Virtual Earthquake and seismology Research Community e-science environment in Europe
Project 283543 – FP7-INFRASTRUCTURES-2011-2 – www.verce.eu – info@verce.eu

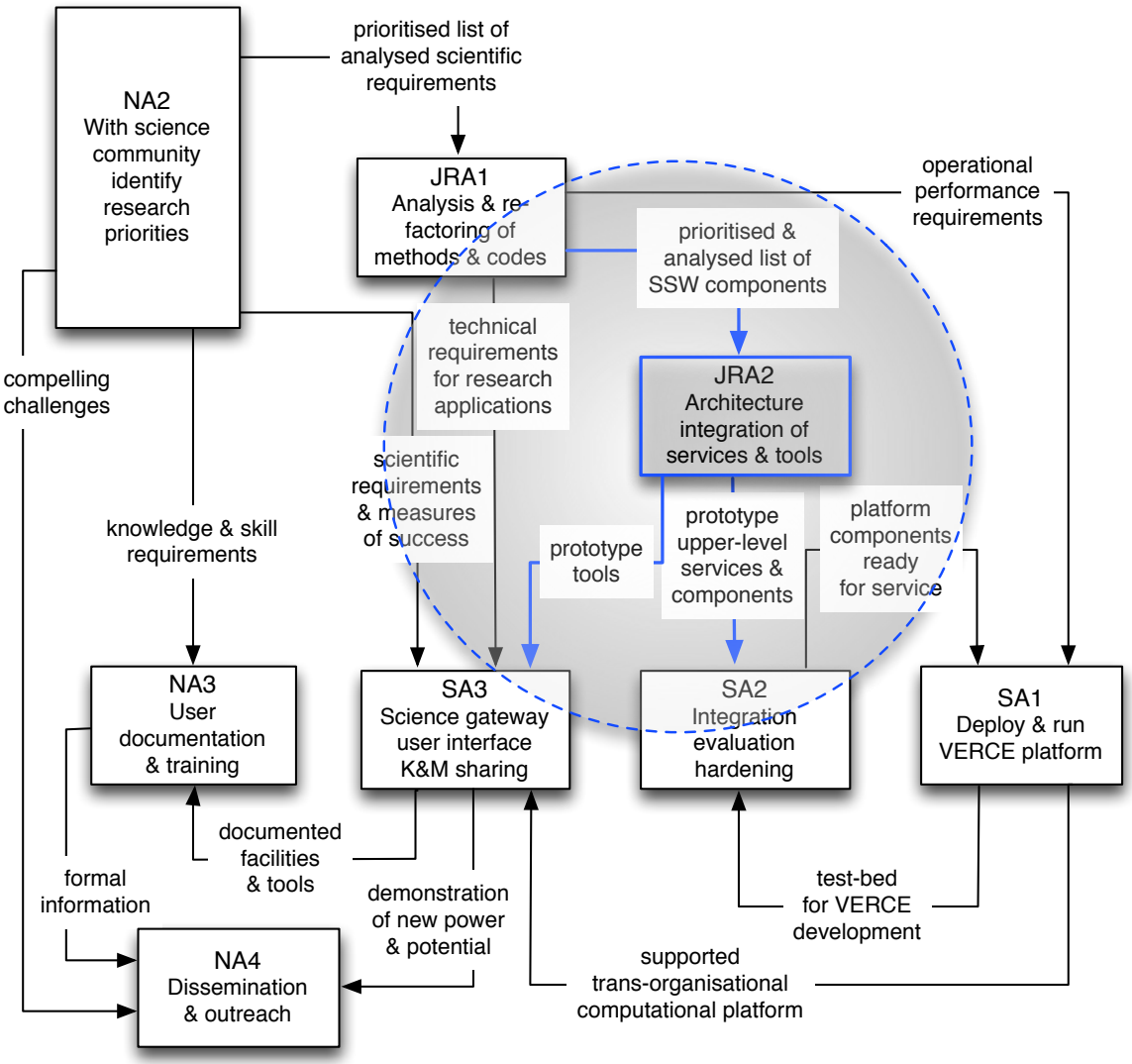
WP9-JRA2: Architecture Components and Perspectives

