Developing Collaborative Software Tools - An Exercise in Collaboration.

Miron Livny
Computer Sciences Department
University of Wisconsin-Madison

A computer science view of a software engineering effort that involves researchers from three different disciplines (computer scientists, physics oriented computer professionals and physicists), who belong to two different kinds of environments (universities and national laboratories), who lead/participate in (at least) eight projects (BaBar, SRB, CMS, D0, …), and is funded by two DoE offices (HENP and MICS).

PPDG Collaborators

California Institute of Technology	**Harvey B. Newman,** Julian J. Bunn, Koen Holtman, Asad Samar, Takako Hickey, Iosif Legrand, Vladimir Litvin, Philippe Galvez, James C.T. Pool, Roy Williams
Argonne National Laboratory	**Ian Foster**, Steven Tuecke **Lawrence Price,** David Malon, Ed May
Berkeley Laboratory	**Stewart C. Loken**, Ian Hinchcliffe, **Doug Olson**, Alexandre Vaniachine **Arie Shoshani,** Andreas Mueller, Alex Sim, John Wu
Brookhaven National Laboratory	**Bruce Gibbard,** Richard Baker, Torre Wenaus
Fermi National Laboratory	**Victoria White**, Philip Demar, Donald Petravick **Matthias Kasemann,** **Ruth Pordes**, James Amundson, Rich Wellner, Igor Terekhov, Shahzad Muzaffar
University of Florida	**Paul Avery**
San Diego Supercomputer Center	**Margaret Simmons,** Reagan Moore,
Stanford Linear Accelerator Center	**Richard P. Mount,** Les Cottrell, Andrew Hanushevsky, Davide Salomoni
Thomas Jefferson National Accelerator Facility	**Chip Watson,** Ian Bird, Jie Chen
University of Wisconsin	**Miron Livny** , Paul Barford, Peter Couvares, Tevfik Kosar

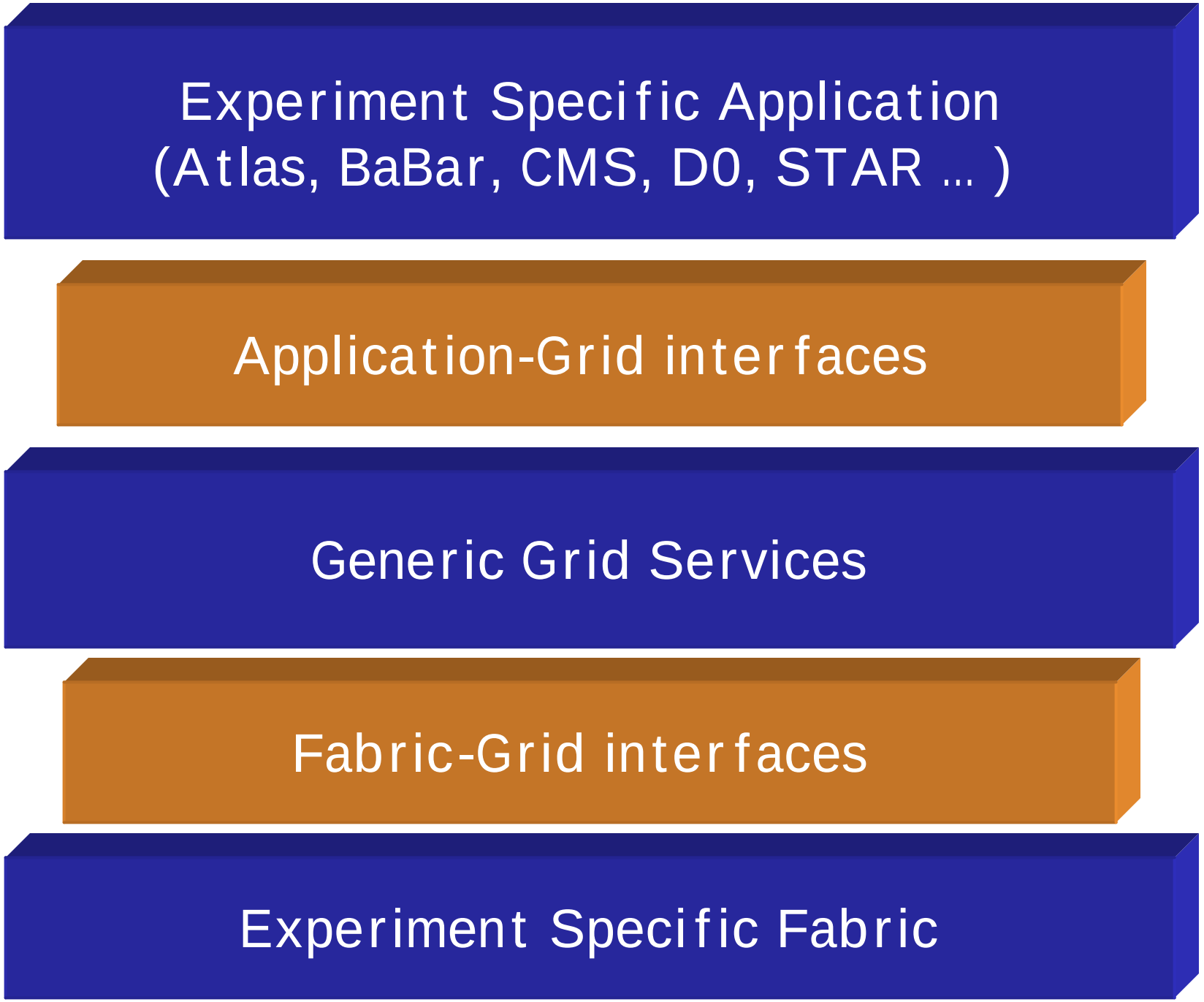
Goals of Phase I (99-01)

