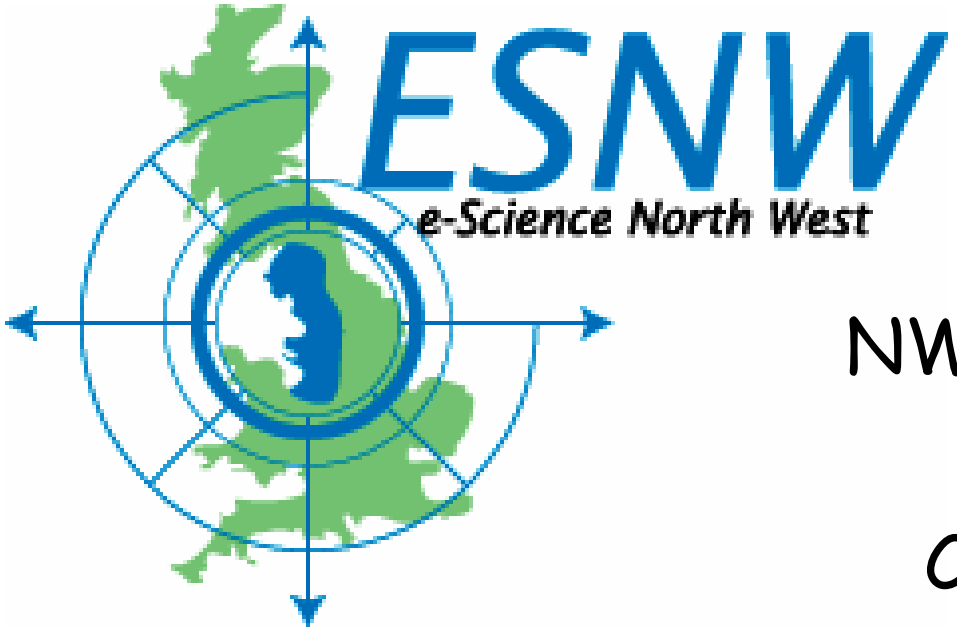


NW-GRID March 2006

John Brooke
Co-Director ESNW

http://www.esnw.ac.uk

Summary of talk

- Structure of NW-GRID
- Some types of application
- Patterns of Grid usage
- Multi-site projects and applications
- Role of data analysis and visualization
- Middleware for NW-GRID

NW-Grid £4.9 million from NWDA 2005-8

THE UNIVERSITY of LIVERPOOL

LANCASTER UNIVERSITY

CLRC

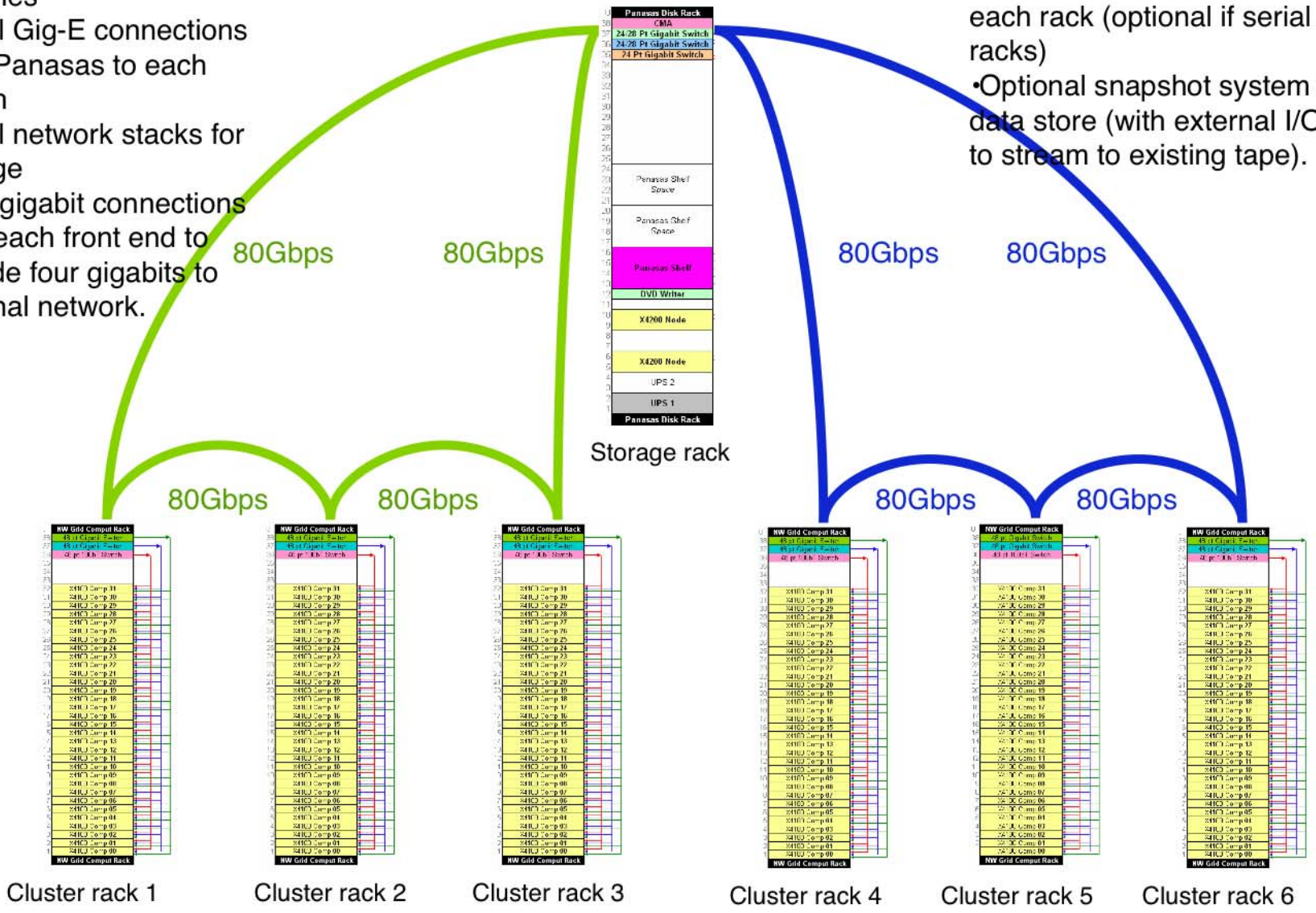
- A regional infrastructure but linked to NGS
- FOUR grid computational nodes linked to rich facilities, for experiment and visualization.
  - Explore bioinformatics databases for organism gene expression analysis
  - Molecular simulation of environmental chemical processes
  - Design of nano-structures with specific material properties
  - Remote access to experimental facilities (e.g., SRS, 4GLS, Jodrell Bank)
  - Statistical modelling of labour market efficiency in social sciences
  - Aiding drug discovery via high throughput screening of molecules

