Astronomy data centres as everyday tools for scientists

The astronomical Virtual Observatory (1)

- Starting point:
  - Heterogeneous, distributed ‘data’ services: archives of observations, value-added data bases, tools, bibliographic data (including e-journals), simulation data
  - Most data freely accessible after proprietary period, most journals after 3 years
- The Virtual Observatory concept

  Seamless and transparent query of data centres

  New analysis and visualisation tools

  A standard structure for data centres to publish their data and services

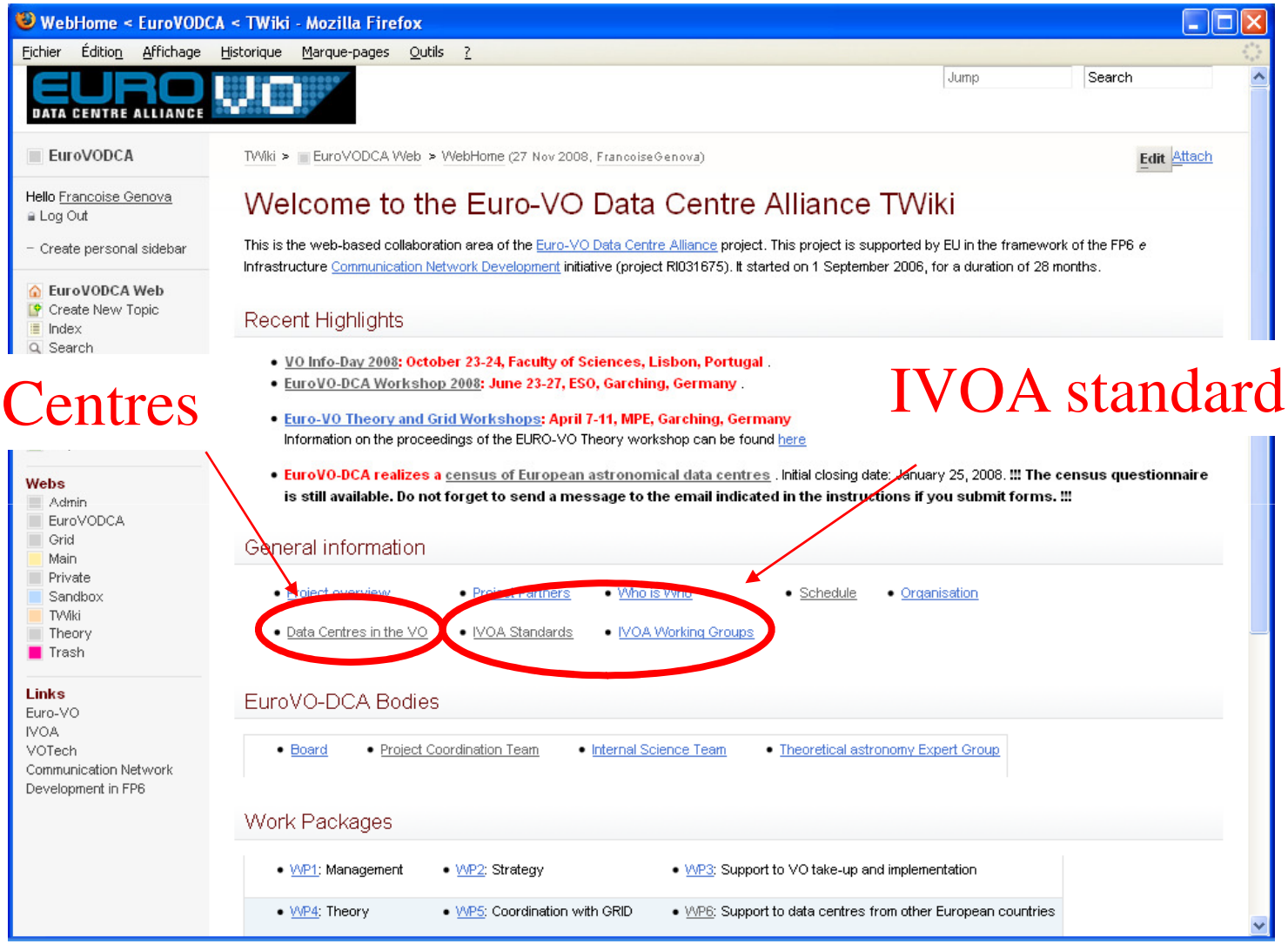
The astronomical Virtual Observatory (2)

