Science on NWGrid using RMCS/eMinerals Infrastructure

Jamie M. Rintelman, Jens Thomas, Rik Tyer, Rob J. Allan, Paul Sherwood
STFC Daresbury Laboratory

Martin T. Dove, Kat F. Austen, Andrew M. Walker, Richard P. Bruin
Department of Earth Sciences, Univesity of Cambridge

Leon Petit
Department of Physics and Astronomy, Univeristy of Aarhus

Marcus C. Durant
School of Applied Sciences, Northumbria University

Outline of Talk

The Hardware: NWGrid

The Middleware: eMinerals infrastructure / RMCS

The Science:

- Experienced Users:
  - Amorphous Silica: Andrew Walker
  - Dichloro Biphenyl, PCDDs Kat Austen
- New Grid Users:
  - Perovskite: Leon Petit
  - QDGA: Jens Thomas

Working together

eMinerals

Science & Technology Facilities Council

CSED, eScience

Univeristy of Aarhus

Northumbria University

NW-GRID

Northwest REGIONAL DEVELOPMENT AGENCY

Science & Technology Facilities Council

LANCASTER UNIVERSITY

THE UNIVERSITY of LIVERPOOL

MANCHESTER 1824

NW-GRID Vision

Technology "tuned to the needs of practicing scientists".

Hooks to other Grid consortia: NGS, WRG

Top end: HPCx and CSAR

Applications and industry

Pharma, meds, bio, social, env, CCPs

Portals, client toolkits, active overlays

Mid range: NW-GRID and local clusters

User interfaces

Sensor networks and experimental facilities

Advanced Network technology

Desktop pools: Condor etc.

Project Aims and Partners

Aims:

- Establish, for the region, a world-class activity in the deployment and exploitation of Grid middleware
- realise the capabilities of the Grid in leading edge academic, industrial and business computing applications
- Leverage 100 posts plus £15M of additional investment

Project Partners:

- Daresbury Laboratory: CSED and e-Science Centre
- Lancaster University: Management School, Physics, e-science and computer science
- University of Liverpool: Physics and Computer Services
- University of Manchester: Computing, Computer Science, Chemistry, bio-informatics + systems biology
- Proudman Oceanographic Laboratory, Liverpool

Hardware – 2006 procurement

From Sun / Streamline computing

Dual core, dual processor AMD Opteron nodes (with at least 8 GB of memory / node)

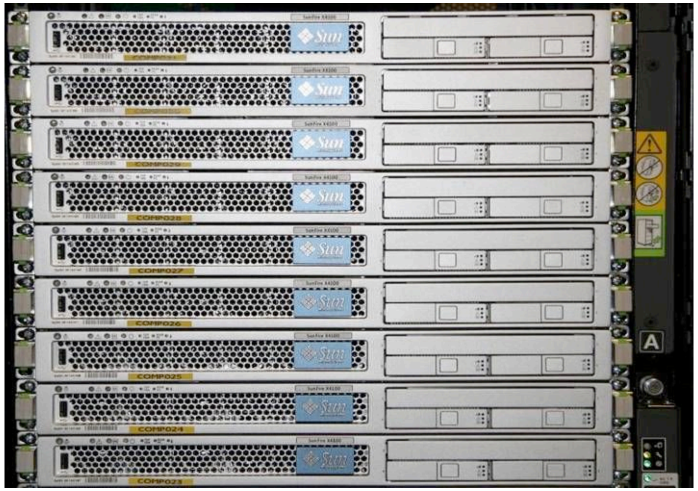
- 96 nodes – Daresbury
- 48 nodes – Lancaster
- 44 nodes – Liverpool
- 25 nodes – Manchester

8 TB Panasas file servers at Daresbury, Lancaster and Liverpool

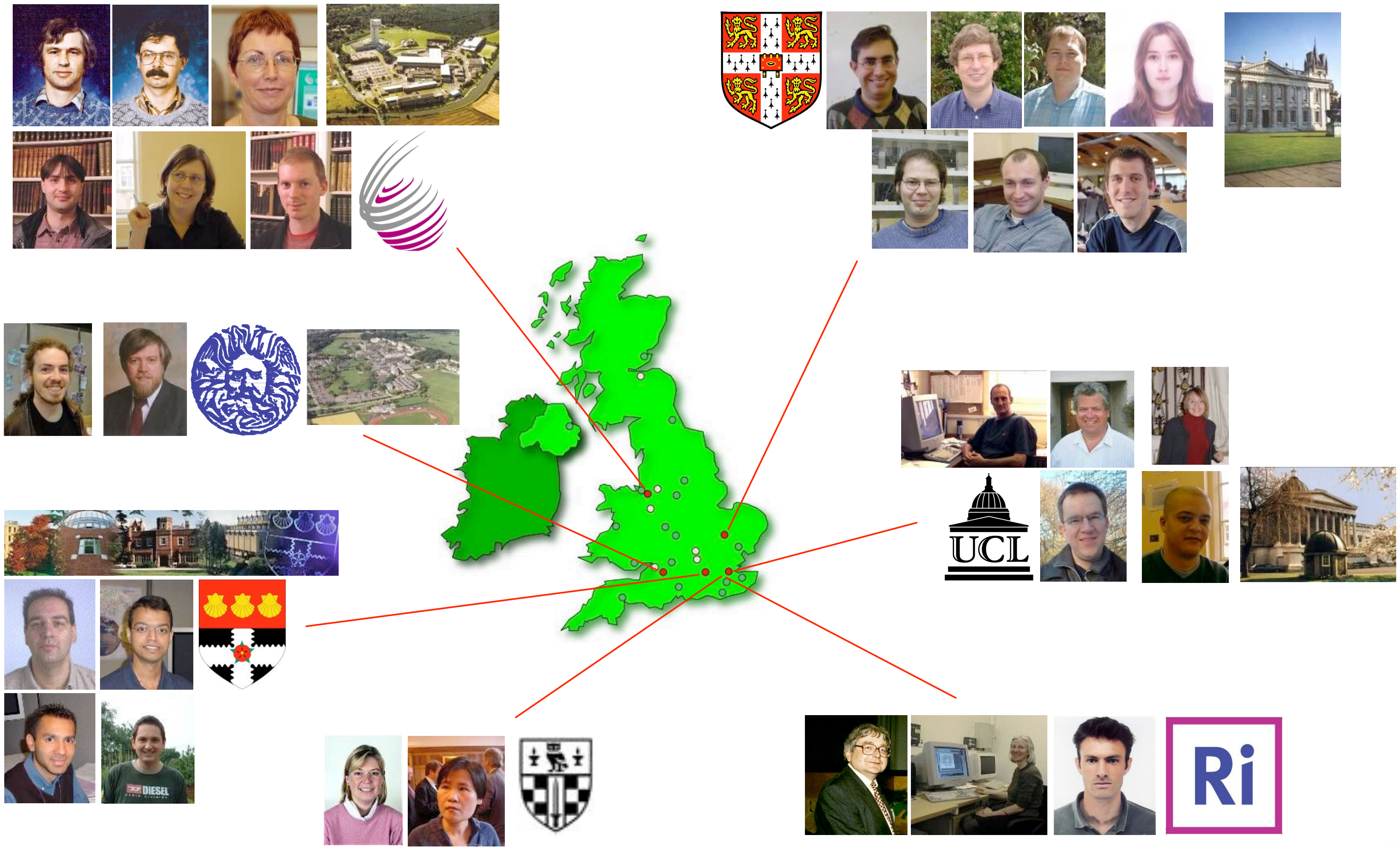
2.8 TB RAID array at Manchester

Separate data and communications GigE interconnect.

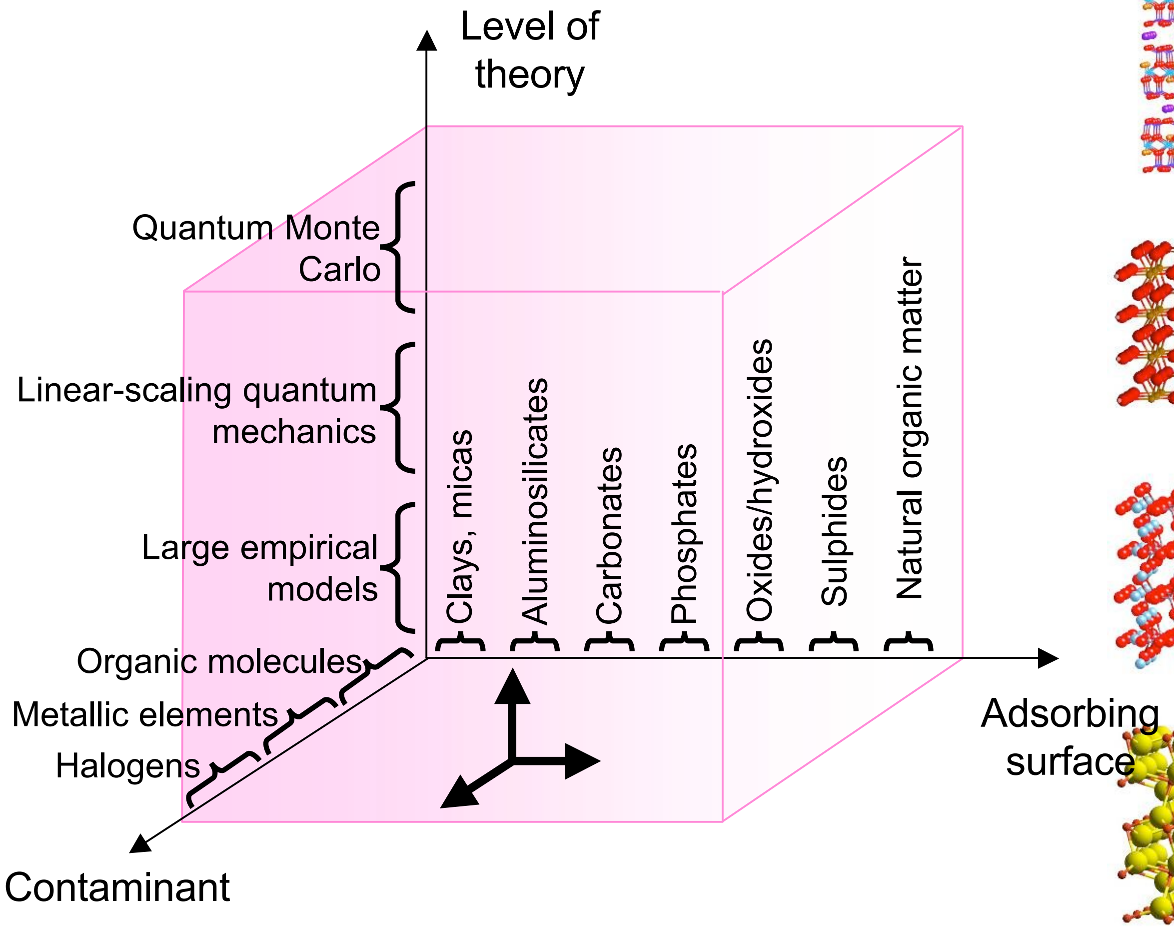
2006 Hardware

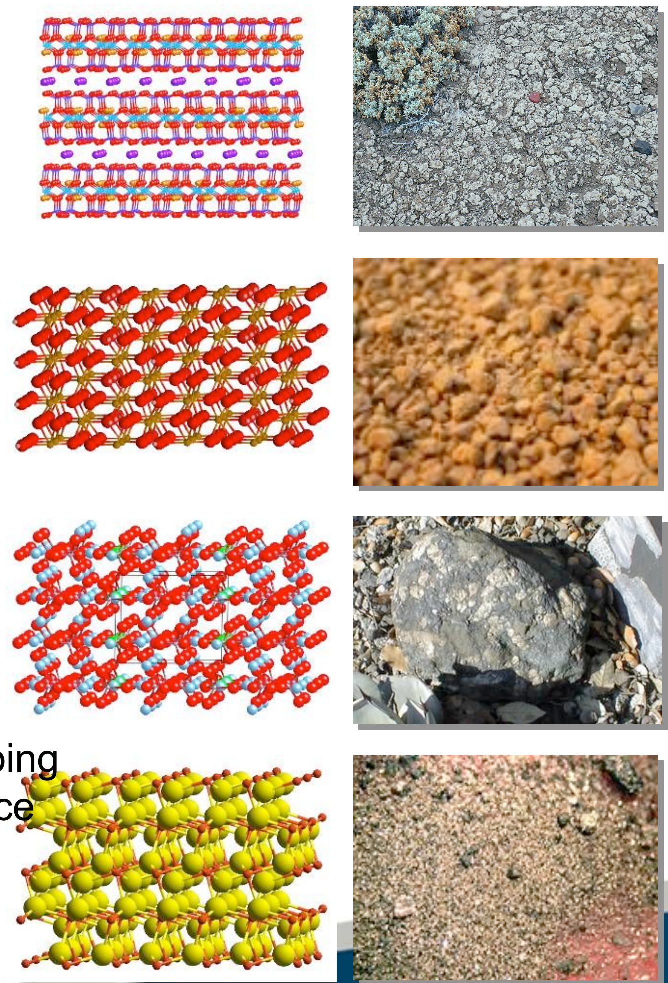


The *e*Minerals team



"Grand Challenge" science and the *e*Minerals VO





Introduction: eMinerals/RMCS framework

- RMCS is the client environment used by eMinerals for desktop job submission and job management
- Require a client environment that provides integrated compute, data and metadata functionality
- Want to be able to interact with the grid from the desktop, i.e.:
  - Tools need to be ‘firewall friendly’
  - Tools need to be ‘lightweight’ and ‘self-contained’
  - Tools should be either web based or integrated into native package management systems

