

The Social Grid

Leveraging the Power of the Web and
Focusing on Development Simplicity

Tony Hey
Corporate Vice President of Technical Computing at Microsoft



successful standards efforts

- TCP/IP versus ISO Protocols
 - ISO Committees disconnected with operational world
 - IETF solved and responded to real operational problems
- MPI community message-passing standard
 - Joint effort between research community and the major vendors (IBM, Meiko, Intel, TMC, Ncube)
- Web
 - Pioneered by research community
 - HTML and HTTP were a pragmatic compromise

the grid community

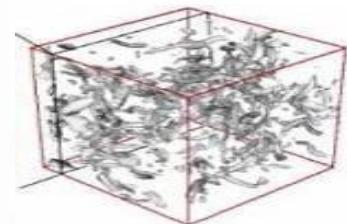
- Initiated by the research community
 - Distributed, collaborative computing
 - HPC Centers
- Successful Grids in industry
 - Finance, Pharma, Oil & Gas
 - Intra-Grids in Business
- Cloud Service providers
 - Google, Amazon, eBay, Yahoo and Microsoft

emergence of a new science paradigm

- Thousand years ago – **Experimental Science**
 - Description of natural phenomena
 - Last few hundred years – **Theoretical Science**
 - Newton's Laws, Maxwell's Equations...
 - Last few decades – **Computational Science**
 - Simulation of complex phenomena
 - Today – **eScience or Data-centric Science**
 - Unify theory, experiment, and simulation
 - Using data exploration and data mining
 - Data captured by instruments
 - Data generated by simulations
 - Data generated by sensor networks
- Scientist analyzes databases/files
➤ Commercial and academic applications



$$\left(\frac{\dot{a}}{a}\right)^2 = \frac{4\pi G\rho}{3} - K \frac{c^2}{a^2}$$



(With thanks to Jim Gray)

motivation

- Grid middleware and Web Services are becoming very complex and hard to use
 - Need to go back and look at needs of users
- Recognize the transition to data-centric eScience
- Provide support for domain scientists
- Focus on user and application requirements
- Need for simplicity and ease of development
- Learn lessons from the Web

the web

- Has tremendous momentum
- The browser is the universal canvas for the delivery of information and functionality
- Web protocols, technologies, and middleware are well supported by the IT industry
- Easy to use collection of ‘Web 2.0’ technologies is maturing
- Emergence of “Software-as-a-service”

users are both consumers and producers

- Users generate content on the Web
 - Blogs, wikis, photographs, videos, etc.
 - They do not have to know HTML
- They form communities
 - Social networks, virtual worlds
- They interact, collaborate, share
 - Instant messaging, web forums, content sites
- They consume information and services
 - Search, annotate, syndicate

scientists today...

- Annotate, share, discover data
- Collaborate, exchange ideas over the Web
- Create communities, social networks
- Use workflow tools to compose services

example - connotea

search

The screenshot shows a web browser displaying the Connotea search results for the query "social networking". The search bar at the top has "social networking" entered. A blue callout box on the right side of the screen contains the text "find resources bookmarked by other users".

Tags used on these bookmarks:

- social networking
- social bookmarking
- social software
- india
- networking
- php
- social computing
- MySQL
- Web 2.0
- developers
- programmers
- Outsourcing
- application
- social tagging
- web
- collaboration
- social networks
- digital library
- emergencies
- health library
- Open
- social apps
- Source
- news
- solutions
- Social

Bookmarks with search terms social and networking

Note: Your search term matches the global tag [social networking](#).

Number of bookmarks per page: 10 | 25 | 50 | 100

Social network - Wikipedia, the free encyclopedia
en.wikipedia.org
Section 4, as of 5/1/2007, is about "Social Networking, Internet Social Networks."
Posted by [ascoppin](#) to [social computing](#) [Web 2.0](#) on [Tue May 01 2007](#) at 19:26 UTC | [info](#)

Social Networking Leaves Confines of the Computer - New York Times
www.nytimes.com
Posted by [library mistress](#) to [networking](#) [Social](#) [social software](#) on [Mon Apr 30 2007](#) at 15:34 UTC | [info](#)

Social-networking sites link Hispanic youth
www.cnn.com
MIAMI, Florida (AP) -- Indie rocker Eric Monterrosa checks his ElHood.com Web page at least three times a day, answering fans, surfing for other new Latin artists and keeping in touch with friends from his native Colombia.
Posted by [msgbeep](#) to [news](#) on [Sun Apr 29 2007](#) at 16:47 UTC | [info](#)

Social networking in the health context
www.ingentaconnect.com
Software and services for creating online social networks.
Posted by [Spiky](#) to [social apps](#) on [Tue Apr 24 2007](#) at 15:58 UTC | [info](#)

Hospital Staffing
Revenue Cycle Staffing including PFS, Pt. Access, Med. Records
www.hrgpros.com

[Report a problem](#)

Related tags:

- social bookmarking
- folksonomy
- bookmarking
- collaborative - tagging
- tagging
- collaborative tagging
- csdl-picasso-folks

Internet | Protected Mode: On

the web application platform

- Service composition as a way to build distributed applications
 - Service-orientation

Yahoo pipes
Aggregate news about “Grid Computing”