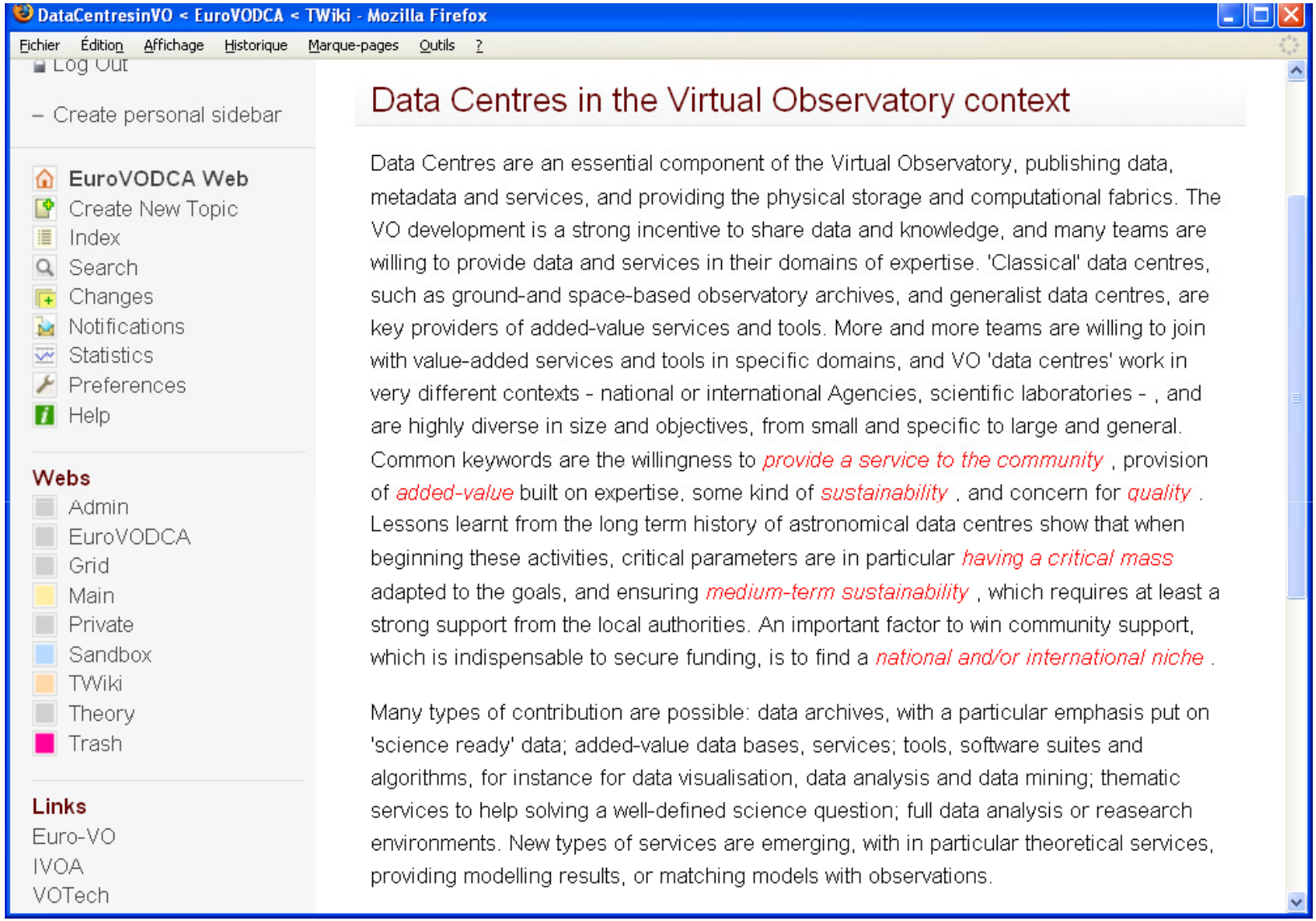
- The International Virtual Observatory Alliance defines discipline-specific interoperability standards
  - Registry of Resources (OAI compliant). disciplinary Semantics, data models, data access layer, query language…
- A thin interoperability layer on top of data and included in services to publish in the Virtual Observatory framework
- Data+Services+Interoperability = *data and service grid*
- Incrementally made available for community usage

