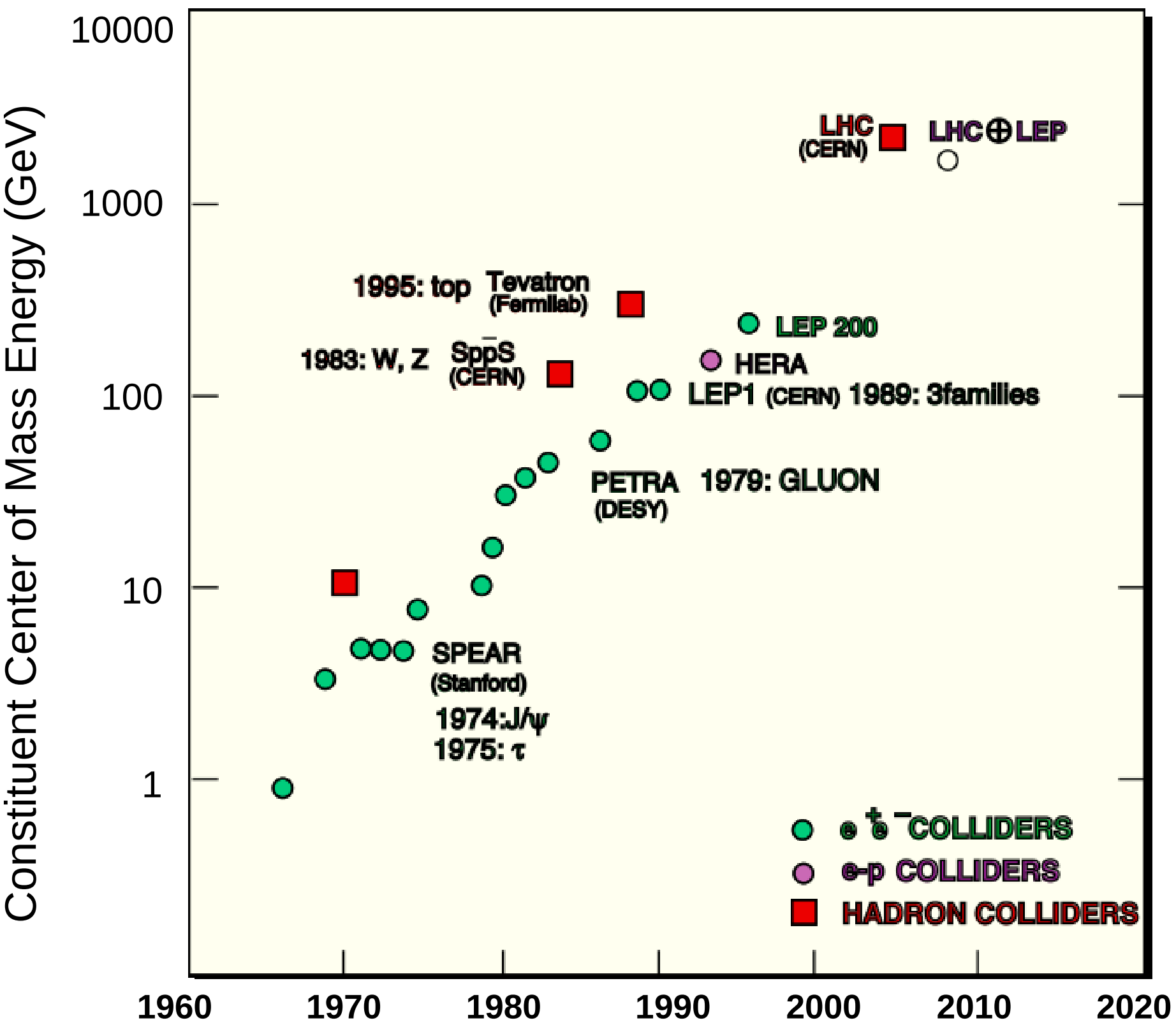
Federating Grid Resources

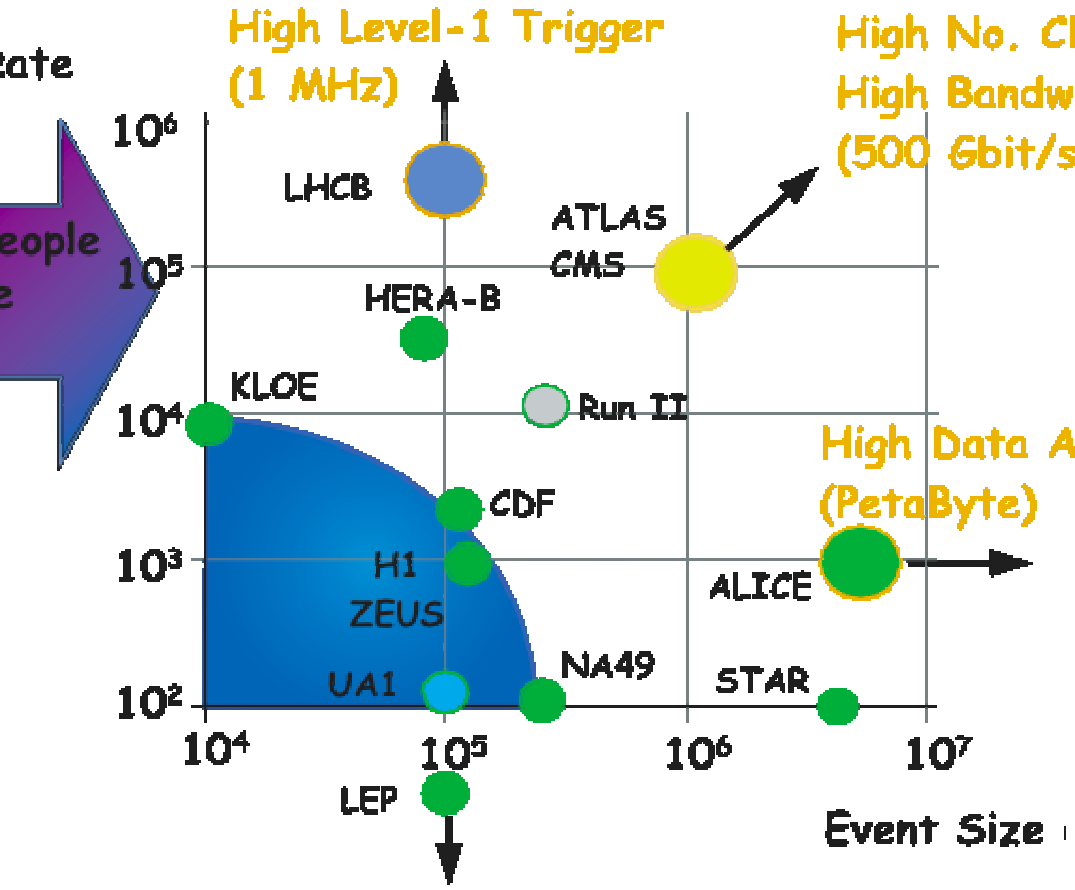
Michael Ernst

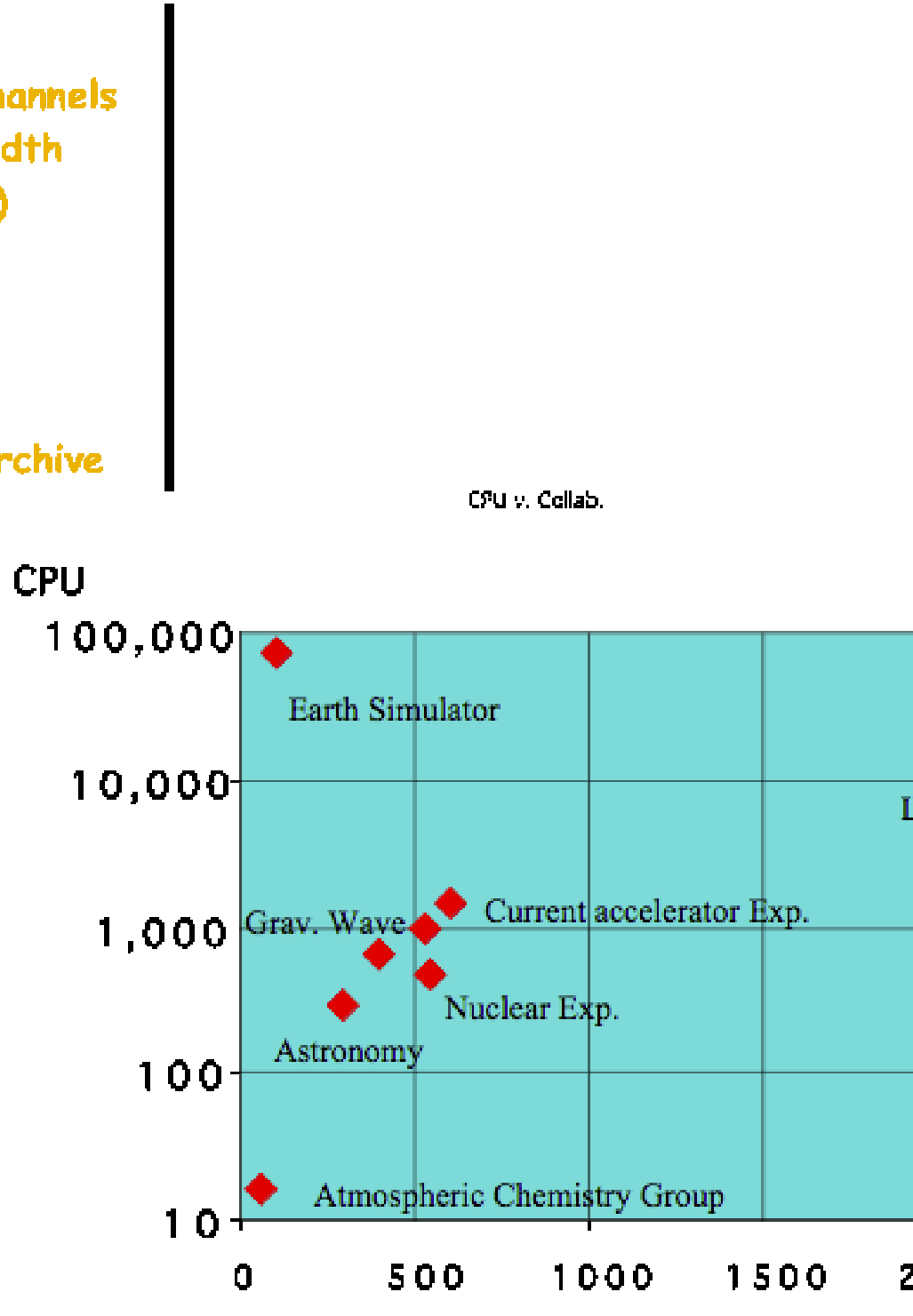
DESY Seminar

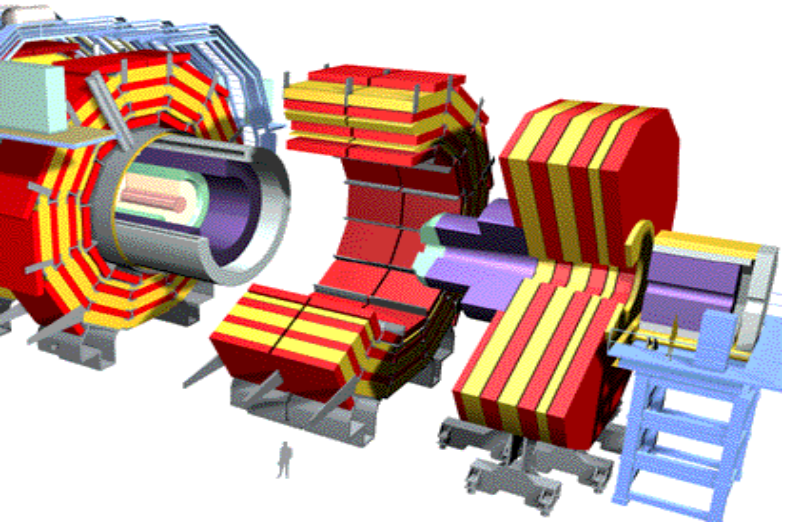
May 3rd 2004


Constituent Center of Mass Energy (GeV)
10000
1000
100
10
1
1960
1970
1980
1990
2000
2010
2020
LHC
(CERN)
LHC ⊕ LEP
1995: top Tevatron
(Fermilab)
LEP 200
1983: W, Z SppS
(CERN)
HERA
LEP1 (CERN) 1989: 3families
PETRA 1979: GLUON
(DESY)
SPEAR
(Stanford)
1974:J/ψ
1975: τ
e+e− COLLIDERS
e-p COLLIDERS
HADRON COLLIDERS

In Data Rate, Data Size, CPU and Number of Scientists







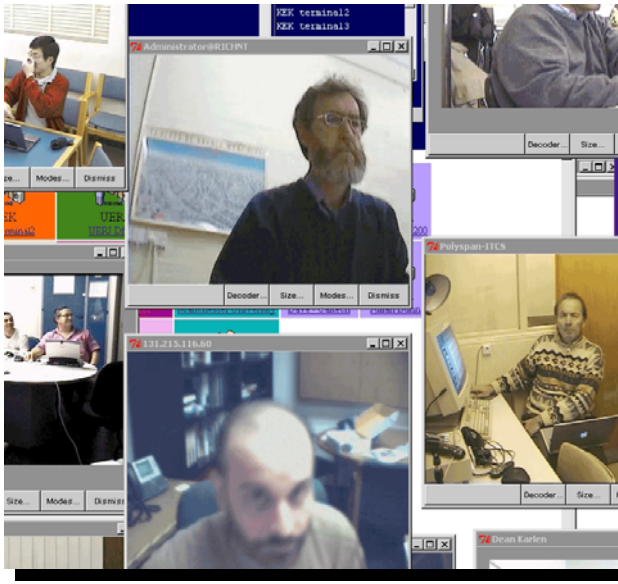
+

e HEP Computing Unprecedented in Scale and Complexity (and Costs)

an Advanced Coherent Global “Information-Infrastructure”

ational and Interdisciplinary Partnerships

wer the Universities to do Research on Physics Data



why we are interested in Grids and enabling Information Techno

global collaboration of thousands of physicists

Provide capabilities to individual physicists and communities of scientists that allow

- To participate as an equal in the research program
- To be fully represented in the Global Experiment Enterprise
