Introduction

40 years soon!!

CDS history

- Founded by INAG/INSU in 1972, agreement with ULP/UdS
- *A far-seeing vision*
  - 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 the data
- Initially stellar data for studying the galactic structure
- 1983 : *Centre de données stellaires* >>
  Centre de Données astronomiques de Strasbourg

CDS role

- Support the international astronomical community in its research tasks (not only collect information), *science driven* (and not technically driven)
- Core task: provide highly used value-added services
- Keywords: quality, scientific and technical relevance, collaborations, networking of expertise and resources
- A unique expertise on data, data dissemination and exchange standards
- Support to projects and a major role in the Virtual Observatory

The context of CDS activities

- CDS got the national label of *Très Grande Infrastructure de Recherche* from the MESR as a result of the 2008 AERES evaluation of the Observatory

  A recognition of its impact and of the quality of the work
- TGIR managed by INSU, agreement with UdS
- A team within the Observatoire Astronomique de Strasbourg

Hervé's talk tomorrow

CDS activities

- Reference, added-value services
  - The basis of CDS service to the community
- Technological watch, R&D
  - One key of long term sustainability
- Virtual Observatory
  - Leadership, interoperability standards, tools
  - European projects
- Participation in projects
- User support, knowledge dissemination, expertise
- *Science*

Scientific activities at CDS

- Active participation of active scientists is another of the keys of success and sustainability: high-level scientific expertise is mandatory to
  - Ensure the quality of content
  - Ensure the relevance of the content and services on the long term, taking into account the constant evolution of astronomy and of users' needs
- The founding fathers were stellar astronomers, now a palette of expertise to cover the different topics of astronomy and multi-wavelength aspects

Bernd's talk

CDS staff

- An integrated team of scientists, « documentalistes » (information scientists) and software engineers
- Symbiosis between the different types of activities
- Relies on the Observatory general services for support activities (computer and network system, administration, logistics)
- The team organisation will be described tomorrow in a specific talk

CDS services

The CDS hub

Astronomical objects

identification, bibliography, data, measurements

Federation of tabular data

catalogues, published tables observation logs, surveys

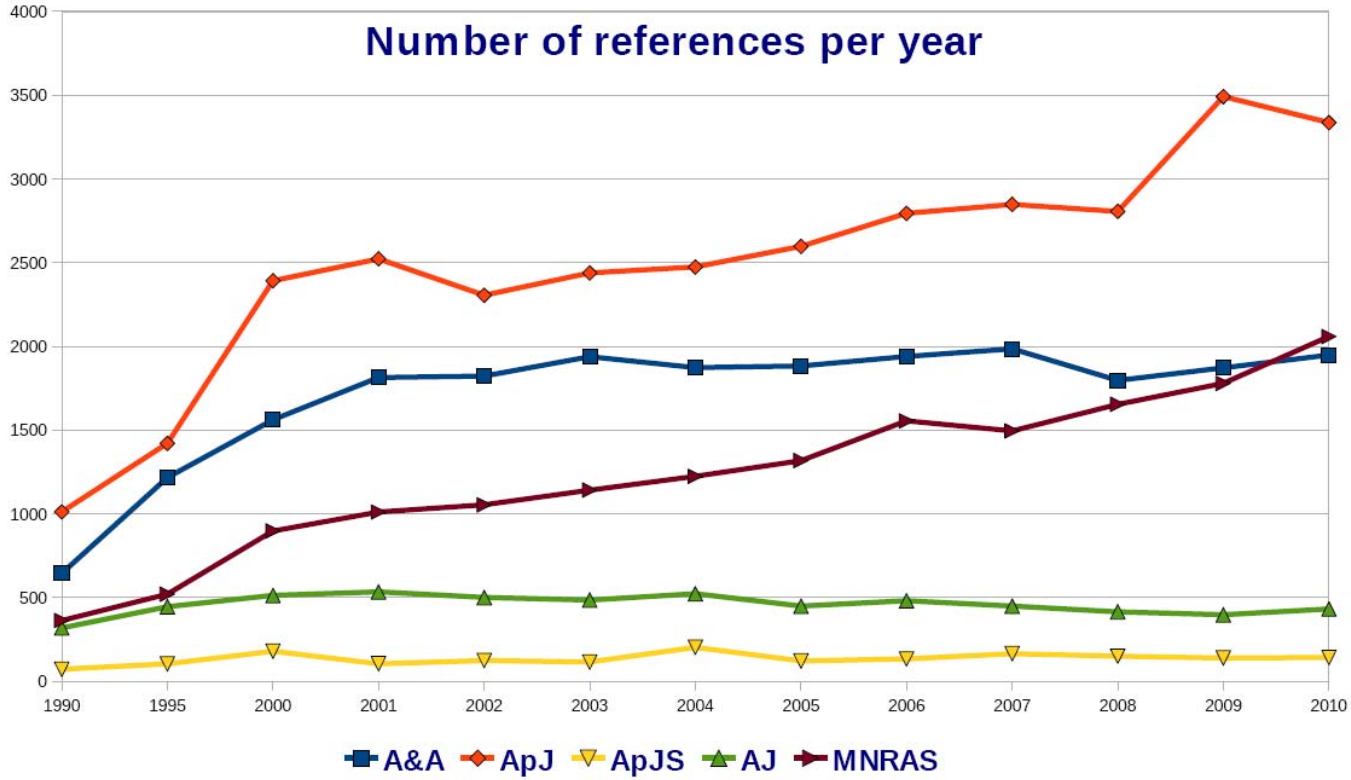
Integration, visualisation, manipulation