Data centres in the VO

- How is this grid constituted?
- Data Centres are essential building blocks, which populate the Virtual Observatory with data and services
- Who are the data and service providers?
- Euro-VO Data Centre Alliance
  - FP6 Coordination Action 2006-2008
  - Supports data providers in their uptake of the VObs framework and gathers feedback from implementatio

  Census of European astronomical data centres


WebHome < EuroVODCA < TWiki - Mozilla Firefox
Fichier Édition Affichage Historique Marque-pages Outils ?
DATA CENTRE ALLIANCE
Jump
Search
EuroVODCA
Hello Francoise Genova
Log Out
Create personal sidebar
EuroVODCA Web
Create New Topic
Index
Search
Webs
Admin
EuroVODCA
Grid
Main
Private
Sandbox
TWiki
Theory
Trash
Links
Euro-VO
IVOA
VOTech
Communication Network Development in FP6
TWiki > EuroVODCA Web > WebHome (27 Nov 2008, FrancoiseGenova)
Edit Attach
Welcome to the Euro-VO Data Centre Alliance TWiki
This is the web-based collaboration area of the Euro-VO Data Centre Alliance project. This project is supported by EU in the framework of the FP6 e Infrastructure Communication Network Development initiative (project RI031675). It started on 1 September 2006, for a duration of 28 months.
Recent Highlights
VO Info-Day 2008: October 23-24, Faculty of Sciences, Lisbon, Portugal .
EuroVO-DCA Workshop 2008: June 23-27, ESO, Garching, Germany .
Euro-VO Theory and Grid Workshops: April 7-11, MPE, Garching, Germany
Information on the proceedings of the EURO-VO Theory workshop can be found here
EuroVO-DCA realizes a census of European astronomical data centres . Initial closing date: January 25, 2008. !!! The census questionnaire is still available. Do not forget to send a message to the email indicated in the instructions if you submit forms. !!!
General information
Who is Who
Schedule
Organisation
Data Centres in the VO
IVOA Standards
IVOA Working Groups
EuroVO-DCA Bodies
Board
Project Coordination Team
Internal Science Team
Theoretical astronomy Expert Group
Work Packages
WP1: Management
WP2: Strategy
WP3: Support to VO take-up and implementation
WP4: Theory
WP5: Coordination with GRID
WP6: Support to data centres from other European countries
Data Centres
IVOA standards


