UK e-Science:
A Vision from Outer Space

Lourdes Verdes-Montenegro

Instituto de Astrofísica de Andalucía (IAA-CSIC, Granada, Spain)

Outline

Outline

My e-Path

Outline

My e-Path

UK e-Science

Outline

My e-Path

UK e-Science

e-Astronomy

Outline

My e-Path

UK e-Science

e-Astronomy

e-Precedents


UK e-Science
e-Astronomy
Outline
UK e-Science
e-Astronomy
e-Precedents
My e-Path
UK e-Science
My e-Path
UK e-Science
e-Precedents
e-Astronomy
My e-Path
e-Precedents
e-Astronomy
e-Astronomy
e-Precedents
My e-Path
UK e-Science
e-Precedents

2003: A “sociological” study of galaxies

HST image (NASA, ESA)

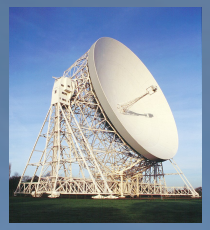
2003: A “sociological” study of galaxies

HST image (NASA, ESA)

2003: A “sociological” study of galaxies

- Objets in the Universe faint and complex
- Gas, stars, dust...
- Different physical conditions

Multiwavelength studies of large samples

2003: A "sociological" study of galaxies

- Objets in the Universe faint and complex
- Gas, stars, dust...
- Different physical conditions

Multiwavelength studies of large samples

1000 ISOLATED GALAXIES

NGC 7217

AMIGA PROJECT

Analysis of Interstellar **M**edium of **I**solated **GA**laxies

MAIN AIM:

NGC 1512 NASA, ESA, Dan Maoz

- To build & analyze a multi-wavelength data base
- Special emphasis on radioastronomy

- To serve as a reference sample

Archive/bibliographic search was a nightmare

MAIN AIM:

NGC 1512 NASA, ESA, Dan Maoz

- To build & analyze a multi-wavelength data base
- Special emphasis in radioastronomy

- To serve as a reference sample

Archive/bibliographic search was a nightmare

MAIN AIM:

NGC 1512 NASA, ESA, Dan Maoz

- To build & analyze a multi-wavelength data base
- Special emphasis in radioastronomy
- To serve as a reference sample

Archive/bibliographic search was a nightmare

The Virtual Observatory

IVOA

Exploitation of astronomical archives:

Interoperability

- Astronomical archives built independently
- Querying to multiple databases painful

Exploitation of astronomical archives:

Interoperability

- Astronomical archives built independently
- Querying to multiple databases painful

- Federation of astronomical databases with homogeneous access protocols
- Need to conform to specific data models

By the way... not such a new problem

By the way... not such a new problem

Liberté,
egalité,
fraternité

By the way... not such a new problem

Liberté,
egalité,
fraternité

and ... Metric System

JUNE 2002

IVOA International Virtual Observatory Alliance

JUNE 2002

IVOA International Virtual Observatory Alliance

2001-2002

JUNE 2002

IVOA International Virtual Observatory Alliance

2001-2002

2004

e-Precedents

2001: earliest e-Science in UK?

Created by John Taylor, the DG of the UK's Office of Science and Technology in 1999

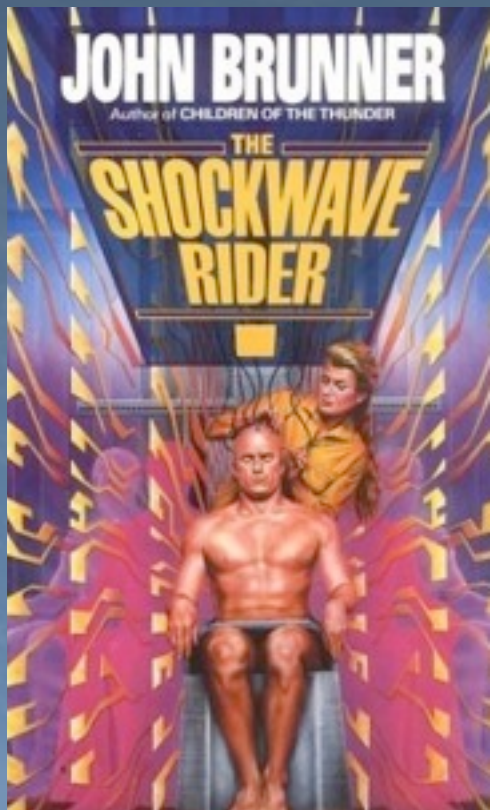
e-Precedents

2001: earliest e-Science in UK?

Created by John Taylor, the DG of the UK's Office of Science and Technology in 1999

But... earlier?

1975. The Shockwave Rider
Story of a future world **(2010)** tied together by a **universal data network**, filled with information overload and corporate domination, and nearly **everything is known about everybody**



2005

My e-Path

JORNADA e-CIENCIA EN ESPAÑA

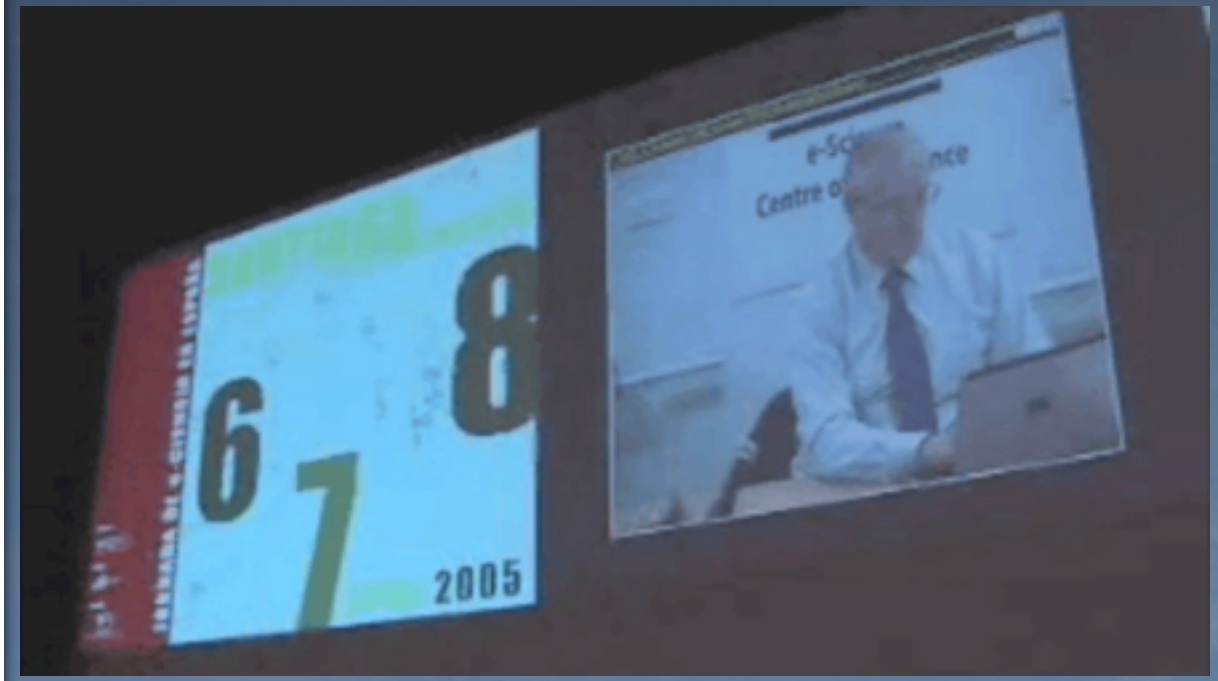
6, 7 y 8 de julio | SANTIAGO DE COMPOSTELA

2005

JORNADA e-CIENCIA EN ESPAÑA

6, 7 y 8 de julio | SANTIAGO DE COMPOSTELA

My e-Path



The consequences...

2006-2007:
Applied for funds, got regional Excellence Project

- e-Science needs of Andalusian groups not yet coordinated

- From +70 participants in 15 research groups + industry to 75 groups in 3 years

- Pioneer regional + bottom-up approach

My e-Path

The consequences...

2010

My e-Path

The consequences...

2010

My e-Path

The consequences...

2010

My e-Path

The consequences...

2010

Anne Trefethen
Oxford e-Research Centre (U. Oxford, Reino Unido)

Guy Wormser
L'institut des Grilles (CNRS, Francia)

Alyssa Goodman
Initiative in Innovative Computing (U. Harvard, EEUU)

Tony Hey
Microsoft External Research Division

Vicente Hernández
Red Nacional de e-Ciencia (UPV)

Javier García Tobío
Centro de Supercomputación de Galicia (CESGA)

2007

- Spanish Network for e-Science supported by Spanish Ministry of Science & Education
- > 900 researchers, 97 groups, 57 institutions
- Coordinated by Vicente Hernández García
- Main mandate: to set up a National Grid Initiative

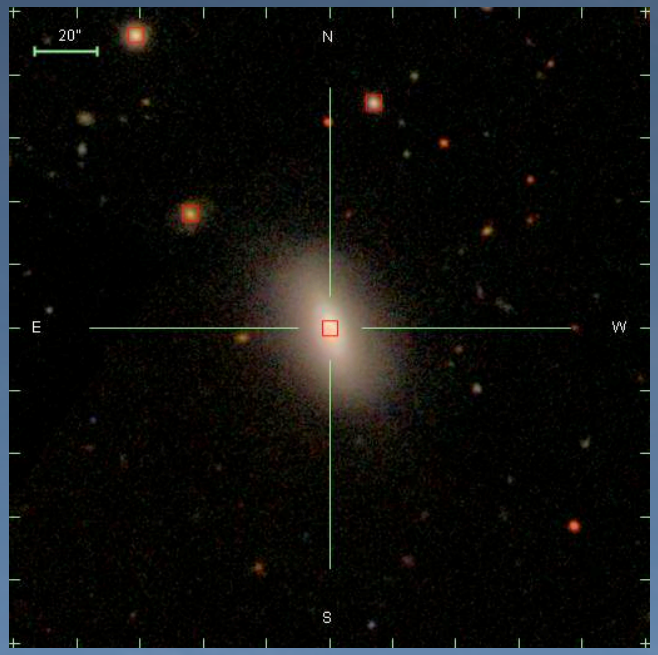
2008

- GRID - CSIC

My e-Path

SLGRID: Starlight code for spectral synthesis running in GRID

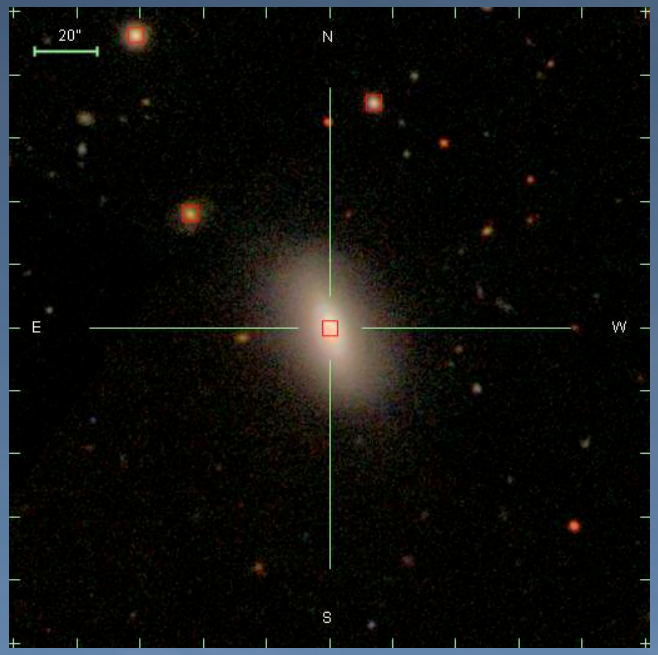
Aim: Search for supermassive BHs in AMIGA galaxies

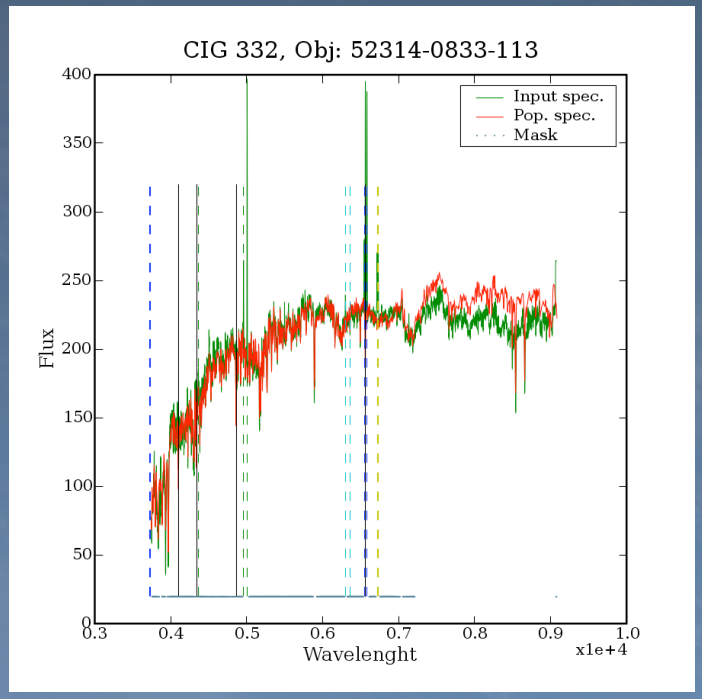


Starlight (Cid-Fernandes et al. 2005; http://www.starlight.ufsc.br/)

