e –Infrastructure Components

Ian Collier

STFC Scientific Computing Department

UKT0 Collaboration Meeting

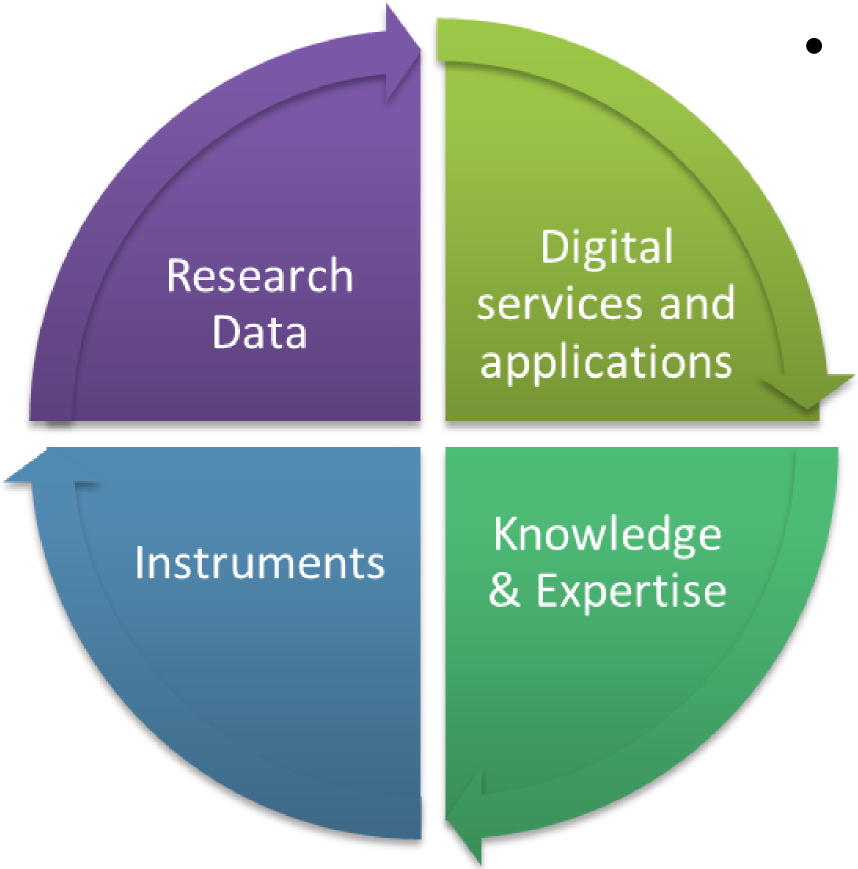
Abingdon 15th March 2018

Science & Technology Facilities Council
Rutherford Appleton Laboratory

Introduction

- A look at (some of) the essential elements of an e-Infrastructure
  - Through the lens of WLCG (effectively a platform running on multiple e-Infrastructures

