e-Science is about Scientists too

The Evolution of the Grid and the Web

David De Roure

University of Southampton

OGF Semantic Grid Research Group

www.semanticgrid.org

Introduction

This is the story of the Semantic Grid

It's a Tale of Two Infrastructures – the Grid and the Web – both evolving

And the most important system of all – the scientists, doing science

Overview

1. The search for the missing link
   - Data
   - Services
   - People
2. Evolution of the Web
3. "Ending the Tyranny of the Grid"
   (or rather, ending the accidental tyranny of the grid mindset)

The Semantic Grid Report 2001

e-Science is about global collaboration in key areas of science and the next generation of infrastructure that will enable it

John Taylor

There are a number of grid applications being developed and there is a whole raft of computer technologies that provide fragments of the necessary functionality. However there is currently a major gap between these endeavours and the vision of e-Science in which there is a high degree of easy-to-use and seamless automation and in which there are flexible collaborations and computations on a global scale.

Us

Missing Link

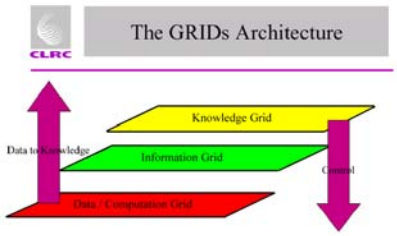
Scientists

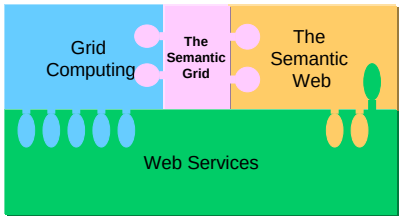
Need something here

Grid Infrastructure

Why Semantic Web?

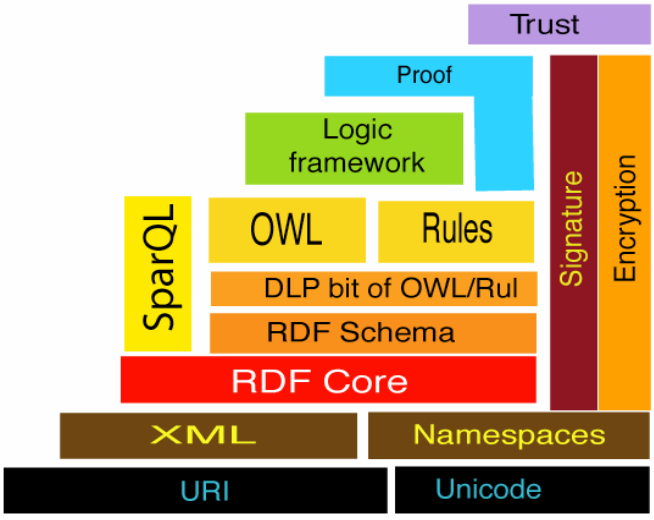
- Huge potential for Science
  - making data reusable, interlinked
  - making connections between decoupled content
  - generating new intelligence
- Automation requires machine processable descriptions
- Grid community talking about metadata and knowledge



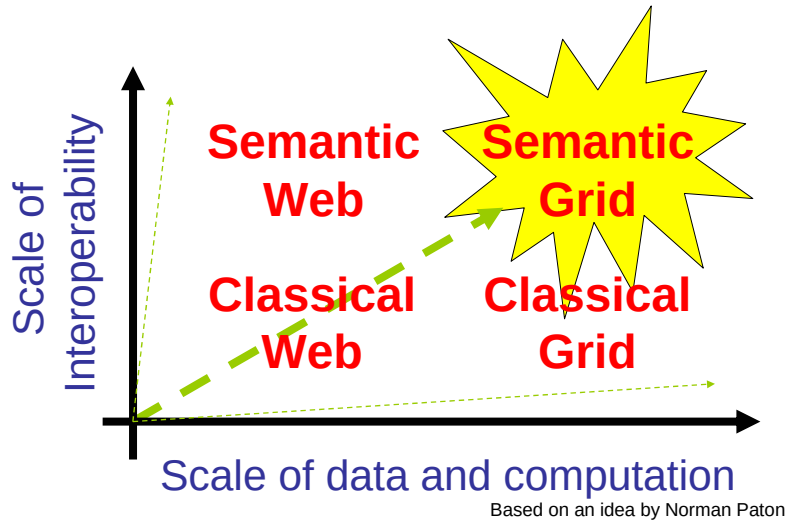


Building bridges

The Semantic Web layer cake



Semantic Grid = Grid + Semantic Web for e-Science



Based on an idea by Norman Paton

Semantic Grid

~~Web~~ Grid

The Semantic ~~Web~~ is an extension of the current ~~Web~~ Grid in which information and services are given well-defined meaning, better enabling computers and people to work in cooperation

Free the data!

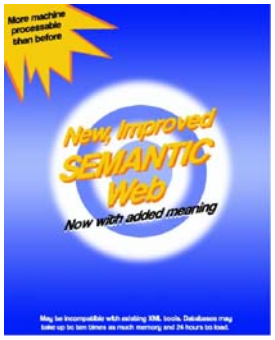
Free the services!

Free the people!

