The GridWay Metascheduler, - a successful Globus incubator project -

Alejandro Lorca

Distributed Systems Architecture Group
Universidad Complutense de Madrid

Universidad Complutense Madrid

La Suma de Todos
Consejería de Educación
Comunidad de Madrid
www.madrid.org

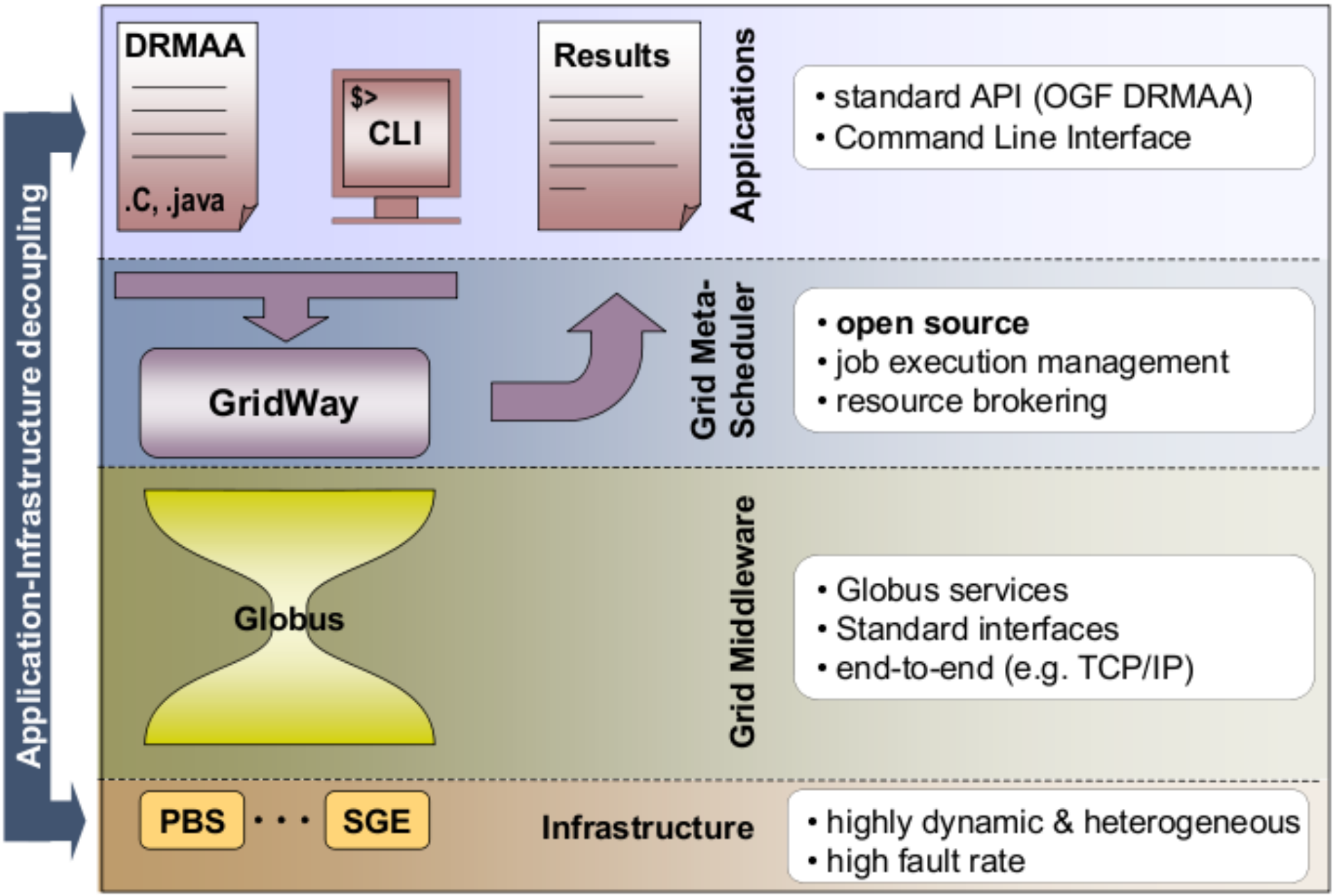
Ministerio de Educación y Ciencia

GridWay is a tool for metascheduling

It creates a virtualization layer on top of Globus services:

- MDS (Information)
- GRAM (Execution)
- GridFTP (Transfer)

What is GridWay? Computational grid architecture



What is GridWay? Benefits

Integration of non-interoperable platforms

- Establishment of a uniform and flexible infrastructure
- Achievement of greater utilization of resources and higher application throughput

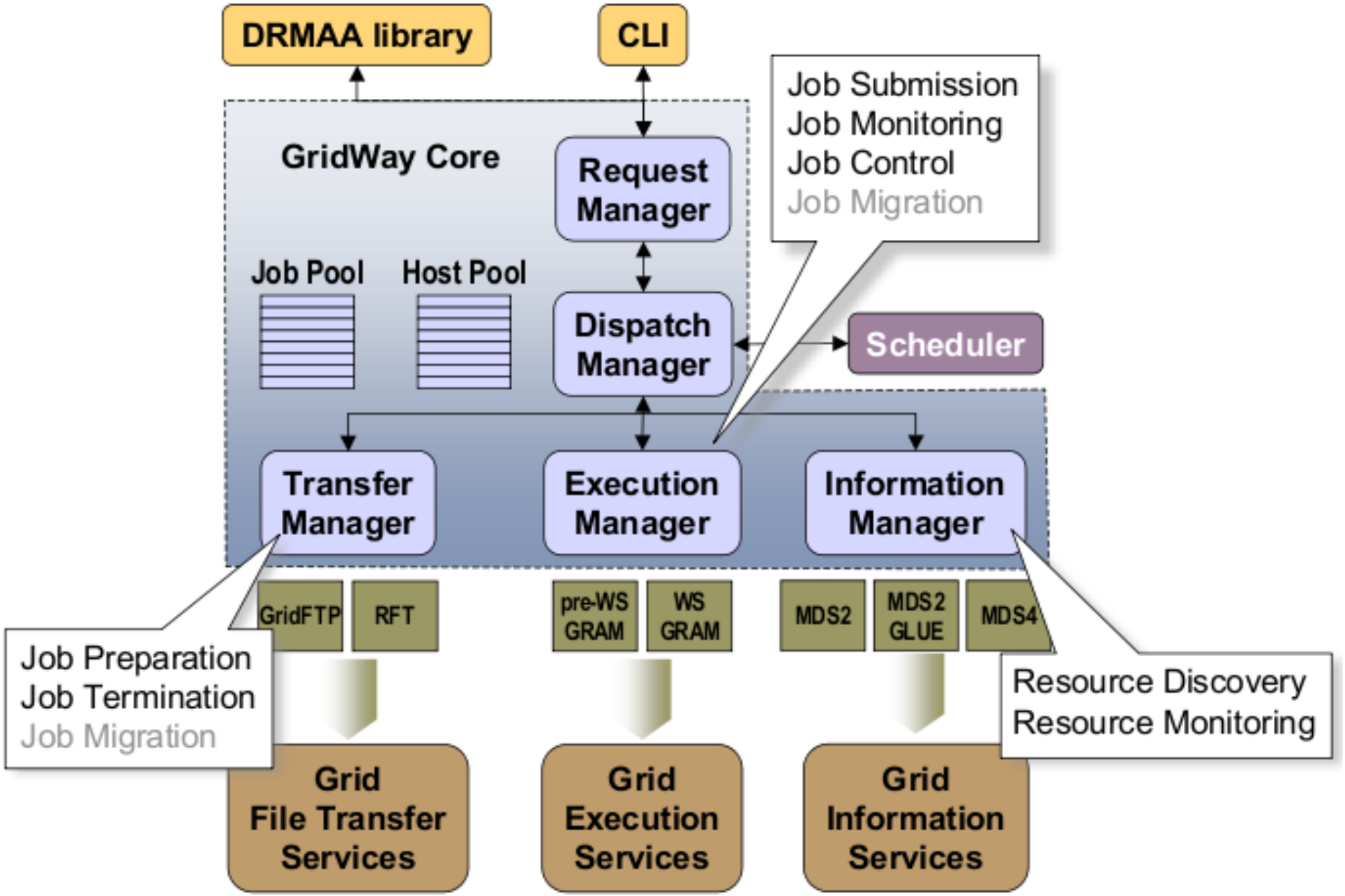
Support for the existing platforms and LRM Systems