Malcolm Atkinson*
University of Edinburgh, UK
Malcolm.Atkinson@ed.ac.uk

Paris 25 April 2013

*Paul Martin, Malcolm Atkinson, Michelle Galea, Iraklis Klampanos, Amrey Krause, Rosa Filgueira, Chee Sun Liew, Alessandro Spinuso, Luca Trani, Siew Hoon Leong and Xiao Wang

JRA2 Interaction within VERCE



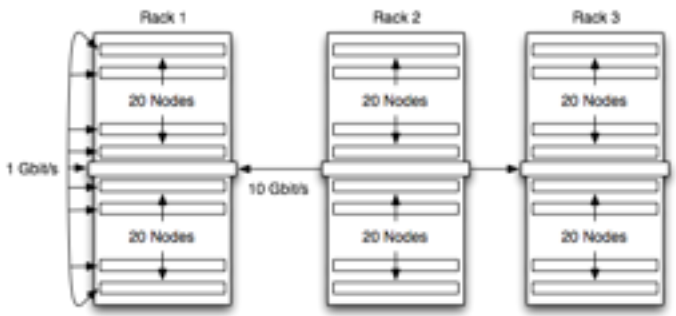
Supporting Scientific Computation

- Scientific advances are proportional to our ability to compute
  - Numerical & Parallel HPC models
- New and growing wealth of data
  - Multiple data sources: instruments, surveys & simulations
  - Diversity in tools, methodologies and needs
  - Diversity in systems and computational methods
- Needs agile development of and experiments with new algorithms
- Automatically manage and combine data and derivatives regardless of volume, type or location
- VERCE requires a marriage of data-intensive & CPU-intensive cultures
  - Developing common computational tools across distributed specialised platforms

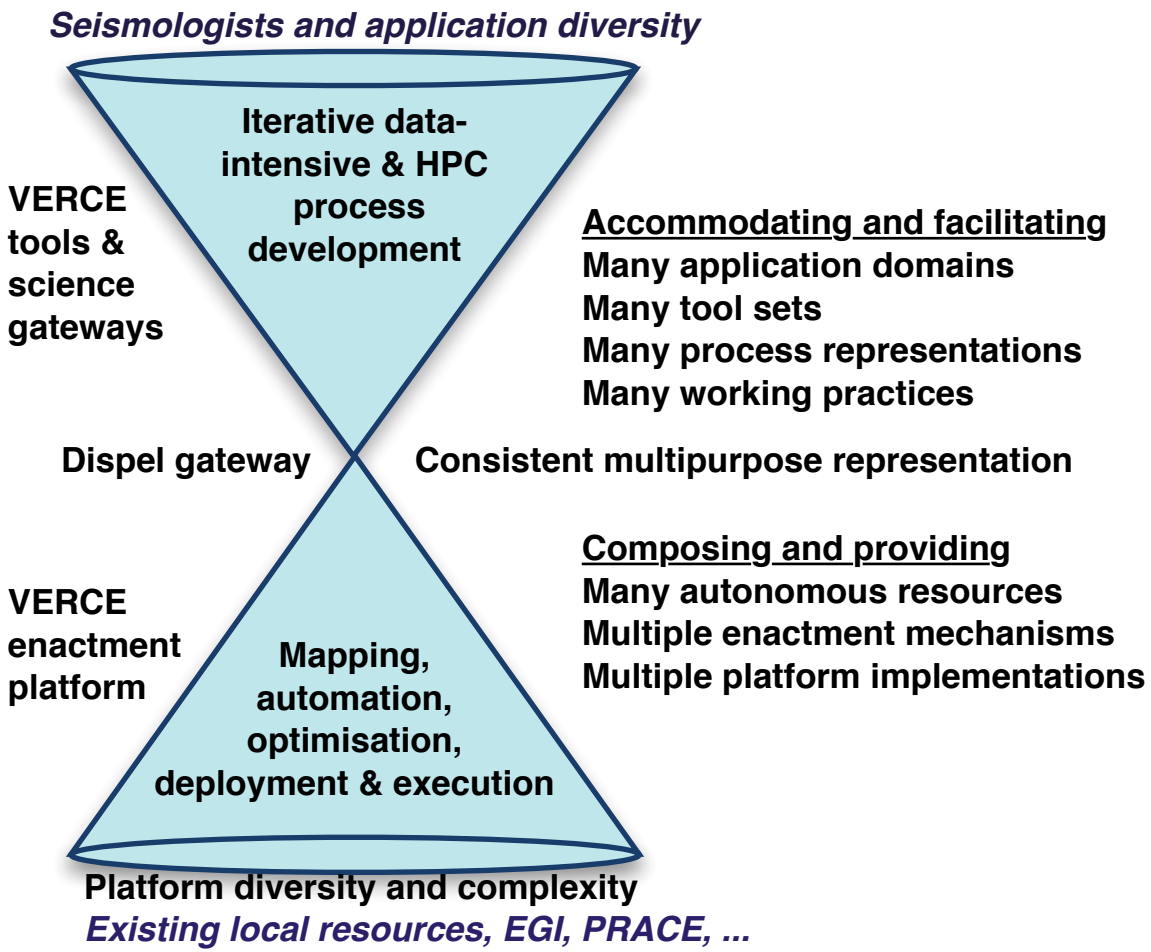
Defining "Data-Intensive"