- To on-demand receive whatever resources and information they need to explore their science inte respecting the collaboration wide priorities and needs

massive computing, storage, networking resources

ncluding “opportunistic” use of resources that are not owned by a particular experiment!

full access to dauntingly complex “meta-data”

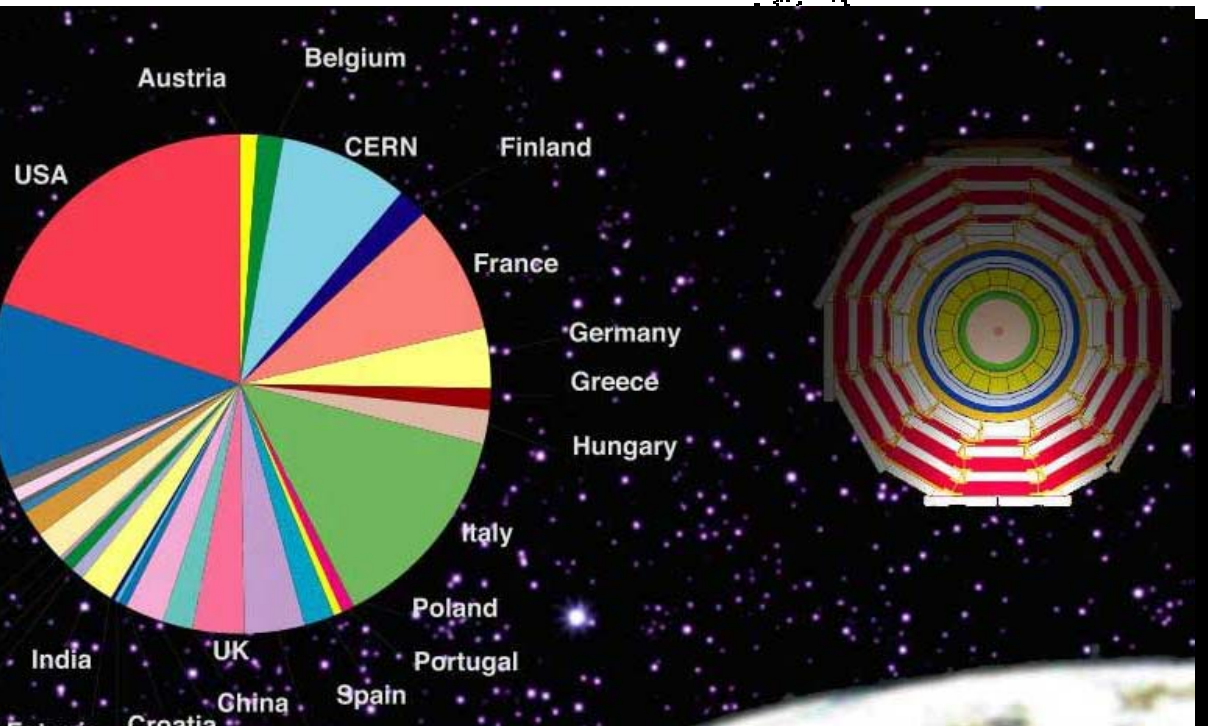
That need to be kept consistent to make sense of the event data

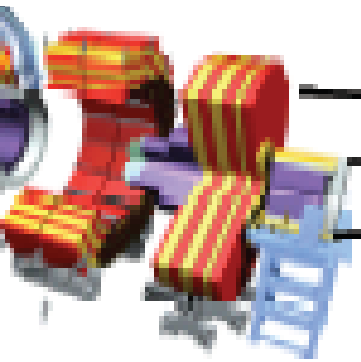
rchical) Distributed Computing Model with multiple Tiers
nal Centers: Managed, fair-shared access to data for Phy
where

aximize total funding resources while meeting the
al computing and data handling needs
alance between proximity of datasets to appropriate resources,
d to the users => Tier-N Model
fficient use of network: higher throughput
Per Flow: Local > regional > national > international
l intellectual resources, in several time zones
Laboratories, universities, remote sites
Involving physicists and students at their home institutions
reater flexibility to pursue different physics interests, priorities, ar
source allocation strategies by region
and/or by common interests (physics topics, subdetectors,…)
anage the System's Complexity
Partitioning facility tasks, to manage and focus resources