Parameter Sweeps

- Large payoff for initial effort becoming “grid enabled”
- Periods when require large numbers of processors, interspersed with quiet periods for analysis
- Need single access point to all compute resources
- Need single access point to data
- Really need to use metadata as primary interface to data
- Running large numbers of simulations requires integrated compute, data and metadata functionality

RMCS / eMinerals Framework

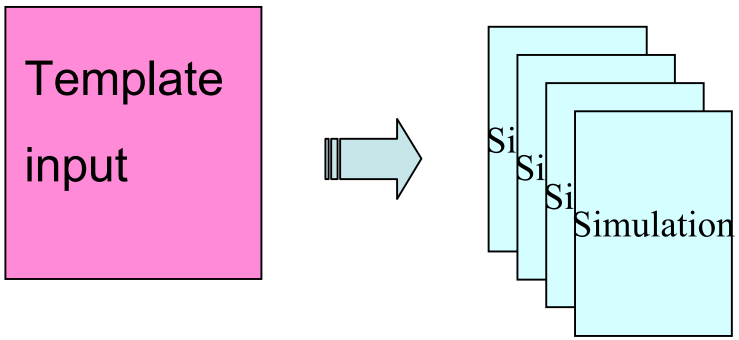
Tools developed organically to facilitate the science

1. Input preparation - Bespoke Scripts
2. Bulk Job Submission - Monty
3. Integrated Compute/Data/Metadata Framework - RMCS
4. Analysis of Results - RGem

Additional Building Blocks:

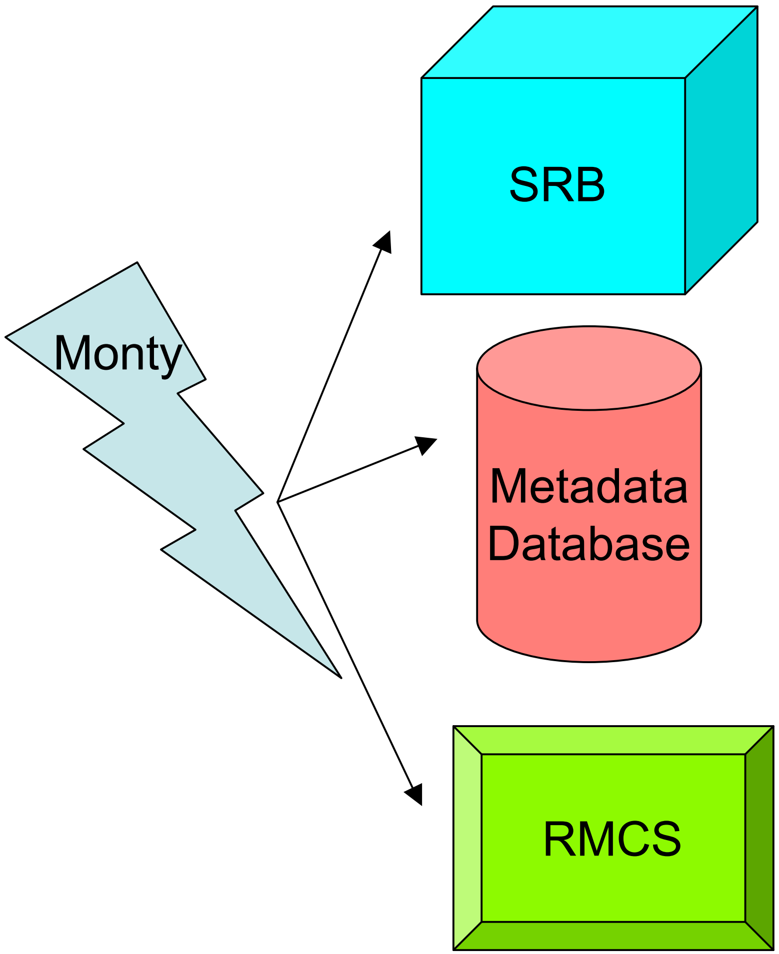
- Storage Resource Broker (SRB) - Data Storage and Collaborative Sharing
- AgentX - XML data, sharing between programs, metadata capture (developed by Phil Couch)

Input preparation



Bespoke scripts to automate generation of input files for parameter sweep type calculations

Monty - Bulk Job Submission



-set up structure in SRB for staging of input files and binary; storage of output files

-set up structure in database for metadata capture

-submit jobs to RMCS

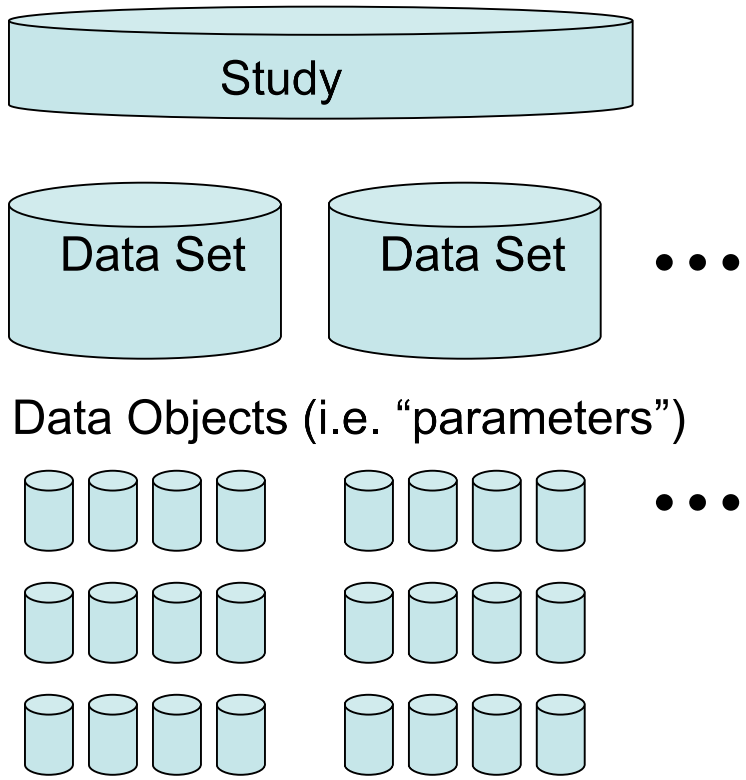
Metadata Database

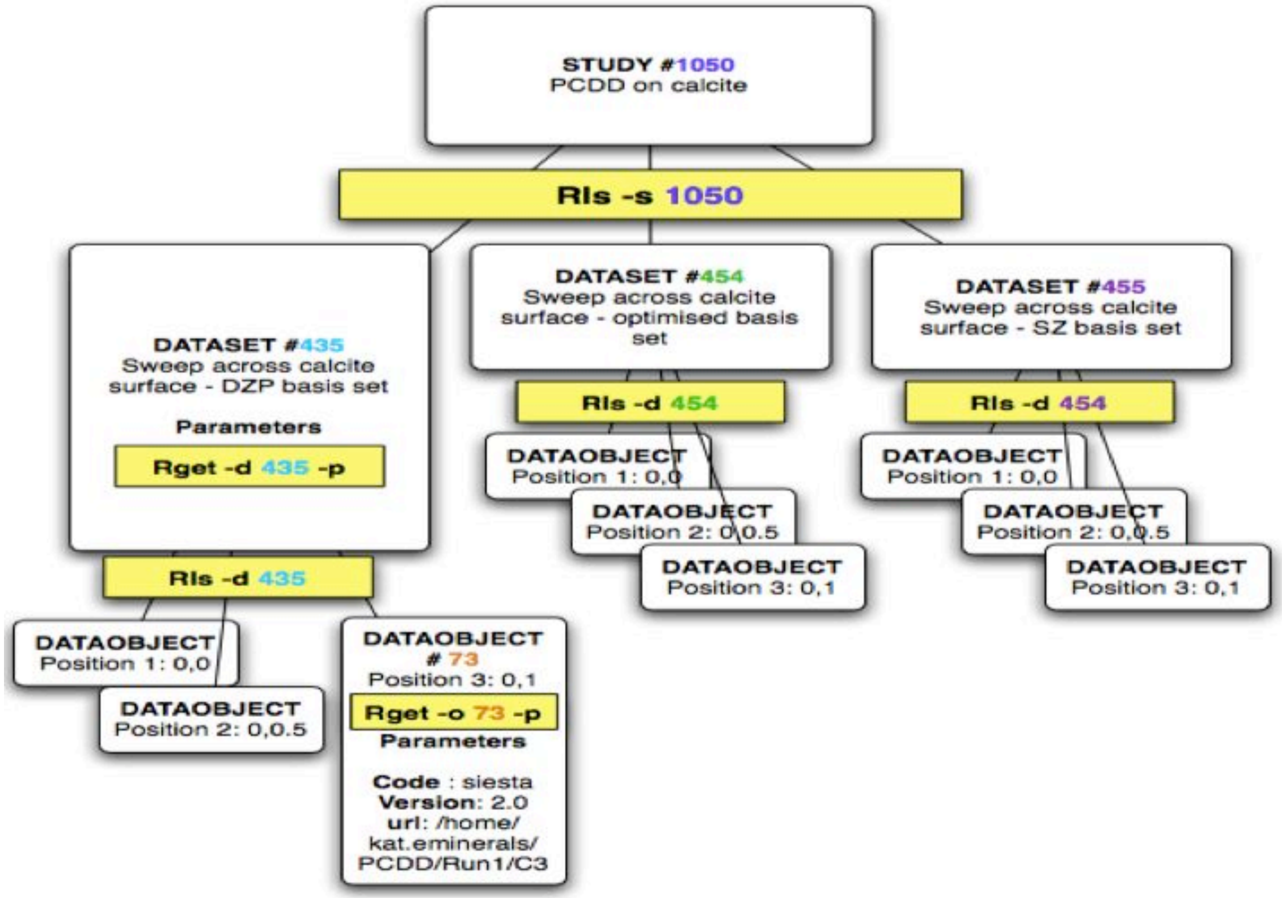
Metadata in database are divided into study, data set, and data object

