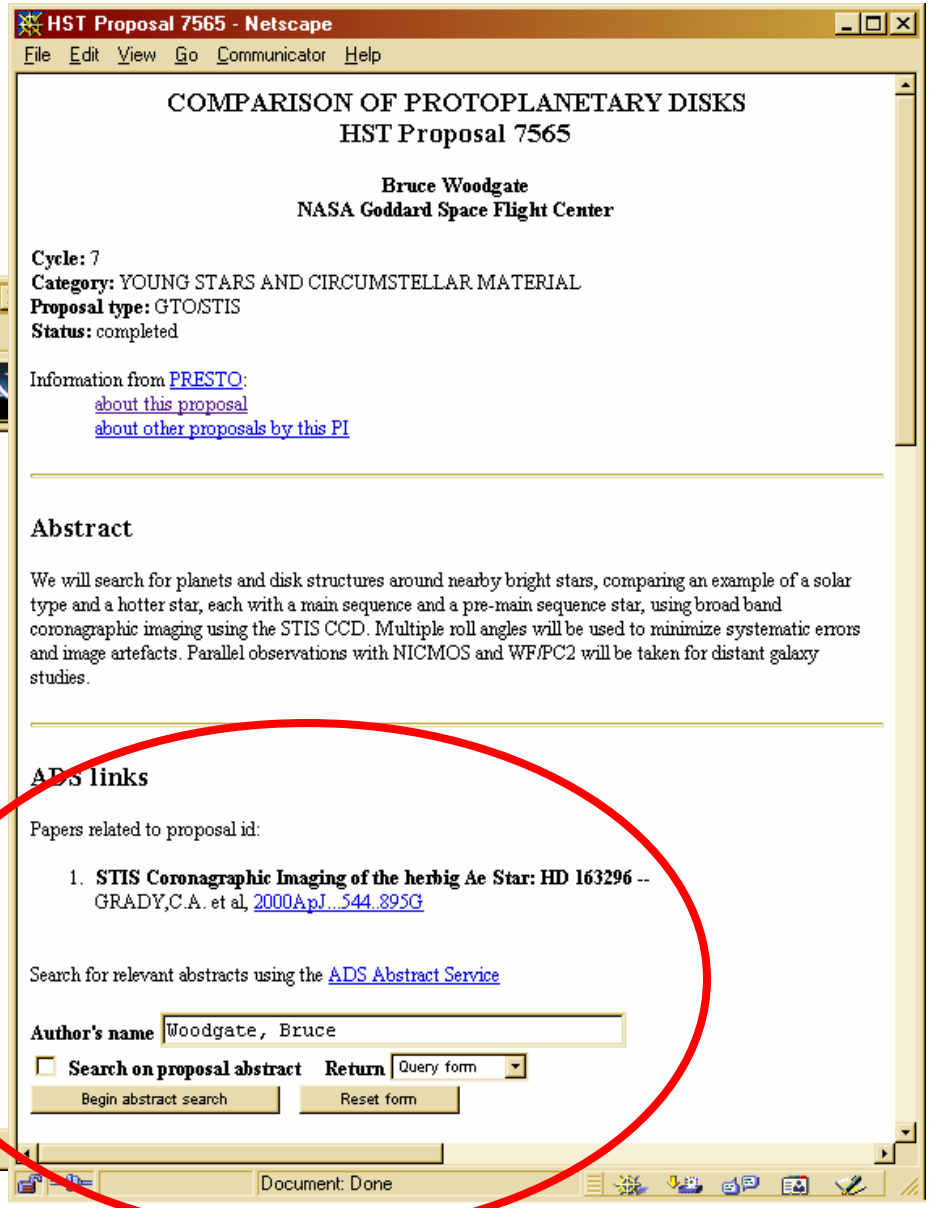


ADS

- *Abstracts*
- *Scanned articles*
- *Links to original on-line paper*
- *Links to other, distributed information e.g. original observations*
- *Also-read articles*

