Development of Data Management Components for Data Analysis in High Energy Physics within the D-Grid HEP Community Project

Project start: 9/2005 Project end: 8/2008

Coordinator: Michael Ernst, DESY

Project Partners:

University Dortmund
DESY Hamburg and Zeuthen

Associate Partners:

Humboldt Universität Berlin
John von Neumann Institute for Computing
Forschungszentrum Karlsruhe GmbH

Partners per Subcontract:

University Freiburg
Konrad-Zuse-Zentrum für Informationstechnik Berlin

Work Package 1: Data Management

1. **Development and Distribution of a *Scalable Storage Element* for the specific requirements of the High Energy Physics Community based on Standard Grid Interfaces**

   (DESY, Unis Dortmund und Freiburg, FZK)

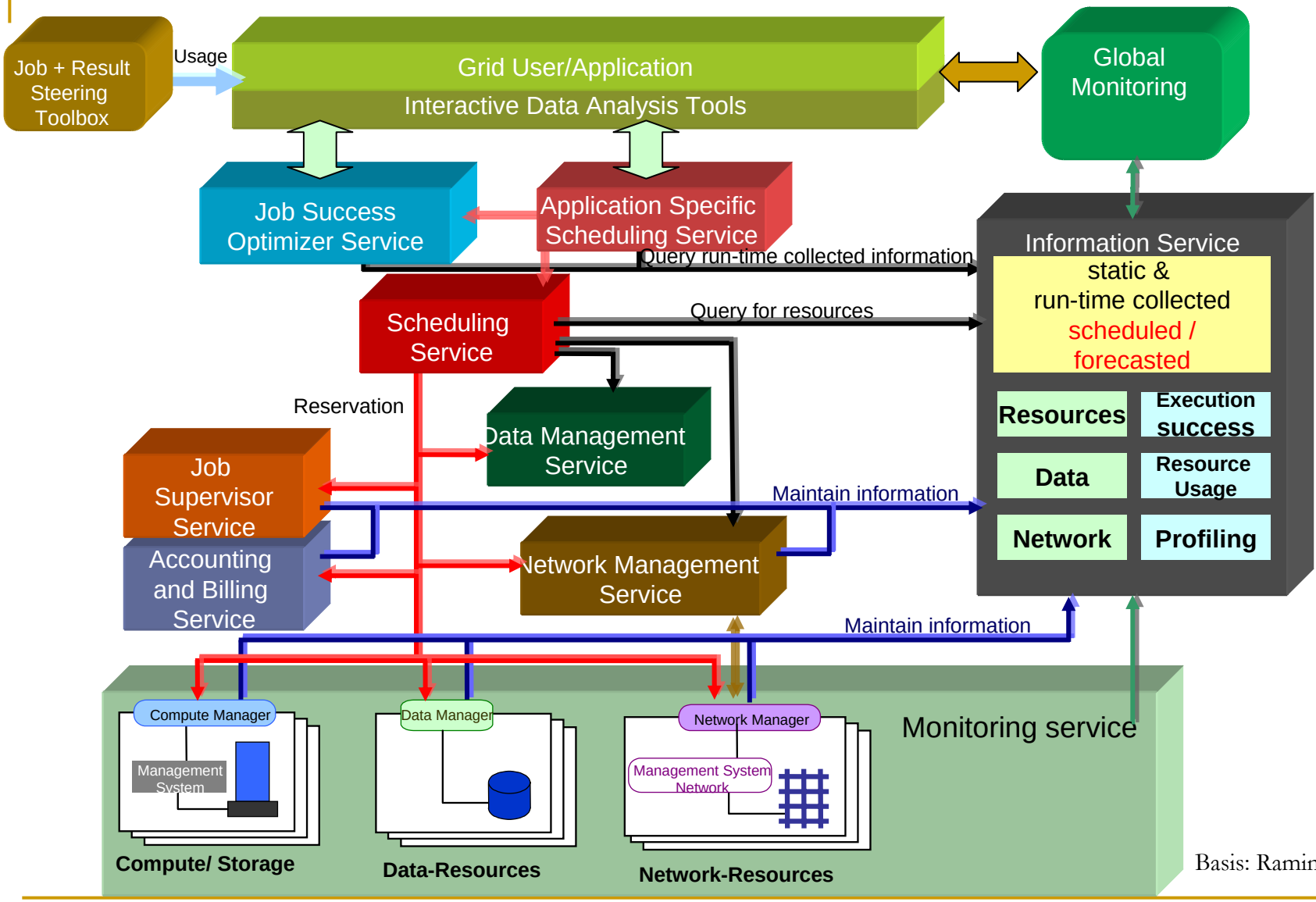
2. ***Optimization of Job Scheduling in Data Intensive Applications***

