The Astrophysical Community and the EGEE

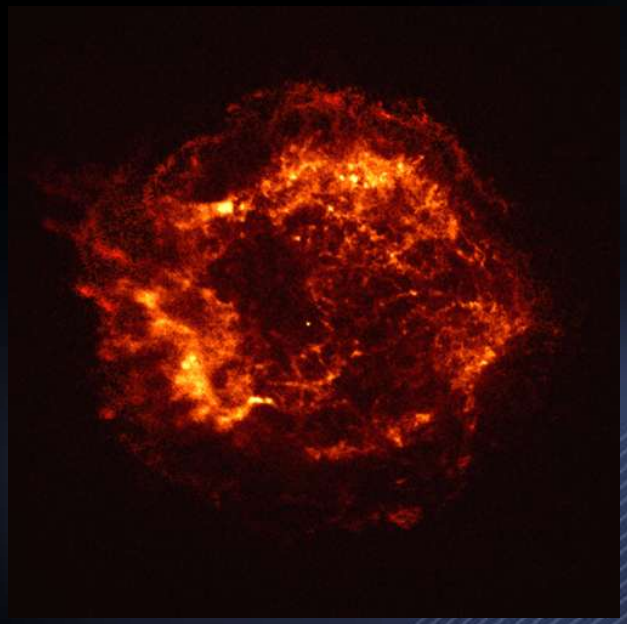
Nicholas A Walton
AstroGrid Project Scientist
University of Cambridge

Astronomy and the Grid

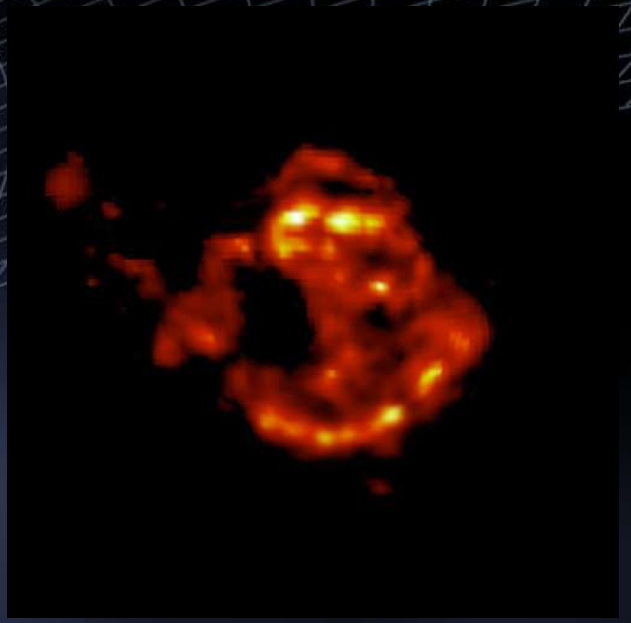
- Astronomy is a BIG international science with similar challenges to other physical science
- Astronomy projects involve:
  - Interaction coordinated research efforts
  - Distributed multi-wavelength teams, resources & data
  - Data volumes with doubling times < 12 months
- Astronomy service organisations need to
  - Provide users with access to software tools, high quality raw and processed data in the face of user computing power and network bandwidths with doubling times > 18 months
- Opportunities for new scientific capabilities

Supernova Remnant Cassiopeia-A – a 300 year old Supernova

The Challenge and Opportunity of multi-Wavelength data:

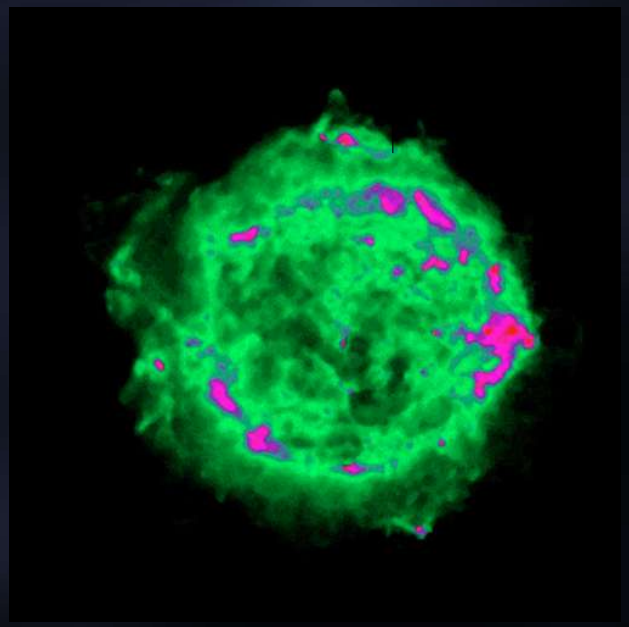


Shocks seen in the X-ray Chandra image



Dust shows in the IR

Heavy elements seen in the optical

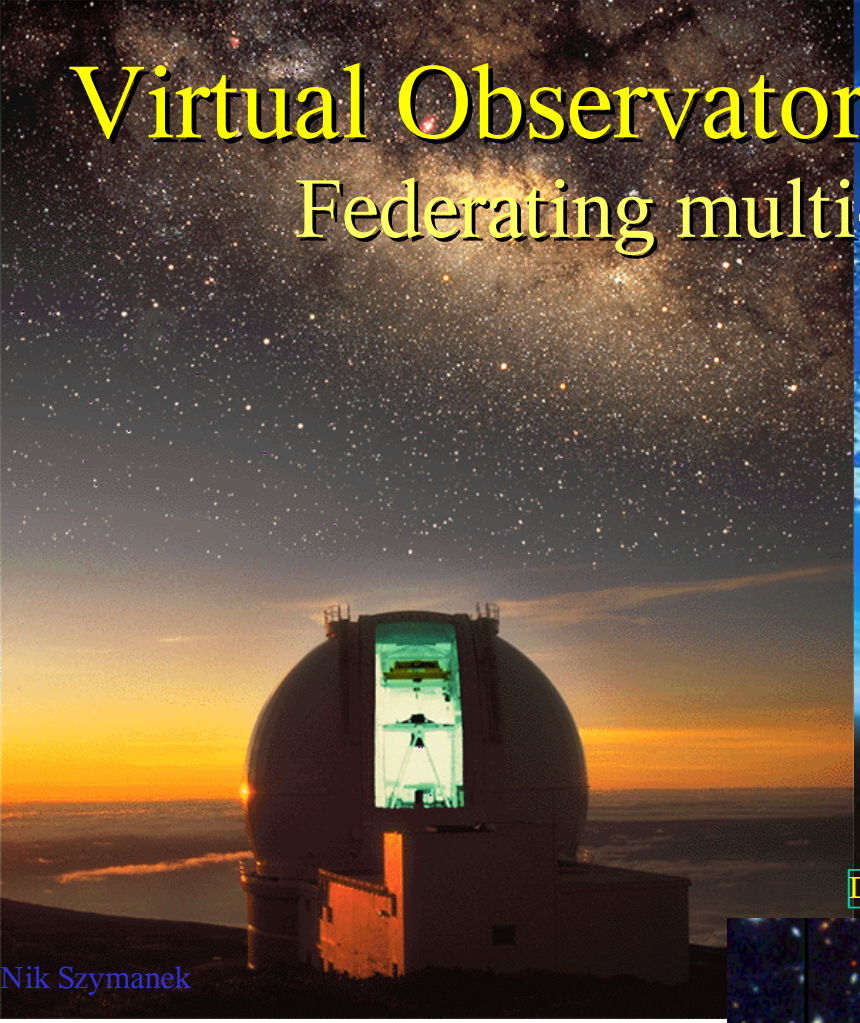


Mapping e^-s in the magnetic field as revealed by Radio data

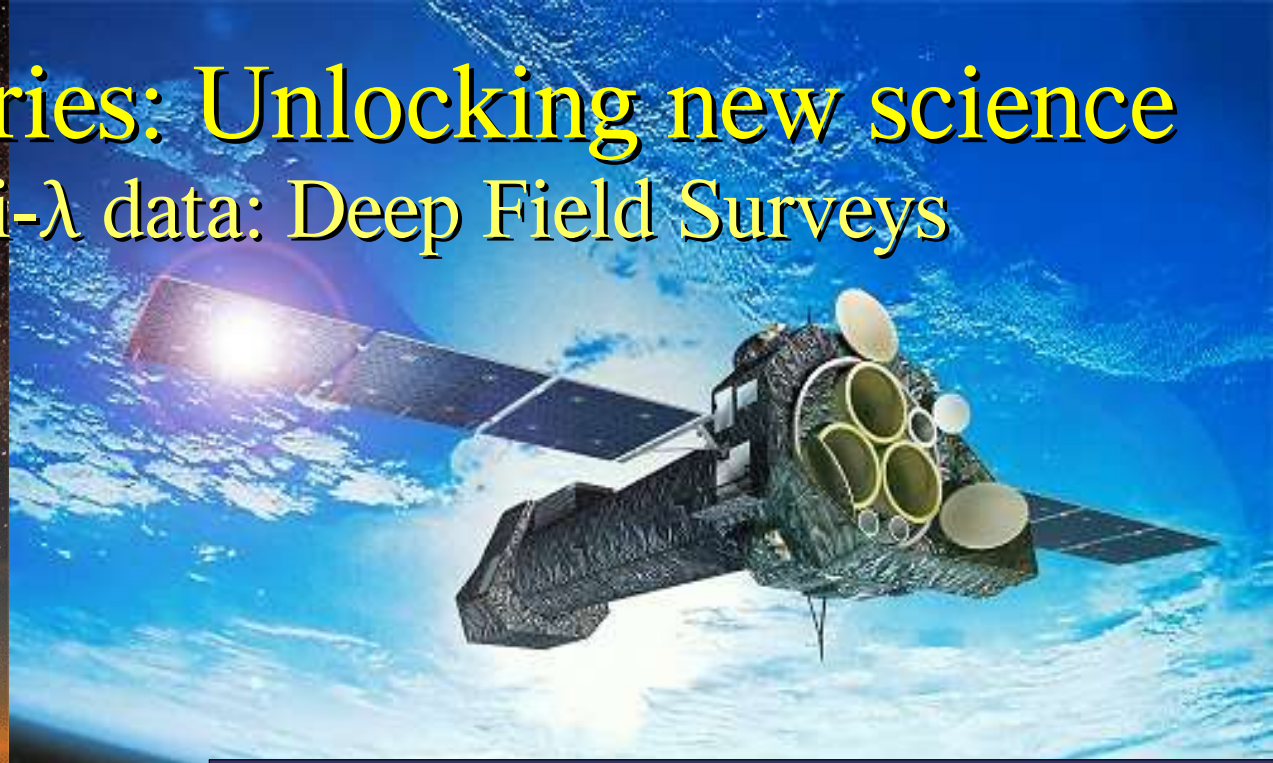
Images from Chandra Science Centre

Virtual Observatories: Unlocking new science

Federating multi-λ data: Deep Field Surveys

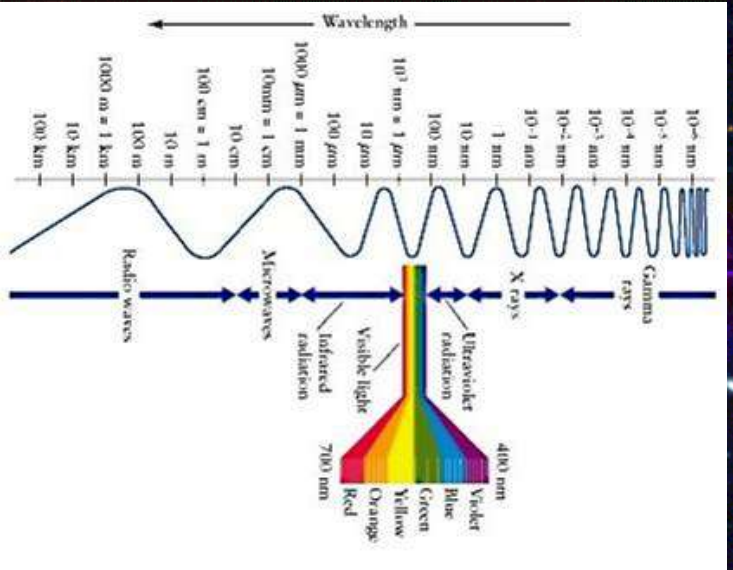


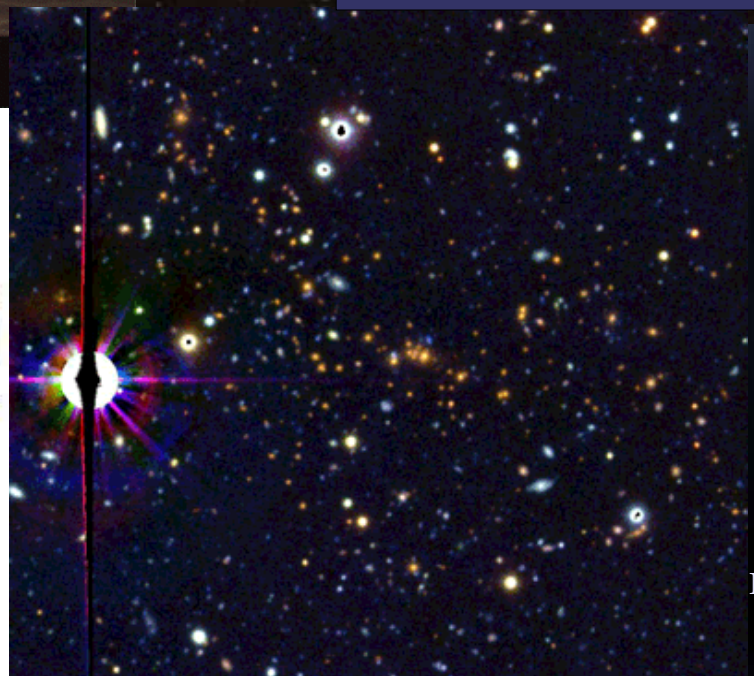
Nik Szymanek



D. Ducros, ESA

Link the X-ray and Opt/IR to understand energetic galaxies at the edge of the Universe





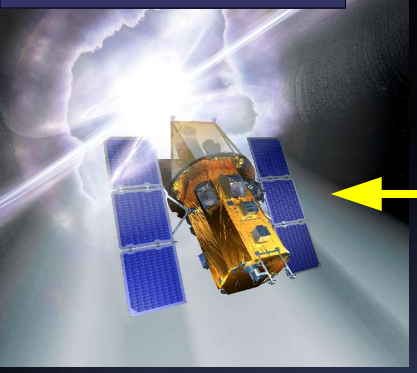
NEP J1716.6+6708: an X-ray cluster at z=0.81: Chandra X-ray image (C. Mullis) overlaid on a deep BRI image (D. Clowe & G. Luppino).

Printed: 18/12/03

AVO
ASTROPHYSICAL VIRTUAL OBSERVATORY

Gamma Ray Bursts

SWIFT satellite observes gamma ray burst



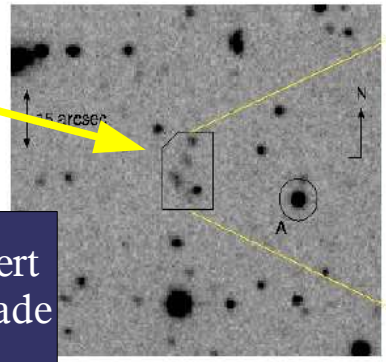
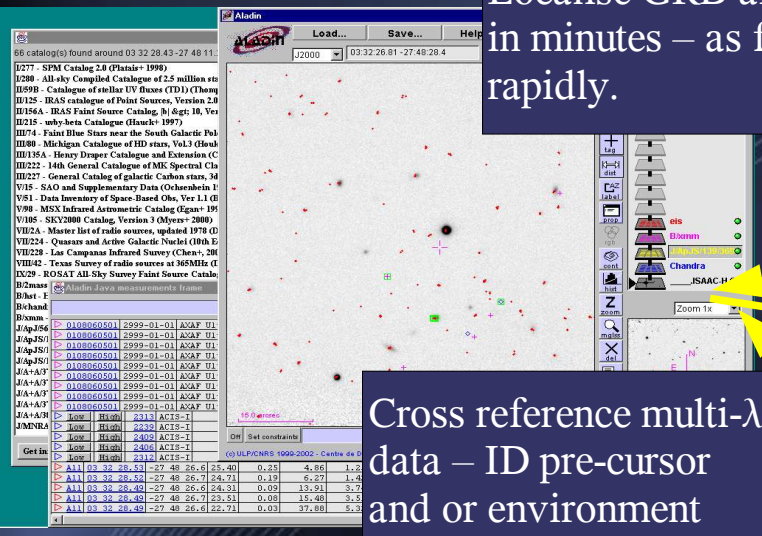


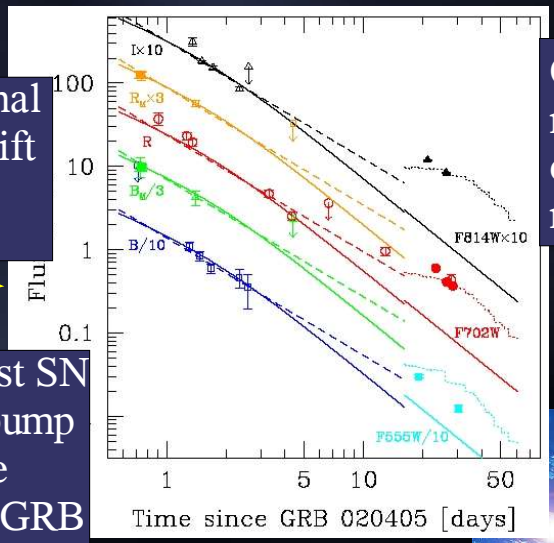
Image from ESO

Interaction with observatory pipe-lines

Localise GRB alert in minutes – as fade rapidly.