The screenshot displays two windows of the Yahoo Pipes application. The left window shows the visual editor with a complex network of nodes (URL Builders) and connections. The right window shows the resulting news feed, which includes articles from various sources such as "Restructuring to delivering Telco 2.0 services", "Elements of High Performance Computing part 3...", "Energy Management", "Myspacers for the World Community Grid!", and "Life, The Universe and Academia". The news feed is titled "Copy of Aggregated News Alerts" and includes a "Configure This Pipe" section where "Grid Computing" is selected.



mashups: composing data and functionality



SensorMap

Functionality: Map navigation

Data: sensor-generated temperature, video camera feed, traffic feeds, etc.

the web application platform

- Services on the cloud
 - Blogging
 - Data processing/transformation
 - Content upload, sharing, discovery
 - Storage, computation, messaging

<http://ecrystals.chem.soton.ac.uk>

Thanks to Jeremy Frey

The screenshot displays a web page from the University of Southampton's Crystal Structure Report Archive. The main title is "Crystal Structure Report Archive". The left sidebar includes links for Home, About, Browse, Search, Register, User Area, and Help. The main content area shows the following details for a compound:

Compound Information:

- Chemical Name: 6,7,9,10,12,13,15,16-Octahydro-benzo-1,4,7,10,13-pentaoxacyclopentadecin
- Author: Simon J Coles, Michael B Hursthouse, Jeremy G Frey and Esther Rousay.
- InChI: InChI=1/C14H20O5/c1-2-4-14-13(3)18-11-9-16-7-5-15-6-8-17-10-12-19-14/h1-4H,5-12H2
- DOI: 10.594/ecrystals.chem.soton.ac.uk/145
- Compound Class: Organic
- Keywords: crown ether
- Depositor Comments
- Creation Date: 07/07/2008
- Deposited By: A. C. Coles
- Deposited On: 2008-07-07

Data collection parameters:

Chemical formula	C ₁₄ H ₂₀ O ₅
Crystallisation Solvent	
Crystal morphology	Plate
Crystal system	Orthorhombic
Space group symbol	Pbc _a
Cell length a	16.4963(18)
Cell length b	8.325(3)
Cell length c	20.061(6)
Cell angle alpha	90.00
Cell angle beta	90.00
Cell angle gamma	90.00

Refinement results:

Solution figure of merit	0.0409
R Factor (Obs)	0.0487
R Factor (All)	0.0977
Weighted R Factor (Obs)	0.1008
Weighted R Factor (All)	0.1192

Validation:

- 04sjc0831.cif (13k)
- 04sjc0831.cml (6k)

Refinement:

- 04sjc0831_res (6k)
- 04sjc0831_xl.lst (34k)

Solution:

- 04sjc0831.prp (6k)
- 04sjc0831_xs.lst (39k)

Other Files:

- 04sjc0831.hkl (702k)
- 04sjc0831.htm (10k)
- 04sjc0831_0h.jpg (57k)
- 04sjc0831_h0h.jpg (85k)
- 04sjc0831_hk0.jpg (88k)
- 04sjc0831_crystal.jpg (17k)

Changelog:

Citation: Coles, S.J., Hursthouse, M.B., Frey, J.G. and Rousay, E. (2004). Southampton, UK: University of Southampton, Crystal Structure Report Archive. (doi:10.594/ecrystals.chem.soton.ac.uk/145)

amazon web services: simple storage service (s3)

- S3 is storage for the Internet
 - Designed to make web-scale computing easier for developers
- Provides a simple Web Services interface to store and retrieve any amount of data from anywhere on the Web
 - ‘CRUD’ philosophy – Create, Read, Update and Delete operations
- Uses simple standards-based REST and SOAP Web Service interfaces
 - Built to be flexible so that protocol or functional layers can easily be added

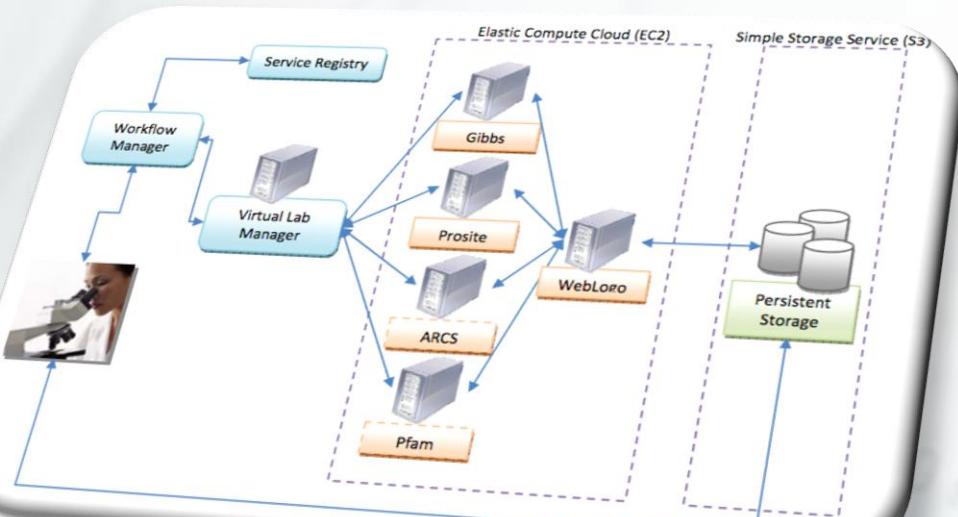
amazon s3 functionality

- Intentionally built with a minimal feature set
 - Write, read, and delete objects containing from 1 byte to 5 gigabytes of data each
- Can store unlimited number of objects
 - Each object is stored and retrieved via a unique, developer-assigned key
- Authentication mechanisms provided
 - Objects can be made private or public, and rights can be granted to specific users
- Default download protocol is HTTP
 - BitTorrent protocol interface is provided to lower costs for high-scale distribution

