

School of Electronics and Computer Science

Semantic Web In Use: Applications and Vocabularies

Dr Nicholas Gibbins nmg@ecs.soton.ac.uk

Overview

Southampton
School of Electronics
and Computer Science

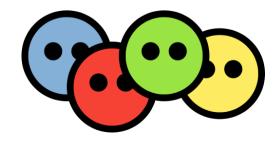
- Social Networking
- Bibliographic Information
- Cultural Heritage
- Business Intelligence
- Public Sector Information
- Research Information
- e-Science



Friend of a Friend (FOAF)

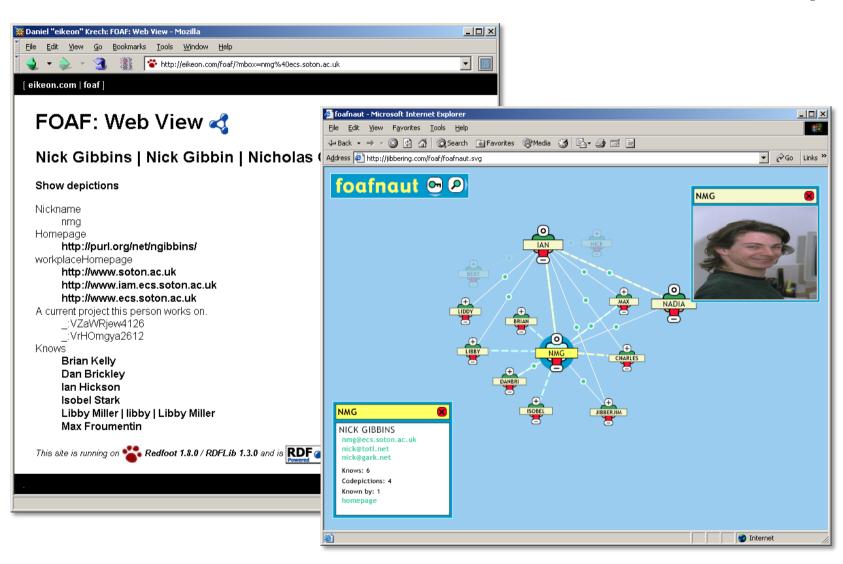


- Semantic Web realisation of "six degrees of separation"
- Fully distributed system
 - Each user writes an RDF file describing themselves and who they know (using the FOAF schema)
 - Files are harvested and aggregated
 - Data is presented in a variety of interfaces
- Avoids siloing of existing social networking systems
- Data from many providers (LiveJournal, etc)



Multiple Interfaces

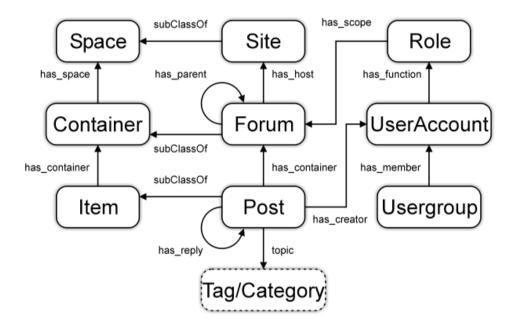




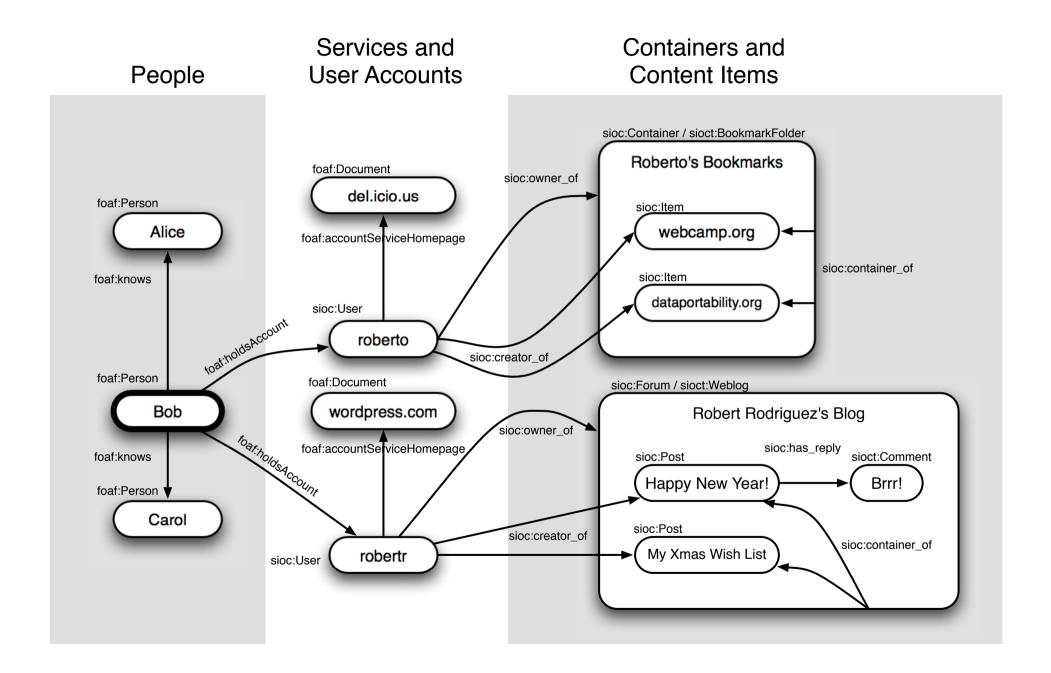
Semantically-Interlinked Online Communities (SIOC)