Study = entire proj

Data Set = group of calculations

Data Objects = piece of data from each calculation

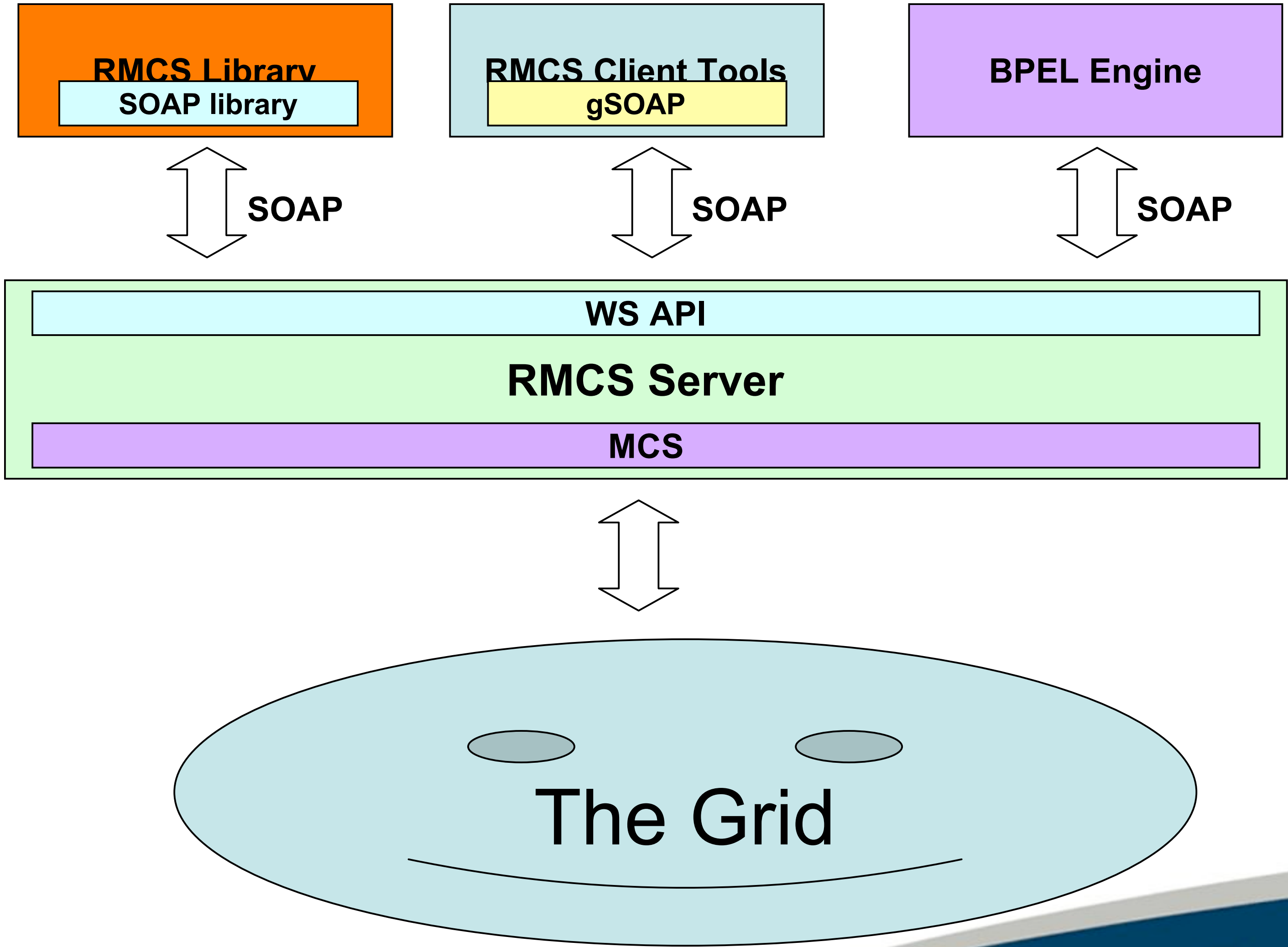



STUDY #1050
PCDD on calcite
Rls -s 1050
DATASET #435
Sweep across calcite surface - DZP basis set
Parameters
Rget -d 435 -p
Rls -d 435
DATAOBJECT
Position 1: 0,0
DATAOBJECT
Position 2: 0,0.5
DATAOBJECT
73
Position 3: 0,1
Rget -o 73 -p
Parameters
Code : siesta
Version: 2.0
url: /home/kat.eminerals/PCDD/Run1/C3
DATASET #454
Sweep across calcite surface - optimised basis set
Rls -d 454
DATAOBJECT
Position 1: 0,0
DATAOBJECT
Position 2: 0,0.5
DATAOBJECT
Position 3: 0,1
DATASET #455
Sweep across calcite surface - SZ basis set
Rls -d 454
DATAOBJECT
Position 1: 0,0
DATAOBJECT
Position 2: 0,0.5
DATAOBJECT
Position 3: 0,1

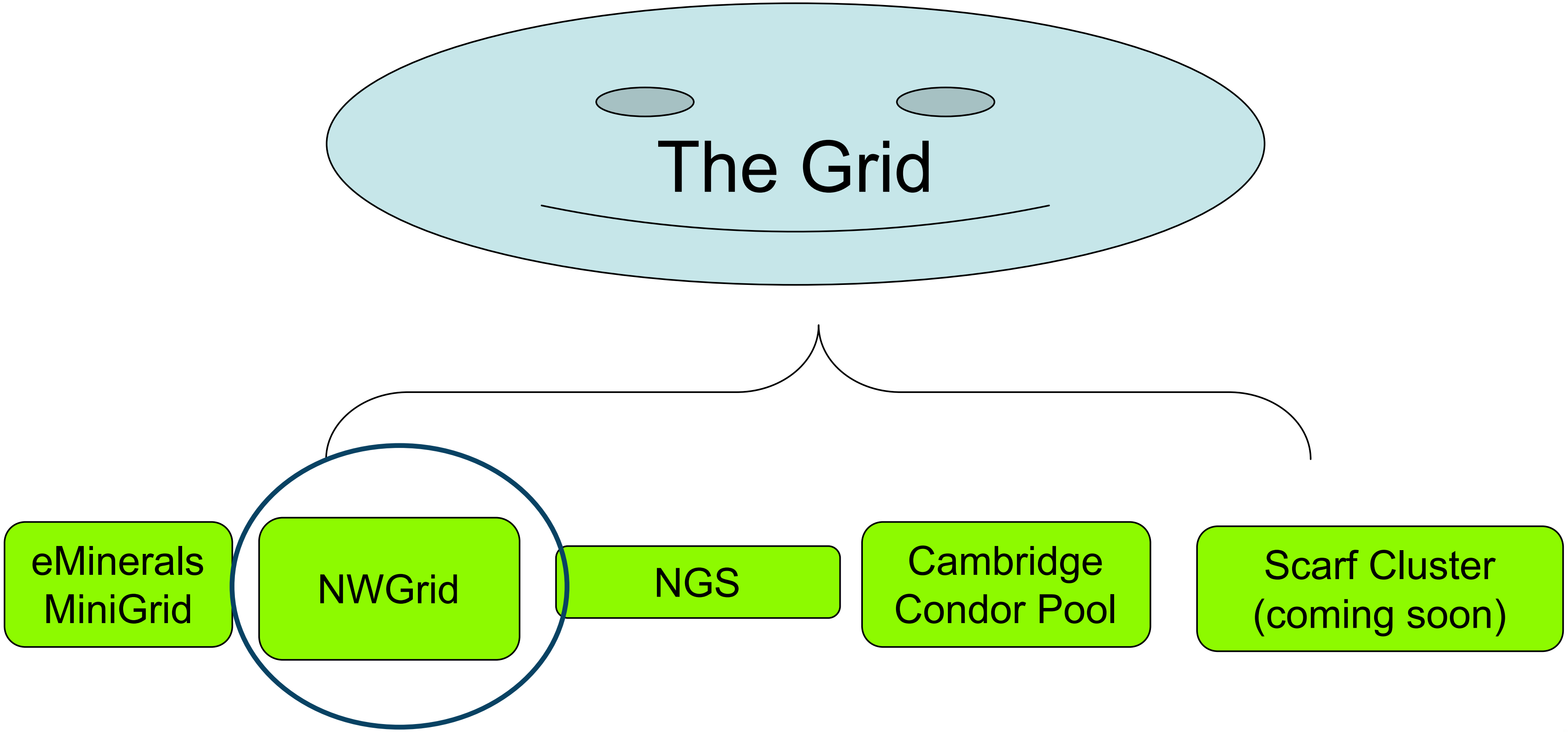
RMCS - integrated compute/data/metadata framework

1. Meta-schedule
2. Stage input files and binary
3. Run job/submit to batch queue
4. Transfer output to SRB
5. Use Rcommands + AgentX to put metadata into database /extract XML data if available

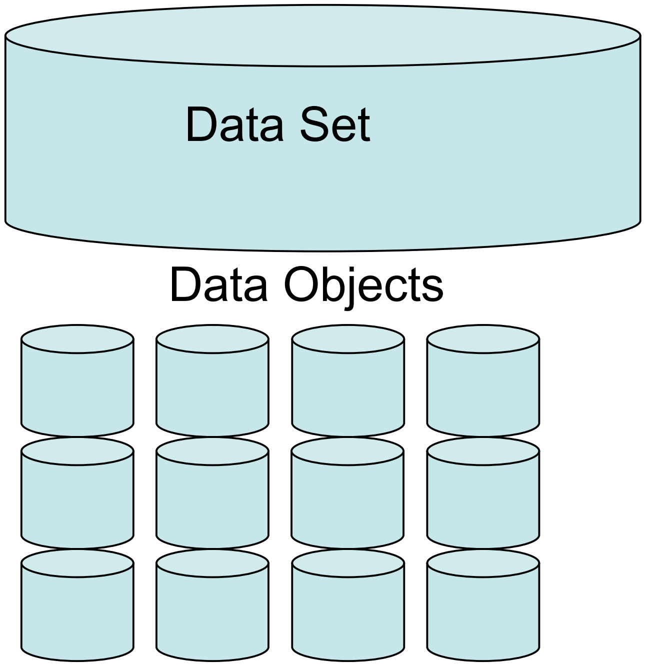
RMCS Architecture



“The Grid”: Real Experience



Rgem - Collect Results



-Analyze Results

-Collect parameters from a chosen dataset -> tab separated file -> graph

The compressibility anomaly in amorphous silica
Andrew M. Walker, Martin T. Dove
Experienced Grid Users

Andrew M. Walker, Martin T. Dove

awal05@esc.cam.ac.uk

Irreversible density change with pressure above ~25GPa

Negative thermal expansion below room temperature

Si increases co-ordination 10-25GPa (amorphous-amorphous phase transition)

Compressibility increases with pressure to ~2GPa

Low temperature anelasticity

Irreversible density change with pressure above ~25GPa

Negative thermal expansion below room temperature

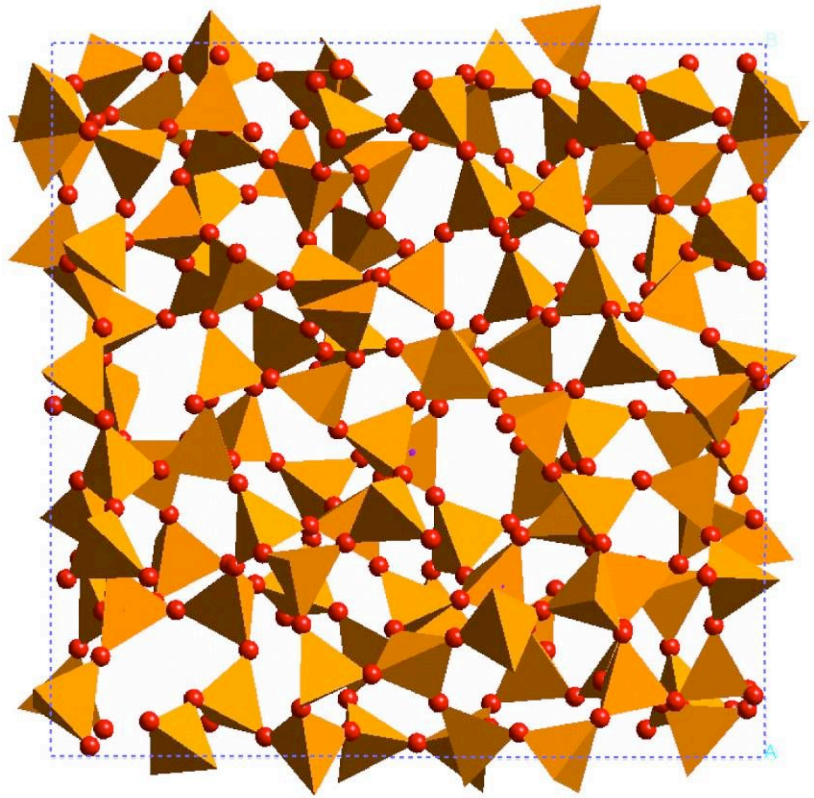
Si increases co-ordination 10-25GPa (amorphous-amorphous phase transition)