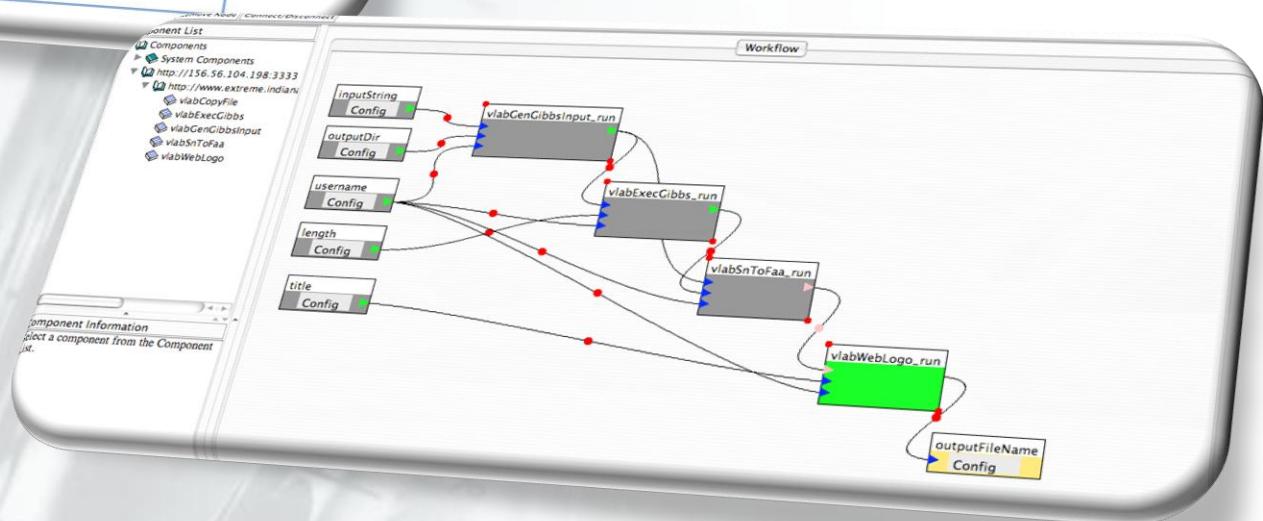
amazon web services: elastic compute cloud (ec2)

- Compute on demand service that works seamlessly with their S3 storage service
- Create Amazon Machine Image (AMI) containing application, libraries and data
- Use EC2 Web Service to configure security and network access
- Use EC2 to start, terminate and monitor as many instances of your AMI as you want
 - Each instance has:
 - 1.7 GHz x86 Processor
 - 1.75 GB RAM
 - 160 GB local disk
 - 250 MB/s network bandwidth
- Used by Catlett and Beckman as capacity computing alternative to TeraGrid ‘SPRUCE’ capability computing for emergency urgent response

a grad student project using s3 and ec2



Gene Analysis Virtual
Lab Experiment
by Jong Youl Choi
at Indiana
(For Beth Plale and Sun Kim)



ogf and standards for grid computing

- Protocol-based integration
 - Based on the Web Services stack
 - WS-* stack has become very complex
 - Danger that Grid-* stack will become too complex
 - Much effort on systems management specifications
 - HPC Basic Profile as the example of a domain-specific specification with industry adoption
- Standards adoption
 - Why are major Grid users/software providers not more engaged with OGF?

grids in industry

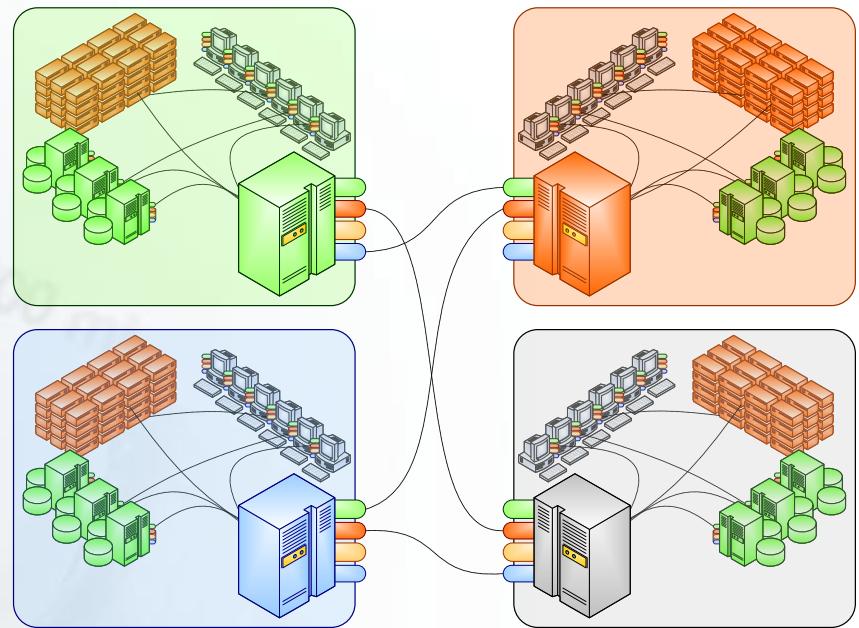
- Google, Amazon, Yahoo, eBay and Microsoft are the major ‘Cloud Platform’ providers
 - All have infrastructures of hundreds of thousands of servers
 - Many large data centers, distributed across multiple continents
 - Have developed proprietary technologies for job scheduling, data sharing and management
 - Care about power consumption, fault tolerance, scalability, operational costs, performance, etc.
 - They are living the “Grid dream” on a daily basis

google as an example

- Estimated 450,000* servers distributed around the world

*source: Wikipedia

- Google File System - highly distributed, resilient to failures, parallel, etc.
- Schedulers and load balancers for the distribution of work
 - Use their 'Map-Reduce' middleware as parallel computational model
- What is financial/competitive advantage for them to adopt standards for these in-house developed solutions?



industry and standards?