- Allocation of grid resources according to management specified policies
- Analysis of trends in resource usage
- Monitoring of user behavior

Familiar CLI and standard APIs

- High Throughput Computing Applications
- Workflows

What is GridWay? Internal architecture



What is GridWay? Features

Workload Management

- Advanced (Grid-specific) scheduling policies
- Fault detection & recovery
- Accounting
- Array jobs and DAG workflows

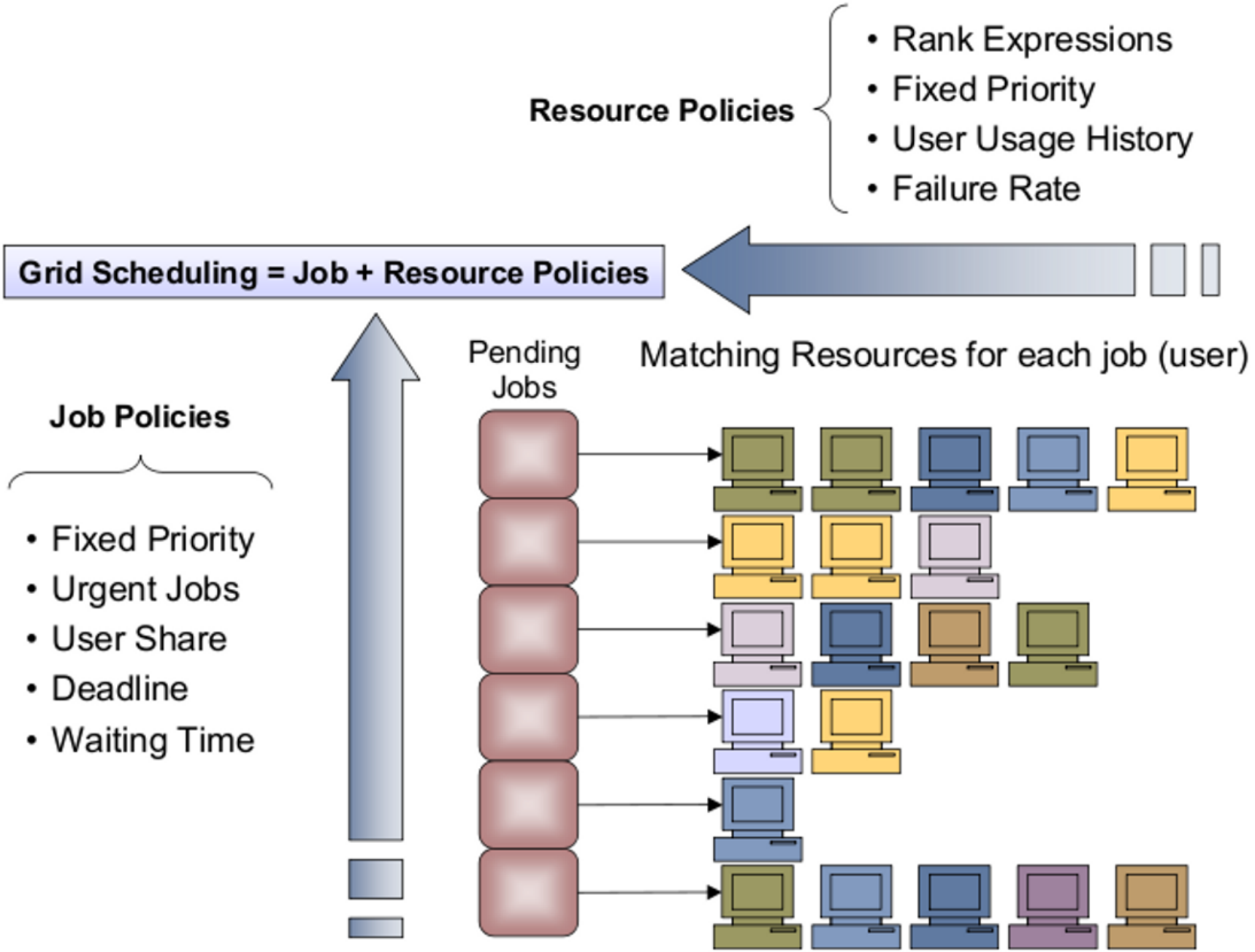
User Interface

- OGF standards: JSDL & DRMAA (C and JAVA)
- Analysis of trends in resource usage
- Command line interface, similar to local LRM Systems

Integration

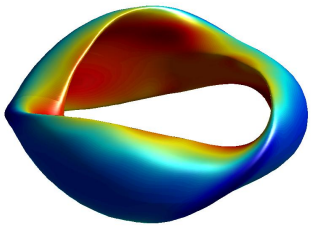
- Straightforward deployment as new services are not required
- Interoperability between different infrastructures

