



Enabling Grids for E-science

JRA1

Middleware Re-engineering

Frédéric Hemmer, JRA1 Manager, CERN
On behalf of JRA1

EGEE 2nd Review
December 6-7, 2007
CERN, Switzerland

www.eu-egee.org
www.glite.org



- **Processes and Releases**
- **Subsystems**
 - Features
 - Deployment Status
 - Short Term Plans
- **Testing Status**
- **Metrics**
- **Summary**

- **Architecture Definition**
 - Based on Design Team work
 - Associated implementation work plan
 - Design description of Service defined in the Architecture document
 - Really is a definition of interfaces
 - Yearly cycle

- **Implementation Work plan**
 - Prototype testbed deployment for early feedback
 - Progress tracked monthly at the EMT

- **EMT defines release contents**
 - Based on work plan progress
 - Based on essential items
 - So far mainly for HTCondor and BioMed
 - Decide on target sites
 - Taking into account enough time for integration

- **Integration produces Release Candidate based on received tags**
 - Smoke Test, Deployment Modules, Configuration
 - Iterate with developers

- **Testing Team**
 - Test Release candidates on a distributed testbed (CERN, RRZN, Hannover, Imperial College)
 - Raise Critical bugs
 - Iterate with developers
- **Once Release Candidate passed functions**
 - Team produces documentation, releases and final packaging
 - Announce the release on the glite Web site and the glite-discuss mailing list.

The focus has been on essential (simple) services and defect fixing – e.g. FTS, R-GMA, VOMS

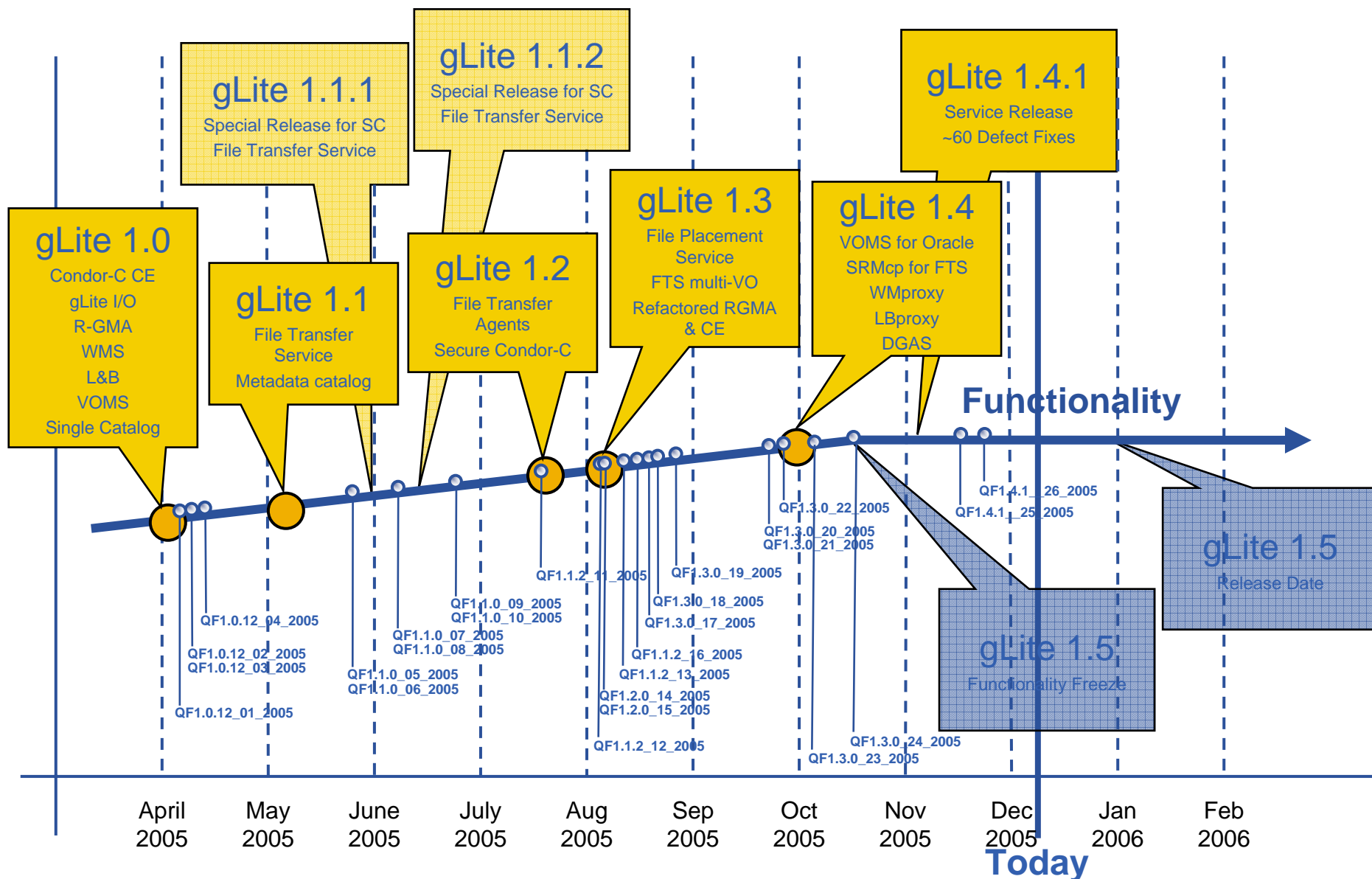
Deployment on Pre-production Service and/or Service Challenges