- Ontology for describing wikis, bulletin boards, blogs, etc
- Designed to interoperate with FOAF









Dublin Core



- A standard set of 15 metadata elements for cross-domain information resource description
- Formally endorsed by ISO, NISO and CEN
- Representation languageneutral
 - Plain XML serialisations in addition to RDF/XML
 - Predates RDF (started in 1995)

- Possibly the most widely used Semantic Web vocabulary
 - Libraries (electronic and physical)
 - Museums
 - Web archives
 - Open Directory Project
 - UK Mirror Service
 - MusicBrainz

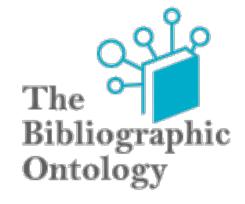
•



Bibliographic Ontology

Southampton
School of Electronics
and Computer Science

- More detailed bibliographic ontology
- Describes types of publication
 - Conference paper
 - Journal article
 - ...
- Builds on Dublin Core
- Used as export format for Eprints 3.2 (more on this later)

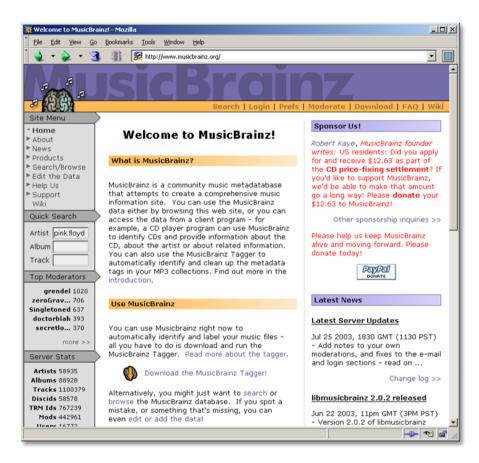


MusicBrainz

Southampton

School of Electronics and Computer Science

- An open replacement for the CDDB Internet CD database
- Complementary to freedb.org
- Community-maintained service
 - Submissions and moderations
- Vocabularies used:
 - Dublin Core
 - Custom ontology for metadata specifically relating to audio recordings



Publishing Standards for Industry Standard Metadata



- Set of metadata vocabularies for automating publishing production processes and content exchange
- Simplified profile of RDF/XML
 - Lower hurdle to adoption
- Commercial users include
 - Time, Inc
 - Lexis-Nexis
 - McGraw-Hill

- Vocabularies used:
 - Dublin Core
 - Custom ontologies for
 - Controlled vocabularies
 - Relational information
 - Resource types
 - Rights management

BBC Programmes

Southampton
School of Electronics
and Computer Science

- Simple ontology for describing programmes
 - Brands
 - Series (seasons)
 - Episodes
 - Broadcast events
 - Broadcast services
- Data describing all BBC programmes



http://www.flickr.com/photos/nicecupoftea/4104234460/

http://purl.org/ontology/po/

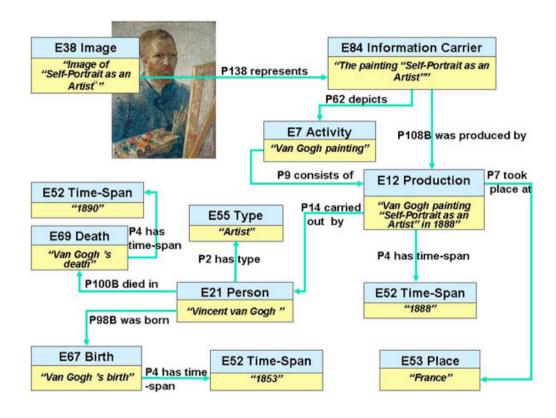
http://www.bbc.co.uk/programmes/developers



CIDOC CRM

- Southampton
 School of Electronics
 - school of Electronics and Computer Science

- Ontology for heterogeneous cultural heritage information
- Represents more than just bibliographic information
- Event-based model (compare with DC's attributebased approach)



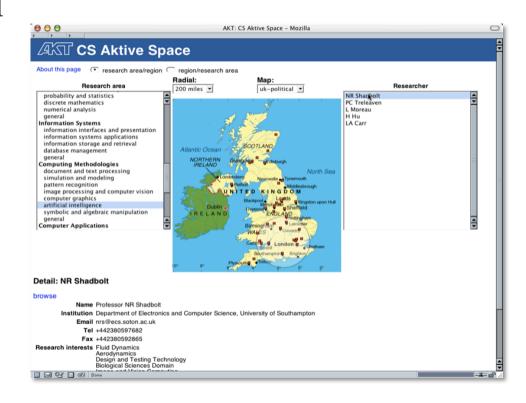


CS AKTive Space (2003)



