Grid Trends

Jennifer M. Schopf

Argonne National Laboratory

UK National eScience Centre

Grids and Changing Science

Collaborative → Project focused, globally distributed teams, spanning organizations within and beyond company boundaries

Distributed & Heterogeneous → Each team member/group brings own data, compute, & other resources into the project

Data & Computation Intensive → Access to computing and data resources must be coordinated across the collaboration

Dynamic Research → Science being addressed is changing as larger data sets can be analyzed and access to addional resources is made possible

Infrastructure must adapt to this new reality

Top Ten Recent and Upcoming Improvements for Globus

the globus alliance

- **New This Year**
  - ◆ 1) Performance improvements for GRAM4
  - ◆ 2) Added functionality for GridFTP
  - ◆ 3) Ease of use tools (MDS4 Trigger, PURSE)
  - ◆ 4) New functionality through Incubators
  - ◆ 5) Introduce, Gridway
- **And in the Upcoming Year**
  - ◆ 6) Updated standards
  - ◆ 7) Managed Object Placement Service (MOPS)
  - ◆ 8) Data Placement Service (DPS)
  - ◆ 9) Swift for workflows
  - ◆ 10) Virtual Workspaces
- **"Globus Now and Next"**
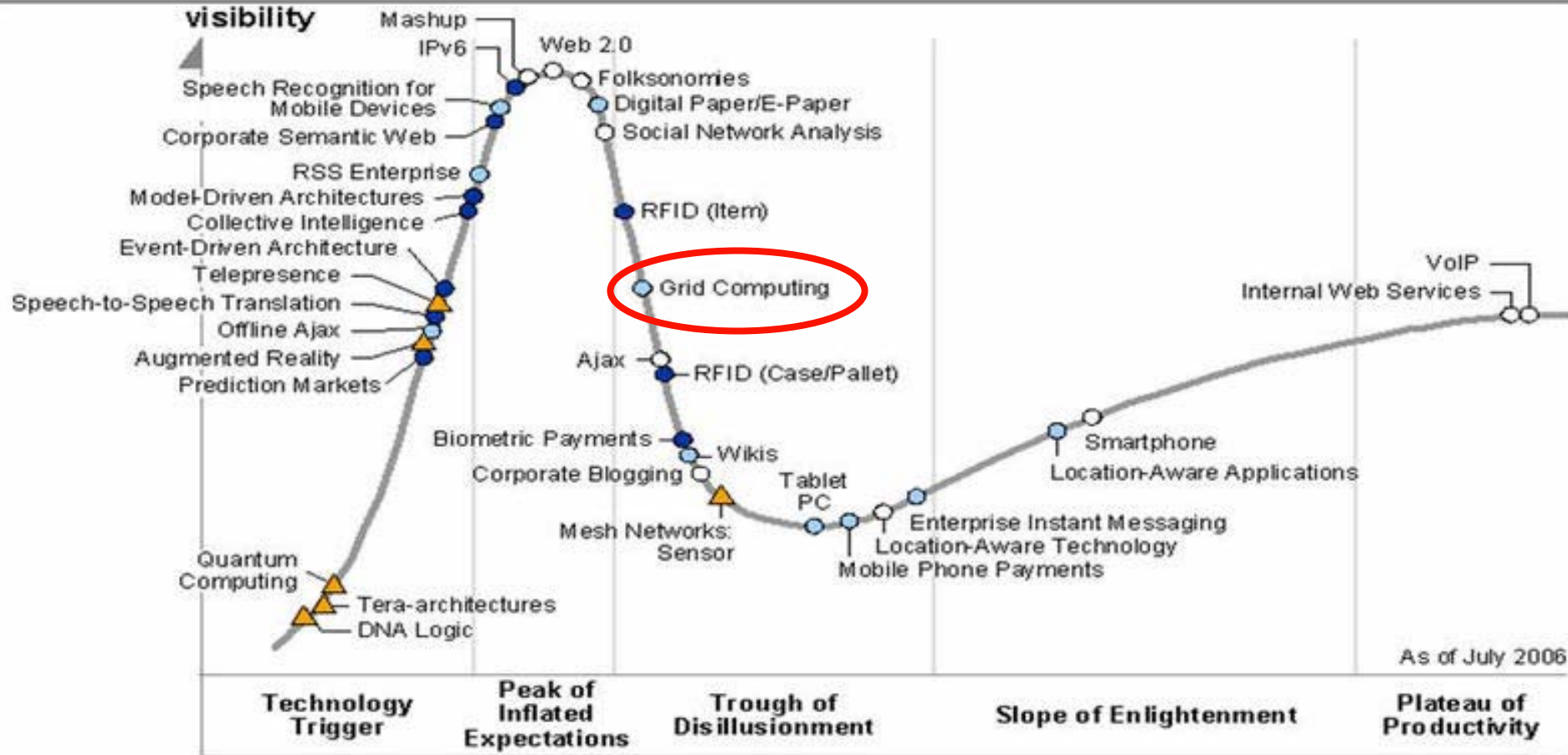 **http://www.mcs.anl.gov/~jms/Talks**