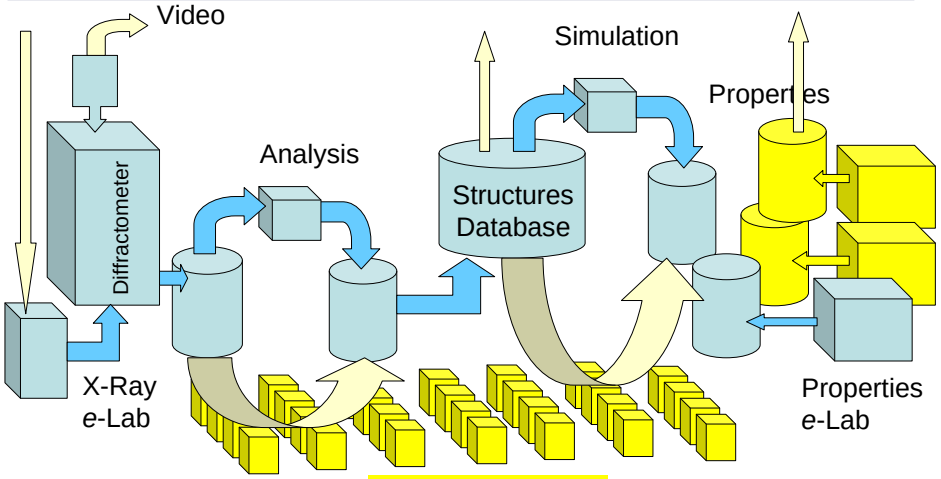
www.semanticgrid.org

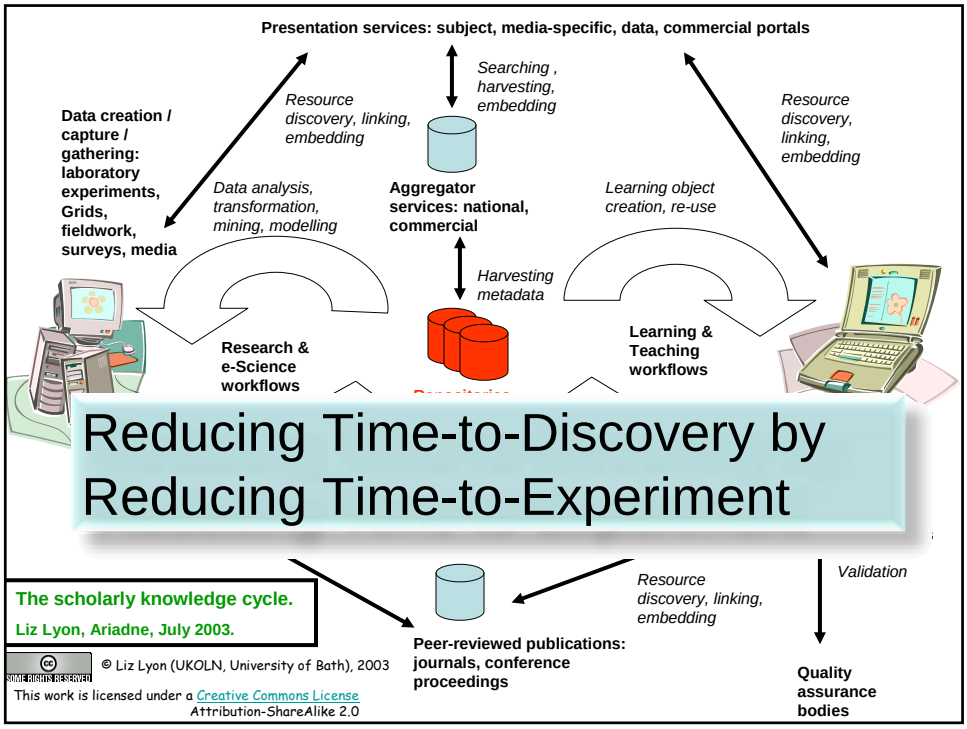
My Chemistry Experiment

Box of Chemists



CombeChem pilot project



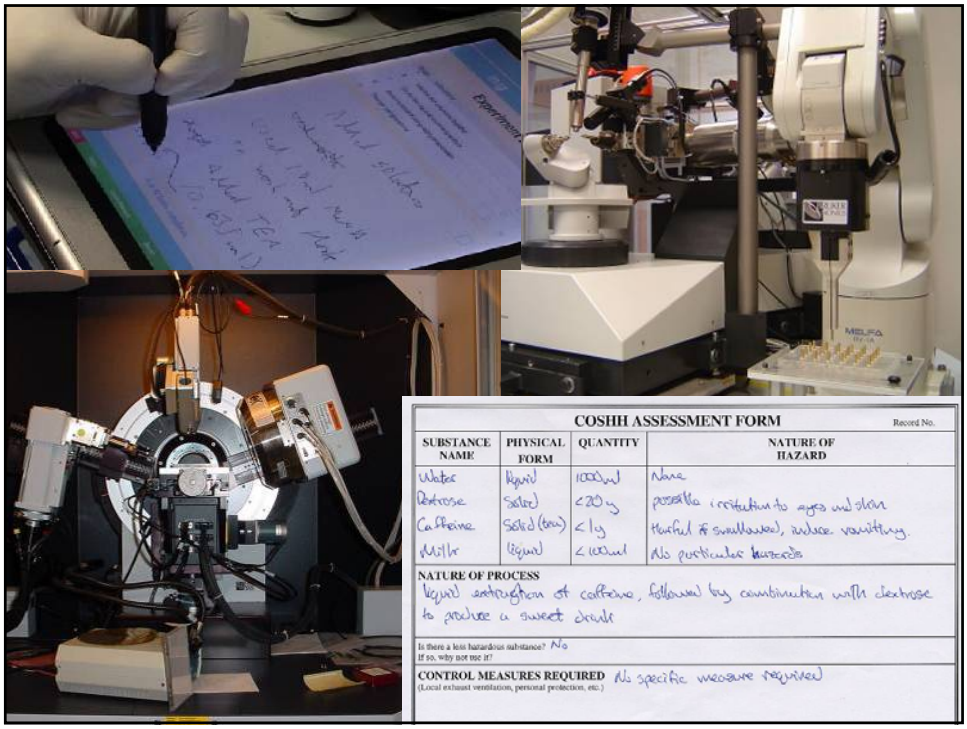


e Science = Record and Reuse. Reuse needs provenance

- The key observation!

  The details of the origins of data are just as important to understanding as their actual values

- "Publication at Source" describes the need to capture data and its context from the outset and maintain a complete end-to-end connection between the laboratory bench and the intellectual chemical knowledge that is published as a result of the investigation.



COSHH ASSESSMENT FORM

Record No.

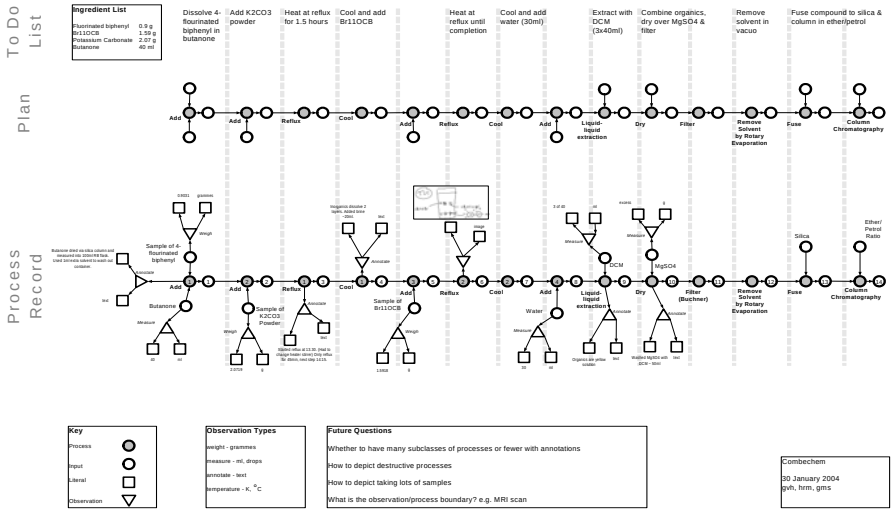
SUBSTANCE NAME	PHYSICAL FORM	QUANTITY	NATURE OF HAZARD
Water	liquid	1000ml	None
Dextrose	solid	<20g	possible irritation to eyes and skin
Caffeine	solid (tea)	<1g	Harmful if swallowed, induce vomitting
Milk	liquid	<100ml	No particular hazards

NATURE OF PROCESS
liquid extraction of caffeine, followed by combination with dextrose to produce a sweet drink

Is there a less hazardous substance? No
If so, why not use it?