SLGRID: Starlight code for spectral synthesis running in GRID

Aim: Search for supermassive BHs in AMIGA galaxies

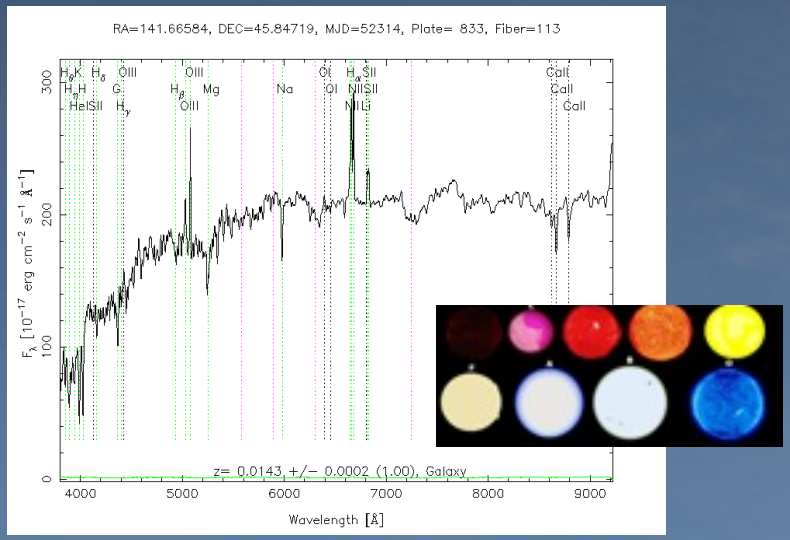




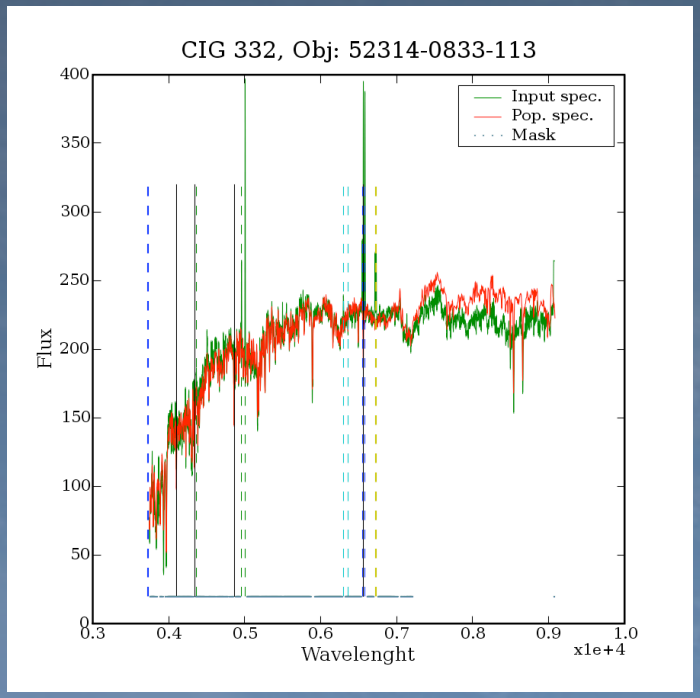
Starlight (Cid-Fernandes et al. 2005; http://www.starlight.ufsc.br/)

SLGRID: Starlight code for spectral synthesis running on the GRID

Aim: Search for supermassive BHs in AMIGA galaxies



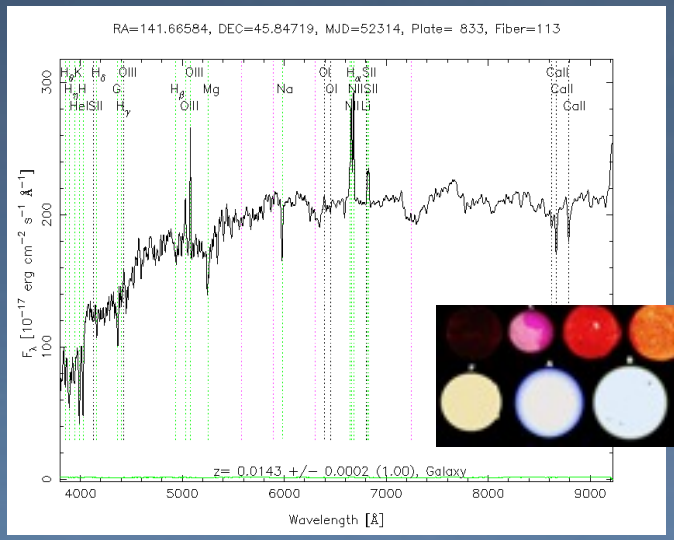
Database of stellar populations



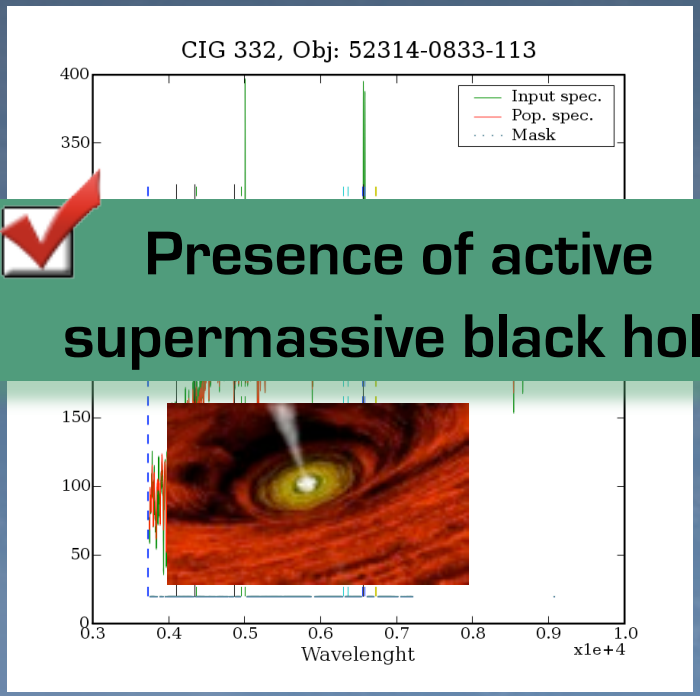
Starlight (Cid-Fernandes et al. 2005; http://www.starlight.ufsc.br/)

SLGRID: Starlight code for spectral synthesis running on the GRID

Aim: Search for supermassive BHs in AMIGA galaxies



Database of stellar populations

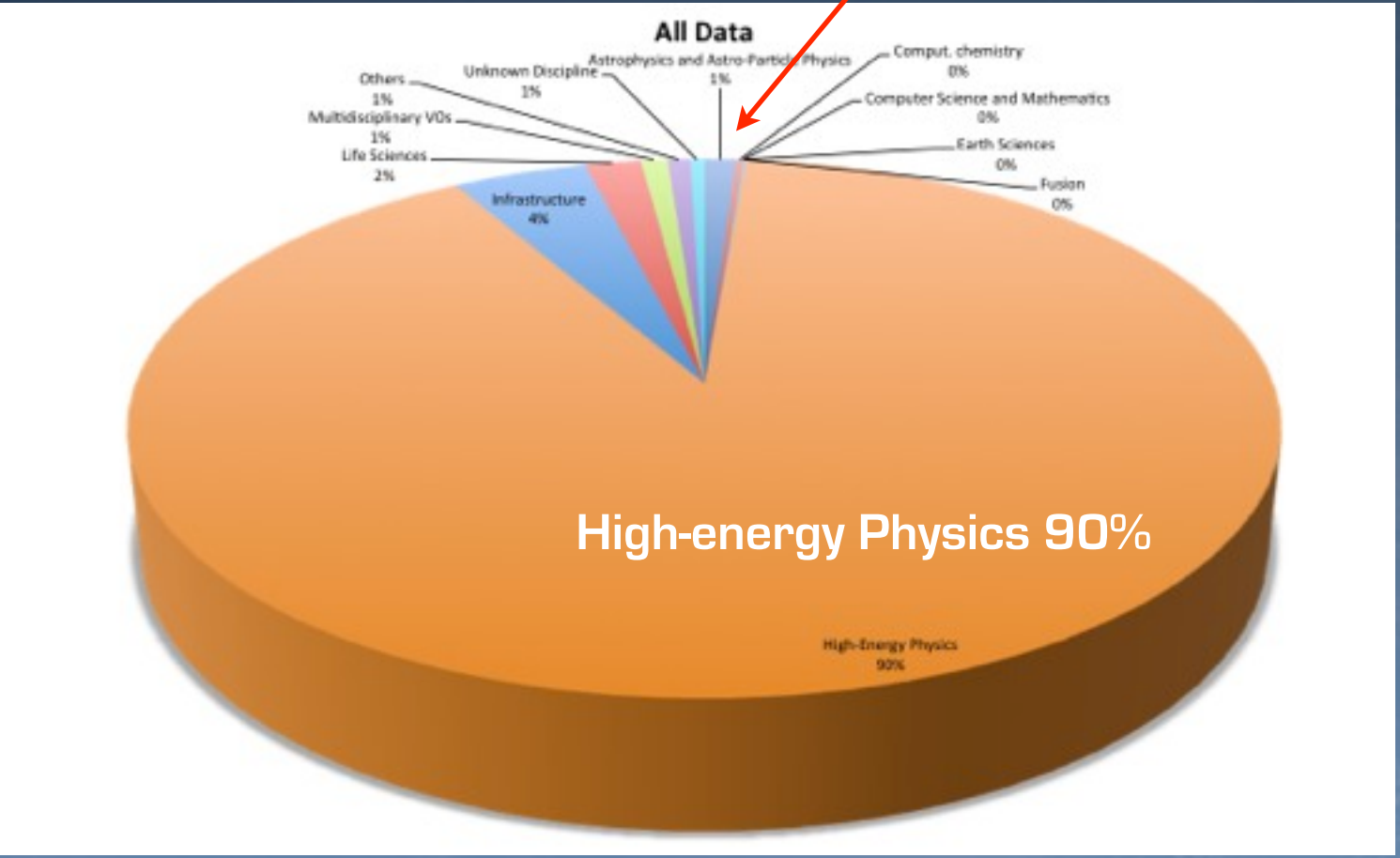


Presence of active supermassive black hole

Starlight (Cid-Fernandes et al. 2005; http://www.starlight.ufsc.br/)