   (Uni-Dortmund (CEI und Physik), DESY – Synergy with C3 Grid)

3. **Development of *Extensible Metadata Catalogs for semantic Data Access***

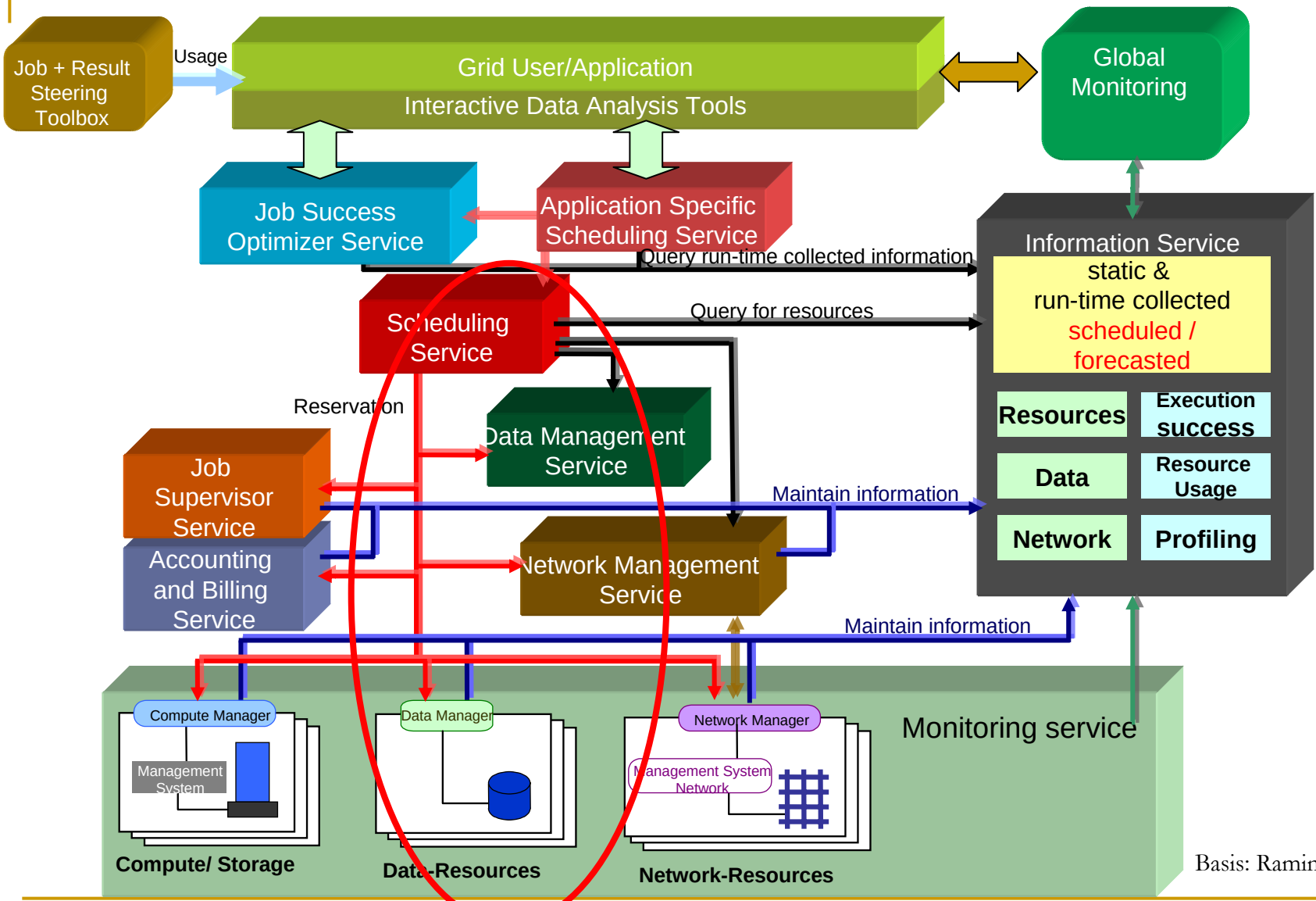
   (DESY, ZIB, HUB, NIC – Possible Synergy with AstroGrid)

