Experiments on the Grid

The UK National Crystallography Grid Service

Jeremy Frey
CombeChem Project
School of Chemistry
University of Southampton, UK

Plan

• UK e-Science & The Comb-e-Chem Project
• Security!
• The National Crystallography Grid Service
  • Architecture
  • Demo
  • Data Grid
• The future

March 2004
UK e-Science programme

Support for collaborative research

Grid Infrastructure

Computational Grid - compute on demand

Data Grid - flow of data

Comb-e-Chem Partners

• Comb-e-Chem
• Bristol
• Chemistry
• IBM
• CCDC
• Pfizer
• GSK
• AZ
• Southamton
• ECS
• Chemistry
• Stats
• Combi Centre
• IUPAC
• RSC

OMII

IT

Innovation

NCS

March 2004
People

**Chemistry (Southampton & Bristol)**
- Mike Hursthouse, Chris Frampton, Jon Essex, Jeremy Frey, Guy Orpen, Stephan Christensen, Thomas Gelbrich, Sam Pepe, Hongchen Fu, Graham Tizard, Suzanna Ward, Lefteris Danos, Jamie Robinson, Kieron Taylor, Chris Woods, Rob Gledhill

**National Crystallography Service (NCS)**
- Simon Coles, Mark Light, Ann Bingham, Peter Horton

**Electronics and Computer Science (Southampton)**
- Dave De Roure, Luck Moreau, Mike Luck, Hugo Mills, Graham Smith, Simon Miles, Nicky Harding, Gareth Hughes, Nick Humphries, Monica Schraefel, Terry Payne

**It-Innovation (Southampton)**
- Mike Surridge, Ken Meacham, Steve Taylor, Daren Marvin

**Statistics (Southampton)**
- Alan Welsh, Sue Lewis, Ralph Manson, Dave Woods

**Rutherford Appleton Laboratory, Atlas Data Centre**
- IBM – Colin Bird, Syd Chapman

March 2004
Plan types:

- Small set of fixed plans (NCS)
- Variable plans, written by chemist (difficult!)
- Ad-hoc, implied by process execution

Continuum of plan types:

Synthetic Chemistry

Phys Chem data analysis

Experiments on the Grid

National Crystallography Grid Service

March 2004
Grid Service – Why?

- The submitter of a sample will have a more detailed knowledge and understanding of it and hence be able to purposefully contribute to the experiment.
- The users of the service are distributed across the UK, yet need to collaborate closely with NCS staff.
- NCS staff are busy with 'demanding' samples and the ability of a user to manage their routine samples greatly relieves this pressure.
- High throughput demands effective sample management and tracking, especially when a user has multiple samples in the system.
- A user can monitor 'out of hours' experiments to ensure they complete successfully and that most effective use of 'instrument time' is made.

March 2004

Security and trust for experiments and data

“On the Internet, nobody knows you're a dog.”

March 2004
Security

- Security is crucial to the successful operation of the NCS Grid Service,
  - authentication of users
  - maintaining the integrity of their data.
- For NCS, we have set up a well-defined trust network, with its own CA and RA.
- All data transfer is encrypted,
  - A user is authorised to access only their own data, or monitor their own experiments.
  - the user’s credentials are mapped to the appropriate authorisation or datasets.

Security Overview

- Security risk management
- Security technology development
  - process-based authorisation extensions
  - WS-Security message processing
- Security implementation
  - operating policies and public key infrastructure
Asset-Based Security

- Identify and value assets
- Identify threats and risks
- Identify and cost defences

Risk Analysis

- Define risk management approach
- Implement defences

Risk Management

Application to NCS Service

- Assets:
  - campus system and network integrity (M/H)
  - sample tracking data (M)
  - experimental result data (L/M)
  - grid service integrity (L/M)
- Risks:
  - system attacks from outside campus (H)
  - systems attacks from inside campus (H)
  - compromise of remote user credentials (H)
  - internal user error (M)
Public Key Infrastructure

- Requirements:
  - be able to authenticate “singleton” remote users
  - be easy to operate by Chemists
  - be secure enough for academic users & Industry?