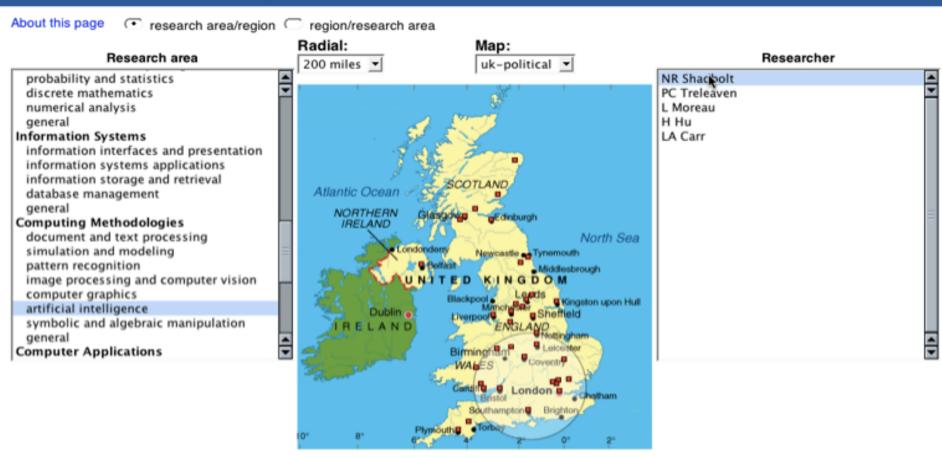
- Repurpose information gathered about CS researchers to help them understand and explore their community
- Data sources:
 - 2001 Research Assessment Exercise submissions
 - Research Council project data
 - Detailed data on personnel, projects and publications harvested for leading CS departments in the UK





000

AKT CS Aktive Space



Detail: NR Shadbolt

browse

Name Professor NR Shadbolt

Institution Department of Electronics and Computer Science, University of Southampton

Email nrs@ecs.soton.ac.uk

Tel +442380597682

Fax +442380592865

Research interests Fluid Dynamics

Aerodynamics

Design and Testing Technology

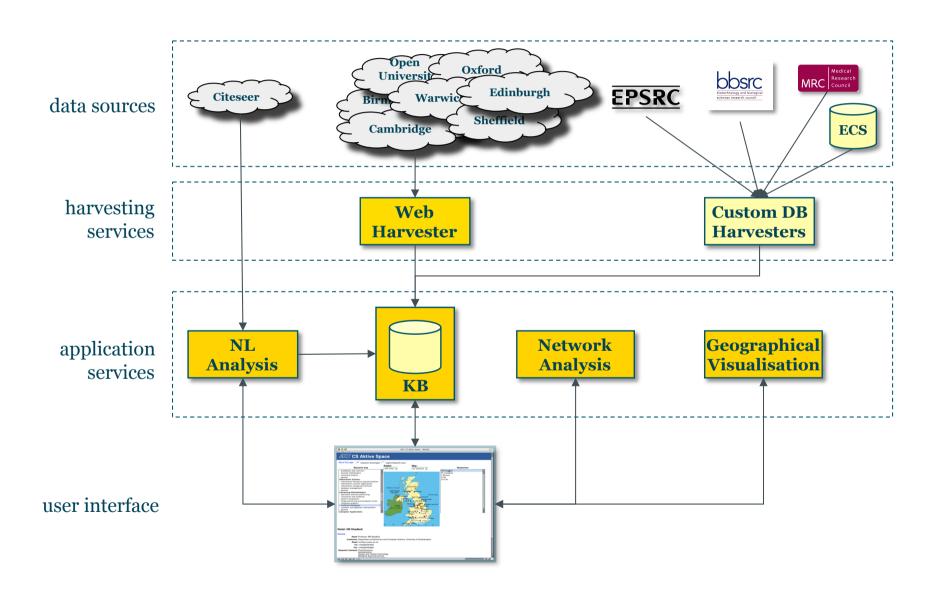
Biological Sciences Domain





System Architecture





dotAC (2009)

Southampton
School of Electronics
and Computer Science

- CS AKTive Space revisited
 - All UK Higher Education
 - Integrated data from institutional research repositories
 - Sustainable live data from research councils
 - Mapping from CERIF
- Extra services
 - Network analysis for related resources (communities of practice)
 - Coreference management identify duplicates in data

http://dotac.info/



AKTive Futures (2004)



- How do we produce a portal for end users within the "strategic decision making" process of an organisation?
- Inform strategic planning and market intelligence
 - Identify relevant drivers
 - Visualise trends in heterogeneous data
 - Present information contextually

Climate and environment

- Storms ahead: weather extremes
- Bio-diversity
- Global transport and industrialisation
- Abrupt climate change
- Sea level rises
- Water shortages
- Kyoto to 'Contraction and Convergence'

Geo-political economy

- US-dominated agenda
- International institutions
- North vs South
- China, India and Asia
- European social agenda
- Scramble for oil in Africa
- US interventionist
- Middle East
- US 'technological sublime'

Terrorism, total surveillance and security

- Global terror
- Future of 'remote wars'
- Everyone is watching
- Human rights
- Privacy and trust

Energy

- Post oil?
- Large to small scale networks
- Hydrogen fuel cells: mobility
- 'Hydrogen highway'
- Costs and pricing
- Demand cuts
- Wind, tidal, solar

Bio-everything

Bio-tech

Bio-remediation

Bio-terror

Bio-error

Bio-agriculture Bio-medical

Bio-informatics

Genomics

Proteomics Bio-systems

Nano-science to nanotech

- Nanobots or goo?
- Materials science
- Nano-machines
- Sensors:RFIDS MEMS
- Drug delivery
- Nano-filtration

Augmented reality, pervasive and mobile computing

- · 'Ambient' computing
- The Grid
- Broadband Mobile
- Wireless
- Bio-inspired computing
- Hybrid iDTV/web
- Remote sensing

Knowledge, Innovation

- The Semantic Web
- Private research
- Public knowledge
- Artificial intelligence
- Intellectual property vs open source
- E-science

Sustainability agenda

- Anti-science: the values debate
- Socio-political conflict over economics and business.
- EU/regulatory drivers
- Public engage vs indifferent
- Corporate social responsibility