Large computational photometric redshift calcs on multi-λ > gives distance



Collate data from multiple telescopes over months - meta data issues

Cross reference multi-λ data – ID pre-cursor and or environment

Compare against SN light curves – bump shows evidence for a SN in the GRB (Price et al, 2002)

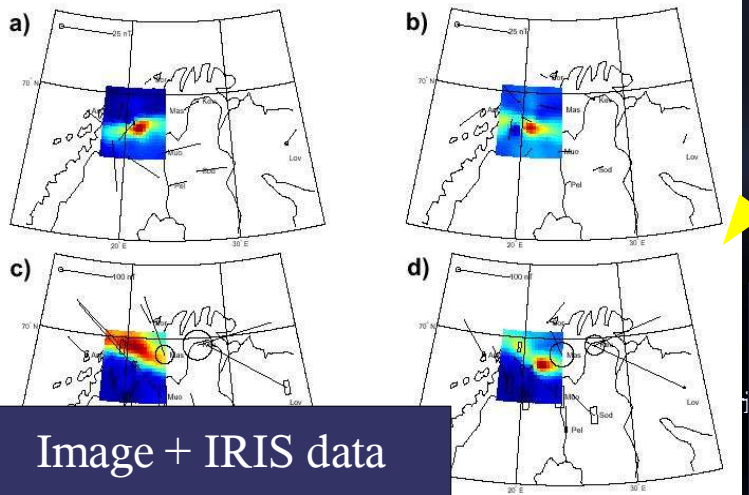


Image + IRIS data

Reprocessing of ionospheric STP data change coords from earth to celestial

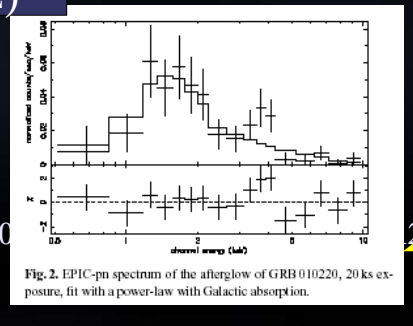
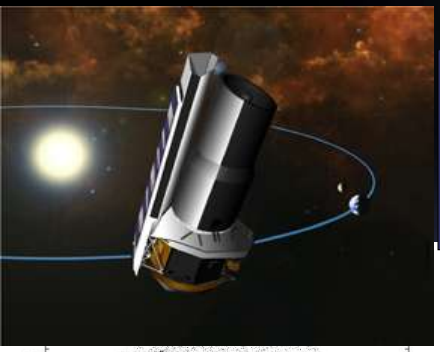


Fig. 2. EPIC-pn spectrum of the afterglow of GRB 010220, 20 ks exposure, fit with a power-law with Galactic absorption.

D. Ducros, ESA

AVO
ASTROPHYSICAL VIRTUAL OBSERVATORY

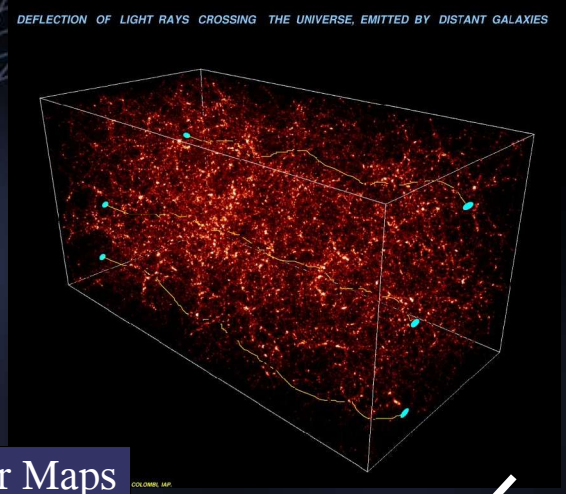
Dark Matter + Large Scale Structure

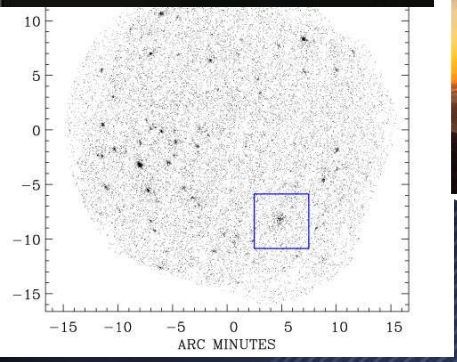


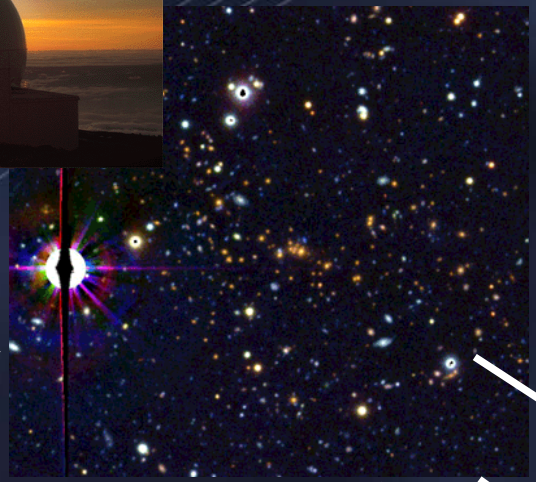
Multiple large image sources: registration & association

Automatic cluster finding techniques

Multi-TB λCDM models, e.g. Millennium Sim







X-ray cluster: Chandra X-ray (Mullis) overlaid on a deep BRI image (Clowe & Luppino).

Generate Shear Maps c.f. λCDM models > DM distribution with redshift

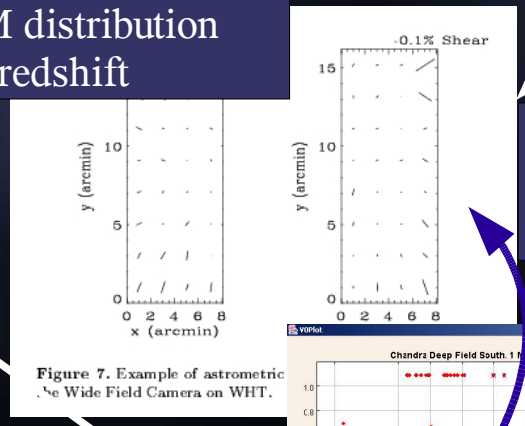
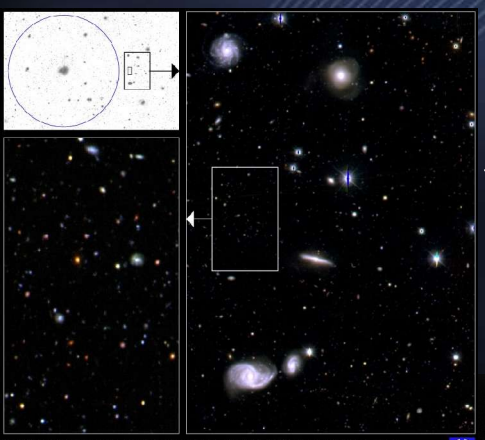
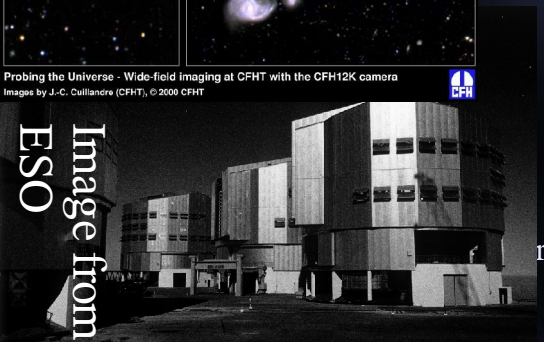


Figure 7. Example of astrometric ...he Wide Field Camera on WHT.

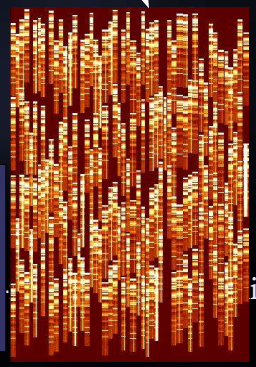
Remove star... correlate gal... with z

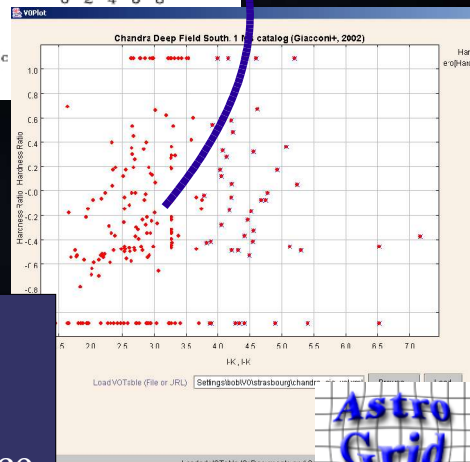




roGrid/EGEE/Pa... ...i

Source ID from multiplexed spectral data





Colour-Colour relationships classification in multi-phase space

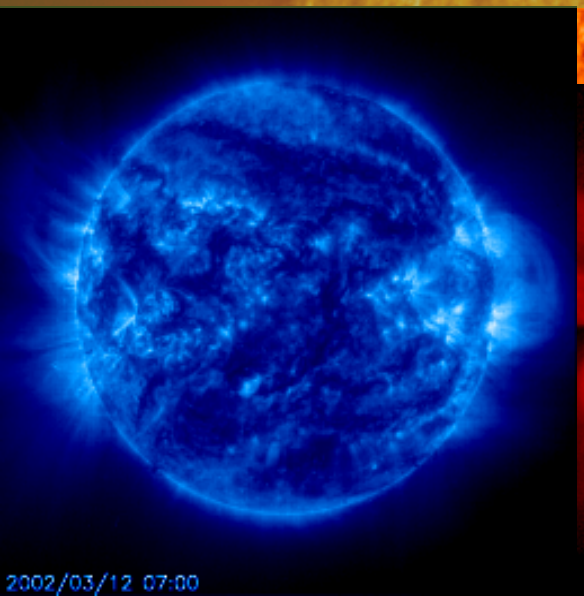
Astro Grid

Virtual Observatory: Understanding Linkages
A Solar-Terrestrial Model

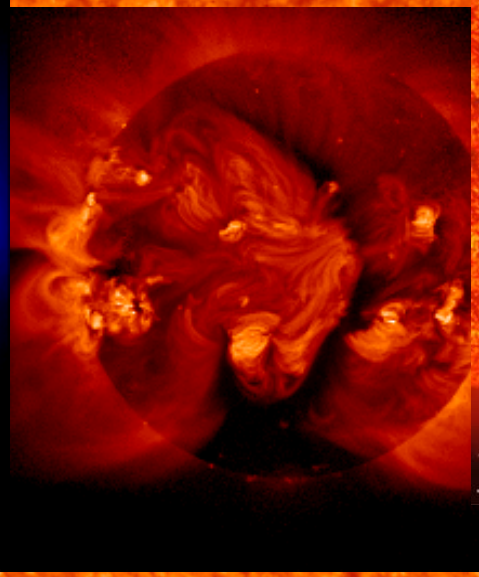
What happens to the Earth's magnetosphere during a coronal mass ejection ?