36 Nations, 159 Institutions, 1940 Scientists and Engineers (Febr

e CMS
aboration





Online System

100-200 MBytes/s

Tier 0

CERN Computer Center > 20 TIPS

2.5 - 10 Gbits/s

Germany

UK

France

USA

2.5

Tier 2

Tier2 Center

2 Center

2 Center

2 C

~0.6 Gbits/s

Tier 3

Institute

stitute

stitute

stitute

1 Gbits/s

Physic

PCs, other portals

Online System

100-200 MBytes/s

Tier 0

CERN Computer Center > 20 TIPS

2.5 - 10 Gbits/s

Germany

UK

France

USA

Peta Scales

Tier 2

Tier2 Center

~0.6 Gbits/s

Tier 3

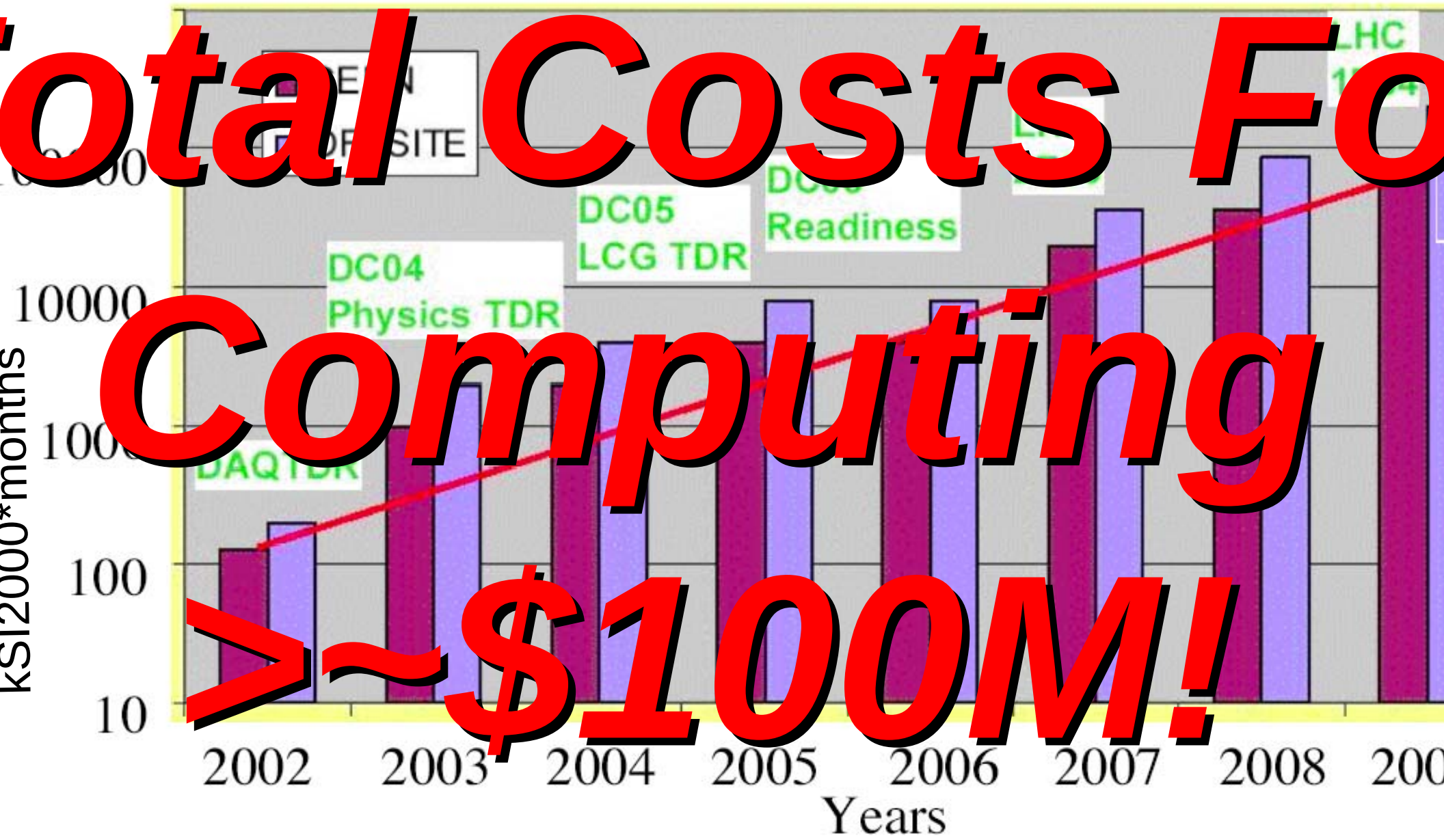
Institute

1 Gbits/s

Physi

r 4

PCs, other portals



Total Costs For Computing >~$100M!

o enable National and Regional Organizations to meet th
tions

ational Funding Agencies pay for computing contributions of proj
n independent set of interfaces leads to partitioning of resources

ore efficient to share large computing facilities with interoperable
terfaces and Environments

self there is no intrinsic value in globally distributing computing
urces

uccess of Grid Computing is based on three fairly simple principl
- National Funding Agencies prefer to spend money in home cou
- Existing local resources (physical infrastructure, HR) can be lev
- plus matching funds and university grants
- Computing Clusters are specified for peak needs and the usage
- has structure => Spare Computing Cycles available somewher

ommon Interfaces are prerequisite to discover available resource
fferent Communities likely to benefit from Grid Computing

oject to build a common grid environment to:

ide the infrastructure and services needed for production and analysis app
ng at scale in a common grid environment.

ide the next phase of the International Virtual Data Grid Laboratory (iVDGL

ide a platform for computer science technology demonstrators.

oject between U.S. ATLAS and U.S. CMS to use a common environment
e LHC Tier-1 and Tier-2 centers:

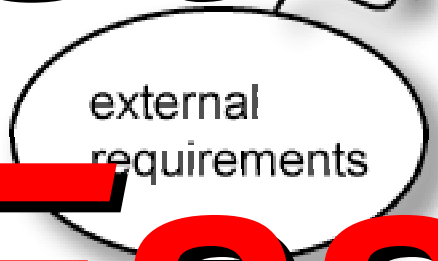
wing software developed by one experiment to be integrated and used by b
ides agreement on policies, principles and procedures for Grid system use.
oring opportunistic use of additional non-HEP computing resources.

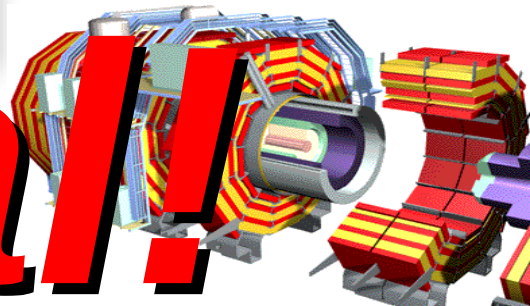
to Demonstrate and Operate a Functioning Multi-Organization Grid:

t well-defined metrics -- a thousand running processes, TBytes/day data tra
CMS, ATLAS, SDSS, LIGO, Biology, CS applications.

S. CMS Grid2003 is a continuation and extension of the existing U.S. CMS
d to participate in a multi-experiment and organization Grid environment.

Development and Integration Test Grids are Essential!





is a Collaborative Team Effort of Application Integrators and Deployers, S
ators and Grid Service Providers and Supporters. Coordinated by a Taskfo
ting the Stakeholders and joint coordinators representing iVDGL & PPDG a
CMS.

ticipants:

ATLAS & CMS

um Grid Projects:

nternational Virtual Data Grid Laboratory (iVDGL), which includes LIGO, SDSS

article Physics Data Grid Collaboratory Pilot (PPDG)

rid Physics Network (GriPhyN)

Telemetry (University of Chicago)

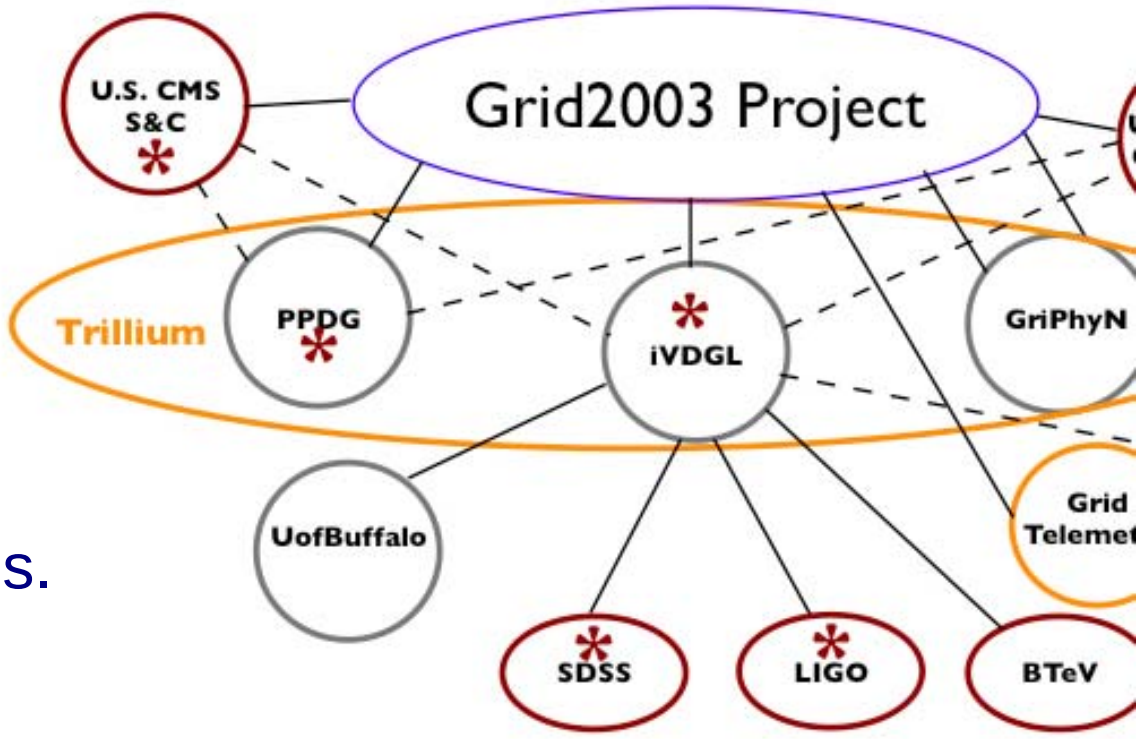
oined During the Project:

Korea

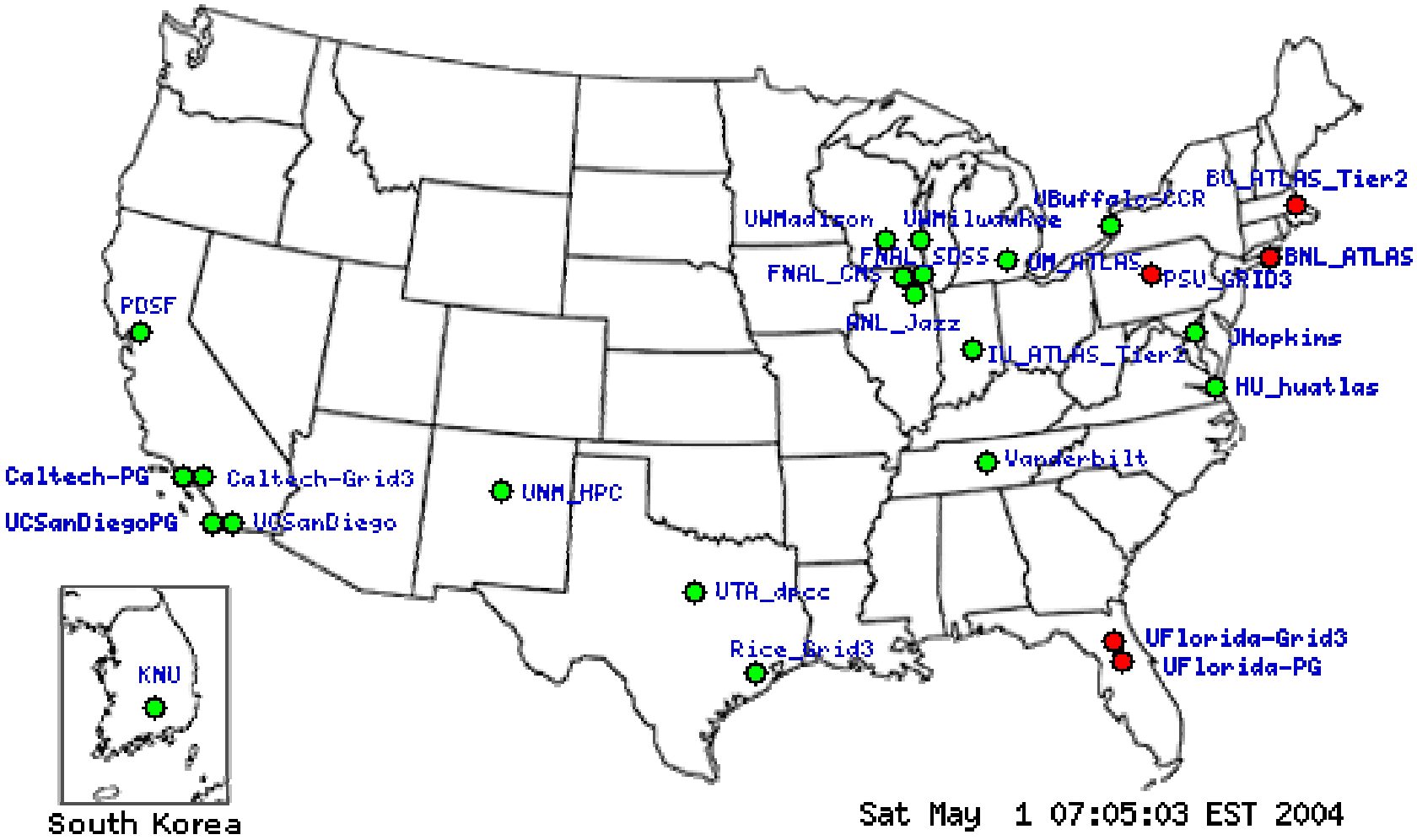
/

Biology - protein sequence analysis.

