



Enabling Grids for E-science

The Data Management Session

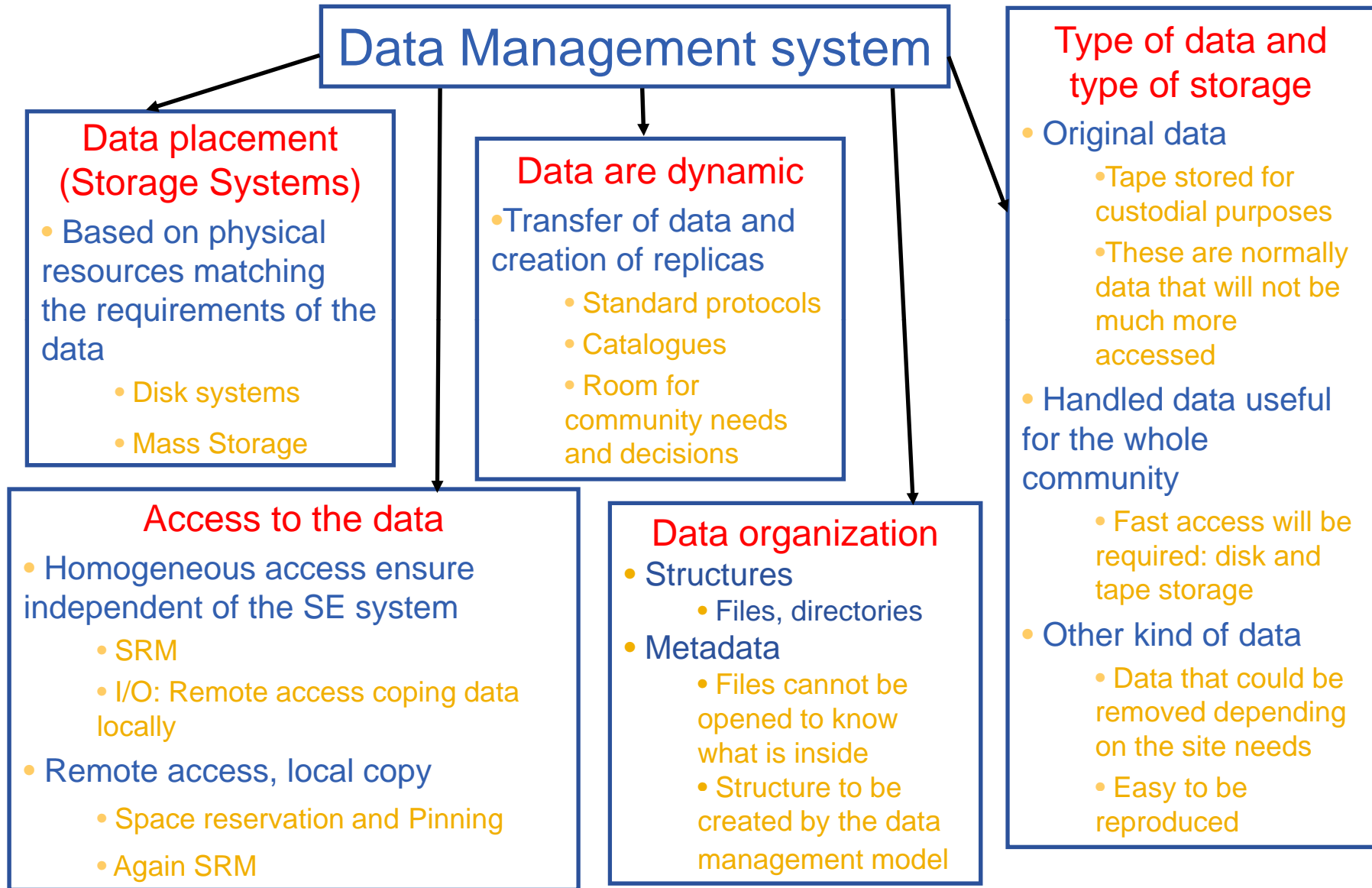
Conveners: Pasquale Pagano (CNR-ISTI) and Patricia Méndez Lorenzo (CERN)

3rd EGEE User Forum. Clermont, Wednesday 13th February 2008

www.eu-egee.org



- In the past distributed computing was focused on Workload management
 - Support of large scale distributed computational tasks
 - Resources fairshare, stability
- Data management is currently the discussions and development hot topic
 - Many Grid applications and VOs need to manage large amount of data (~PB)
 - High Energy Physics: ~10PB/year, ~10M files/year
 - Earth observations, satellite imagery (NASA, UNO agencies)
 - Distributed on many sites with different storage systems, access protocols
 - Developments towards standard data access
 - Described by multiple and heterogeneous metadata
 - From LFN (few years ago) to the metadata (present and future direction)
 - Different security requirements
 - While HEP data are worldwide read accessible, privacy is a fundamental key is some other communities
 - Reliability and scalability
 - Inaccessible storage is more damaging than inaccessible CPU



- Session highlights
 - Full day – long session in this forum, 11 presentations
 - Morning – Metadata
 - Afternoon – Storage interoperability, database access, storage system design
- We remind the speakers to respect the 20min per talk (including discussion) - our contingency are the coffee/lunch breaks