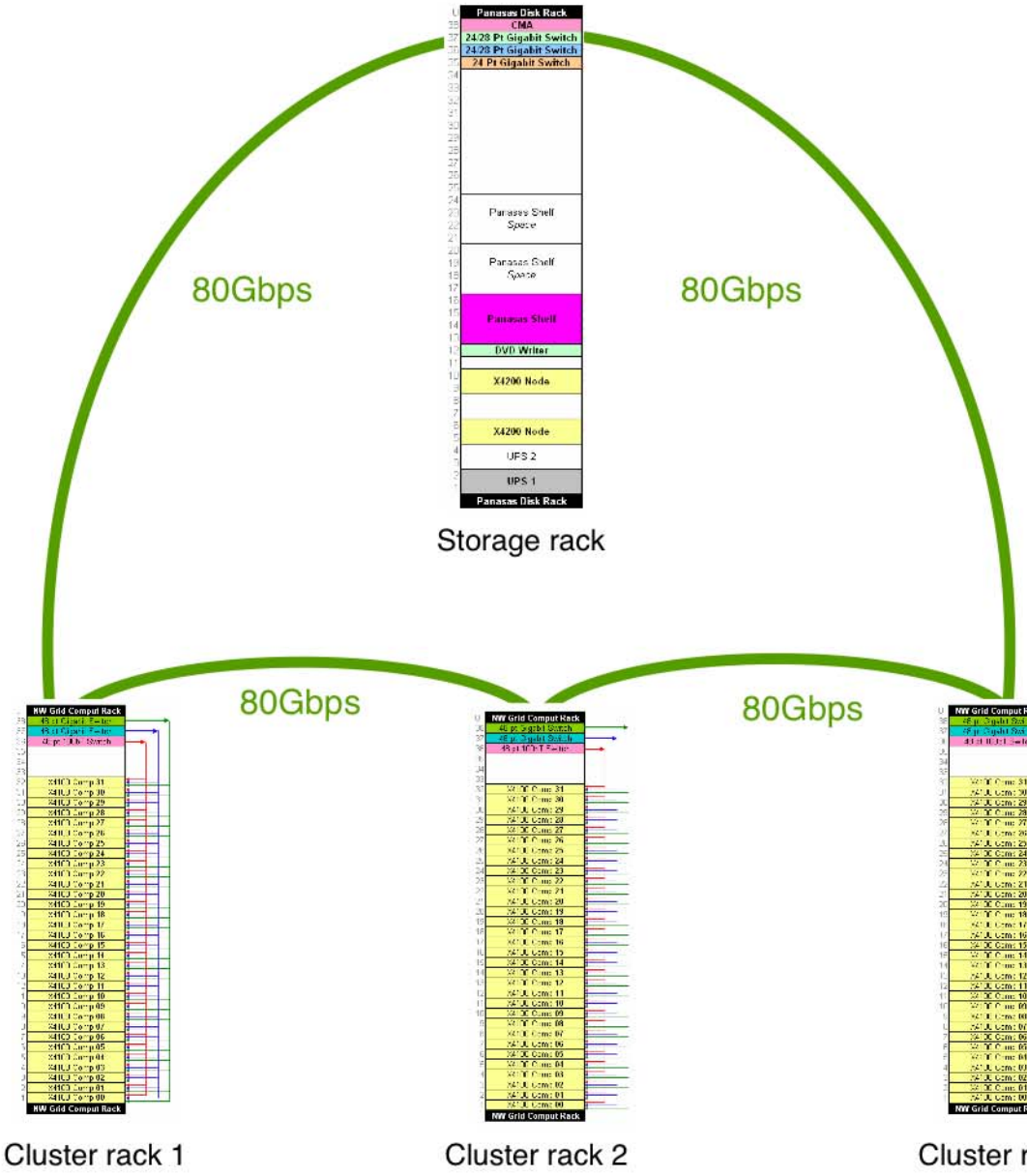
e-Science dti

- Gig-E connections from each front end to both switches
- Dual Gig-E connections from Panasas to each switch
- Dual network stacks for storage
- Two gigabit connections from each front end to provide four gigabits to external network.

Lancaster

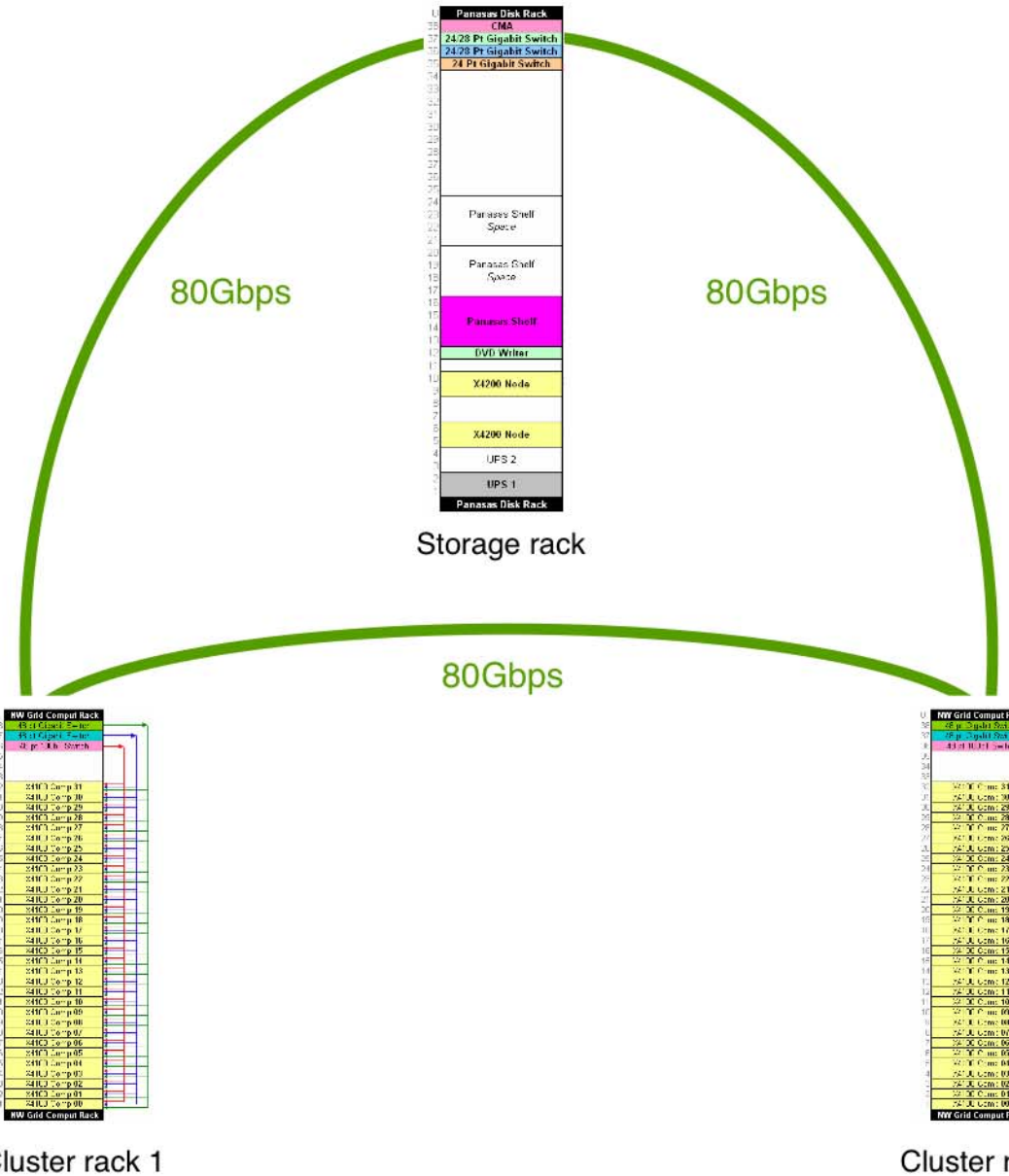
- six compute node racks
- dedicated MPI network in each rack (optional if serial racks)
- Optional snapshot system for data store (with external I/O to stream to existing tape).