- Feedback from larger number of sites and different level of competence
- Raise Critical bugs as needed
- Critical bugs fixed with Quick Fixes when possible

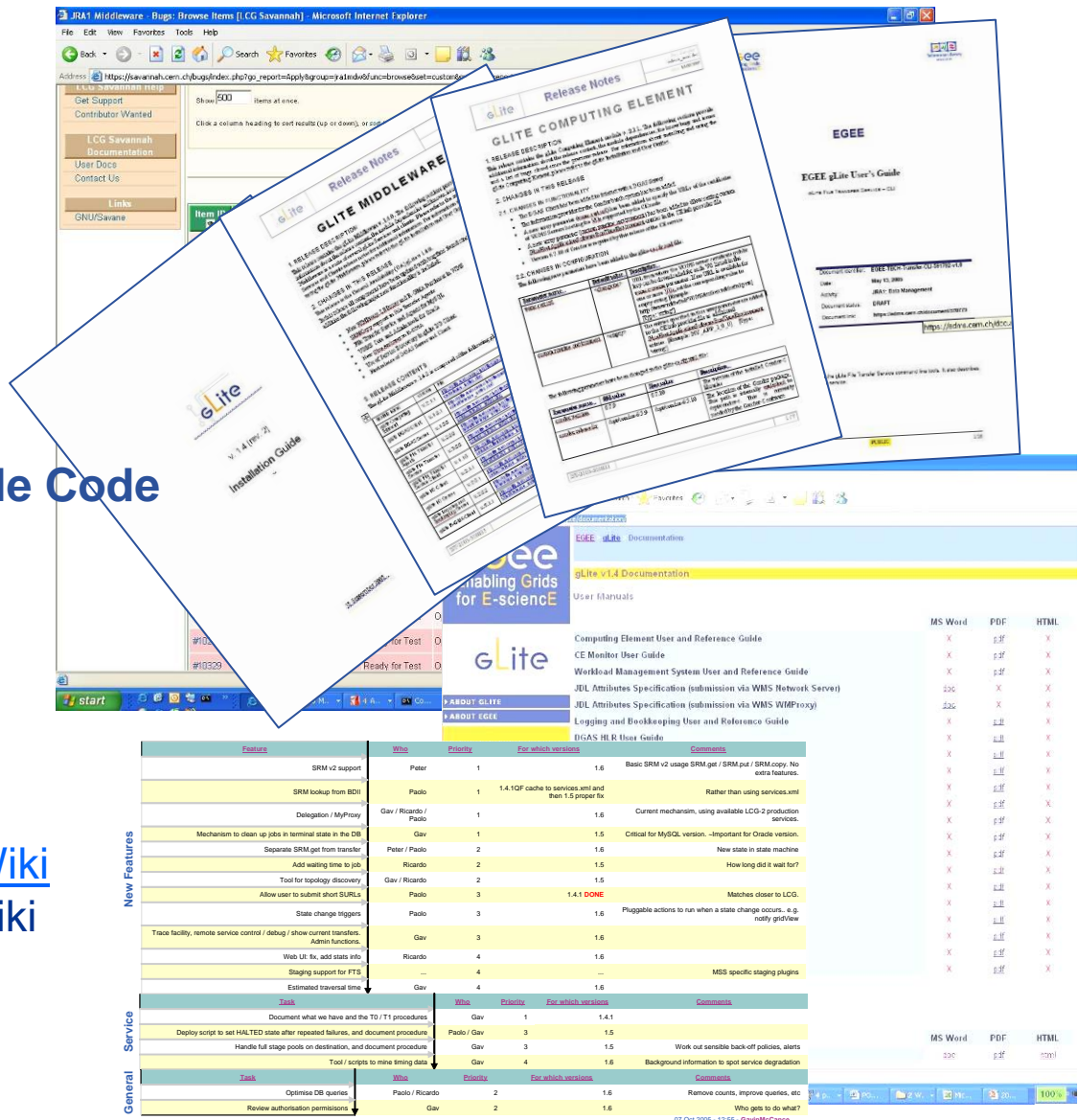
Deployment on Production of selected set of Services

- Based on the needs (deployment, applications)
- Today FTS clients, R-GMA, VOMS

gLite Releases and Planning



- [Installation Guide](#)
- [Release Notes](#)
 - [General](#)
 - [Individual Components](#)
- [User Manuals](#)
 - [With Quick Guide sections](#)
- [CLI Man pages](#)
- [API's and WSDL](#)
- [Beginners Guide and Sample Code](#)
- [Bug Tracking System](#)
- [Tutorials](#)
- [Mailing Lists](#)
 - [gLite-discuss](#)
 - [Pre-Production Service](#)
- [Other](#)
 - [Data Management \(FTS\) Wiki](#)
 - [Pre-Production Services Wiki](#)
 - [Public](#) and [Private](#)
 - [Presentations](#)
 - [Posters](#)

