TCB Requirements

Report on requirements submitted and reviewed by EGI communities

Michel Drescher (Michel.Drescher@egi.eu)

01 November 2012

Version 2

TABLE OF CONTENTS

1 Introduction

This technical report provides an overview of requirements managed at the TCB level that were submitted or reviewed by EGI communities after requesting clarification, in accordance with [TCBReq]. The information given herein is intended as material for discussion at the upcoming 14th TCB meeting [TCB-14] for the participants to decide how to proceed with each of the requirements discussed in this document.

2 Submitted and reviewed requirements

At the time of writing this document (26 October 2012), more than 30 requirements were submitted and reviewed for re-evaluation by the TCB. A swift glance at these requirements showed that TCB status and information was not accurate, and that the TCB requirements dashboard [TCBDashb] both displayed requirements that shouldn't have been shown, and was not listing requirements that should have been included.

To get an accurate overview of requirements tracked in the dashboard (thus also adjusting the submitted and reviewed columns), all requirements that have been changed in the last 6 months were assessed and adjusted appropriately, where obvious comments and attached artefacts indicated such an action. This included adjusting the TCB status and the priority of the requirements as indicated in the original requirements collection submissions. This work has been conducted collaboratively with Gergely Sipos and Peter Solagna solving TCB Action 13/01[1]. The remaining eleven submitted and reviewed requirements are summarised in Table 1 below.

EGI Capability		ID	Requirement	Prio	Type
Info.	Information Discovery	3279	Top-BDII must scale with number of sites	5	nf
		3326	Automatic configuration of CE capacity parameters	3	nf
Storage	Storage Management	3230	6.4 Data lifetime management	5	f
		881	Requirement for VO renaming / migration	3	f
		1187	Some new features for Data Mgmt and Sharing	3	f
		2538	Quick identification of users affected in an SE intervention	2	f
Compute		2652	Support multi-VO services, association to multiple topBDIIs	2	f
Cross-cutting		2563	Mandatory variables in configuration files should be clearly identified	4	nf
		1384	DOS attacks	3	nf
		1381	Improving automatic services configuration	3	nf
		1203	Uniform API	0	f

Table 1: Overview of submitted and reviewed requirements (f = functional, nf = non-functional requirement)

[1] As a result of this, work, the search queries implemented in the Wiki will be simplified. Meanwhile an RT based dashboard for TCB members is available at http://go.egi.eu/TCBRequirements.

Grouping the requirements by related EGI Capability and ordering them by priority shows that the two most demanded categories are requirements *cross-cutting* EGI Capabilities, and *Storage Management* functions. This is enforced by examining the relative demand as the product of the number of requirements and the sum of priorities assigned, as indicated in Table 2 below.

EGI Capability	# Requirements	Sum of priority points	Relative demand
Storage	4	13	52
Cross-cutting	4	10	40
Information	2	8	16
Compute	1	2	2

Table 2: Relative EGI demand for improvements expressed as the product of columns 2 and 3

The following subsections provide an analysis of the eleven requirements, providing suggestions for how to proceed with each requirement individually.

2.1 Storage Management requirements

Four requirements deal with functional enhancement requests for storage management services: Two requirements (#881, #2538) relate to situations where owners may need to take care of their files residing on a storage location before a management action on the pertinent storage service can take place. One requirement (#3230) describes the need of a capability to implement and manage data retention policies. The remaining requirement (#1187) describes distributed data integration requirements specifically for the UNICORE platform. The following subsections discuss the individual requirements.

2.1.1 Data lifetime management (#3230)

Analysis:

This requirement is a container for several related community specific requirements. It describes a feature allowing communities to implement and manage data retention policies on the storage services they are using. Originally referring to two VO-specific requirements (#2024 / LSGC VRC, and #2877 / WeNMR VRC), WeNMR has accepted the solution that was provided to them earlier.

The VOs affiliated with both VRCs use storage services provided by (alphabetically listed) CASTOR, dCache, DPM, StoRM and.

Recommendation:

The SRMv2.2 specification is a standardised interface definition published 2008 by the OGF as GFD.129 [GFD.129]. It is available online as a HTML version at [SRMv2.2].

The SRMv2.2 specification provides all the necessary functionality to implement a data retention policy on any storage service fully implementing GFD.129 (e.g. section 2.1 “srmReserveSpace”

operation). Further, all four storage services affected by this requirement are reported to fully implement the SRMv2.2 interface[2].