What is an e-Infrastructure



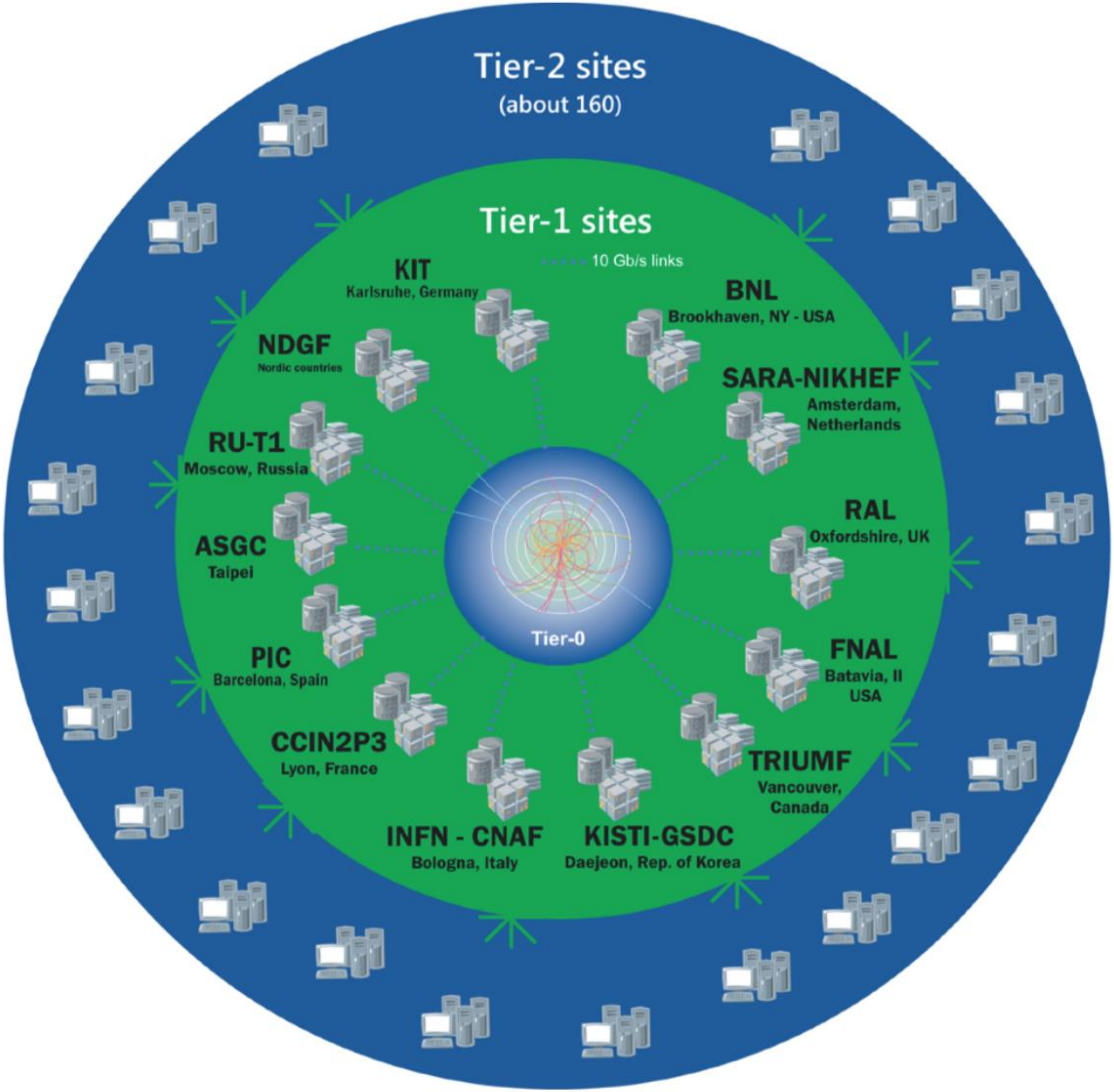
- e-Infrastructure = underpinning IT
  - Tools and Services
  - Data
  - Compute
  - Networks and Connectivity

WLCG

- WLCG = Worldwide LHC Computing Grid
  - Storage and CPU beyond that run by the experiments and by CERN itself
- Effectively a Platform as a Service running on two main infrastructures:
  - EGI (formerly European Grid Infrastucture)
  - Open Science Grid (mainly in US)
- In turn federate national and institutional infrastructures

Scale of WLCG

- Pledged WLCG resources for 2017
  - CPU 5.1 million HS06 (~500,000 processors)
  - Disk 396 PB (immediate access)
  - Tape 590 PB (ie robot tape libraries)
- Most countries oversupply, some significantly
  - eg UK supplies twice as much disk as pledged
- ~170 sites in ~40 countries across 6 continents


Tier-2 sites
(about 160)
Tier-1 sites
10 Gb/s links
KIT
Karlsruhe, Germany
BNL
Brookhaven, NY - USA
NDGF
Nordic countries
SARA-NIKHEF
Amsterdam,
Netherlands
RU-T1
Moscow, Russia
RAL
Oxfordshire, UK
ASGC
Taipei
Tier-0
FNAL
Batavia, Il
USA
PIC
Barcelona, Spain
CCIN2P3
Lyon, France
TRIUMF
Vancouver,
Canada
INFN - CNAF
Bologna, Italy
KISTI-GSDC
Daejeon, Rep. of Korea

Essential components

- Distributed compute and storage resources
  - With clearly defined interfaces (do not all have to be identical)
- An effective network
- Information systems
  - What resources and services are available where
- Operational security
  - Security policy, monitoring, incident response
- Authentication & authorisation
  - Including management of groups of users
- Resource usage accounting
- A mechanism for allocating/sharing resources
- Availability monitoring/testing
- Mechanism for distributing application software
- Ticketing system