Use of European Grid Infrastructure

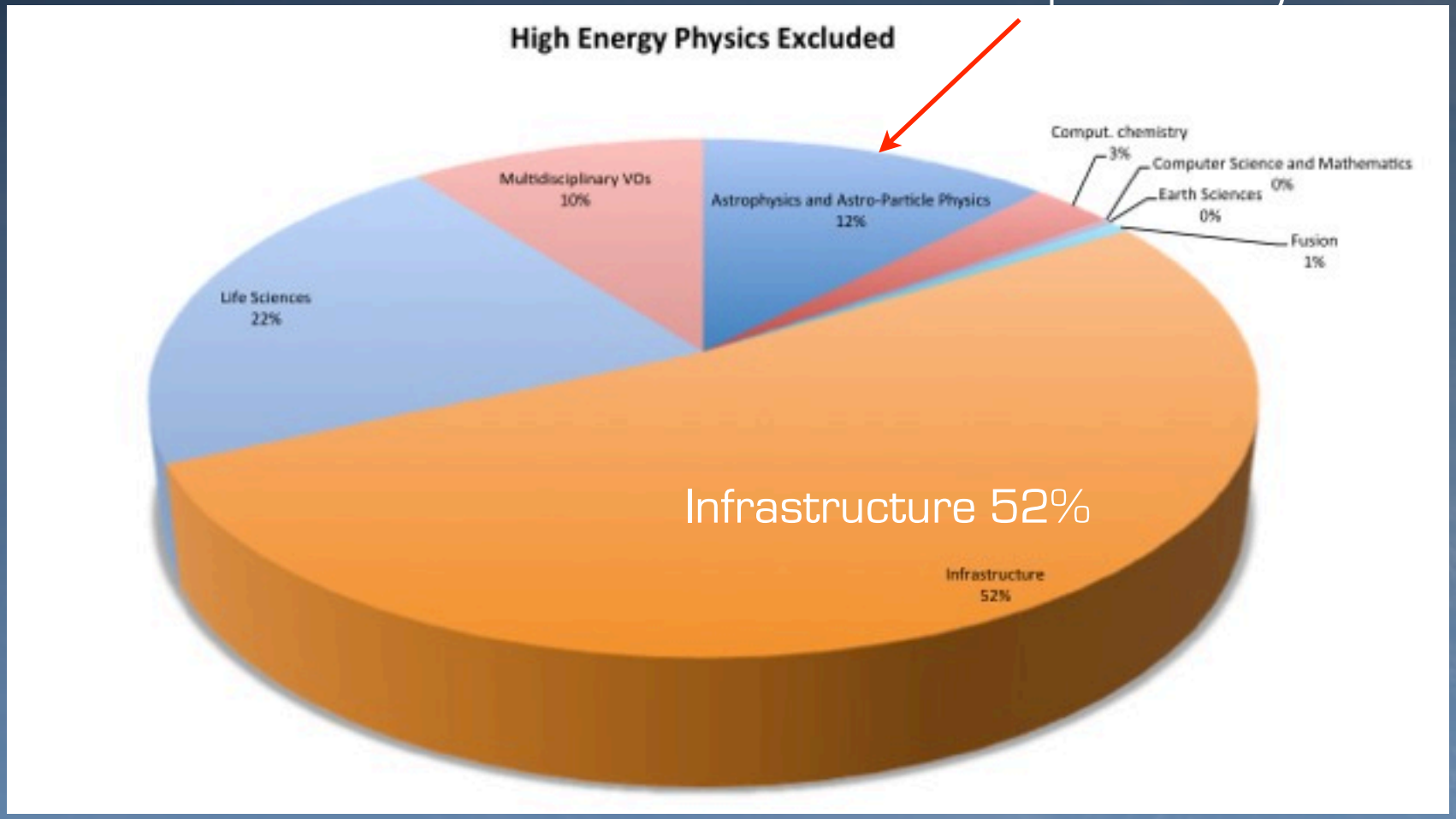
Astrophysics & Astroparticle Physics 1%



Job submissions: June 2010 - May 2011
Obtained using EGI statistics tool http://accounting.egi.eu/

Use of European Grid Infrastructure

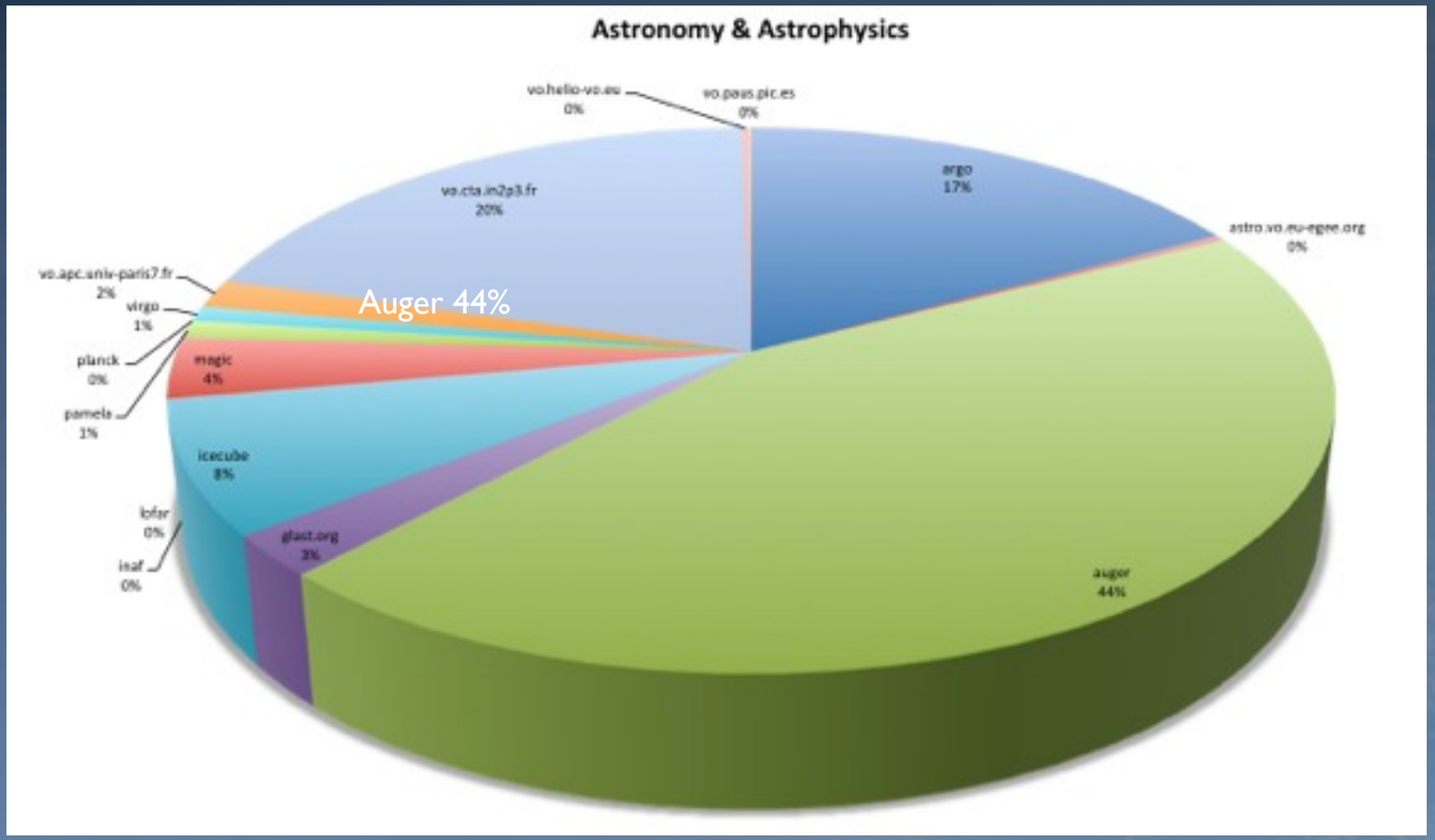
Astrophysics & Astroparticle Physics 14%



Job submissions: June 2010 - May 2011

Obtained using EGI statistics tool http://accounting.egi.eu/

European e-infrastructure not used much by the A&A community



e-Astronomy

Numerical simulations of galaxies

- AMIGA is about galaxy evolution and environment

My e-Path

- Numerical simulations help to identify features produced by interactions
- Particles representing gas+stars+etc
  - Evolving with time
    - Under physical laws

GRID computing in Astronomy

GRID computing in Astronomy

The Antennae, KPNO

Bob and Bill Twardy/Adam Block/NOAO/AURA/NSF

The Antennae, KPNO

Bob and Bill Twardy/Adam Block/NOAO/AURA/NSF

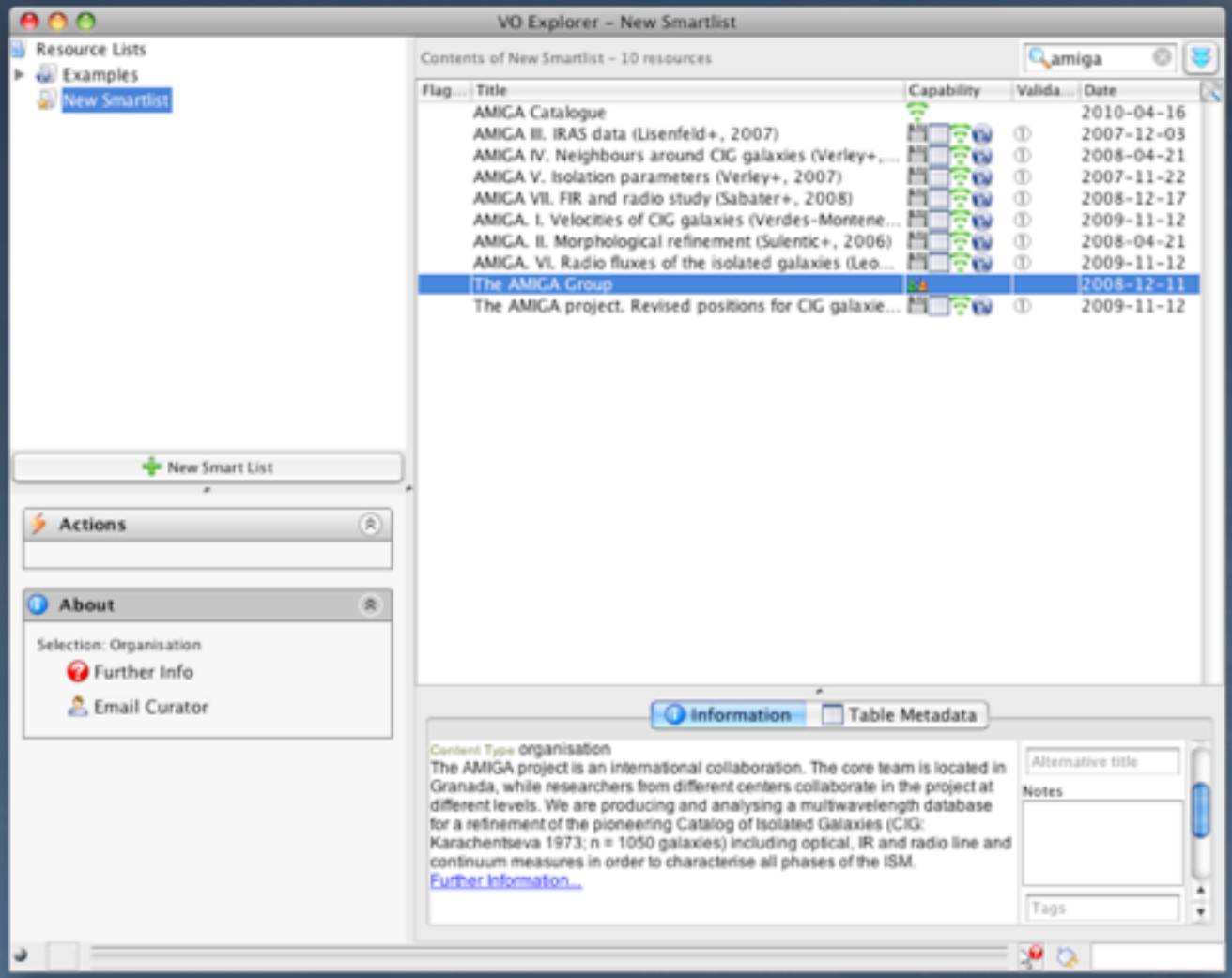
C. Mihos, L. Hernquist

C. Mihos, L. Hernquist

2006:

From publication...

VO Desktop
Astrogrid



2006:

From publication... to the design of Virtual Observatory standards for radioastronomy

MOST VO EFFORTS SO FAR IN OPTICAL ASTRONOMY

2006:

From publication... to design of VO standards for radioastronomy

- 1st VO compliant data model for radio
- Implemented at IRAM-30m, one of today's largest and most sensitive mm telescopes

2006:

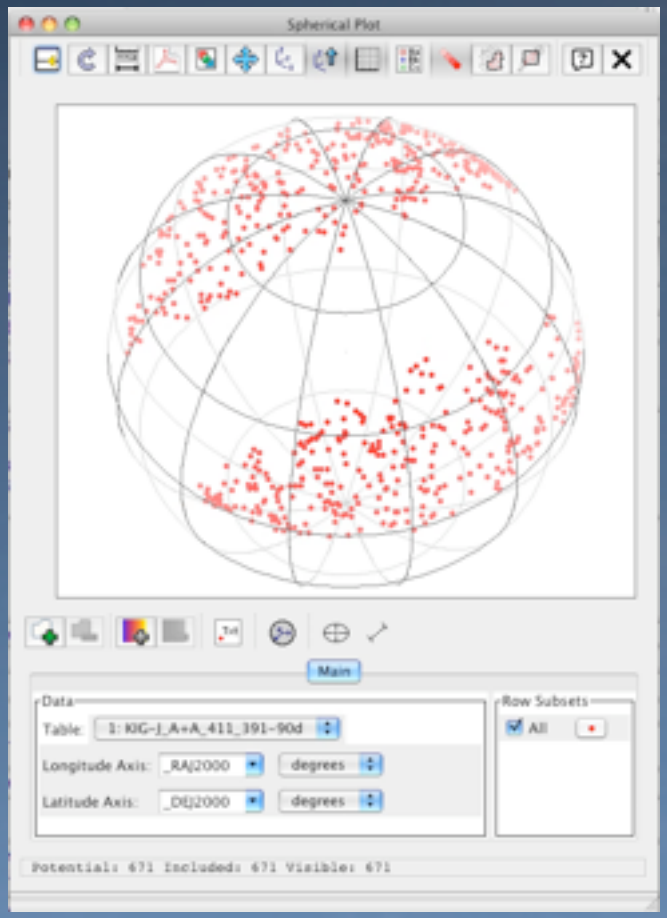
From publication... to design of VO standards for radioastronomy