Compressibility increases with pressure to ~2GPa

Low temperature anelasticity

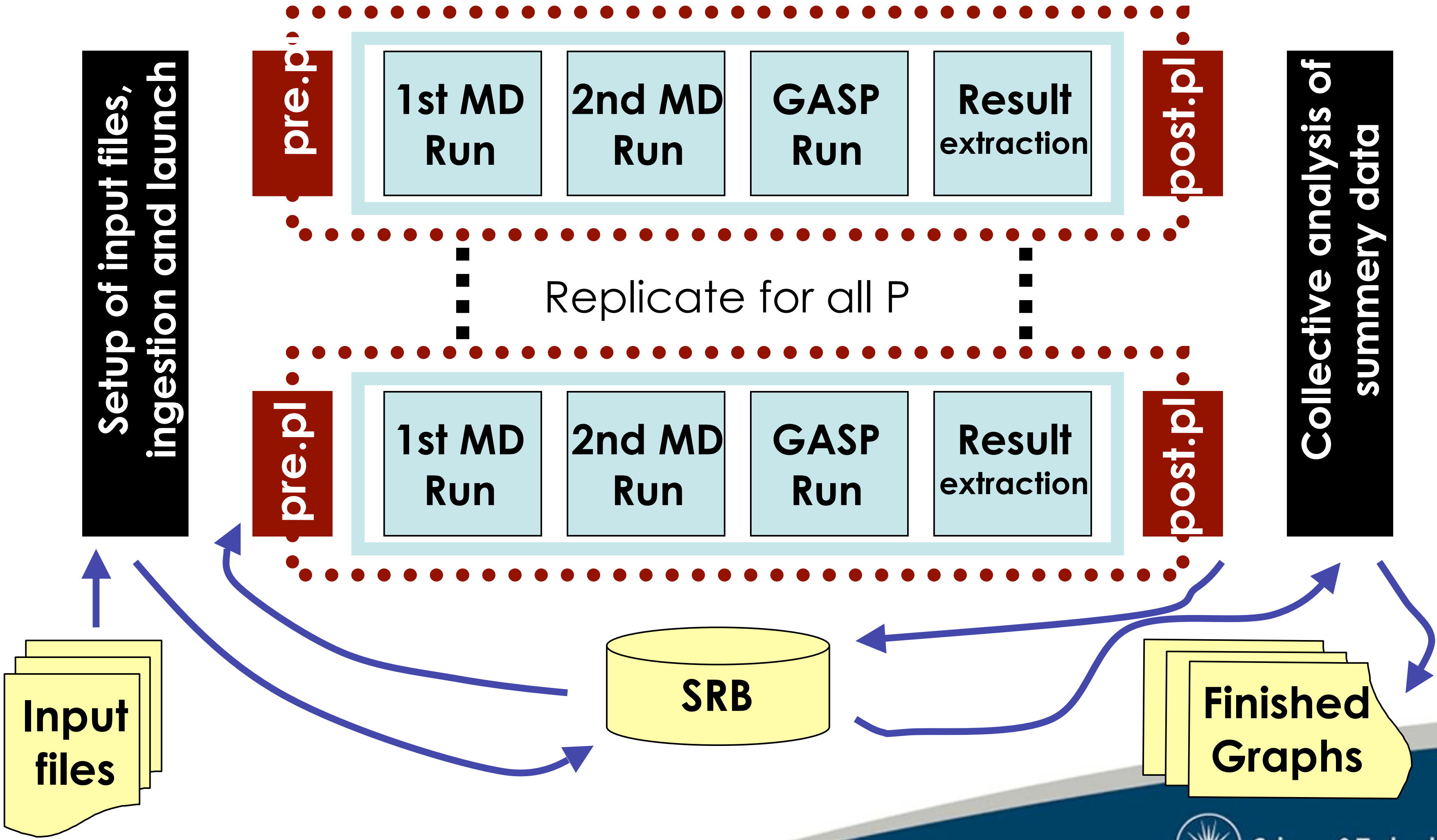
Approach

- Perform molecular dynamics calculations over a range of pressures
- Examine structure and dynamics for evidence of network softening
- Use three networks with perfect connectivity (no defects) and two potential models

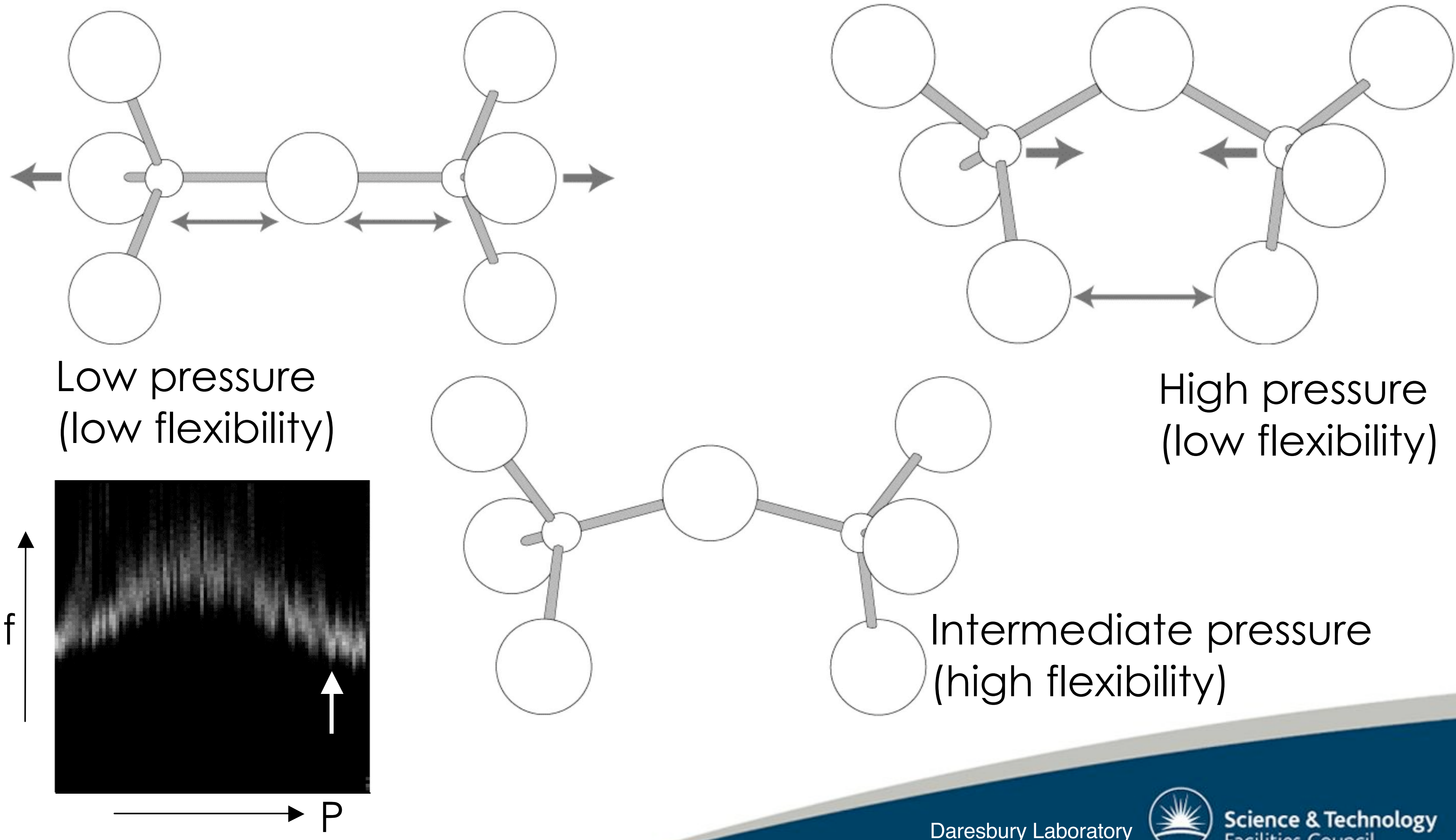


Wooten *et al.* (1985) *PRL* **54**, pp. 1392-1395

Grid computing



Network flexibility

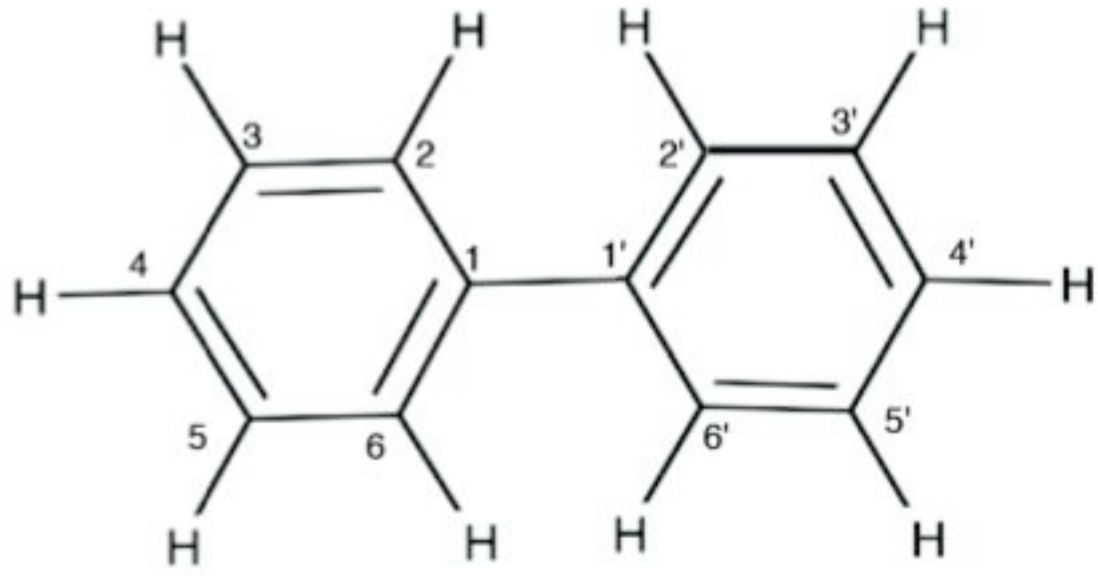


Conclusions

- The compressibility anomaly in silica glass appears to be caused by a changes in the flexibility of the tetrahedral framework
- A phase transition is not required
- The mechanism does not require an amorphous state
- Currently investigating the anomalous thermal expansion

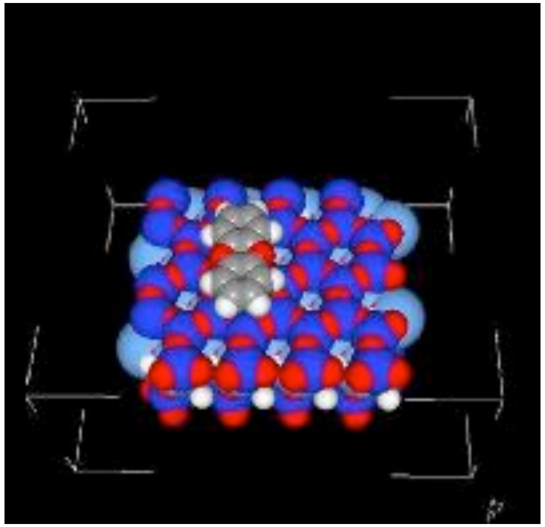
Walker *et al.* (2007) *J. Phys: Con. Mat.* **19** art. no. 275210
Dove *et. al.* (2006) *Mol. Simul.* **32** pp. 949-952

Investigation of the Torsional Behavior of 2,2'-Dichloro Biphenyl - Kat Austen
GAMESS-UK, SIESTA



Adsorption of Polychlorinated Dibenzo-p-Dioxins onto Mineral Surfaces - Kat Austen