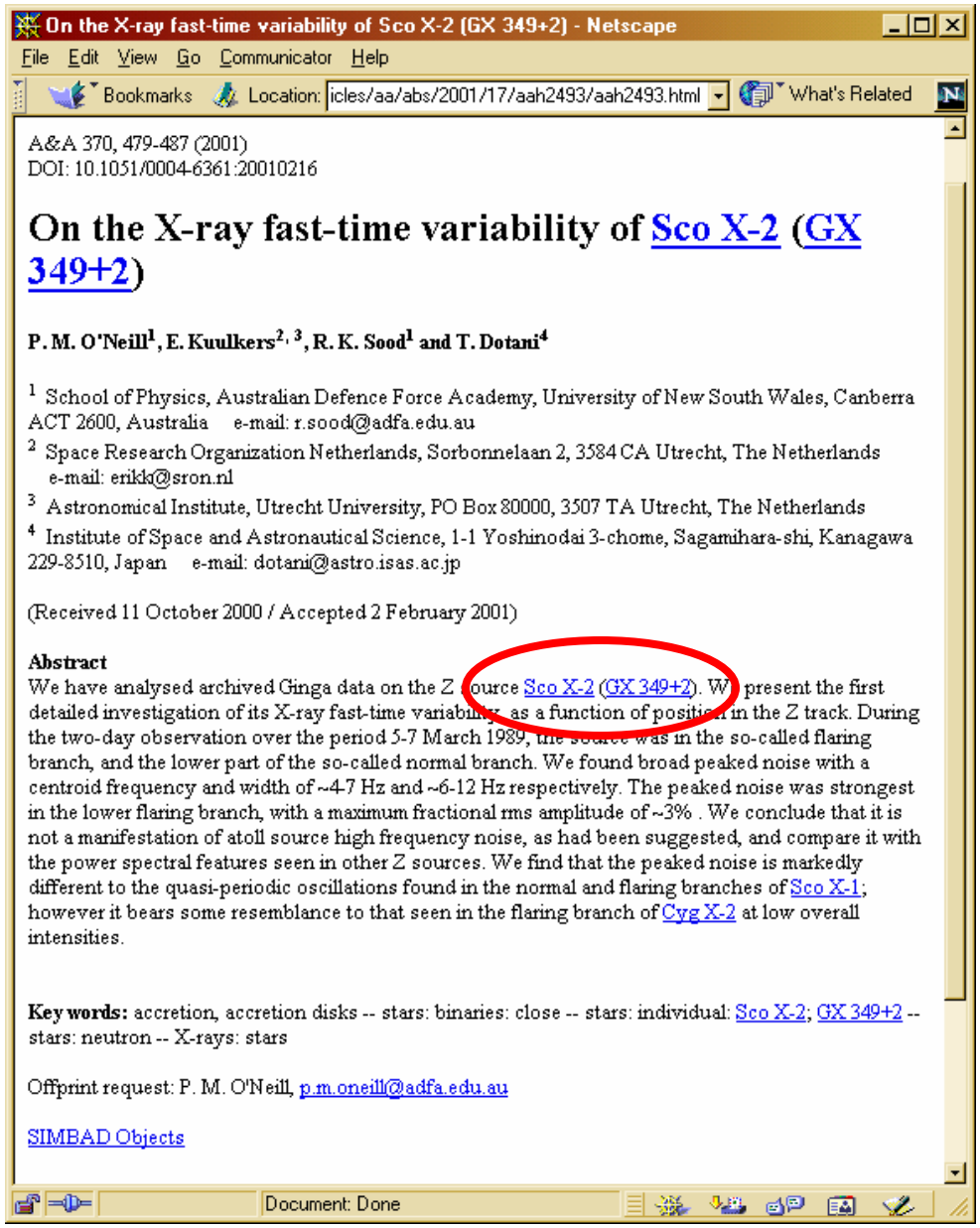


HST archive
From observation
To publication

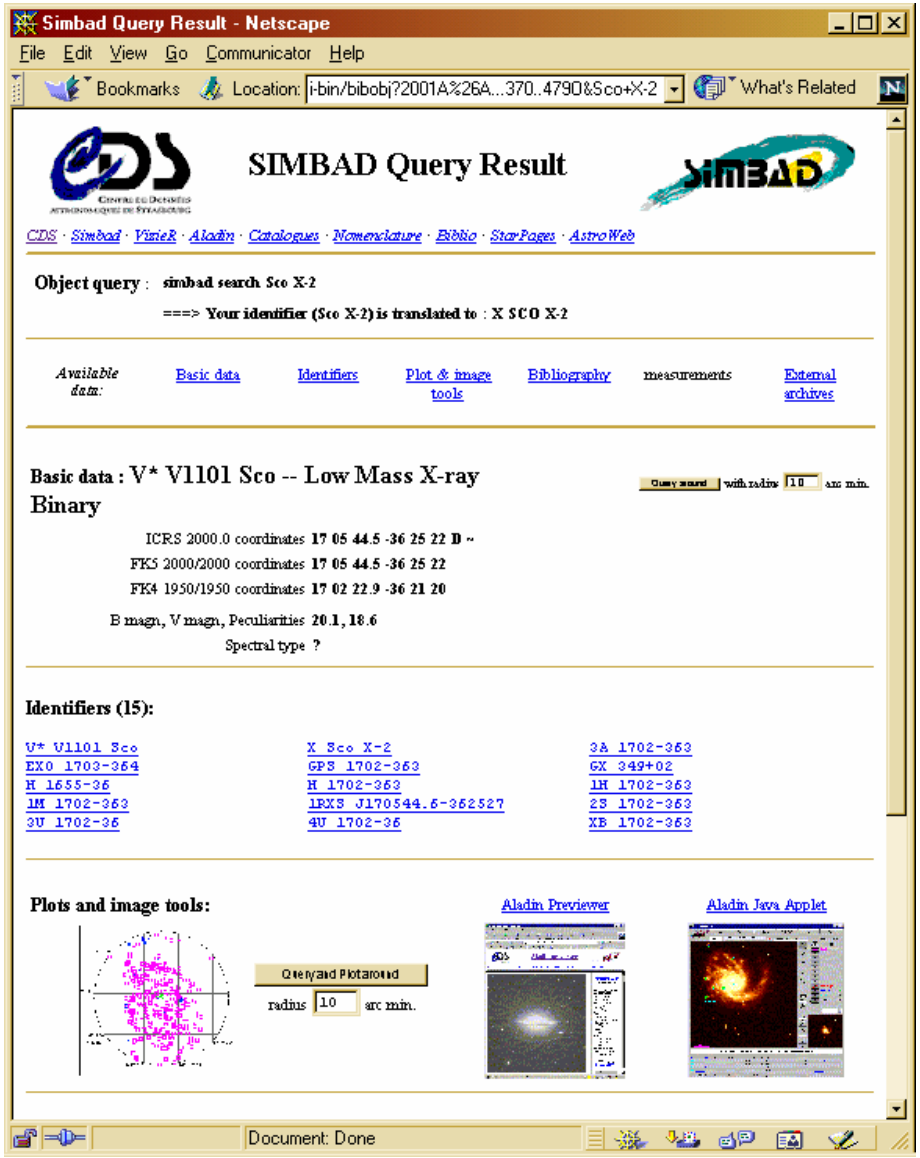


A&A on line

From publication to SIMBAD and more



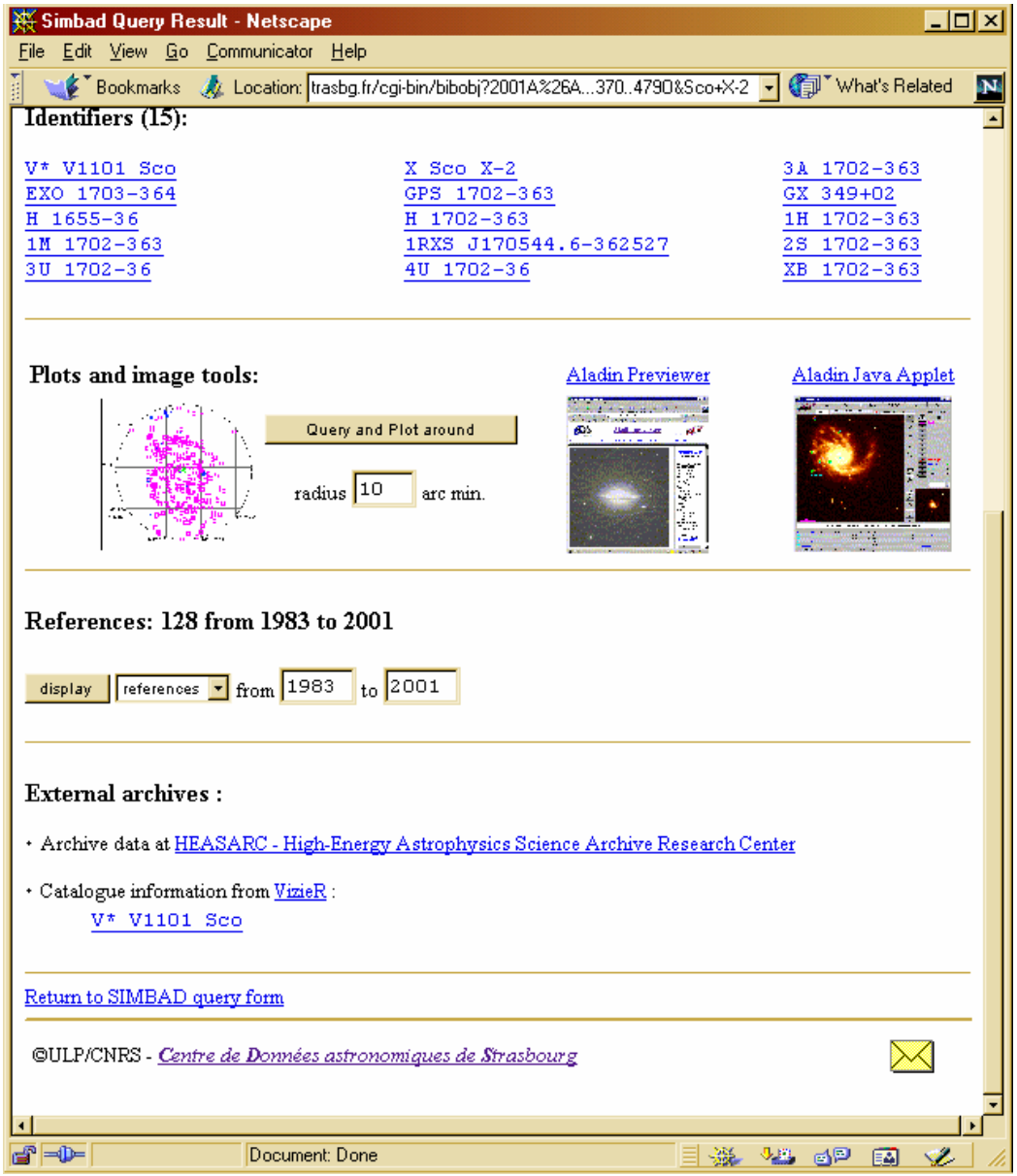
Simbad
Information
About the object



Links

To high energy observations in HEASARC,

To Catalogues



Lessons learnt (1)

- *De facto* standard
- Bottom-up approach: Cooperation between all the actors + snowball effect

  Journals, ADS, data centers, archive centers
- Easy-to-build link

  but contents / validation are fundamental

Lessons learnt (2)

- Standard discussed long before Internet, because two data bases needed to exchange bibliographic information
- Bibcode is better adapted to ‘classical’ publication in journals, human readable
- Gateway to other bibliographical standards (e.g. DOI): correspondance table

Data federation
Tabular data in astronomy

A common description for tabular data

- Reference catalogues
- Published tables
- Large surveys
- Catalogues of observations in archives