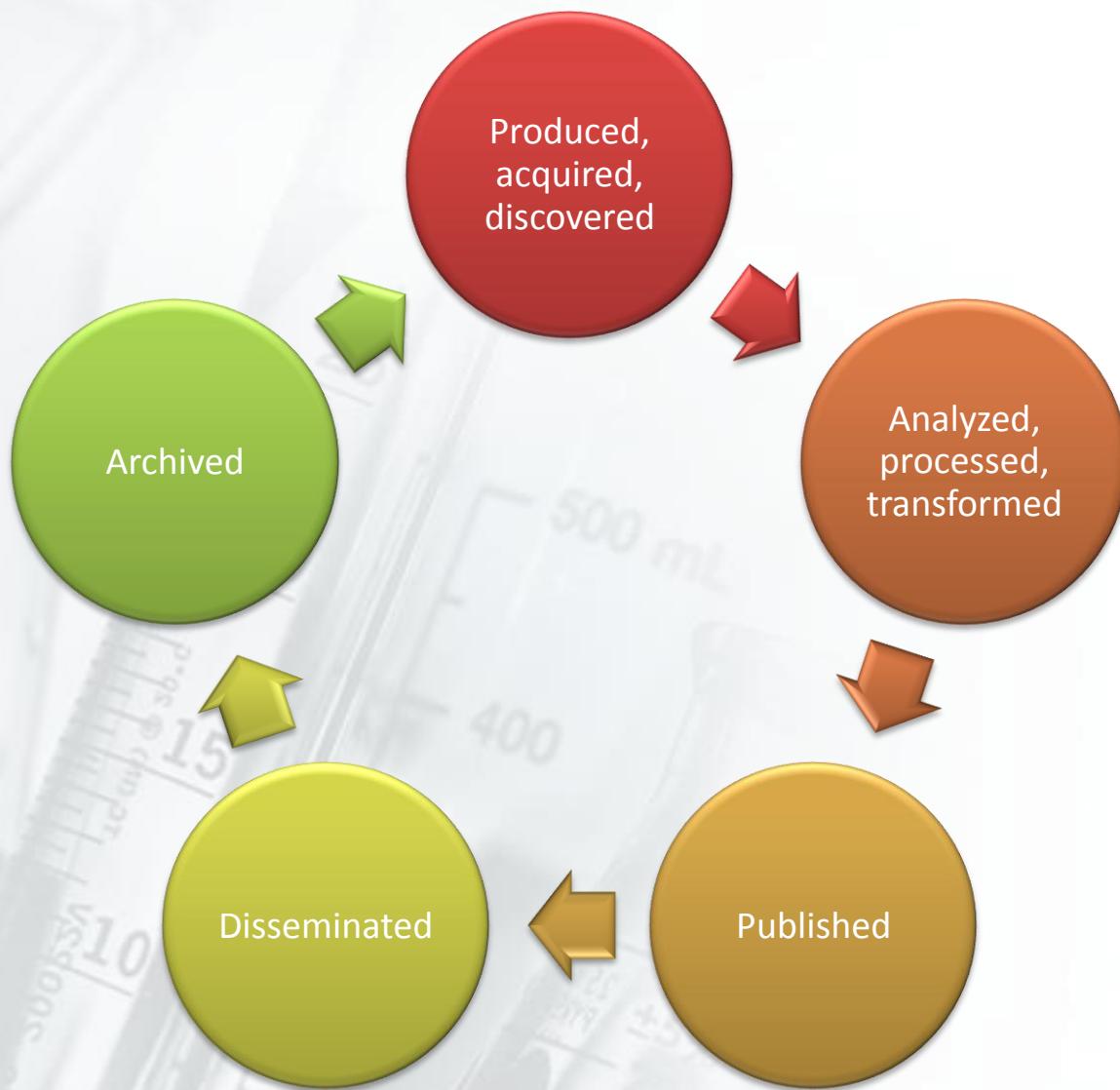
- Why are they not more involved in Grid standards?
 - See competitive advantage in their proprietary infrastructure and proprietary services
- No need for interoperation between companies
 - Will not see map-reduce tasks from Google being offloaded to Amazon's infrastructure any time soon
 - Security and privacy are still major concerns
 - Probably still cheaper for them to maintain their own infrastructure than to offload to someone else's
- There are some possibilities for standardization
 - WS-Management standards could reduce the operational cost of large collections of commodity infrastructure components - motherboards, disks, network hubs, etc.

