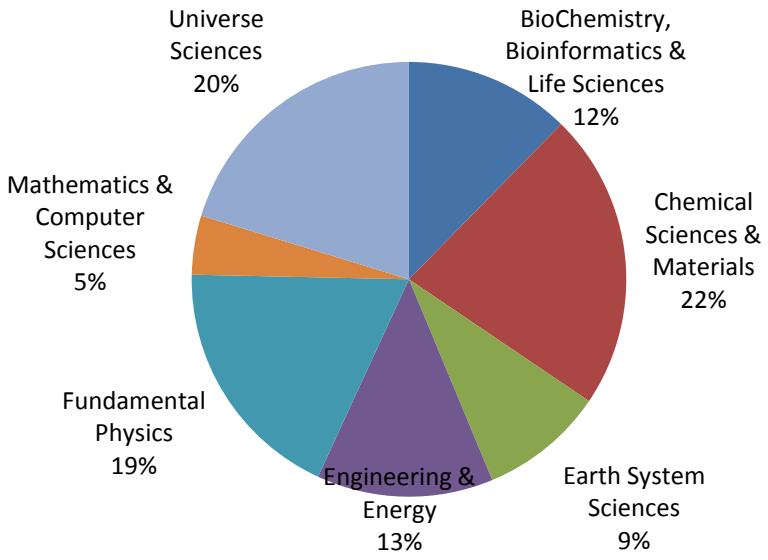
Understanding PRACE in the light of Data Sharing and Interoperability & RDA Relevance

Morris Riedel, Jules Wolfrat et al.
Juelich Supercomputing Centre,
SURFsara

Brussels, 17 October 2014

PRACE driving Science

- PRACE offers scientific communities access to large-scale HPC systems based on peer-reviewed-based scientific cases
  - Application enabling tailor scalable solutions to large-scale computational needs



- Simulation Sciences
  - Demand for sharing and transferring
  - Interoperability: EUDAT, XSEDE, EGI
- Data-driven sciences
  - e.g. in XSEDE-PRACE Earth science cases
  - RDA Big Data Analytics interest group: bottom-up understanding of methods for analytics/analysis in science cases

RDA
RESEARCH DATA ALLIANCE

Top 5 of User requirements

- Support for the use of Persistent IDentifiers (PIDs)
  - Enabling data sharing & replication (e.g. iRODS, GridFTP) in registered data domain
  - Keep directories in sync with data repositories: ‘trust to delete PID assigned data‘
- Quality of data and analysis
  - Sharing of quality metadata & data for re-use & reproducability (new RDA group)
  - Towards peer-review-based community steering in data solutions & storage needs
- High performance data transfers
  - between PRACE sites or between PRACE and external sites (EUDAT, XSEDE)
  - After completion of a project users have to transfer data from the PRACE site
- Use of (real-time) analysis/analytics tools
  - E.g. Ferret, NCO, NCL and ncview, piSVM, R, …
- Federated AAI
  - domain-specific ESFRIs, EUDAT, EGI, XSEDE (US), …

Thanks

PRACE welcomes user requirements and is willing to provide effort for fulfilling these