CONTROL MEASURES REQUIRED (Local exhaust ventilation, personal protection, etc.)
No specific measure required

The RDF Graph



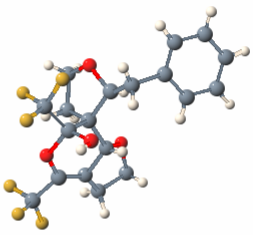
2-Benzyl-4',6'-bis(trifluoromethyl)-6'-hydroxy-2',3',7',7a'-tetrahydro-6'H-spiro-(furan-3,7'-(furo(3,2-c)pyran))

S. J. Coles, J. M. Mellor, A. H. El-Sagheer, E. E. -D. M. Salem and R. N. Metwally.

University of Southampton

$C_{19}H_{18}F_6O_4$

CCDC Code: WOSSUS
IChI Code: C19H18F6O4,20-17(21,22)13-12-6H2-8H2-28-15H(12)16(18(26H,29-13)19(23,24)25)7H2-9H2-27-14H(16)10H2-11-4H-2H-1H-3H-5H-11
(google for ichi)
Compound Class: Organic
Keywords: additions; Grignard reagents; trifluoromethylketones
Creation Date: 13 September 1999
Deposited By: Susanne L. Huth
Deposited On: 30 July 2004



Data collection parameters

Chemical formula	C19 H18 F6 O4

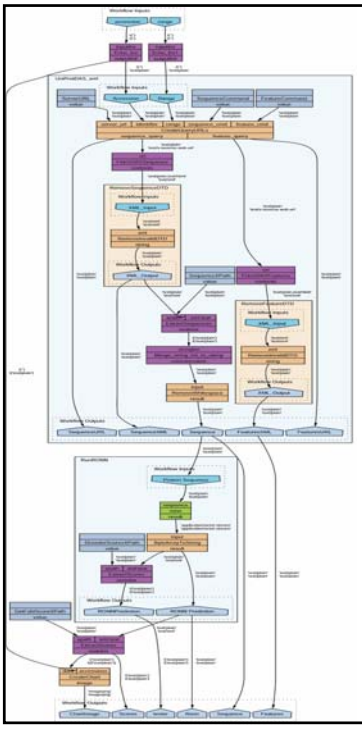
Available Files

Final Result

99sot029_data/99sot029.CIF	14k
99sot029_data/99sot029.cml	7k

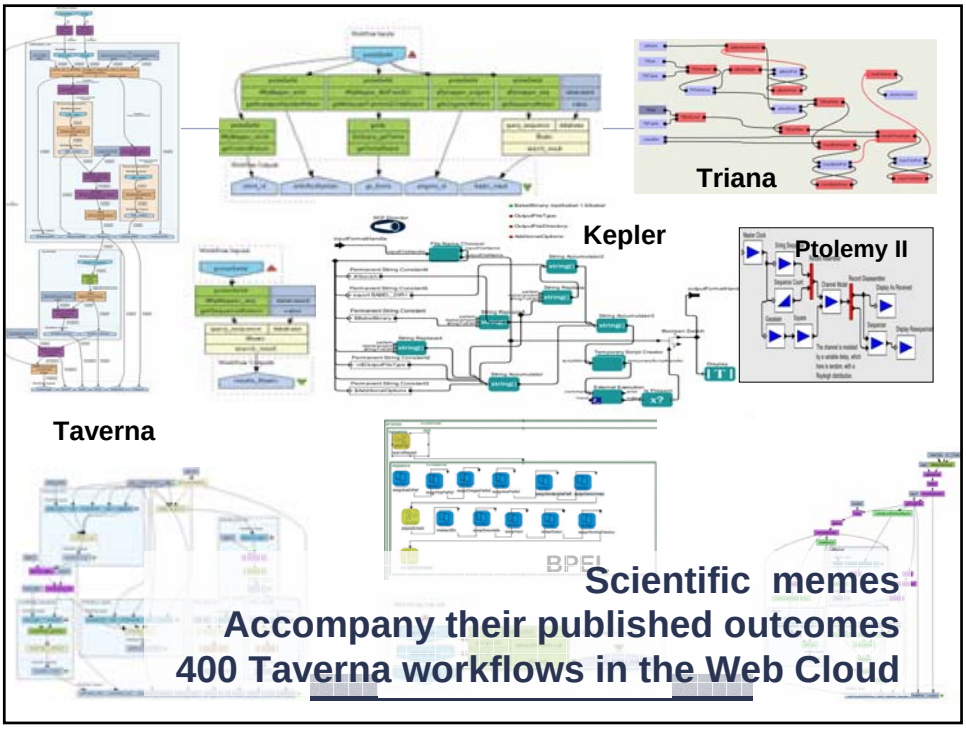
CombeChem Principles

- It's a Semantic DataGrid
- Linking decoupled data "In the Wild" (3rd party sources)
- Publish don't warehouse
- Think Holistic – we're working in the context of the Scholarly Knowledge Cycle
- Power of Provenance
- A little Semantics goes a long way
- Empowerment versus requirements capture

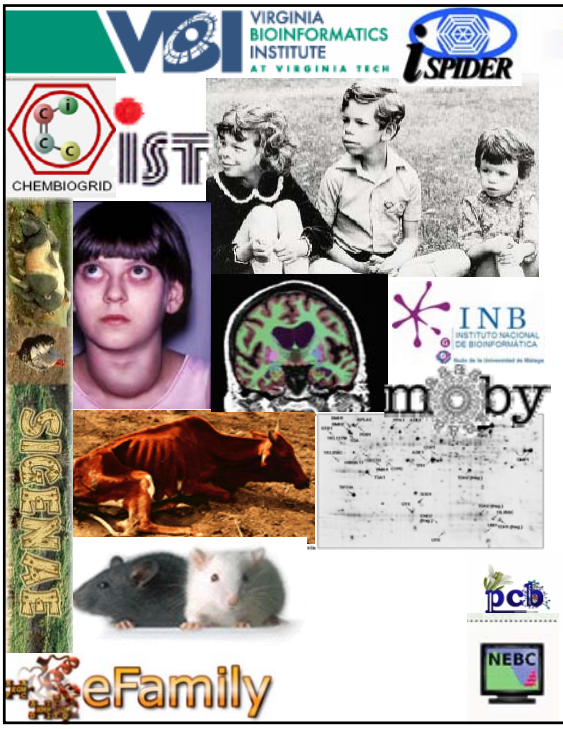


E. Science laboris

- Workflows are the new rock and roll.
- Machinery for coordinating the execution of (scientific) services and linking together (scientific) resources.
- The era of Service Oriented Applications
- Repetitive and mundane boring stuff made easier.
- The challenge for biology is complexity and heterogeneity, not so much compute.



