The many faces of

SIMBAD home:
The Centre de Données astronomiques de Strasbourg

The CDS was created in 1972 by the Agency in charge of French ground-based astronomy (INAG, now INSU) as the *Stellar Data Center*

Initial charter:

- Collect 'useful' data on objects, in electronic form
- Improve them by critical evaluation and combination
- Distribute the results to the international community
- Conduct research using these data

  Gather stellar data to study the galactic structure

Main objectives of CDS

– Data curation

– Added-value services - local expertise on data

Provision of science tools to the community

In 1983: *Strasbourg Data Centre*

– Beyond stellar data

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

(Data ⇨ information)

SIMBAD

- SIMBAD begun before the CDS, as *Catalog of Stellar Identifications* (CSI), created in 1971
- Starting point: cross-identification of a few fundamental stellar catalogues (HD, SAO, GC, …)

  + bibliography, measurements from the catalogues
- Renamed SIMBAD in 1981

  Set of Identifications, Measurements and Bibliography for Astronomical Data
- Extended to extragalactic objects in 1983

  A homogenized view of astronomical objects across astronomy sub-disciplines

BICDS #2, 3: CSI, catalogues, collaborations

INFORMATION BULLETIN N° 2

DECEMBER 1971

Editor : J. JUNG - Observatoire de Strasbourg - 11, rue de l'Université -
67 - STRASBOURG - France

Contents :

INFORMATION BULLETIN N° 3

JULY 1972

Editor : J.JUNG - Observatoire de Strasbourg
11, rue de l'Université
67000 - STRASBOURG - FRANCE

CONTENTS:

F. Genova, 2006/10/17, ADASS 2006

SIMBAD hw/sw system (1)

- Evolution of hardware, software and languages, and of technical constraints and possibilities

- From the beginning: queriable from the distance (from batch punchcards to client/server and the Web)

- Four main releases: 1971, 1981,1990, 2006
  - From IBM mainframes to stations to PCs
  - From batch queries to interactive mode
  - From IBM dependant to home-made DBMS to PostgreSQL
  - Gained independance from hardware, operating systems, DBMS, vendors and developer

F. Genova, 2006/10/17, ADASS 2006

SIMBAD hw/sw system (2)

- SIMBAD 4
  - More flexibility to include new data and implement new functionalities
  - Full search capabilities
  - Look-and-feel for users not too different from SIMBAD 3 but with many more functionalities
  - The hidden face: a brand new graphical updating interface for SIMBAD librarians and astronomers

SIMBAD content (1)

- Specialized librarians who scan the journals and enter data and astronomers who provide scientific expertise
- Two main sources of information:
  - Systematic scanning of journals (90 journals)
  - Case by case: reference catalogues, including some observation logs
- Also decided case by case: systematic 'cleaning' of the somehow heterogeneous data entered from the literature
    e.g. X-ray objects from older satellites to prepare for XMM/Chandra
- Among recent evolutions: Notes, to keep track of important information about the object

SIMBAD content (2)

- Basic strategy: all objects found in published papers are entered in SIMBAD

- Data from published tables and catalogues

  'Long' tables are documented and entered in VizieR (in collaboration with journals and data centres)

  Semi-automated inclusion of catalogues and tables in SIMBAD with systematic cross-identification and check by an expert – not all tables are entered in SIMBAD

SIMBAD in the CDS hub (1)

- SIMBAD is not an isolated service
- From the beginning, two main aspects in CDS data gathering:
    - Cross-identifications and bibliography - *SIMBAD*
    - Catalogues (with description) – *catalogue service*, in collaboration with the other data centres, plus *VizieR* (browsing capability) from 1996
- Reference images and Aladin visualizer from 1998

SIMBAD in the CDS hub (2)

- Change in scale with the very large surveys and the fast increase in the number of lists of objects published in the litterature: complementarity between SIMBAD and VizieR

- Visualisation and comparison with distributed archives and services with Aladin

- By-product from SIMBAD: The *Dictionary of Nomenclature* of Astronomical Objects – contains complementary additional information linked to object names ('In source', list object type, instruments, …)