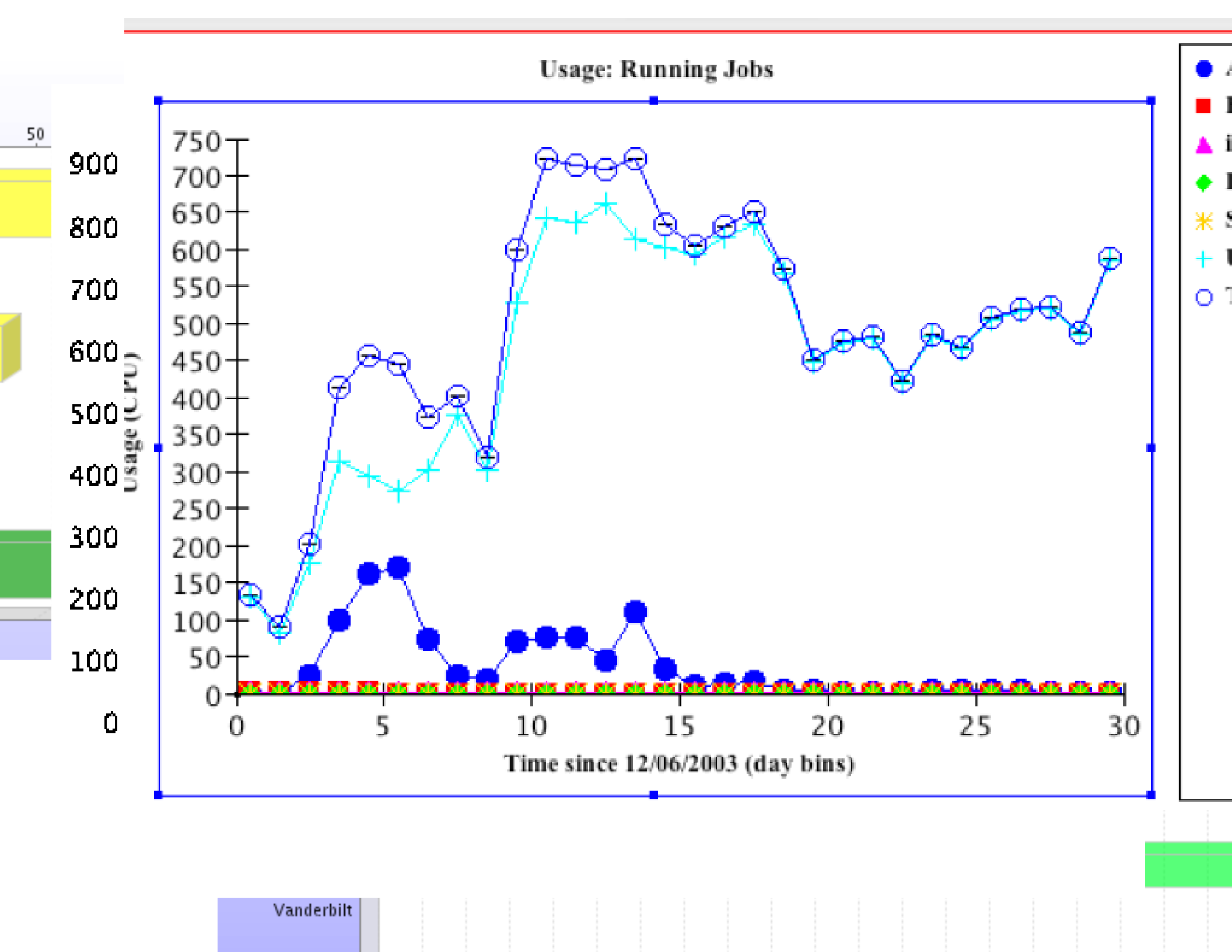
ersity of Buffalo - Biology -



anizations Managed as 4 Virtual-Organizations.
plications including 4 from U.S. LHC.
es including U.S. ATLAS and U.S. CMS Tier-1 & Tier-2 c
CPUs.



(site availability generated


Usage: Running Jobs
Usage (CPU)
Time since 12/06/2003 (day bins)
0
5
10
15
20
25
30
0
50
100
200
300
400
500
600
700
800
900
0
50
100
150
200
250
300
350
400
450
500
550
600
650
700
750

Vanderbilt

Rob Gardner	Mike Wilde	Ian Fisk	Scott Koranda	Rich Baker
Jim Annis	Peter Couvares	Jorge Rodriguez	Leigh Grundhoeffer	Nickolai Kuropatine
Xin Zhao	John Hicks	Ed May	Alain Roy	Brian Moe
Fred Luerhing	Iowna Sakrejda	Yuri Smirnov	Marco Mambelli	Anzar Afaq
Suresh Singh	Carey Kireyev	Alain DeSmet	Jerry Gieraltowski	Doug Olson
Brian Tierney	Saul Youssef	Anne Heavey	Terrence Martin	Andrew Zahn
Scott Gose	Vijay Sekhri	Dantong Yu	Lawrence Sorrillo	Yong Xia
Rob Quick	Michael Ernst	Greg Graham	Bobby Brown	Bockjoo Kim
Jens Voekler	Ruth Pordes	Matt Allen	Yujun Wu	Lisa Giacchetti
Joe Kaiser	Erik Paulson	George Fekete	Dan Engh	Kihyeon Cho
James Letts	Tim Thomas	John Weigand	Iosif Legrand	Mark Green
Craig Prescott	Nosa Olomu	Ben Clifford	Dan Bradley	Timur Perelmutov
Patrick McGuigan	Shawn McKee	Guarang Mehta		

on in Grid3 was at the level of 58 people. 8 worked full time. 10 worked half
strators worked quarter time.

ffort was at about the estimated 7 FTE-years (17 FTE equivalents for 5 mo

etrics were defined in July and measured in November.
e Goal was to operate a Month. Organizations left their resources in Grid20
rformance Metrics were met, but not the Efficiency Metrics.

letric	**Goal**	**Achieved**	**Comments**
umber of CPUs	500	2700	More than 60% of the CPUs are non-dedicated facilities. They are shared with local users
umber of Users	10	102	Most job submissions are by a small number of administrators.
umber of Sites	20	26	Complexity metric.
umber of Sites running oncurrent Applications	10	17	Demonstrates policies in use.
ata Transferred per ay	2-3 TB	4 TB	GridFTP test application. Grid stable during tes
ercentage of esources Used	90%	40-50%	% = Nodes used compared to those available t jobs. Note many nodes used by local jobs.
fficiency of Job ompletion	75%	75% and variable	
umber of Concurrent obs	1000	1100	< 50% of 2700. but many used by Local Jobs. are not measuring this metric correctly yet.
			Grid3 “Meltdown” due to revoked certificate (~ times)

itecture:

ilities: Execution and storage. Include non-dedicated, shared resources. No
uirements on nodes.

vices: Processing, storage, account management, information, monitoring, s
figuration, operations.

lications: Installed dynamically without site administration. Application admi
onsible for the deployment, operation and monitoring of their applications.

Middleware:

ual Data Toolkit (VDT - Globus, Condor, NSF Middleware Initiative, EDG sc

3 extensions for VO (European Data Grid VO Management System), Inform
ension to Glue Schema) and monitoring (Globus MDS, U.S. CMS MonaLisa
glia, MDViewer).

er middleware (e.g. data management) is different for and the responsibility
ication.

g, *Installation and Configuration:* Used iVDGL/ATLAS Pacman for packagi
on of middleware and applications. Goal to make installation simple and mir

ng Service

dor-G or Chimera Virtual Data System submissions through Globus-Gram keeper to one of 4 standard batch systems (PBS, Condor, LSF, FBSng).

nsfer Services

FTP interfaces on all sites through gateway systems

les are transferred into processing sites using globus-url-copy.

pplication specific transfer of Results back

- Basic: GridFTP based Data Movement to Permanent Storage System
- Advanced: Managed Data Storage and Data Access Services based on SRM/dC

ation and Accounts

certificates; Globus Gridmapfiles; Unix Account for Each VO.