What is GridWay? Job scheduling

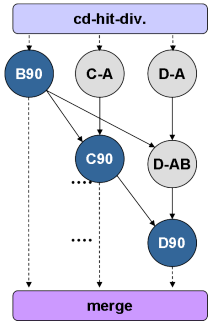


What is GridWay? Application porting

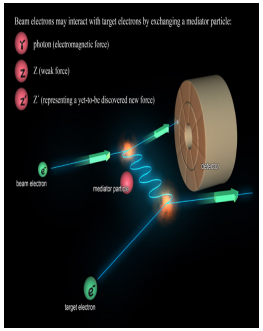
Different scientific-domain



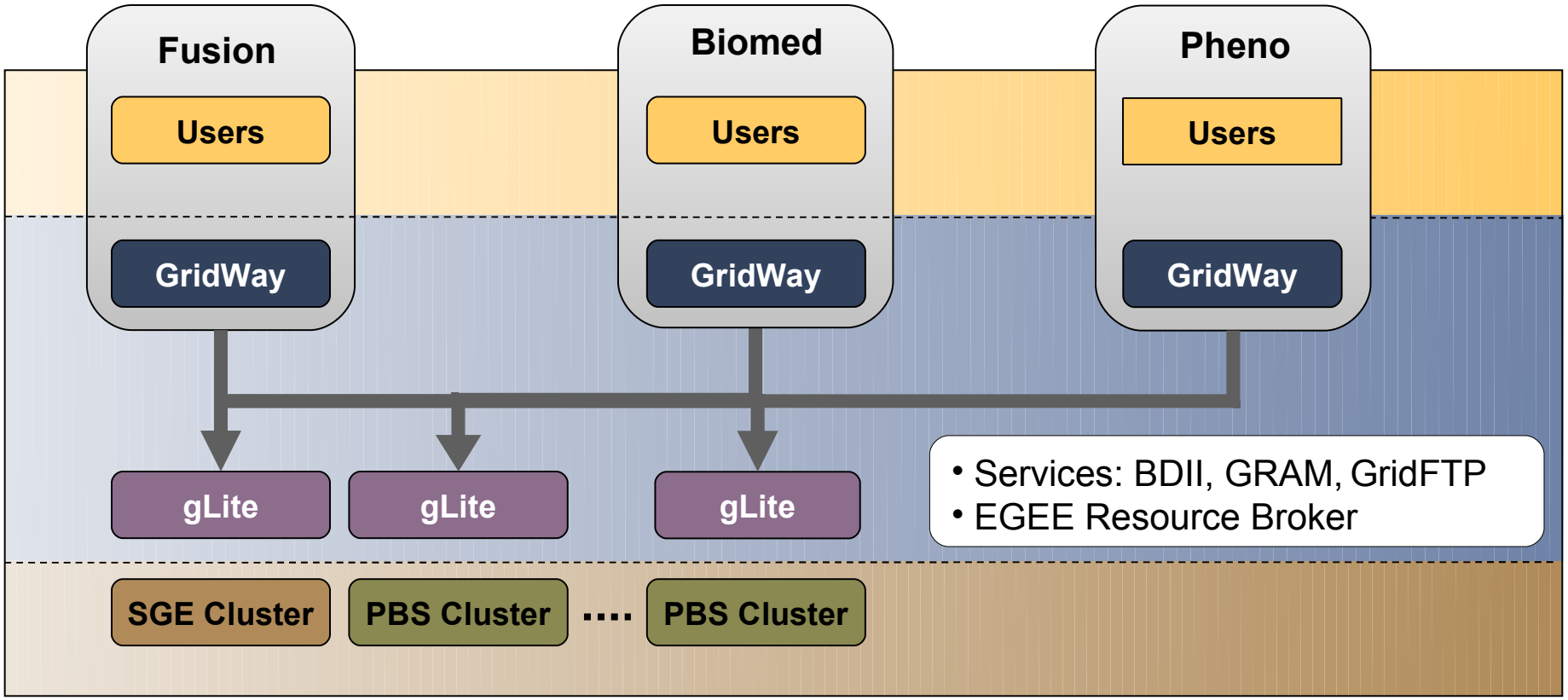
Massive **Ra**y **Tra**cing



CD-HIT workflow



AITALC code
Bhabha



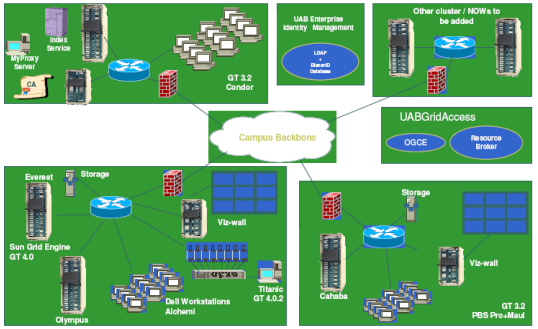
What is GridWay? Infrastructure examples

European Space Astronomy Center

- Data Analysis from space missions (DRMAA)
- Site-level meta-scheduler
- Several clusters

UABGrid, University of Alabama at Birmingham

- Bioinformatics applications
- Campus-level meta-scheduler
- 3 resources (PBS, SGE and Condor)

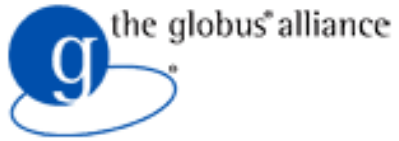


AstroGrid-D, German Astronomy Community Grid

- Collaborative management of supercomputing resources & astronomy-specific resources
- Grid-level meta-scheduler (GRAM interface)
- 22 resources @ 5 sites, 800 CPUs

Globus established as open source project

- founded in 1996
- became Globus Alliance in 2005



  - based in Apache Jakarta model
  - governance by merit (meritocracy)
  - infrastructure configured for global community use

Globus development principles: Guidelines

1) Roles and Responsibilities: Users/Contributors/Commiters/Chair

2) Communication: between users and developers via mailing lists

3) Decision Making: do-ocracy, action items (plans) and voting

4) Source Repositories: CVS, Licensing, patches...

5) Globus Management Committee Bylaws: Defines the roles and responsibilities of the Globus Management Committee.

6) Globus Philosophy: 6 principles: Collaborative, commercial-friendly, high-quality, respect, standards, security.

7) The Globus Incubator: How to join the Globus Alliance

8) Other Globus Entities: Conferences, Security, PR, License committees

9) Globus Intellectual Property: Describes who controls intellectual property

10) About these Guidelines: Relationship between original Apache Jakarta and Globus guidelines