Acronym	Use Format		Year	1st Author	Obj. Type
(LMC M)	[MYM2001]	NNN LMC M HHMMm+DDMM	2001	MIZUNO N.+	CO Cloud
M		M NNN	1850	MESSIER C.	(Opt)
M	M	1-NN 2-NN 3-NN 4-NN	1946	MINKOWSKI R.	PN
M		M NNN	1975	MAFFEI P.	V*
(M)		GCM +LL.ll+BB.bb	1981	GUSTEN R.+	MCld
(M)	[GVC73]	{M} R.N	1973	GIOVANELLI R.+	Concentration
(M)	[H68]	{M} {M} R	1968	HULSBOSCH A.N.M.	HVC
(M)	[M59]	NN	1959	MANOVA G.A.	Em*
(M)	[M61a]	NN	1961	MINKOWSKI R.	G in ClG
(M)	[MAG95]	NNN	1995	MINNITI D.+	GCl
(M)		Mills HH+DA	1952	MILLS B.Y.	(Rad)
(M)	[MLV92]	NNNNNN	1992	MAGNIER E.A.+	*
(M)		MM NN	1965	MORAN M.	(Rad)
(M)		MSH HH+D-NN	1958	MILLS B.Y.+	(Rad)
(Mess)		M NNN	1850	MESSIER C.	(Opt)
Messier		M NNN	1850	MESSIER C.	(Opt)
(Min)	M	1-NN 2-NN 3-NN 4-NN	1946	MINKOWSKI R.	PN

F. Genova, 2006/10/17, ADASS 2006

Partnership

- A constant of CDS: collaboration with other partners: archive providers, other data centres (catalogues) and services (NED), ADS, journals
  - Data exchange >> interoperability standards
  - Tools e.g. name resolver, Aladin
  - Links
- Electronic journals – contents (TOC, tables) and links – added-value on published information 'processed published information'
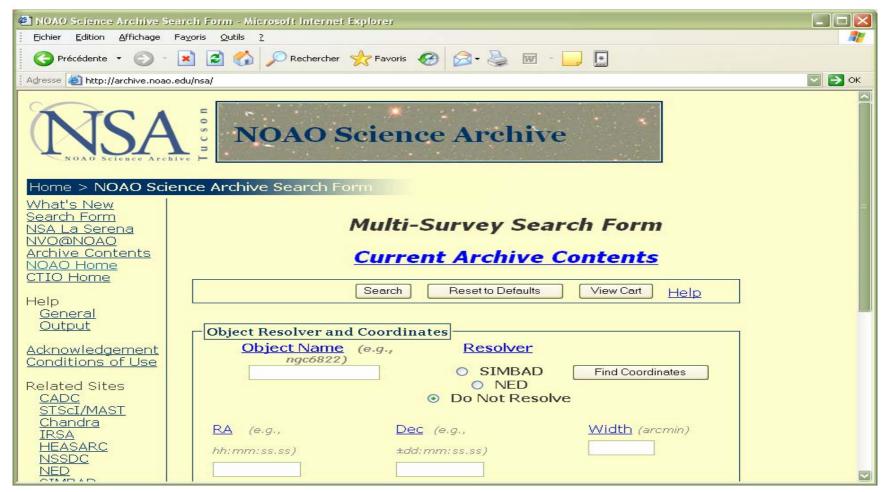
Interoperability

Interoperability an issue long before the VO!

e.g.

Data exchange with NED: definition of the bibcode/refcode (e.g., 2006A&A...447...89T), used and extended by ADS and the journals after the advent of the Internet – a key for the very rapid networking of astronomical bibliographic resources

Ready long before the Web and publishers' agreement on DOI and is human readable

SIMBAD client/server (1992): name resolver for observatory archives and the ADS
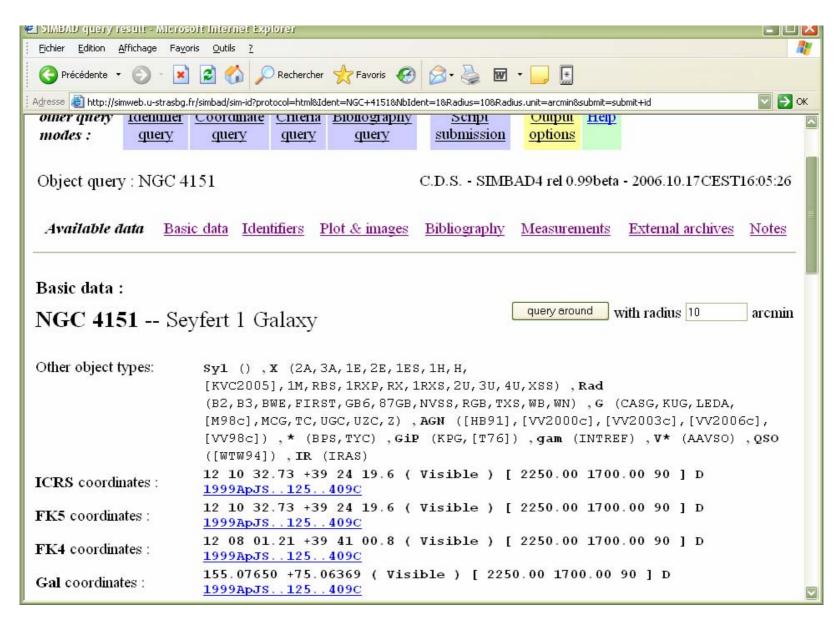
Key evolutions (1)

