Application Development on Grids

Gergely Sipos
sipos@sztaki.hu

MTA SZTAKI
Hungarian Academy of Sciences

Acknowledgements

- **Slides made by Mike Mineter for EGEE Training and User Induction:**
  - http://www.egee.nesc.ac.uk

- **Many discussions with EGEE colleagues, in particular**
  - Emidio Giorgio, University of Catania and INFN, Italy
  - Richard Hopkins & Guy Warner, TOE, Scotland
  - Gergely Sipos, SZTAKI, Hungary

Definition

What is a grid application?

Software that interacts with grid services to achieve requirements that are specific to a particular VO or user.

Goals of this talk

- **A few reminders to ease us into the course**
- **This talk maps the landscape – a high-level view of application development in Grids**
  - Practicals will explore specific features in that landscape

Contents

- **Part 1:**
  - Review of concepts: grids
  - Types of Grid applications
  - Challenges to researchers who write applications
- **Part 2:**
  - Some ideas to have in mind

Grids: a foundation for e-Research

- Enabling a whole-system approach
- Collaborative research / engineering / public service …

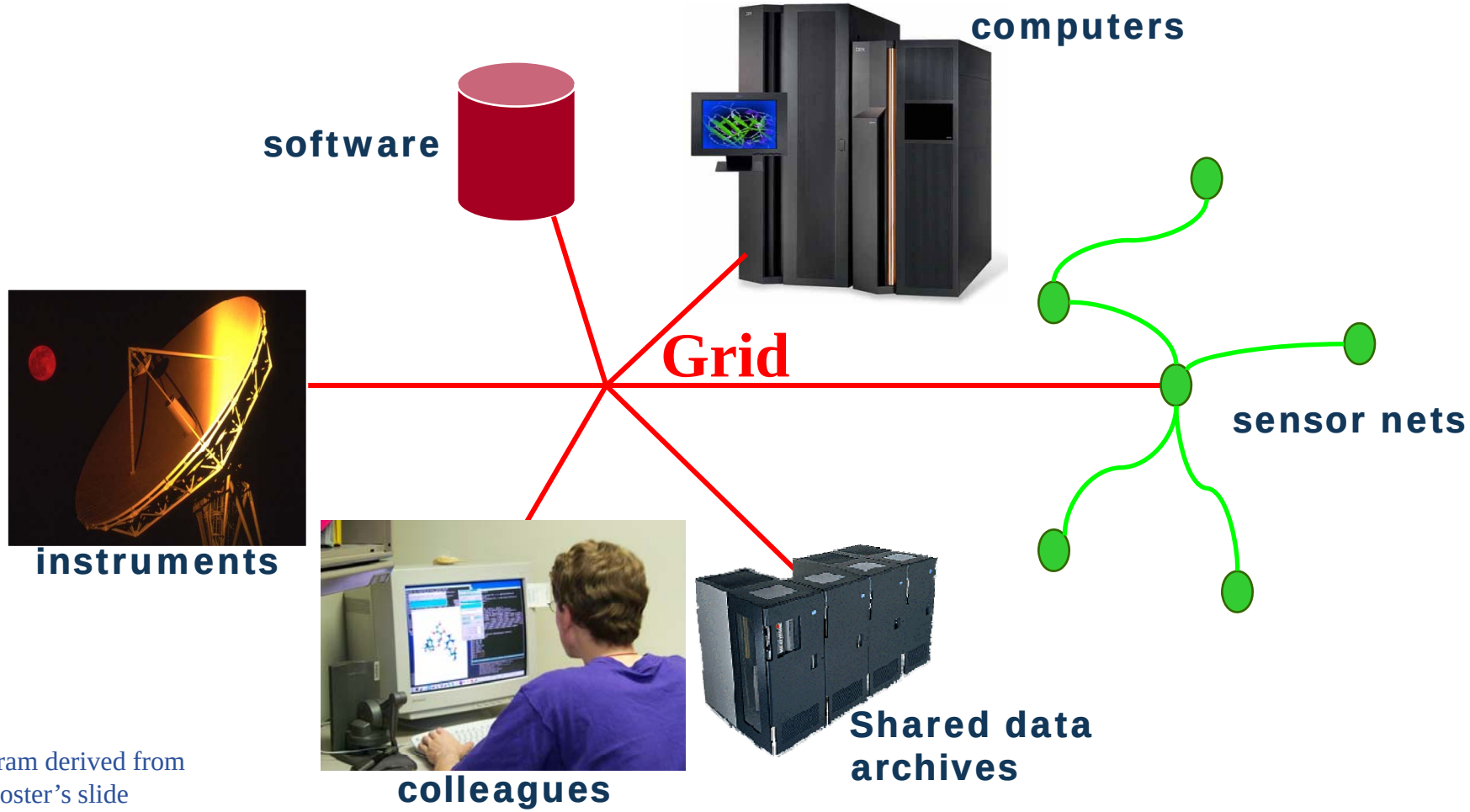


Diagram derived from Ian Foster's slide

e-Infrastructure: crossing boundaries

Grids: across administrative domains