customer demand can drive interoperability solutions

Example of Security

- Customers want secure access/confidentiality no matter what service they use
- No competitive advantage in doing simple security in a proprietary way
 - Customer demand is forcing Microsoft's Passport, Google's ID, Liberty Alliance, OpenID to interoperate
- More complicated 'VO' scenarios have yet to be proven widely useful
 - Require cross organization trust, credential delegation, id federation
 - Work has already started in this space e.g. WS-Trust, WS-Policy
 - VOs required for some research collaborations
 - Perhaps B2B scenarios like Supply Chain Management can provide commercial driver for their wide adoption

scientific data on the web is...



eScience: data is easily accessible

The collage consists of four overlapping screenshots from environmental science websites:

- AmeriFlux Network:** Shows a map of North America with green and blue shading representing ecosystem fluxes. A sidebar provides information about the network's objectives, organization, and participant sites.
- LTER Network:** Features a map of the United States with various research sites marked. It includes sections on the history of LTER, its mission, and news items.
- National Weather Service Hydrologic Information Center:** Displays a map of the US and surrounding islands with green dots representing gauge locations. It includes a search bar and a sidebar with climate-related links.
- National Climatic Data Center:** Shows a map of the world with various climate-related research projects like Co-Weetea Hydrologic, Cedar Creek Study, and Bonanza Creek. It also features a "Web Services API" interface for data retrieval.

With thanks to
Catharine van Ingen

eScience: data is easily shareable

The website presents data from the Sloan Digital Sky Survey, a project to make a map of a large part of the universe. We would like to share the beauty of the universe, and share with you our excitement as we build the largest map in the history of the world.

SkyServer Tools

- Famous places
- Get Images
- Visual Tools
- Explore
- Search
- Object upload
- CasJobs

Science Projects

- Basic
- Advanced Challenges
- For Kids
- Games and Contests
- Teachers
- Links to other projects

Info Links

- About Astronomy
- About the SDSS
- About the SkyServer
- SDSS Data Release 5
- SDSS Project Website
- Open SkyQuery
- Images of RC3 Galaxies

Contact Us

News

The **SDSS** is supported by

What's new in DR5
What's new on this site
Known problems
More...

Help

SDSS DR5 Image List Tool

obj list page 1

274-51913-230 J103915.59-003918 275-51910-275 J104412.23+000900

278-51900-112 J11222.08-001518 278-51900-225 J10821.84-001218

281-51614-230 J112426.16-002537.2 282-51658-167 J11355.51-003

249-51699-582 J10268.88+641221.6 363-51703-328 J170256.87+603346.8

358-51818-349 J172343.24+570025.1 387-51791-72 J000258.56+000831.1

389-51795-481 J001529.76+003623.9 390-51900-196 J002043.91-002623.9

390-51900-464 J002143.68+001745.5

SpecObj

AB Spectra

SpecLine

XCredStat

Spectrum

Plate

FITS

NED search

SIMBAD search

ADS search

Notes

Save in Notes

Show Notes

Print

Drawing options

- Grid
- Label
- PhotoObjs
- SpecObjs
- Targets
- Outline
- BoundingBox
- Fields
- Masks
- Plates
- InvertImage

SDSS J113459.47+002509.1

GALAXY ra=173.747818, dec=0.419213, ObjId = 588848900446814264

mode PRIMARY

status TARGET PRIMARY OK_STRIPE OK_SCANLINE PSEGMENT RESOLVED

flags STATIONARY MOVED BINNED1 CHILD

PrimTarget TARGET_GALAXY

nm	rerun	camcol	field	obj	rowc	colc
756	44	4	387	58	549.4	1974.6
19.55	18.04	17.55	17.33	17.19		
18.04	17.54	17.54	17.55	17.99	17.55	
0.08	1.796	parentid	nChild			

SpecObjID = 79597814924967936

plate	mid	fiberid	z	zErr	zConf	specClass	ra	dec	fiberMag_r	objId
282	51658	494	0.000	0.00006	1	STAR	173.74778	0.41924	17.81	588848900446814264
						#Status	XCORR_HC			
						#Warning	NOT_GAL			
						PrimTarget	TARGET_GALAXY			
						SecTarget				
						eClass	-0.015			
						emZ	0.000			
						emConf				
						xcZ	0.000			
						xcConf	1			