g Services

em and application level monitoring

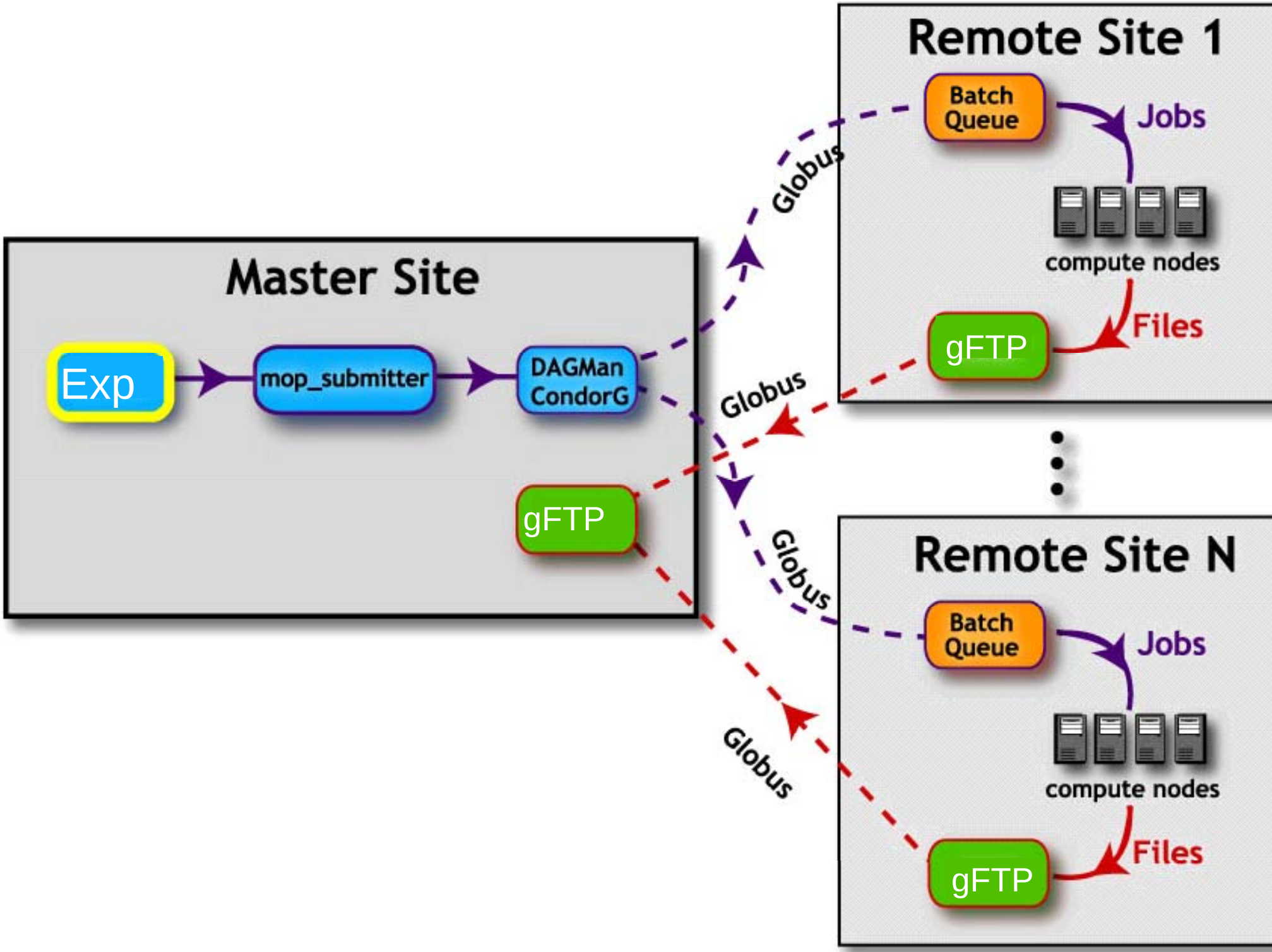
otors between different implementations

rganization Management Services

IS server + administration interface per VO.

scripts to automatically generate gridmapfiles.

on Services


Remote Site 1
Batch Queue
Jobs
compute nodes
gFTP
Files
Globus
Master Site
Exp
mop_submitter
DAGMan CondorG
gFTP
Globus
Globus
Globus
Remote Site N
Batch Queue
Jobs
compute nodes
gFTP
Files

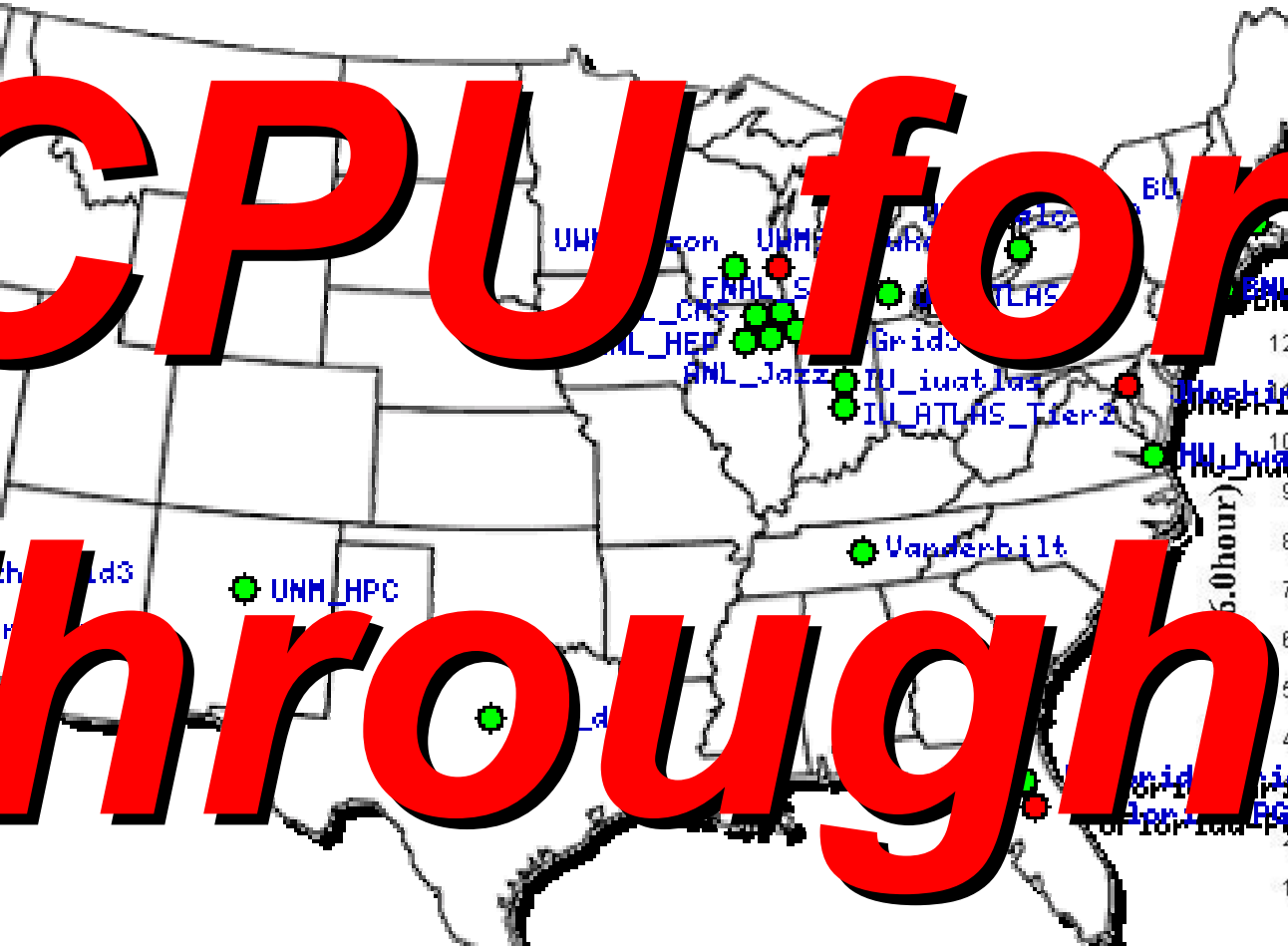
Grid “Regional Center” used dedicated US CMS resources for

missing Grid3 resources for CMS Simulation Production

100% Gain In

CPU for CMS

through Grid



MS Production Facilities stay part of Grid3

MS simulation production fully compatible with Grid3 environme
ll essential US CMS required functionalities have “production qu
upport

rtunistic use of non-CMS resources continues to be ssful!

fter SC2003 milestone, Grid3 kept running successfully
ome local configuration issues, did not impact overall stability
oks like a successful strategy (effectively doubling resources)

peration of U.S. Grid with the rest of LHC Grid?
tions model that would support development and new releas
echnology cycles (data analysis!) on Grid3 and LCG/EGEE?

has Demonstrated:

erogeneous facilities can be used in a common Grid Environment.

operate and use a Multi-Organization Grid with distributed ownership and a coherent system.

rid of over ~20 Facilities can be robust and performant for simple productio ications.

feasibility of the strategy of federating and sharing resources - Open Scien lmap.

e Plans:

rate the current infrastructure for Data Challenges

lve the common Grid Environment for increased capability and performance

t longer term Engineering of Services and Capabilities aligning with and co e Open Science Grid.

tinue interoperability and joint projects with the LCG

utions participating in Grid2003 will continue to contribute their resources a
of the shared common infrastructure in accordance with the MOU with
GL. The details of a model for ongoing Operations will be worked out over
ext few month.

3+ will Operate and Evolve the current Grid Environment:

Upgrade infrastructure to new versions of middleware and applications;

Follow up on recommendations from Lessons Learned document
robustness and hardening, extended monitoring, operations infrastructure)

Continue adding and integrating a (SRM/dCache based) Storage Element.

borate on further projects to Engineer and Deploy the Next Phase of Grid
ces and technologies

as part of the continuation of the existing Grid projects in the US and LCG

n conjunction with the Open Science Grid engineering & blueprint

ing strategy of Interoperability and Joint Projects with the LHC puting Grid Project on many fronts.

orative Efforts in particular on

ommon Virtual Data Toolkit (VDT) delivery and support team.

ata Movement and Storage Management: U.S. CMS demonstration etween Grid2003 and Cern. Using GridFTP, dCache, Storage Resource anagement (SRM).

ob Execution: U.S. ATLAS Grid3 application submission to LCG sites using himera Virtual Data System (VDS).

erge of Information Attributes (GLUE Schema extensions) from Grid2003 nd LCG.

Collaborative Efforts:

rtual Organization Management Project (VOX) collaboration with European ata Grid and LCG Security Working Group.

ontributions to and from the wider CMS and ATLAS software and omputing deliverables.

resentations at and discussions with LCG committees (Grid Deployment oard, Project Oversight Board, Software Computing Committee, Project xecution Board)

articipation in High Energy Physics Joint Technical Board and Global Grid

e putting “real”, multi-organizational Grids to work

not talk about EGEE

G HEP Collaboration (or parts of it) fundamentally is a “physicists
zation”

R&D project with an operational component

not a sustained IT organization or a “distributed computing center”

eed to build the partnerships and need to address the organizational issues

how to build the supporting structures to run a truly distributed, engineered

managed, robust and resilient, accountable and secure service for data ac

and analysis on Peta-scales!

g with centers, universities and projects to formulate a
nap towards the “Open Science Grid”