Scientific memes
Accompany their published outcomes
400 Taverna workflows in the Web Cloud

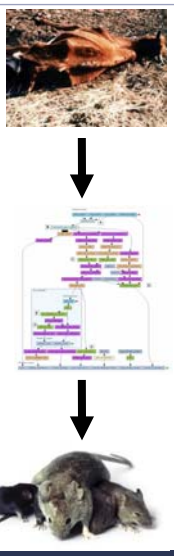


myGrid Taverna

- *34,632 sourceforge downloads.*
- *Ranked 210 in sourceforge activity (06 June 07)*
- *1,500 per month, 200+ sites*
- *Users throughout UK, USA, Europe, and SE Asia*
- Systems biology
- Proteomics
- Gene/protein annotation
- Microarray data analysis
- Medical image analysis
- Heart simulations
- High throughput screening
- Phenotypical studies
- Phylogeny
- Text mining
- Plants, Mouse, Human
- Astronomy

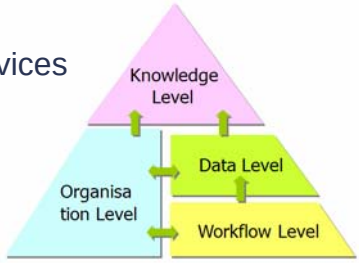
Recycling, Reuse, Repurposing

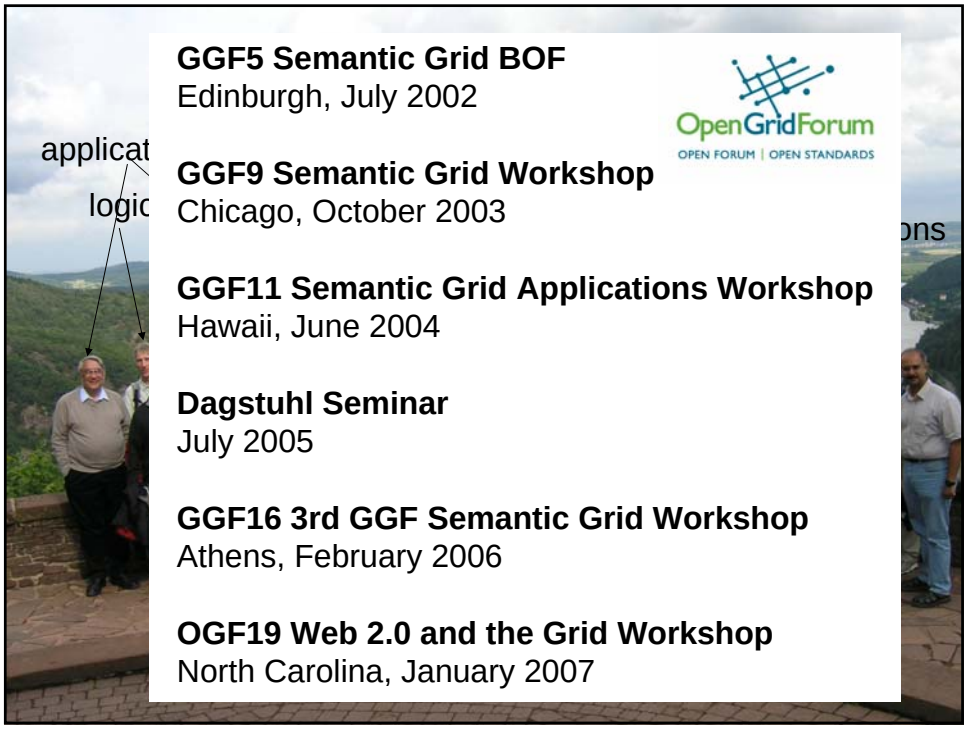
- Paul meets Jo
- Trypanosomiasis cattle workflow reused without change
- Identified the biological pathways involved in sex dependence in the mouse model, previously believed to be involved in the ability of mice to expel the parasite
- Previously a manual two year study, by Jo, of candidate genes had failed to do this



myGrid Principles

- Reuse, reuse and reuse – workflows transcend their application
- In the Wild – Integrating 3rd party services
- A little Semantics goes a long way
- Power of Provenance
- Users add value
- "Come As You Are". Your desktop app.
- Understand the rewards system of stakeholders
  - Jam Today and more / better Jam Tomorrow





GGF5 Semantic Grid BOF
Edinburgh, July 2002

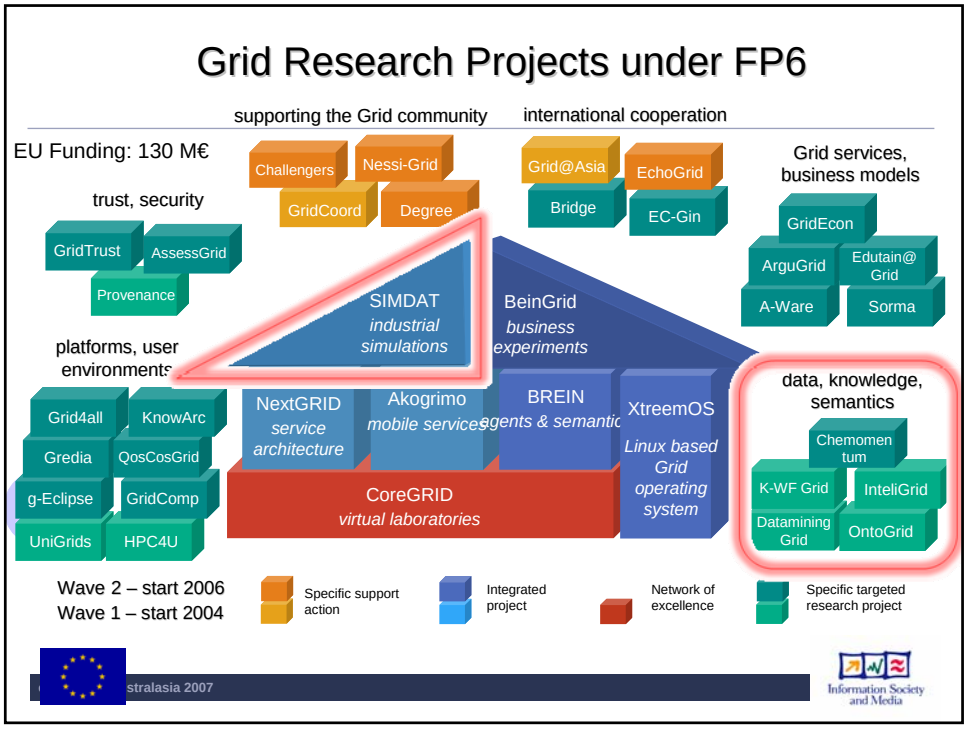
GGF9 Semantic Grid Workshop
Chicago, October 2003