Demographics, migration and ageing

- Cures for everything?
- Longer and better?
- Ageing West, pensions crisis, inter- 'Care world' generational tension
- Acceleration of migration

Socio-cultural

- Values, attitudes and lifestyles: fast trends
- •• Anti-corporate
- 'Open source' values
- New trust matrix: NGOs
- Ethical consumers
- Family life
- · Young vs old
- Boomers age
- Hypermobility, transport and urban/rural futures
- · Health 'economy'
- · Culture of fear and anxiety

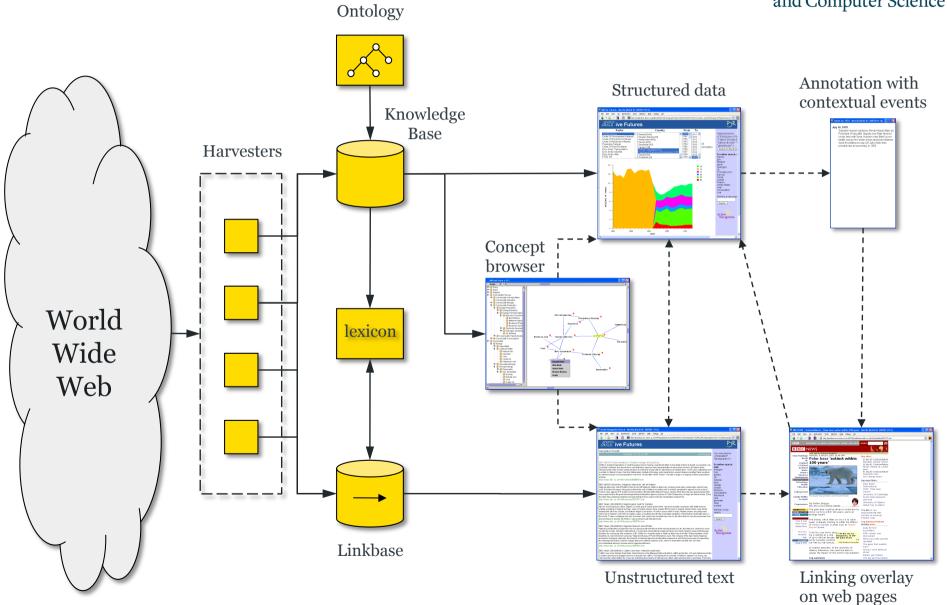
Media

- Conflicts over 'simple/certain' vs 'complex/uncertain
- Plural
- Fragmented public media and discourse
- Single issue moral panics
- Smart mobs
- Mobile opinion formers

System Architecture



School of Electronics and Computer Science



Market Blended Insight (2006-2010)

- Application of Semantic Web technologies to marketing strategies for the B2B sector
- Real data, real B2B processes to ensure real scenarios for the undertaken research
- Workflow:
 - Identify market segments
 - Identify value chains and key relationships
 - Identify opportunities and channels of communication
 - Refine and iterate















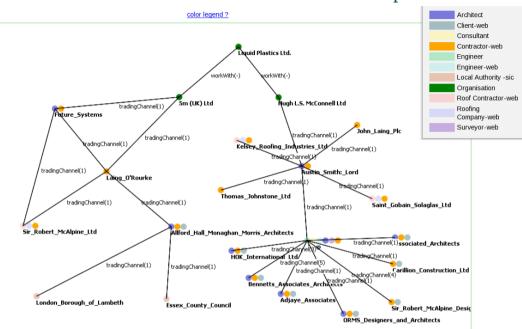
http://mbi.ecs.soton.ac.uk/

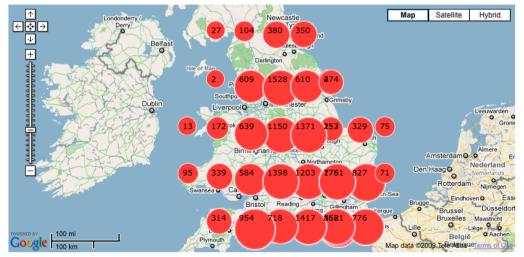
Data Sources and Visualisations

Southampton

School of Electronics and Computer Science

- UK business backbone:
 - ~4M companies
- Ordnance Survey Address Layer
- UK SIC(92) industrial classifications
- Other sources:
 - Local publications
 - Trade publications
 - Contact data for ~300K companies, extracted from company web sites
 - London Gazette (all UK insolvencies)





Good Relations



- Standardised vocabulary for product, price, and company data
- Represents:
 - Prices
 - Accepted payment methods
 - Opening times
 - Delivery methods and charges
 - Warranty promises and scope

- Integrates with:
 - UNSPSC vocabulary for offerings
 - SIC codes for company type





AKTive PSI (2005)



- Use advanced knowledge management technology to improve the delivery of policy and public services across Government
- Build up a detailed picture of life in two London Boroughs, using as broad a collection of Public Sector Information as possible