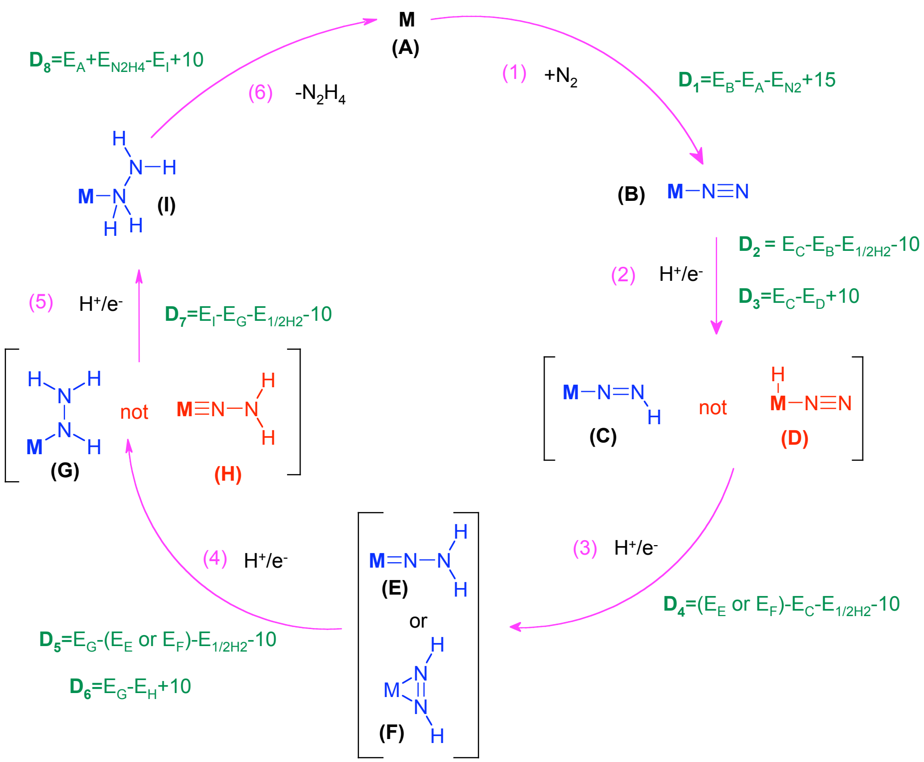
SIESTA 32-64 proc/job



Quantum Directed Genetic Algorithm

- Marcus Durrant, Paul Sherwood and Jens Thomas

New Grid Users



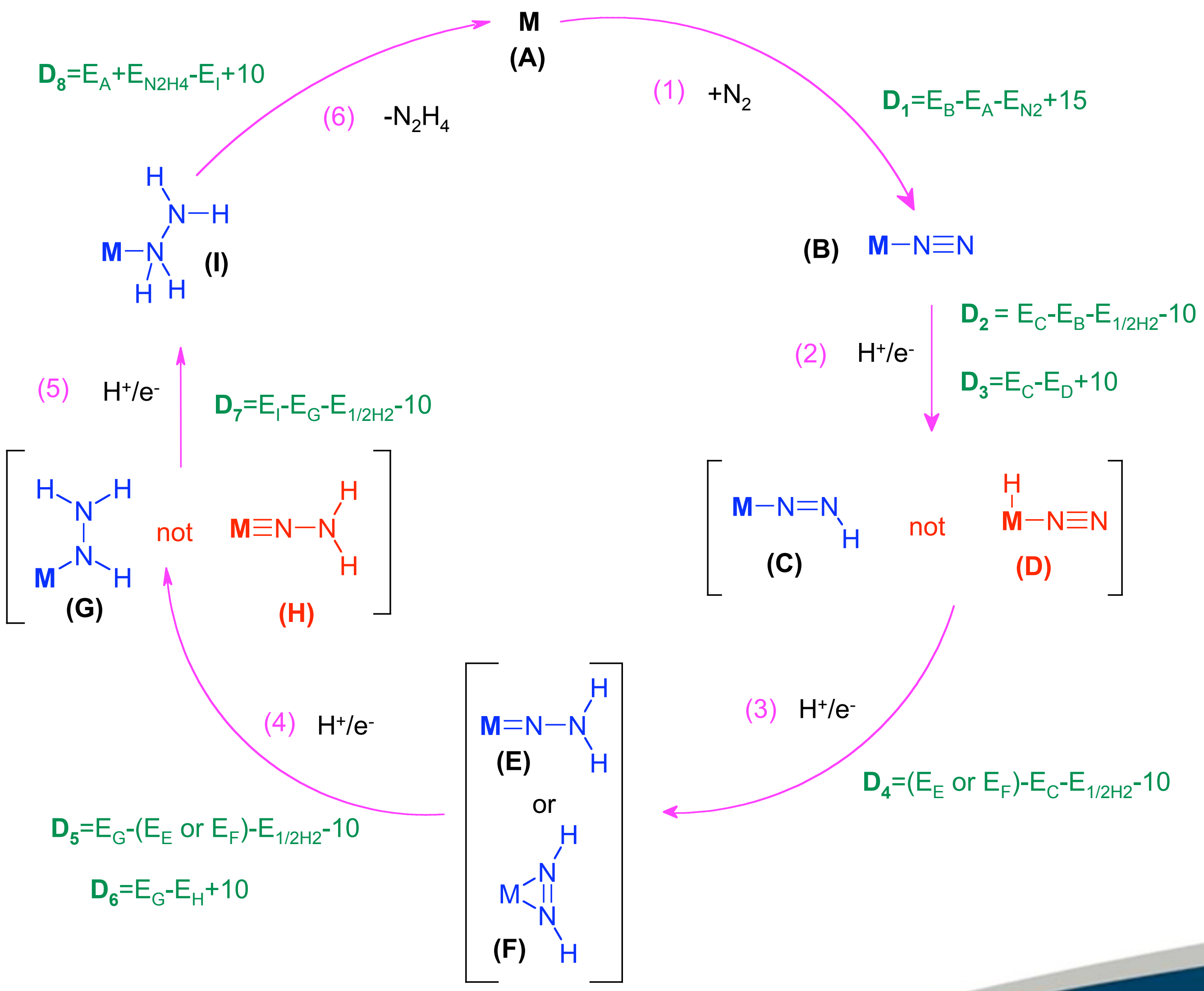
AIM: Find the most effective transition metal complex for catalyzing the reduction of N_2 to N_2H_4

Quantum Directed Genetic Algorithm

- Marcus Durrant, Paul Sherwood and Jens Thomas

- The project aims to find the most effective transition metal complex for catalysing the reduction of N_2 to N_2H_4.
- Reaction energies for each step in the cycle are calculated and the most successful complexes go through to the next 'round'.
- The complexes are 'bred' and 'mutated' to create new complexes that will hopefully combine the most successful attributes of their parents.
- The process is repeated until a complex with a desired efficacy is found.
- Reaction energies are calculated using GAMESS-UK running numerous jobs concurrently and in parallel