Daresbury

- three compute node racks
- single or dual MPI networks per rack
- high-bandwidth storage network across all racks
- option of two switches in storage rack for resilience, one port from front end and two from Panasas to each switch
- quad Gig-E connection from Panasas to storage network
- Gig-E connection from front end systems to storage network
- Two available front end gigabit from each front end to give fout gigabit uplinks from cluster to outside network



Storage rack

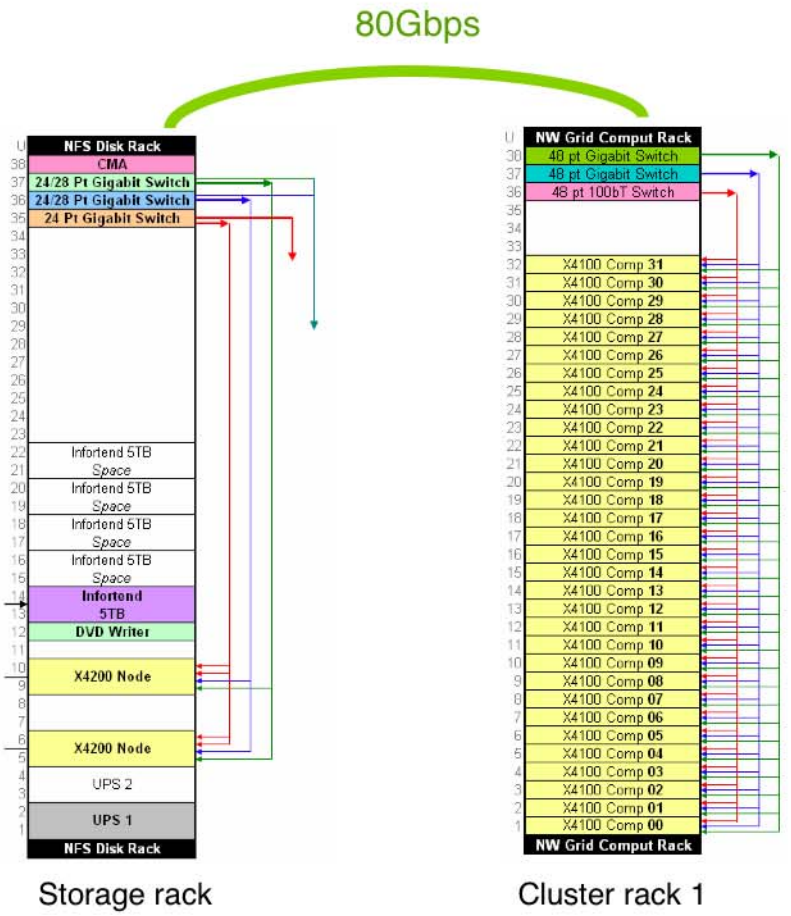
Cluster rack 1

Cluster rack 2

Liverpool

- two compute node racks One full , one part full.
- dedicated MPI network per rack
- high-bandwidth storage network across all racks
- quad Gig-E connection from Panasas to storage network
- Gig-E connection from front end systems to storage network
- two gigabit connections from each front end to provide four gigabits to external network.

Manchester



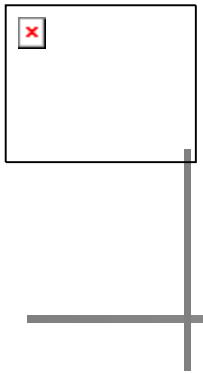
Storage rack

Cluster rack 1

- one part full compute node rack with dedicated MPI network
- high-bandwidth network to storage rack
- dual Gig-E connection from NFS server to storage network
- Gig-E connection from front end systems to storage network
- Two gigabit connections from each front end to provide four gigabits to external network.

Contacts with user base

- Regular weekly meetings with NW-GRID projects
- Monthly ESNW management meetings.
- Weekly meetings myGrid/OMII-UK
- Contacts with NIHBI, MCISB, Chemistry, Turing Institute, Jodrell Bank



Types of application - I

- Computational Chemistry (currently majority user) codes Gaussian, Amber.
- Protein sequencing, eFungi, MPIBlast, IntProScan
- Distributed query processing, OGSA-DQP eFungi. ISPDR

Types of application - II

- Text and data-mining, NaCTeM
- Processing of instrumental data, trials on NGS of Jodrell Bank Pulsar Data, interest from bio-sciences, spectrometer output, microarray data, medical image processing.