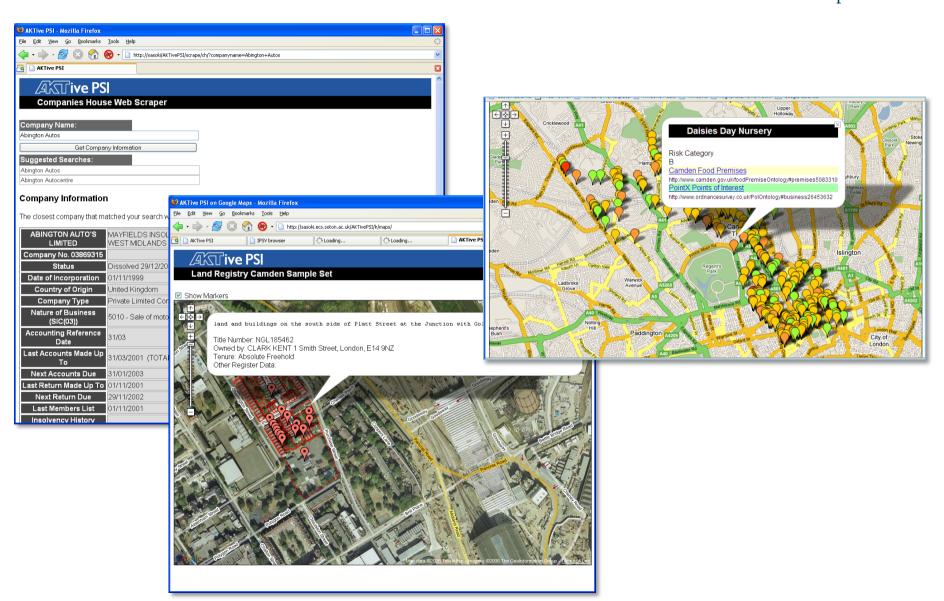




Visualisations and Mashups



School of Electronics and Computer Science



Integrated Public Sector Vocabulary



- Taxonomy for classifying public sector information resources
- Developed with the backing of the ODPM and the Cabinet Office e-Government Unit
- Available in a variety of formats, but RDF/XML is the definitive format
- Uses Dublin Core and SKOS

IPSV Top Level Headings



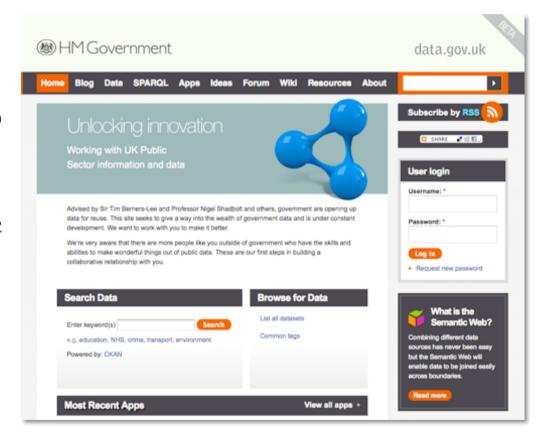
- Business and industry
- Economics and finance
- Education and skills
- Employment, jobs and careers
- Environment
- Government, politics and public administration
- Health, well-being and care
- Housing

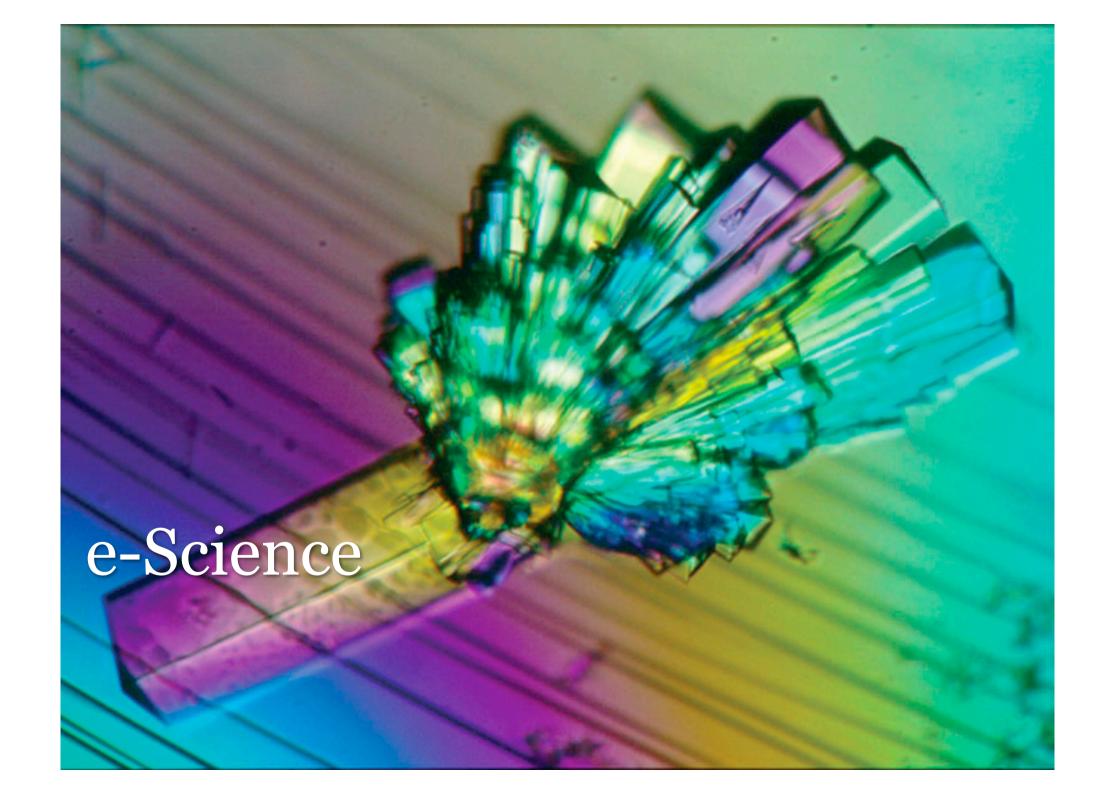
- Information and communication
- International affairs and defence
- Leisure and culture
- Life in the community
- People and organisations
- Public order, justice and rights
- Science, technology and innovation
- Transport and infrastructure

data.gov.uk (2010)