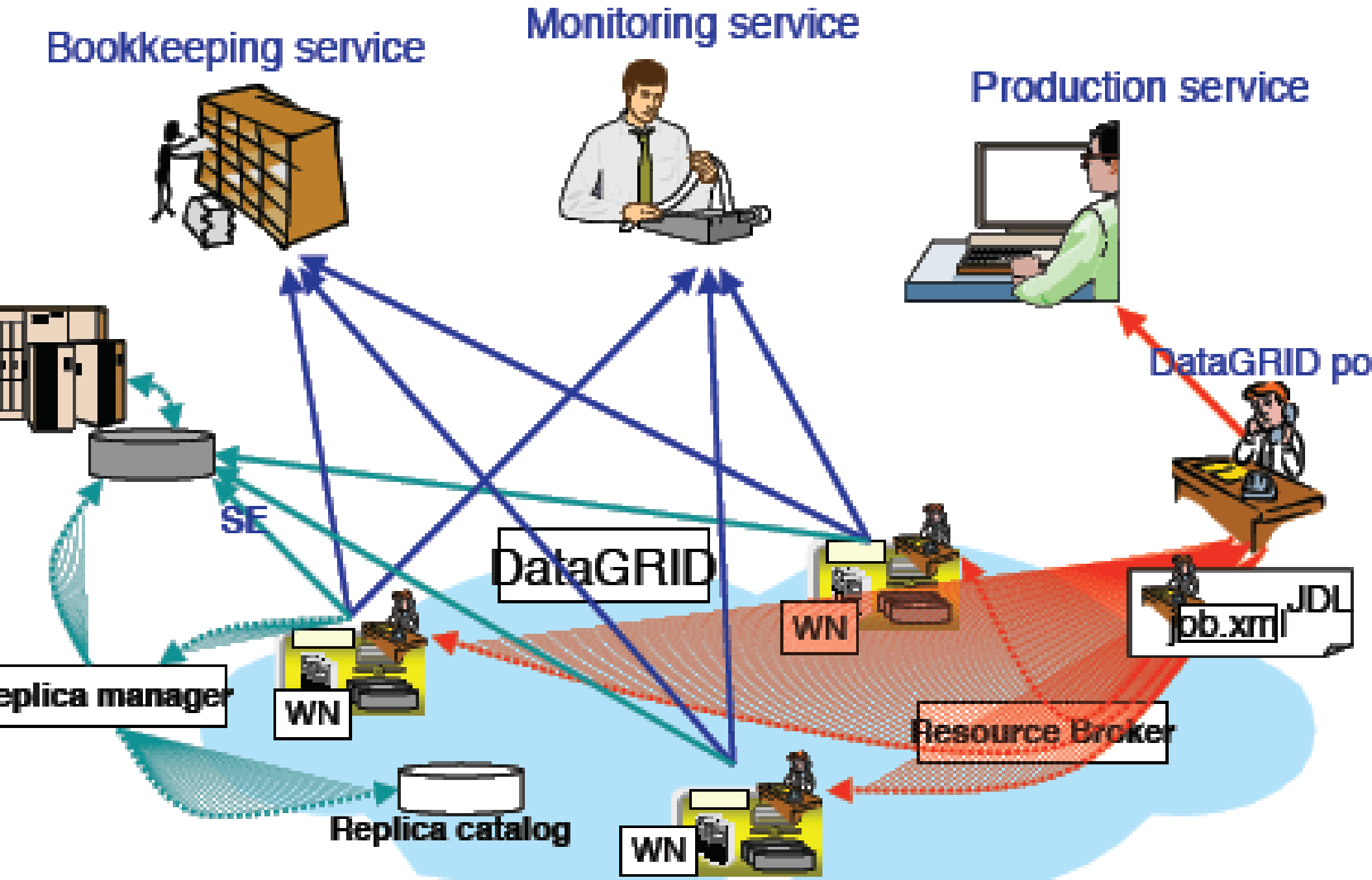
HC as an exemplar global science project to drive the creation of a US Natio
source for science – the Open Science Grid

S. grid middleware and services the basis for international initiatives to buil
oduction grids for science

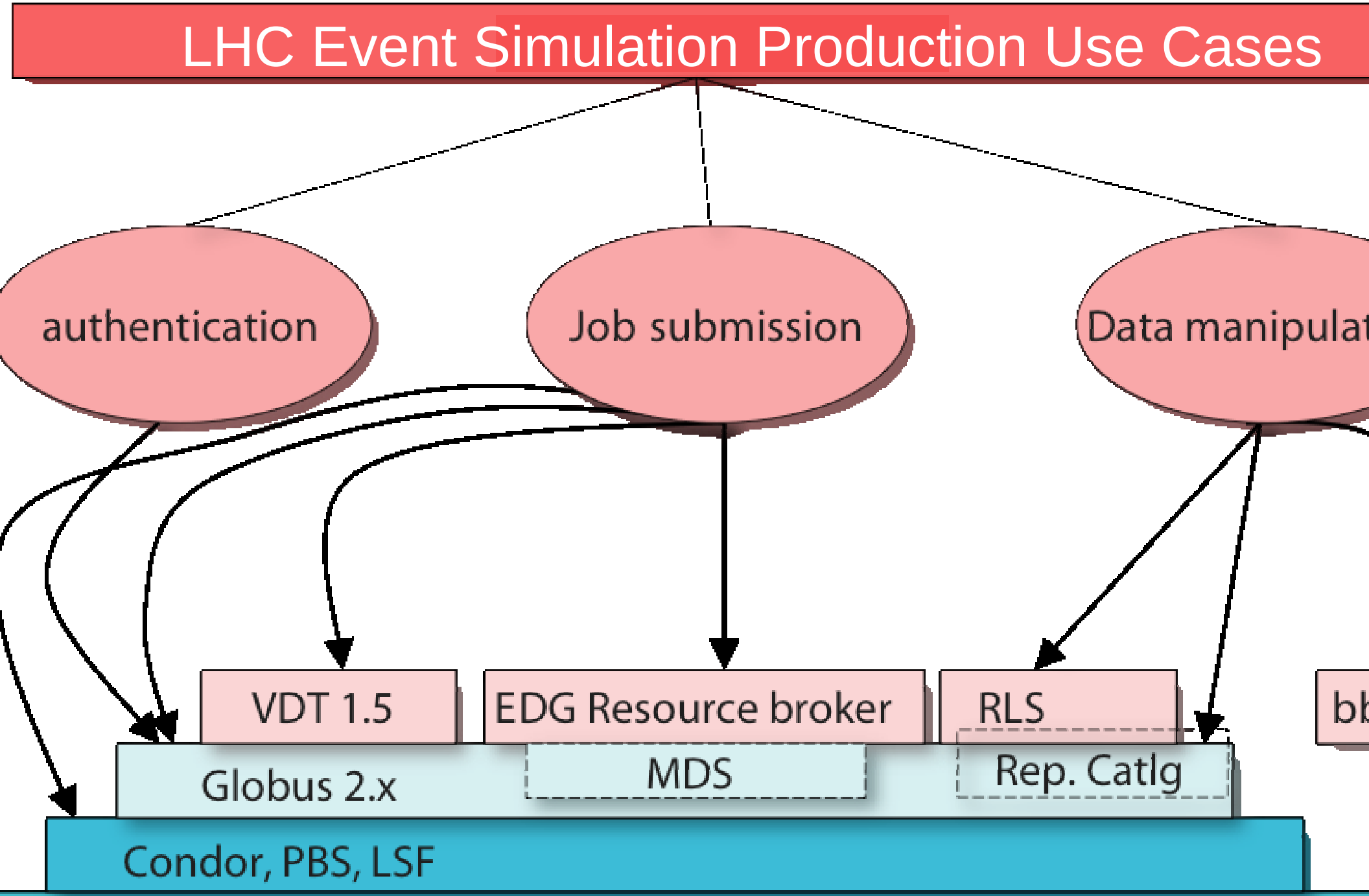
me has come to federate the (U.S.) resources and to continue to “lead”
onstructing a global grid

uild the OSG so it is open to other sciences and complements and interope

services for processing, to be further developed



LHC software stack for Event Simulation Production



Architecture: (H. Newman)

Above the Collective Layer

t's Application Codes

econstruction, Calibration, Analysis

ents' Software Framework Layer

Modular and Grid-aware:

rchitecture able to interact effectively

ith the lower layers (above)

lications Layer

ers and algorithms that govern system operations)