- 1st VO compliant data model for radio
- Implemented at IRAM-30m, one of today's largest and most sensitive mm telescopes

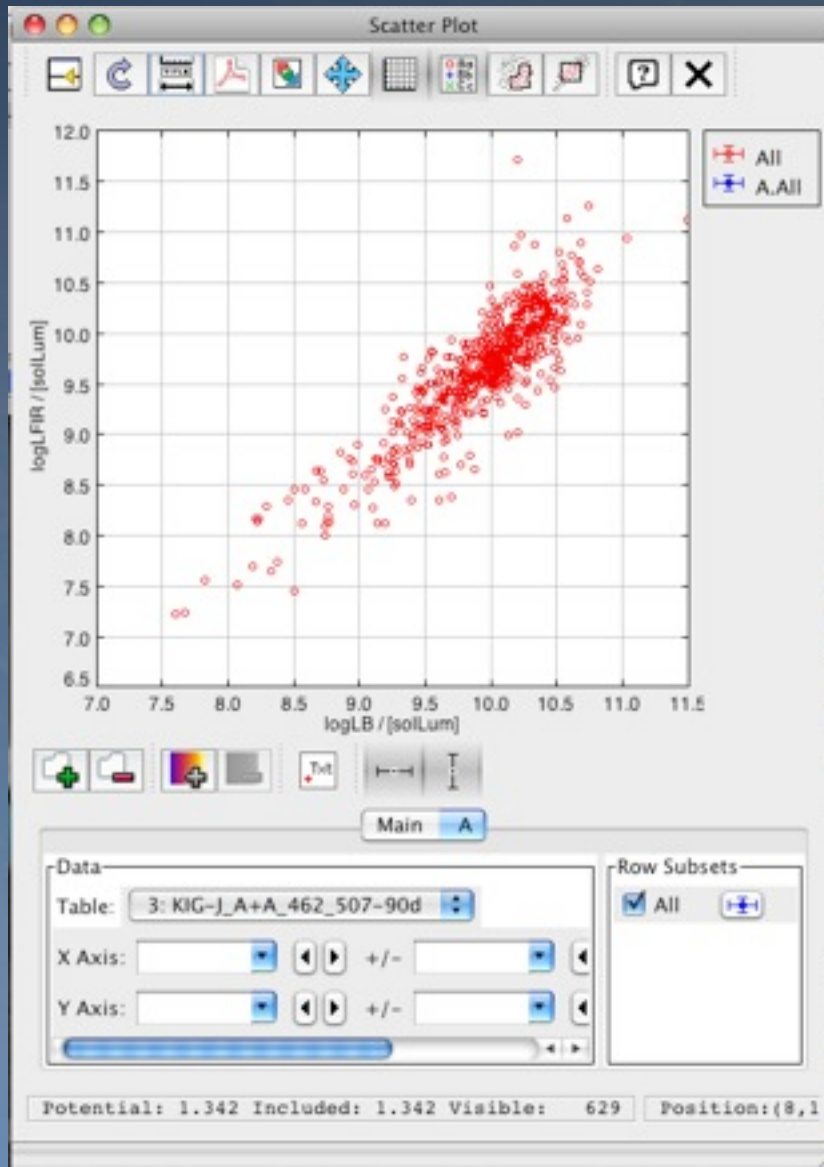
PhD 2009

UK e-Science

Users of tools

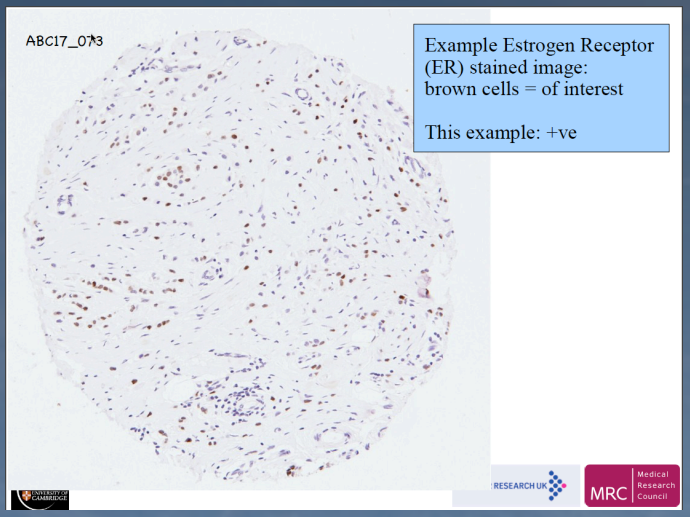


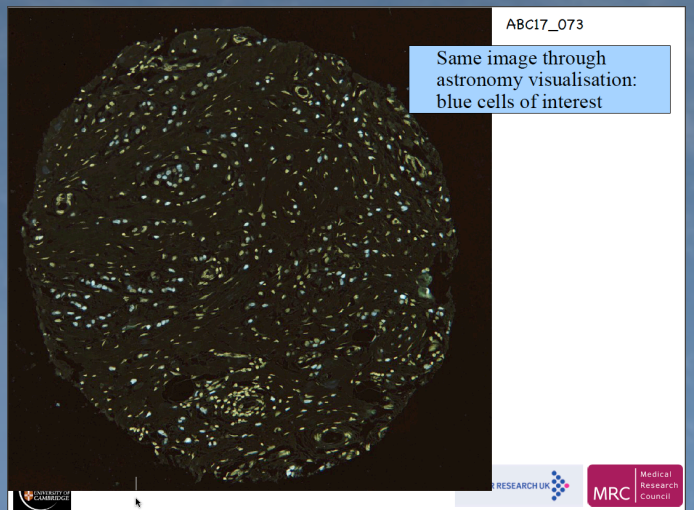
TOPCAT
Astrogrid



PathGRID: VO Tools used beyond Astronomy!

Analysis of Tissue Micro Array histopathology image data





AstroGrid:

- image analysis pipeline
- graphical workflows via Astro-Taverna
- scripting via Python

From Nic's Walton talk at UK All hands 2009

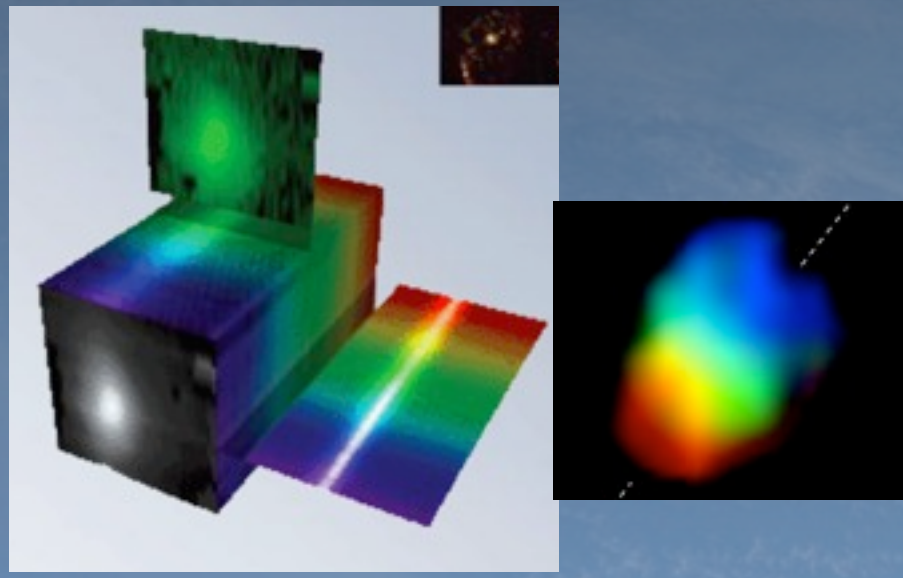
My e-Path

2009:

- Astronomical data frequently 3D
- No VO standards for 3D data
- No VO tools for 3D data

- Kinematical modeling tools for cubes being integrated in the VO

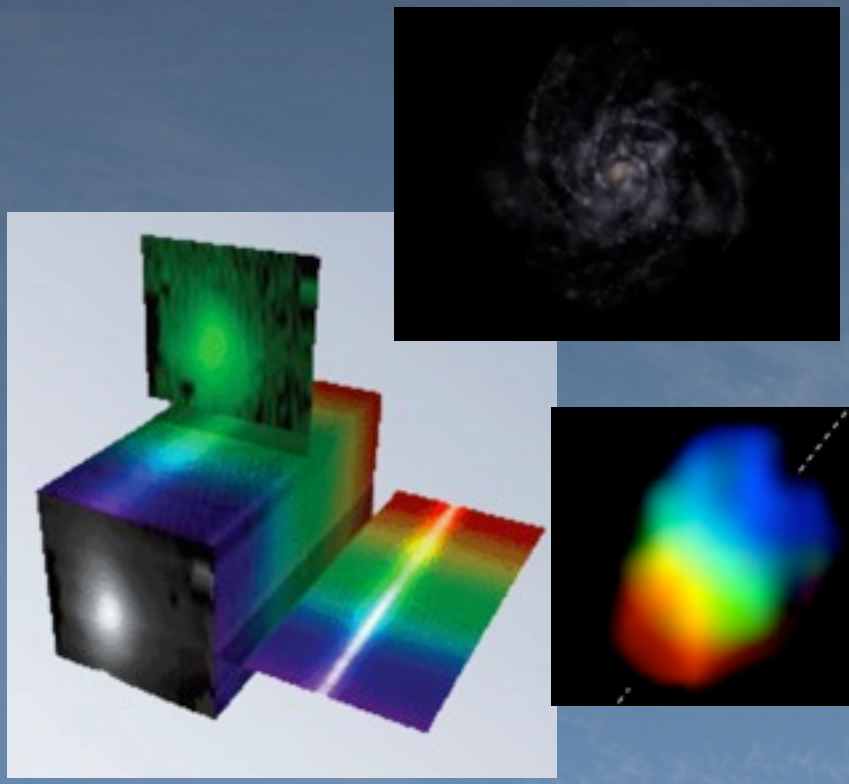
- 3D data services



2009:

- Astronomical data frequently 3D
- No VO standards for 3D data
- No VO tools for 3D data

- Kinematical modeling tools for cubes being integrated in the VO



- 3D data services

Astronomical data growing exponentially

- Public data (after short proprietary period)
- Large area surveys: full sky

Astronomical data growing exponentially

- Public data (after short proprietary period)
- Large area surveys: full sky
- New generation of radiointerferometers