Sloan Digital Sky Server/SkyServer
<http://cas.sdss.org/dr5/en/>

eScience: services expose functionality

BLAST service delivered through a Web browser

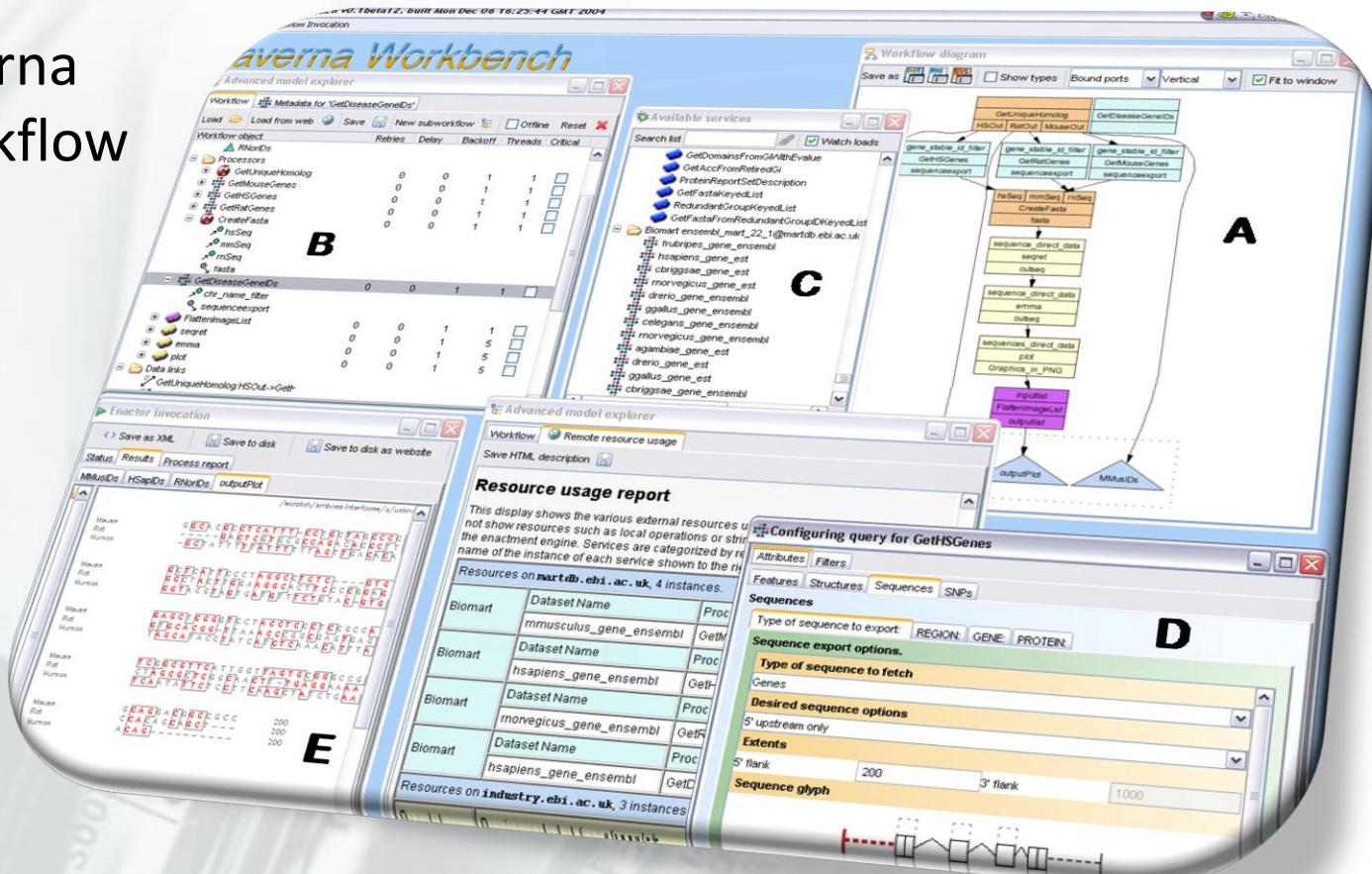
BLAST service (WSDL) that can be integrated into an application

The image displays several screenshots of the BLAST service interface:

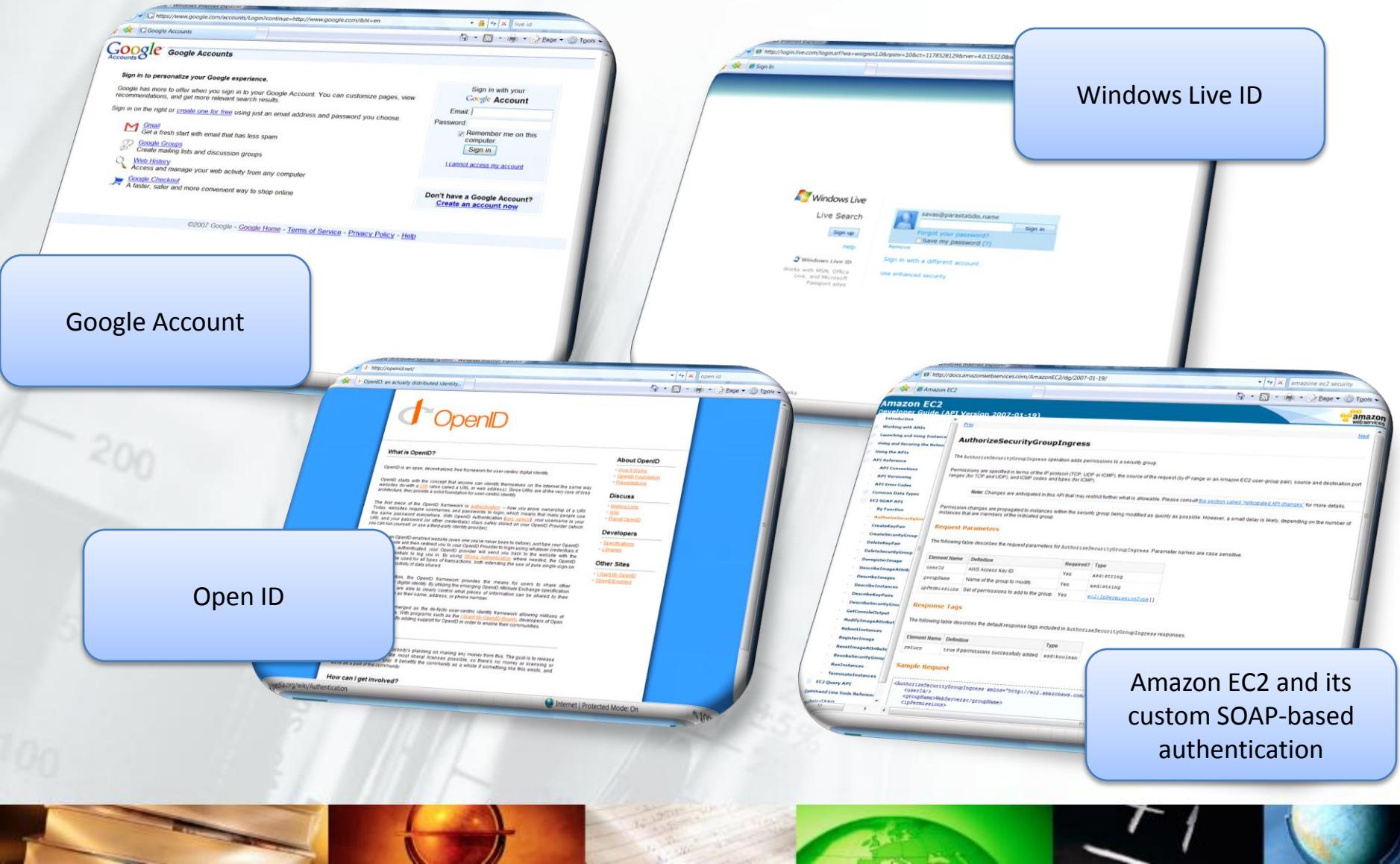
- Top Left:** A screenshot of the BLAST search interface in a Windows Internet Explorer browser. It shows the main search form with fields for Query sequence, Database, and Search Set.
- Top Right:** A screenshot of the BLAST search interface in a Mozilla Firefox browser, showing a more detailed view of the search parameters and results.
- Middle Left:** A screenshot of the BLAST search interface in a Microsoft Edge browser, showing the search results page.
- Middle Right:** A screenshot of a web page showing the WSDL (Web Services Description Language) definition for the BLAST service. The XML code defines the service endpoint, message definitions, and complex types like `EntrezQuery` and `EBIParams`.
- Bottom Left:** A screenshot of the BLAST search interface in a Microsoft Edge browser, showing a 'Tip of the Day' box with the text: "Use Genomic BLAST to see the genomic context".