Event imaged by space based solar observatory

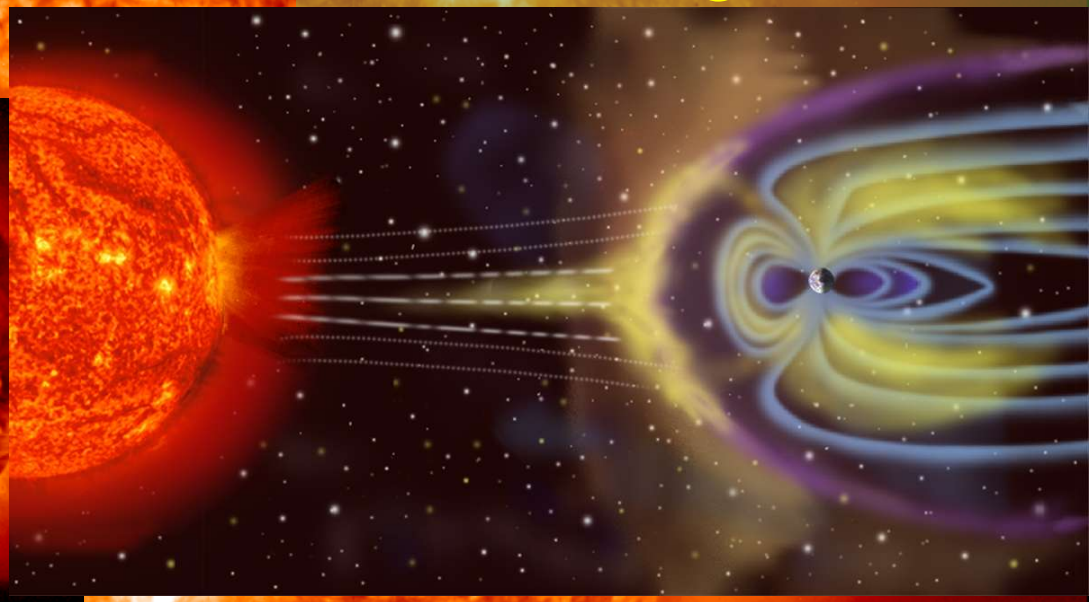
Effect detected later by satellites and ground radar



SOHO/EIT – EUV



Yohkoh – Xray

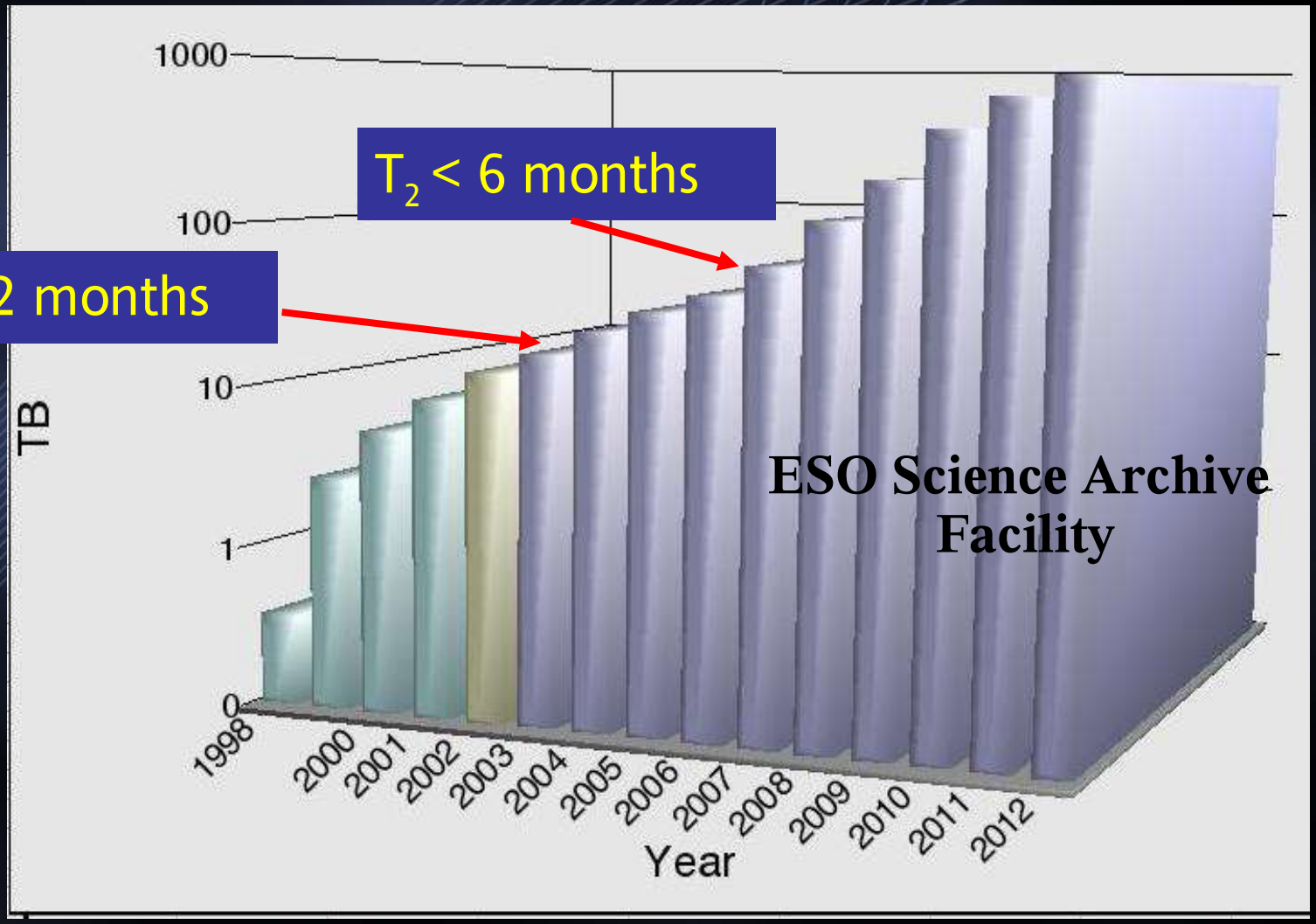


NASA: Living With a Star – http://lws.gsfc.nasa.gov

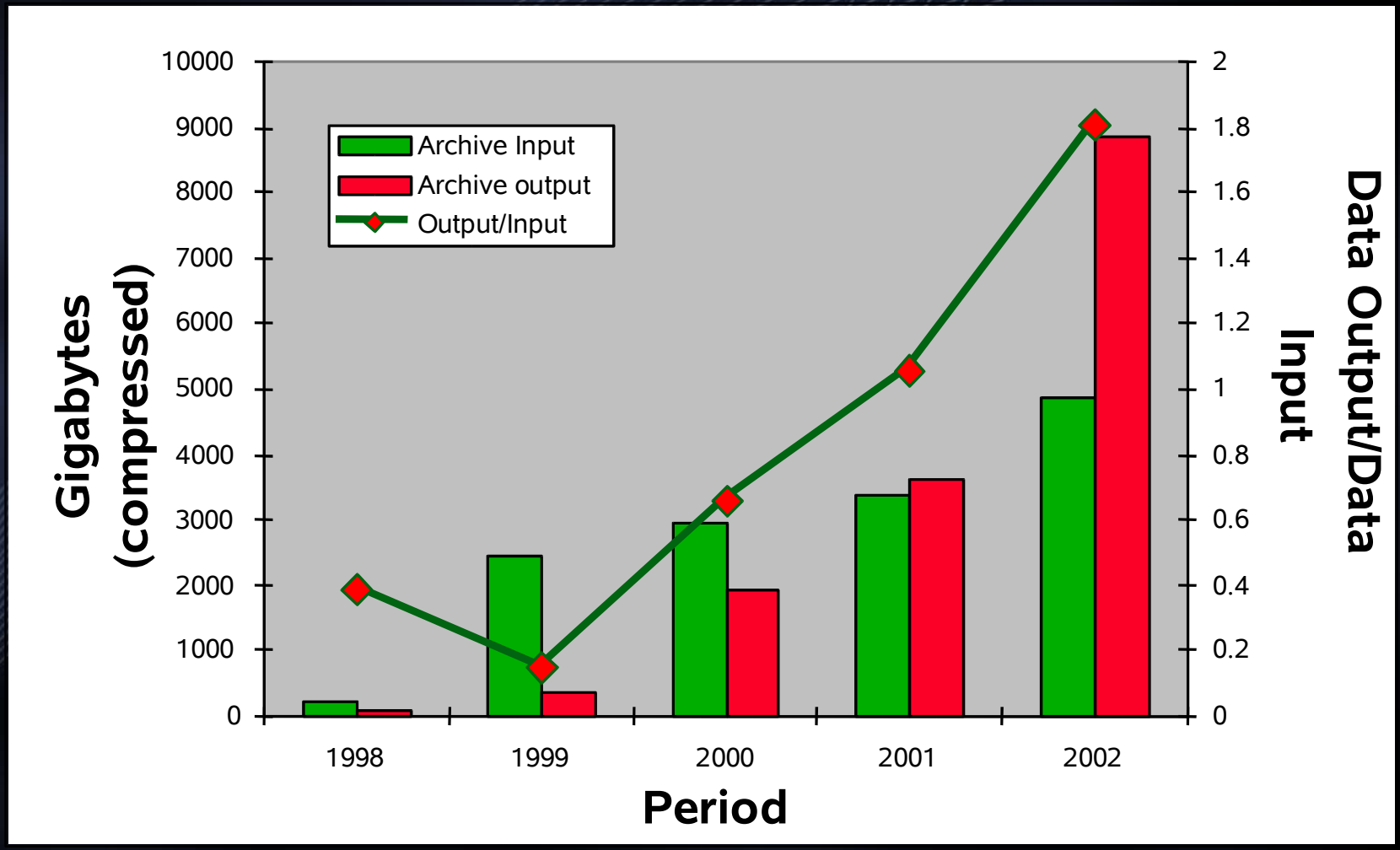
The Need for Virtual Observatories: Managing Technological Change