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

Elements of strategy

- Maintain the services at the highest possible level in terms of content and functionalities
- Add functions to the core services in line with our expertise, the users' needs and R&D results, e.g.,
  - Aladin « Healpix » mode (Planck) opens a new way to use Aladin, and Healpix will be used for other usages by CDS (cross-match) and is an entry point to Gaia
  - A new change in paradigm with Web 2.0/3.0: fist implementation in the services
- Change in scale (increase of publication volume, very large surveys)

  Accept the complementarity between VizieR (exhaustive) and SIMBAD (not complete, more added-value) and put it into action, e.g.
  - Sesame as a common name resolver, which also includes NED
  - Selection of data for SIMBAD on scientific criteria (for lists)


Number of references per year
4000
3500
3000
2500
2000
1500
1000
500
0
1990
1995
2000
2001
2002
2003
2004
2005
2006
2007
2008
2009
2010
A&A
ApJ
ApJS
AJ
MNRAS

SIMBAD snippets (1)

- The reference database for identification and bibliography of astronomical objects, providing a homogenized view across astronomy subdisciplines
- The current version of the software (the fourth major update since 1972) has been operational since 2006
  - From a home-made object-oriented database to PostgreSQL
    Increased maintenability, flexibility and searchability
  - Constant evolutive maintenance
  - Specific implementations to speed up simple queries by other services (ADS, Aladin)
  - Adding less controlled data: CDS team notes (2002), user annotations (2010)
- Mirror copy at CfA, maintained by the ADS (more and more elaborate links between the services and excellent collaboration)

Marc's talk

snippets (2)

- Contents

« Content's » talks

  - A team of highly qualified documentalistes and astronomers
  - Semi-automated entry of objects (paper text and selected tables), validation by a specialist
  - April 2011: 5.000.000 objects, 15.000.000 identifiers and 8.000.000 reference citations

	2003	2008	2011
Objects	3.000.000	3.900.000	5.400.000
Identifiers	8.300.000	11.750.000	15.124.000
References	140.000	216.000	253.000
Citations	4.100.000	5.750.000	8.180.000

***illustration: peut-être image composite avec
l'interface web, une interface name resolver
dans une archive, une utilisation dans le proto ADS***

snippets (1)

Francois/Gilles' talk

- The reference database for tabular data from astronomical catalogues and tables published in scientific papers (CDS is the data curator)
- Tables and their description, which links physical and astronomical content
- Relational database system
  - Sybase for the master database, queries distributed on a local cluster
  - PostgreSQL for some of the mirror copies (CDS, ADS, INASAN, UKIRT), the others use Sybase (CADC, NAOJ, NAOC, UCam).
- Specific system for very large catalogues (10^9 objects) for efficient queries by position

snippets (2)