That said, the requested functionality is already implemented on a generic level, allowing customers to implement and enforce a wide variety of individual data retention policies. This allows for the following suggestions for proceeding with the requirement in order of preference):

1. **ENDORSE** – Providing data retention functionality may be considered a strategic capability for EGI, this be incorporated into the EGI Standards portfolio and technical roadmaps. Storage service technology providers will have to exercise their diligence ensuring that SRCv2.2 is fully implemented. A tertiary action may be to develop documentation/best practices on how to implement and enforce a data retention policy using SRMv2.2
2. **RETURN** – The necessary functionality is already present. LFC-specific enhancements described in the original requirements documentation (see [TCB-10]) should be pursued through GGUS enhancement requests.

2.1.2 Requirement for VO renaming / migration (#881)

Analysis:

The requirement is describing two related but different user stories, i.e. renaming a VO, and merging two (or more) VOs in to one.

The description indicates that, on a technical level, a "natural key" identifies a VO, instead of a surrogate technical identifier. This has implications on both managerial and technical levels as described in the requirement document. Both managerial use cases however are expected to occur within EGI and are considered essential for efficient VO-based data management. The current implementation, however, requires the same technical solution as described as "download-to-scratch-upload" in the original submission.

Recommendation:

Although both use cases described in the original requirement submission are related, they are different in the underlying technical implications: While renaming a VO ideally incurs one management action (change the name of the VO), merging several VOs into one will cause the change of ownership and affiliation of at least the data associated with the discontinued VO to the prevailing VO.

This difference indicates two different solutions for this requirement to be fulfilled. For example, if a technical identifier were used in the infrastructure level to identify a VO, this technical identifier may stay the same in a simple VO renaming operation (where the name would be used for display purposes only), implementing the first use case. To implement the second use case, the VO identifier (whether the name or a technical identifier) needs to change one way or the other on affected storage elements. On the other hand, SRMv2.2 ([GFD.129]) provides means by which the second use case (VO merge or migration) may already be satisfied, but not known how to use the SRM functionality properly.

Since the potential implications can be fairly disruptive, the recommendation is to:

- **CLARIFY** – request further clarification to the EGI community to split this requirement into two individual requirements reflecting these different (a) Implement technical surrogate VO

[2] See https://sdm.lbl.gov/srm-wg/doc/SRM.v2.2.html#_Toc241633135 referring to Appendix I in [SRMv2.2]

identifier across EGI, and (b) Implement storage service operations to change VO ownership of data. For requirement (b) SRM features may be further explored before considering re-submission to the TCB.

2.1.3 Some new features for Data Management and Sharing (#1187)

Analysis:

This requirement describes how to provide distributed data sharing features and services for the UNICORE platform.

The description of the actual requirements remains unclear, e.g. whether to integrate with existing solutions.

Recommendation:

Two alternative ways of proceeding with this requirement are possible (in order of preference)

1. **RETURN** – The requirement is clearly scoped around the UNICORE platform hence this requirement should be pursued through GGUS enhancement requests.
2. **CLARIFY** – Further clarify to which extent this requirement may affect other services offered in EGI, and appropriately amend the requirement description before re-submitting it to the TCB.

2.1.4 Quick identification of users affected in an SE intervention (#2538)

Analysis:

This requirement describes the use case of decommissioning a storage element, and derives from this the requested feature of providing a single command to retrieve the owning users of all files affected by a storage element management action.

The description already states that although the functionality is satisfied by the existing implementations, it is "inconvenient" and should be optimised by a single command.

On the other hand, the provided text describes a fairly specific way of managing SE interventions that may not be implemented everywhere across EGI.

Recommendation:

The requirement description already states that a solution exists. Alternative ways of managing such SE intervention may involve active use of SRM commands (e.g. srmGetSpaceTokens, srmGetSpaceMetaData) in a variety of ways specific to EGI research communities. Therefore the following is recommended:

- **RETURN** – There is sufficient functionality available to satisfy this requirement on an infrastructure level, allowing research communities to implement a variety of managerial processes on top of them. In case a "best practice" emerges across research communities, these may be expressed as LFC-related enhancement requests through GGUS.

2.2 Cross-cutting requirements

Three requirements describe non-functional properties for services deployed in EGI. The remaining requirement describes a uniform API across EGI services.

2.2.1 Mandatory variables in configuration files should be clearly identified (#2563)

Analysis:

This requirement describes the need to clearly identify configuration variables that must be set to appropriate values before the pertinent service will deliver its intended function. Optional variables should be set to default values. It is actually already the description of how this requirement may be implemented.

