TCB Requirements

Report on requirements submitted and reviewed by EGI communities

Michel Drescher (Michel.Drescher@egi.eu)

07 December 2012

Version 1

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 15th TCB meeting [TCB-15] for the participants to decide how to proceed with each of the requirements discussed in this document.

At the 14th TCB meeting [TCB-14] most of the pending requirements were analysed, and either sent back to the originating communities for clarification and re-engineering, or taken forward for Technology Providers to assess possible implementation effort and impact. Some were rejected by the TCB for the originating requestor to pursue using alternative means.

However, while preparing documentation on submitted and reviewed requirements for TCB-14, one requirement slipped the common attention, and was not discussed at the last TCB meeting.

2 Submitted requirement

Currently, there is only one requirement waiting for its analysis at the TCB, as summarised in Table 1 below.

	EGI Capability	ID	Requirement	Prio	Type
Data	Database Access	926	Distributed relational data management	5	nf

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

2.1 Distributed relational data management (#926)

Analysis:

Distributed relational data management, with ACL-based access control, are needed. Some databases are distributed over multiple sites, making a mediation and federation layer necessary.

GRelC is providing a grid DB interface but it provides neither data mediation nor fusion for using multiple DBs simultaneously. OGSA-DAI is probably closer to providing such functionality, but it requires manually implementing mediation and pushing requests to a mediated database.

Two sites host databases containing similar data (but that may structured along different schemas), e.g. data entities related to a same medical study. A grid user would like to query both DBs simultaneously as if it would be a single one (thus requiring DB schemes alignment and data fusion in virtual tables).

Recommendation:

The requirement refers to two common design patterns in data integration designs, namely data warehousing (tight coupling between distributed databases using an "extract, transform, load" logic), and virtualised schemas (using wrappers that translate queries to local data on demand).

While both references are technically achievable, the submitted requirement leaves one important issue unanswered – the *semantic integration problem*. Semantic integration refers to the intellectual exercise (and subsequent technical implementation) of mapping correlated data of different origin

(i.e. different databases) and different structure and representation into a unified representation of the whole. This effort, however, is intrinsically *domain specific*. Since domain-specific data models lead to disparate data representations, whether local databases or virtualised schemas, it is expected that user communities will have very disparate configuration and performance requirements.

From an infrastructure perspective, it makes sense to investigate a "Database as a Service" deployment model, a specialisation of a "Storage as a Service" model that would support the hosting of the source schemas, as well as a database service that is capable of hosting virtualised schemas. RDBMSs exist that support virtualised schemas.

Therefore, it is recommended to request further clarification from the originating community whether such a potential service scenario is indeed what was requested, or to clarify the requirement further before reconsideration.

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 2 provides a summary of the recommended TCB decision.

ID	Requirement	Prio	§	Suggested state[1]
926	Distributed relational data management	4	2.1	In Clarification

Table 2: Overview of proposed TCB decision for each submitted and reviewed TCB requirement.

4 References

[TCBReq]	TCB Requirements Management Process, https://documents.egi.eu/document/440
[TCB-15]	15th TCB meeting (F2F), 14 December 2012, http://go.egi.eu/TCB-15
[TCB-14]	14th TCB meeting (F2F), 6 November 2012, Amsterdam, NL, http://go.egi.eu/TCB-14

---

[1] 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.