A Broader Question

- What do users want?

- How can we make Grids accessible to end users?

- User Requirements Gathering
  - ◆ 2 UK-centric efforts
  - ◆ July 2004, December 2007
  - ◆ 25+ and 35+ groups

Gartner Hype Cycle

Trough of Disillusionment

- Solutions have been oversold
  - ◆ Globus is a set of building blocks, not an application specific solution
  - ◆ Middleware is often still very complicated
  - ◆ Users aren't finding tools they need
- Everything takes longer than you think it should
- Grids *are* helping research and science, but often not as much as they could

6

Moving to Enlightenment

- Technology
- Social-political
- Application-oriented approaches

- For each topic, I'll walk through an example, talk about current vs future, and give my estimate of when it might happen

Technology

- Virtualization
- Ease of Use
- Security

Virtualization

- Vision of the Grid:
  - Plug in and get the services you need
  - Just like electricity
  - Doesn't matter what resource is supplying it, or where it is, just use the "juice"
- Concrete example, a use might ask..
  - "Run my job, finish by lunch"
  - "Get a data set that has these attributes"
  - "Tell me when that simulation will finish"

Where are we today?

- "Run my job, finish by lunch" becomes
  - ◆ Run my job on this exact machine
  - ◆ With these data files transferred
  - ◆ I think it will take 2 hours, the queues have been slow lately, so I should make sure I send this off by 9am, or earlier if I want to be safe in having results for 2

Where are we today?

- "Get a data set that has these attributes" becomes
    - Given a set of attributes, give me a set of logical file names
    - Given those, map them to physical file names
    - Given physical placements of the file, figure out which one is easiest to access
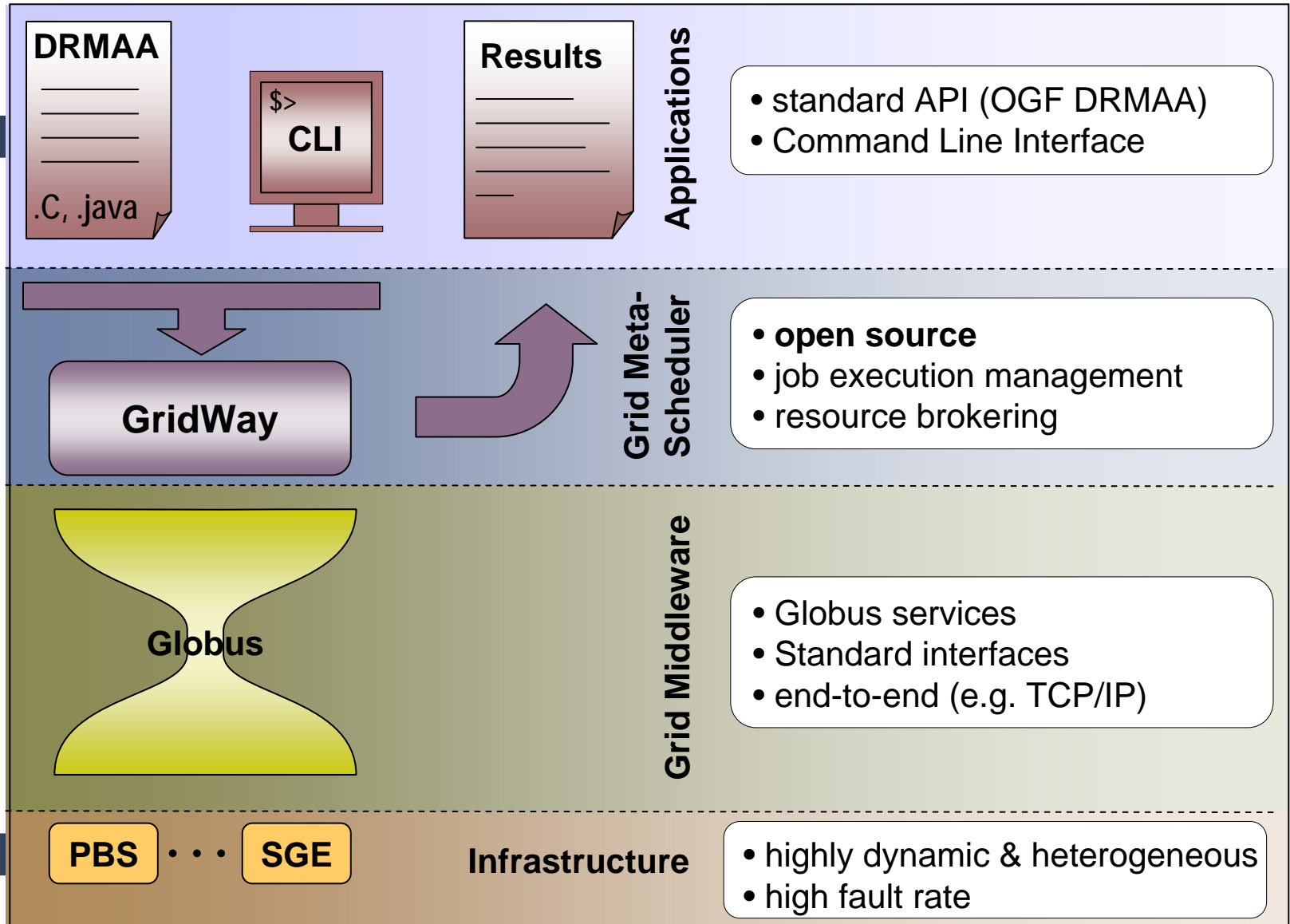   - Copy the file to my machine