Task	Who	Priority	For which versions	Comments
SRM v2 support	Peter	1	1.6	Basic SRM v2 usage SRM get / SRM put / SRM copy. No extra features.
SRM lookup from BDI	Paolo	1	1.4.10Q cache to services.xml and then 1.5 proper fix	Rather than using services.xml
Delegation / MyProxy	Gav / Ricardo / Paolo	1	1.6	Current mechanism, using available LCG-2 production services.
Mechanism to clean up jobs in terminal state in the DB	Gav	1	1.5	Critical for MySQL version. -Important for Oracle version.
Separate SRM get from transfer	Peter / Paolo	2	1.6	New state in state machine
Add waiting time to job	Ricardo	2	1.5	How long did it wait for?
Tool for topology discovery	Gav / Ricardo	2	1.5	1.5
Allow user to submit short SURLS	Paolo	3	1.4.1 DONE	Matches closer to LCG.
State change triggers	Paolo	3	1.6	Pluggable actions to run when a state change occurs. e.g notify gridview
Trace facility, remote service control / debug / show current transfers, Admin functions	Gav	3	1.6	1.6
Web UI: fix, add stats info	Ricardo	4	1.6	1.6
Staging support for FTS	—	4	—	MSS specific staging plugins
Estimated traversal time	Gav	4	1.6	1.6

Task	Who	Priority	For which versions	Comments
Document what we have and the T0 / T1 procedures	Gav	1	1.4.1	1.4.1
Deploy script to set HALTED state after repeated failures, and document procedure	Paolo / Gav	3	1.5	1.5
Handle full stage pools on destination, and document procedure	Gav	3	1.5	Work out sensible back-off policies, alerts
Tool / scripts to mine timing data	Gav	4	1.6	Background information to spot service degradation

Task	Who	Priority	For which versions	Comments
Optimise DB queries	Paolo / Ricardo	2	1.6	Remove counts, improve queries, etc
Review authorisation permissions	Gav	2	1.6	Who gets to do what?

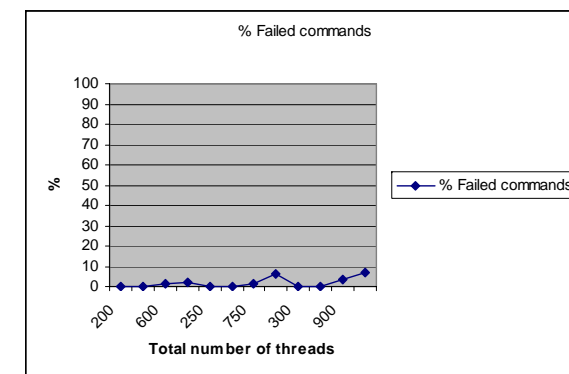
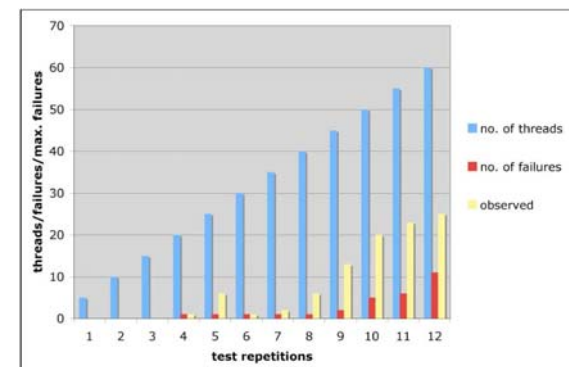
- **Different strategies depending on potential users**
 - FTS
 - Working directly with (part of) Service Challenges team
 - Daily meetings
 - R-GMA
 - Daily reports for Job Monitoring/GridFTP log monitoring
 - Weekly meetings
 - HEP Task Forces
 - Helping experiments to use gLite from their frameworks
 - Assessing functionality performance of the components they are interested in
 - Sometimes evaluating unreleased/new components/features
 - *From developments environments*
 - Weekly (mostly) meetings
 - BioMedical applications
 - Focused Data Management exercise
 - Weekly phone calls
 - Respective developers working hand-in-hand
 - DILIGENT project
 - Relatively loose coupling
 - 10 meetings, but very effective collaboration
 - Results reported at and the EGEE 4th conference
 - Operations/PPS
 - Weekly meeting to address most critical defects
- **Side effects**
 - Sometimes the bypass of the formal process
 - E.g. install rpm's instead of using configuration scripts and following installation guide/release notes
 - Example: FTS
 - <https://uimon.cern.ch/twiki/bin/view/LCG/FtsServerInstall13>
 - Does not always help improving the installation but helps in deploying quickly
 - *Defects are usually reported upon deployment on the PPS*
 - Unreleased components are sometimes exposed as fully functional
 - While having only been installed in one place
 - Potentially causing frustration for early users
 - However helps in defining useful functionality and improve performance
 - *E.g. factor of 12 in improving matchmaking has been identified*
 - Costs significant efforts in JRA1

- **Most Services rely on GSI and MyProxy**
 - Still using well understood GT2 implementation
 - Authentication can be expensive
 - Several subsystems provide bulk operations

- **VOMS**
 - Manages VO Membership
 - Provides support for Groups and Roles
 - Support for MySQL and Oracle DB backend
 - Included in the VDT
 - Support for many other clients than SLC3

- **VOMS Admin**
 - Support for Oracle and MySQL back ends
 - VOMS ADMIN (Oracle) still problematic
 - Support issues clarified

- **Deployed on the Production Infrastructure**
 - Interfaced with OSG's VOMSRS



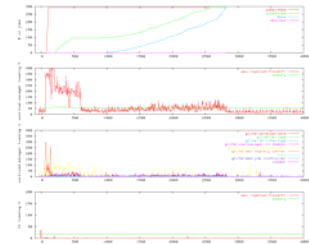
- **Logging and Bookkeeping**
 - Tracks jobs during their lifetime (in terms of events)
 - LB Proxy
 - Provides faster, synchronous and more efficient access to LB services to WMS services
 - Support for “CE reputability ranking”
 - Maintains recent statistics of job failures at CE’s
 - Feeds back to WMS to aid planning
 - Considered to be included in the VDT