Semantics, ontologies: across disciplines

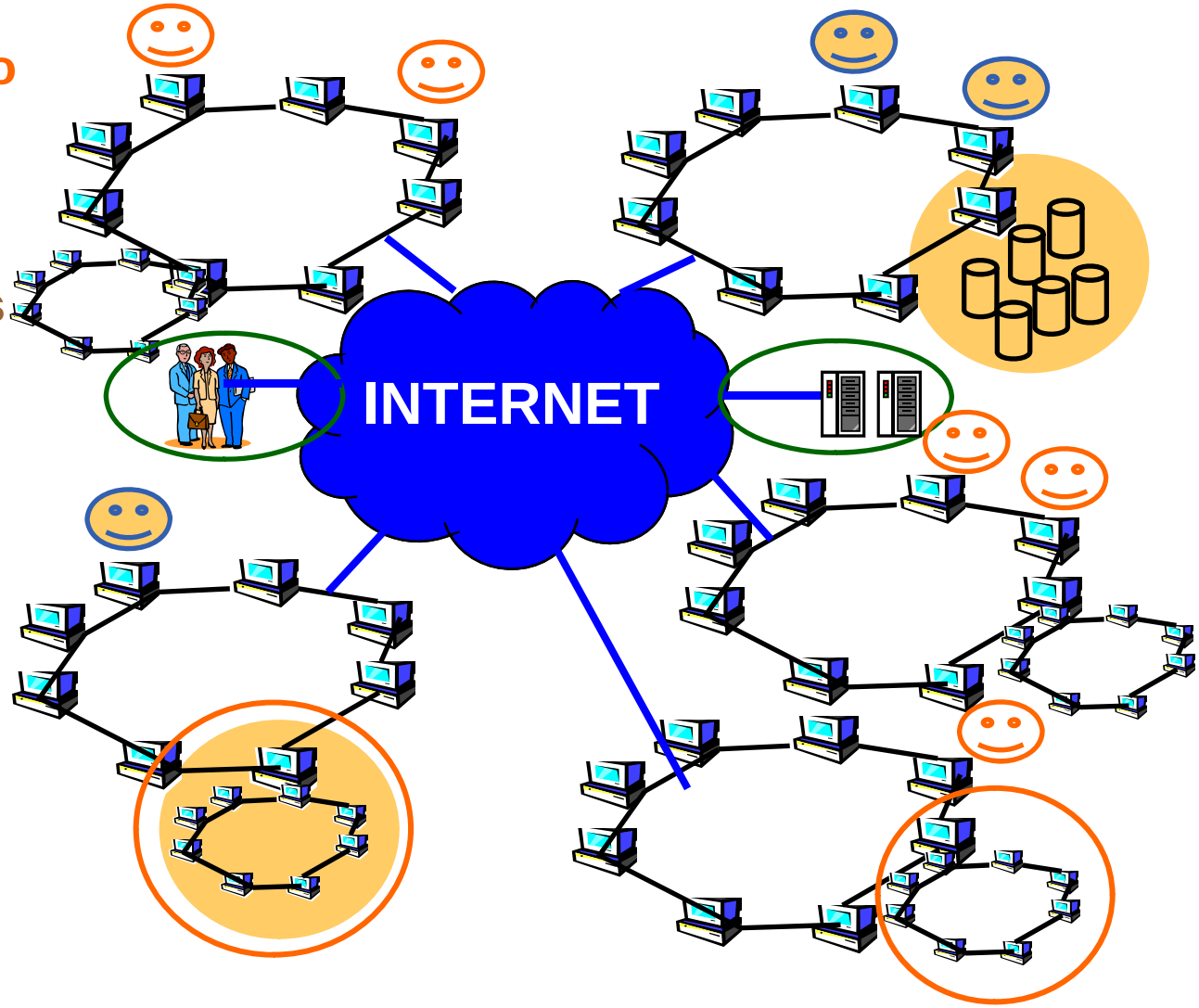
Resource-orchestration

Networks: across geographical distance

Storage, (“curation”): across time

Typical current grid

- **Virtual organisations negotiate with sites to agree access to resources**
- **Grid middleware runs on each shared resource to provide**
  - Data services
  - Computation services
  - Single sign-on
- **Distributed services (both people and middleware) enable the grid**



Where are we now? –users' view

Networks

Web

Grids

Arts
Humanities
e-Soc-Sci

Sciences,
engineering

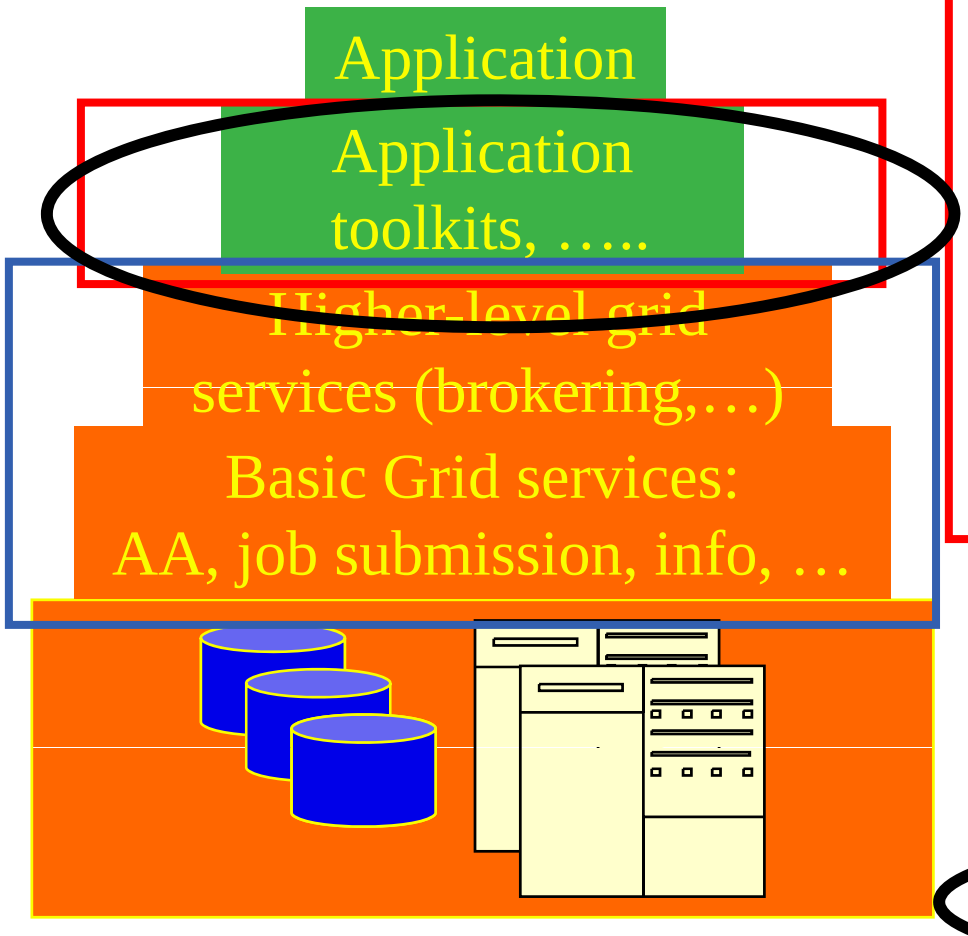
Early production grids:
International - EGEE