Where are we today (cont)

- General agreement we have basic functionality
  - Tell me what this set of resources look like
  - Run this job on that resource
  - Transfer this file
  - Globus (among others) does give these basic building blocks (mostly)
- General agreement general functionality isn't enough by far

Step Forward: GridWay

Application-Infrastructure decoupling

DRMAA

.C, .java

CLI
$>

Results

Applications
- standard API (OGF DRMAA)
- Command Line Interface

GridWay

Grid Meta-Scheduler
- **open source**
- job execution management
- resource brokering

Globus

Grid Middleware
- Globus services
- Standard interfaces
- end-to-end (e.g. TCP/IP)

PBS · · · **SGE**

Infrastructure
- highly dynamic & heterogeneous
- high fault rate

GridWay: http://www.gridway.org

13

Step Forward: CEDPS Data Placement Service

- Tie together several lower level tools
- In addition perform distribution or placement plan generated by higher-level service
- Provide feedback to higher level placement services on the outcome of the placement workflow
- Call on lower-level services to coordinate

- Release 1.0 available Oct 1, 2006
- http://www.cedps.net/wiki/index.php/Data

14
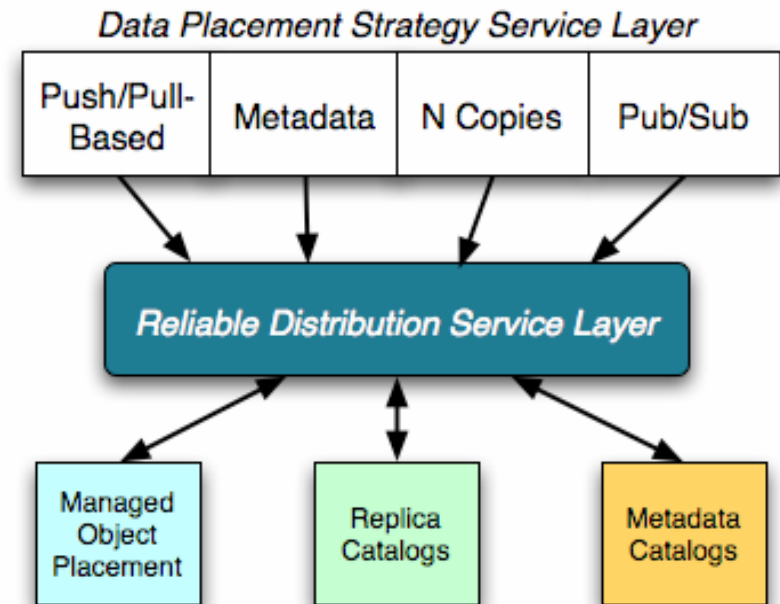
Virtualization is happening.. slowly

- Some higher level tools to move users further from services are being developed

- General concept of service-oriented grids is being accepted

- Service Level Agreements (SLAs) are coming into place

- When will this be resolved?
  - ◆ Tools to tie together functions, very soon
  - ◆ True vision? Not this year...