DataCentresinVO < EuroVODCA < TWiki - Mozilla Firefox
Fichier Édition Affichage Historique Marque-pages Outils ?
Log Out
– Create personal sidebar
EuroVODCA Web
Create New Topic
Index
Search
Changes
Notifications
Statistics
Preferences
Help
Webs
Admin
EuroVODCA
Grid
Main
Private
Sandbox
TWiki
Theory
Trash
Links
Euro-VO
IVOA
VOTech
Data Centres in the Virtual Observatory context
Data Centres are an essential component of the Virtual Observatory, publishing data, metadata and services, and providing the physical storage and computational fabrics. The VO development is a strong incentive to share data and knowledge, and many teams are willing to provide data and services in their domains of expertise. 'Classical' data centres, such as ground-and space-based observatory archives, and generalist data centres, are key providers of added-value services and tools. More and more teams are willing to join with value-added services and tools in specific domains, and VO 'data centres' work in very different contexts - national or international Agencies, scientific laboratories - , and are highly diverse in size and objectives, from small and specific to large and general. Common keywords are the willingness to provide a service to the community , provision of added-value built on expertise, some kind of sustainability , and concern for quality . Lessons learnt from the long term history of astronomical data centres show that when beginning these activities, critical parameters are in particular having a critical mass adapted to the goals, and ensuring medium-term sustainability , which requires at least a strong support from the local authorities. An important factor to win community support, which is indispensable to secure funding, is to find a national and/or international niche .
Many types of contribution are possible: data archives, with a particular emphasis put on 'science ready' data; added-value data bases, services; tools, software suites and algorithms, for instance for data visualisation, data analysis and data mining; thematic services to help solving a well-defined science question; full data analysis or reasearch environments. New types of services are emerging, with in particular theoretical services, providing modelling results, or matching models with observations.

EuroVO-DCA Census

- An inclusive definition
- Five questionnaires
  - Identification of the data centre
  - Observational archives and data products
  - Services/Tools/Software suites
  - Theoretical archives
  - Theory services
- Level of granularity left to respondants (e.g., data from each different instruments on a large telescope could be considered as a different data product, or all data from one telescope or one space mission as one single data set)

Results (1)

- 68 ‘data centres’ answered
- Data archives: 134 (ESO: 24; ESAC:10)
  - Covers all fields of astronomy and all types of astronomical data (e.g., tables published in journals)
  - Most provide ‘science ready’ data
  - Wide variety of size
  - Most use or plan to use VObs standards

Results (2)

- Services: A rich population (67 responses from 36 institutes) of very diverse value-added services, from complete bibliographies and libraries of astronomical objects to services dealing with specific object types, and of sofware tools with generic (access to images, spectra, tabular data, data publication, etc) or specific goals

Results (3)

- Theory data & services:
  - A solid population of theory-related services (a new topic for the VObs) with 15 institutes identifying 23 archives and services, including huge and small simulations
  - Nearly all archives have a fully open data access policy
- The provision of data & services has clearly been strongly encouraged by the development of the VObs (in particular for theory but not only)

The astronomy knowledge grid (1)

- A huge diversity of repositories:
  - large services provided by international agencies, with archives of the large ground-based and space instruments
  - large systematic surveys of the sky, results of large simulations
  - generalist data bases and services
  - smaller contributions of scientific teams which share their expertise
- Some of these services are are widely used by scientists to access to bibliography, data and tools

The astronomical knowledge grid (2)

- The VObs framework is fully international and allows all astronomers with transparent access to data and services wherever they are located (full international interoperability)
- Strong international links also at the data centre level (the Canadian data centre has answered the census, some services are mirrored in other countries or contain copies of data from other countries in particular to ensure security)

Conclusions (1)

- A fully distributed data curation model with no central point
  - Agencies responsible for large infrastructures provide data archives
  - Established data centres provide value-added services and tools
  - Now smaller, motivated actors are appearing
- On-line data & service availability has changed the way astronomers do research and has been critical for the development of ‘multi-wavelength’ astronomy (now a significant fraction of published papers)

Conclusions (2)

- This was made possible in particular
  - by the development of community standards, which began more than 30 years ago (common data format),
  - by the advent of the Internet for facilitating on-line access to data,
  - now by the development of the VObs to allow seamless access to data

Conclusions (3)

- Widely recognized that projects (observations, theory) need an archive & data distribution
  - Much better when foreseen from the beginning
  - Cost should be included in the project cost
- Long term support required for large projects
- Small teams: local or regional support centres
- Critical to remain very close to user needs
- Remain close to expertise at least during the project active life (data & service description, data updates & re-engineering, evolution of models, etc)