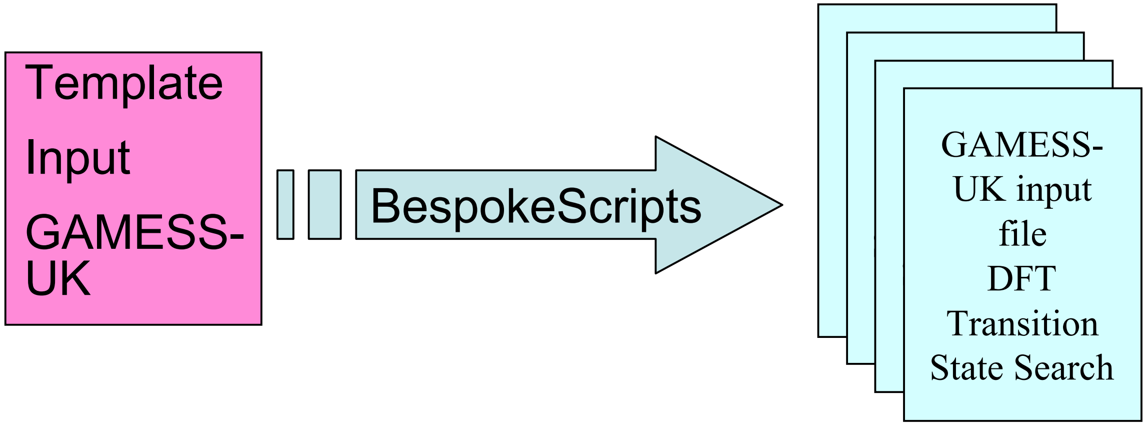
Catalytic Cycle

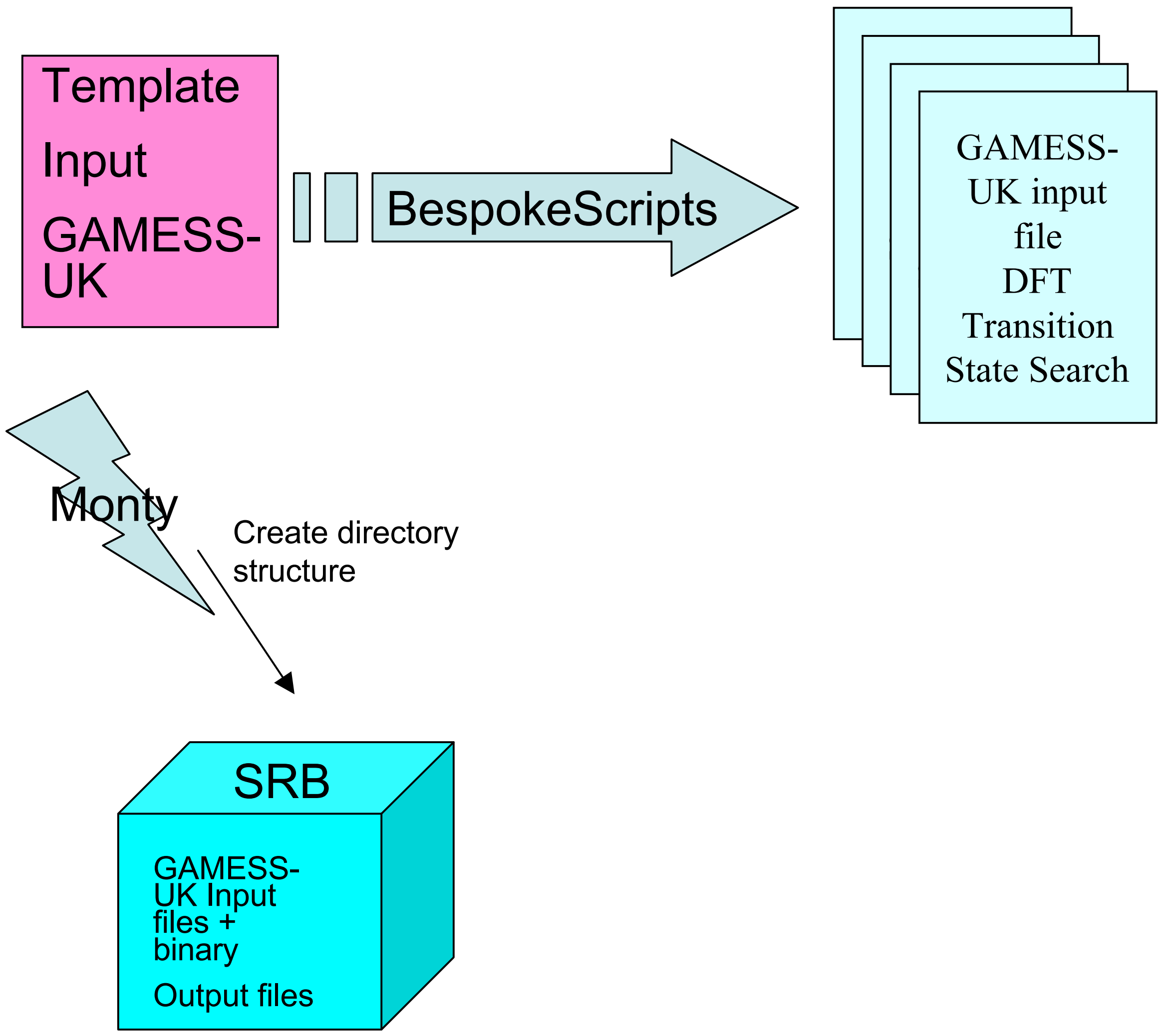


Template

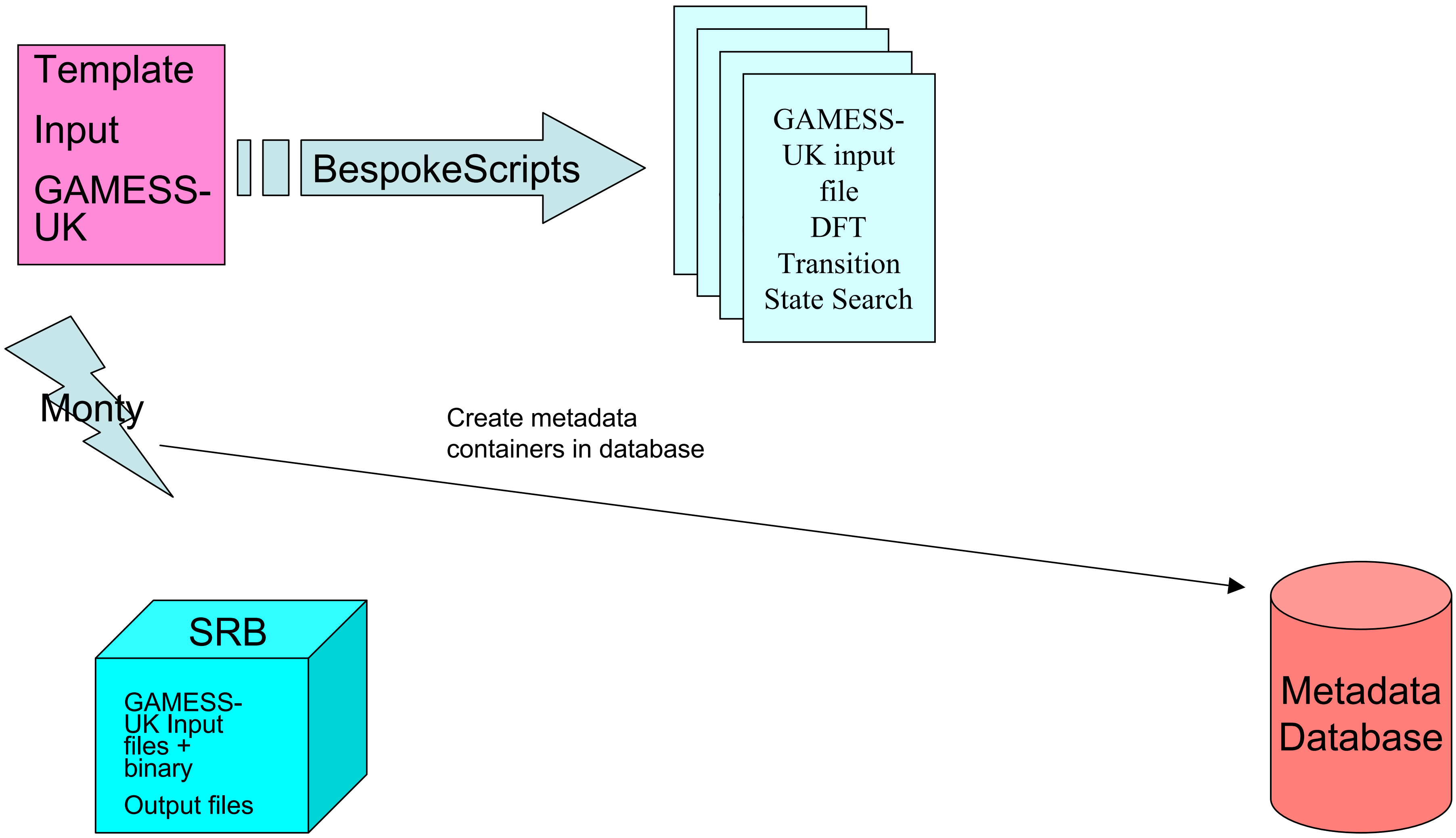
Input

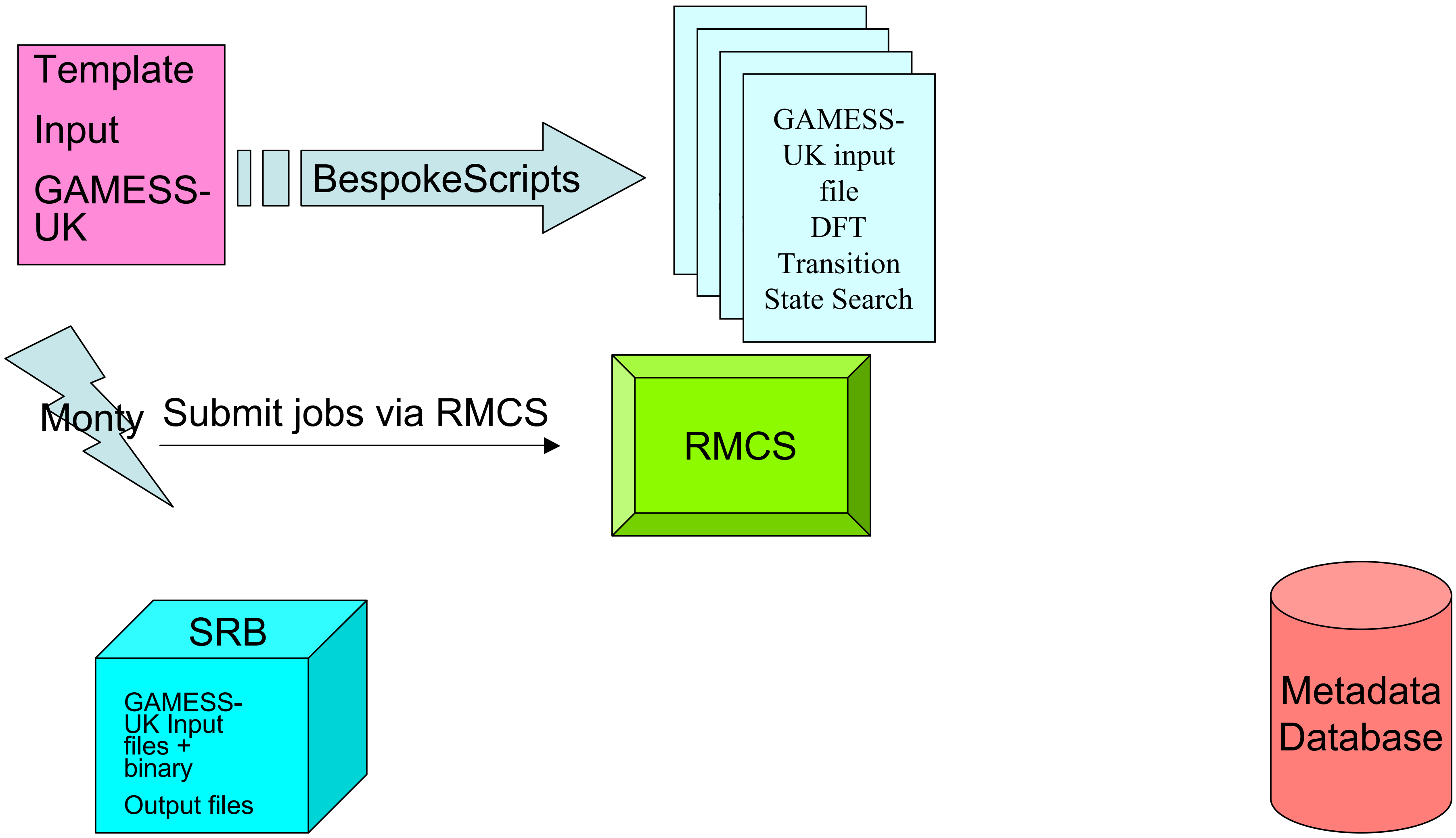
GAMESS-UK

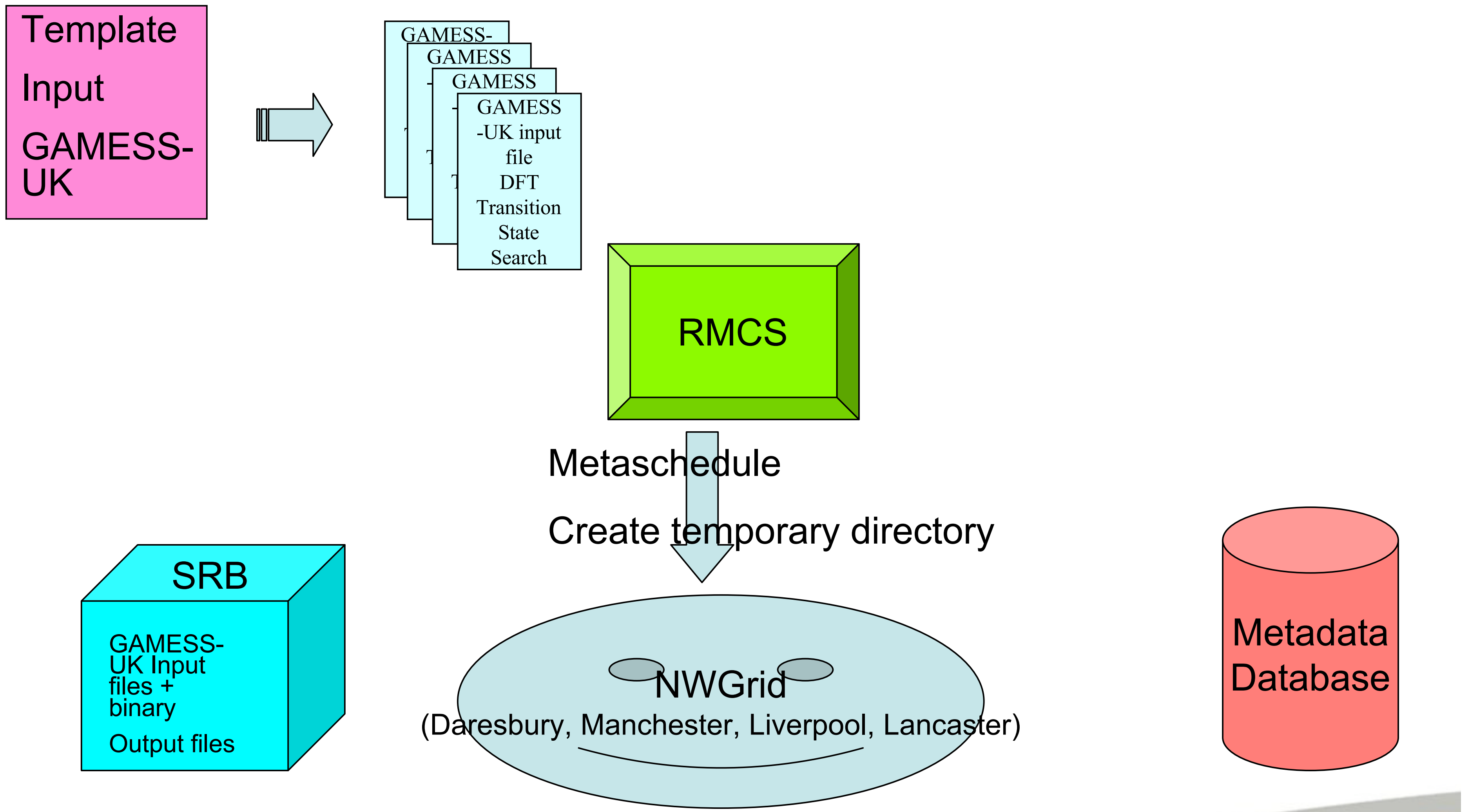

Template
Input
GAMESS-UK
BespokeScripts
GAMESS-UK input file
DFT
Transition State Search

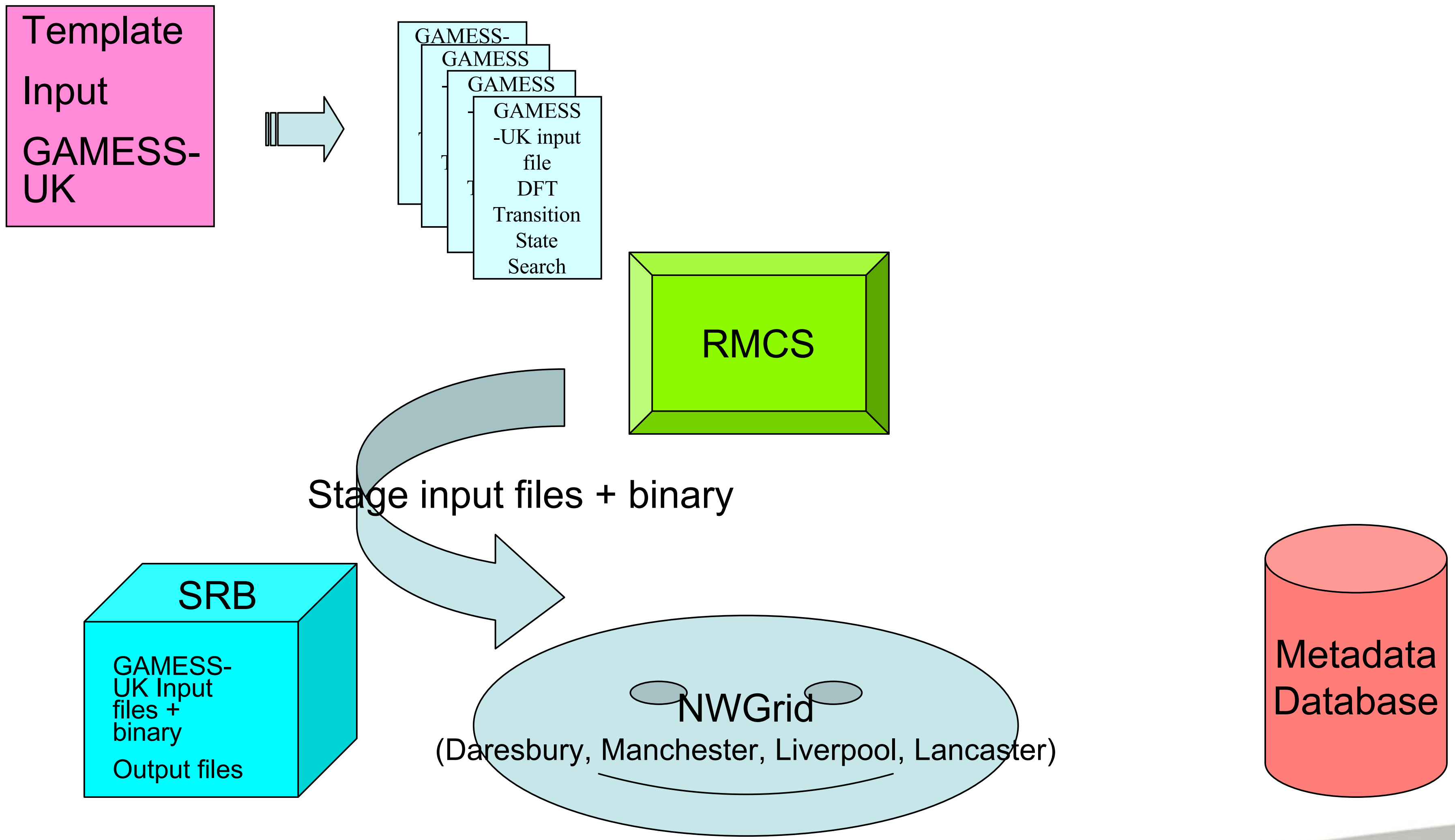


Template
Input
GAMESS-UK
BespokeScripts
GAMESS-UK input file
DFT Transition State Search
Monty
Create directory structure
SRB
GAMESS-UK Input files + binary
Output files