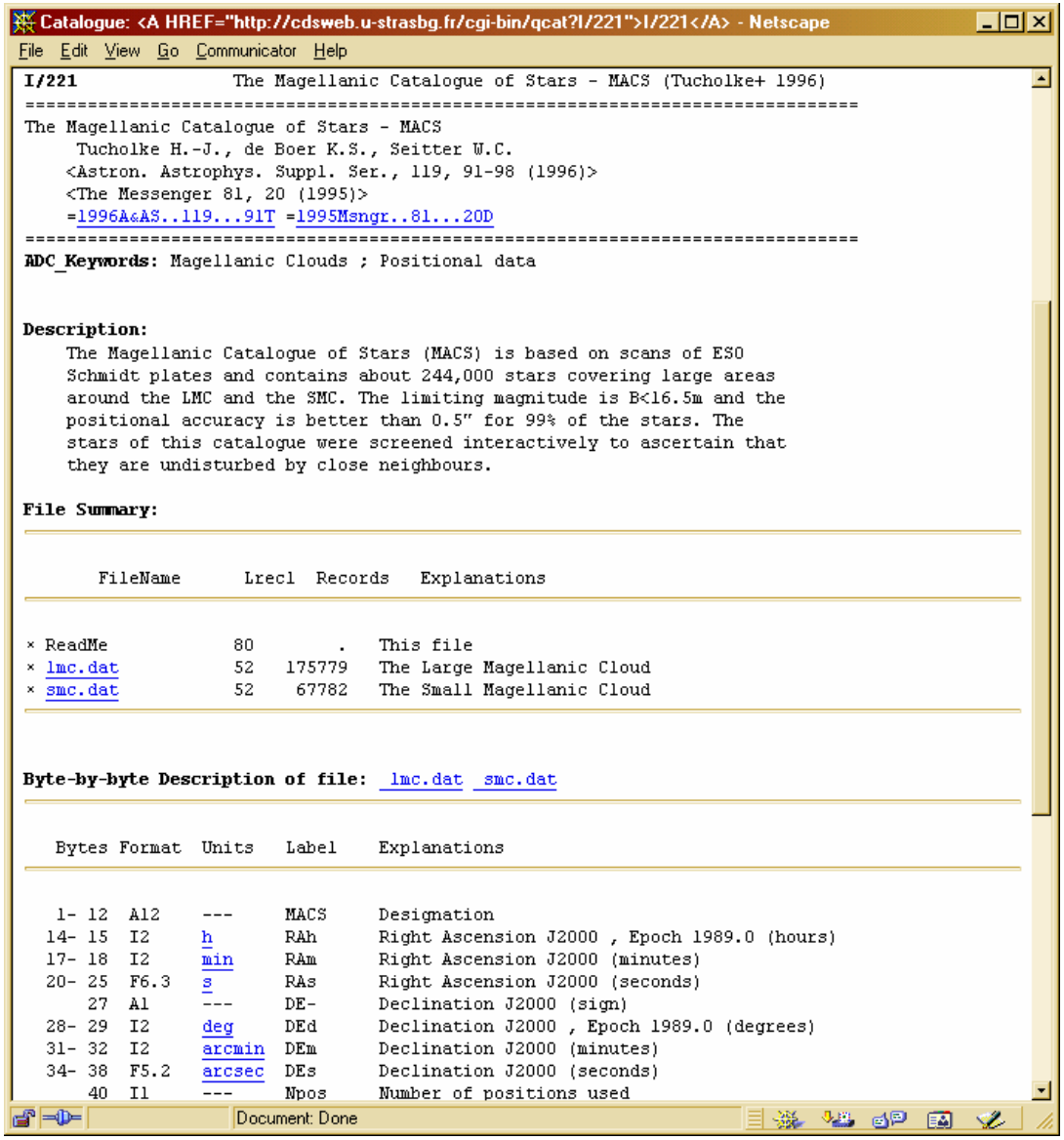
ReadMe

physical organization ⇔ *contents*

ReadMe

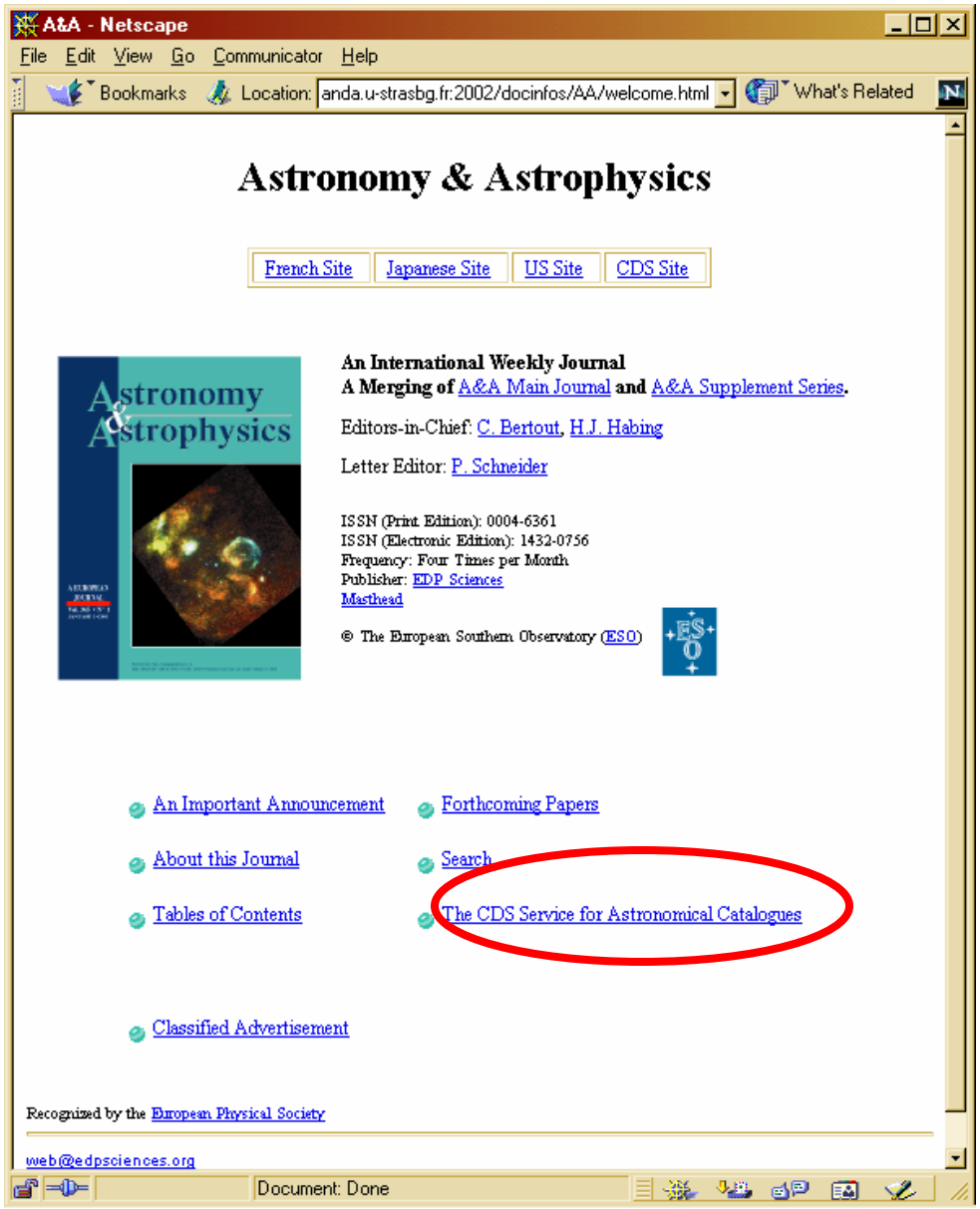
For published tables:
check of
data homogeneity
before publication