Ease of Use

- Users will only come when they have decent tools
  - Simple enough for "easy" use
  - Robust enough for "stupid" use
  - Still allow work-arounds for "hard-core" use

- Users are hampered by software that doesn't do what they need it to
- Globus is NOT an end-to-end solution

What is needed

- Closer ties between tool builders and user
  - Tool builders still creating "cool" solutions to problems that don't exist
  - Users still not communicating what they need – or ignoring "not built here" solutions when available
- When will this be resolved?
  - More and more common to see cross-discipline teams
  - First steps – but a long way to travel

Grid Security

- Without security we can't have a Grid

- EVERYTHING needs to be secure-

    - ◆ Who can run on a machine

    - ◆ File transfers

    - ◆ What data does someone have access to (program data, system data)

    - ◆ Who can access which services?

Security vs. Usability

- Users want security but don't want to deal with it
- If security is hard- it won't be used
- Grid security often based on public key infrastructure (PKI), which can be notoriously difficult for users to work with
- Many extras to consider
  - Multiple certificates? Group access? Dynamic policy changes? Scalability?

- Without security no one will really use the Grid

Current work – there's a lot of it!

- Grid-Shibboleth interactions
  - Tying PKI into Shib, which is used on many higher education campuses
- VOMS
  - Community-level group membership system
  - Allows the VO to centrally manage roles
- PURSe
  - And other tools that wrap lower level tools with simple interfaces are becoming more common
- When will this be resolved?
  - This is an area that will ALWAYS have active research and development

Moving to Enlightenment

- Technology
- Social-political
  - Communication
  - Standards for Interoperability
- Application-oriented approaches

Socio-political Issues

- Hardest problems are often not technical ones
    - Multiple administration domains means multiple policies
    - Multiple countries means multiple communication styles
    - Decisions are often made on non-technical basis

Communication Between Software Providers and End Users

- Software providers and end users often communicate quite differently
  - Confusion between the possible and the probable
  - End-users aren't a single voice
  - Software providers often want to talk about the new and shiny, not the true and tested
- Still need broad outreach and evangelizing about what exists – not what MAY exist in the future

What to do?

- Ongoing efforts to continue better communication are needed to build a global community
- Training and lines of communication
- Constructive criticism, reporting of errors, etc – just saying "Globus Is Bad" simply isn't helpful

- When will this be resolved?
  - Better than it used to be – although needs ongoing attention

Need for Standards

- Need for standard APIs and protocols to allow easier
  - Access to data sources
  - Registration of data
  - Archiving tools
- Standards for what information is available
- Standards for what that information means
- Standards for communication of errors
  - This is in part what inspired Globus's move to Web services!

Standards and Interoperablity

- What's the real goal behind standards?
  - Interoperabilty!
- Without standard interfaces, languages, schemas, etc we cannot have multiple implementations that work together
- What do you mean by interoperate?
  - Share data?
  - Share job submissions?
  - Share accounting data?
  - Share accounts?
  - Have a common broker?
  - Have a common software stack?
  - Have common environments or testing?

Where are we

- **Standards are hard**
  - ◆ Need buy in from many communities
- **Grid Interoperablity Now (GIN) effort**
  - ◆ Technology wasn't used by applications
  - ◆ Effort still ongoing
- **When will this be resolved?**
  - ◆ Pair-wise interoperability is happening now, and expanding
  - ◆ *Real* interoperability will happen when users demand it

Moving to Enlightenment

- **Technology**

- **Social-Political**

- **Application-Oriented Approaches**
    - ◆ User Communities
    - ◆ Grids Become Data Grids