Types of use:

Service-oriented, workflow, "legacy" data

High throughput, new data

The vital layer



Where computer science meets the application communities!

VO-specific developments built on higher-level tools and core services

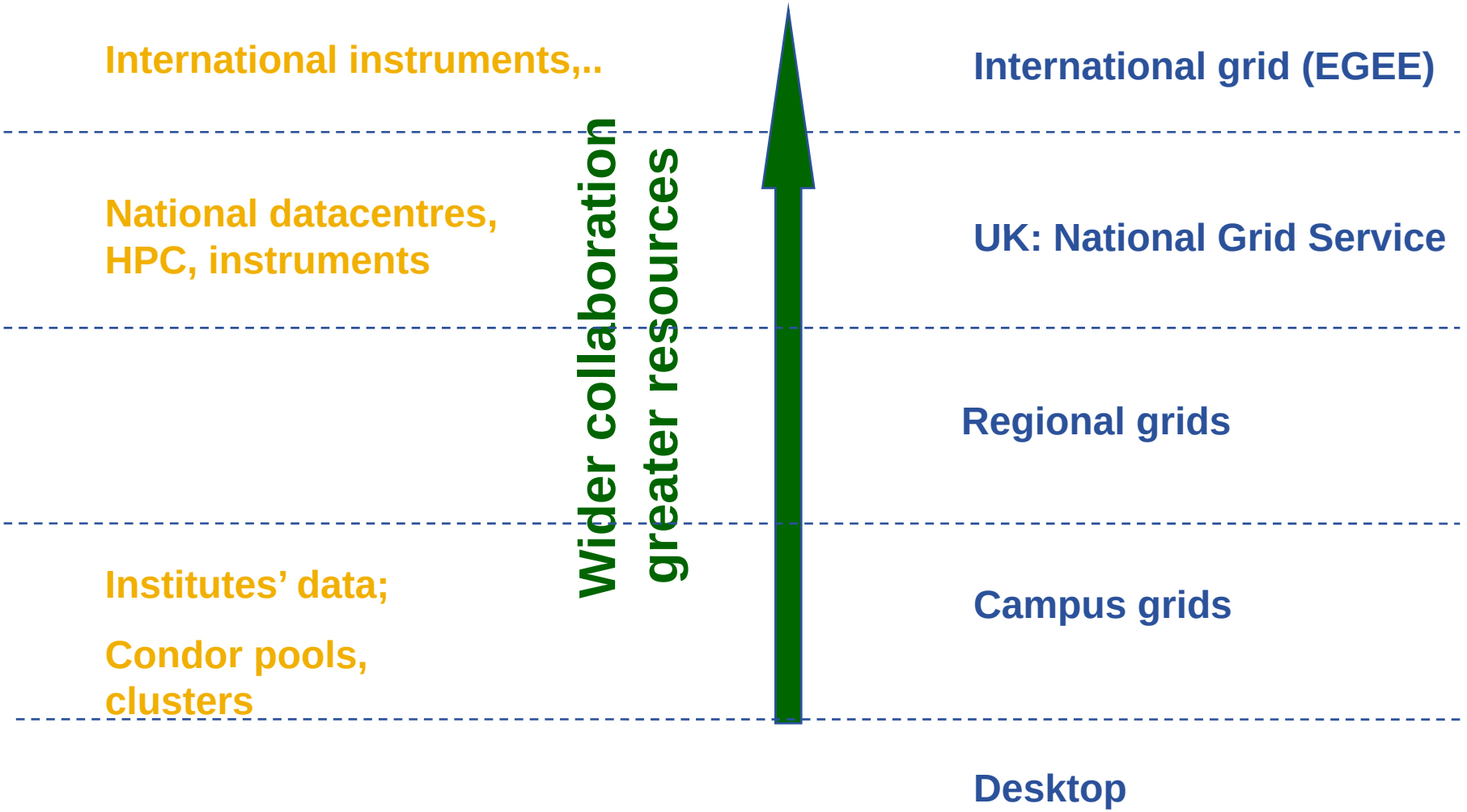
Makes Grid services useable by non-specialists