olicy and priority metrics

Workflow evaluation metrics

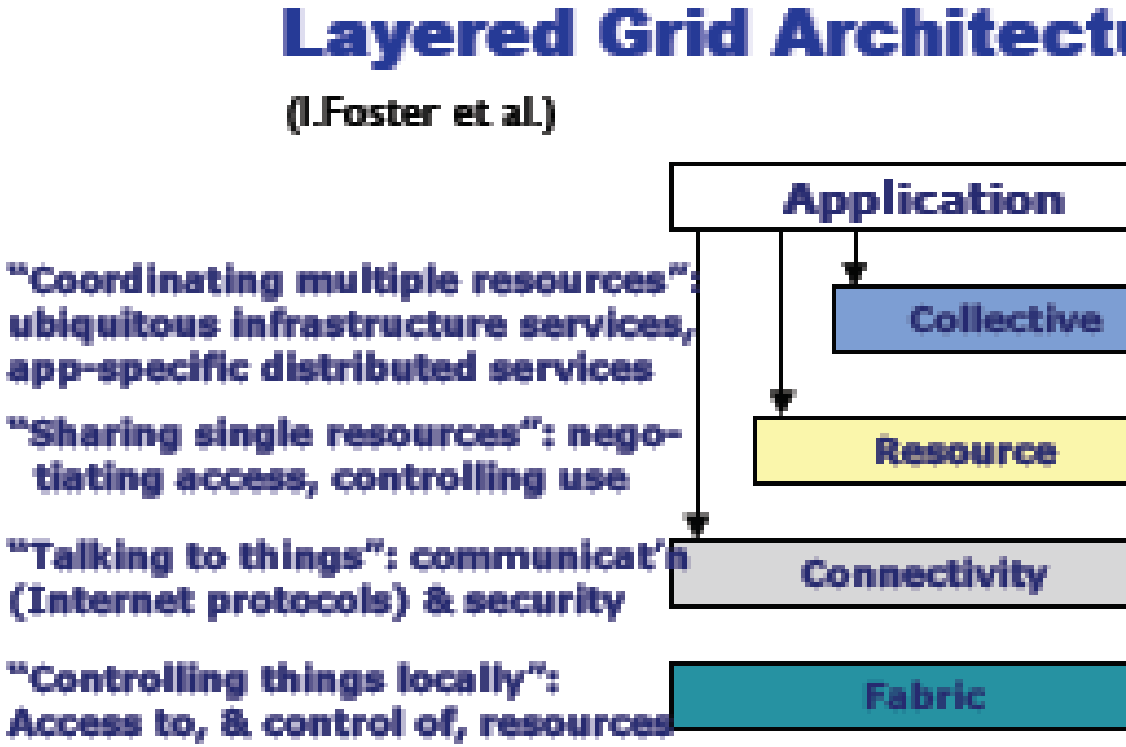
ask-Site Coupling proximity metrics

nd-to-End System Services Layer

Workflow monitoring and evaluation mechanisms

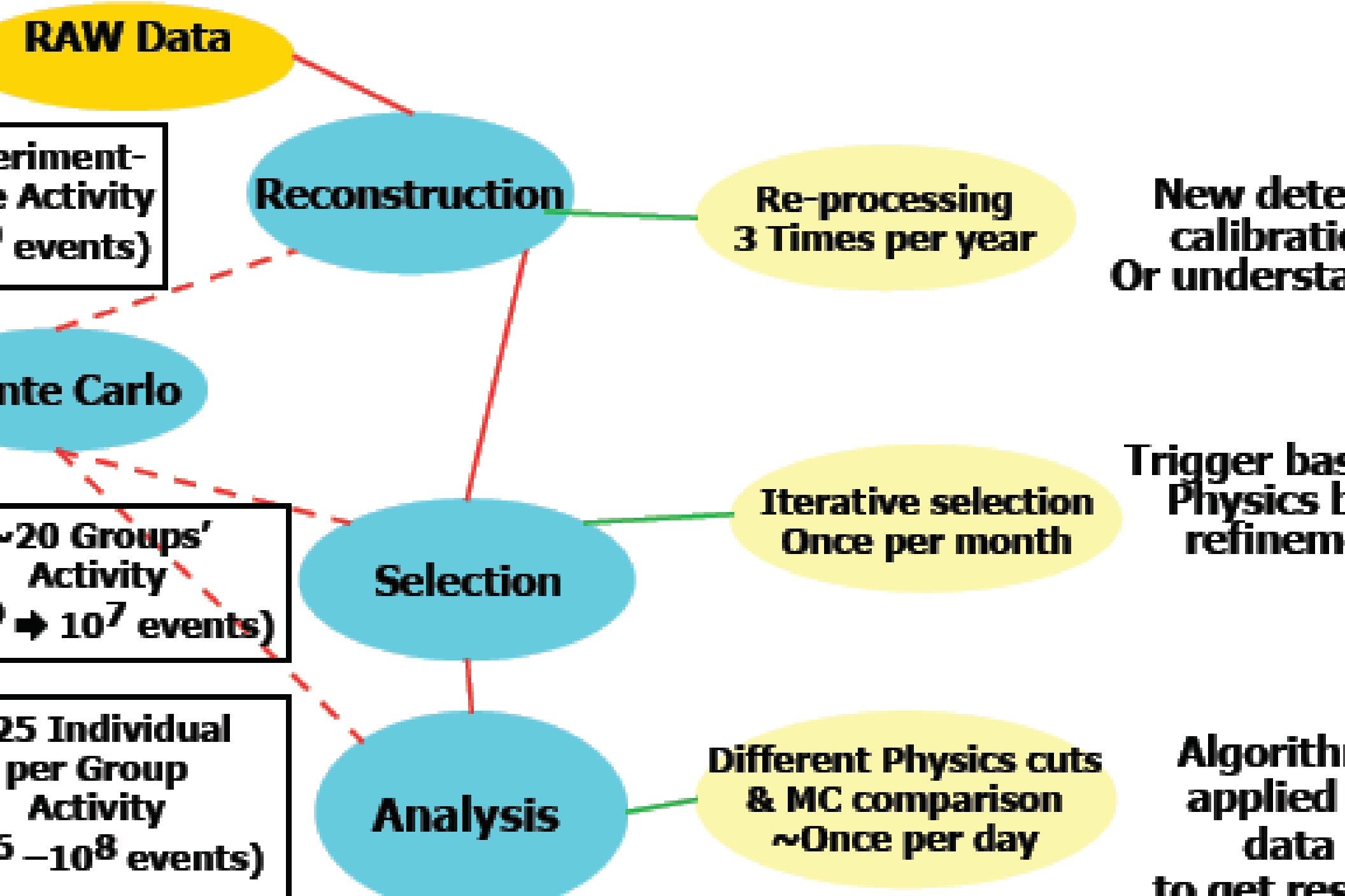
rror recovery and long-term redirection mechanisms

ystem self-monitoring, steering, evaluation and optimization mechanisms

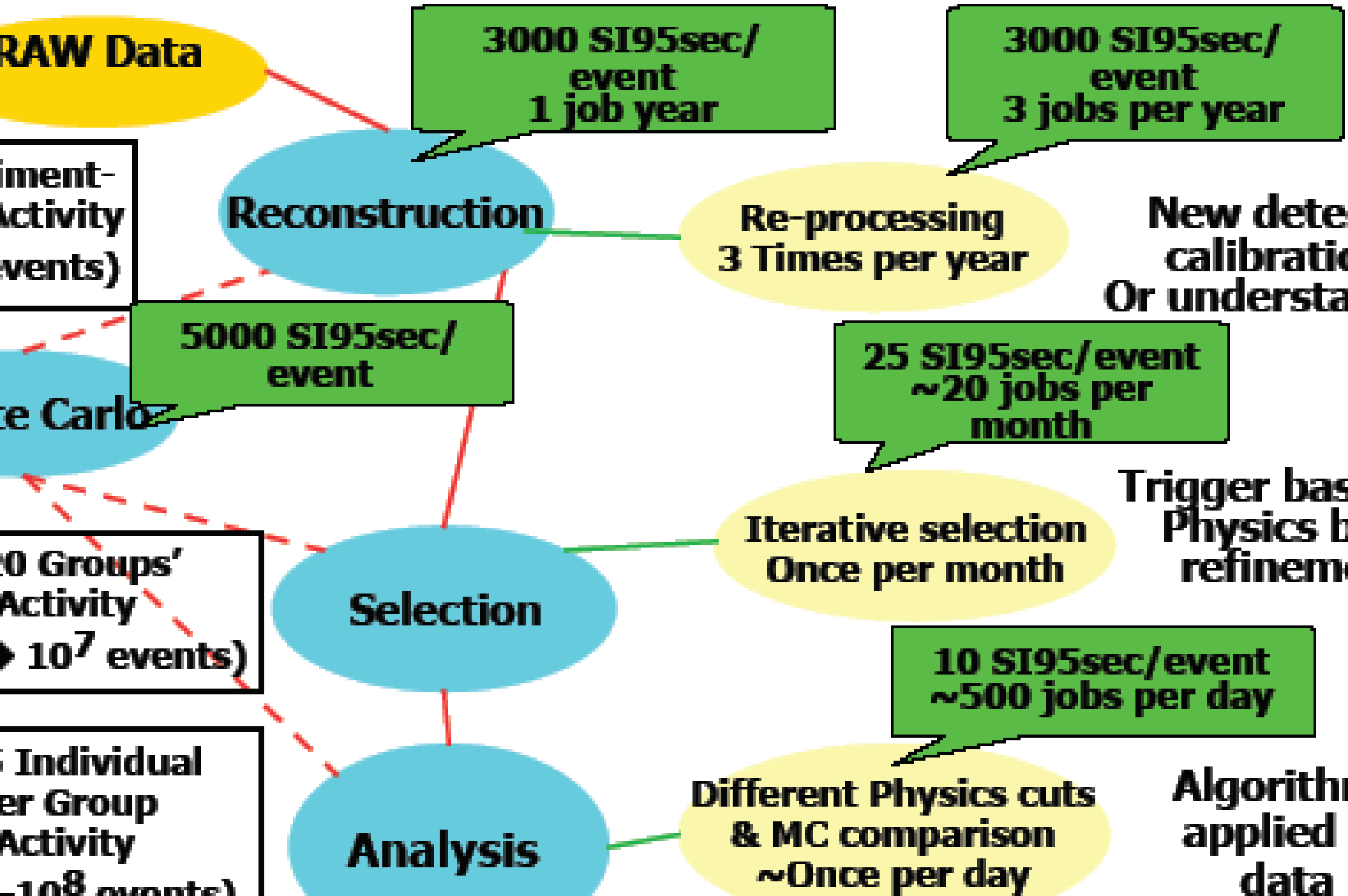


forward to physics data analysis means a significant paradigm s
m well-defined production jobs => interactive user analysis
m DAGs of process => “Sessions” and state-full environments
m producing “sets of files” => accessing massive amounts of data
m files => data sets and collection of objects
m using essentially “raw data” => complex layers of event representation
m “assignments” from a central repository => Grid-wide queries
m “user registration” => enabling sharing and building communities
e (Grid) technologies ready for this?
ere needs to be a tight inter-play between prototyping the analysis services
veloping the “lower level” services and interfaces => ARDA Prototype
are going to be the “new paradigms” that will be exposed to the u
er “data analysis session” transparently extended to a distributed system?
but requires a more prescriptive and declarative approach to analysis
of services for “collaborative” work?
new paradigms beyond “analysis”

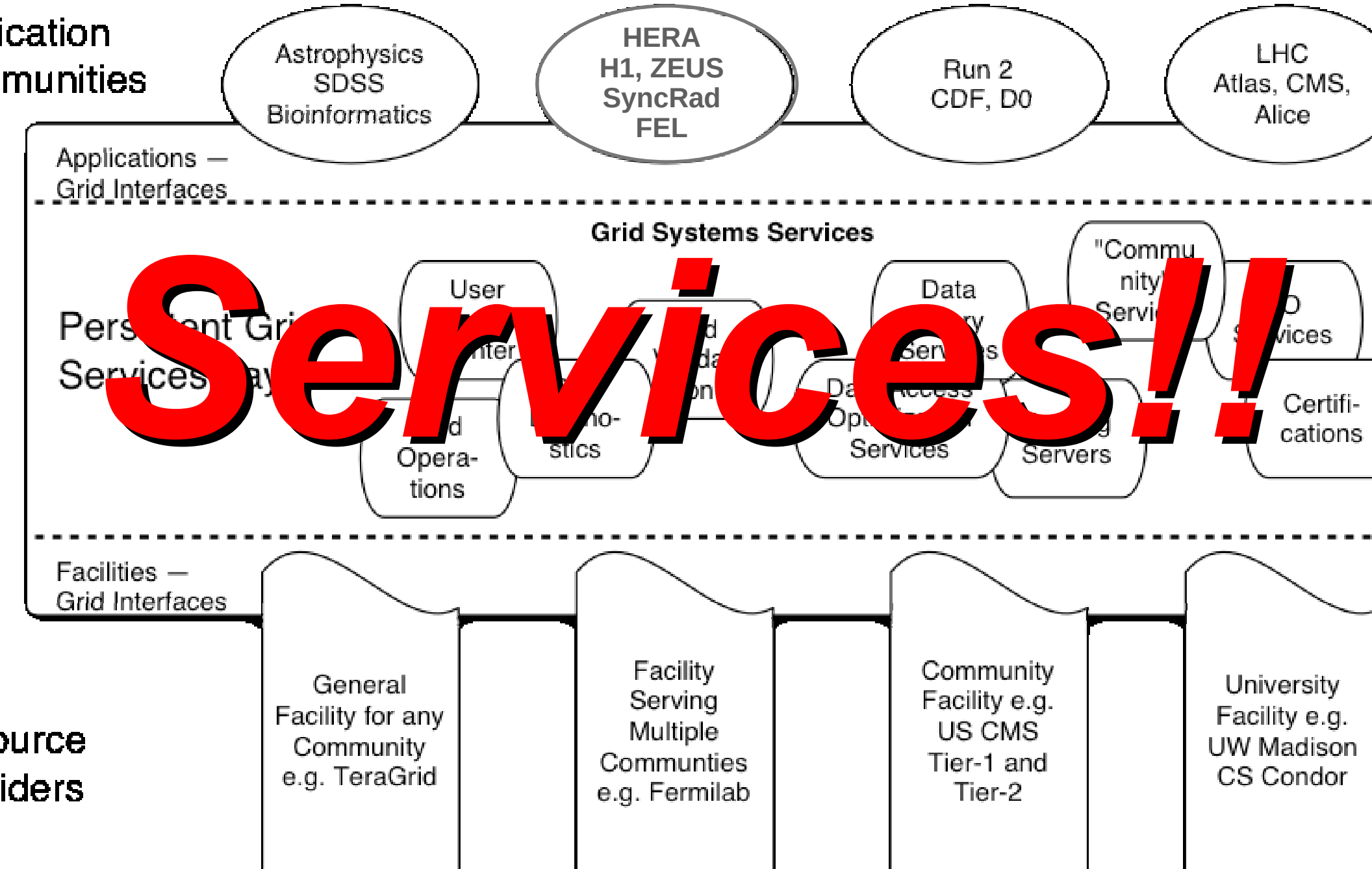
Hierarchy of Processes (Experiment, Analysis Groups,Individuals)