Projects 2003-2005

Projects	**2003-2005**		
	Staff	Capital (£)	Rec (£)
Manchester			
UniGrids	2		100K
GRIP			
RealityGrid	4		24K
GODIVA	1		60K
MyGRID	6		360K
ESNW	4		240K
RAHWL	1		15K
GridOneD	2		100K
CoMoS	2		120K
ISPIDER	2		120K
e-Fungi	5		100k
BEACON	2.5		150K
Local e-Science Support	1		60K
Industrial	6		360K
Sub-total	38.5	0	1729 K

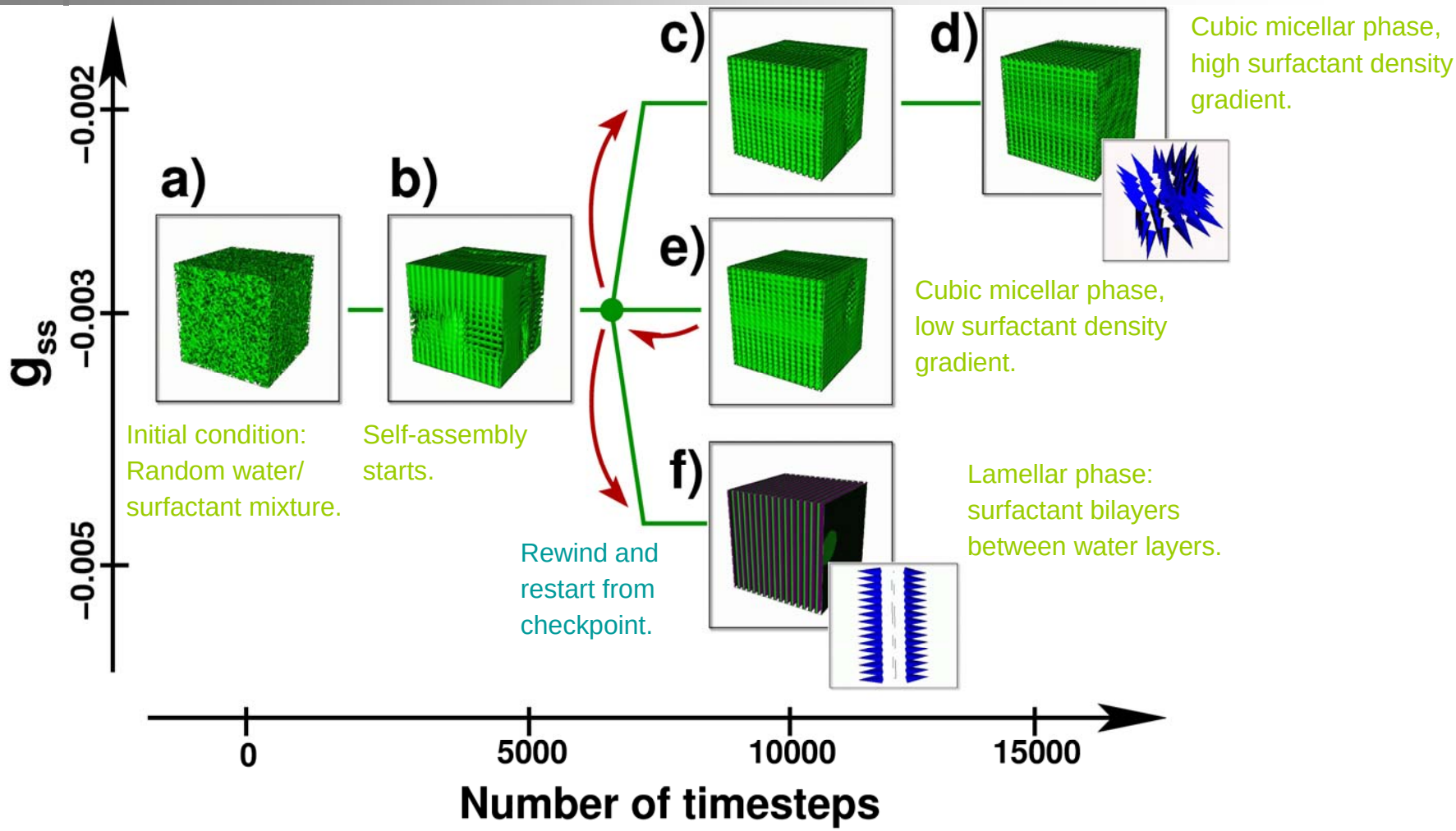
Multi-site projects and apps

- Beacon project, Mcr-Lpool, medical imaging, Mcr will have BIRN node.
- MIMAS datasets being Grid enabled, possible links to Lancaster work??
- Investigate steering of other sites codes.

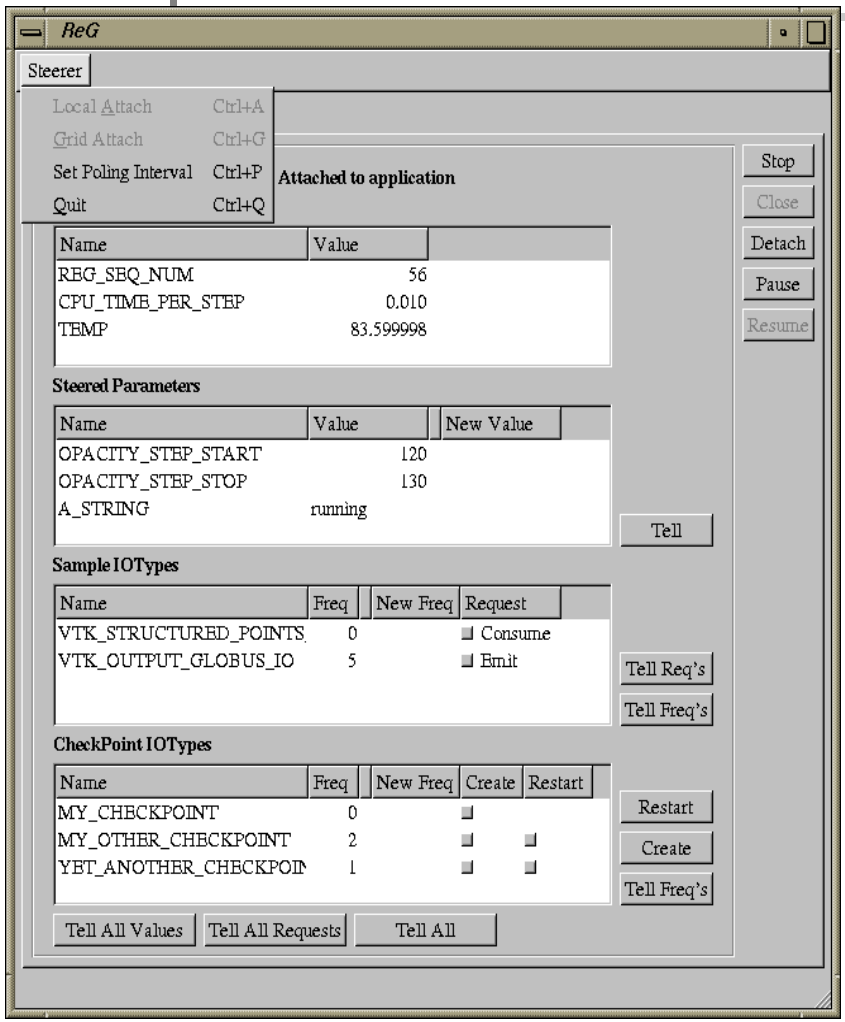
Application technology - Computational Steering