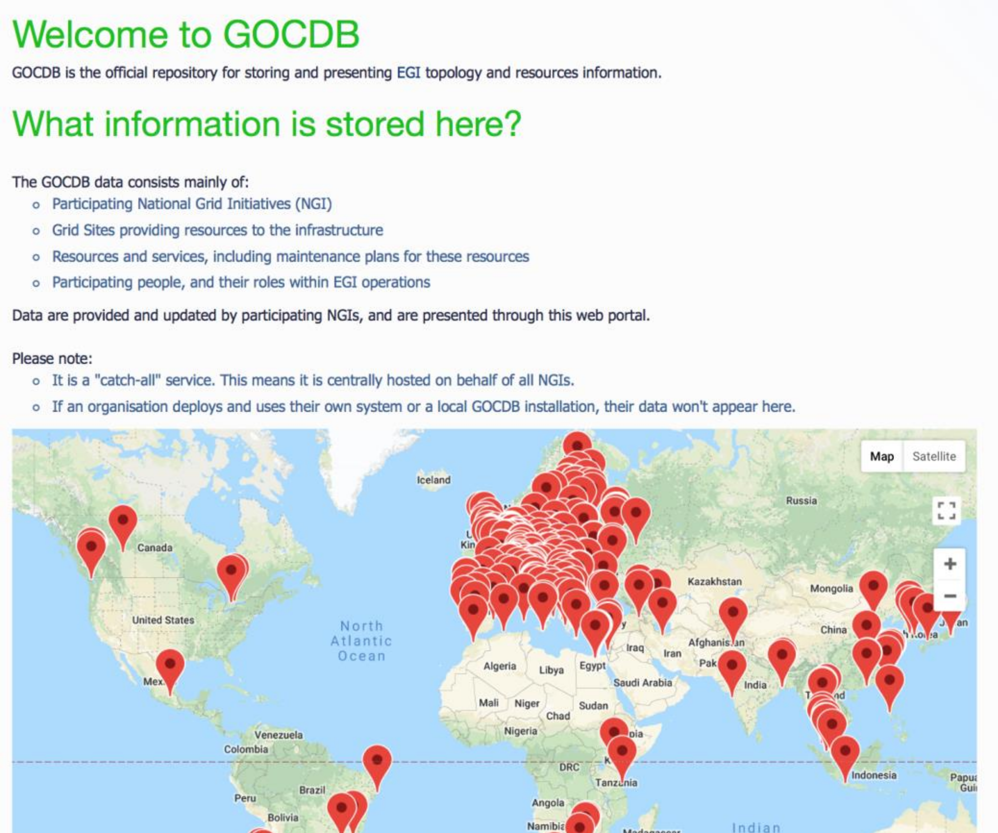
Essential components

- Different functions are met in different places:
  - Centrally run by e-Infrastructures
  - Centrally by WLCG
  - By the experiments
  - At the sites
- And some functions have moved between these over time

Central Services

Information system - GOCDB

- GOCDB is Grid Operations Centre Database – service registry
  - Used to manually register sites and services within sites
  - Used to declare downtimes of services
  - Has programmatic API to allow experiments to gather information automatically



Welcome to GOCDB

GOCDB is the official repository for storing and presenting EGI topology and resources information.

What information is stored here?

The GOCDB data consists mainly of:

- Participating National Grid Initiatives (NGI)
- Grid Sites providing resources to the infrastructure
- Resources and services, including maintenance plans for these resources
- Participating people, and their roles within EGI operations

Data are provided and updated by participating NGIs, and are presented through this web portal.

Please note:

- It is a "catch-all" service. This means it is centrally hosted on behalf of all NGIs.
- If an organisation deploys and uses their own system or a local GOCDB installation, their data won't appear here.

- It's essential to have some uniform way of keeping track of sites and what they make
- Especially true of downtimes

Ticketing system - GGUS

- GGUS is Global Grid User Support
  - Distributed ticketing system (cf JIRA, RT etc)
  - Includes site admins, experiment ops people, and middleware developers
- Hard to overstate importance
  - Used by experiments when handing over between different people on shift
  - Allows regional projects (eg GridPP) to see how responsive the sites are to problems
  - Exposes common problems
  - It allows us to shunt problems off to the developers when it's the software at fault ... or vice versa – in one system

Resource accounting - APEL

- APEL is the accounting framework
  - Defines job accounting records (semantically and in terms of text files, XML, ...)
  - Provides tools to publish and aggregate numbers
  - Used by EGI and OSG
- Used by experiments to compare with pledged resources
  - Provides objective motivation for sites to keep things working efficiently
- Tied in with GOCDB site information
- Recent extensions for cloud resources & storage

Availability testing – SAM/ETF

- SAM is Site Availability Monitoring
  - Now Experiment Testing Framework (ETF)
  - Run by WLCG itself
  - Sends test jobs to sites for the experiments
- Long tradition of this going back to ~2002
- Monitoring is the only way to keep sites up
  - (Things break silently and still look ok to site admins)
- Since test scripts are supplied by experiments, they can test aspects of the site that matter to them
- Also useful for validating new or upgraded services
- Used in monthly site Availability and Reliability Reports sent to each country - again, very motivational!