The actual requirement that is masked by this description is to prevent services from offering their pertinent functionality unless it is properly contextualised (i.e. all mandatory configuration variables are set). Configuration files are a popular means to contextualise a service before it starts; however there are other models of managing/contextualising a service. While configuration files implement an "eager service contextualisation" model, services offering an administration interface for runtime configuration and management may implement a "lazy service contextualisation" model, where the service starts into a manageable state, allowing to set missing variables online, and switch to an operative state once all mandatory variables are set to meaningful values.

Irrespective of the actual contextualisation method used, it is important to minimise if not prevent configuration mistakes to avoid unnecessary effort spent on analysing service misbehaviours.

Recommendation:

Since services deployed in the EGI predominantly use the configuration file based contextualisation method the following is recommended:

- **ENDORSE** – EGI should endorse this requirement further clarifying/re-scoping it to prevent successful service startup until all mandatory configuration values are properly set to values that can be consumed by the pertinent service.

2.2.2 DOS attacks (#1384)

Analysis:

This requirement describes a scenario where properly authenticated users accidentally (i.e. without genuine intent) satiate a service's resources so that the service is incapable of accepting further requests.

This requirement is different from real Denial of Service (DoS) attacks in that the latter have the malignant(!) intention of rendering the targeted service in denial of service requests. Also, real life DoS attacks employ several spoofing technologies (e.g. identity, source location of attack) to mask the originator of the attack, whereas accidental resource satiation usually has all essential real information available.

As indicated in the history of the respective RT tracker item, the real requirement behind this is to implement meaningful, configurable resource consumption constraints.

Recommendation:

- CLARIFY – Request re-scoping of the requirement and clarification on details of resource consumption constraints are necessary.

2.2.3 Improving automatic service configuration (#1381)

Analysis:

Originally, the submission requested an automatic way of providing contextual information to YAIM. After re-scoping the request based on an initial feedback from EMI, it turned into a general automation of service configuration request – being effectively rejected by EMI as well[3].

Indeed, the requirement's scope is too general in its description. There are many different types of information necessary to configure services, e.g. performance-related parameters, service integration parameters (e.g. remote service endpoint addresses), and many more.

As such, this requirement is related to #2563 (see section 2.2.1 above) and #3326 (see section 2.3.2 below). In a larger picture, however, this requirement touches issues in the technical architecture of a distributed services platform, for example dynamic services registration and discovery, and (automated) platform bootstrap; and how it interacts with the EGI Core Infrastructure Platform.

Recommendation:

In the current form, this requirement is not yet suitable for digestion by Technology Providers and its implications are beyond the intended scope. Therefore it is suggested to:

- **CLARIFY** – Within EGI, clarify the implications on the EGI e-Infrastructure strategy and technical architecture of a platform approach, before resubmitting to the TCB.

2.2.4 Uniform API (#1203)

Analysis:

The originating community would like to see a client API implemented/provided to control services. The EMI project has stated that a uniform API across all its services will not be possible, but has promised appropriate API documentation available with the delivery of EMI-3.

On the other hand, the SAGA project is providing a client side API that unifies APIs across various Grid middlewares such as gLite, UNICORE and Globus. In fact, proper SAGA connectors for Globus services will be provisioned for deployment in the EGI. Client-side SAGA technology is available for use, too.

Recommendation:

Since the SAGA project is providing suitable technology, the following is suggested:

- **ENDORSE** – The TCB should endorse this requirement and ask the SAGA project to develop a provisioning and deployment plan for its client side technology satisfying this requirement.

2.3 Information management requirements

2.3.1 The BDII must scale with the number of sites (#3279)

Analysis:

This is a scalability requirement for the top-BDII technology provided by the EMI project. However, the details provided in the original submission concern technical implementation issues instead of scalability requirements.

[3] see https://rt.egi.eu/rt/Ticket/Display.html?id=1381#txn-131154

Recommendation:

This requirement concerns top BDII only hence should be dealt with using appropriate GGUS enhancement requests:

- **RETURN** – submit scalability enhancements for top-BDII through GGUS after improving the details on scalability and performance requirements, not technical implementations.

2.3.2 Automatic configuration of CE capacity parameters (#3326)

Analysis:

This requirement requests the automation of CE capacity configuration for information providers, reducing manual intervention and error. Further, the description requires that CE information providers acquire HEP SPEC 06 information from the underlying batch sub system.