Grids provide the compute and data storage resources

Production grids provide these core services.

Focus of this course

The many scales of grids



Little interoperability across these scales of grids – yet.

Types of grid applications

Complexities of grid applications

1. **Simple jobs – routine run in batch mode**
   - gLite command line clients (WMS tricks and tips today)
   - gLite APIs (gLite APIs tomorrow)
   - advanced submission frameworks (Ganga today)
2. **Job invokes grid services**
   - To read & write files on grid storage (GFAL API tomorrow)
   - Monitoring (R-GMA API tomorrow)
   - For outbound connectivity – interactive jobs
   - To manage metadata (AMGA today)
   - …
3. **Complex jobs**
   - An environment controls multiple jobs and files on users' behalf
     - High-level services
     - Portals with workflow and parametric sweep support (P-GRADE tomorrow)
     - Software written for the VO (or for the user)
     - …

Invocation of applications

- **From the UI (User's Interface to a grid)**
  - Command Line Interfaces / Scripts
  - Higher level tools
- **From desktop applications**
  - Use Grids without awareness of them!
  - APIs (write your own client)
- **From portals**
  - For recurring tasks: "core grid services" as well as application layer
  - Accessible from any browser
  - Tailored to applications or to community needs

Characteristics of VOs

- **What is being shared?**
  - resources of storage and/or compute cycles (typically regional, national grids: SEE-GRID, VOCE, …)
  - software and/or data (biomed VO, fusion VO, …)
  - Both (comp. chemist members of VOCE)

- **Distinct groups of developers and of users?**
  - Some VOs have distinct groups of developers and users…
    - Biomedical applications used by clinicians,….
  - …. Some don't
    - Physics application developers who share data but write own analyses
  - Effect: need to
    - hide complexity from some VOs
    - expose functionality to other VOs

Different Goals for App. Development

- **I need resources for my research**
  - I need richer functionality
    - MPI, parametric sweeps,…
    - Data and compute services together…
    - How to parallelize existing algorithms?
    - How to write/generate parallel code?
- **I provide an application for (y)our research**
  - How!?
    - Pre-install executables ?
    - Hosting environment?
    - Share data
    - Use it via portal?
- ***We* provide applications for (y)our research**
  - Also need:
    - Coordination of development
    - Standards
    - …

Engineering challenges increasing

Challenges to researchers who write grid applications

Challenges

- **Research software is often**
  - Created for one user: the developer
  - Familiarity makes it useable
  - Short-term goals: Used until papers are written and then discarded
- **Grid applications are often used**
  - by a VO
  - Without support from developer
  - In new contexts and workflows
- **Grid application developers are**
  - In a research environment
  - Yet their s/w must have:
    - Stability
    - Documentation
    - Useability
    - Extendability
  - i.e. Production quality

Need expertise in:

- **software engineering**
- **application domain**
- **grid computing**

Consequences

- **Team work!**
- **Engaged in world-wide initiatives – reuse, don't make your own! Cross disciplines for solutions.**
- **From research to production software: ~5 times the effort.**
  - "80% of the time for last 10% of the functionality & reliability"
- **Standardisation is key**
  - For re-use, for dynamic configuration of services,..
  - Both for middleware and domain specific (e.g. GEON)
- **Need to follow a deliberate development process**
  - Waterfall? Rapid prototyping?
  - Requirements engineering, design, implementation, validation, deployment
  - Engaged with the user community

Summary – part 1

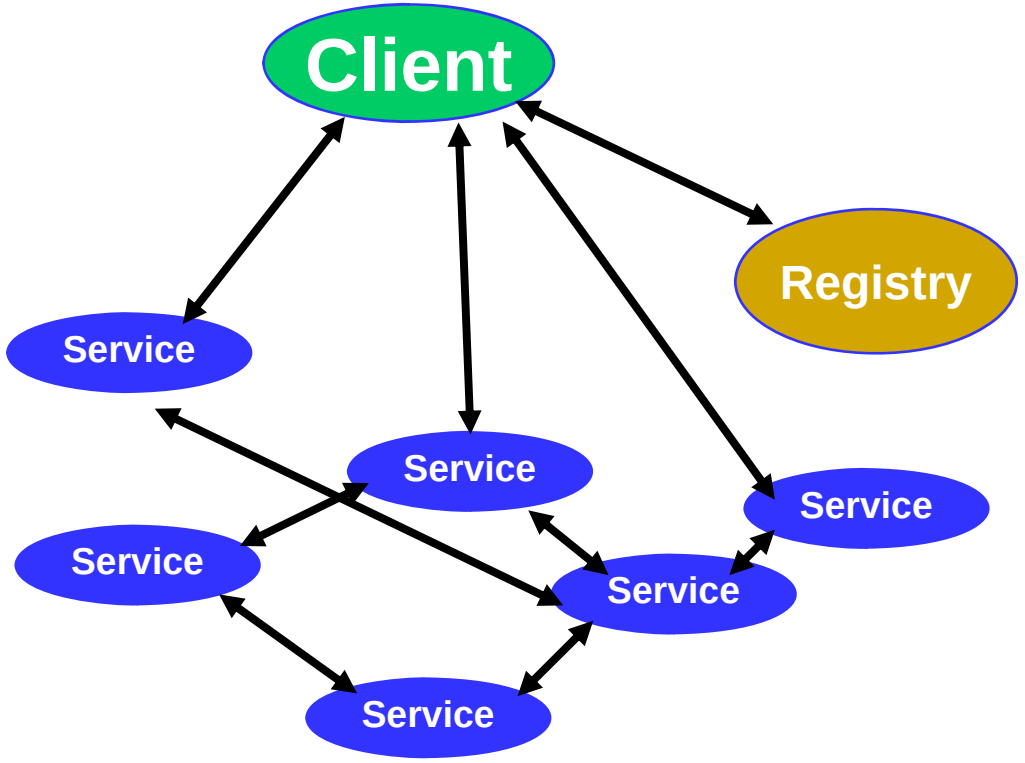
- **“Ask not what ‘the Grid’ can do for you, but what you can do in a Virtual Organisation”**