sent by authors and checked or prepared by editorial office

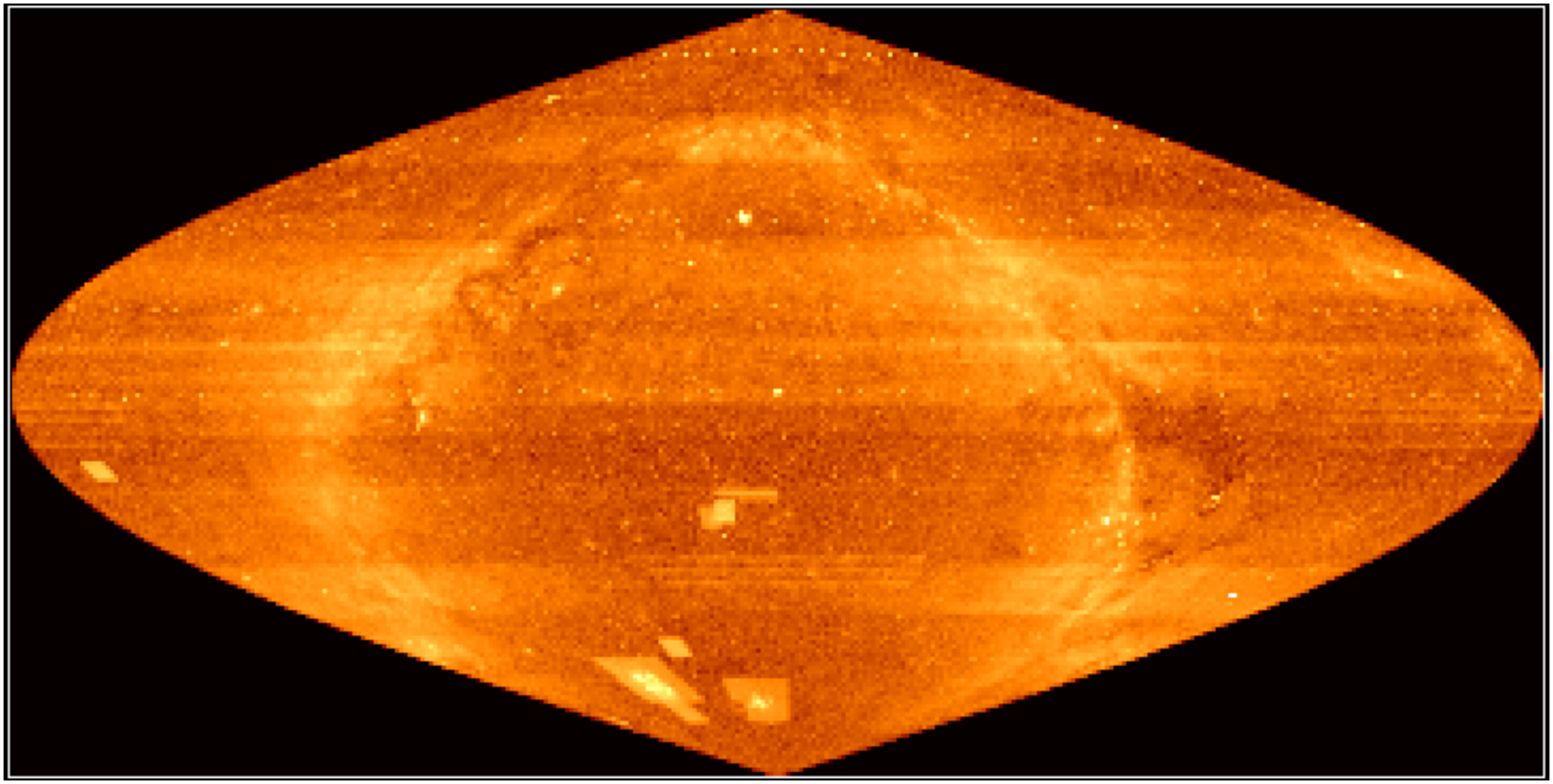


Tables published in articles are usable data!

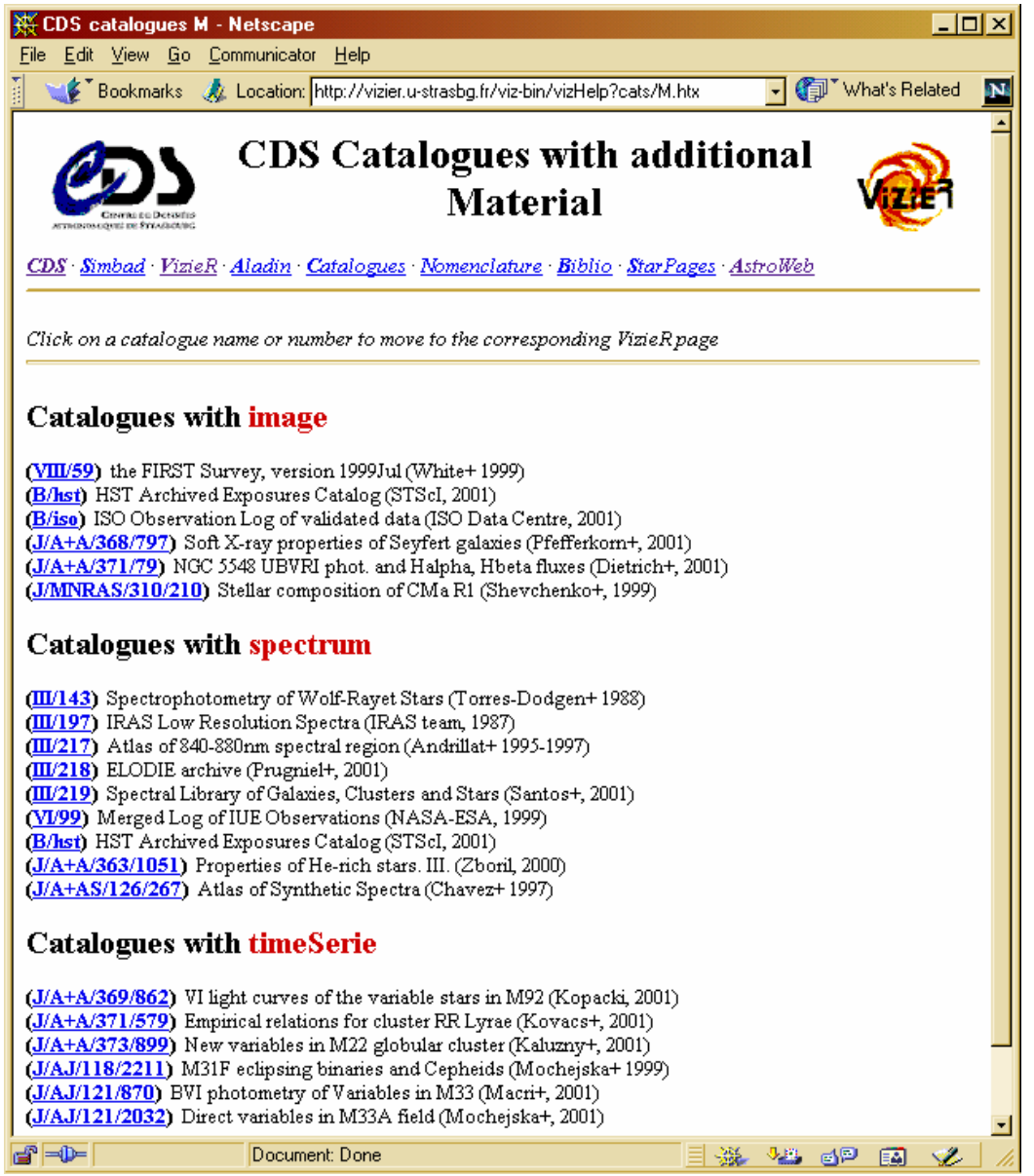
Narrow collaboration between CDS and journals