eScience: services can be composed

Taverna Workflow



eScience: data and services can be accessed securely



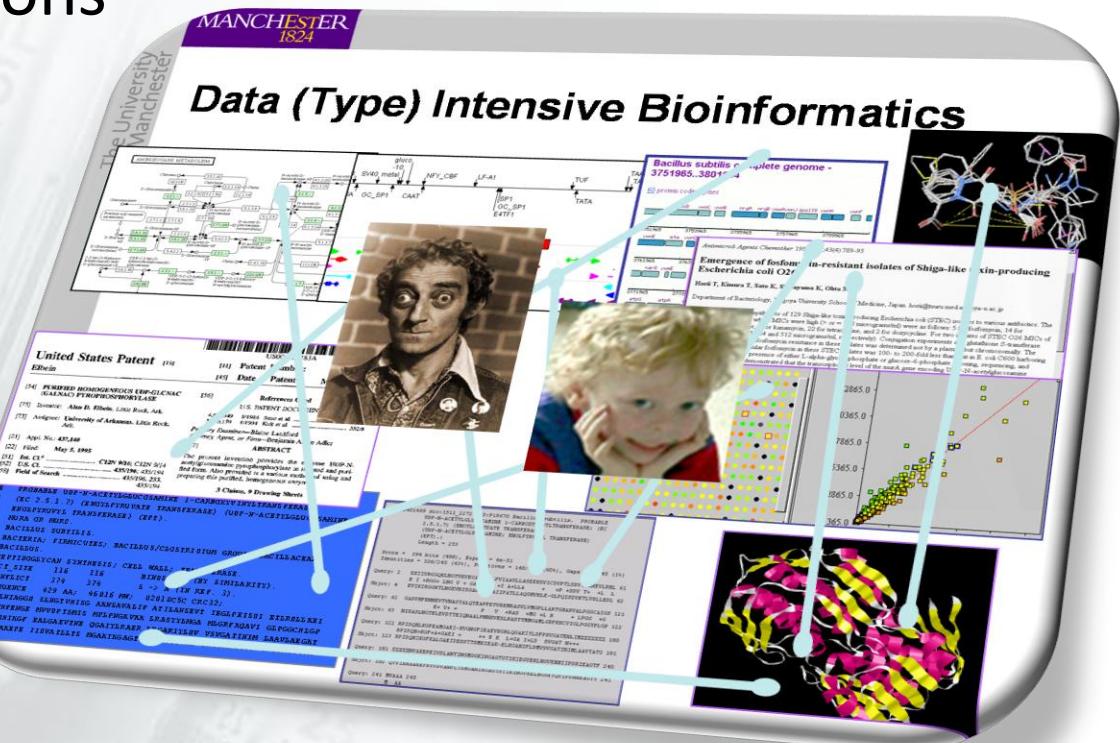
security on the web

- HTTPS
 - XML Encryption
 - WS-Security
 - Single-sign on
 - Windows Live ID
 - Open ID
 - Google ID
 - Virtual Organizations through portals
 - Aggregation of services into one portal, with single authentication
- Grid Security?
 - Need to cater for many different user communities

eScience: knowledge can be created/published/archived/discovered

- Semantic relationships between different data
- Semantic descriptions of services
- Annotations
- Provenance
- Repositories
- Ontologies