Architecture and Services of HEP Community Grids



Basis: Ramin Yahyapour

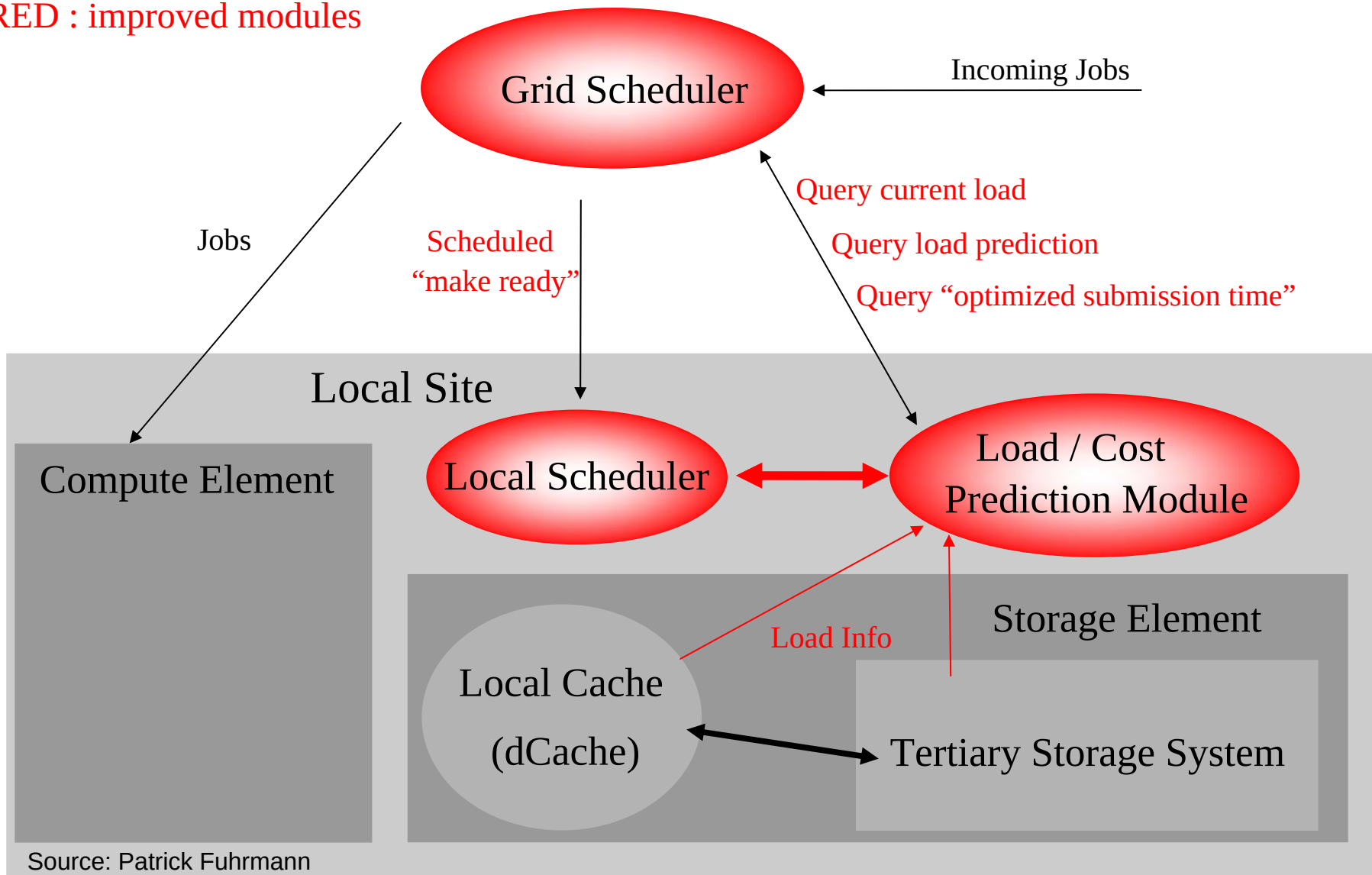
Architecture and Services of HEP Community Grids



Basis: Ramin Yahyapour

Improved Scheduling

RED : improved modules



Source: Patrick Fuhrmann

Task Breakdown and Milestones

- Tasks according to Spread Sheet
- Milestones (as they are relevant so far)
  - Scalable Storage Element

Project Month	Deliverable	Milestone
6	In HEP/LCG Environment easily installable and maintainable SE based on existing dCache functionality	Distribution of SRM-dCache Package to Partner Institutions

  - Optimized Job Scheduling

Project Month	Deliverable	Milestone
6	Overall Concept and agreed Communication Interface between Data Management and Scheduling	Complete Overall Concept