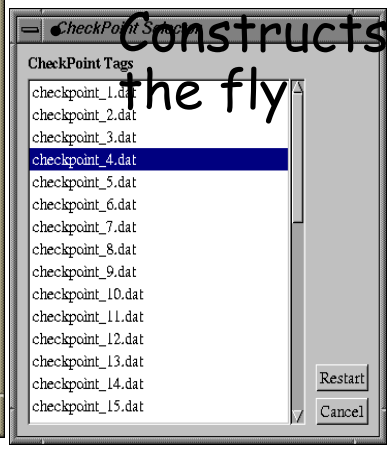
- Simulations can generate in days data that takes months to understand
- Problem: to efficiently explore and understand the parameter spaces of materials science simulations
- Computational steering aims to short circuit *post facto* analysis
  - Brute force parameter sweeps create a huge data-mining problem
  - Instead, we use computational steering to navigate to interesting regions of parameter space

Exploring parameter space through computational steering



Steering client



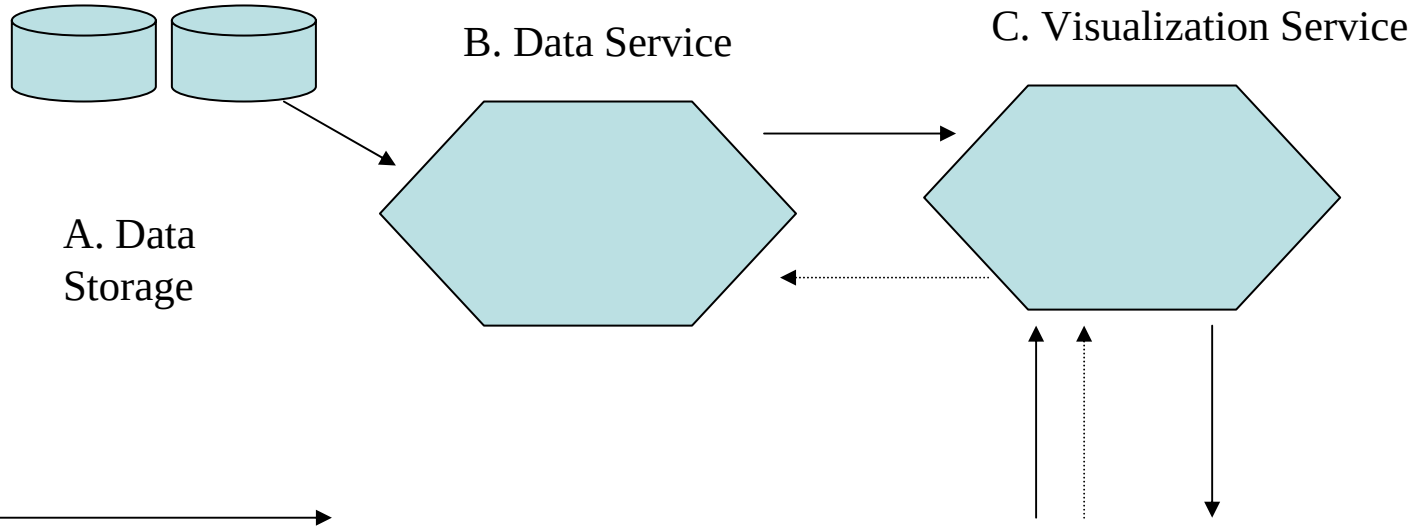


- Built using C++ and Qt library – currently have execs. for Linux and IRIX
- Attaches to any steerable RealityGrid application
- Discovers what commands are supported
- Discovers steerable & monitored parameters
- Constructs appropriate widgets on the fly

Application technology - exploration of very large datasets

- Models and experiments generate huge volumes of data.
- How can this be understood?
- In this application we focus particularly on the use of visualization to aid human intuition.

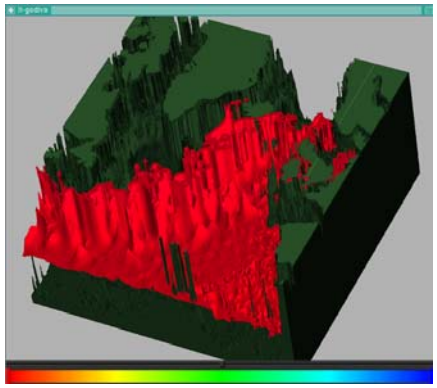
Locality-aware architecture



Sends data to

Requests data from

D. Client viewer, rendering either locally or by using visualization service



Visualization tools

- SGI PRISM at Mcr linked to NW-GRID, remote rendering service, possibility of remotely displaying visualization applications
- Projects investigating visualization on GPU cards, price/performance makes attractive but need customized application development.