- In 2009, Tim Berners-Lee and Nigel Shadbolt appointed as advisors to HMG on opening up public data
- data.gov.uk website went public on 21 Jan 2010
 - 2890 datasets from central and local government
 - Uses SW technologies to represent data





e-Science



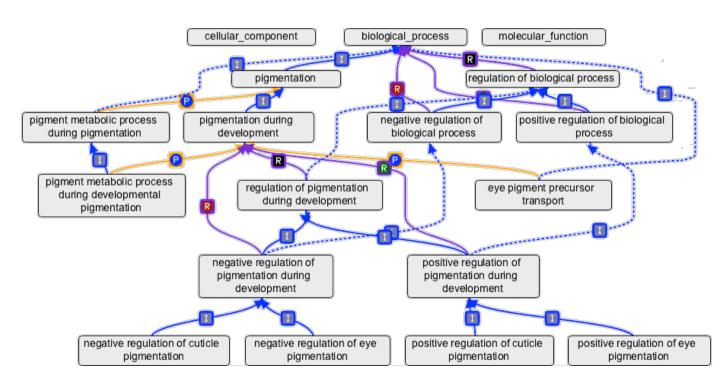
- e-Science characterised as:
 - Large-scale science
 - Distributed global collaborations
 - Very large data collections
 - Very large scale computing resources
- Data integration will be a major issue
 - Capture, publish, reuse data
 - Agreed vocabularies for data exchange

Gene Ontology Consortium



- Controlled vocabularies for the consistent description of gene products
- Simplified profile of RDF/XML

- Multi-view ontology
 - Molecular function
 - Biological process
 - Cellular component



http://www.geneontology.org/



[Term]

id: CL:0000540

name: neuron

def: "The basic cellular unit of nervous tissue. Each neuron consists of a body\, an axon\, and dendrites. Their purpose is to receive\, conduct\, and transmit impulses in the nervous system." [MESH:A.08.663]

xref_analog: FBbt:00005106

xref_analog: FBbt:00005146

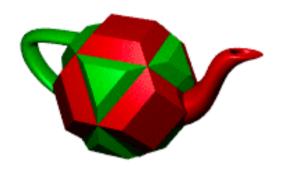
is_a: CL:0000393! electrically responsive cell

is_a: CL:0000404! electrically signaling cell

relationship: develops_from CL:0000031! neuroblast

The Smart Tea Project





- Improving the information environment for chemists – both within and beyond the lab
- Supporting chemists in the preparation, execution, analysis and dissemination of their work





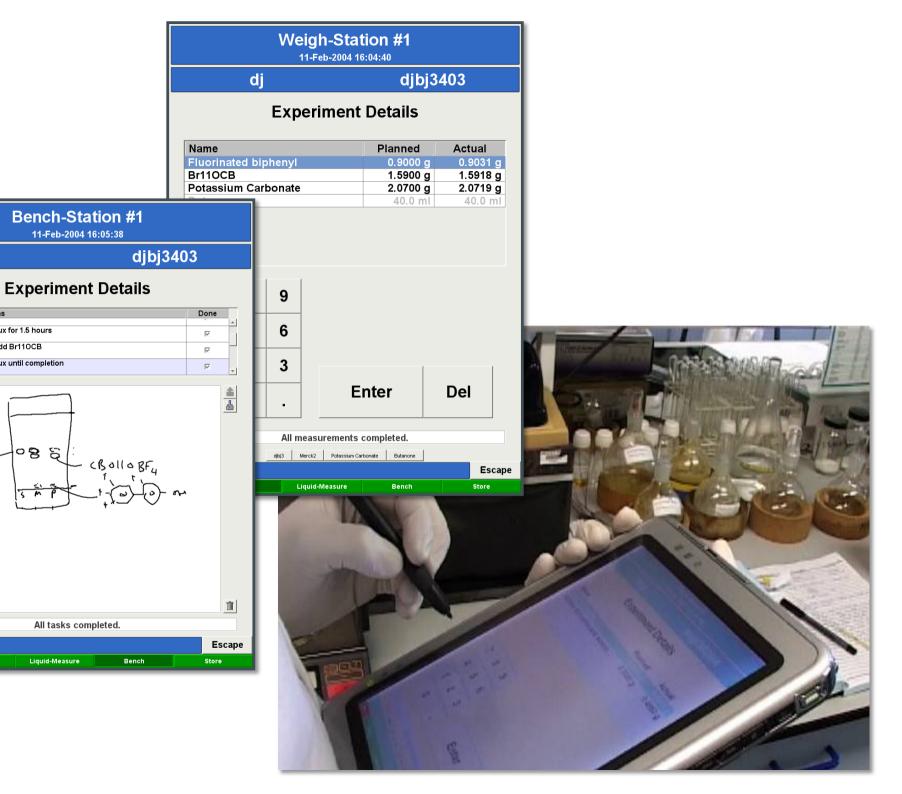


Data Capture: The Lab Notebook



School of Electronics and Computer Science

2-21-39	Prep. of 179 g amiline 1750 cc dil. H2SO4 (168 cc cour. H2SO4) Swith rapid stiming
	137g Na NO2 in 1050 cc of water
	2443 in 7000 cc ag. containing 100g NaOH. 3000 cc ag. containing 240g NaOtt.
	179 g of aniline (1.93 moles) was dissolved in 1750cc of dil. H2SOy conty. 2.98 moles H2SO4 (168 cc conc. H2SO4). Cooled to zero with rapid stirring. The with end of dropping funnel below surface, 137 g (1.98) moles of NaNO2 in 1050cc of water were added very sloroly, rate being so, that no large excess of nitrous acid at any time. Allowed to stir for two hours at 0°, at which time a slight starch iodide paper test.