- Close collaboration with the journals

  On-line publication of « long » tables for A&A since 1993!
- A change in paradigm: published, printed tables are re-usable data
- Many tables come with « attached » data, stored at CDS or elsewhere
- 9.100 catalogues in 2011

  6.500 in 2008, 3.800 in 2003

Quelle illustration??? Sed? Nouvelle interface? une table avec des donnees additionnelles (ou ca plutot dans l'expose 'strategie')?

snippets (1)

Pierre's talk

- Reference software dedicated to the integration, visualisation and manipulation of images and catalogues provided by most astronomical data servers in the world (CDS, ESA, ESO, NASA, CADC, …)
- Continuously evolving, many new functionalites (huge images, huge data cubes, photometry, image convolution, cross-match, usage in scripts, …)
- Used by ESA, STScI, NED, CADC to visualize images
- The VO image portal, able to interact with other VO tools such as Topcat
- Major recent evolution: usage of Healpix, which provides a hierarchical view of data with fast zooming capabilities

  Starting point : usage of Healpix in Planck – our wish to offer an efficient tool to a new user community (also WMAP)

snippets (2)

- A new way of using the tool, also adapted to building views of the full or part of the sky from data obtained by individual projects (they are offered the possibility to build easily a local Healpix database which they can opened for usage by all Aladin users, or by their collaborators, or keep for themselves)
- CDS image databases
  - The ‘historical’ reference image database (*** TB), in which image data sets are fully documented (« hierarchical data tree »)
  - A rapidly growing collection of Healpix views of the sky from different projects
- Current evolution of the image database into a metadata database allowing to start exploration from the Healpix view and to get full information about the images of interest
- Opens the way for usage for CFHTLS and other projects, including Gaia

illustration?

Technological watch
R&D activities

R&D strategy (1)

- A fundamental activity for medium/long term sustainability, to be maintained continuously in spite of the operational constraints
- Driven by the data centre needs (and NOT technology driven)
- Deal with the evolution of astronomy and take advantage of the evolution of technology
- In-house activities, managed by permanent staff in general working with contractors or trainees
- A significant fraction of the time of engineers and « instrumentalistes »

R&D strategy (2)

- Take advantage of projects
  - Since 2001, a series of European projects

    VO oriented R&D which has also strongly influenced the services
  - Problems to get ANR funding on this kind of R&D

    No success since the ACIs in 2001-2005 in spite of several proposals, some in collaboration with IT labs
- Products
  - New services, new functionalities
  - VO standards and tools

R&D talk
Mark's talk

A few examples

- Expertise in dealing with textual information + Dictionary of Nomenclature: *DJIN*
- Aladin + image expertise (LSIIT long term collaboration) + Planck (a new type of usage): *Aladin Healpix*
- Long term expertise with catalogues + VizieR + large surveys: *cross-match* – soon to come
- The rapid emergence of Web 2.0/3.0 : *annotations*, *CDS portal* (‘*mashup*’), *mobile interface*… to be continued

Participation in projects

Strategy for project participation

- Driven by the data centre needs and expertise
- Support to projects at different levels
  - Customized usage of services (e.g., XMM, Planck)
  - Data distribution (e.g., CFHTLS)
  - Interoperability
  - Counselling
- We have chosen to participate fully in the VO endeavour – a strategic choice at the beginning of the VO

The CDS and the VO

>> « CDS in the VO talk »

- The CDS has been a precursor of the VO in many respects
- It has been a major actor of the VO since the emergence of the concept in 1999-2000
  - A major role in the IVOA activities: international coordination of activities and definition of the standards which are the VO framework
  - CDS services are important building blocks of the VO
  - Leadership role at the European and national level

A viewgraph presented to the CDS SC in Nov. 2001

Interoperability standards

- Essential ingredients of the global VO

  Information retrieval, information exchange, integration of query results, common tools

- AVO/NVO collaboration

  Visit of R. Williams (VO architect)

  → Starting point of the first VO standard, VOTable (March 2002)

- A roadmap for the VO Garching meeting

  (May 2002)