The VizieR ‘tabular’ data mine



Links to observational data (images, spectra, time series), often distributed in observatory archives

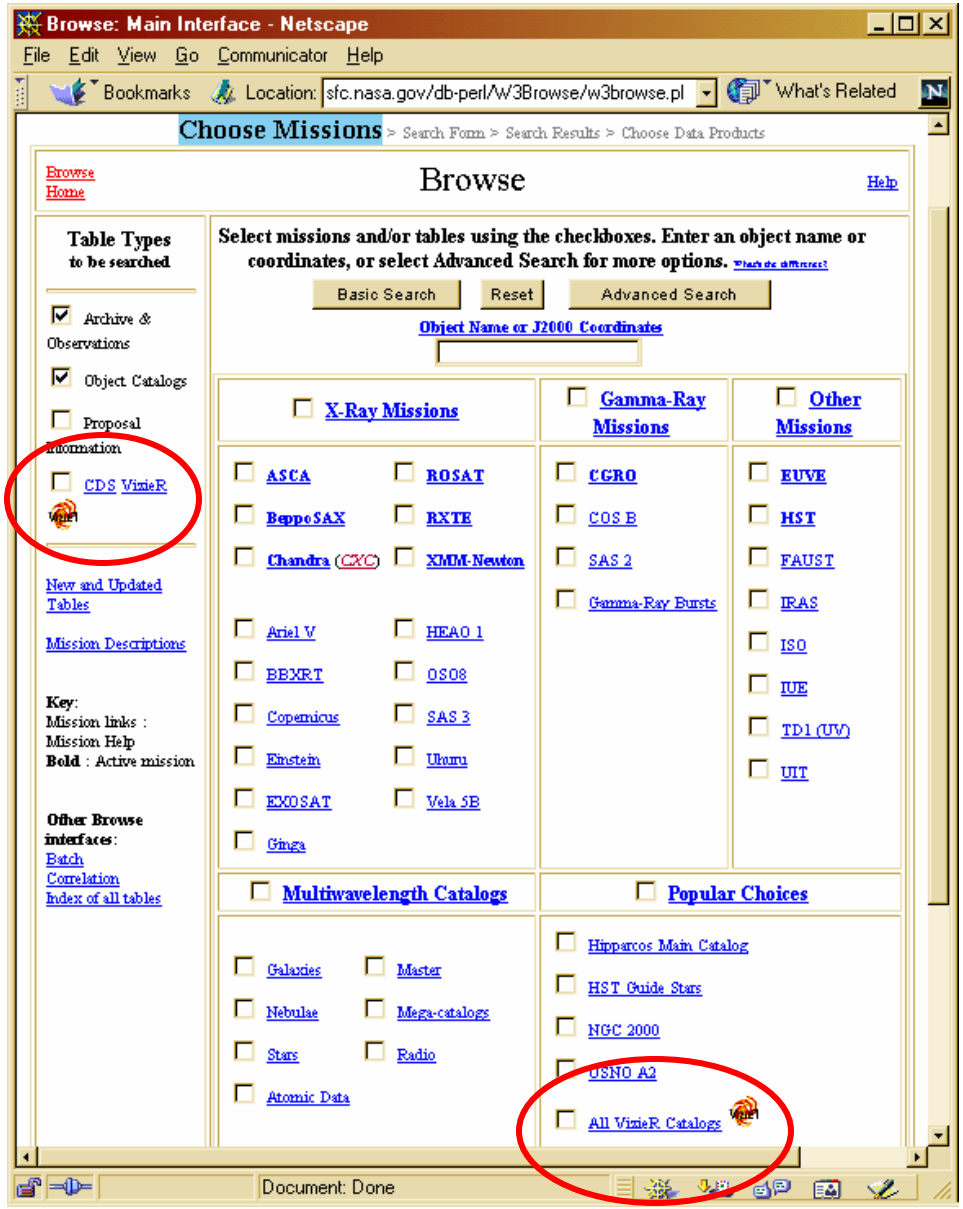


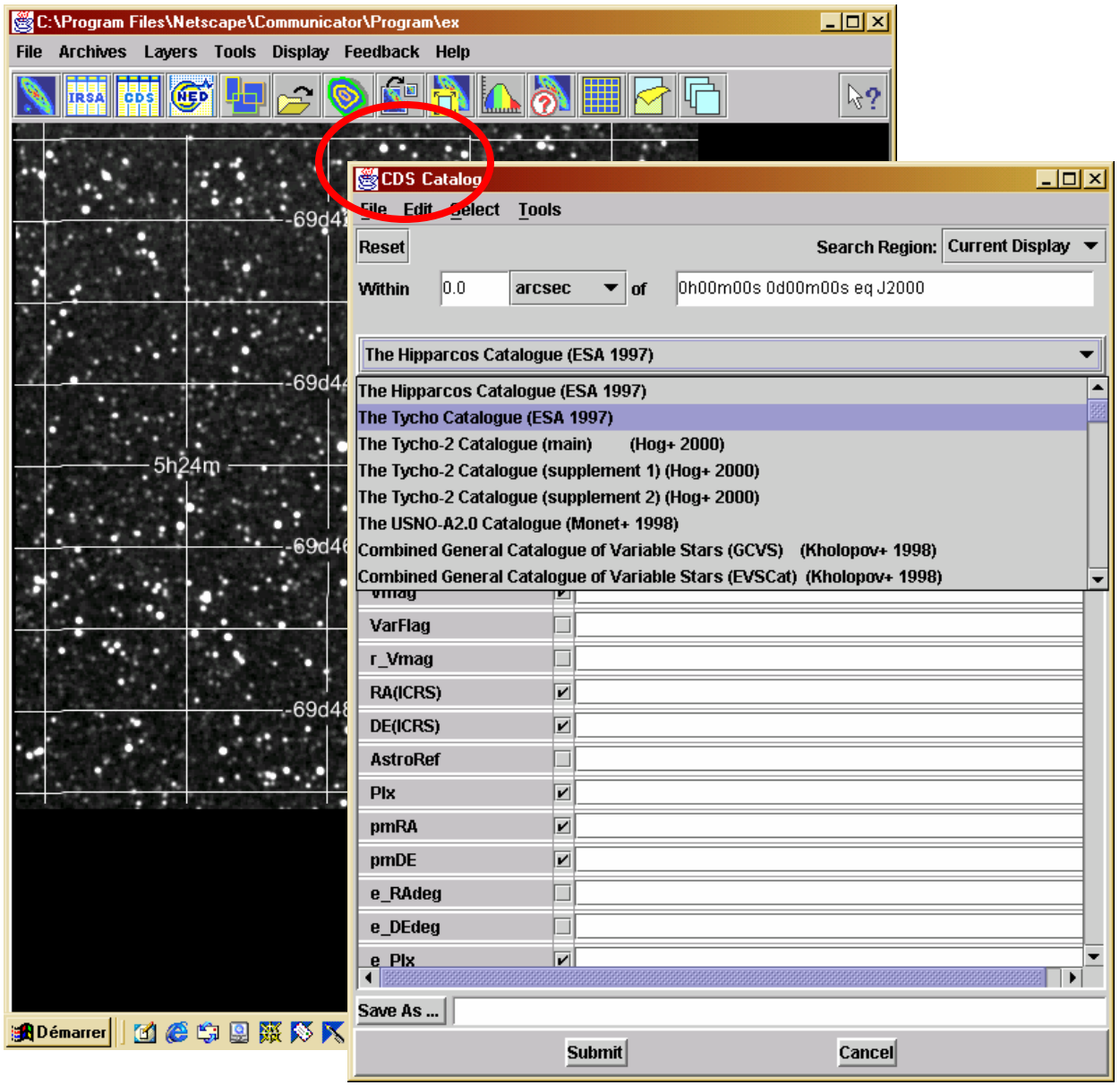
Towards data integration

- XML
- Questions
  - Minimize the ‘tagging overhead’
  - Management of binary ‘blobs’
  - Assess/distribute information about data quality
  - Unit knowledge/conversion
- *Astrores*, then VOTable

HEARSARC Browse

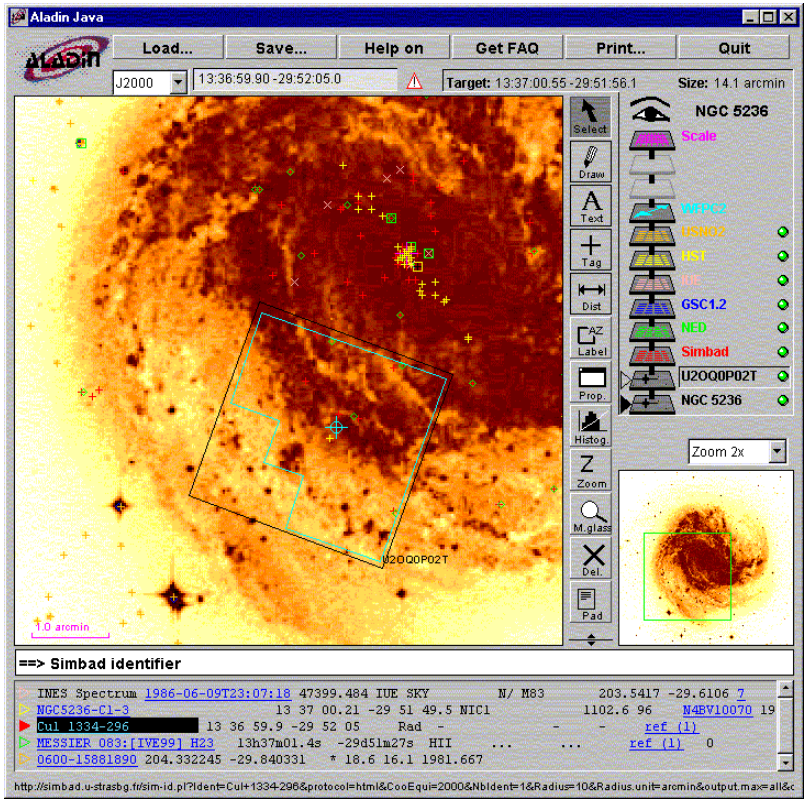
Astrores in action!




C:\Program Files\Netscape\Communicator\Program\ex
File Archives Layers Tools Display Feedback Help
IRSA
CDS
NED
-69d42
-69d44
5h24m
-69d46
-69d48
CDS Catalog
File Edit Select Tools
Reset
Search Region: Current Display
Within 0.0 arcsec of 0h00m00s 0d00m00s eq J2000
The Hipparcos Catalogue (ESA 1997)
The Hipparcos Catalogue (ESA 1997)
The Tycho Catalogue (ESA 1997)
The Tycho-2 Catalogue (main) (Hog+ 2000)
The Tycho-2 Catalogue (supplement 1) (Hog+ 2000)
The Tycho-2 Catalogue (supplement 2) (Hog+ 2000)
The USNO-A2.0 Catalogue (Monet+ 1998)
Combined General Catalogue of Variable Stars (GCVS) (Kholopov+ 1998)
Combined General Catalogue of Variable Stars (EVSCat) (Kholopov+ 1998)
VarFlag
r_Vmag
RA(ICRS)
DE(ICRS)
AstroRef
Plx
pmRA
pmDE
e_RAdeg
e_DEdeg
e_Plx
Save As ...
Submit
Cancel
Démarrer