- Generally
  - A computational task is data-intensive if you have to think hard about an aspect of data handling to make progress
  - distribution, permissions and rules of use, complexity, heterogeneity, rate of arrival, unstructured or changing structure, long tail of small and scattered instances, size of data, number of users, ethical and legal constraints

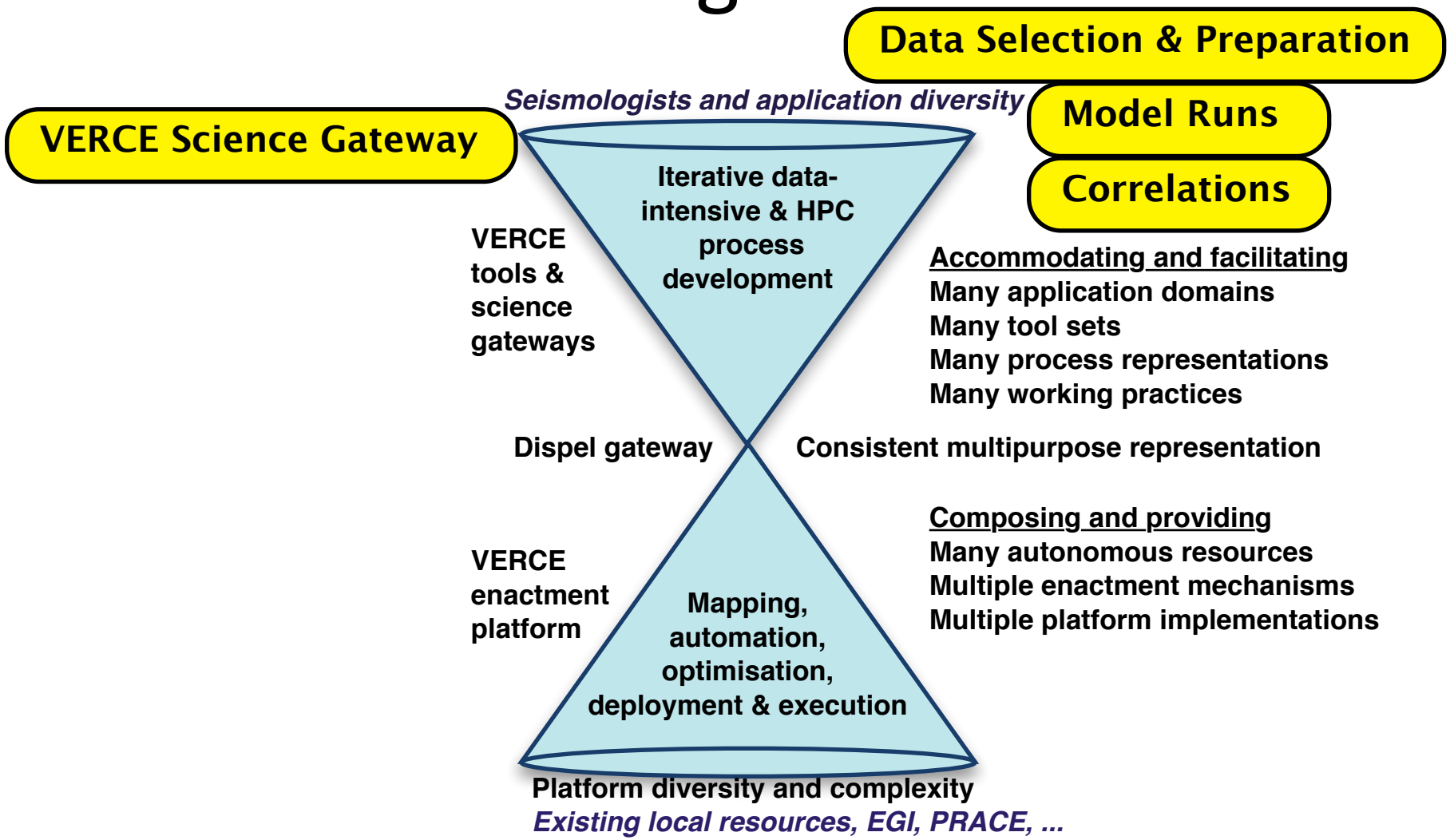
- Quantitatively
  - The computation's Amdahl numbers are close to 1
  - bits transferred in or out of memory / CPU operations in unit time
  - 1 I/O operation / 1000 CPU operations