User interface

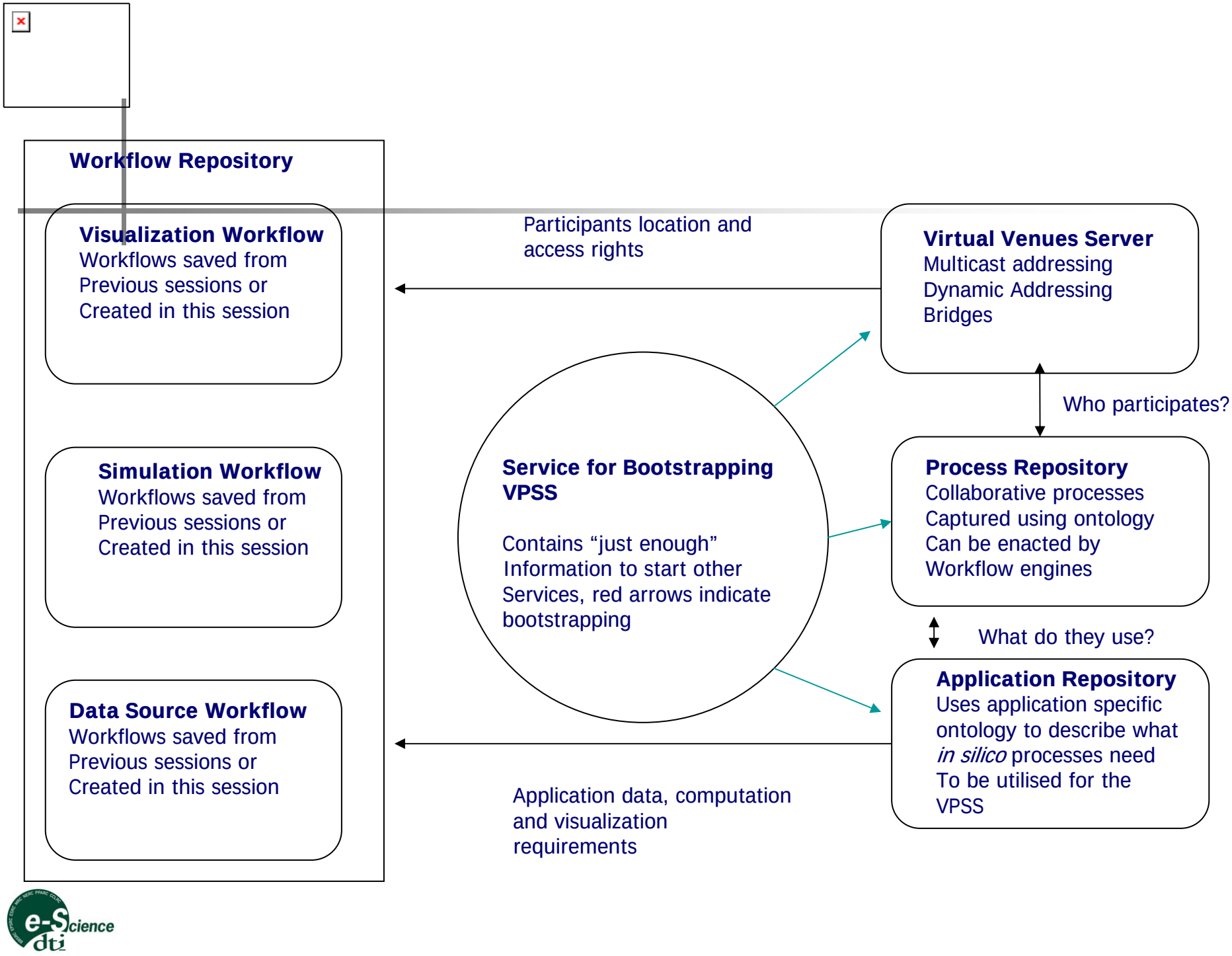
- Portals using e.g. Gridsphere
- Pervasive computing, lightweight clients.
- Industrial users want to use current applications as they look now, do not want to change interfaces.
- Secure transfer of data and analysis tools, e.g. workflows.

Collaborative working

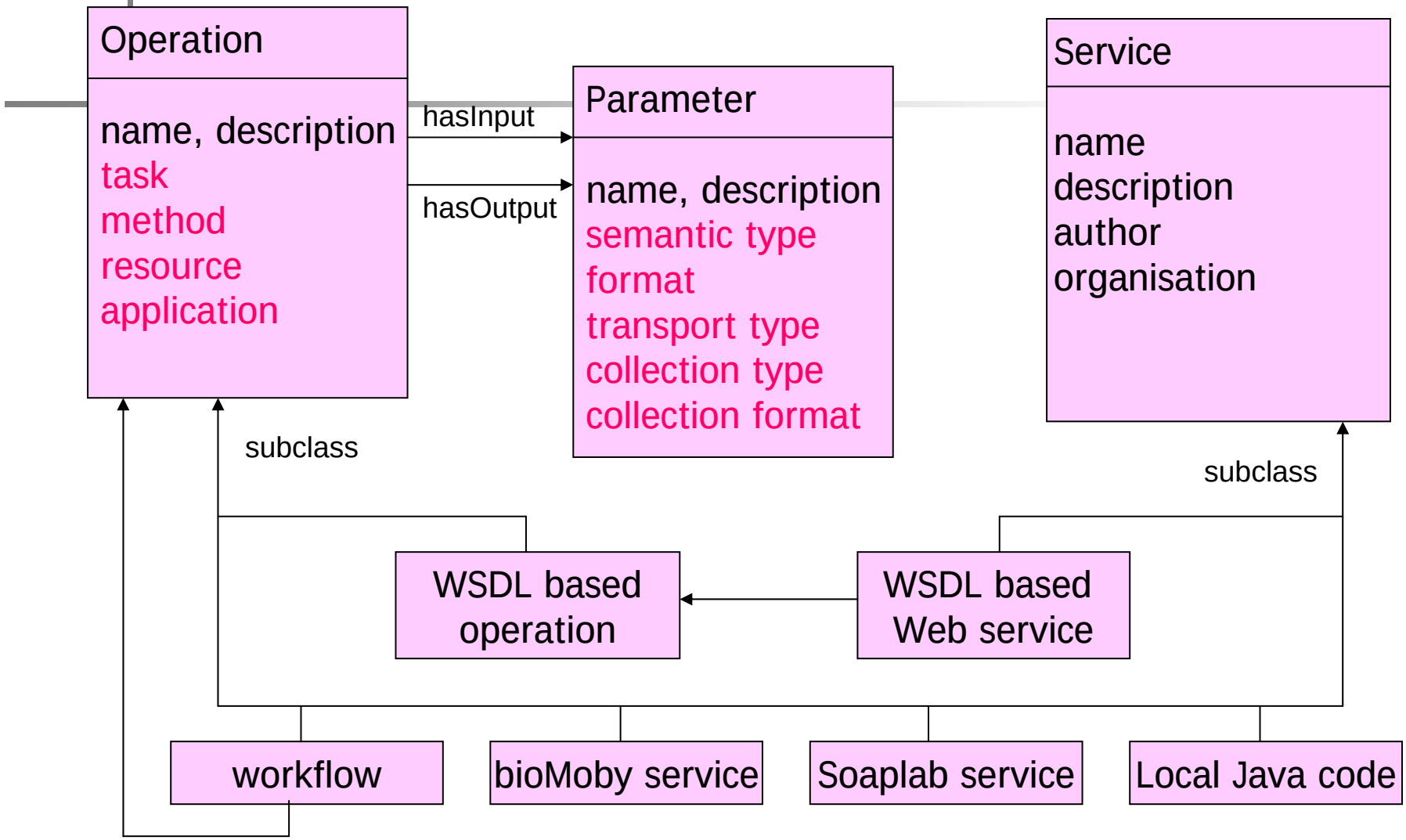
- Access Grid level only a start
- Want to integrate applications into meetings, either synchronous or asynchronous
- Collaborative, interactive visualization.
- Sharing data and results, e.g. trees generated in parameter space steering.