- **CE**
 - Service representing a computing resource
 - CE moving towards a VO based local scheduler
 - BLAH
 - More efficient parsing of log files (these can be left residing on a remote machine)
 - Support for hold and resume in BLAH
 - To be used e.g. to put a job on hold, waiting for e.g. the staging of the input data
 - Condor-C GSI enabled
 - CEMon
 - Major reengineering of this service
 - Better support for the pull mode; More efficient handling of CEMon reporting
 - Security support
 - Possibility to handle also other data
 - E.g. a GridIce plugin for CEMon implemented
 - Included in VDT and used in OSG for resource selection

- **GPbox**
 - XACML-based policy maintainer, parser and enforcer.
 - Can be used for authorisation checks at various levels.

- **WMS**

- Backward compatibility with LCG-2
- WMPProxy
 - Web service interface to the WMS
 - Allows support of bulk submissions and jobs with shared sandboxes
- Support for shallow resubmission
 - Resubmission happens in case of failure only when the job didn't start running - Only one instance of the user job can run.
- Support for MPI job even if the file system is not shared between CE and WNs
- Support of R-GMA as resource information repository to be used in the matchmaking besides bdII and CEMon
- Support for DLI and StorageIndex Data management interfaces
- Support for execution of all DAG nodes within a single CE - chosen by user or by the WMS matchmaker
- Support for file peeking to access files during the execution of the job
- Initial integration with G-Pbox - considering simple AuthZ policies
- Initial support for pilot job
 - Pilot job which "prepare" the execution environment and then get and execute the actual user job



- **DGAS Accounting**

- Accumulates Grid accounting information about the usage of Grid resources by users / groups (e.g. VOs) for billing and scheduling policies
- CEs can be instrumented with proper sensors to measure the resources used

- **Job provenance**

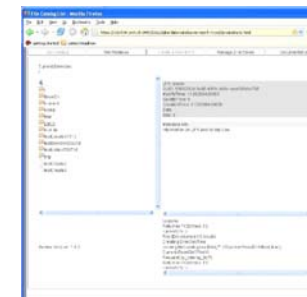
- Long term job information storage
- Useful for debugging, post-mortem analysis, comparison of job executions in different environments
- Useful for statistical analysis

- **WMS, CE, LB, DGAS are considered for inclusion in the next LCG-2 release**

- Currently deployed on the Pre-production service and DILIGENT testbed
- Tested on many private instances

- **FiReMan catalog**

- Resolves logical filenames (LFN) to physical location of files (URL understood by SRM) and storage elements
- Oracle and MySQL versions available
- Secure services, using VOMS groups, ACL support for DNs
- Full set of Command Line tools
- Simple API for C/C++ wrapping a lot of the complexity for easy usage
- Attribute support
- Symbolic link support
- Exposing ServiceIndex and DLI (for matchmaking)
- **Separate catalog available as a keystore for data encryption ('Hydra')**
- Deployed on the Pre-Production Service and DILIGENT testbed

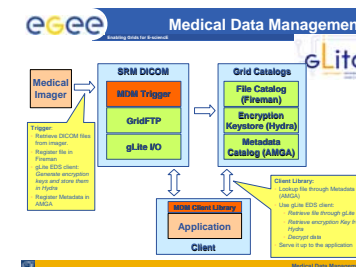


- **gLite I/O**

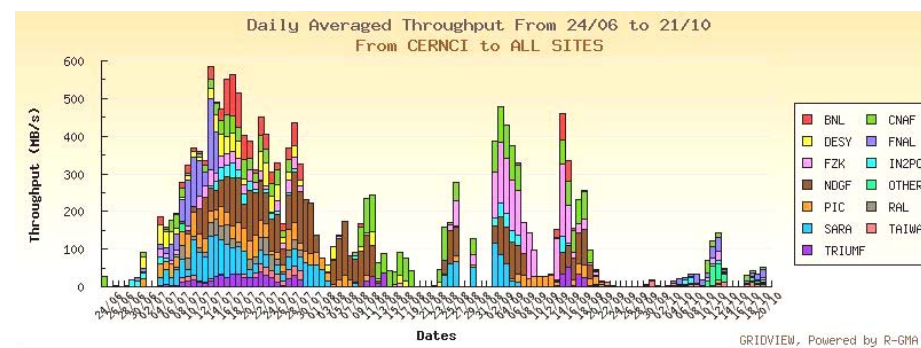
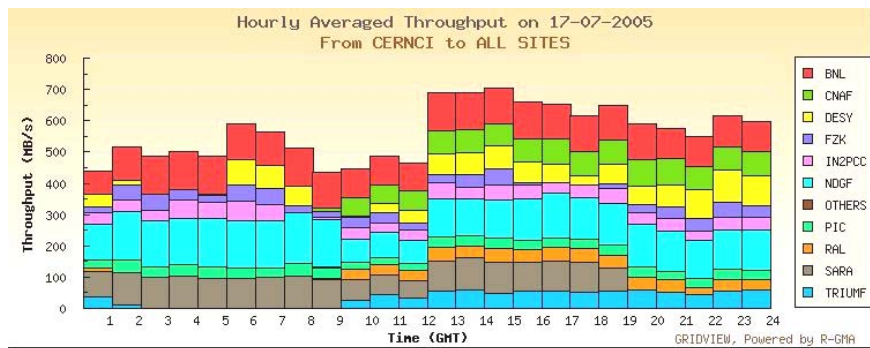
- Posix-like access to Grid files
- Castor, dCache and DPM support
- Added a remove method to be able to delete files
- Changed the configuration to match all other CLI configuration to service-discovery
- Improved error reporting
- Has been used for the BioMedical Demo
 - **Encryption and DICOM SRM**
- Deployed on the Pre-Production Service and the DILIGENT testbed