Design, experiment, evaluate and prototype HENP specific data transfer and data staging software tools for Grid environments.

- Overcame cultural differences
- Overcame architectural differences
- Overcame technical differences
- Leverage existing software and HW

Goals of Phase II (01-04)

- Deliver (in a timely fashion) vertically integrated Grid enabled capabilities to the experiments
- Collect and document existing distributed computing techniques and practices used by the different experiments
- Identify, implement and package common Grid enabled software tools for the HENP community.


Experiment Specific Application
(Atlas, BaBar, CMS, D0, STAR …)
Application-Grid interfaces
Generic Grid Services
Fabric-Grid interfaces
Experiment Specific Fabric

Approach

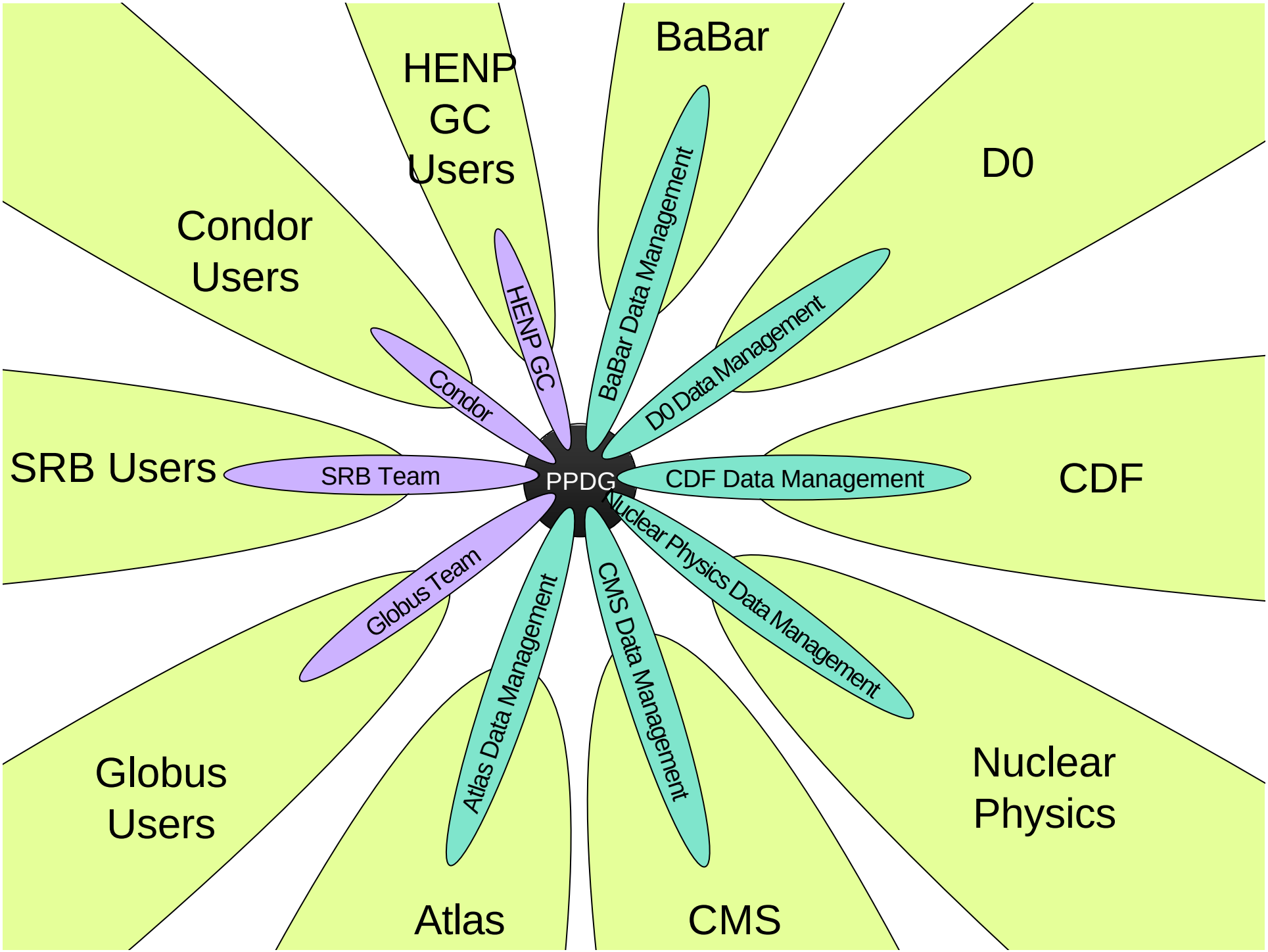
- Identify short (1 year), mid (2years) and long (3 years) term Grid computing needs of experiments
- Identify common Computer Science areas and groups with interest and capabilities in these areas.
- Define, size and budget joint (experiment-CS) sort-mid projects.
- Identify and outline mid-long term joint activities.

Computer Science Foci

- Data Transfer for placement and replication - Managable, reliable, recoverable, efficient, …. (GridFTP)
- Replica (meta-data) catalogs - … scaleable
- Mass storage access and management – (HRM)
- Job description and control - … (ClassAd)
- Caching and staging services - … (DRM)
- Monitoring and trouble shooting - … robust
- Common Interfaces (collect, define document)

Specifics (short term)

- Grid Data Management Pilot ++ (CMS)
- Production job management (DZero)
- Production data replication services (2 sites) (STAR)
- Production distributed data analysis (2 sites) (BaBar)
- Production distributed data services (Atlas)
- …


HENP
GC
Users
BaBar
Condor
Users
D0
HENP GC
BaBar Data Management
Condor
D0 Data Management
SRB Users
SRB Team
PPDG
CDF Data Management
CDF
Nuclear Physics Data Management
Globus Team
Atlas Data Management
CMS Data Management
Globus
Users
Nuclear
Physics
Atlas
CMS

The three steps

- Formulation of problem and possible solution(s).
- Development and implementation of software tools.
- Deployment of tools in production environments.

Formulation

In all three cases, effort started as a pure Computer Science (CS) research activity.

- Established as CS research activity with its own stream of funding.
- Provides a foundation for Ph.D. work in CS.