will culminate with SKA

VLA, USA

1000 -1500 antennas x15m in 5km
1000 -1500 antennas x15m up to 3000 km

My e-Path

e-Astronomy

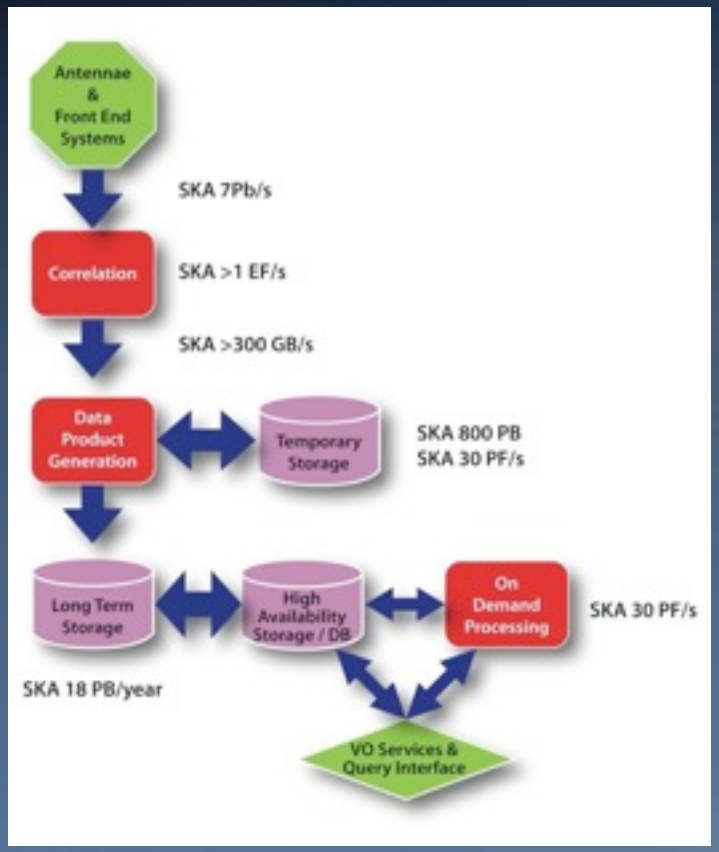
- Processing needs = 10*9 top range PCs

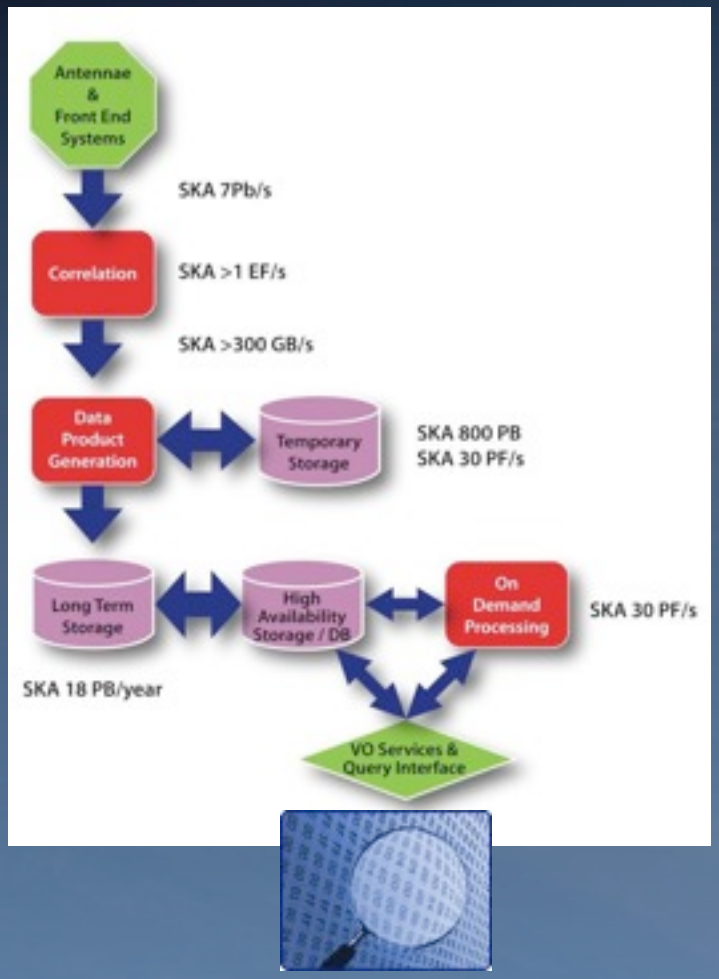
- 1 day = annual data product of all mankind

- Aperture arrays = 250 times the current Global Internet traffic

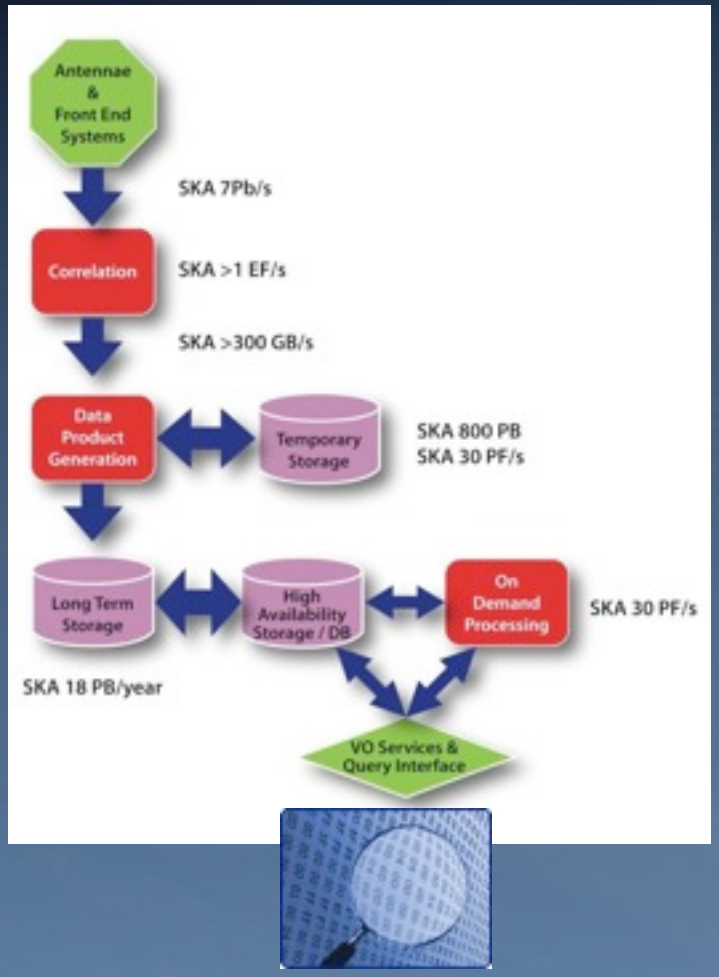
- Typical survey 1000 cubes = 5 days read time @ 10GB/sec

e-Astronomy


Antennae & Front End Systems
SKA 7Pb/s
Correlation
SKA >1 EF/s
SKA >300 GB/s
Data Product Generation
Temporary Storage
SKA 800 PB
SKA 30 PF/s
Long Term Storage
High Availability Storage / DB
On Demand Processing
SKA 30 PF/s
SKA 18 PB/year
VO Services & Query Interface



BUT A NEW CHALLENGE STARTS HERE



BUT A NEW CHALLENGE STARTS HERE

Extraction of scientifically relevant information from huge data volumes

- Visualization of enormous catalogs into multiD parameter spaces
- Efficient packaging of scientific methodology
- Collaborative science

Transfer of knowledge to society

Extraction of scientifically relevant information from huge data volumes

- Visualization of complex multiD data in parameter spaces
- Efficient packaging of scientific methodology
- Collaborative science

Transfer of knowledge to society

A disruptive change in the methodology required

BUT A NEW CHALLENGE STARTS HERE

An e-Science solution for the SKA

An e-Science solution for the SKA

My e-Path

e-Astronomy

R Dimensions

- Repeatable and reproducible
- Replayable, refreshable and roll-backable
- Reusable and Reliable
- Retrievable and referenceable
- Repurposeable

R Dimensions

- Repeatable and reproducible
- Replayable, refreshable and roll-backable
- Reusable and Reliable
- Retrievable and referenceable
- Repurposeable

Scientific workflows, collaborative tools

Publication beyond the PDF, preservation

UK e-Science My e-Path e-Astronomy

Workflows for ever Project (started December 2010)

Advanced Workflow Preservation Technologies for Enhanced Science

Funded under FP7 ICT-2009-6

Participants:

iSOCO (PI), UPMadrid, U. Manchester, Poznan PSNC, U. Oxford, Leiden U. Medical Center, IAA-CSIC

Development of standards for workflow preservation, which will enable workflow' Rs dimensions

UK e-Science My e-Path e-Astronomy

Workflows for ever Project (started December 2010)

Advanced Workflow Preservation Technologies for Enhanced Science

Funded under FP7 ICT-2009-6

Participants:

iSOCO (PI), UPMadrid, U. Manchester, Poznan PSNC, U. Oxford, Leiden U. Medical Center, IAA-CSIC

Development of standards for workflow preservation, which will enable workflow' Rs dimensions

Use cases: Astronomy and Biomedicine

Wf4ever: Preservation of Scientific Workflows

- **Research Object**:

  packages workflow **descriptions**, **provenance** of their **executions**, and links to all the **related resources**

- Models for workflow abstraction –> **classification, indexing**
- Strategies for **sharing and reusing** workflows or **fragments (from different sources)**
- Mechanisms for personalized workflow **recommendation**
- Methods and tools to proactively **evaluate** workflow information quality

Cross-fertilization

Wf4ever: Preservation of Scientific Workflows

- **Research Object**:

  packages workflow **descriptions**, **provenance** of their **executions**, and links to all the **related resources**

- Models for workflow abstraction –> **classification, indexing**
- Strategies for **sharing and reusing** workflows or **fragments (from different sources)**
- Mechanisms for personalized workflow **recommendation**
- Methods and tools to proactively **evaluate** workflow information quality

Cross-fertilization

Key issues in Wf4ever:

Provenance

Needs to track it for data, services, tools

Semantic Web

"I have a dream for the Web [in which computers] become capable of analyzing all the data on the Web – the content, links, and transactions between people and computers. A 'Semantic Web', which should make this possible, has yet to emerge, but when it does, the day-to-day mechanisms of trade, bureaucracy and our daily lives will be handled by machines talking to machines."

– Tim Berners-Lee, 1999

Linked data

UK e-Science e-Precedents

- **Provenance: 1986** (Becker & John M. Chambers, cf Luc Moreau)
- **Semantic Web: 2001** (Berners-Lee, Scientific American)
- **Linked data: 2006**

And before in UK?

UK e-Science e-Precedents

- Provenance: 1986 (Becker & John M. Chambers, cf Luc Moreau)
- Semantic Web: 2001 (Berners-Lee, Scientific American)
- Linked data: 2006

And before in UK?

Heuristic ALgorithmic computer ... 9000

Arthur C. Clarke (1964)

2001, A Space Odyssey

- Experience-based techniques
- For problem solving, learning, and discovery
- Where an exhaustive search is impractical

Still we are necessary...

Where are we now?

- Scientific workflows:

  Social and life sciences, engineering, chemistry, music, etc: intensive use

UK e-Science e-Astronomy

Where are we now?

- Scientific workflows:

  Social and life sciences, engineering, chemistry, music, etc: intensive use

  Astronomy

Helio project on-going

Virtual Observatory for solar physics

Data access + sharing + description of the knowledge in the field (via ontologies), and their processes (via workflows)

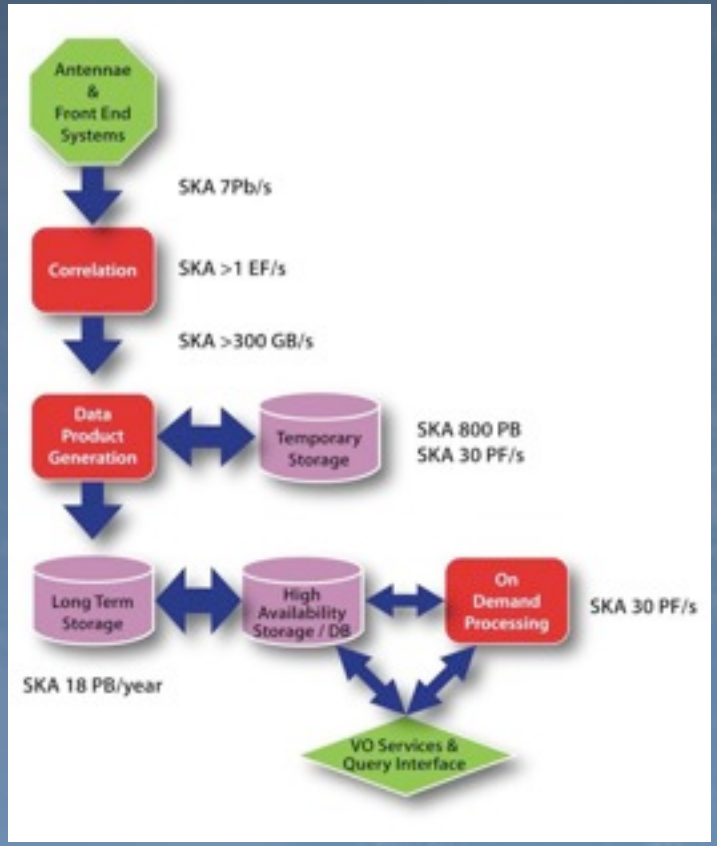
UK e-Science e-Astronomy

Where are we now?

- Scientific workflows:

  Social and life sciences, engineering, chemistry, music, etc: intensive use

  Astronomy



Helio project on-going

Virtual Observatory for solar physics

Data access + sharing + description of the knowledge in the field (via ontologies), and their processes (via workflows)

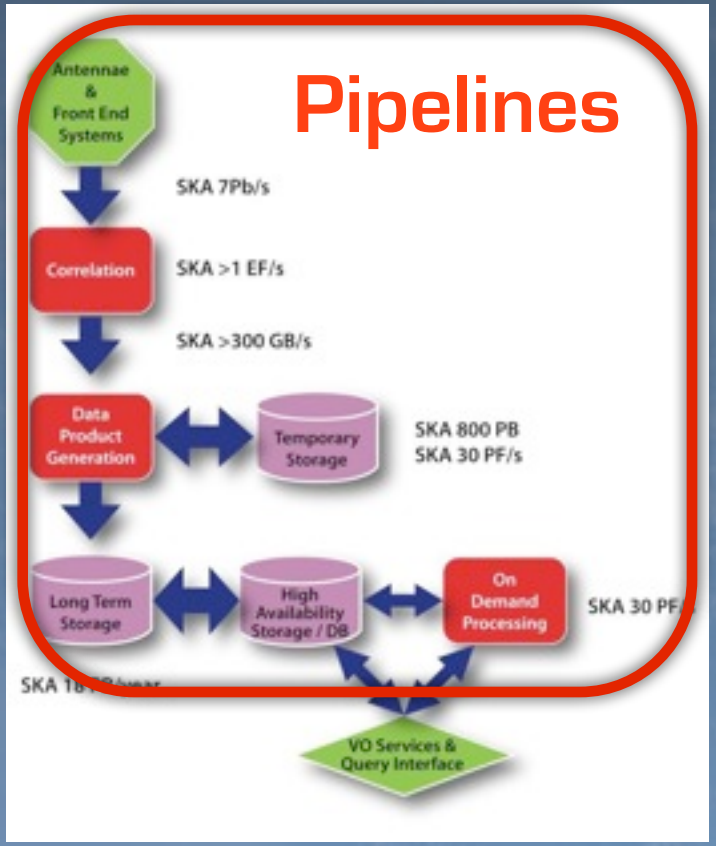
UK e-Science e-Astronomy

Where are we now?