User Communities

- Traditionally, we started with the physicists
  - ◆ Hard core users (heroic users)
  - ◆ Large computational problems
  - ◆ Already had strong national and international collaborations
- This is growing and changing as understanding of how the resources can be used to further science and research are better understood
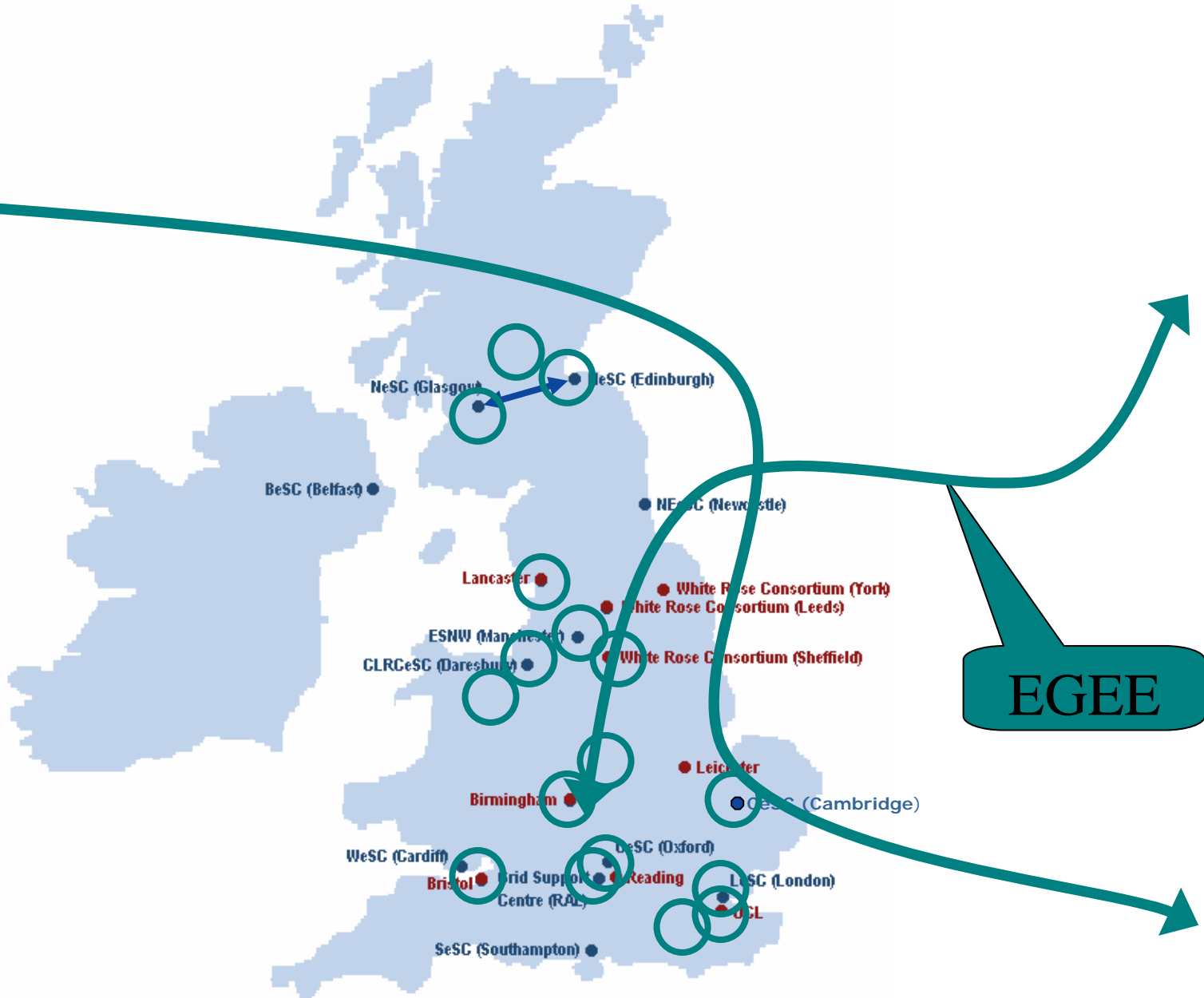
UK eScience Centres

30

High Energy Physicists

the globus alliance

ATLAS
CMS
D0
Star
QCD
Lattice Gri
GridPP

EGEE

NeSC (Glasgow)
NeSC (Edinburgh)
BeSC (Belfast)
NEeSC (Newcastle)
Lancaster
White Rose Consortium (York)
White Rose Consortium (Leeds)
ESNW (Manchester)
White Rose Consortium (Sheffield)
CLRCeSC (Daresbury)
Leicester
Birmingham
CeSC (Cambridge)
WeSC (Cardiff)
CeSC (Oxford)
Bristol
Grid Support
Reading
LeSC (London)
Centre (RAL)
UCL
SeSC (Southampton)