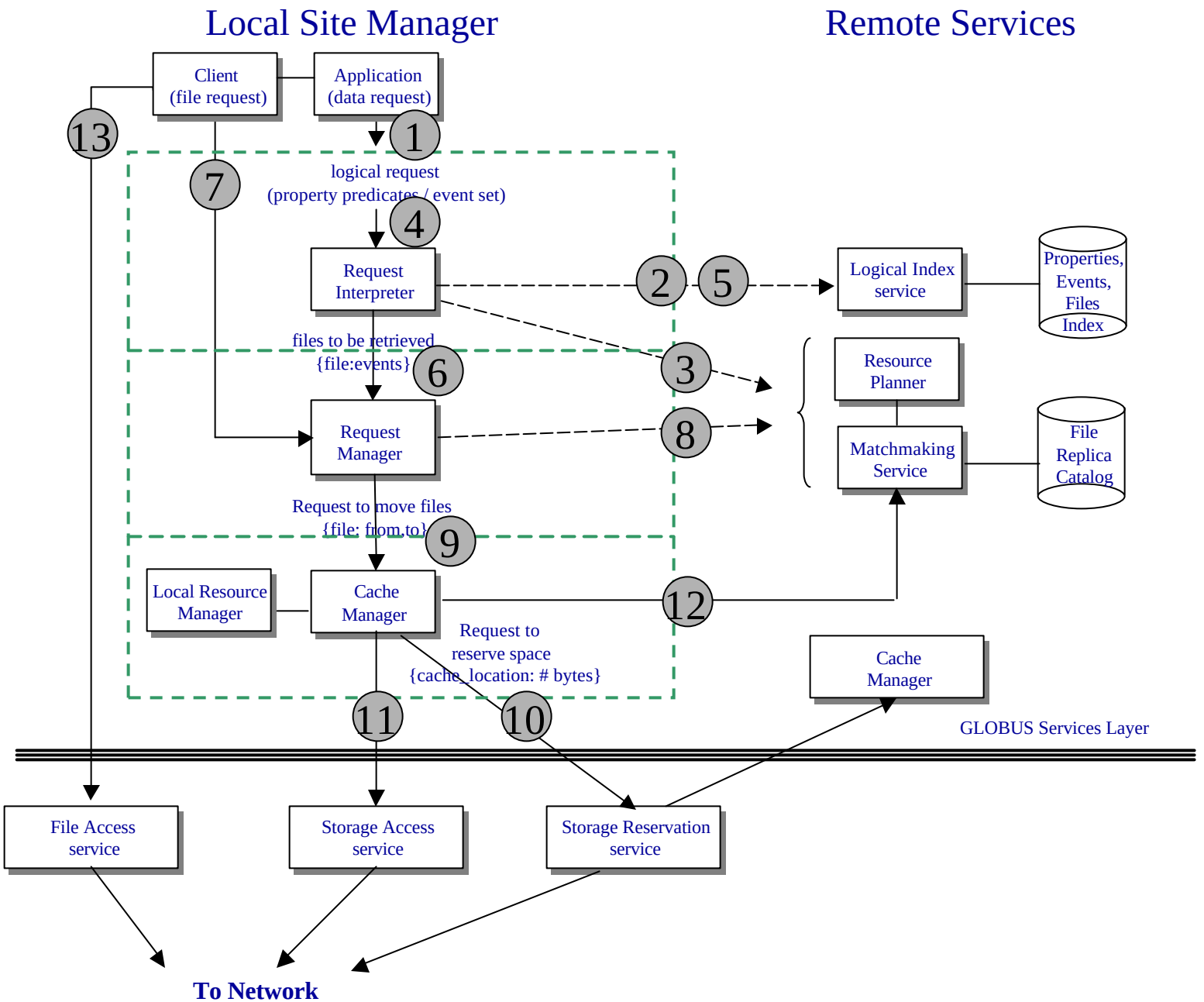


Fig 1: Architecture for the general scenario - needed APIs

Development

A collaboration (jointly funded) between Domain and CS scientists.

- Identify domain specific requirements.
- Balance generic vs. domain/case specific approaches.
- Develop appropriate abstractions and translate them to application and user interfaces.
- Identify achievable goals.

Deployment

A joint effort to put the software tools in the hands to domain scientists and have them use it in their daily research.

- Has to solve a real problem.
- Has to make their life better/simpler.
- Has to be robust.
- Has to be supported.

Expectations

- Software quality (Domain from CS):
  - Robustness.
  - Support.
  - Expected life time.
- Time commitment (CS from Domain):
  - Articulate requirements.
  - Experimentation.
  - Provide feedback.

www.ppdg.net

PPDG