Distributed Data & Integration
Breakout Session

- What are the major challenges?
  - Lack of a common ontology
  - Inefficiency of current grid solutions
  - Bandwidth
  - Security
- These issues are not specific to data mining – we should seek solutions from the grid and database communities.

Ontologies

- Problem:
  - ∄ a common ontology to make joins between distributed data feasible without hand-crafted solutions.
- State of the art:
  - Partial, bottom-up solutions such as UCDs
- What should be done next?
  - Work towards completing ontologies within each discipline
  - Cross-discipline ontologies?
- What can the SIG do to help?
  - Establish groups to develop ontologies for those disciplines that lack them

Efficiency and bandwidth

- Problems:
  - Lack of!
  - Inefficiency of current grid solutions for federating data
  - Inefficiency of XML databases
- State of the art:
  - BinX?
- What can the SIG do to help?
  - Access to higher bandwidth and related expertise - UKLight?

Security

- Problems:
  - Current solutions not *seen as secure enough*
  - *Adding security can degrade performance still further*

Misc.

- What else can the SIG contribute?
  - Define standard tests/criteria for benchmarking software
  - Form a group to look at the data integration problem in detail
  - Raise a challenge to be met – an X-prize (beer, pint, one) for data integration?

Comments, suggestions, corrections to John Taylor, jdt@roe.ac.uk