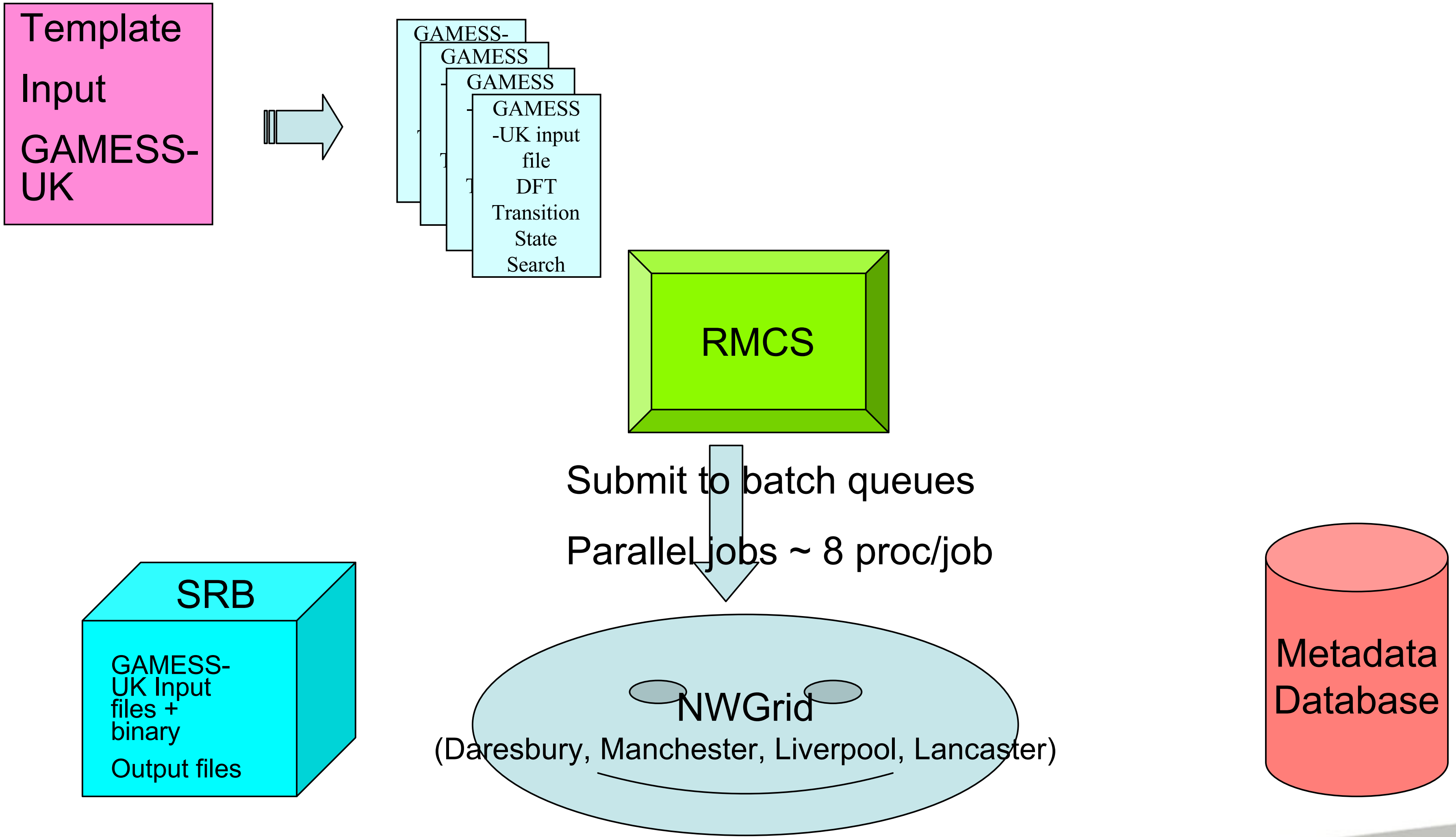

Template
Input
GAMESS-UK
BespokeScripts
GAMESS-UK input file
DFT
Transition
State Search
Monty
Create metadata containers in database
SRB
GAMESS-UK Input files + binary
Output files
Metadata Database

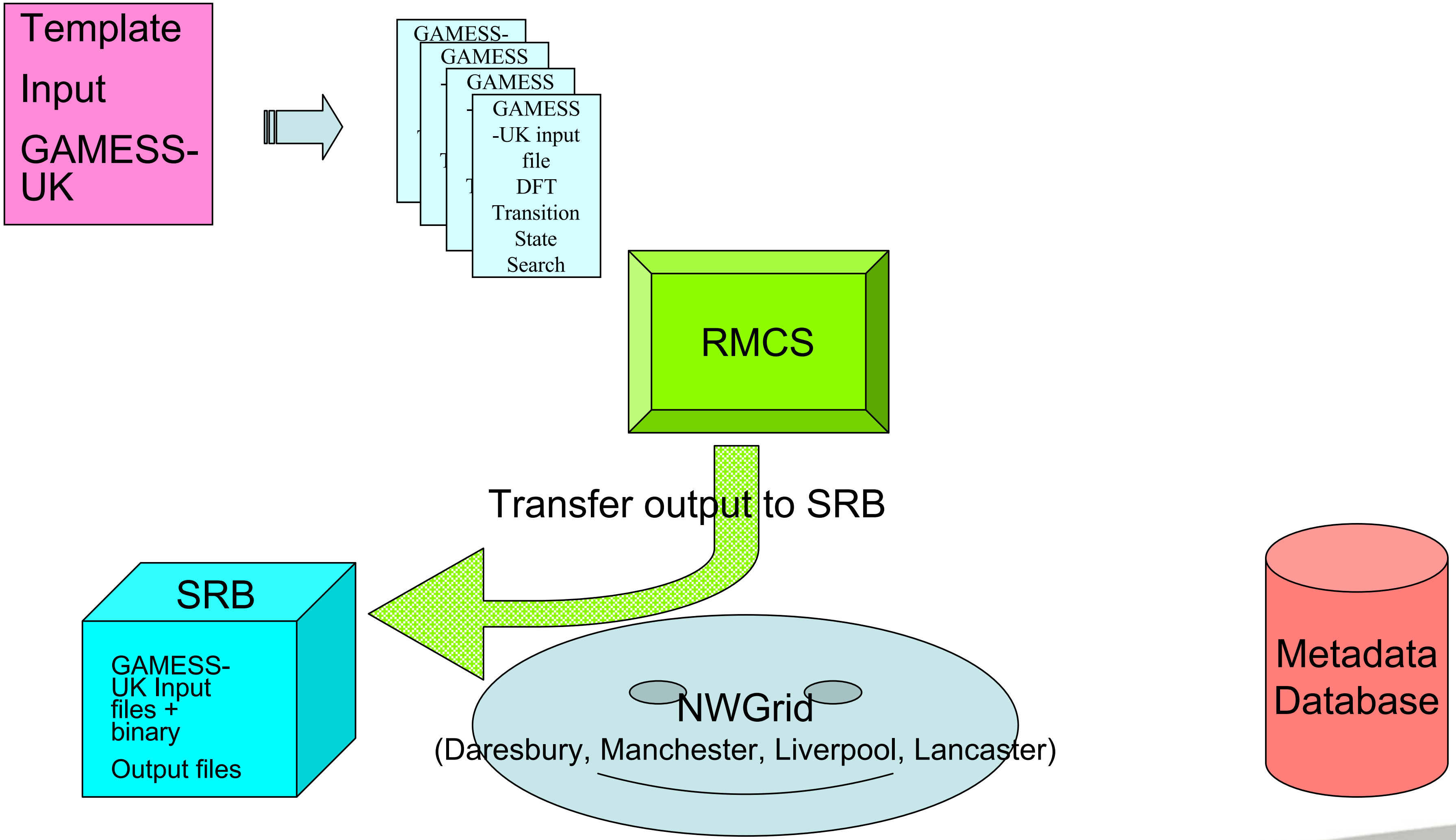

Template
Input
GAMESS-UK
BespokeScripts
GAMESS-UK input file
DFT Transition State Search
Monty
Submit jobs via RMCS
RMCS
SRB
GAMESS-UK Input files + binary
Output files
Metadata Database

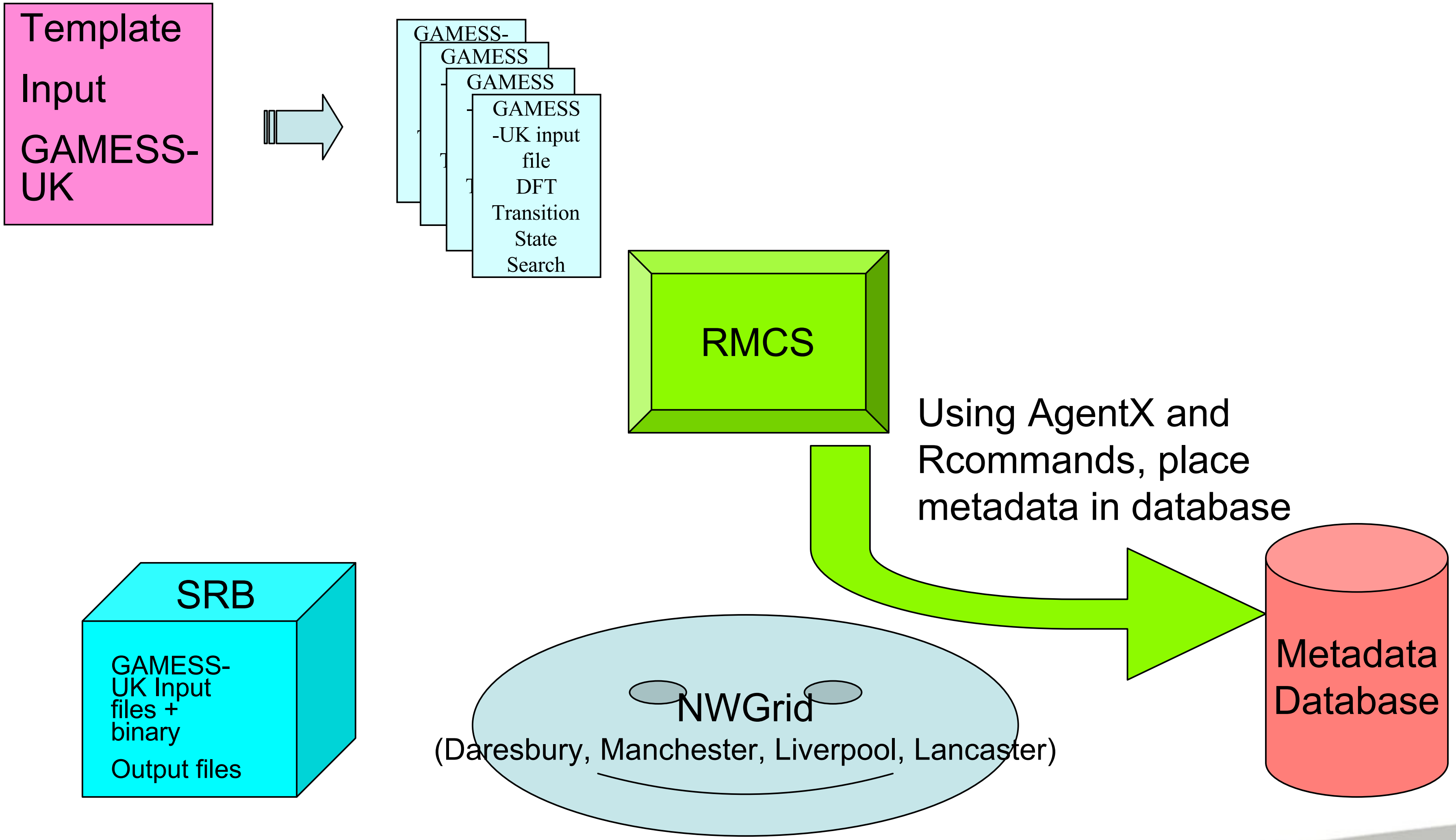

Template
Input
GAMESS-UK
GAMESS-UK input file DFT Transition State Search
RMCS
Metaschedule
Create temporary directory
SRB
GAMESS-UK Input files + binary
Output files
NWGrid
(Daresbury, Manchester, Liverpool, Lancaster)
Metadata Database