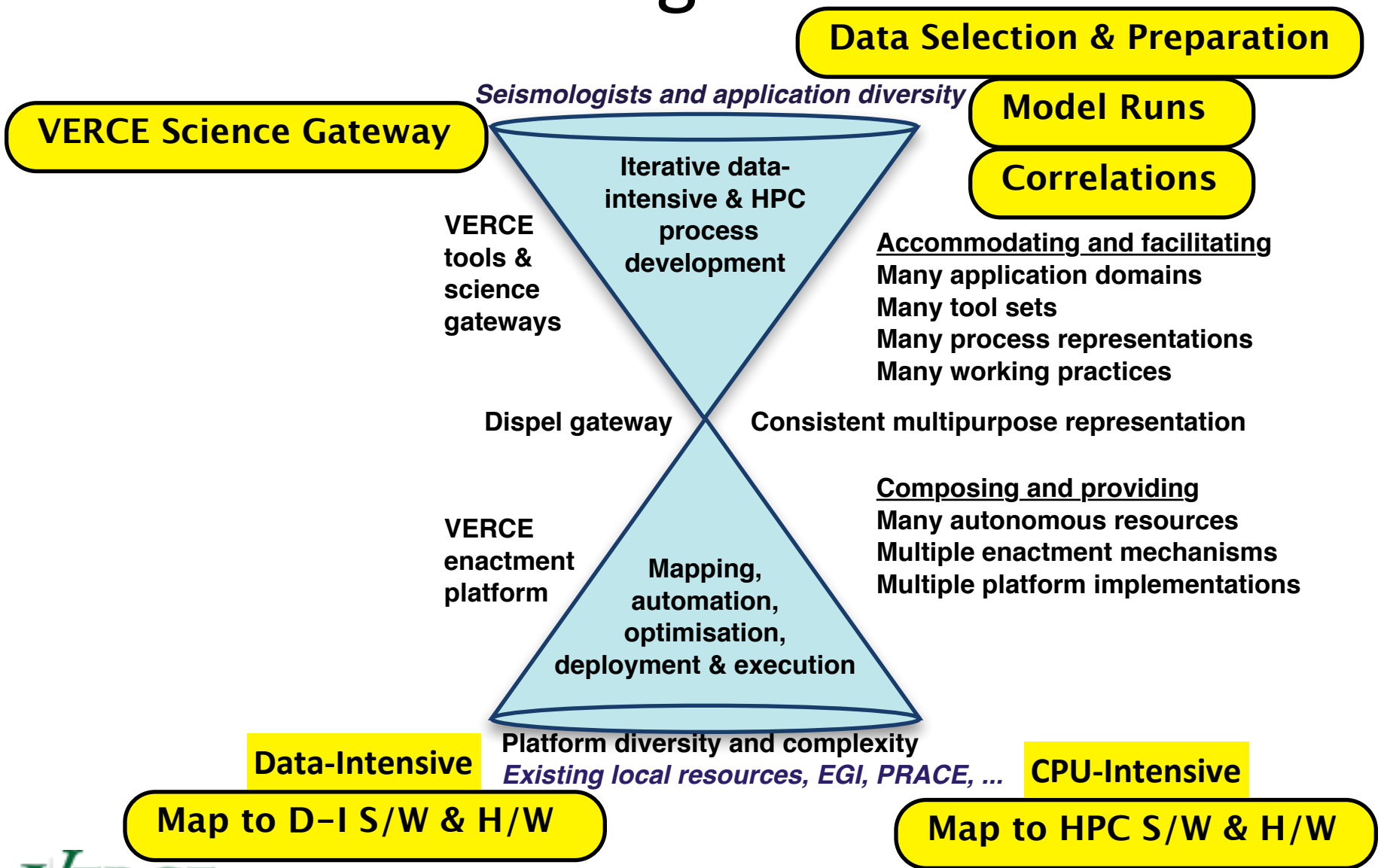
The Hourglass Model



The Hourglass Model

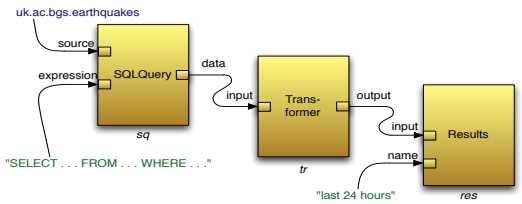


The Hourglass Model

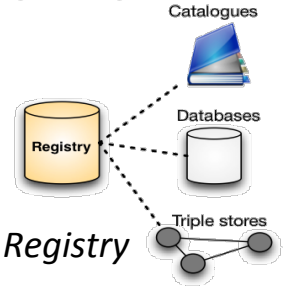


Architectural Elements

Core components



Workflow language



Registry

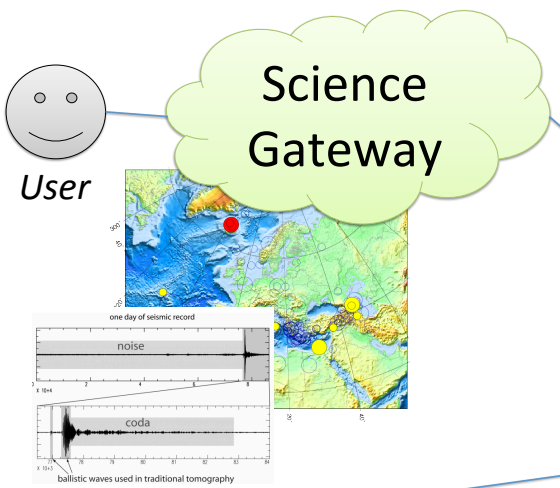
Enactment-Execution

Science Gateway

- Customisable workflows
- Localised views of data and programs
- Re-use of methods and data

– VERCE Platform –

- Movement between local grids, stores and HPCs.
- Transparent management of derivative data, redundancy, provenance