Others

- FTS – File Transfer Service
  - Used by individual experiment data management tools
- CVMFS – CernVM Filesystem – software distribution
  - Globally distributed virtual filesystem
  - Distributes **all** WLCG experiment software to **all** sites (and clouds)
  - Replaced very cumbersome and fragile process involving more or less powerful software servers at all sites plus privileged jobs that installed latest software
    - Had to tag sites with available versions and only send jobs ot sites that had the right versions

Based on http, fuse, squid, sqlite

User group management - VOMs

- VOMs – the Virtual Organization Membership Service
  - Structured, simple account database
  - Defines specific capabilities and roles for users

Site Services

Site Services

- Site services broadly divide into job execution and storage
- Multiple implementations in production
- Jobs: CREAM, ARC, HTCondor-CE
  - All with their own APIs!
  - You also need to run a batch system: PBS/Torque, HTCondor, Grid Engine, LSF
  - Also a number of approaches to suporting VMs and containers
- Storage: DPM, dCache, StoRM
  - All provide SRM web service for management
  - And xrootd, GridFTP, HTTP for low level transfers
- Other services at sites publish information, in or out
  - APEL (accounting), Argus (authorization), BDII (detailed service descriptions)

Worker Nodes

- WLCG maintains a high degree of uniformity between the machines that execute the jobs
  - Whether batch nodes or in VM systems like Clouds
- So mostly SL6.x (rebuild of RHEL 6)
  - CentOS 7.x increasing rapidly
  - Support for containers allowing greater choice here
- For managing the experiments' software distribution CernVM-FS (cvmfs) has been invaluable
  - Wide-area readonly filesystem with strong versioning to present a coherent view to jobs
  - Also the basis of the root filesystem with the OS in CernVM-based VMs