The requirement is missing some information. The description admits that currently the CE capacity is estimated – presumably, more of an approximation than estimation – of the actual value. It then is normalised for accounting purposes. The requirement refers to HEP SPEC 06 in an ambiguous way; it does not state whether the data is to be provided using the HEP SPEC 06 data model, or whether the HEP SPEC 06 benchmark is to be run regularly to actually acquire the cluster's capacity. A related EMI tracker item[4] raises similar concerns.

Recommendation:

In its current form, the requirement is difficult if not impossible to assess for potential implications. It is not clear whether and how the required information is stored in the underlying batch sub systems, and how this interacts with cluster reconfiguration processes – to obtain accurate values, the HEP SPEC 06 benchmark would have to be run on a reconfigured cluster that is in planned downtime etc.

- **CLARIFY** – Provide more details on how HEP SPEC 06 would be used in this context, how accurate the information has to be (estimation, approximation, accurate, …)

2.4 Compute requirements

2.4.1 Support for multi-VO services (#2652)

Analysis:

This requirement requests to extend WMS so that an administrator/operator can configure more than one top BDII to fetch information from. Several use cases are provided that underpin this requirement.

Recommendation:

This requirement has a clear and exclusive scope around WMS (although the RT ticket mentions potential scope for other higher-level services). It therefore should be tracked using GGUS enhancement requests assigned to the appropriate SU:

- **RETURN** – Pursue this WMS enhancement request using GGUS.

[4] https://savannah.cern.ch/task/?26578

3 Conclusions

Section 2 provides an analysis for each requirement the TCB is due taking a decision whether to endorse and proceed with for potential implementation, ask for more information, or to return the requirement to the originating community. Table 3 provides a summary of the recommended TCB decision.

ID	Requirement	Prio	§	Suggested state[5]
3230	6.4 Data lifetime management	5	2.1.1	Endorsed
881	Requirement for VO renaming / migration	3	2.1.2	In Clarification
1187	Some new features for Data Mgmt and Sharing	3	2.1.3	Returned
2538	Quick identification of users affected in an SE intervention	2	2.1.4	Returned
2563	Mandatory variables in configuration files should be identified	4	2.2.1	Endorsed
1384	DOS attacks	3	2.2.2	In Clarification
1381	Improving automatic services configuration	3	2.2.3	In Clarification
1203	Uniform API	0	2.2.4	Endorsed
3279	Top-BDII must scale with number of sites	5	2.3.1	Returned
3326	Automatic configuration of CE capacity parameters	3	2.3.2	In Clarification
2652	Support multi-VO services, association to multiple topBDIIs	2	2.4.1	Returned

Table 3: Overview of proposed TCB decision for each submitted and reviewed TCB requirement.

The recommended decisions as summarised in Table 3 spread approximately evenly over the three decisions the TCB may take (i.e. to endorse, request clarification, or return). This indicates potential for improvement in the requirements engineering processes preceding the TCB submission:

- Improve the decision making process whether a requirement is truly in scope for the TCB to manage
- Improve the requirements engineering process and techniques eliciting the technical details and context of a requirement before submitting to the TCB

The quality and scope of the requirements suggested for endorsement and clarification enforce the direction and intent of the TCB requirements process: To coordinate strategic requirements and technical evolution in the areas of

- Standard adoption (#3230, #1203, #881)
- Service operation (#2563)
- Good service abstraction level allowing generic services (#881)
- Service management and operation (#1187, #1381, #3326)

This direction aligns well with EGI's move towards a platform oriented e-Infrastructure model.

[5] For those requirements where more than one recommendation is given, only the preferred recommendation is provided here. The respective sections list all recommendations in order of preference.

4 References

[TCBReq]	TCB Requirements Management Process, https://documents.egi.eu/document/440
[TCBDashb]	TCB requirements tracking dashboard, https://wiki.egi.eu/wiki/Track_UMD_Requirements
[TCB-14]	14th TCB meeting (F2F), 6 November 2012, Amsterdam, NL, http://go.egi.eu/TCB-14
[TCB-11]	11th TCB meeting (F2F), 24 April 2012, Amsterdam, NL, http://go.egi.eu/TCB-11
[TCB-10]	11th TCB meeting (F2F), 10 February 2012, http://go.egi.eu/TCB-10
[TCB-4]	4th TCB meeting, 28 February 2011, http://go.egi.eu/TCB-4
[GFD.129]	The Storage Resource Manager Interface Specification v2.2, OGF, 2008, http://www.ggf.org/documents/GFD.129.pdf
[SRMv2.2]	SRMv2.2 (GFD.129) as online HTML version, https://sdm.lbl.gov/srm-wg/doc/SRM.v2.2.html#_Toc241633135