- Scientific workflows:

  Social and life sciences, engineering, chemistry, music, etc: intensive use

  Astronomy



Helio project on-going

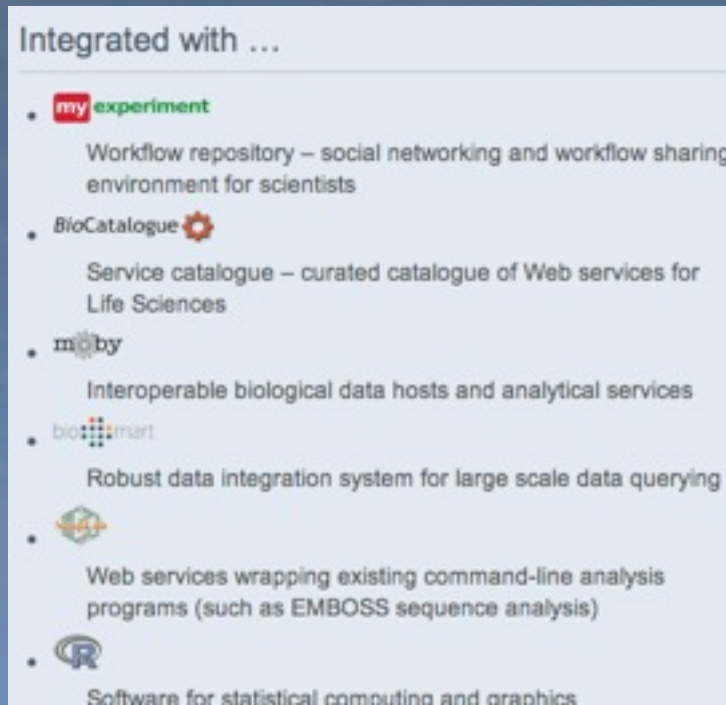
Virtual Observatory for solar physics

Data access + sharing + description of the knowledge in the field (via ontologies), and their processes (via workflows)

Where are we now?

- Building tools: TAVERNA

Modularity + capability to encapsulate methodologies, allow scientists



Where are we now?

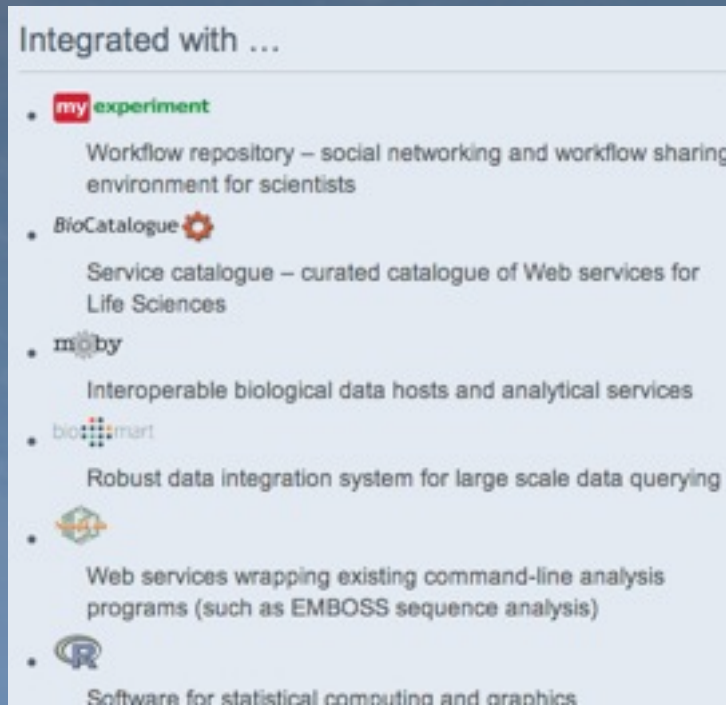
- Building tools: TAVERNA

Modularity + capability to encapsulate methodologies, allow scientists

to reuse and share them

myExperiment

(http://www.myexperiment.org/)



Workflows

Search filter terms

« previous 1 2 3 ... 137 next »

Sort by: Rank

Showing 1367 results. Use the filters on the left and the search box below to refine the results.

Search

Filter by type

- Taverna 1 557
- Taverna 2 522
- RapidMiner 112
- Bioclipse Scri… 29
- GWorkflowDL 24
- LONI Pipeline 20
- Kepler 17
- BioExtract Ser… 13
- Trident (Packa… 10
- Chemistry Plan 7

Filter by tag

- example 178
- mygrid 103
- localworker 99
- bioinformatics 89
- benchmarks 77
- cheminformatics 65
- protein 60
- ebi 56
- BLAST 49
- pathway 47

Taverna 2

Pathways and Gene annotations for QTL region - Mouse (v6)

View

Download (v6)

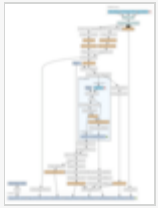
Created: 19/11/09 @ 18:18:52 | **Last updated:** 05/04/11 @ 10:59:25

Credits: Paul Fisher

License: Creative Commons Attribution-Share Alike 3.0 Unported License

Original Uploader

Paul Fisher



This workflow searches for genes which reside in a QTL (Quantitative Trait Loci) region in the mouse, Mus musculus. The workflow requires an input of: a chromosome name or number; a QTL start base pair position; QTL end base pair position. Data is then extracted from BioMart to annotate each of the genes found in this region. The Entrez and UniProt identifiers are then sent to KEGG to obtain KEGG gene identifiers. The KEGG gene identifiers are then used to searcg for pathways in the KEGG path...

Rating: 4.5 / 5 (2 ratings) | **Versions:** 6 | **Reviews:** 0 | **Comments:** 4 | **Citations:** 1

Viewed: 1916 times | **Downloaded:** 400 times

Tags (13):

New/Upload

Workflow GO

Log in / Register

Username or Email:

Password:

Remember me:

OR

Use OpenID:

(eg: name.myopenid.com)

Log in

Need an account?
Click here to register

Forgot Password?

Popular Tags

25 tags
[All Tags]

benchmarks | bio2rdf | bioinformatics | BLAST | cheminformatics | data integration | ebi | example |

Fisher

input of: a chromosome name or number; a QTL start base pair position; QTL end base pair position. Data is then extracted from BioMart to annotate each of the genes found in this region. The Entrez and UniProt identifiers are then sent to KEGG to obtain KEGG gene identifiers. The KEGG gene identifiers are then used to searcg for pathways in the KEGG path...

Rating: 4.5 / 5 (2 ratings) | **Versions:** 6 | **Reviews:** 0 | **Comments:** 4 | **Citations:** 1

Viewed: 1916 times | **Downloaded:** 400 times

Tags (13):

data-driven | disease | genotype | kegg | mouse | nbiconworkflows | pathway | pathway-driven | pathways | phenotype | qtl | shim | subworkflow

---

Taverna 1

Original Uploader

Marco Roos

BioAID_DiseaseDiscovery_RatHumanMouseUniprotFilter (v4)

View

Download (v4)

Created: 15/12/08 @ 20:46:09 | **Last updated:** 26/01/11 @ 14:43:31

Credits: Marco Roos AID

License: Creative Commons Attribution-Share Alike 3.0 Unported License

This workflow finds disease relevant to the query string via the following steps: 1. A user query: a list of terms or boolean query - look at the Apache Lucene project for all details. E.g.: (EZH2 OR "Enhancer of Zeste" +(mutation chromatin) -clinical); consider adding 'ProteinSynonymsToQuery' in front of the input if your query is a protein. 2. Retrieve documents: finds 'maximumNumberOfHits' relevant documents (abstract+title) based on query (the AIDA service inside is based on Apache's Luce...

Rating: 4.0 / 5 (2 ratings) | **Versions:** 4 | **Reviews:** 0 | **Comments:** 2 | **Citations:** 0

Viewed: 3390 times | **Downloaded:** 483 times

Tags (9):

AIDA | BioAID | biorange_nl | disease | pkna | protein | text_mining | text_mining_network | VL-e

---

Taverna 1

EBI_InterProScan (v3)

View

Filter by tag

- example 178
- mygrid 103
- localworker 99
- bioinformatics 89
- benchmarks 77
- cheminformatics 65
- protein 60
- ebi 56
- BLAST 49
- pathway 47

Filter by user

- Alan Williams 210
- Paul Fisher 89
- Antoon Goderis 82
- Peter Li 64
- Hamish McWil… 52
- Francois Belleau 43
- Franck Tanoh 27
- Andreas Hohe… 26
- Anja Le Blanc 25
- Egon Willigha… 24

Filter by licence

- by-sa 895
- by 273
- by-nd 188
- GPL 7
- CC0 3
- LGPL 1

Filter by group

- myGrid 214
- SabrOndexPr… 32
- helio 26
- MediGRID 22
- TextGrid 21
- BauVOGrid 20
- CDK-Taverna 18

Click here to register
UK e-Science

Popular Tags

25 tags

[All Tags]

benchmarks | bio2rdf | bioinformatics | BLAST | cheminformatics | data integration | ebi | example | gene | graph | impact | kegg | Kegg Pathways | localworker | mygrid | ondex | pathway | pathways | phenotype | protein | pubmed | sequence | taverna | text mining | workflow

Astronomy is ready: lifecycle is digital

Submission of proposals

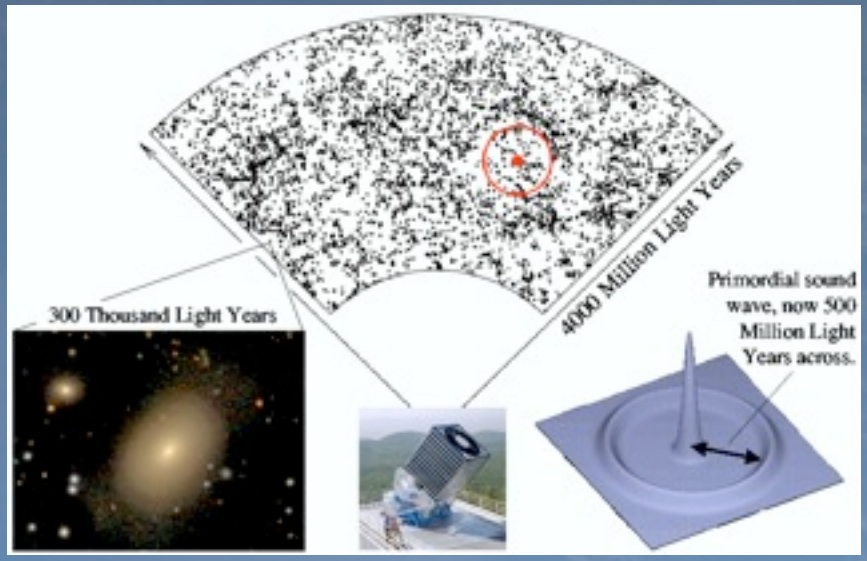
- Pipelines for automated data reduction:

  Science Ready Data

- Access policy:

  often public after a

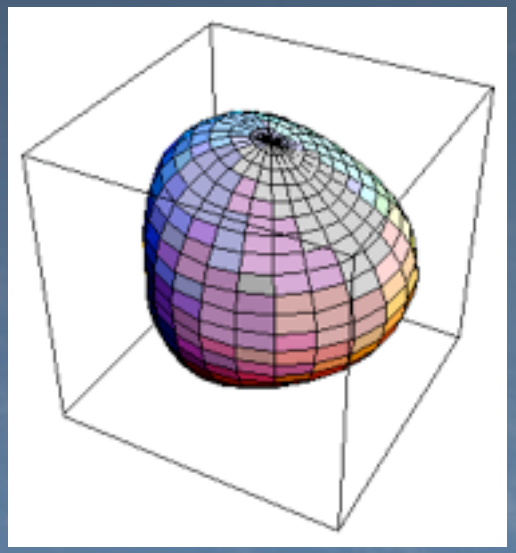
  short period



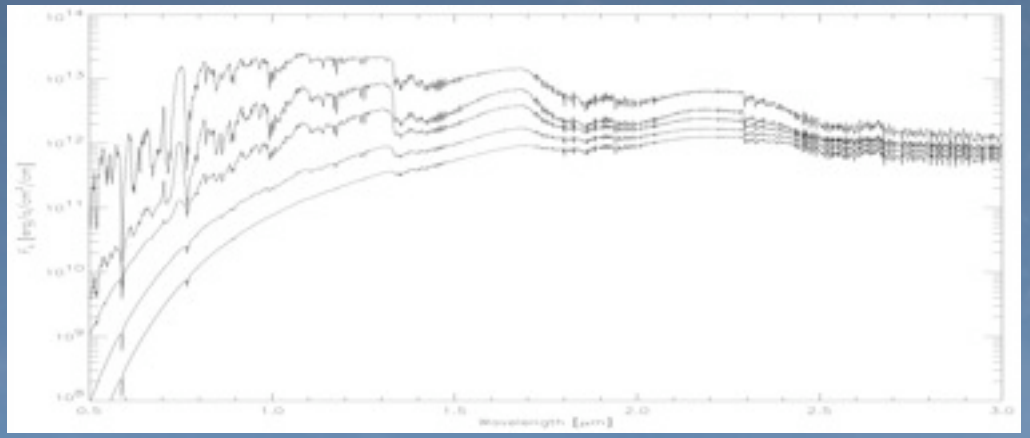
- Intensive use of databases

Numerical simulations

Stellar models



Astronomical Libraries



PDF publication: Bibliographical archives

- Vizier provides interlinking to Bibliographic Archives

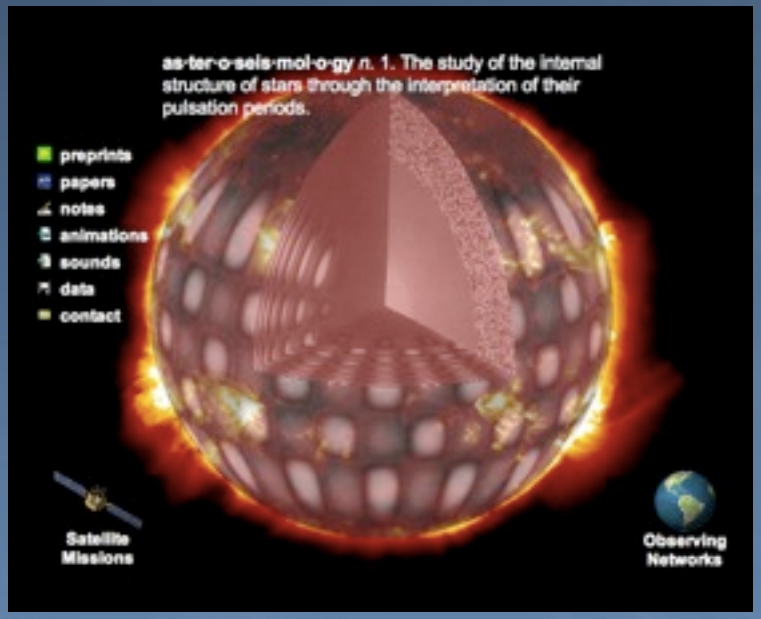
30000 queries/day on average in 2006

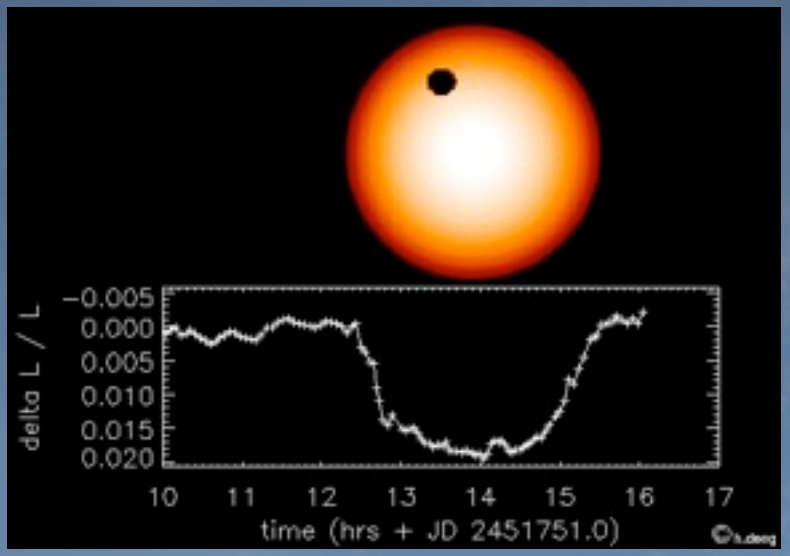
- ADS provides interlinking to Astronomical Objects DB

- Long term history with GRID of Instruments

To avoid gaps in time domain observations:

asteroseismology & search for exoplanets





- Long term history with GRID of Instruments

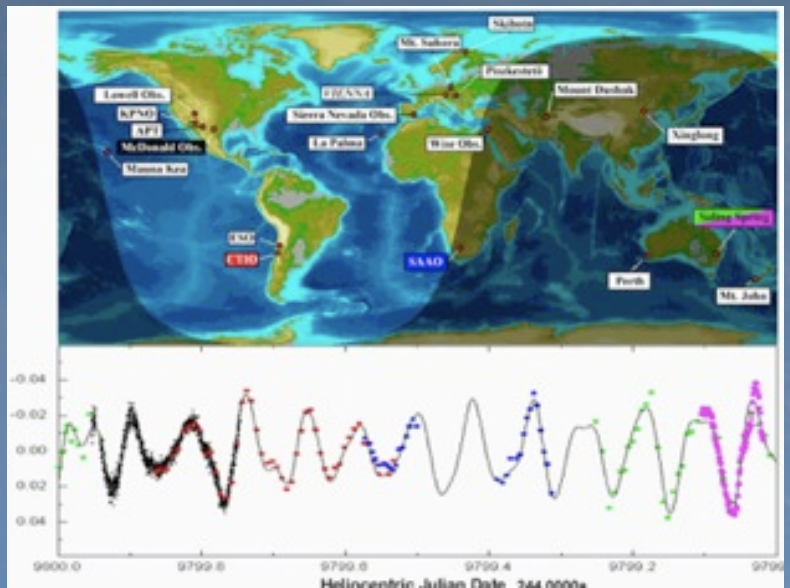
To avoid gaps in time domain observations:

asteroseismology & search for exoplanets

Coordinated observations: a telescope network

The Delta Scuti Network (DSN)

The Delta Scuti Network (DSN) is a collaboration of astronomers from all around the globe who observe and study short period variable stars. It was founded by Michel Breger at the University of Texas (McDonald Observatory) and at the University of Vienna (Institute of Astronomy) in 1983 with the goal to improve the frequency solutions of multiperiodic Delta Scuti stars.



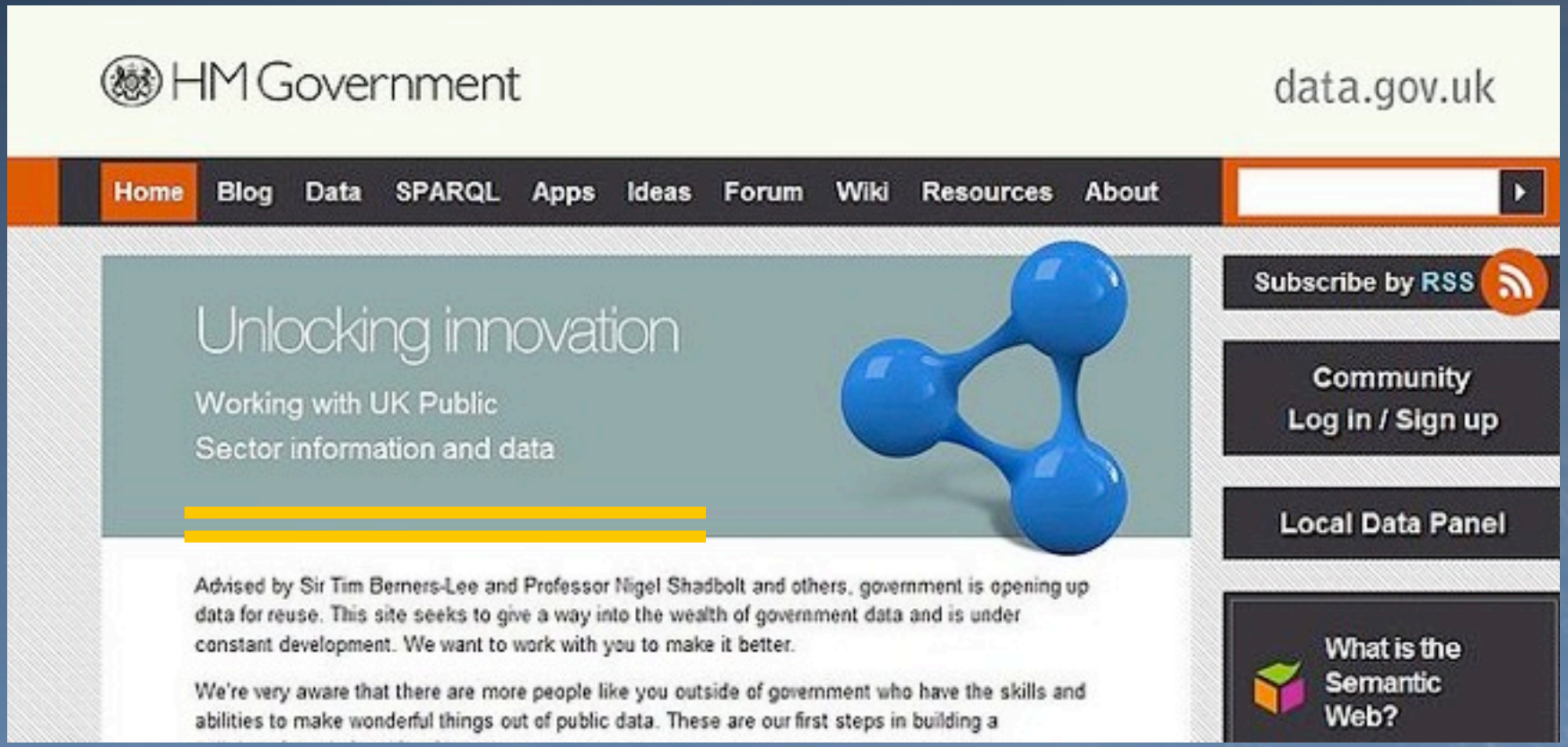
How should e-Science return to citizens?

How should e-Science return to citizens?

- Improving life by doing better science and bringing it faster to the society
- Delivering information
- Make them aware of scientific methodology
- Make them part of the scientific methodology

Delivering information

- Linked Government data



"Information should be in the public domain unless there is a good reason not to - not the other way around." T. Berners-Lee

Make them aware of scientific methodology

Make them aware of scientific methodology

International Year of Astronomy: 800 students measure the Earth's radius

On 26 March, 800 students throughout Spain took part in the same experiment: they repeated the experiment that Eratosthenes performed in 240 BC to measure the Earth's radius.

Global collaboration is the clue!

Make them part of the scientific methodology

1900

- Frank Chapman (Ntl Audubon S) ornithologist
- Counting birds on Xmas instead of killing them!
- 27 observers, 25 places (USA & Canada)

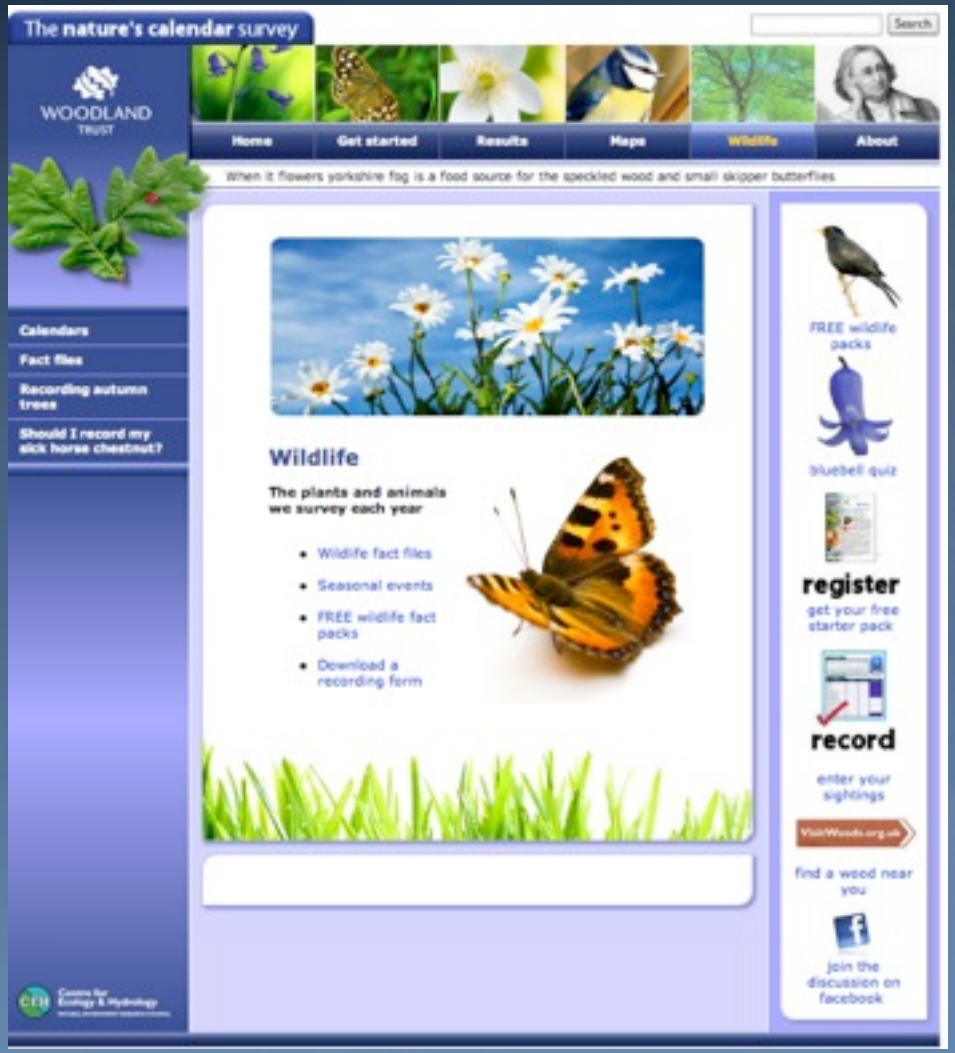
Make them part of the scientific methodology