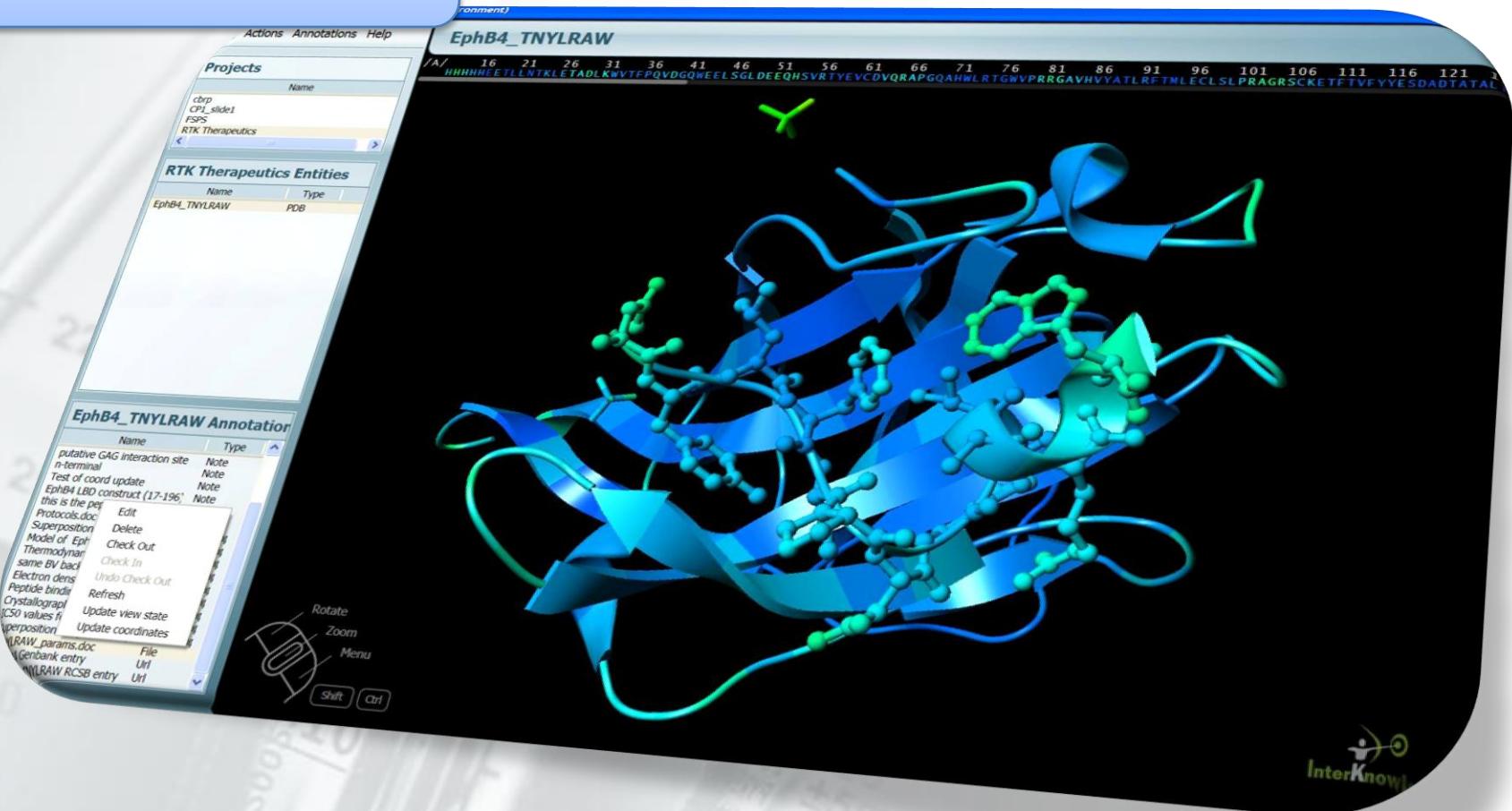
myGrid



eScience: new software tools can make a difference

SCRIPPS

(see www.microsoft.com/science)



a social grid built around the web?

Cross-organization Grids (Collaboration)

- Trust
- Policies
- Federation
- etc

Intra-organization Grids (Cluster)

- HPC Basic Profile
- Scheduling
- Data storage
- etc

Data Center Grids

- Management
- Virtualization
- Quality of service, resource reservation
- etc

Social Grids

- Cloud services, social networking, mashups, semantics
- Existing, simple technologies
- Scientific/technical computing
- End-user /application, not infrastructure, requirements as a priority

social grids and the web

- Focus on **solutions** for scientific/technical computing and **not just on infrastructure**
- Focus on “data-centric eScience”
 - Help domain experts define formats for representing and annotating domain-specific data
- Keep it simple, build on known Web technologies
 - Solutions that “just work” without the need for complicated middleware platforms
 - Leverage only existing, Web infrastructure (HTTP, XML, simple Web Services, services in the cloud)

call to action

- Need for a new type of Grid
 - Domain-specific
 - Provide support for scientists to manage their domain's data
 - Engage with them to define domain-specific data formats
 - Educate them on the use of the Web as a platform for managing their domain's information
 - Demonstrate the use of semantics to manage knowledge
 - User-driven requirements
 - Scientists are our largest audience, let's help them

Microsoft[®]

Your potential. Our passion.[™]