Bio-Medical Community

CancerGrid
eDiamond
myGrid
Integrative
Mouse Atlas

Support Services
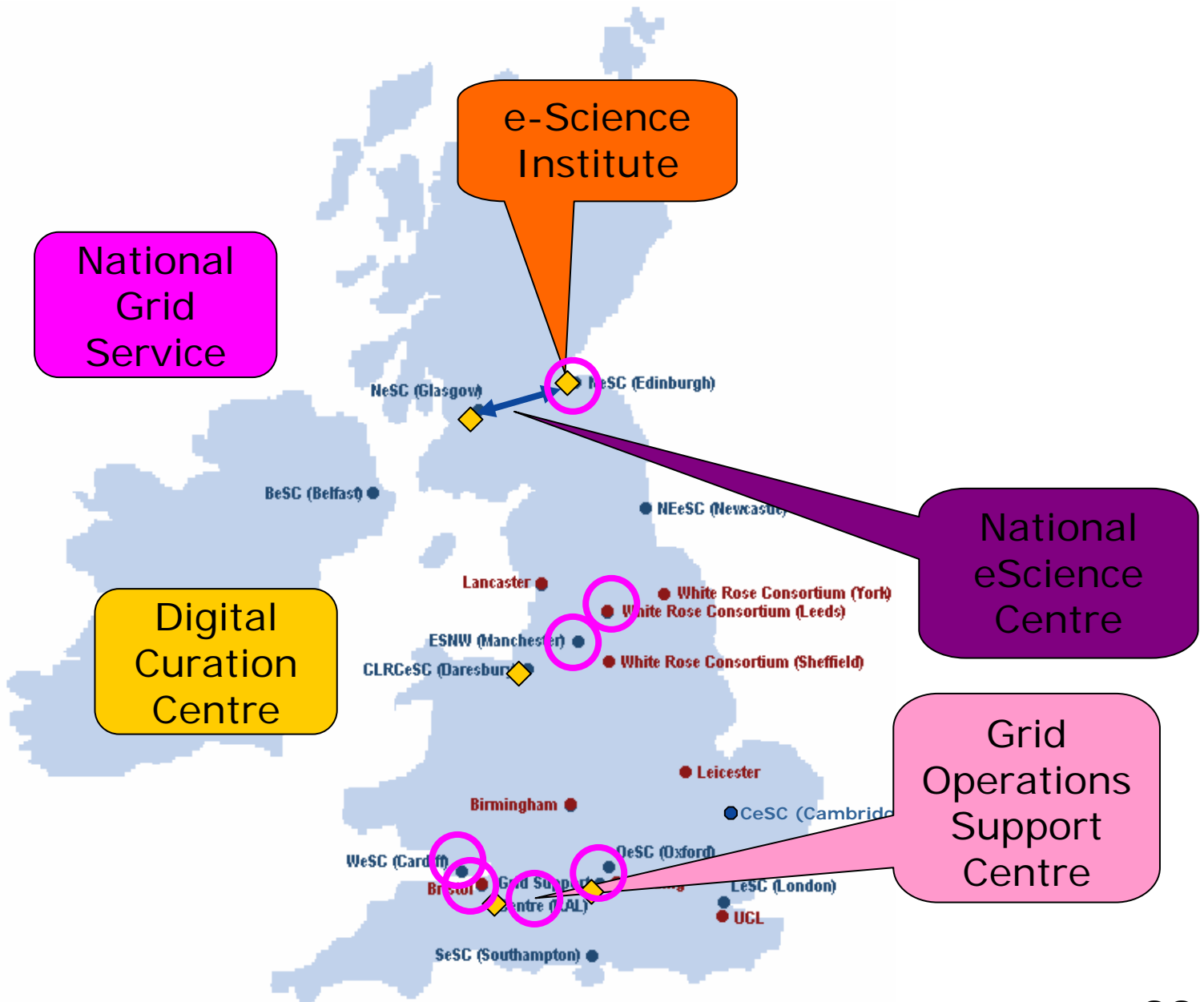
New Application-Focused Centres