- The massive Growth of Data
  - Number + size of telescopes
    - Optical: ESO's 4x8m VLT, 2x8m Gemini
    - X-ray: XMM-Newton
    - sub-mm: ALMA
  - Increase in size and multiplex capabilities of instrumentation:
    - Infra-Red: VISTA > 100 GB/nights
    - Radio: e-Merlin > data rates ~320 Gbps

Astronomical Data Growth



Data Utilisation

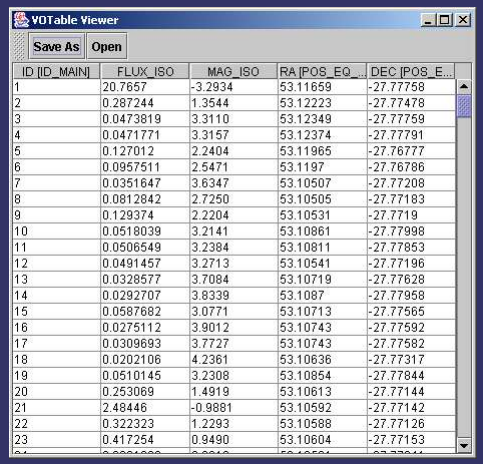


The Astrophysical Virtual Observatory

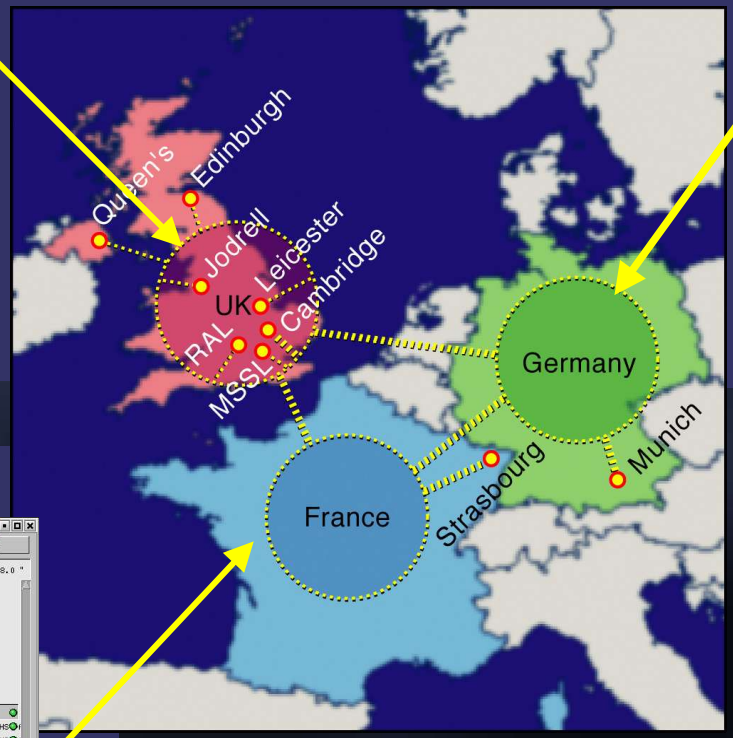
AVO 1st Light Prototype Components

an integrated one tool product

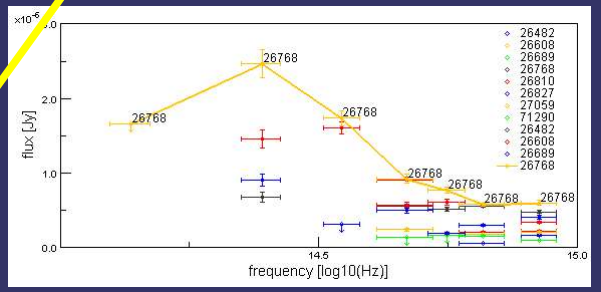
Astronomy Catalogue Extractor (ACE)



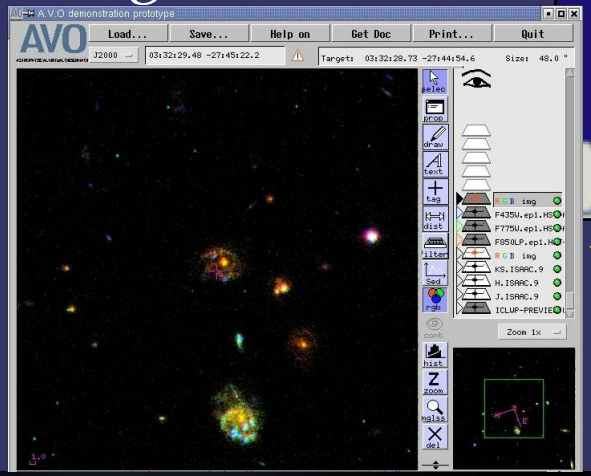
ID [ID_MAIN]	FLUX_ISO	MAG_ISO	RA [POS_EQ_...]	DEC [POS_E...]
1	20.7657	-3.2934	53.11659	-27.77758
2	0.287244	1.3544	53.12223	-27.77478
3	0.0473819	3.3110	53.12349	-27.77759
4	0.0471771	3.3157	53.12374	-27.77791
5	0.127012	2.2404	53.11965	-27.76777
6	0.0957511	2.5471	53.1197	-27.76786
7	0.0351647	3.6347	53.10507	-27.77208
8	0.0812842	2.7250	53.10505	-27.77183
9	0.129374	2.2204	53.10531	-27.7719
10	0.0518039	3.2141	53.10861	-27.77998
11	0.0506549	3.2384	53.10811	-27.77853
12	0.0491457	3.2713	53.10541	-27.77196
13	0.0328577	3.7084	53.10719	-27.77628
14	0.0292707	3.8339	53.1087	-27.77958
15	0.0587682	3.0771	53.10713	-27.77565
16	0.0275112	3.9012	53.10743	-27.77592
17	0.0309693	3.7727	53.10743	-27.77582
18	0.0202106	4.2361	53.10636	-27.77317
19	0.0510145	3.2308	53.10854	-27.77844
20	0.253069	1.4919	53.10613	-27.77144
21	2.48446	-0.9881	53.10592	-27.77142
22	0.322323	1.2293	53.10588	-27.77126
23	0.417254	0.9490	53.10604	-27.77153