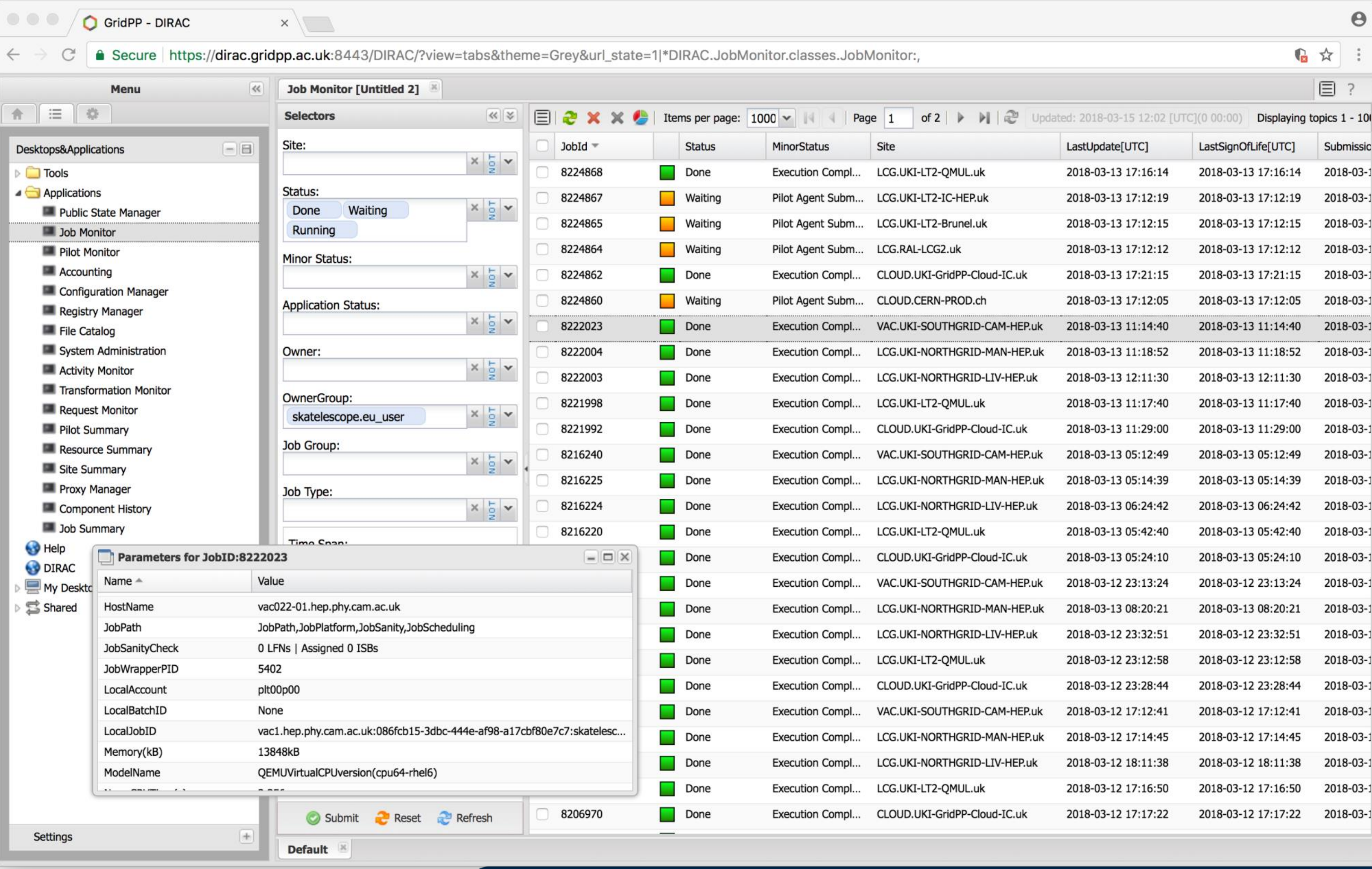
Experiment Central Services

Experiment Grid Frameworks

- Hope in EDG was that generic software could be written to manage jobs and data for all four experiments
  - Experiments would just need bookkeeping systems specific to their data formats
  - And agents to create jobs en-masse to process their datasets
- Limitations in the initial solutions led all four experiments to develop their own job and data management systems
  - Some components (eg HTCondor, FTS) are used by more than one experiment, but they're still mutually incompatible
- Some work has been done to generalise frameworks and remove experiment specifics (eg DIRAC and BigPanda)

Experiment Grid Services

- Typically
  - A job management system (DIRAC WMS, BigPanda, ...) that users and production systems submit jobs to
  - A data management system (DIRAC DMS, Rucio, Fedex, ...) that uses lower level tools to move data around
  - Their own information system to keep track of sites, job queues etc (DIRAC CS, AGIS, ...)
  - An automated monitoring system to temporarily test and ban problem sites (DIRAC RSS, Hammercloud, ...)
  - Various dashboards used by people on shift to look for problems and debug things going wrong
    - Then write a GGUS ticket etc



https://dirac.gridpp.ac.uk:8443/DIRAC/?view=tabs&theme=Grey&url_state=1|*DIRAC.JobMonitor.classes.JobMonitor:,

- Tools
- Applications
  - Public State Manager
  - Job Monitor
  - Pilot Monitor
  - Accounting
  - Configuration Manager
  - Registry Manager
  - File Catalog
  - System Administration
  - Activity Monitor
  - Transformation Monitor
  - Request Monitor
  - Pilot Summary
  - Resource Summary
  - Site Summary
  - Proxy Manager
  - Component History
  - Job Summary
- Help
- DIRAC
- My Deskto
- Shared

Site:

Status: Done, Waiting, Running

Minor Status:

Application Status:

Owner:

OwnerGroup: skatelescope.eu_user

Job Group:

Job Type:

Time Span:

Items per page: 1000 Page 1 of 2 Updated: 2018-03-15 12:02 [UTC](0 00:00) Displaying topics 1 - 10