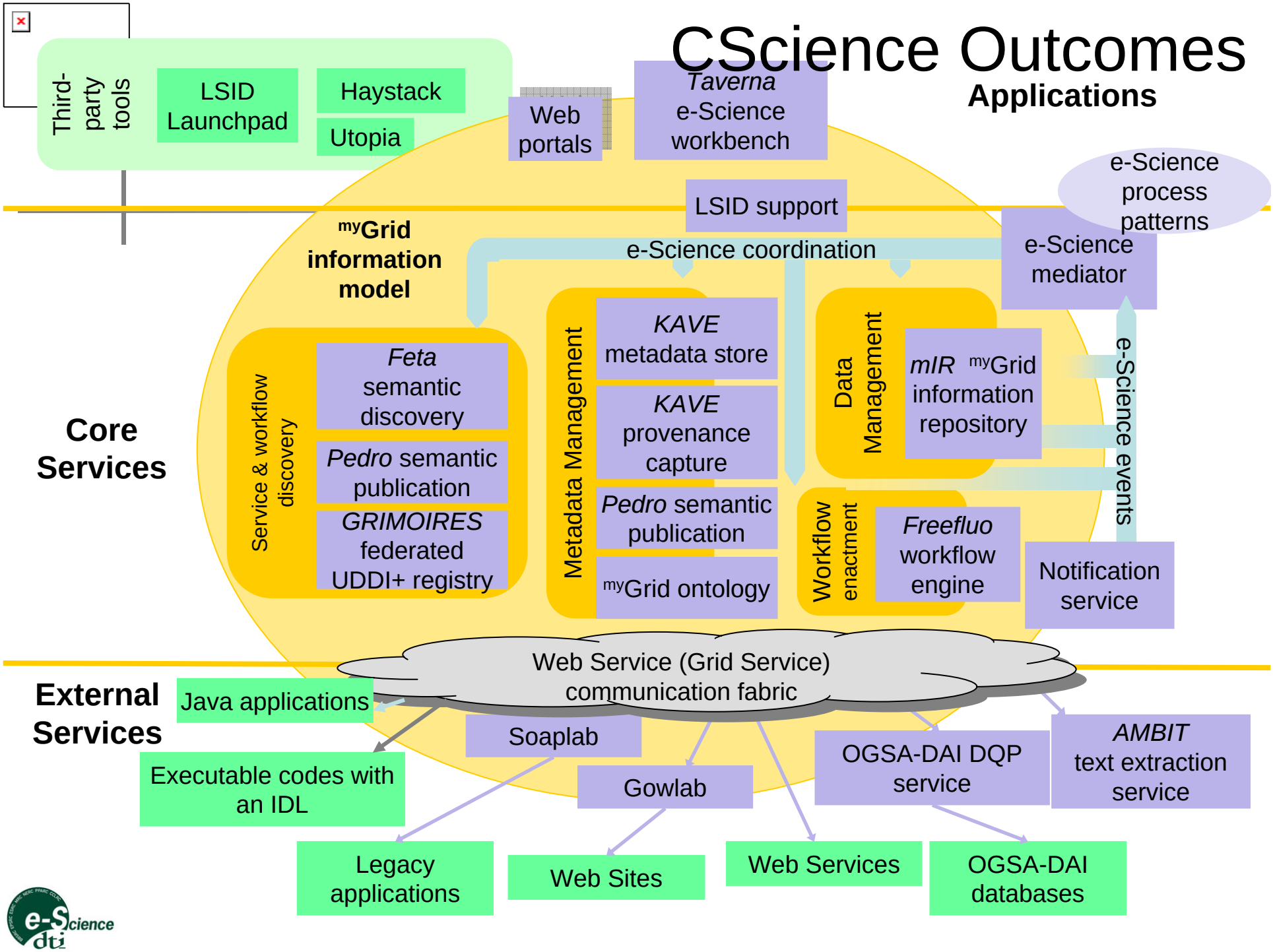
Projects 2003-2005

Projects	**2003-2005**		
	Staff	Capital (£)	Rec (£)
Manchester			
UniGrids	2		100K
GRIP			
RealityGrid	4		24K
GODIVA	1		60K
MyGRID	6		360K
ESNW	4		240K
RAHWL	1		15K
GridOneD	2		100K
CoMoS	2		120K
ISPIDER	2		120K
e-Fungi	5		100k
BEACON	2.5		150K
Local e-Science Support	1		60K
Industrial	6		360K
Sub-total	38.5	0	1729 K

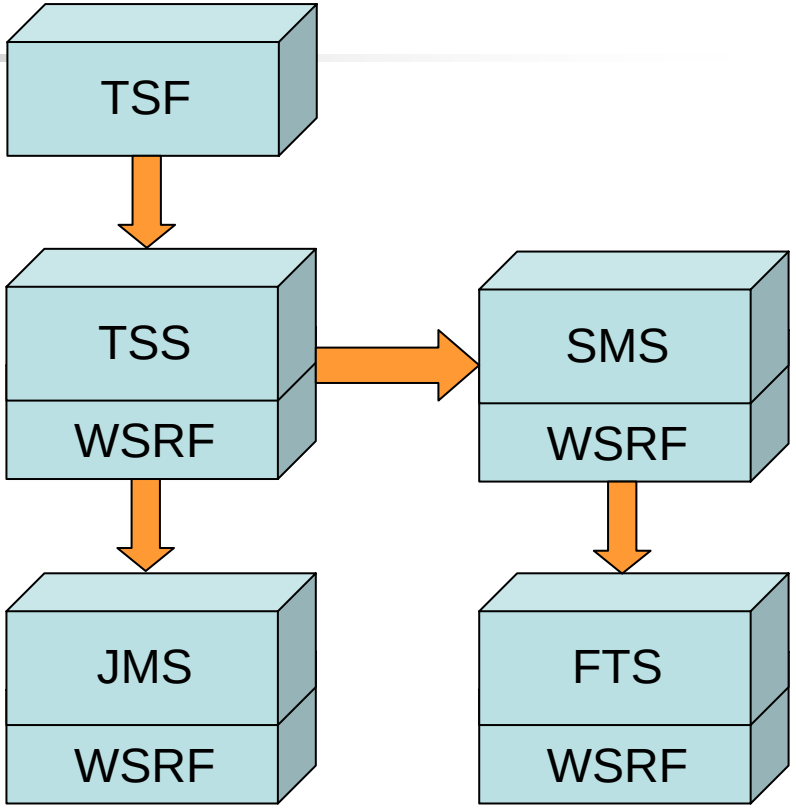

Workflow Repository
Visualization Workflow
Workflows saved from
Previous sessions or
Created in this session
Simulation Workflow
Workflows saved from
Previous sessions or
Created in this session
Data Source Workflow
Workflows saved from
Previous sessions or
Created in this session
Participants location and
access rights
Application data, computation
and visualization
requirements
Service for Bootstrapping
VPSS
Contains “just enough”
Information to start other
Services, red arrows indicate
bootstrapping
Virtual Venues Server
Multicast addressing
Dynamic Addressing
Bridges
Who participates?
Process Repository
Collaborative processes
Captured using ontology
Can be enacted by
Workflow engines
What do they use?
Application Repository
Uses application specific
ontology to describe what
in silico processes need
To be utilised for the
VPSS