  → IVOA created at this meeting

European projects leadership

- Euro-VO Data Centre Alliance, FP6, Coordination Action, 2006-2008, 1.7 M€
- Euro-VO Astronomical Infrastructure for Data Access, FP7, I3, 2008-2010, 2.7 M€
- Euro-VO International Cooperation empowerment, FP7, Coordination Action, 2010-2011/12, 210 k€
- Also participation in VO-TECH, FP6, Design Study, 2004-2009, part CDS ~700 k€ (and previously in FP5 AVO)

EuroVO-AIDA

- ***a summary of that particular project may be useful and remains to be done***

European projects impact

- Major impact on CDS services
- Major tool to build European-wide cooperation and a common view on VO standards

- The European projects have enabled many of the VO standards and VO-enabled services and tools

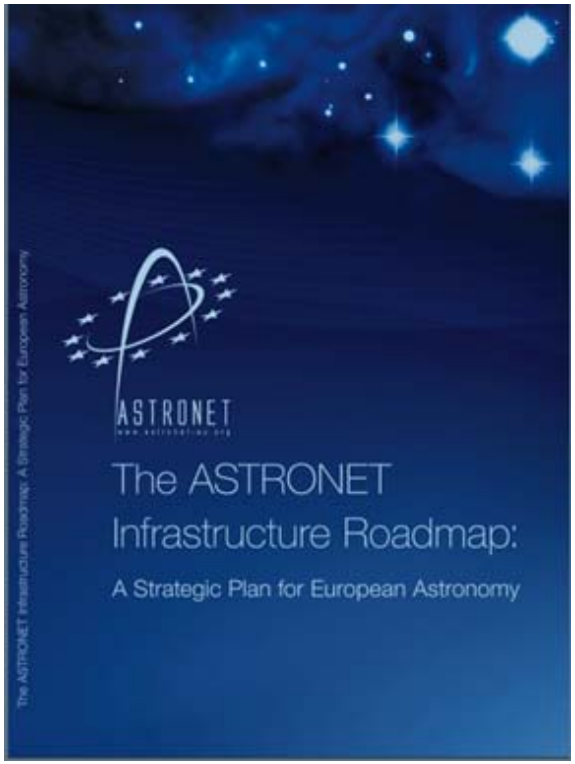
- Excellent technical collaborations in particular with INAF on Semantics, and with ESAC for the usage of Aladin in astronomical missions including Herschel and Planck, but also SOHO (solar data)

- *But no sustainability of European funding*

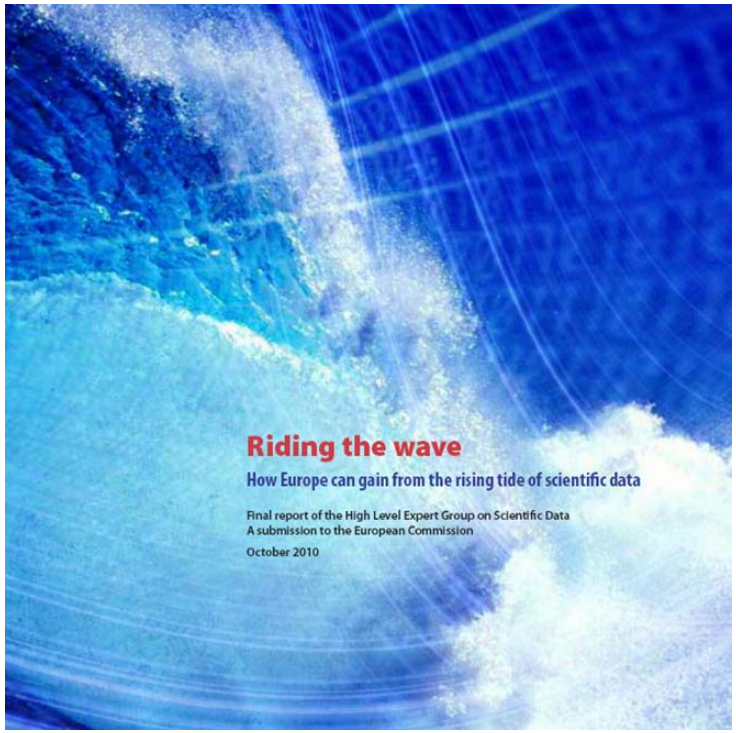
User support, knowledge dissemination, expertise