Services Integration Layer

Scientific Catalogues

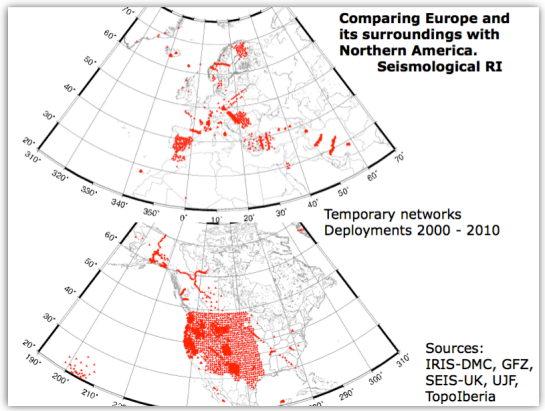
Data and metadata Stores

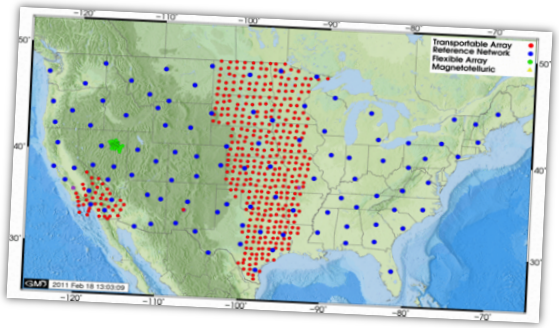
Computing Resources

VERCE-wide Registry of components, Provenance

Synergies: EUDAT, PRACE, EGI, SCI-BUS, etc.

Globus, UNICORE, OGSA-DAI, SAGA, etc.



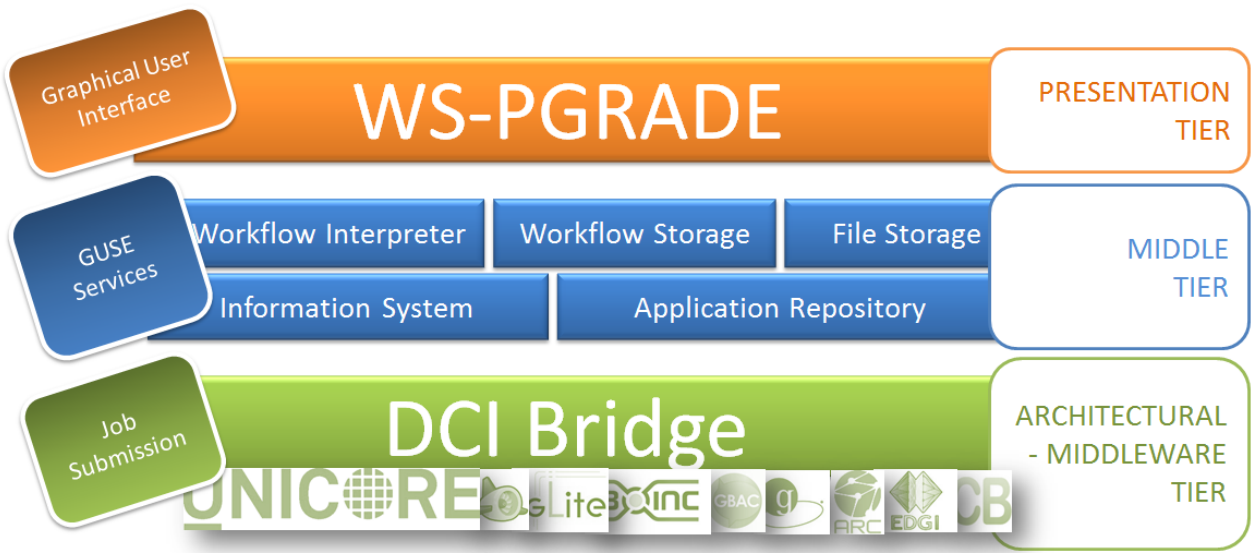


Architectural Components

- **Science gateway**
  - SCI-BUS, ER-FLOW: external, under evaluation
  - Liferay-specific components: partially operational, under development
- **Data management**
  - iRODS: external, under evaluation
  - MonetDB service: external, under evaluation
  - Distributed file systems: to be evaluated
- **Job submission**
  - JSAGA: external, operational
  - DCI-Bridge: external, under evaluation
  - Globus toolkit, Unicore/X and prerequisites: operational
- **Dispel components**
  - Gateway: operational
  - OGSA-DAI: operational
  - Storm: external, under evaluation
  - Registry: partially operational, under development
  - Provenance: partially operational, under development
- **Various**
  - GridFTP, ObsPy, SeisSol, Specfem3D: operational