Incubation: GridWay history

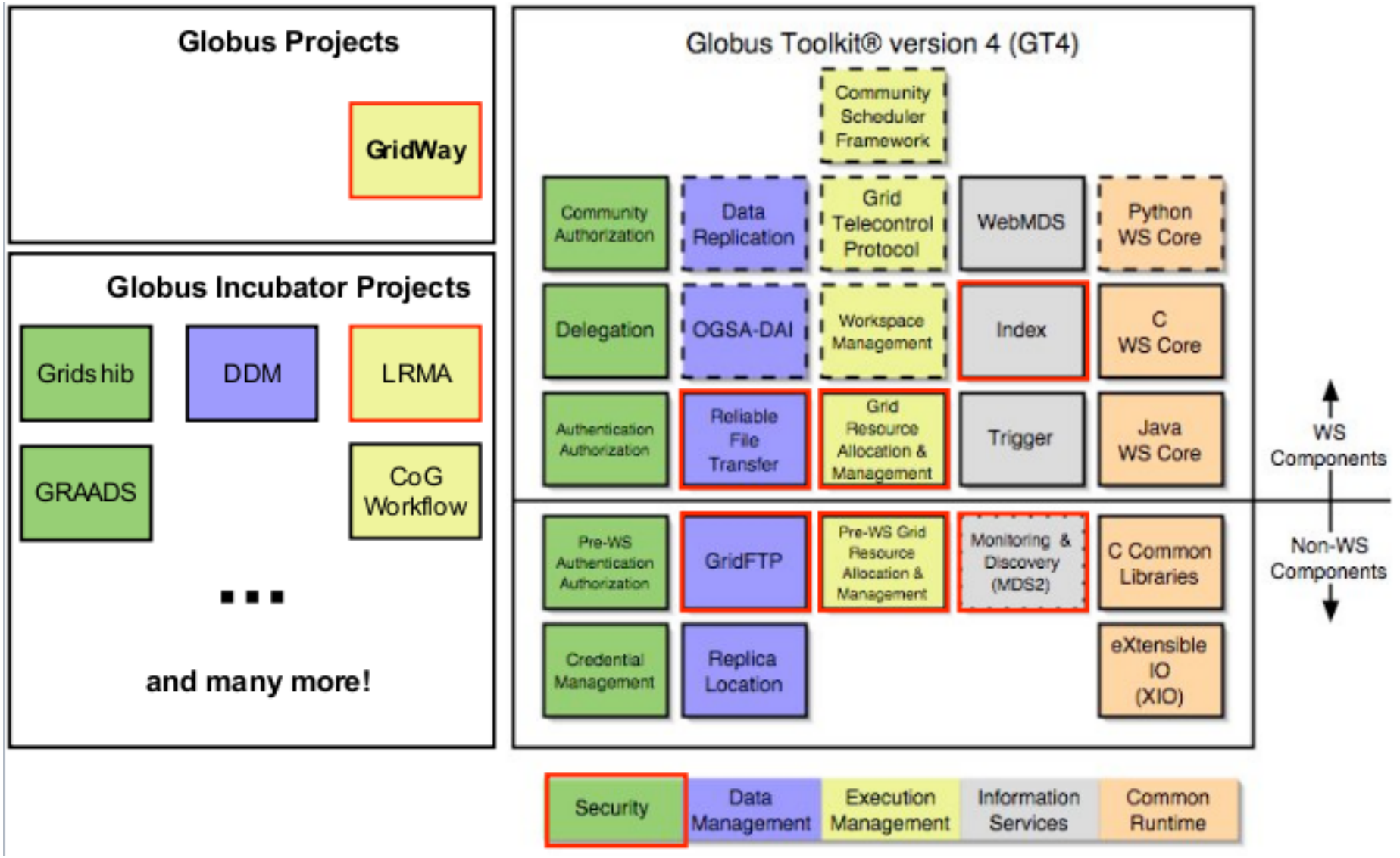
- Started in **2002**, as a research-only effort
- First open source release (v4.0) in **January 2005** (Apache license v2.0)
- In **2006 started the INCUBATION**
- **Ended in January 2007** (the **1st** incubator to become a Globus project)
- In June **2007** GridWay became part of the **Globus Toolkit**
- Since January 2005, more than 2000 downloads from 107 different countries, ~25% are private companies and ~75% are universities and research centers.