- Actions towards the astronomical community (dissemination on the usage of CDS services and of the VO) Mark's and Caroline's talks
- In France, expertise/counselling
  - Astronomical projects (CoRoT, etc)
  - For other disciplines of INSU (e.g., seismology information system)
  - For other Institutes of CNRS (e.g., IN2P3, the ADONIS human sciences portal for INSHS)
- More generally, CDS (and astronomy!) at the forefront for the sharing of scientific data

Examples of expertise at European level



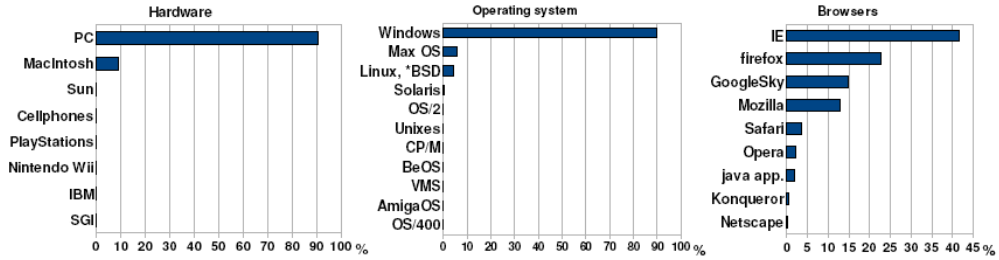
Astronet Roadmap



High Level Expert Group on Scientific Data

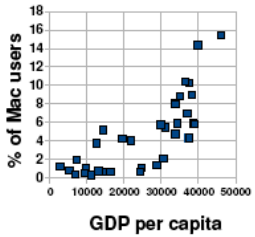
Impact

- SIMBAD
  - 2003: ~20.000 queries/day
  - 2007: 70.000 queries/day
  - 2010: 238.000 queries/day
- VizieR
  - 2005: 31.000 queries/day
  - 2007: 90.000 queries/day
  - 2010: 246.000 queries/day
- Aladin
  - 2003: 2.100 queries/day
  - 2007: 9.000 queries/day
  - 2010: 16.500 queries/day (500.000/month) + local installs

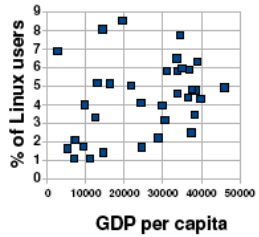


SIMBAD users' operating system (statistics produced in 2009)

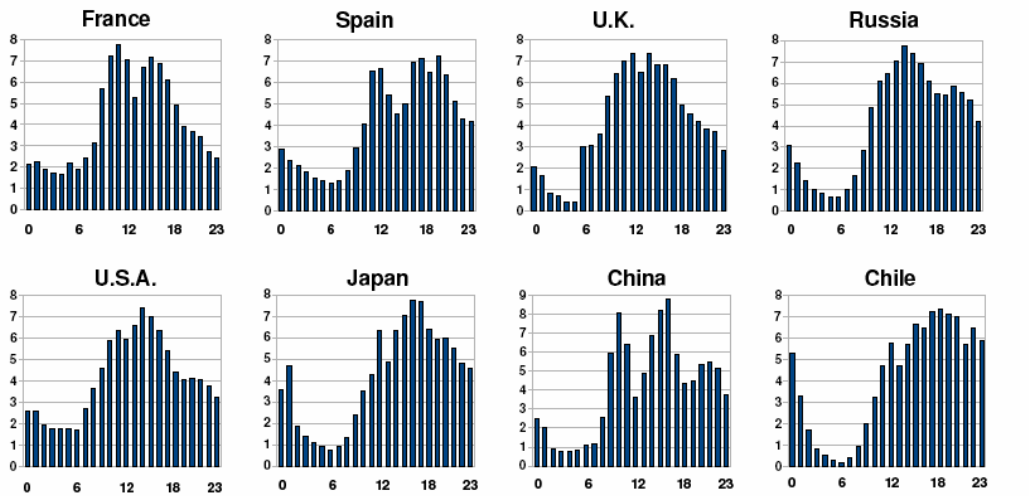
There is a correlation between the GDP (Gross Domestic Product) per capita [ref 2] and the percentage of Mac Os users in these countries



This correlation does not exist for Linux users...