Spectral Energy Distribution Utility

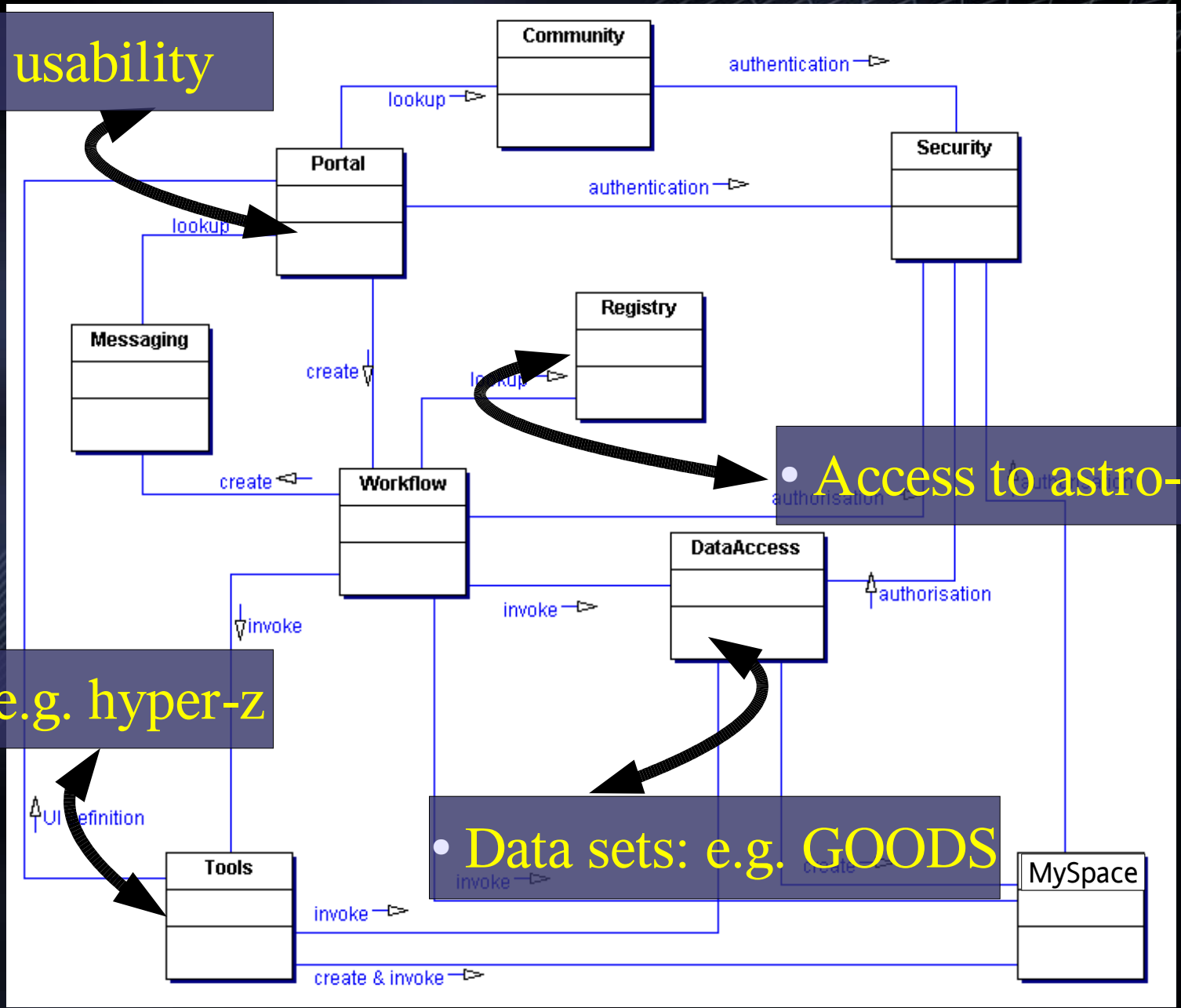


Data Integration



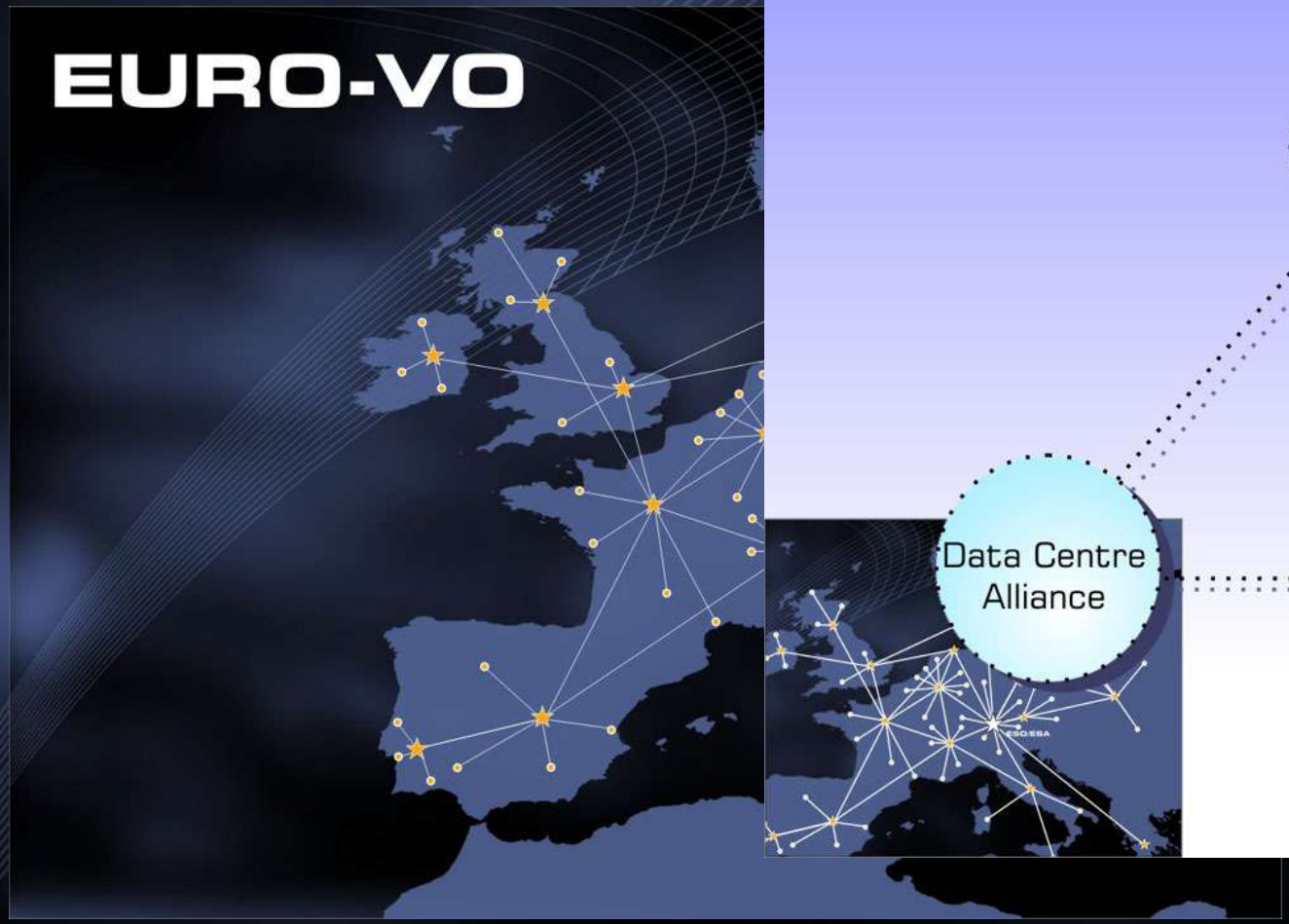
AstroGrid/AVO Service Model

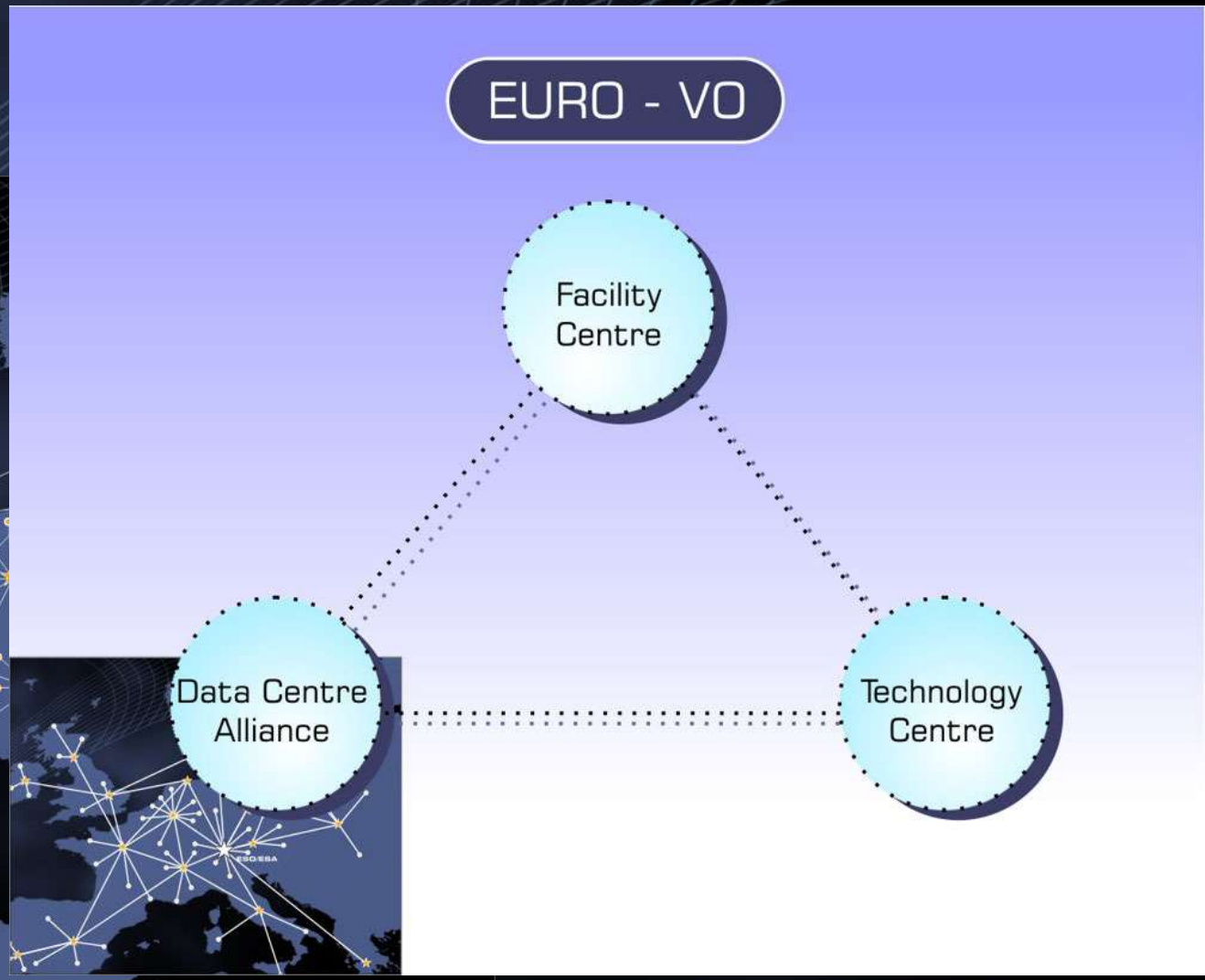
- GUI: usability



- Access to astro-data
- Tools: e.g. hyper-z
- Data sets: e.g. GOODS

The European Virtual Observatory





The International Virtual Observatory Alliance

Mission

To facilitate the international coordination and collaboration necessary for the development and deployment of the tools, systems and organisational structures necessary to enable the international utilisation of astronomical archives as an integrated and interoperating virtual observatory