National Centre for e-Social Science

National Institute for Environmental e-Science

Arts and Humanities e-Science Support Centre

NeSC (Glasgow)

NeSC (Edinburgh)

BeSC (Belfast)

NEeSC (Newcastle)

Lancaster

White Rose Consortium
White Rose Consortium (Leeds)

ESNW (Manchester)

CLRCeSC (Daresbury)

White Rose Consortium

Birmingham

CeSC (Cambridge)

WeSC (Cardiff)

OeSC (Oxford)

Bristol

Grid Support Centre (RAL)

Reading

London

UCL

SeSC (Southampton)

Changing User Community

- Adapting tools to new users can be challenging

- If you're building tools
  - Talk to users, early and often
- If you're a user-
  - Tell the toolmakers what you like and don't like
  - Be constructive
  - Offer to alpha test

Technology Focus is Changing as Well as User Community

- Originally, application users (scientists) wanted to run big jobs
  - These were the obvious candidates coming from HPC
  - These were the people willing to put up with poor interfaces, buggy software, no support
  - These were some of the people writing the software
- Obvious use case when Grid were first being looked at

...then came the data

- Replicating very large data sets

- Accessing distributed databases

- Understanding data provenance

- Finding files (or parts of files) with certain attributes

Sharing Large-Scale Data

- By far the largest concern of the users we spoke with (Dec 2006)
- How to share data with colleagues
  - Within their project or their wider community
  - Software, results, or other data
- Long-term storage and curation
  - Annotate files with metadata about the contents and provenance
  - Support search and reanalysis at a later date
  - Need for Metadata standards

When will the data problem be resolved?

- Some applications are already doing this quite happily
- Additional work will continue for next several years to
  - ◆ Increase usability
  - ◆ Tie together additional services
  - ◆ Curation issue
- This is where much of the growth of Grids is likely to be

Summary

- **Technology**
  - ◆ Virtualization
  - ◆ Ease of Use
  - ◆ Security
- **Social-political**
  - ◆ Communication
  - ◆ Standards for Interoperablity
- **Application-oriented approaches**
  - ◆ User communities
  - ◆ Grids becoming Data Grids

Summary

- Grids are here, and the problems they solve aren't going away

- Need to concentrate our efforts to move from the Trough of Disillusionment to the Slope of Enlightenment

- Open issues abound – lots of interesting work still to come!
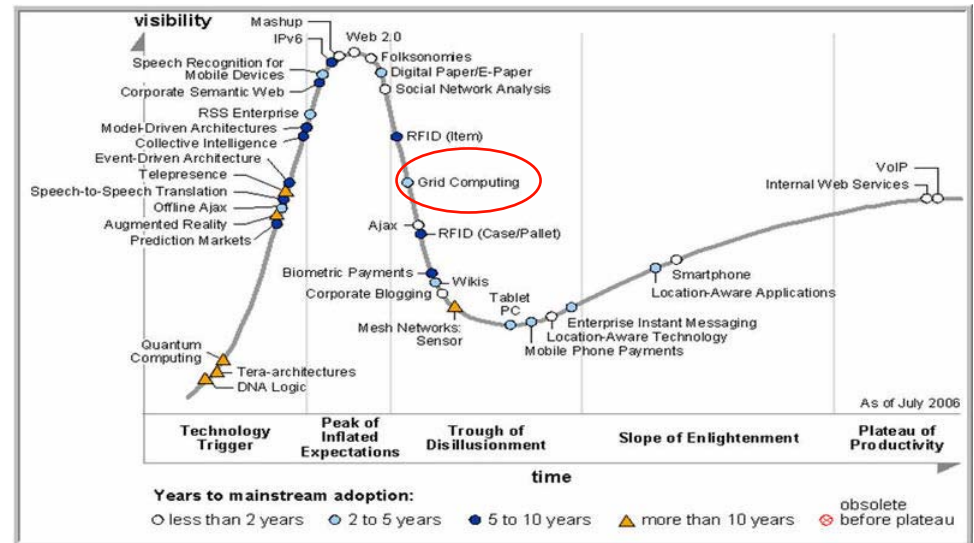
For More Information

Jennifer M. Schopf

- jms@mcs.anl.gov
- www.mcs.anl.gov/~jms

Globus Information:

- http://www.globus.org
- http://dev.globus.org

Talks:

- "Globus Now and Next"
  http://www.mcs.anl.gov/~jms/Talks
- This talk online soon as well