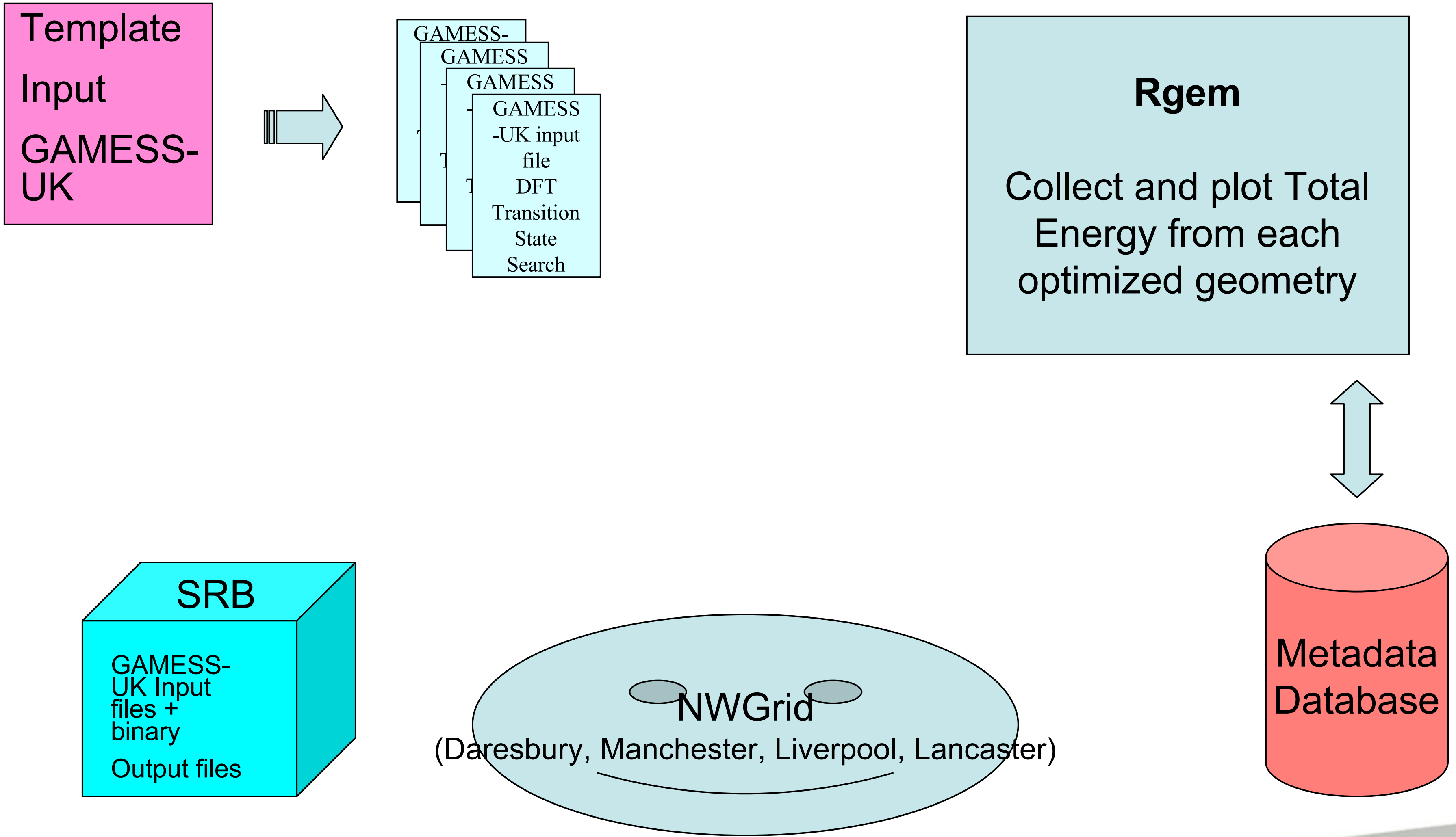
Template
Input
GAMESS-UK
GAMESS-
GAMESS
GAMESS
GAMESS
-UK input
file
DFT
Transition
State
Search
RMCS
Stage input files + binary
SRB
GAMESS-UK Input files + binary
Output files
NWGrid
(Daresbury, Manchester, Liverpool, Lancaster)
Metadata Database

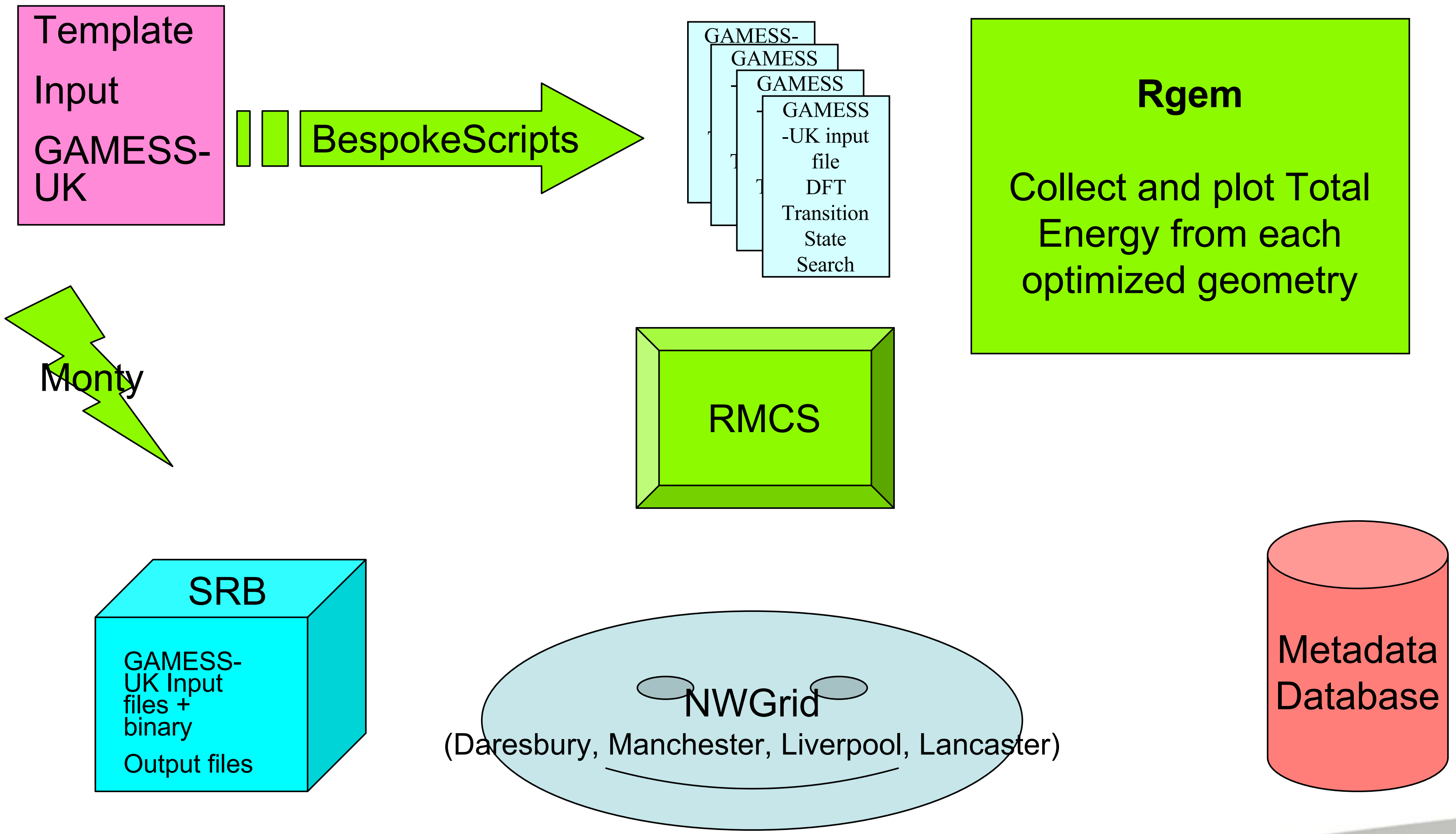

Template
Input
GAMESS-UK
GAMESS-
GAMESS
GAMESS
GAMESS
-UK input
file
DFT
Transition
State
Search
RMCS
Submit to batch queues
Parallel jobs ~ 8 proc/job
SRB
GAMESS-UK Input files + binary
Output files
NWGrid
(Daresbury, Manchester, Liverpool, Lancaster)
Metadata
Database


Template Input GAMESS-UK
GAMESS-UK input file DFT Transition State Search
RMCS
Transfer output to SRB
SRB
GAMESS-UK Input files + binary
Output files
NWGrid
(Daresbury, Manchester, Liverpool, Lancaster)
Metadata Database



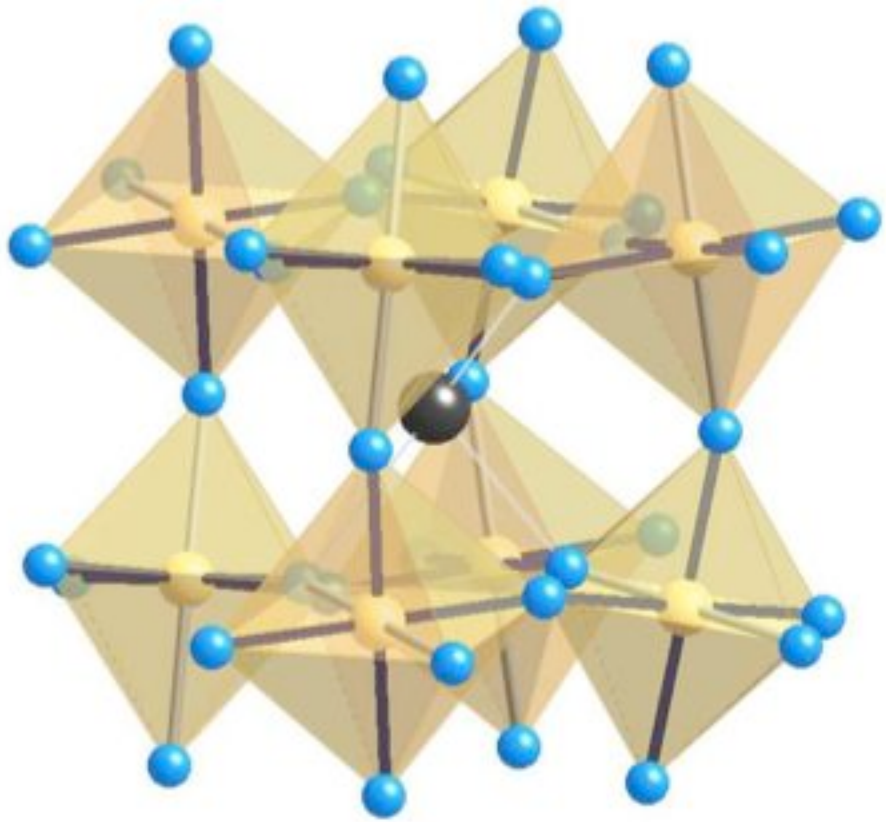
Template
Input
GAMESS-UK
GAMESS-
GAMESS
GAMESS
GAMESS
-UK input
file
DFT
Transition
State
Search
RMCS
Using AgentX and Rcommands, place metadata in database
SRB
GAMESS-UK Input files + binary
Output files
NWGrid
(Daresbury, Manchester, Liverpool, Lancaster)
Metadata Database


Template Input GAMESS-UK
GAMESS-
GAMESS
GAMESS
GAMESS -UK input file DFT Transition State Search
Rgem
Collect and plot Total Energy from each optimized geometry
SRB
GAMESS-UK Input files + binary
Output files
NWGrid
(Daresbury, Manchester, Liverpool, Lancaster)
Metadata Database


Template
Input
GAMESS-UK
BespokeScripts
GAMESS-
GAMESS
GAMESS
GAMESS
-UK input
file
DFT
Transition
State
Search
Rgem
Collect and plot Total Energy from each optimized geometry
Monty
RMCS
SRB
GAMESS-UK Input files + binary
Output files
NWGrid
(Daresbury, Manchester, Liverpool, Lancaster)
Metadata Database

Manpower reduced from several weeks to several hours through use of eScience, RMCS framework!

Total Energy Investigations of Perovskite Structured Materials - Leon Petit



Daresbury Laboratory
Science & Technology Facilities Council

Conclusions

- Grid computing can make computational science easier, more effective
- With right tools, new users can be "converted"
- Grids like NWGrid with fast processors, large compute and data capacity expand the scope of grid computing to include concurrent, parallel jobs