- SIMBAD 4 much more flexible
  - Full search capabilities
  - Give more information to users, e.g.,
    - Object types: each object name gives an object type from the Dictionary of Nomenclature, and all are displayed – multi-wavenlength information; helps the user to make up his/her mind when conflicts
  - With quality in mind, implementation of new functionalities require some validation, e.g.,
    - Hierarchy: 'Object in' from the Dictionary of Nomenclature + validation

F. Genova, 2006/10/17, ADASS 2006

Key evolutions (2)

- Dynamic cross-match: in addition to cross-matches kept in SIMBAD, propose possible cross-match and allow to compare data, using in particular catalogues available in VizieR and user provided catalogues
  - Cross-match tool in Aladin, soon implemented as a separated tool (AVO, VO-TECH)
  - SPECFIND (B. Vollmer et al.)
    - cross-match tool for radio catalogues using the source physical parameters (cross-match, hierarchy and associations)
    - extended in the VO context (VO-TECH project) to become a *dynamic* general SED builder for radio using VO technologies (UCD) – implies to enter new metadata for VizieR

F. Genova, 2006/10/17, ADASS 2006

Key evolutions (3)

- Increased synergy with VizieR and other CDS services

- VO R&D spin-off, in particular 'Intelligent resource discovery' (one of the Design Studies of the VO-TECH project), MDA (project in French 'Massive Data' IT program). In R&D phase:
  - Usage of Object-type ontology for information retrieval in Simbad (led by INAF A. Preite Martinez - Poster)
  - Object name recognition – *experts still needed for validation!*
    - Expected by-product: *proper link with NED taking into account the nomenclature differences*

F. Genova, 2006/10/17, ADASS 2006

Long term sustainability (1)

Dealing with the endless and ever increasing data flow without sacrificing quality

- Evolve the procedures to focus staff work on added-value tasks

  e.g., negociation with the journals when they became electronic, to obtain the Table of Contents

  This saved typing time and increased quality – but all references are still checked with the printed publication

Semi-automated procedures (TOC, 'raccord') - always keep validation by an expert

- Play the complementarity between the CDS services

Long term sustainability (2)

- Dealing with evolution of astronomy, technology and the 'political' context
- Very different time scales, activities and constraints have to be managed
  - Building of the content
  - Software development (data bases, user interfaces)
  - Operations
  - R&D on software and methods
  - Projects/collaborators schedules
- Evolutive strategy definition 'agile and ready to respond to unexpected situations, awareness' (NR yesterday)

Long term sustainability (3)

- Community support mandatory, which implies to be and remain science driven

- R & D mandatory to ensure the long term technical pertinence
  - Keep in mind that the objective is to improve the services
  - Use of new technologies not too early, not too late
  - Use projects (AVO, VO-TECH, MDA, …) for preparing medium/long term evolution (e.g., the huge influence of the VO development on Aladin)

Main lessons learnt

- Quality is a must and this is a very long term activity
  Routine is the worst enemy!
- Key importance of the integrated team, astronomers, librarians and software engineers
  - astronomers and engineers: strategy has to take into account the two points of view
  - can be an excellent specialization for librarians which are losing their traditional tasks
- Transmission of expertise and evolution of s/w and procedures to hand on the torch – e.g., major evolutions of SIMBAD and VizieR s/w

Now and in the future: the VO

- Long term keywords
  - Value-added services and tools
  - Interoperability

  fit well with the VO context
- VO is a change in scale and a remarkable new opportunity
- SIMBAD and the other CDS services are building blocks of the VO, with Aladin one of the VO portals
- More data, more services around, new functionalities, new types of usage