- **AMGA MetaData Catalog**

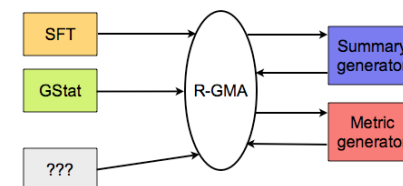
- NA4 contribution
 - Result of JRA1 & NA4 prototyping together with PTF assessment
 - Used by the LHCb experiment
 - Has been used for the BioMedical Demo



- **Reliable file transfer**
- **Full scalable implementation**
 - Java Web Service front-end, C++ Agents, Oracle or MySQL database support
 - Support for Channel, Site and VO management
 - Interfaces for management and statistics monitoring
 - Gsiftp, SRM and SRM-copy support
- **Has been in use by the Service Challenges for the last 5 months.**
 - Evolved together with the Service Challenges Team
 - Daily meetings
- **FTS evolved over summer to include**
 - Support for MySQL and Oracle
 - Multi-VO support
 - GridFTP and SRM copy support
 - MyProxy server as a CLI argument
 - Many small changes/optimizations revealed by SC3 usage
- **FTS workshop with LHC experiments on November 16**
 - Issues, Feedback and short term plans

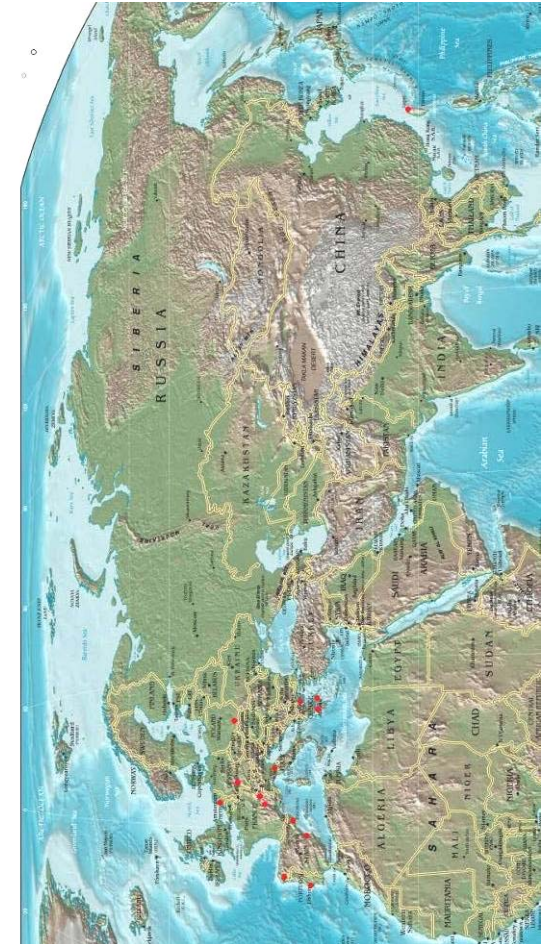


- **R-GMA**
 - Essentially bug fixes & consolidation
 - Merging LCG & gLite code base
 - Secure version
- **Used in production as monitoring data aggregator**
 - Job status published from L&B every 5 minutes
- **Interfaced from the WMS ISM**
- **Service Discovery**
 - Was not part of gLite 1.0
 - An interface has been defined and implemented for 3 back-ends
 - R-GMA
 - BDII
 - Configuration File
 - Command Line tool end users
 - Used by WMS and Data Management clients
- **Services still using BDII as the Information System**
 - Pre-Production Service has started to use R-GMA



- **Computing Element**
 - Gatekeeper, WSS (Globus)
 - Condor-C (Condor)
 - BLAH (EGEE)
 - CE Monitor (EGEE)
 - Local batch system (PBS, LSF, Condor, SGE, BQS)
 - DGAS (EDG/EGEE)
- **Workload Management**
 - WMS (EDG/EGEE)
 - Logging and bookkeeping (EDG/EGEE)
 - Condor-C (Condor)
 - Job Provenance (EGEE)
- **Storage Element**
 - File Transfer/Placement (EGEE)
 - glite-I/O (AliEn)
 - GridFTP (Globus)
 - SRM: Castor (CERN), dCache (FNAL, DESY), DPM (CERN), other SRMs
- **Catalog**
 - File and Replica Catalog (EGEE)
 - Metadata Catalog (EGEE/NA4)
- **Information and Monitoring**
 - R-GMA (EDG/EGEE)
 - Service Discovery (EGEE)
 - BDII
- **Security**
 - VOMS (DataTAG, EDG/EGEE)
 - GSI (Globus)
 - LCAS/LCMAPS (EDG/EGEE)
 - Authorization for C and Java based (web) services (EDG/EGEE/Globus)
 - GPBox (EGEE)
 - WSS (Globus)
- **UI (Various)**

- **Production**
 - FTS
 - R-GMA (Monitoring & Accounting Data Aggregation)
 - VOMS/VOMS Admin
- **Preproduction Service**
 - 14 sites
 - CERN, CNAF, PIC CE's are connected to the production worker nodes
 - ~ 1M Jobs submitted
 - FTS, WMS/LB/CE, FireMan, gLite I/O (DPM, Castor), R-GMA
- **Others**
 - DILIGENT has deployed a number of those services as well