Part 2: Some ideas to have in mind

Move to service oriented grids

Service orientation – software components that are…

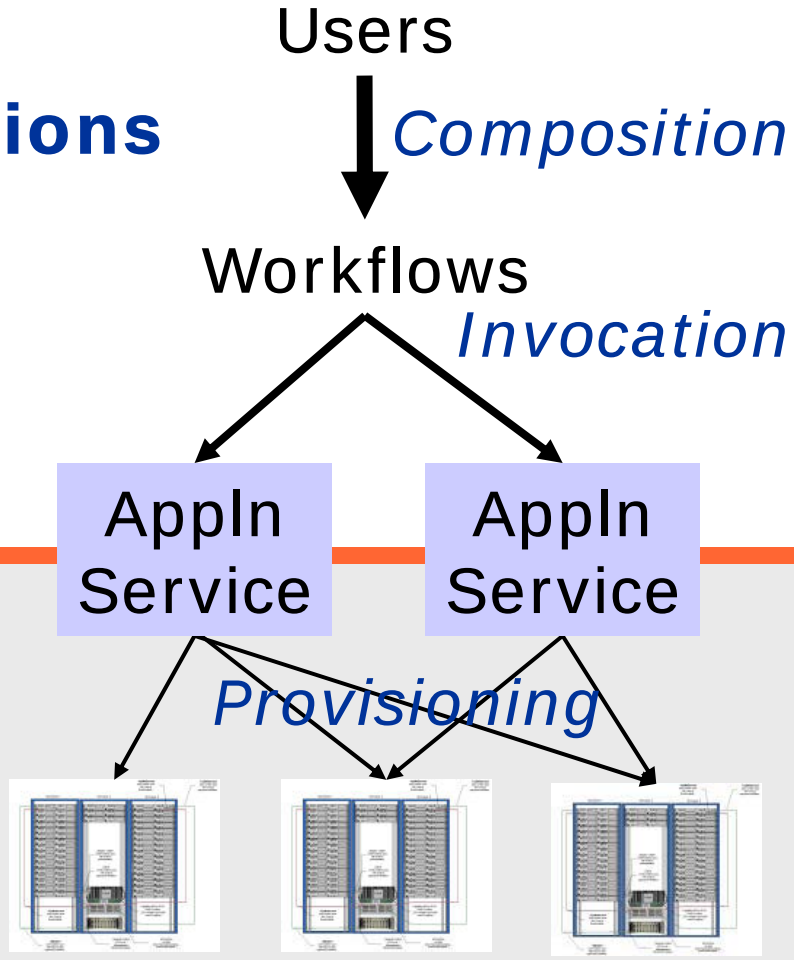
- Accessible across a network
- Loosely coupled, defined by the messages they receive / send
- Interoperable: each service has a description that is accessible and can be used to create software to invoke that service
- Based on standards (for which tools do / could exist)
- Developed in anticipation of new uses



the globus alliance
www.globus.org

Service-Oriented Systems: The Role of Grid Infrastructure

- Service-oriented **applications**
  - Wrap applications as services
  - Compose applications into workflows
- Service-oriented **Grid infrastructure**
  - Provision physical resources to support application workloads



"The Many Faces of IT as Service", ACM Queue, Foster, Tuecke, 2005

The changing face of research

- **Moving toward "utility view" – computation and data services provided by a grid**
  - Applications that can be instantiated on grid resources
- **Effect: people work in their specialisms….**
  - Researcher does research!
  - Service providers provide services!
  - Resource managers manage resources!

- **Further information: Ian Foster's talk at ISSGC 2006**
  - Go to the ICEAGE interface to the Digital Library
    - http://baillie.lib.ed.ac.uk/
  - Search on "Service-Oriented Science"

Summary

- **Grid application development demands more than routine researcher's software engineering**
- **In this course:**
  - Scripting with gLite services
  - Methods in gLite for:
    - Metadata management, application monitoring
  - Introduction to gLite APIs
  - Two prominent examples of higher level tools: Ganga, P-Grade
    - Workflow, parametric jobs, interoperability across grids, GUI
- **"service orientation" is coming**
  - It is more than "the way to build grids"…
  - Services that can be orchestrated for research
  - It impacts HOW research is done
    - Specialisms of researcher, service provider, resource manager
- **Higher level tools are needed**

Information sources and support groups for Application Developers

- **EGEE application domains and VOs (NA4)**
  - http://egeena4.lal.in2p3.fr
- **Grid Application Support Centre** 

  - www.lpds.sztaki.hu/gasuc
- **Recommended External Software Packages for EGEE Communities (RESPECT)**
  - Coming soon at http://egeena4.lal.in2p3.fr