1900

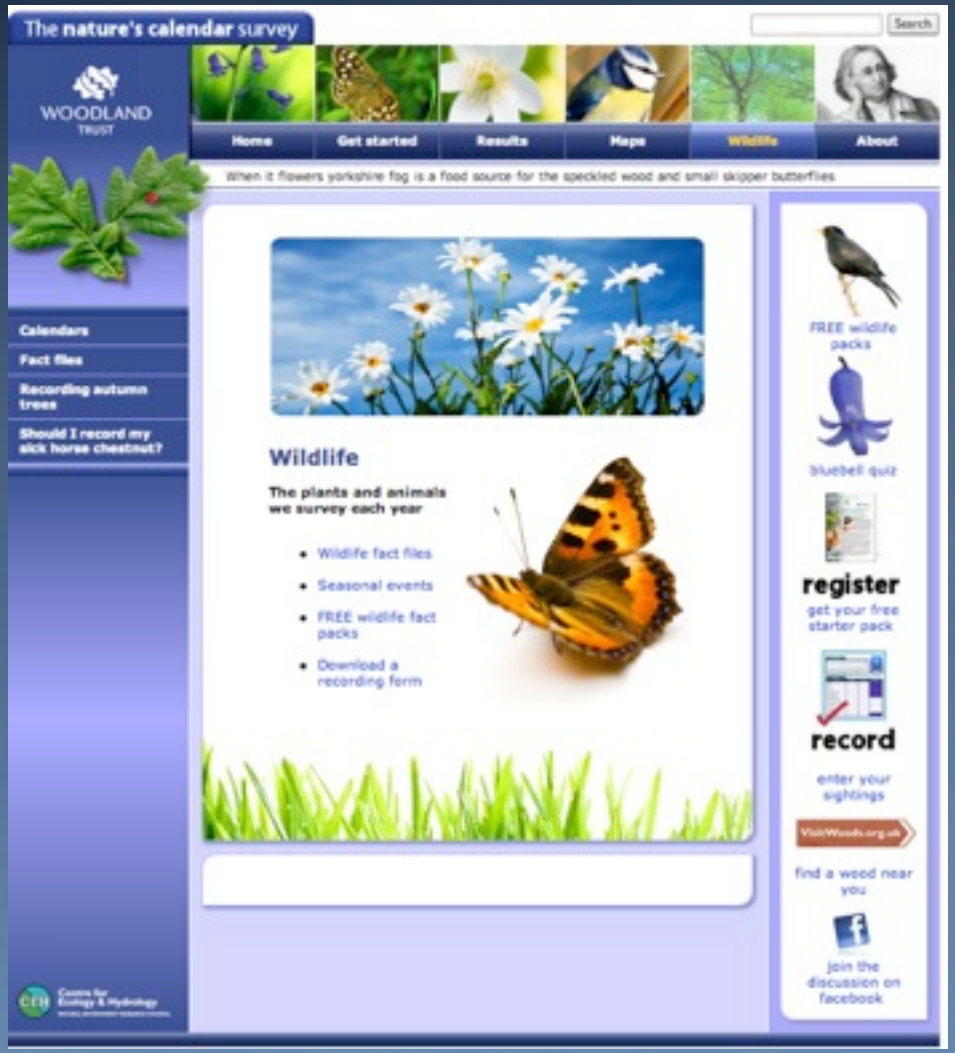
- Frank Chapman (Ntl Audubon S) ornithologist
- Counting birds on Xmas instead of killing them!
- 27 observers, 25 places (USA & Canada)
- Still continuing

eBIRD

2000: Nature's Calendar Survey



2000: Nature's Calendar Survey



- 50.000 people across UK
- Phenology: times of recurring natural phenomena
- Woodland Trust + Centre for Ecology & Hydrology

A review of spring 2010 – Richard Smithers and Tim Sparks

Readers may remember media interest in the weather during early spring 2010 because January and February were the coldest for 30 years. Unsurprisingly some spring events expected at this time of year, such as blackthorn flowering, were consequently late.

However, monthly temperatures in March and April were above average and as a result some events expected in these months were relatively early; for instance butterfly emergence and leafing in some trees. Overall, mean February-April temperatures were still warmer than both 2001 and 2006.

Some species showed rather unexpected responses, which are detailed below.

Weather

Monthly spring temperatures have been quite variable over the last five years (see Figure 1). Spring 2010 got off to a chilly start with the coldest January since 1987 and coldest February since 1991. Average temperatures for January-February were 1.7°C below the 30-

Events

In spring 2010, colder temperatures until mid-March resulted in subsequent events being later than our baseline year of 2001 (the year spring temperatures were very close to the 1961-1990 30-year average). However, warmer temperatures in the second half of March appear to have reversed the situation for events from early April onwards.

Events that on average occurred later than in 2001 were:

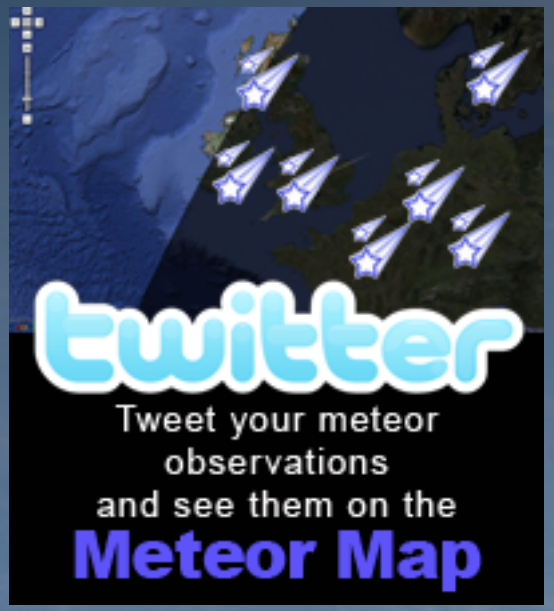
- **Lesser celandine first flower** on 20 April, five days late and 17 days later than 2009
- **Hawthorn budburst** on 27 March, 4 days late and 16 days later than 2009

- **Hawthorn first leaf** on 6 April, 3 days late and 15 days later than 2009

2009: Twitter Meteor Watch

- 3 nights around the peak of the Perseid meteor shower
- > 20,000 people on Twitter

New media for the new generation.



GALAXY ZOO: NEXT TALK!

How should e-Science return to citizens?

How should e-Science return to citizens?

- Improving life by doing better science and bringing it faster to the society 
- Delivering information 
- Make them aware of scientific methodology 
- Make them part of the scientific methodology 

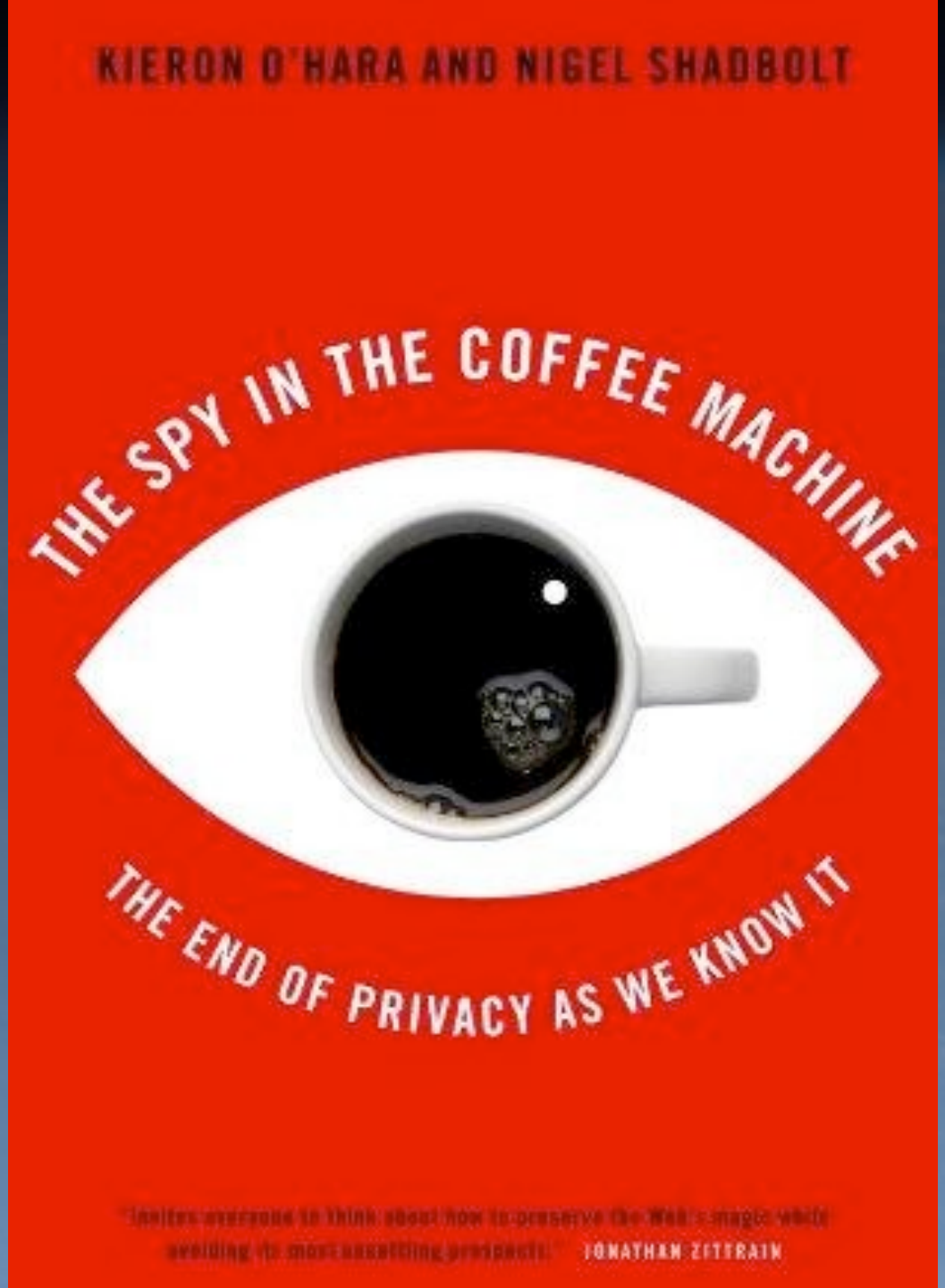
How should e-Science return to citizens?

- Improving life by doing better science and bringing it faster to the society 
- Delivering information 
- Make them aware of scientific methodology 
- Make them part of the scientific methodology 

Are there any drawbacks...?


UK e-Science

UK e-Science


KIERON O'HARA AND NIGEL SHADBOLT
THE SPY IN THE COFFEE MACHINE
THE END OF PRIVACY AS WE KNOW IT
JONATHAN ZITTRAIN

Eagle eye, 2008. D. J. Caruso

Eagle eye, 2008. D. J. Caruso

NET NEUTRALITY

"Threats to the Internet, such as companies or governments that interfere with or snoop on Internet traffic, compromise basic human network rights."

Berners-Lee

NET NEUTRALITY

"Threats to the Internet, such as companies or governments that interfere with or snoop on Internet traffic, compromise basic human network rights."

Berners-Lee

DE-CENTRALIZATION OF SCIENCE

THANK YOU

AND BY THANK YOU

...I MEAN

AND BY THANK YOU

...I MEAN

THANK YOU

FOR MAKING ALL

THESE HAPPEN

AND BY THANK YOU

...I MEAN

AND BY THANK YOU

...I MEAN

THANK YOU

FOR MAKING ALL

THESE HAPPEN