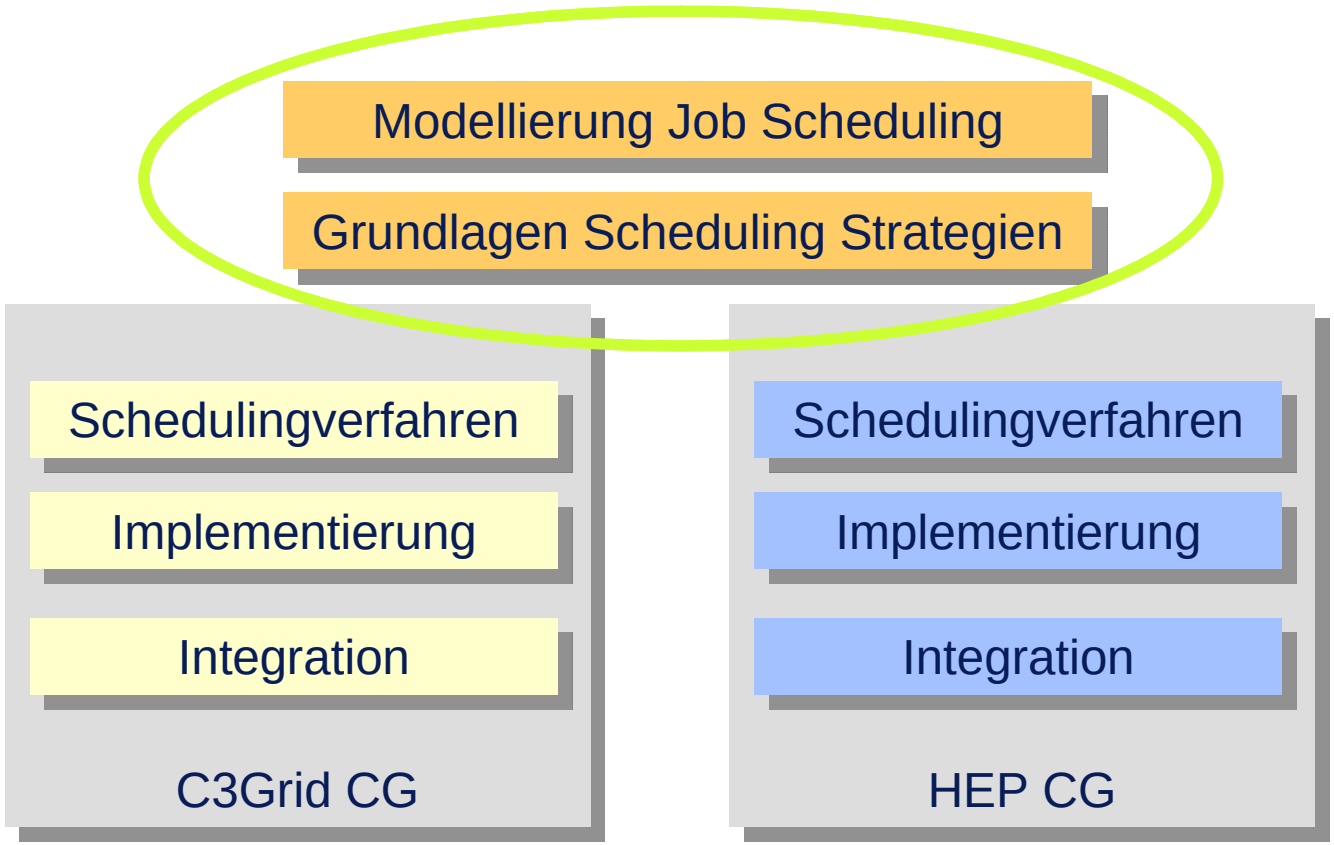
Task Breakdown and Milestones

- Extensible Meta Data Catalog

Project Month	Deliverable	Milestone
12	Create Mark-up description of Physics Data (XML schema)	Full description of data structures that is agreed within the Community

Synergies w/ other Community Projects

- Job Scheduling in C3 and HEP
  - Savings: 18 Months



Summary (1/2)

- WP1/ Data Management within HEP Community Project on track
  - Milestones reached so far (detailed presentations on subsections will follow)
    - Scalable Storage Element
      - Martin Radicke (DESY) for the dCache Team
    - Optimized Job Scheduling
      - Lars Schley (Uni-Do/IRT)
    - Extensible Meta Data Catalog
      - Dirk Pleiter (DESY & NIC)
  - Some areas ahead of plan
  - dCache Scalable Storage Element as it currently exists is already used world-wide and an important component in LHC Computing
  - Have met the EGEE Workload Management System Developers to investigate collaboration potential and find a non-intrusive way to integrate the extended scheduling mechanism

Summary (2/2)

- WP1/ Data Management within HEP Community Project on track
  - Goals as outlined in project plan will – according to current understanding - be met
  - No changes to project plan
    - Concepts and Strategies worked out and are well aligned with expectations expressed by community
  - Components developed are expected to be used as part of the DGI infrastructure and the international LHC computing environment