IVOA: Developing Key VO Stds

Interoperability Standards

- Registries
- Data Models
- Uniform Content Descriptors
- Data Access Layer
- VO Query Language
- VOTable

http://www.ivoa.net/forum/

Astronomical Grid Computing Community
The Astro-RG @ Global Grid Forum

- IVOA backed Astronomy 'application area' at the GGF
  - Opportunity to influence the 'grid' community stds process
  - Shape deliverables from major providers: IBM, MS, etc
- Forming after successful kick-off BOF at GGF9
- Co-Chairs:
  - Reagan Moore (SDSC/ NVO)
  - Nic Walton (IoA/ AstroGrid)
- See http://www.ggf.org and GridForge pages for docs:
  - http://forge.gridforum.org/projects/astro-rg/
- Formal start at GGF10 in Frankfurt (spring 2004)

Scenarios for Grid use in VO's

Our options 3+4 fit EGEE

1. No Grid, no way!

2. Grid throughout.

3. Grid services as leaf nodes.

4. Leaf nodes + pervasive GSI + pervasive GridFTP.

Use of Grid Leaf Nodes

- Use standard, OGSA services (e.g. OGSA-DAI, MMJFS). Must be:
  - Supported
  - Standardised
  - High-quality
- Use OGSI, GridFTP, GSI to get in. Must have:
  - Multi-source, multiplatform middleware
  - High-quality middleware
  - Authorisation support

Astronomy Requirements on System

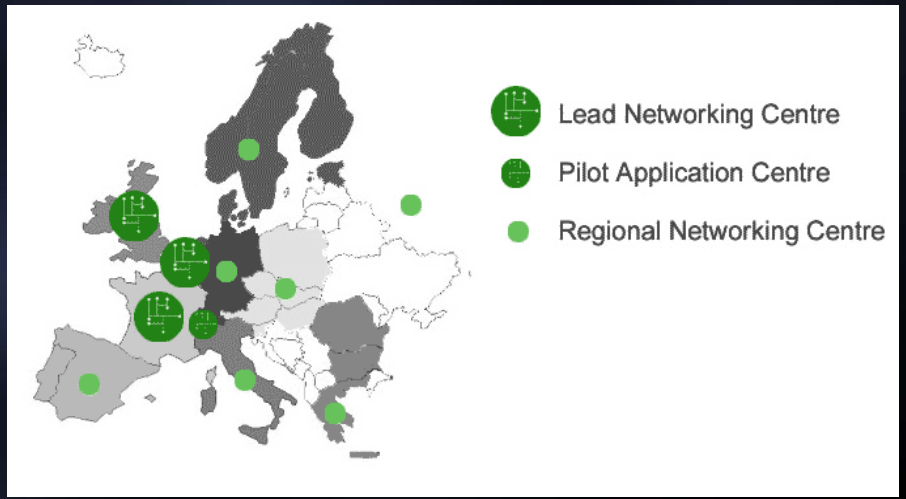
- System has to *work*!
- Has to be standardised (astro input at GGF).
- Has to be documented *properly*.
- Be efficient in developer time
- Know that OGSI etc. will survive on timescales ~ 5 years.
- *Quality*, not breadth of function.

EGEE addresses these issues: provides a R^3 system

A Perfect Partner for the VO

Key EGEE offer meets the needs of the EURO-VO

- On demand computing
- Pervasive access
- Large scale resources
- Sharing of software and data
- Improved support

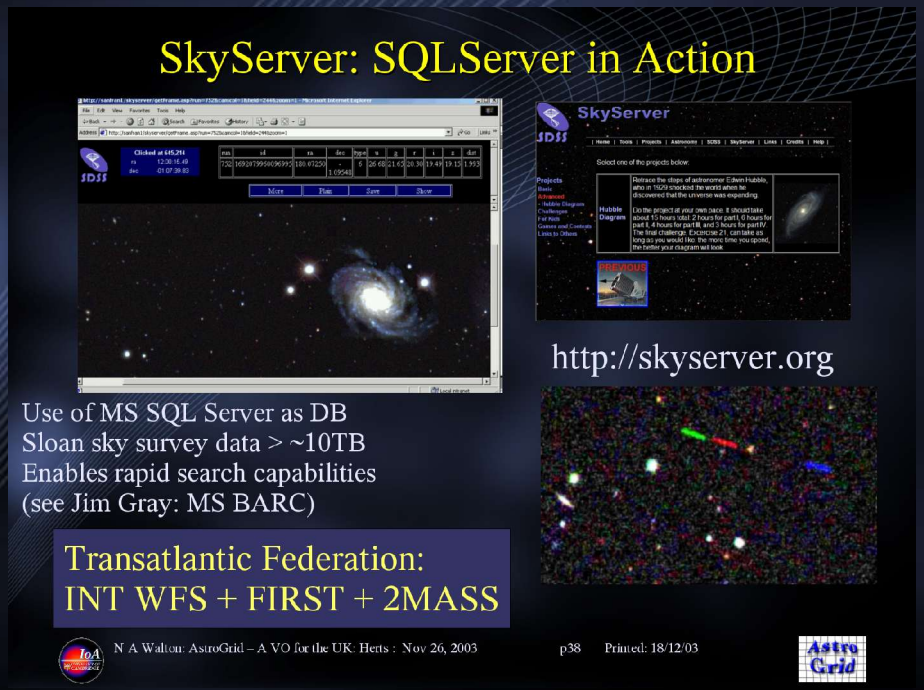


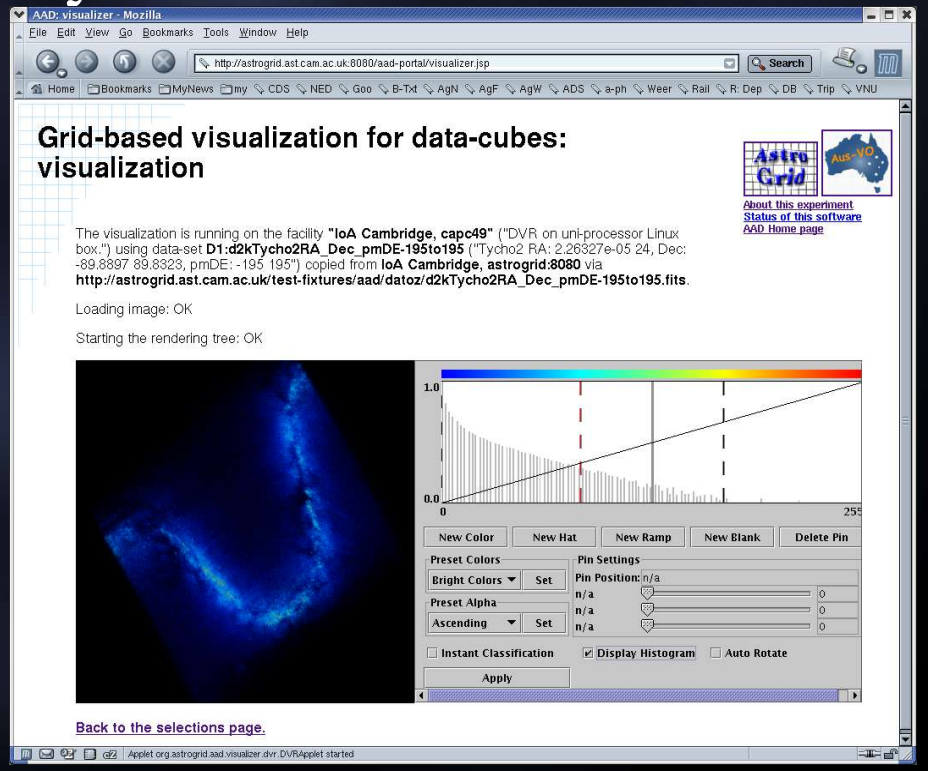
Euro-VO – Application Rich

- The Astrophysical community has a wide range of computational challenges to address, all of which potentially enabled by access to distributed EGEE resources:
  - Compute intensive
    - e.g. Mass scale image convolution, N-body simulations
  - Data intensive
    - e.g Mining large scale sky surveys – petabyte data systems
  - Network intensive
    - e.g. Federation of heterogeneous, distributed data resources
  - All of the above
    - e.g. Instrument data reduction pipelines