Collaborating with other projects

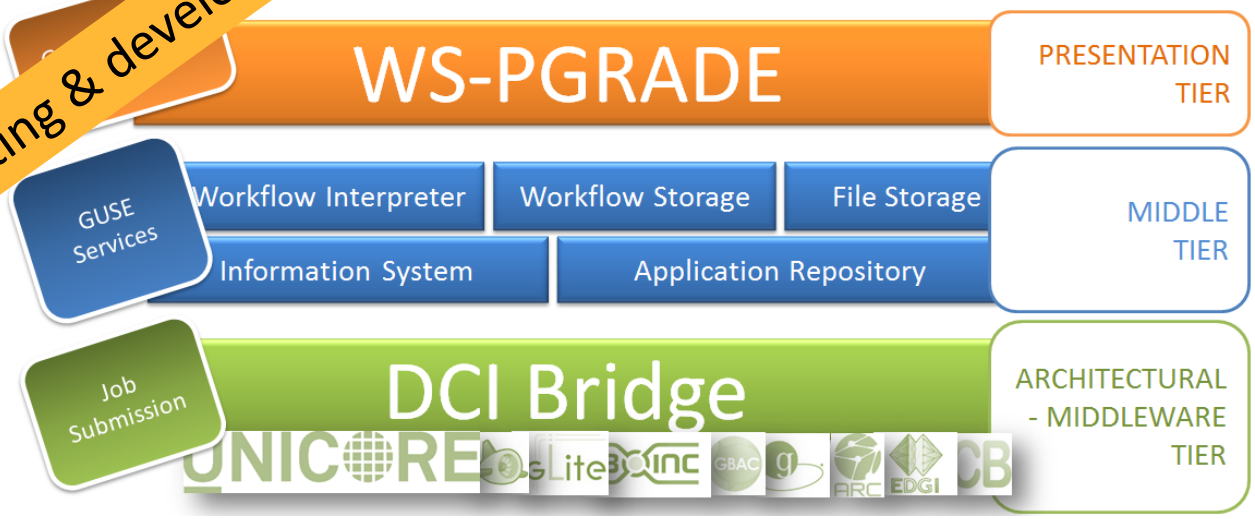
- EUDAT
- EGI/PRACE/EUDAT Synergy
- SCI-BUS
- ER-Flow



Collaborating with other projects

- EUDAT
- EGI/PRACE/EUDAT Synergy
- SCI-BUS
- ER-Flow

Investigating, selecting & developing opportunities for shared technical solutions



Targets in next 6 Months

- Perform data-intensive & CPU-intensive experiments at significant scales
- Establish the registry for
  - automation & optimisation
  - personalisation
  - resource management
- Initial model for defining work-resource-scientist relationships
- Exploit critical collaboration synergies

Roadmap: 12 Months

- Well-engineered *optimised* mapping of use-cases to platforms
  - both HTC & HPC resources
  - diversity of local resources & data sources
  - reliability at scale
  - performance at scale
- Prototype of sophisticated multi-faceted VERCE Science Gateway
  - Multiple kinds of seismic research
  - Integrated multi-purpose visualisation
  - personal data spaces
  - scientists' personal profiles
- Open up to selected external users for testing

Summary

- JRA2 deliverables & milestones have all been achieved
- Current architecture has supported a working system
  - for data-intensive use cases
  - for cpu-intensive use cases
- All members of the task forces have now experienced the architecture
  - as a framework to integrate across platforms
  - as a specification of the future technical requirements
- Six-month targets
  - stress testing; critical model of complex distributed computations in registry, performance, focused alliances
- Twelve-month targets
  - engineering, optimisation, access to diverse research capabilities, on all sites

Thank you!

contact: Malcolm.Atkinson@ed.ac.uk

more information: www.verce.eu

Question Time