- Analysis of existing NCS authentication:
  - uses personal knowledge of user community
  - uses contextual information (e.g. EPSRC project codes)
  - lightweight for both NCS and their customers

- Public key infrastructure developments:
  - Comb-e-Chem certification policy agreed
  - procedures developed for NCS to certify remote users
  - operational responsibility transferred to Chemistry

March 2004
Implementation

Do we know them?
Is this allowed?
Allow that next

Message Handler
Auth

Process and respond

Action and Response
Message Records
Authorisation Records
Certificate Store

ECSES Architecture

Global GRID
GRID COMPUTE SERVER
GASS
NGS GATEWAY SERVER

SOAP/HTTPS/PGP

3rd Party Database
Lab Database

Remote Client Site
DMZ
Lab Resources
Firewall Management Issues

Remote Client Site  |  Campus Network  |  Lab DMZ

\[\text{Globus GRID} \rightarrow \text{GRID COMPUTE SERVER} \]
\[\text{SOAP/HTTPS/PGP} \rightarrow \text{GASS} \]
\[\text{3rd Party Database} \rightarrow \text{Lab Database} \]

e-worries

\[\text{GTi} \rightarrow \text{WSRF} \]

Must ensure this is not a problem for applications
NCS - How it works

- The Status Service determines the client’s Distinguished Name (DN) from their NCS certificate
  - queries the Sample Database
  - list of all samples submitted by the client
  - they may only see their own samples.
- client’s browser, shows the status of each sample.
  - The client may regularly track the progress of their samples within the NCS system.
- Once a sample enters the Running state,
  - a link is made available to the Control Service, whereby the client may monitor the running experiment.

Experiment Control Services

- The Control Service provides the client with a portal to their running X-ray diffraction experiment
  - opportunity to observe the experiment in progress, and to steer it if they wish.
- The display is continuously updated,
  - reflects the current state of the experiment
  - prescans, unit cell determination, full data collection and data processing.
- Scanned images and other raw data are collected by the diffractometer, and published via the portal
  - enabling the client to make informed decisions at each stage whether to continue the experiment, etc.
Experimental Steering

The client can examine the images returned for the prescan, to determine the quality of the crystal. They may then choose to reject the crystal if they wish. A timeout ensures that the experiment may continue, if the client is not present (or takes too long to decide!).

Once the experiment has completed, a summary of the results is automatically published and linked to the Sample Status page.
National Crystallography Service: NCS

SYNTHON Project

Voronoì Project

Property Prediction Services

Structure Determination Services

Single Crystal Structure Determination

Powder Diffraction

Single Molecule Shape Determination
NCS WORKFLOW

Data Grid

- Issues of interaction with existing databases
- Graphical front ends currently make for complex query architecture over a grid
- This is being addressed and will employ a workflow system with full provenance
Increased Remote Control?

- Remote control of equipment
  - Interaction with people & equipment
- Safety Critical Systems
  - External & Internal control of systems can lead to safety conflicts
  - Safety critical software (avoid at this stage)
Video – fun but useful with limited bandwidth

VNC ideal for connecting existing programs – developing Web Service/SOAP VNC but how much control?
Very useful inside the service – but not suitable for deployment over the grid

In the future will employ Access Grid techniques
Safety!

High Power Laser  X-Ray Target

Interlock Systems

Mission Critical Software

Future

• Smart Lab front end
  • Transfer URI as well as sample giving access to the materials preparation
  • Capture of crystal selection process with user interaction.
  • More automated data analysis
  • Automatic publication (xtl-prints)
HT-Raman

- Instrument & Data Service
- Long scans – need for remote user control

Web sites?

www.combechem.org
www.soton.ac.uk/~xservice
Comb-e-Chem

Changing the way we work

E-Lab: X-Ray Crystallography
E-Lab: Combinatorial Synthesis
E-Lab: Properties Measurement

Quantum Mechanical Analysis
Properties Prediction
Data Mining, QSAR, etc.
Design of Experiment

Data Provenance
Data Streaming
Visualisation
Agent Assistant

Laboratory Processes
Samples
Properties DB

March 2004