Hierarchy of Processes (Experiment, Analysis Groups,Individuals)



Grid Layer "Abstraction" of Facilities — Rich with Services!



m extracts a subset of the datasets from the virtual file catalogue
lata conditions provided by the user.

m splits the tasks according to the location of data sets.

alancing between local data access and data replication.

n sub-jobs and submit to Workload Management with precise job
ptions

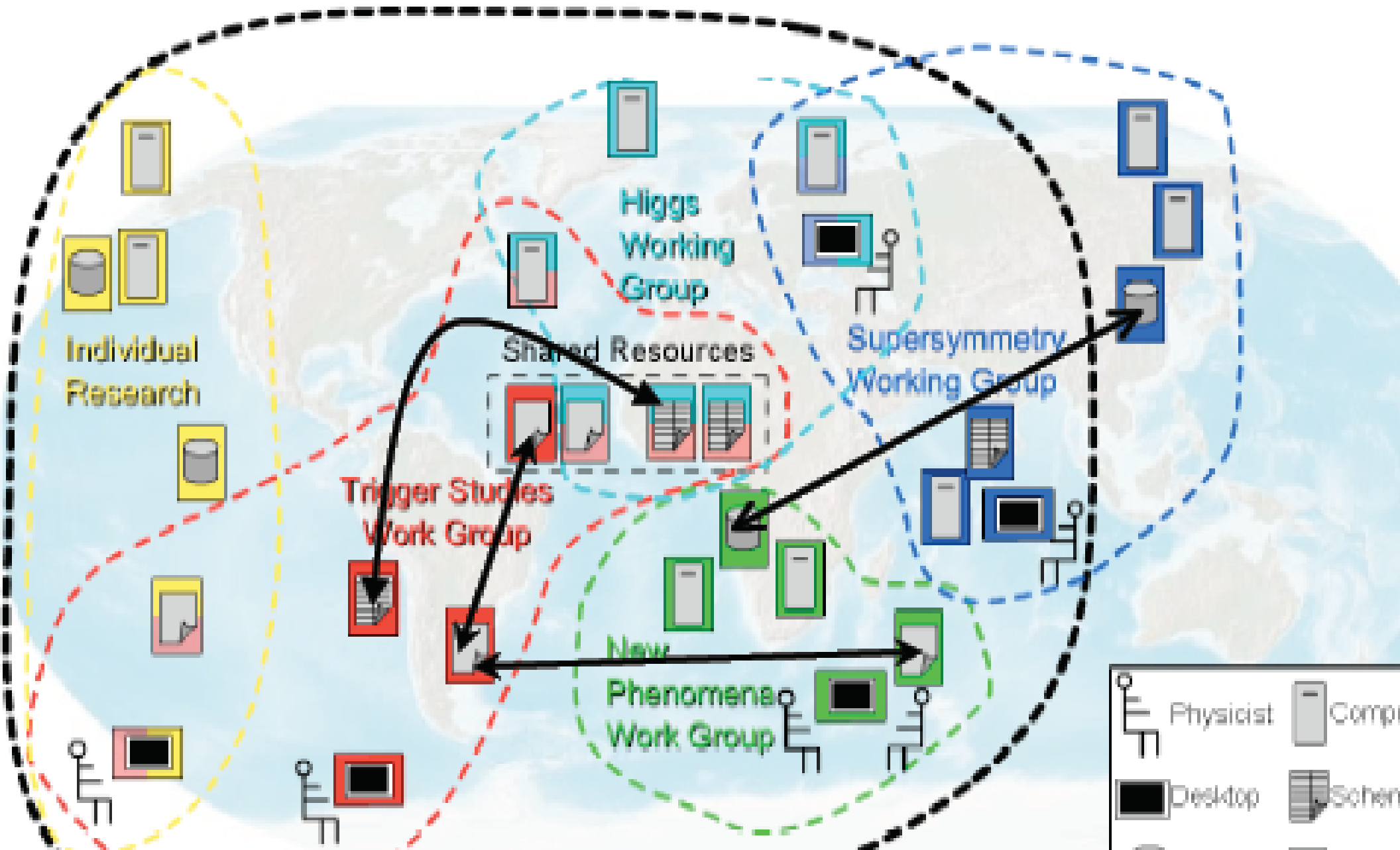
Jser can control the results while and after data are processed

t and Merge available results from all terminated sub-jobs on re

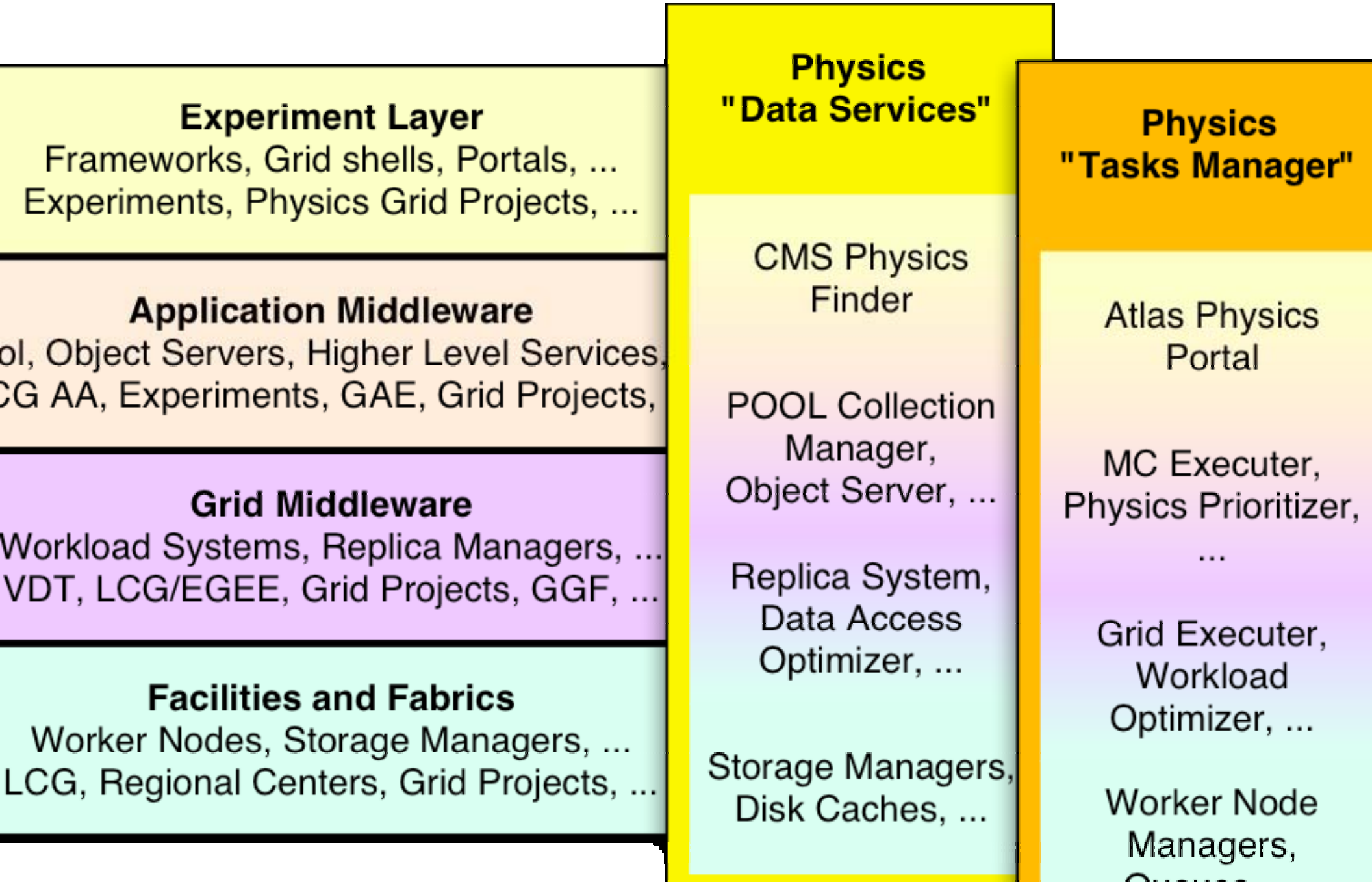
sis objects associated with the analysis task remains persistent i
nvironment so the user can go offline and reload an analysis tas
late, check the status, merge current results or resubmit the sam
nodified analysis code.

Communities of Scientists Using the Grid for Distributed Analysis

Infrastructure for sharing, consistency of physics and calibration data, softw



iment's "Services" go end-to-end!



What is analysis anyway?

Monday, 10 May, 2004
Massimo Lamanna (CERN)
“The ARDA Project: Grid Analysis Prototypes of the LHC Experiments”

’s HEP Collaborations are getting ready for the big challe

xpected to lead to discoveries of new elementary particles and nc
aviors of the fundamental forces — and a decades-long scientific
ram

lopted a globally distributed computing model, to enable science
al scale, and to enable scientists worldwide to be full part of the
ected breakthrough discoveries

any technology and organizational challenges ahead

orating with computer scientists and other scientific
unities

onstruct a global cyberinfrastructure of international computing gr
at will be used by thousands of scientists
bal to enable scientific collaborators to work together as co-locate
eers, and to create new capabilities to empower the individual sc

# of Research for Grid Federations

erface Languages and Standard Protocols

- Experiments need to be able to describe Interfaces needed by applications
- Current model where Computing Resource Providers encourage to install suites of software limits ability to flexibly deploy Grid Services for Applicati

onitoring and Information Providing

- to track VO usage of resources
- to indicate to incoming users what the likely priority is they will receive
- much richer information provider is needed to convey information to optimi users to enable intelligent scheduling decisions

uthentication – A lot of work has been done, but …

- Consensus needs to be built on how tools are used and administered

uthorization and Privilege – Almost nothing exists

- Closely related to priority and quota setting

erating Resources to form a Federated Grid is necessary

g together Computing Resources from Different Commun

only then that the Power of Grid Enabled Distributed Co

begin to be realized.

important for the Computing Providers and the Experime

ed to work within the Framework of Federated Resource

beginning of Development.

eloping a Homogeneous Distributed Computing System

hes the Experiments very little about Operating in a Grid

ronment, and more importantly gains them very little in te

nputing Resources.

Thank you