- **Revision of the Architecture, Design and Work plan documents**
 - <https://edms.cern.ch/document/594698/>
 - <https://edms.cern.ch/document/573493/>
 - <https://edms.cern.ch/document/606574/>
- **Advanced Reservation**
 - Architecture defined
 - <https://edms.cern.ch/file/508055/2-2/EGEE-JRA1-AR-508055-v2-2.pdf>
 - Integration with WMS prototyped
 - <http://agenda.cern.ch/askArchive.php?base=agenda&categ=a052420&id=a052420s3t5/transparencies>
- **OMII evaluation**
 - <where is the URL>
- **GT4 evaluation**
 - <https://edms.cern.ch/document/672123/>
- **Interfacing of ProActive to gLite**
 - Demonstrated at the 2nd Grid [PlugTests event](#)
 - Hands on with gLite
 - <http://agenda.cern.ch/askArchive.php?base=agenda&categ=a055911&id=a055911s1t9/transparencies>
- **Development of a new Web Services based CE**
 - CREAM: <http://grid.pd.infn.it/cream/field.php>
- **Tutorials and Schools**
 - gLite Installation & Certification Training Event (CERN, Switzerland)
 - <http://agenda.cern.ch/fullAgenda.php?ida=a053710>
 - GRID'05 EGEE Summer School (Budapest, Hungary)
 - <http://www.egee.hu/grid05/>
 - GGF International Grid Summer School
 - <http://www.dma.unina.it/~murli/GridSummerSchool2005/>
 - CERN School of Computing 2005
 - Grid Track
 - <https://edms.cern.ch/document/605400/1>
- **WSS**
 - Prototype of the integration of the Globus Works Space Services
 - Joint Globus/Condor/EGEE paper submitted at <Erwin please help>
- **glogin**
 - CrossGrid's glogin has been demonstrated with gLite
- **PGrade**
 - PGrade Portal interfaced to gLite
- **Continuously managed Prototype testbed**

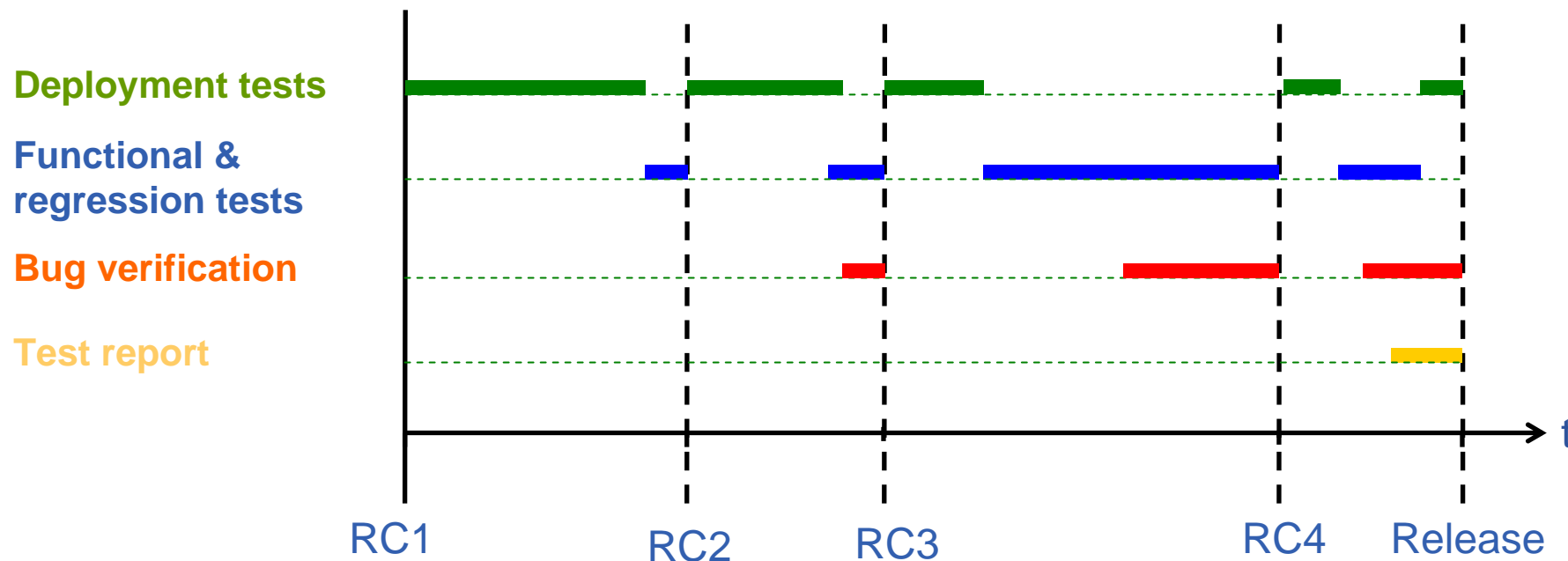
- **Distributed testbed: 3 sites**

- CERN
- Imperial College
- RRZN Uni Hannover
- (RAL and NIKHEF stopped contributing)