dj

Instructions

CBOUB

Weigh

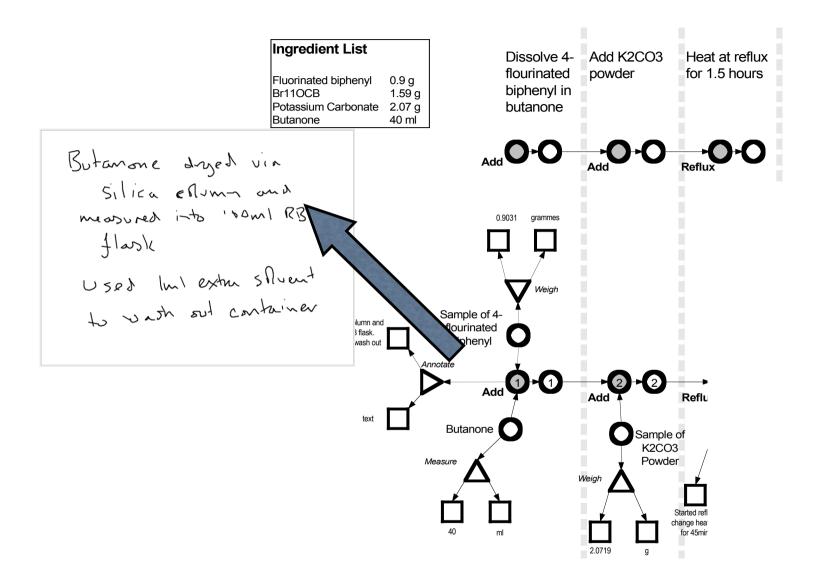
Heat at reflux for 1.5 hours

Heat at reflux until completion

380

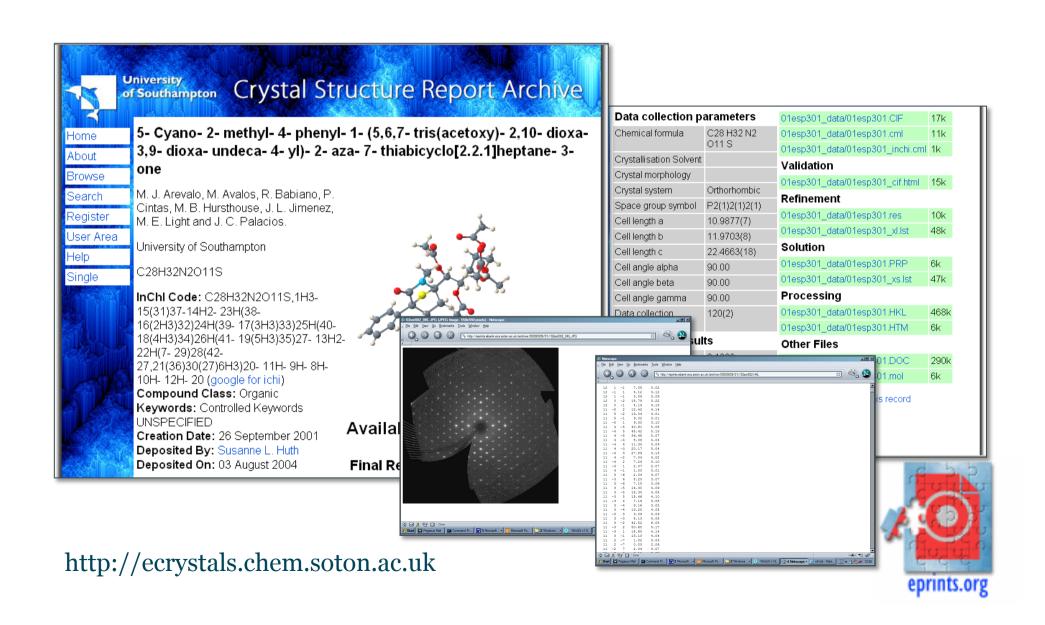
Liquid-Measure

Cool and add Br110CB



Publish and Reuse





Exchange Vocabularies



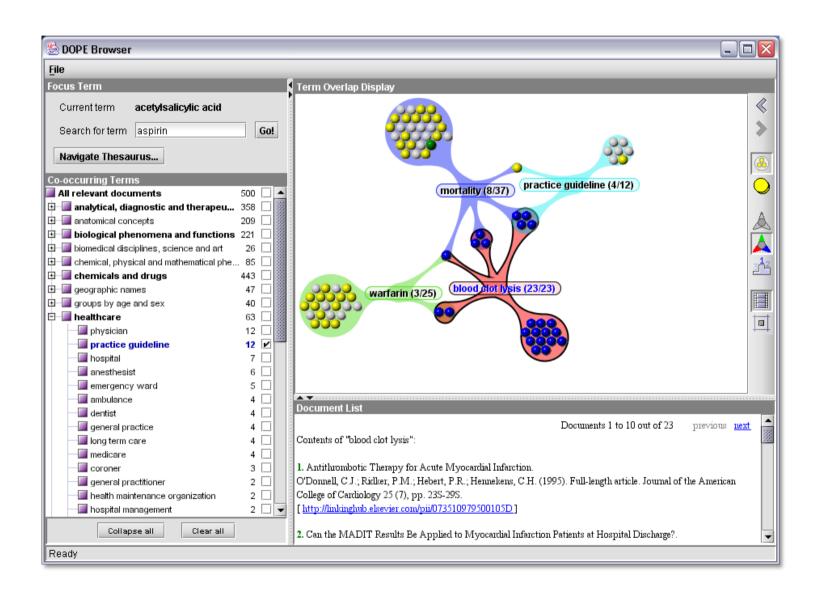
- BioPax Ontology (biological pathways)
 - Metabolic and signalling pathways, molecular interactions
- Gene Ontology (genes and gene products)
 - Molecular function, cellular component, biological process
- NCI Cancer Ontology
 - Diseases, drugs, anatomy, genes