Layering models




CScience Outcomes
Applications
Third-party tools
LSID Launchpad
Haystack
Utopia
Web portals
Taverna e-Science workbench
e-Science process patterns
LSID support
myGrid information model
e-Science coordination
e-Science mediator
Core Services
Service & workflow discovery
Feta semantic discovery
Pedro semantic publication
GRIMOIRES federated UDDI+ registry
Metadata Management
KAVE metadata store
KAVE provenance capture
Pedro semantic publication
myGrid ontology
Data Management
mIR myGrid information repository
e-Science events
Workflow enactment
Freefluo workflow engine
Notification service
Web Service (Grid Service) communication fabric
External Services
Java applications
Soaplab
Executable codes with an IDL
Gowlab
OGSA-DAI DQP service
AMBIT text extraction service
Legacy applications
Web Sites
Web Services
OGSA-DAI databases

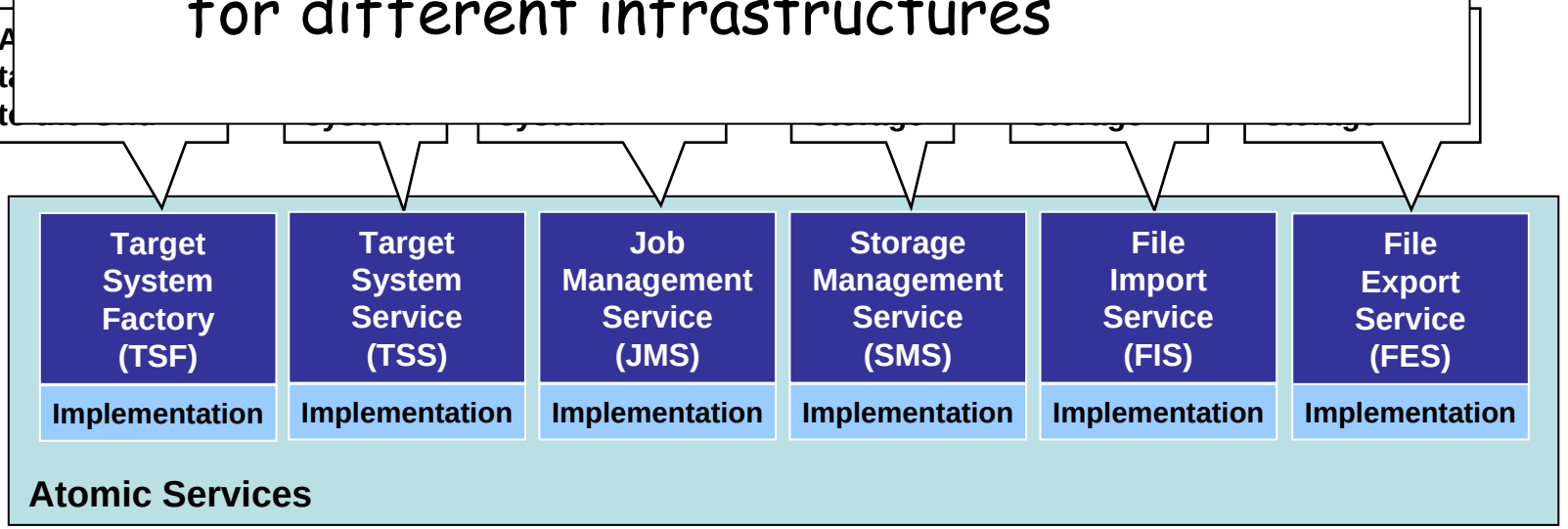
Atomic Services



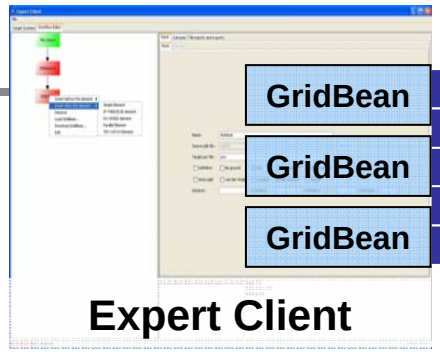
- Unicore basic functions
  - Site Management (TSF/TSS)
    - Compute Resource Factory
    - Submit, Resource Information
  - Job Management (JMS)
    - Start, Hold, Abort, Resume.
  - Storage Management (SMS)
    - List directory, Copy,
      Make directory,
      Rename, Remove.
  - File Transfer (FTS)
    - File import, file export
- Standardization
  - JSDL WG UniGrids and NAREGI collaborated with OMII (GridSAM), RealityGrid
  - Atomic Services are input to the OGSA-BES WG

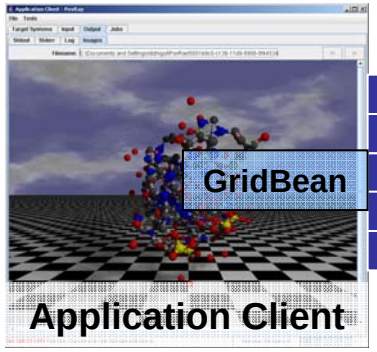
Atomic Services Overview

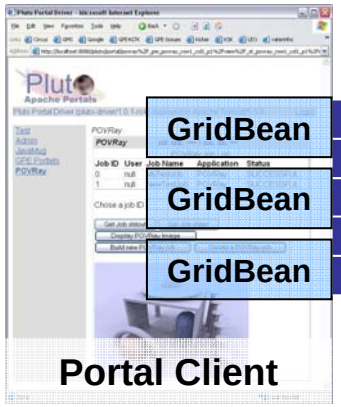
- Atomic service interfaces define basic set of operations and properties that have to be available on a Grid
- Different implementations of interfaces for different infrastructures

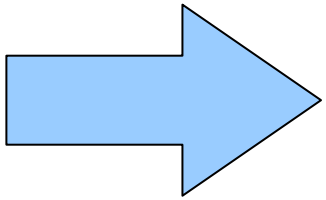


GPE as interoperability framework









Atomic Service Client API

Atomic Services

UNICORE/GS

Globus Toolkit 4

China Grid Support Package

Other OGSA-compliant Grid servers

Web info

http://www.nw-grid.ac.uk