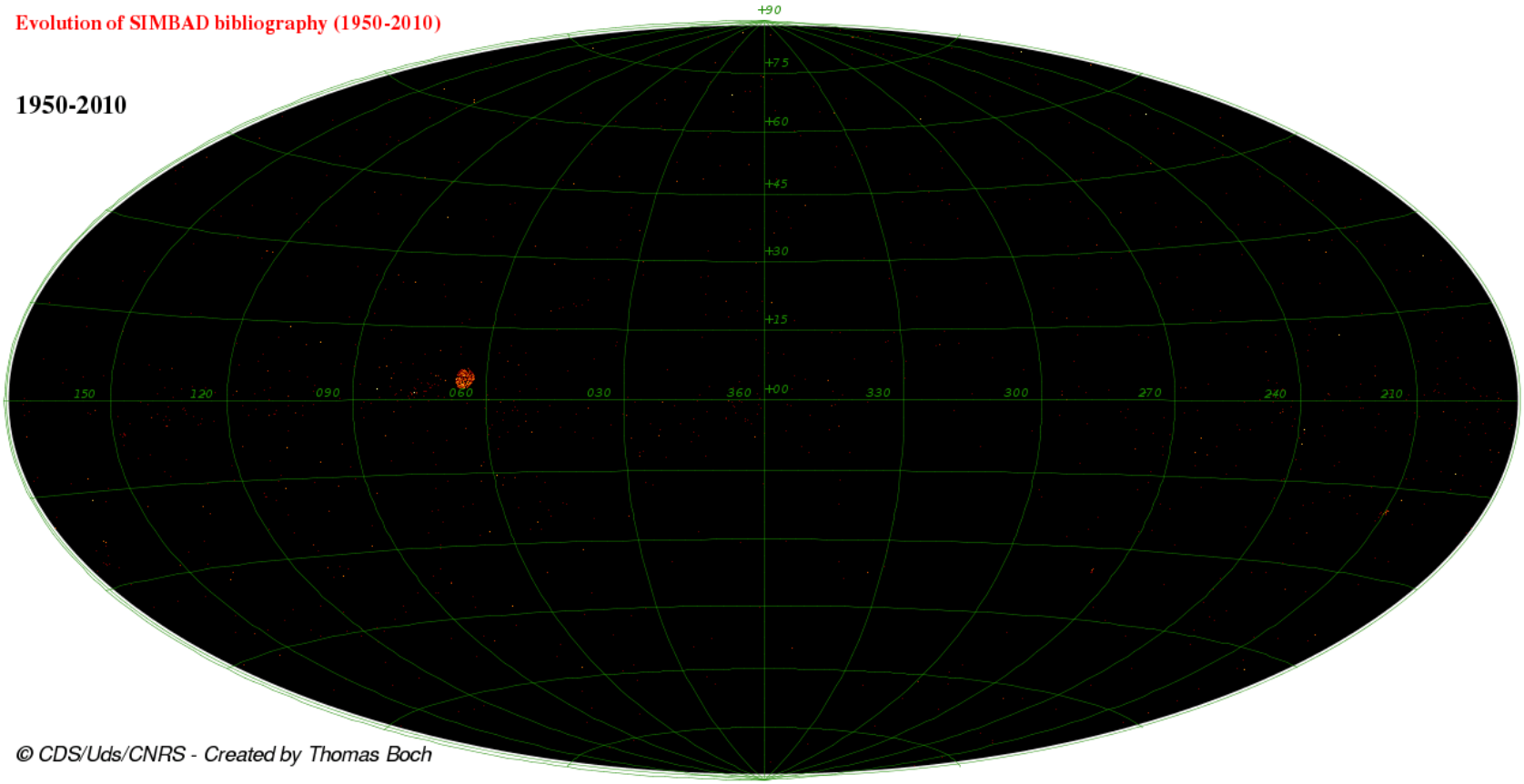


Working day of SIMBAD usage (local time)



Evolution of SIMBAD bibliography (1950-2010)

1950-2010



© CDS/Uds/CNRS - Created by Thomas Boch