Incubation: GridWay modifications

- **Guidelines were adopted**
  - Commiter board: I.M.Llorente (chair), R.S.Montero and E.Huedo
  - Open source philosophy and Globus rules
- **Infrastructure** has been used
  - Globus Mailing lists: gridway-user, gridway-announcements, gridway-dev, gridway-commit
  - Globus CVS repository for source code
  - Bugzilla for bug report and feature requests. http://bugzilla.globus.org
  - GridWay versions from 5.0.1 till 5.2 (full project)

- Effort was **rewarded** as full Globus project by January 2007
- One of the **three** Execution Projects altogether with **GRAM** and **MPICH-G2** in Globus Toolkit 4
- Simplification has occurred in new Globus version 5.0 (talk by S. Tuecke)

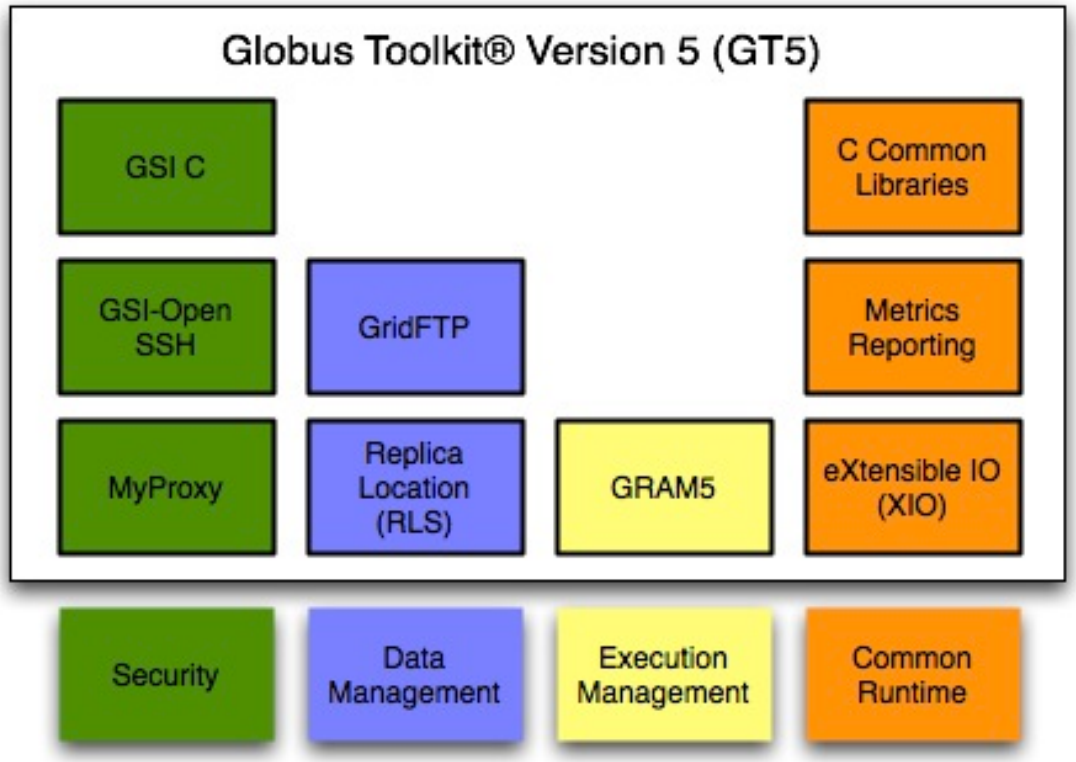
Relationship to other Globus Projects



Conclusions

- Globus and GridWay were independent projects focussing on grid computing
- GridWay became a successful Globus project in 2007 after incubation:
  - Synergies were compatible
  - GridWay adopted Globus rules
  - GridWay profitted from Globus infrastructure
- Process extremely positive, providing a richer toolkit to a **broader** user community

Globus 5 (Jan 2010) has been much simplified:



Initiative for Globus in Europe will have a deep impact in european middleware infrastructures

- The GridWay Metaschedular is a **mature** project:

  1) Globus GridWay is a packaging flavour

  2) Created **own** bug portal, repositories and mailing lists (more dynamic interaction).

  3) Globus infrastructure used for Globus issues.

1) The GridWay Metaschedular is a **mature** project:

   1) **Porting** to Globus Toolkit 5 still to be done

   2) Interaction with **clouds** is on the roadmap

UNIVERSIDAD COMPLUTENSE MADRID

dsa-research.org

Thank you for your attention!