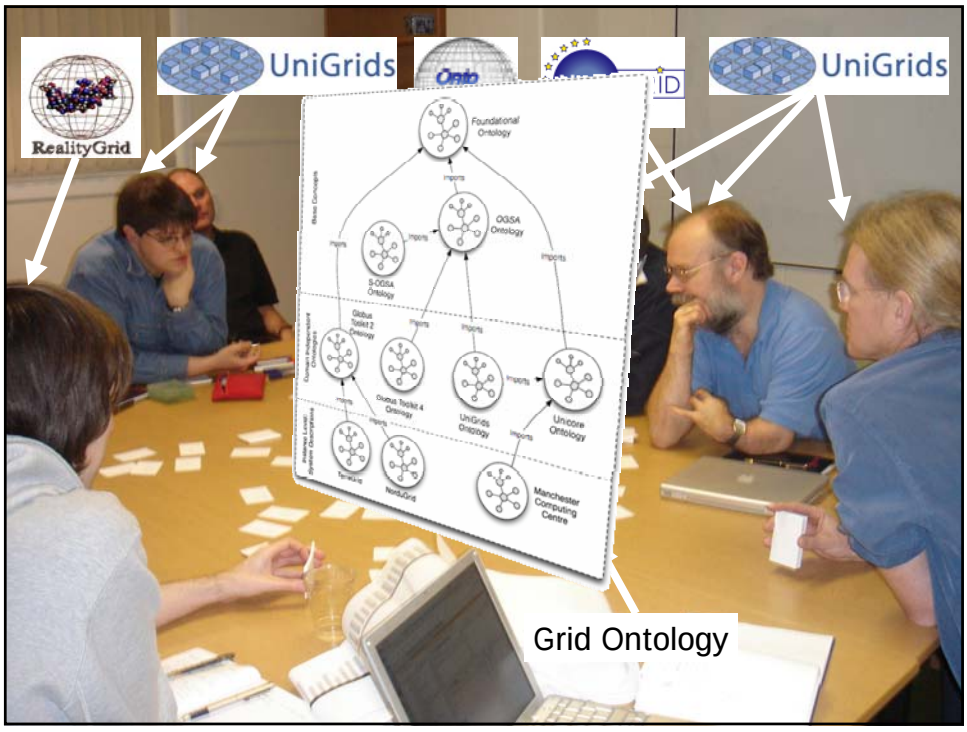
GGF11 Semantic Grid Applications Workshop
Hawaii, June 2004

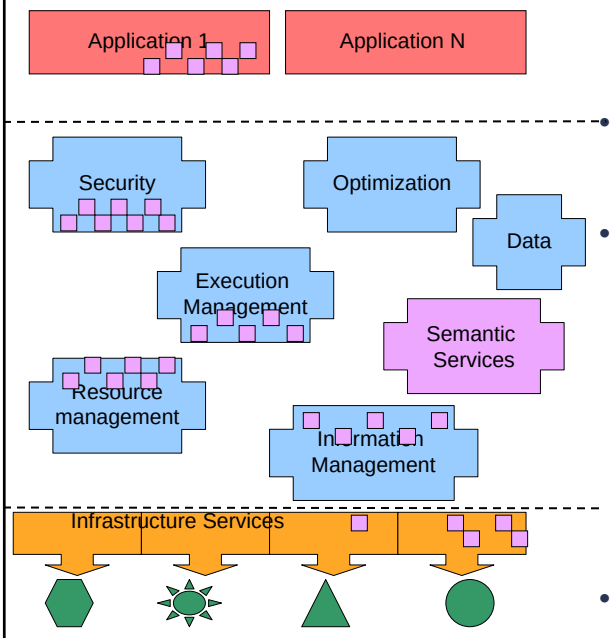
Dagstuhl Seminar
July 2005

GGF16 3rd GGF Semantic Grid Workshop
Athens, February 2006

OGF19 Web 2.0 and the Grid Workshop
North Carolina, January 2007


Grid Research Projects under FP6
supporting the Grid community
international cooperation
EU Funding: 130 M€
Challengers
Nessi-Grid
GridCoord
Degree
Grid@Asia
EchoGrid
Bridge
EC-Gin
Grid services, business models
GridEcon
ArguGrid
Edutain@ Grid
A-Ware
Sorma
trust, security
GridTrust
AssessGrid
Provenance
SIMDAT industrial simulations
BeinGrid business experiments
platforms, user environments
Grid4all
KnowArc
Gredia
QosCosGrid
g-Eclipse
GridComp
UniGrids
HPC4U
NextGRID service architecture
Akogrimo mobile services
BREIN agents & semantic
XtreemOS Linux based Grid operating system
CoreGRID virtual laboratories
data, knowledge, semantics
Chemomen tum
K-WF Grid
InteliGrid
Datamining Grid
OntoGrid
Wave 2 – start 2006
Wave 1 – start 2004
Specific support action
Integrated project
Network of excellence
Specific targeted research project
stralasia 2007
Information Society and Media


RealityGrid
UniGrids
UniGrids
Grid Ontology



Semantic OGSA

- Semantic Grid Reference Architecture
- Mixed ecosystem of Grid and Semantic Grid services
  - Services ignorant of bindings
  - Services binding aware but unable to process them
  - Services binding aware and capable of processing
- Everything is OGSA compliant

NGG3

Service Oriented Knowledge Utility

The architecture comprises services which may be instantiated and assembled dynamically, hence the structure, behaviour and location of software is changing at run-time

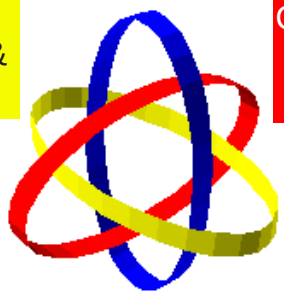
Services are knowledge-assisted ('semantic') to facilitate automation and advanced functionality

A utility is a directly and immediately useable service with established functionality, performance and dependability

Dan Atkins

Transformative Application- to enhance discovery & learning

Provisioning- Creation, deployment and operation of advanced CI

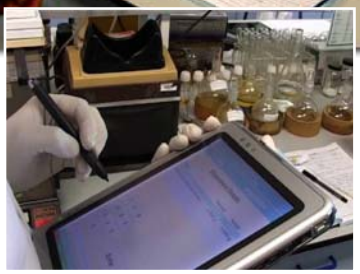


R&D to enhance technical and social dimensions of future CI systems

CombeChem Smart Tea

the Smart Tea Project

www.smarttea.org



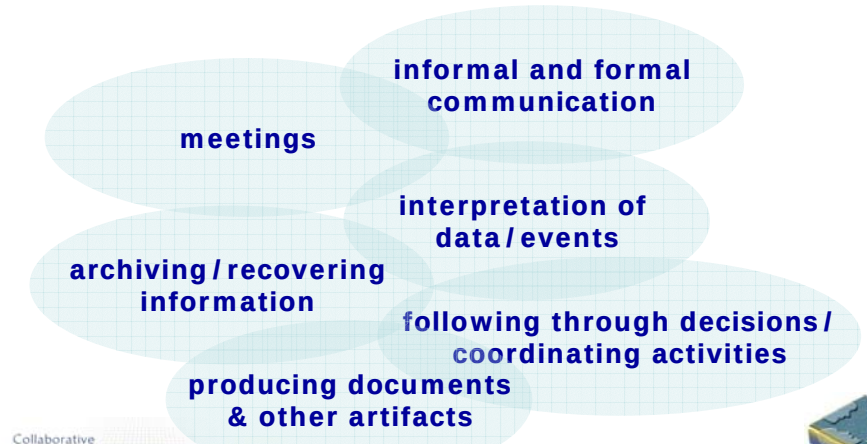
Bioinformatics is not Chemistry

There are many pieces, from many boxes, but no box, and no lid with a complete picture of what the puzzle is supposed to be.