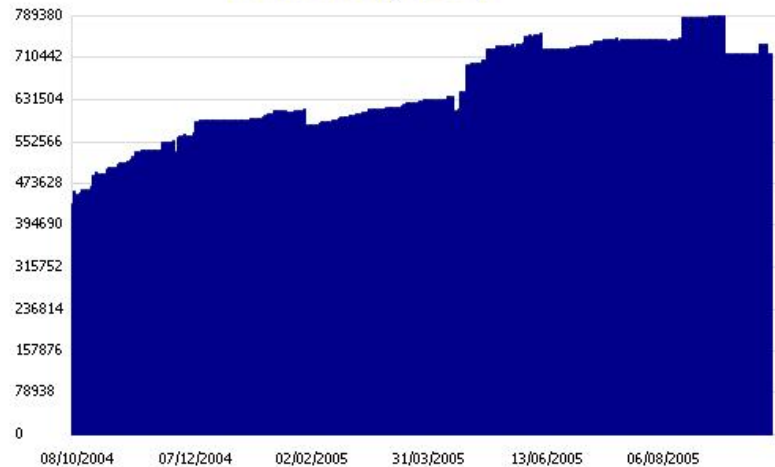
Imperial College
London

R|R|Z|N|

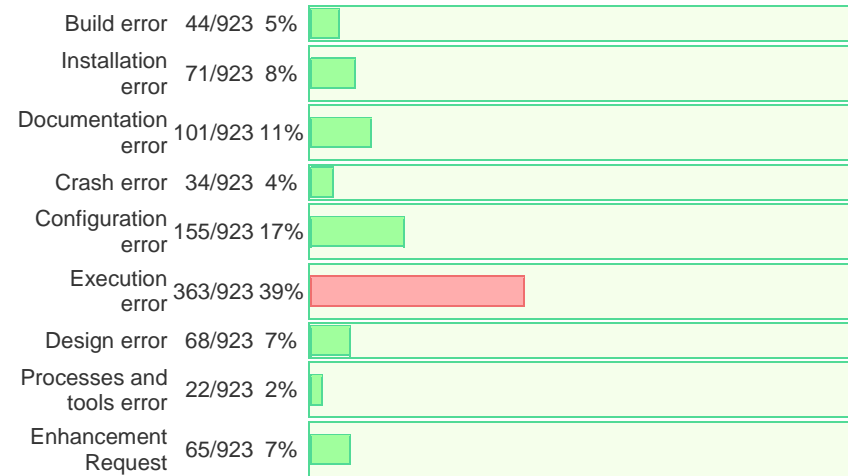


	Deployment	Test suite	Stress	Bug verification	Test report
CE	✓	✓	✓	✓	✓
DGAS	✓	✗	✗	✗	✓
Fireman	✓	✓	✗	✗	✓
FTS	✓	✓	✗	✗	✓
I/O	✓	✓	✓	✓	✓
LB	✓	in preparation	✗	✗	✗
R-GMA	✓	✓	✓	✗	✓
SD	✓	✗	✗	✗	✗
VOMS	✓	✓	✓	✗	✓
WMS	✓	✓	✓	✗	✓

Code Size (SLOC)

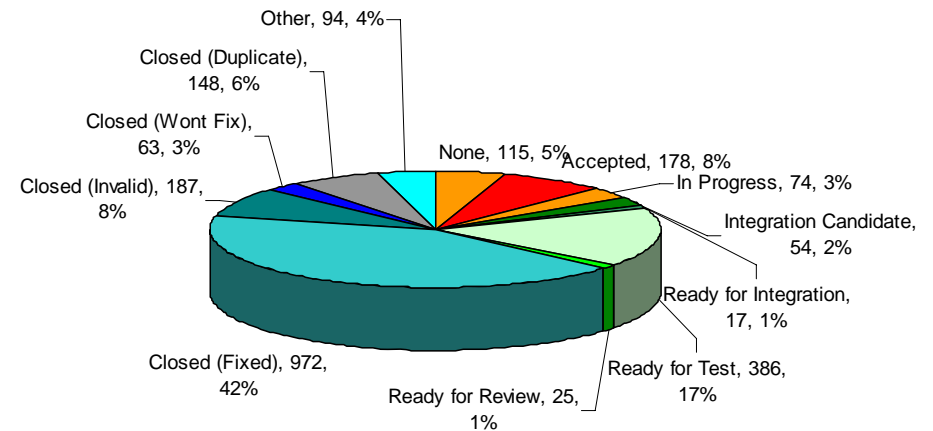
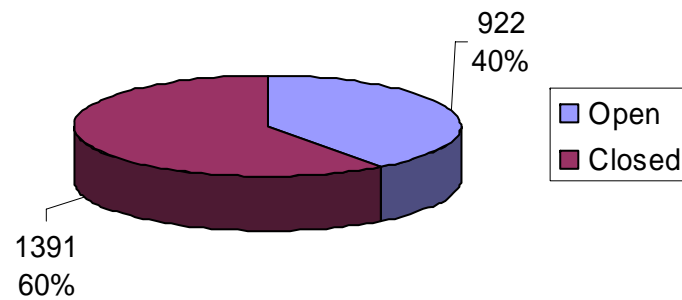