JobId		Status	MinorStatus	Site	LastUpdate[UTC]	LastSignOfLife[UTC]	Submissio
8224868	🟩	Done	Execution Compl...	LCG.UKI-LT2-QMUL.uk	2018-03-13 17:16:14	2018-03-13 17:16:14	2018-03-1
8224867	🟧	Waiting	Pilot Agent Subm...	LCG.UKI-LT2-IC-HEP.uk	2018-03-13 17:12:19	2018-03-13 17:12:19	2018-03-1
8224865	🟧	Waiting	Pilot Agent Subm...	LCG.UKI-LT2-Brunel.uk	2018-03-13 17:12:15	2018-03-13 17:12:15	2018-03-1
8224864	🟧	Waiting	Pilot Agent Subm...	LCG.RAL-LCG2.uk	2018-03-13 17:12:12	2018-03-13 17:12:12	2018-03-1
8224862	🟩	Done	Execution Compl...	CLOUD.UKI-GridPP-Cloud-IC.uk	2018-03-13 17:21:15	2018-03-13 17:21:15	2018-03-1
8224860	🟧	Waiting	Pilot Agent Subm...	CLOUD.CERN-PROD.ch	2018-03-13 17:12:05	2018-03-13 17:12:05	2018-03-1
8222023	🟩	Done	Execution Compl...	VAC.UKI-SOUTHGRID-CAM-HEP.uk	2018-03-13 11:14:40	2018-03-13 11:14:40	2018-03-1
8222004	🟩	Done	Execution Compl...	LCG.UKI-NORTHGRID-MAN-HEP.uk	2018-03-13 11:18:52	2018-03-13 11:18:52	2018-03-1
8222003	🟩	Done	Execution Compl...	LCG.UKI-NORTHGRID-LIV-HEP.uk	2018-03-13 12:11:30	2018-03-13 12:11:30	2018-03-1
8221998	🟩	Done	Execution Compl...	LCG.UKI-LT2-QMUL.uk	2018-03-13 11:17:40	2018-03-13 11:17:40	2018-03-1
8221992	🟩	Done	Execution Compl...	CLOUD.UKI-GridPP-Cloud-IC.uk	2018-03-13 11:29:00	2018-03-13 11:29:00	2018-03-1
8216240	🟩	Done	Execution Compl...	VAC.UKI-SOUTHGRID-CAM-HEP.uk	2018-03-13 05:12:49	2018-03-13 05:12:49	2018-03-1
8216225	🟩	Done	Execution Compl...	LCG.UKI-NORTHGRID-MAN-HEP.uk	2018-03-13 05:14:39	2018-03-13 05:14:39	2018-03-1
8216224	🟩	Done	Execution Compl...	LCG.UKI-NORTHGRID-LIV-HEP.uk	2018-03-13 06:24:42	2018-03-13 06:24:42	2018-03-1
8216220	🟩	Done	Execution Compl...	LCG.UKI-LT2-QMUL.uk	2018-03-13 05:42:40	2018-03-13 05:42:40	2018-03-1
	🟩	Done	Execution Compl...	CLOUD.UKI-GridPP-Cloud-IC.uk	2018-03-13 05:24:10	2018-03-13 05:24:10	2018-03-1
	🟩	Done	Execution Compl...	VAC.UKI-SOUTHGRID-CAM-HEP.uk	2018-03-12 23:13:24	2018-03-12 23:13:24	2018-03-1
	🟩	Done	Execution Compl...	LCG.UKI-NORTHGRID-MAN-HEP.uk	2018-03-13 08:20:21	2018-03-13 08:20:21	2018-03-1
	🟩	Done	Execution Compl...	LCG.UKI-NORTHGRID-LIV-HEP.uk	2018-03-12 23:32:51	2018-03-12 23:32:51	2018-03-1
	🟩	Done	Execution Compl...	LCG.UKI-LT2-QMUL.uk	2018-03-12 23:12:58	2018-03-12 23:12:58	2018-03-1
	🟩	Done	Execution Compl...	CLOUD.UKI-GridPP-Cloud-IC.uk	2018-03-12 23:28:44	2018-03-12 23:28:44	2018-03-1
	🟩	Done	Execution Compl...	VAC.UKI-SOUTHGRID-CAM-HEP.uk	2018-03-12 17:12:41	2018-03-12 17:12:41	2018-03-1
	🟩	Done	Execution Compl...	LCG.UKI-NORTHGRID-MAN-HEP.uk	2018-03-12 17:14:45	2018-03-12 17:14:45	2018-03-1
	🟩	Done	Execution Compl...	LCG.UKI-NORTHGRID-LIV-HEP.uk	2018-03-12 18:11:38	2018-03-12 18:11:38	2018-03-1
	🟩	Done	Execution Compl...	LCG.UKI-LT2-QMUL.uk	2018-03-12 17:16:50	2018-03-12 17:16:50	2018-03-1
8206970	🟩	Done	Execution Compl...	CLOUD.UKI-GridPP-Cloud-IC.uk	2018-03-12 17:17:22	2018-03-12 17:17:22	2018-03-1

Parameters for JobID:8222023

Name	Value
HostName	vac022-01.hep.phy.cam.ac.uk
JobPath	JobPath,JobPlatform,JobSanity,JobScheduling
JobSanityCheck	0 LFNs | Assigned 0 ISBs
JobWrapperPID	5402
LocalAccount	plt00p00
LocalBatchID	None
LocalJobID	vac1.hep.phy.cam.ac.uk:086fcb15-3dbc-444e-af98-a17cbf80e7c7:skatelesc...
Memory(kB)	13848kB
ModelName	QEMUVirtualCPUversion(cpu64-rhel6)