Complex Data Mining/ Visualisation: Application Example 1

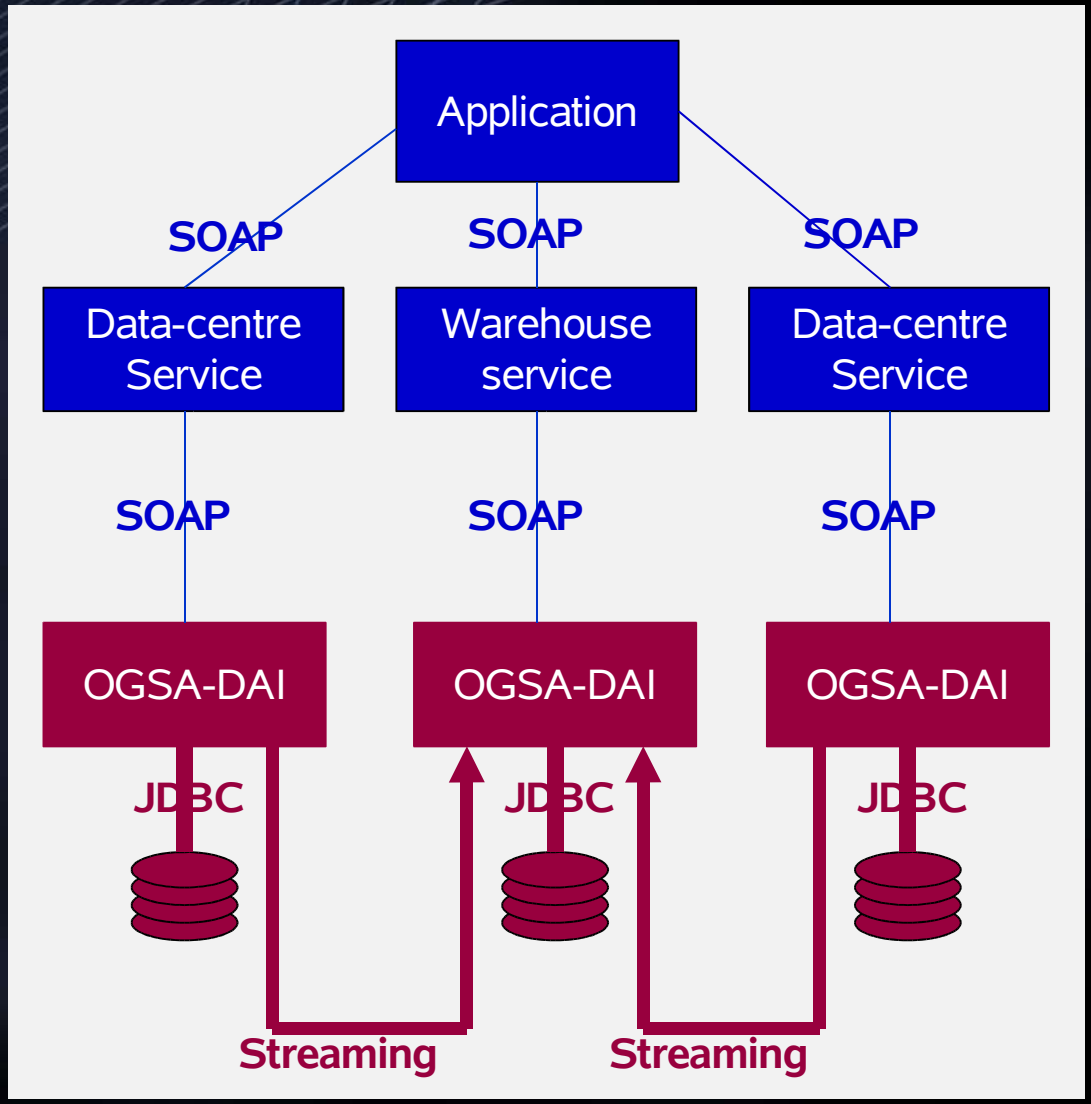
- Mine large heterogeneous databases: e.g. ESO, ALMA, OWL, GAIA, etc
  - Some queries > creation of dynamic data warehouses
  - Ideal for use on large e.g. Petabyte Store
- Server based visualisation





Data grid for DB tables: Implementation Scenario

- AstroGrid/AVO creating a Grid Data Warehouse – deploying on UK eScience Grid – component of euro wide EGEE structure
- Data grid (red) complements compute grid (blue).
- OGSA-DAI instances stream DB tables.
- Tables accumulate in warehouse DB for data-mining.
- OGSA-DAI runs on any grid; may be remote from astronomy services.



Astro Missions

- Euro-VO provides access to data, information from all major astronomy missions
- Missions produce mission specific data products
- Science exploitation
  - Mission specific products
  - Mining multi-mission products
- Strong programme in Europe
  - ESO – 4x8m telescopes
  - OWL – 100-m optical telescope (2015)
  - Planck – precision CMB (2007/8) Example mission
  - GAIA – precision astrometry (2010-12)

Planck: A Possible Scenario

Enabling Grids for E-science in Europe

Planck simulations and processing are an ideal application to evaluate the power of the European Grid Infrastructure

Grid-enabled environment: possible scenario

- The Planck user requests to download, through a user interface, a specific set of all-sky simulated data under certain scientific hypotheses, and using a selected mission and instrument configuration
- The environment understands if such a simulation has been already produced and, if so, it allows the user to access the data
- If no data are available, then suitable computing facilities should be selected from a pool of available resources to produce the data the user will eventually be able to access
- Data could be processed locally or, if needed, in a distributed way throughout the Grid once again by selecting the computing facilities from those available on the Grid infrastructure

C. Vuerli, Astrophysical Technologies Group (ATG), INAF - OATs
18 december 2003

ASTROPHYSICAL VIRTUAL OBSERVATORY

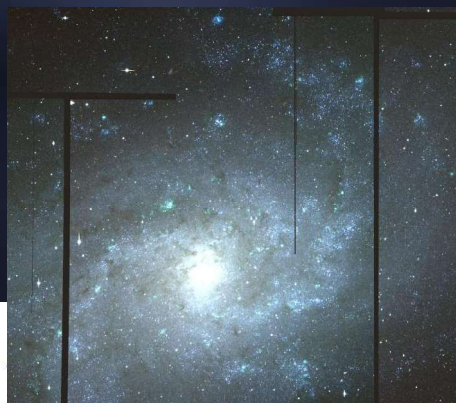
Planck: Added Value

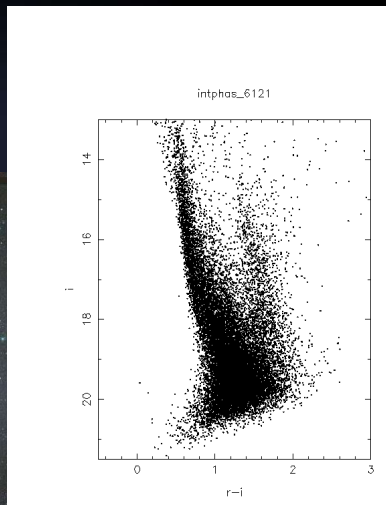
- Simulations code and produced simulated data are transparently and easily accessible to the Planck community through the Grid User Interface. Planck users can ask for specific simulated data and, in case, run the application (the Simulations Pipeline) to produce them
- Pipeline runs may require considerable computing power capabilities. By using the Grid, pipeline runs will be disseminated over the continental Grid infrastructure so that computing power shortages of single institutes can be easily overcome. Moreover, the exploitation of computing resources is optimized
- Simulated data may be remarkable in size (e.g. frequency and component maps, and especially time series may be huge). Simulations results may be transparently spread over different SEs, and, from there, retrieved by the gridified Pipeline
- Because of their intrinsic parallelism, simulations applications should gain great advantage when run over the Grid infrastructure

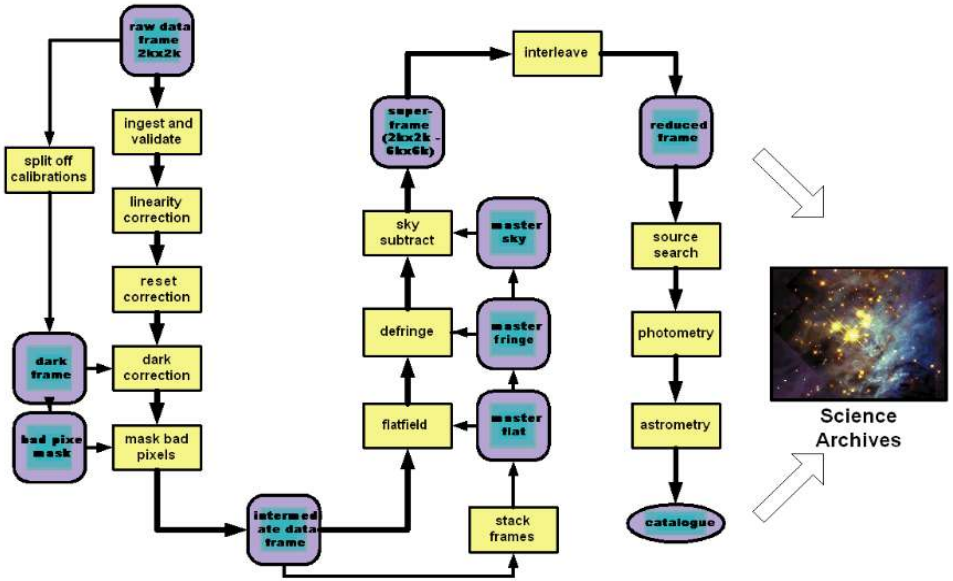
C. Vuerli, Astrophysical Technologies Group (ATG), INAF - OATs
18 december 2003

VO Pipeline Access: Application Example 2

- Major new facilities – advanced processing pipelines
  - e.g. IR survey telescopes VISTA/WFCAM: >100Gb/night – compute intensive
- Run on dedicated h/w, but for …
- Peak demand
  - Offload to ext. resources
- Process coarse grained
  - 'beowulf' ready
- Meet demands of user runs
  - Distributed user base







A Perfect Partner for the VO

Key EGEE offer meets the needs of the EURO-VO

- On demand computing: astro demands on dynamic workflows support via peak load access
- Pervasive access: supports Euro distributed community
- Large scale resources: meets the heavy 'peta' scale demands from astronomy
- Sharing of software and data: confluence with VO s/w
- Improved support: interact with Euro-VO support

The Astrophysical Community and the EGEE

Nicholas A Walton
AstroGrid Project Scientist
University of Cambridge