(and many others from other disciplines)

Advanced Visualisations



School of Electronics and Computer Science



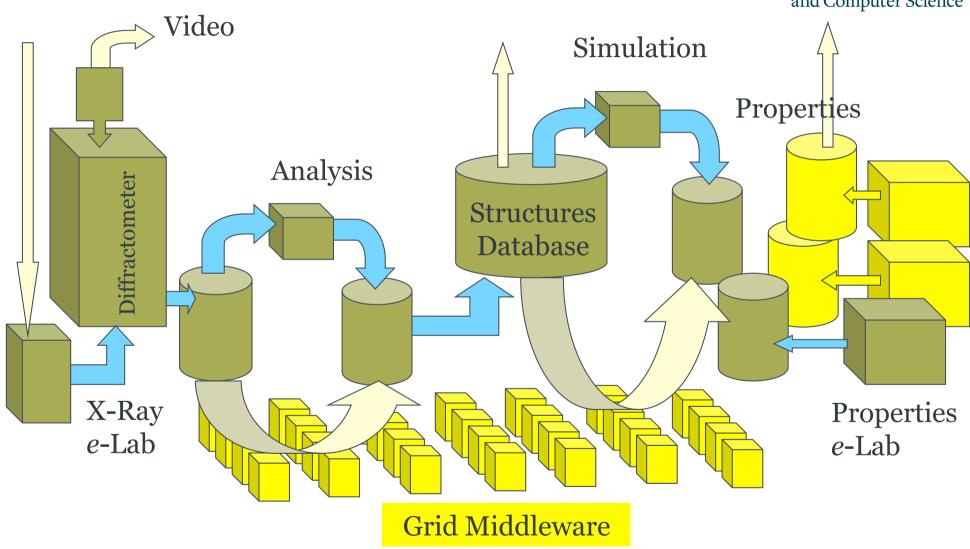
Grid Computing



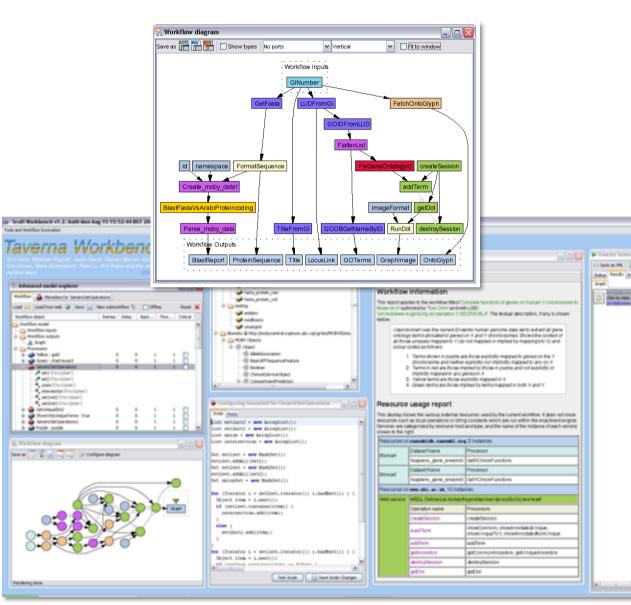
- e-Science applications typically have very high computational requirements
- Grid Computing provides an infrastructure for
 - Flexible, secure, coordinated resource sharing
 - Dynamic collections of individuals, institutions and resources
 - Virtual organisations
 - Workflow management
- Social computing, in effect



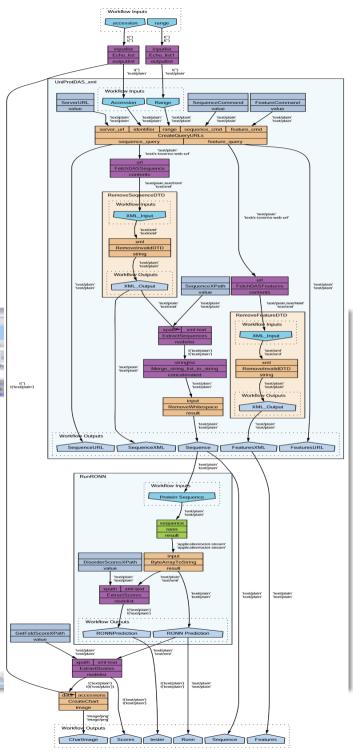
School of Electronics and Computer Science



http://www.combechem.org/



http://www.mygrid.org.uk



The Semantic Grid



- Grid Computing + Semantic Web
- Information and services are given a well-defined meaning
 - Uses SW technologies OWL, RDF, etc
 - Ontologies for describing services
- Better enables computers and people to work together
 - Requires coordination and planning capabilities found in agent technologies