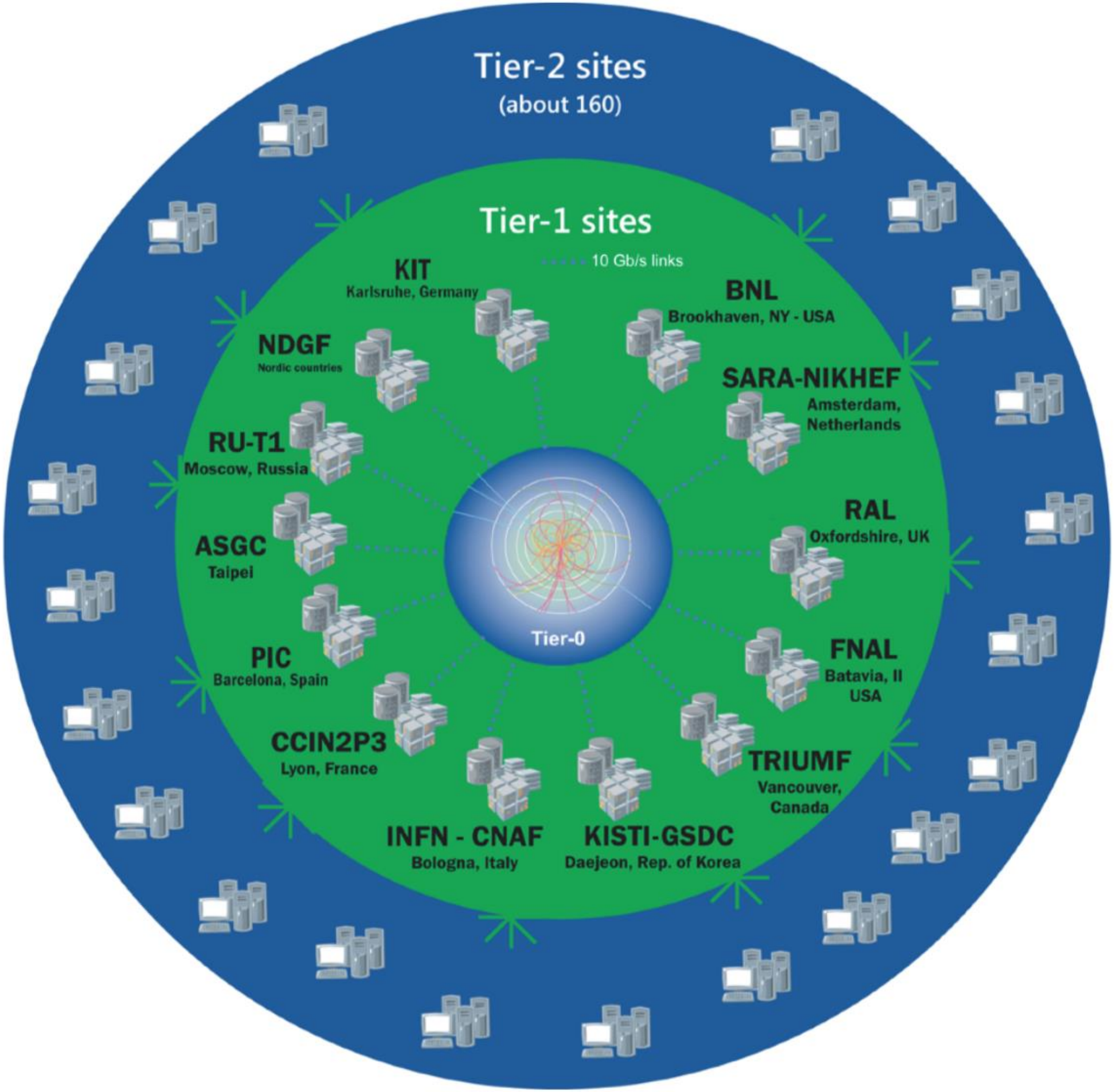
Whole Collaboration (WLCG) activities

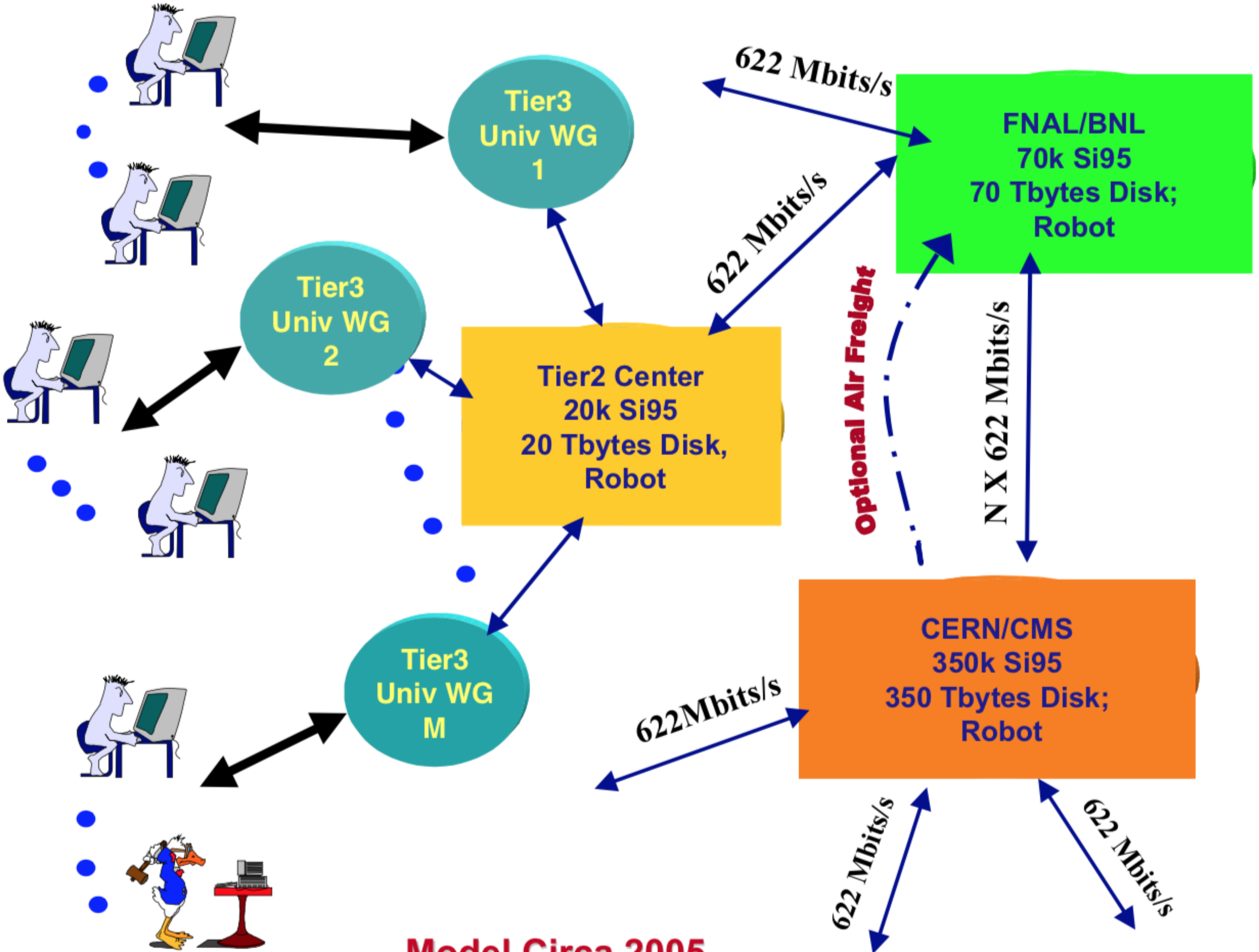
- Grid Deployment Board (GDB)
  - Monthly meeting of site and experiment representatives about current status/developments, often preceded by a topical pre-GDB
  - Used to be very operational – much more focussed on evolution
- WLCG "daily" operations meetings
  - Now just once a week. What is wrong at the moment, with reports by experiments and larger sites
- WLCG Operations Coordination
  Monthly meeting to take a longer view of operations
  - Various task forces and working groups spawned by above activities to look at particular topics
- All meetings use video conferencing, and usually a meeting room at CERN and maybe larger sites

Summary

- These examples are drawn from WLCG – but the functions described are (mostly) required in any e-Infrastructure
- Some functions may move between the eInfrastructure and the platform

- Uses EGI and OSG infrastructures underneath