- Planning? No.
- Metadata an afterthought

Carole Goble

Key collective activities in e science

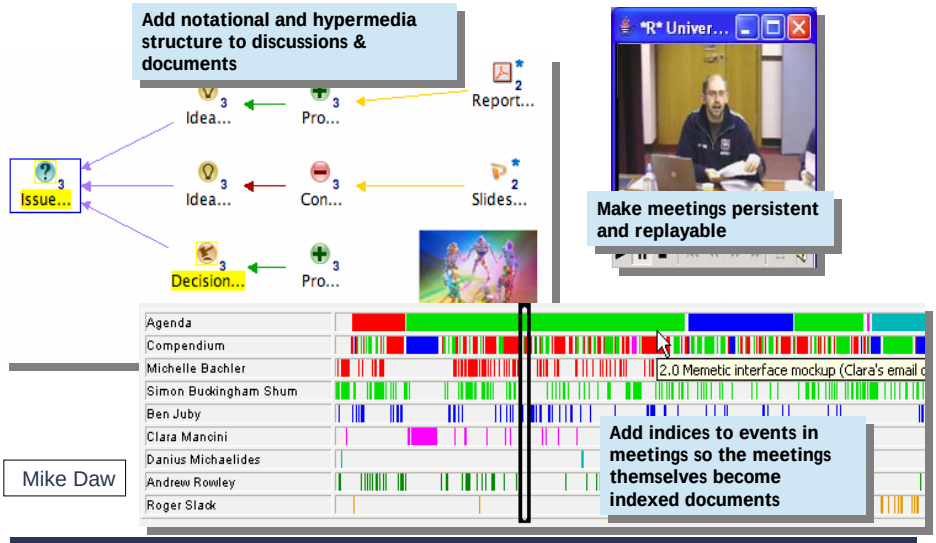


Collaborative ADVANCED KNOWLEDGE TECHNOLOGIES CoAKTinG in the Grid

http://www.aktors.org/coakting/

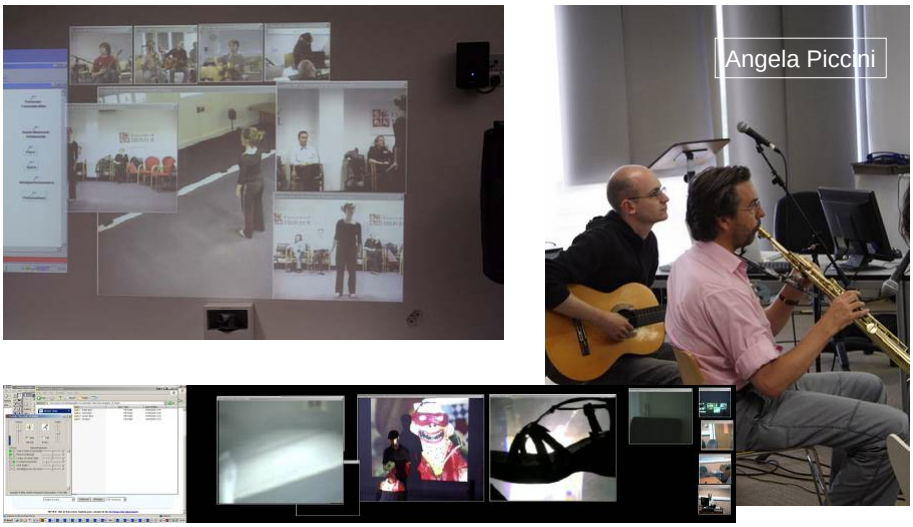
memetic

How can we weave together distributed discourse and documents?



performativity, place, space

Workshop 1: Kyra Norman and Orchestra Cube

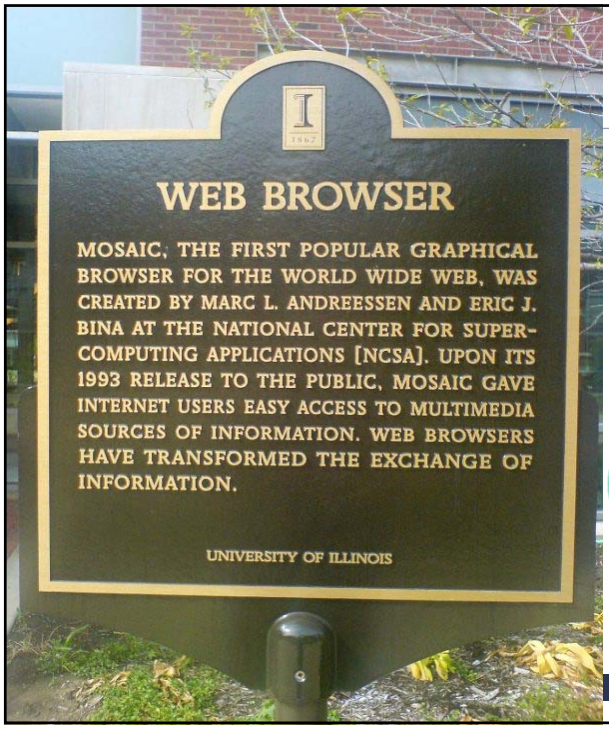


The Collaborative Semantic Grid, CTS 2006

Stephen Downie

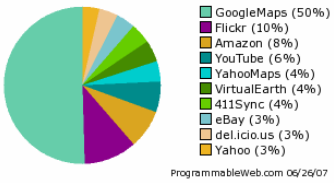
People Principles

- e Science as sense making
- Supporting formal and informal scientific process
- Collaboration over artefacts
- Scaling up from project to community



Evolution

amazon web services™



ProgrammableWeb.com 06/26/07

Web 2.0 Design Patterns

1. The Long Tail
2. Data is the Next Intel Inside
3. Users Add Value
4. Network Effects by Default
5. Some Rights Reserved
6. The Perpetual Beta
7. Cooperate, Don't Control
8. Software Above the Level of a Single Device


O'REILLY

What Is Web 2.0

Design Patterns and Business Models for the Next Generation of Software

Read this article in:

- Chinese

- http://www.oreillynet.com/pub/a/oreilly/tim/news/2005/09/30/what-is-web-20.html

GRID today

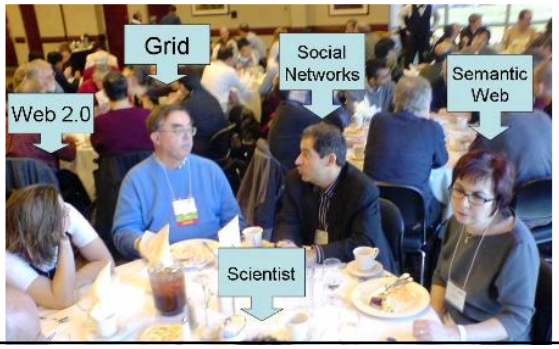
The Leading Source for Global News and Information from the evolving Grid ecosystem, including Grid, SOA, Virtualization, Storage, Networking and Service-Oriented IT

February 19, 2007

Special Features:

Grid Meets Web 2.0 at OGF19

When Google-Gadget Award-winner Pamela Fox told Grid developers at OGF19 that she had written some code at 3 a.m. on Sunday morning and it had 6000 users by Tuesday, the OGF audience knew they had to pay attention. Add to this the fact that the first time she used a Web API was eight months before.