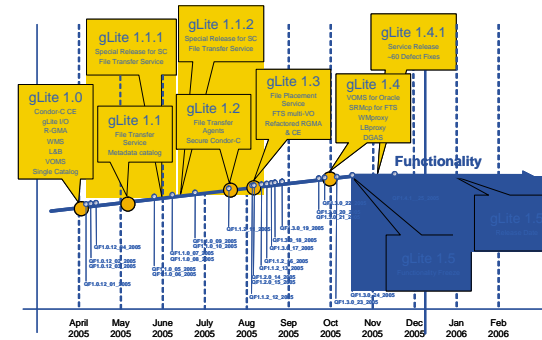
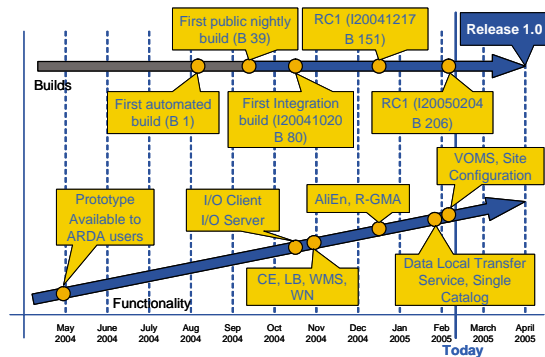
Copyright (c) 2004 EGEE



Bugs by status

Total open/closed bugs (Total = 2313)

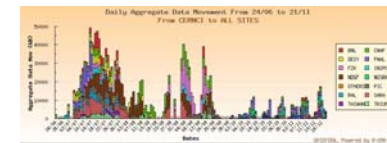




- Total complete builds done: 208
- Number of subsystems: 12
- Number of CVS modules: 343
- FTS: Preview in Release 1.0
- Simple Metadata Catalog
- VOMS on MySQL
- Re-engineered WMS
- Manual component Testing

- Total complete builds: 641, 236 (HEAD)
- Number of subsystems: 21
- Number of CVS modules: 454

- FTS
- AMGA Catalog agreed by PTF
- VOMS on Oracle
- Web Services bulk job submission
- Testing semi-automated, reports



- **Design Team**
 - Small international group of competent people understanding each other
 - Task Queue, Condor-C integration in WMS, Storage Index, Data & Job Management security models, WSS, future VO scheduler, etc...
- **VDT**
 - VOMS, LB, CEMon are in or scheduled (using NMI processes)
- **Collaboration in particular with University of Wisconsin/Madison**
 - Not only Condor, also NMI, relations with OSG, etc..
 - Significant (not reported) manpower dedicated to gLite related issues
 - Has been instrumental in the ETICS project proposal

Standardisation Body	Area	Working Group	Contributor	Role
GGF	Architecture	OGSA-WG, Resource Management Design Team	Sergio Androozzi	External contributor
GGF	Architecture	OGSA-WG	Abdeslem Djaoui	Member
DMTF	Resource Information Modeling	Core and Devices WG	Sergio Androozzi	External contributor
		CRM	Massimo Sgaravatto, Luigi Zangrando, Erwin Laure, Miron Livny, Ian Foster,...	Members
GGF	Management	UR-WG and GESA-WG	Andrea Guarise, Rosario Piro	External contributors
GGF	Security	OGSA-AUTH	Vincenzo Ciaschini	External contributor
GGF	Data	GSM-WG	Peter Kunszt	Chair
GGF	Data	OGSA-D-WG	Peter Kunszt	Member
GGF	Data	OREP-WG	Peter Kunszt	External contributor
GGF	Data	INFOD-WG	Abdeslem Djaoui	Member
GGF	Data	INFOD-WG	Steve Fisher	Secretary

Accomplishments

- **gLite “brand name”**
- **Services offering significant (basic) functionalities required by Applications**
 - Components included in VDT
- **Collaboration with DILIGENT**
- **Collaboration with US**
- **First ever storage encryption solution for BioMedical demonstrated**

Issues

- **Complex software suite**
 - Many fixes and patches
- **Integration & Testing understaffed**
- **Multi-platform support**
- **Multiple reporting lines**
- **Integration & Testing perceived as slowing down the process**

- **Continue Defect fixing as required**
- **Complete gLite 1.5 release**
 - Including Documentation, Installation Guide, Release Notes, Testing reports
 - Forming DJRA1.6 deliverable
- **Converge with SA1 the LCG and gLite middleware releases to a single distribution called gLite**
 - Being discussed with Operations
 - EGEE-II startup timeframe
 - Tentatively named gLite 3.0

- **Reviewers comments on**
 - gLite should be a stepping stone towards more robust standards based Grid infrastructure; Collaboration with other M/W R&D initiative to select additional components
 - Continuous participation in GGF, collaboration with OSG, NorduGrid, CRM Initiative
 - gLite use many external dependencies coming from other Middleware initiatives
 - Collaboration on interoperation of Condor-C and GT4
 - Prototype of interoperation of the CE and Globus WSS
 - gLite make use of SRM's developed by other initiatives
 - JRA1 has proposed Unicore and Shibboleth interoperation in EGEE-II
 - Co-scheduling timetable should be announced, but no implementation before end of EGEE
 - This is still pretty much R&D
 - JRA1 has proposed an Architecture for Advanced Reservation
 - JRA1 has prototyped Advanced Reservation of Storage and Network Resources

- **gLite releases have been produced**
 - Tested, Documented, with Installation and Release notes
 - Subsystems used on
 - Service Challenges
 - Pre-Production Services
 - Production Service
 - Some components included in the VDT
 - And by other communities (e.g. DILIGENT)
 - Special effort for helping early adopters in place
- **gLite processes are in place**
 - Closely monitored by various bodies
 - Hiding many technical problems to the end user
- **gLite is more than just software, it also about**
 - Processes, Tools and Documentation
 - International Collaboration



Lightweight Middleware for
Grid Computing

www.glite.org