  - In practice experiments talk to sites directly through WLCG structures (and EGI ticket system)
- Common grid services at sites became less important than central services run by experiments
- Human-to-human systems proved much more important expected
  - Ticket systems, regular ops meetings, conferences
  - Automation only gets you so far


Tier-2 sites
(about 160)
Tier-1 sites
10 Gb/s links
KIT
Karlsruhe, Germany
BNL
Brookhaven, NY - USA
NDGF
Nordic countries
SARA-NIKHEF
Amsterdam, Netherlands
RU-T1
Moscow, Russia
RAL
Oxfordshire, UK
ASGC
Taipei
Tier-0
FNAL
Batavia, Il USA
PIC
Barcelona, Spain
CCIN2P3
Lyon, France
TRIUMF
Vancouver, Canada
INFN - CNAF
Bologna, Italy
KISTI-GSDC
Daejeon, Rep. of Korea



Tier3
Univ WG
1

Tier3
Univ WG
2

Tier3
Univ WG
M

Tier2 Center
20k Si95
20 Tbytes Disk,
Robot

FNAL/BNL
70k Si95
70 Tbytes Disk;
Robot

CERN/CMS
350k Si95
350 Tbytes Disk;
Robot

622 Mbits/s

622 Mbits/s

Optional Air Freight

N X 622 Mbits/s

622Mbits/s

622 Mbits/s

622 Mbits/s

Model Circa 2005