OASIS

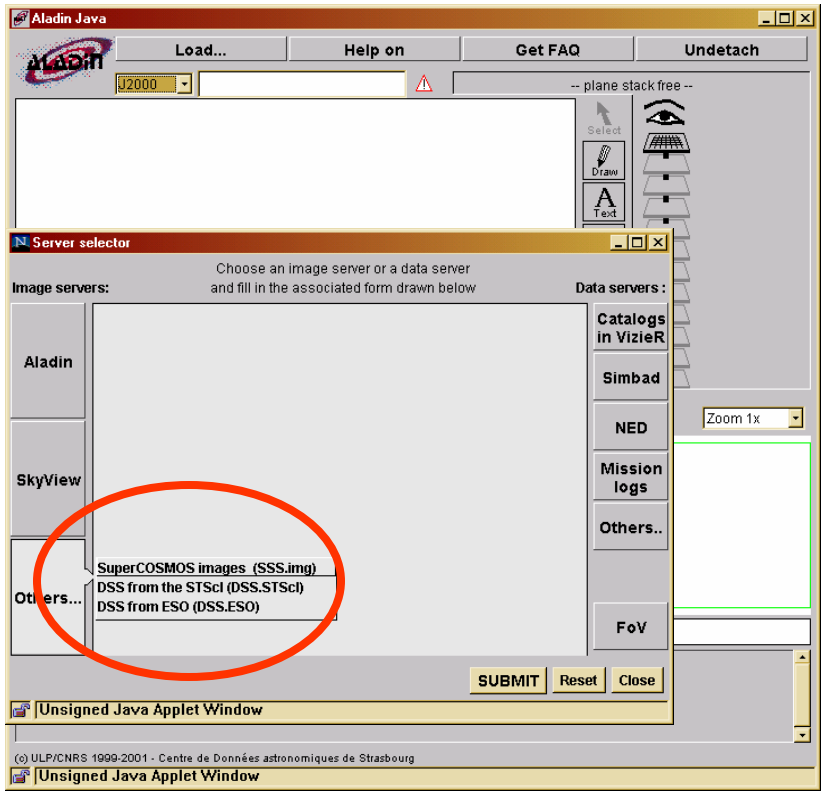
Data integration with



NGC 5236

- DSS image
- HST observation FOV
- SIMBAD and NED
- GSC, USNO A2
- IUE observations

ALADIN access to distributed information



- Distributed
  - Image databases
  - Catalogues
- Predefined
  - Supercosmos
  - HEASARC
  - SkyView
- User defined

The Virtual Observatory

« an enabling and coordinating entity to foster the development of tools, protocols, and collaborations necessary to realize the full scientific potential of astronomical databases in the coming decade »

NVO White Paper, juin 2000

Science driven

VO components (1)

- Network infrastructure
- GRID (computer AND data GRID)
- High capacity storage
- Management of very large data sets
- Access to distributed information
- Information retrieval tools
- Data Mining
- Statistical tools on very large distributed data sets

Virtual Observatory projects

- A priority in Europe (OPTICON) and USA (Decadal Survey)
- Several conferences
  - Virtual Observatories of the future (CalTech)
  - Mining the sky (Garching)
  - AstroGRID workshop (Belfast)
- Europe: AVO; USA: NVO; AstroGRID; Australia; Germany; …

VO components (2)

- Interoperability
  - Access to data
  - Query result integration
  - ⇨ Exchange standards
  - ⇨ Metadata
- Open question: define rules allowing people to deposit their data in the VO (quality?)

The AVO ASTROPHYSICAL VIRTUAL OBSERVATORY project (1)

- EC RTD project

  PI P. Quinn (ESO)

  ESO, ESA, AstroGRID, CDS, Terapix, Jodrell Bank

- Phase A proposal
- Shared cost project

The AVO project (2)

Three work areas

- Science Use Cases and Requirements (ST-ECF)

cf 

- Interoperability deployment and demonstration
- Technology needs (AstroGRID)
  - GRID systems
  - Scalable storage and computation
  - Databases

Interoperability prototype

Interoperability deployment and demonstration

Ground- and space, multiwavelength, multitechnique archives

Tools

VizieR (data federation)/Aladin (data integration)

cross-id … ESO/CDS Data Mining Project

Interoperability prototype

Products

- Working prototype open for community usage
- Running of science cases for evaluation
  - ⇨ Science driven: results, feedback !
- Improvement of federation and integration tools
- Counsel to archives on interoperability implementation
- Standards

Interoperability WG

Membership: Data managers

- Study *cost effective tools and standards* for improving access and data exchange to/from data archives and information services (minimal workload)
- Discuss the technical results of the AVO interoperability prototyping
- optical/IR (EC OPTICON), radio (EC RadioNet)
- International partnership beyond EU

  USA, Canada, Australia

The International VO alliance

Coordination between the projects

- A common roadmap
- First milestone (April 15th, 2002)

  VOTable V1.0

- VO meeting, Garching, June 2002

Information retrieval

Where is the data of interest for me?

- Practical Objective

  Define a minimal and ‘controlled’ set of metadata

  *To be done **with** archive/service managers*

- AstroBrowse/AstroGLU; ISAIA
  - Knowledge/maintenance of query syntax

    GLU (Générateur de Liens Uniformes)

The GLU

- GLU: a distributed/mirrored service directory
  - 31 synchronized repositories, 11 countries, 5 providers
- GLU can build queries by adjusting automatically the syntax as required by services (coordinate conversion, splitted fields)
- GLU is a toolkit managing HTML pages with dynamically updated links
  - dealing with mirror sites
  - dealing with keywords describing resources
- Hot topic: GLU vs WSDL/UDDI

Information retrieval (2)

- Proposal for query syntax

  ASU Astronomical Server URL

- Proposal for contents description

  UCD Uniform Content Descriptor

- Data quality

  Avoid illegitimate usage of data

UCDs

= Uniform Content Descriptor

- Developed in the frame of the ESO/CDS Data Mining project
- Hierarchical knowledge tree (2 000 items)
- Implemented for the 100,000 columns in VizieR (cats, publ. tables, logs, surveys)
- To be tested and extended: access tool

- AVO/Astrogrid/NVO common collaborations
  - VOTable (VOData as generalization?)
    - XML schemas
  - Metadata issues
    - Query mechanisms
    - Data service/resource directories
    - Data service/resource capabilities
  - Metadata dictionary
  - Data models
  - Identity, authentication
  - Science scenarios
  - User interfaces, portals??
- AVO/Astrogrid/NVO common milestones

List of topics

Conclusion

- Astronomy is at the forefront for data diffusion and networking
- The on-line services are everyday tools for scientists
- Building an international virtual observatory is an important endeavour for the coming years
- Partnership is critical
- Discipline standard/tools vs generic ones is a constant debate