Fox's Web 2.0 developers' tutorial -- entitled "Web 2.0 Mashups: How People Can Tap into the "Grid" for Fun & Profit" -- was one of several invited talks at the workshop organized through the eScience OGF function, on Web 2.0 and the Grid organized at OGF19 by David De Roure. "In Grid and Web 2.0 we see different approaches to building interoperable systems. The workshop was the first crucial opportunity to see what Grid can learn from the successes of Web

Geoffrey Fox

Mashups are workflow (and vice versa)

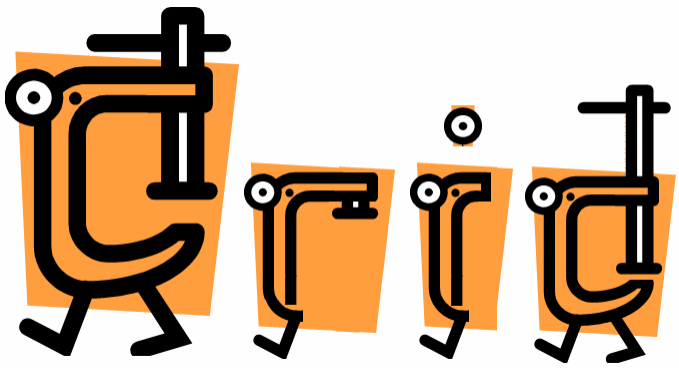
Portals are start pages and portlets could be gadgets

So there is more or less no architecture difference between Grids and Web 2.0 and we can build e-infrastructure or Cyberinfrastructure with either architecture (or mix and match)

We should bring Web 2.0 People capabilities to Grids (eScience, Enterprises)

We should use robust Grid (motivated by Enterprise) technologies in Mashups

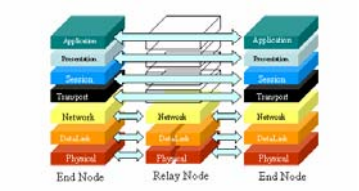
See Enterprise 2.0 discussion at http://blogs.zdnet.com/Hinchcliffe/



Too sophisticated for its own good?

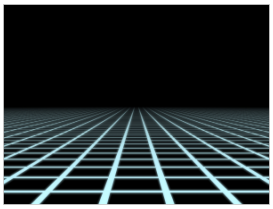
The Grid Mindset

- Provide an advanced infrastructure to enable researchers to do exciting new things
- Middleware hides the complexity of underlying systems, and this needs standards
- Layered model – success is invisible infrastructure



When Grids go bad

- Overengineering of standards
- Assumption that users will come
- Divorces computation from content provision
- Service provider mentality
  - users seen as consumers of services not producers of value



NB These oppose the characteristics of Web 2.0

The Grid Problem

The Web 2.0 community decided Web Services are too complicated so they use HTTP instead.

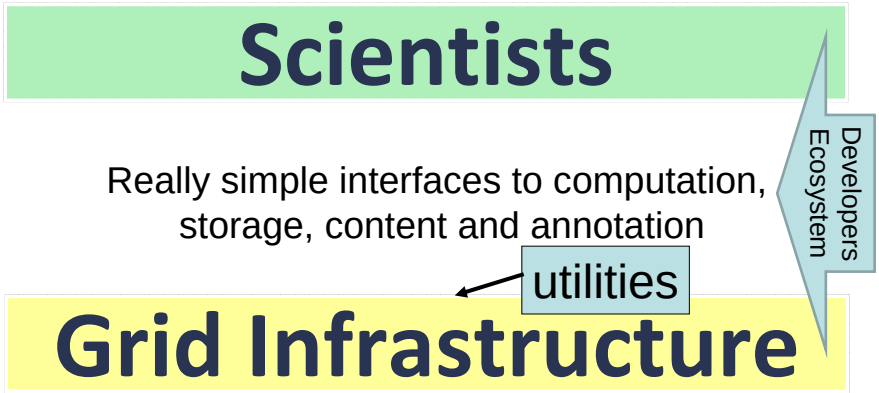
The Grid community decided Web Services aren't complicated enough so they invented OGSA.

Matthew Dovey

When Web goes bad

- Semantic Web is not a quick win
  - Learning curve for concepts and tools
- Return on Investment more visible “in the large”
  - Need to bootstrap to get value

Missing Link



JISC Virtual Research Environments

VRE 1

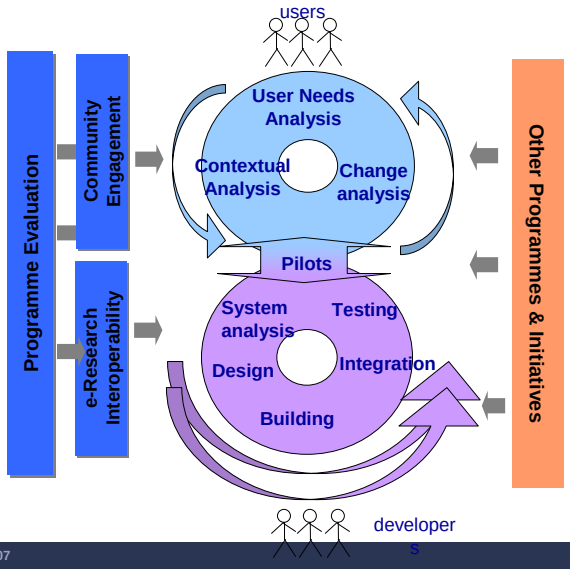
- Technology-focused
- Experimental
- Diverse design & development approaches
- Stand-alone solutions

VRE 2

- User- & research practice-focused
- Developmental
- Unified design & development approaches
- Integrated solutions

- Collaboration
- Supporting small & large-scale research
- Support for single-disciplinary and multi-disciplinary research

Virtual Research Environment Approach



myexperiment beta

from the myGrid team who brought you Taverna.

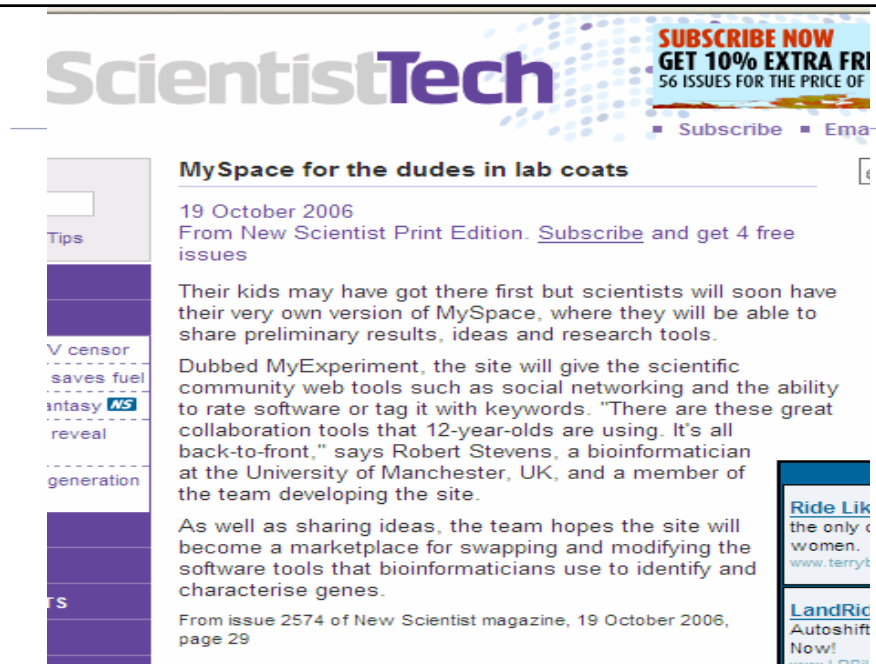
myExperiment makes it really easy for the next generation of scientists to contribute to a pool of scientific workflows, build communities and form relationships. myExperiment enables scientists to share, re-use and repurpose workflows and reduce time-to-experiment, share expertise and avoid reinvention.

> *"Their kids may have got there first but scientists will soon have their very own version of MySpace, where they will be able to share preliminary results, ideas and research tools."*
> — New Scientist Tech, October 2006.

myExperiment introduces the concept of a workflow bazaar; a collaborative environment where scientists can safely publish their creations, share them with a wider group and find the workflows of others. Workflows can now be swapped, sorted and searched like photos and videos on the web.

Carole Goble carole.goble@manchester.ac.uk

David De Roure dder@ecs.soton.ac.uk



ScientistTech

SUBSCRIBE NOW
GET 10% EXTRA FR
56 ISSUES FOR THE PRICE OF

Subscribe Ema

Tips

V censor
saves fuel
antasy NS
reveal
generation

ts

MySpace for the dudes in lab coats

19 October 2006
From New Scientist Print Edition. Subscribe and get 4 free issues

Their kids may have got there first but scientists will soon have their very own version of MySpace, where they will be able to share preliminary results, ideas and research tools.

Dubbed MyExperiment, the site will give the scientific community web tools such as social networking and the ability to rate software or tag it with keywords. "There are these great collaboration tools that 12-year-olds are using. It's all back-to-front," says Robert Stevens, a bioinformatician at the University of Manchester, UK, and a member of the team developing the site.

As well as sharing ideas, the team hopes the site will become a marketplace for swapping and modifying the software tools that bioinformaticians use to identify and characterise genes.

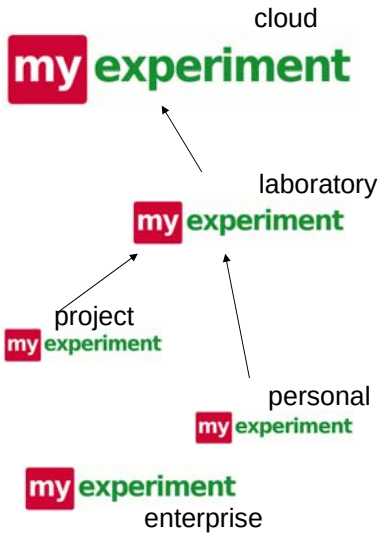
From issue 2574 of New Scientist magazine, 19 October 2006, page 29

Ride Lik
the only
women.
www.terryb

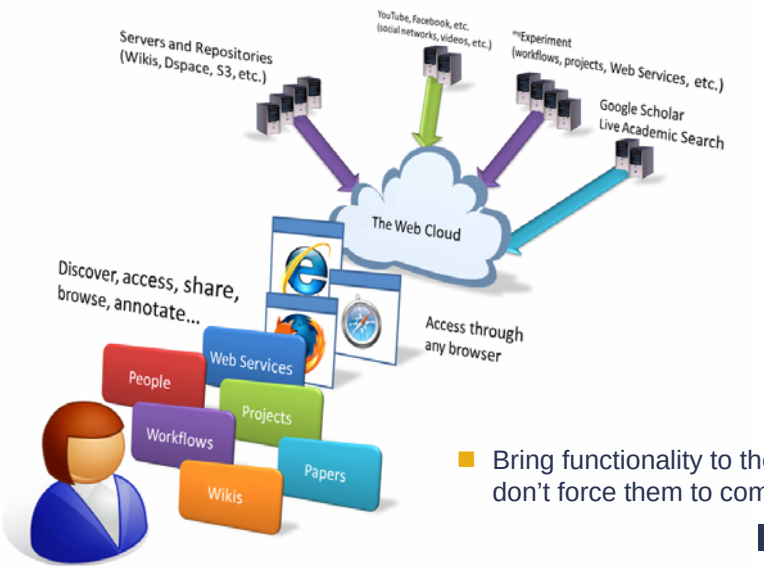
LandRic
Autoshift
Now!

Warehouse or Federation

- Community web site, and Distributed stores
- Multiple myExperiments
- Mixed identity regimes: an identity authority
- Publish what I want when I want within the group I want
- Open Archives Initiative
  http://www.openarchives.org/
- Leveraging from the CombeChem project
  http://www.combechem.org/



Cooperate don't control



- Bring functionality to the user – don't force them to come to you

The Future

- A mixture of Grid, Semantic Web and Web 2.0
- The Grid is about linking things up so that people can do new stuff, so we need to empower people to do *functionality mashups*
- Use Semantic Web technologies (RDF and Ontologies) to assist
  - Mashing up of data
  - Finding and using services
  - Empowering people
  - Working with live feeds, ...
- The Grid community can learn from Web 2.0 in terms of how developers and users engage with the new capabilities
  - bring new functionality to the users rather than expecting them to come to it, and enable them to participate
  - Web 2.0 is compatible with Grid in that it requires robust services underlying it

Messages

- Semantic Grid – *metadata management* and *automation through annotation* – is needed more and more to work with decoupled, disconnected, distant resources In the Wild
- e is for *Empowering Scientists* not just *Enabling Science*
  - Think vertically as well as horizontally!
  - Rise above the pieces, harness the new capabilities, create an ecosystem of participation

Semantic Grid Research Group

semanticgrid.org

See the Call for Participation for the OGF21 Grid and Web 2.0 Workshop

OGF21 Seattle, October 15-19, 2007

David De Roure

dder@ecs.soton.ac.uk

Carole Goble

Geoffrey Fox

Marlon Pierce

Credits, Links and Contacts

- Slides
  - Stephen Downie, Liz Lyon, Geoffrey Fox, Jeremy Frey, Carole Goble, Angela Piccini , Savas
- Teams
  - CombeChem
  - myGrid
  - myExperiment
  - CoAKTinG/Memetic

David De Roure

dder@ecs.soton.ac.uk

Carole Goble

carole.goble@manchester.ac.uk

JISC