Case Statement
for
QoS/DataLC Working Group

WG Charter

Mission

To reduce the likelihood of misunderstanding of a research community's storage requirements, or of a storage provider's service.

To facilitate dialogue between a research community and multiple storage providers, and between a storage provider and multiple research communities.

To maximise the scientific output of a research community with a fix budget by allowing them to use the cheapest storage that supports their requirements and to automate data management tasks that are predictable.

Stakeholders

The organisations responsible for procuring storage capacity for researchers that either work with sufficiently large amount of data that storing it themselves is impractical, or that form a distributed community using distributed resources. Such organisations are often universities and other research institutes.

Storage technology providers (vendors and developers) that create storage solutions for research communities.

Storage providers that offer a service to multiple research communities.

Brokers that allow research communities to discover the best storage provider for storing their data.

Agents that provide an enhanced storage service by aggregating resources from multiple storage providers.

Public organisations that procure data storage.

Anyone with archival responsibility.

Goals

Provide a common vocabulary that may be used by both researcher communities and storage providers, or by storage service providers and storage technology providers (vendors and developers) to:

- Describe the Quality of Service the researchers expect and that the storage providers will

deliver. The vocabulary may be used both within documents (such as SLAs) and computer interactions (both person-to-computer and computer-to-computer).

- Describe known, predicable transitions that data will undergo throughout its life-cycle so that a research community may delegate responsibility for managing such transitions.

Non-goals

The group will not describe the process of deciding whether some offered service is satisfactory for a particular communities desired QoS. Any such process will likely be domain-specific.

The group will not describe any network protocol that allows discovery of available QoS. While embedding the vocabulary within existing network protocols may be desirable, different communities have adopted different protocols, so it is likely that multiple protocols must be considered.

The group will also not describe what activity may be necessary by the storage provider to achieve the desired QoS. Such activity will likely be technology specific.

The group will not describe highly domain- or project- specific QoS aspects where there is little chance of those descriptions being reused by others.

Value Proposition, Beneficiaries and Key Impacts

Value proposition

This group will establish a vocabulary for describing expected Quality of Service (QoS) and Data LifeCycle (DataLC) that is applicable to multiple scientific disciplines.

This will ease the dialogue between stakeholders when negotiating an acceptable service. It may allow a research community to compare offers from multiple storage providers, including commercial offerings.

It will allow automated brokering between a research community and a storage provider, where acceptable storage is provisioned on demand.

It will allow storage providers to reduce costs by allowing them to provision storage that closely matches a research community's desired QoS; additional guarantees often translate to additional costs.

A research community can store data over multiple provisioned storage capacity with distinct QoS. This allows the community to achieve an aggregate, heterogeneous service that best matches the data while minimises cost.

It will help brokering agencies when mediating between storage providers and research communities.

It may allow aggregate storage services where some broker service provides improved Quality of Service by aggregating storage from multiple storage providers.

Beneficiaries

The key beneficiaries are the research communities and storage service providers. Research communities may obtain a better QoS, for a given budget constraints, by better describing which aspects of storage are important and which are not. Storage service providers may make better offerings, potentially reducing costs, since they will understand better the needs of their research communities.

Key Impacts

Research communities will have a lower overall costs for their storage requirements, or a "better" overall service for the same budget.

Engagement with existing work in this area

Work has started in the Storage Networking Industry Associate (SNIA) Cloud Storage Initiative (CSI) through the Cloud Data Management Interface (CDMI) protocol.

Incorporate the output from the "Practical Policy" Working Group.

Fold vocabulary into the "Data Foundation and Technology" Interest Group.

Examine the "Matrix of use-cases and functional requirements for research data repository platforms" document, provided by "Repository Platforms for Research Data" Interest Group and extract potential QoS attributes.

Identify other existing QoS descriptions from stakeholders.

Work Plan

This Working Group will work on-line using suitable collaborative tools, with video conference meetings every two weeks, and face-to-face meetings during Plenary meetings. Additional face-to-face meetings may be organised if appropriate.

Wherever possible, consensus is achieved through discussion aimed to establish a common position with which all members feel comfortable. Should some point of discussion fail to converge after being discussed in three bi-weekly meetings, an acceptable multi-choice question is phrased and members may vote on the proposal. The choice receiving the simple majority of votes wins. If no clear winner is available then the chair will select between the choices receiving the largest share of the votes using some fair, random process.

Wherever possible, conflicts will be addressed by incorporating disparate views as alternative ways of describing QoS and DataLC, through a consensus-driven model. To facilitate reuse, common subsuming concepts are introduced to mark the commonality between different approaches.

Roughly once every two months, the chairs will evaluate current activity against the desired goals to identify areas where the group is deviating from the charter. If this happens and there is sufficient interest in that alternative direction then a separate Working Group is formed to continue that activity and this group may remain focused.

The work is organised into the following activity and milestones:

April 2017	Starting session of the Working Group (session participants)
April – June 2017	An initial set of QoS descriptions, based on existing QoS standards and community-driven use-cases for QoS. Commonalities are identified (members of the Working Group).
June 2017	Provide initial set of QoS descriptions to allow feedback from stakeholders (members of the Working Group).
July – Sept 2017	Updated set of QoS descriptions, based on feedback (members of the Working Group).
September 2017	Working Group meeting during P10 for creating primer document and to prepare next steps (session participants).
Sept – Dec 2017	Vocabulary updated to describe DataLC transitions (members of the Working Group).
Jan – April 2018	Stakeholder commenting on vocabularies (co-chairs, Stakeholders, Members of Working group).
April 2018	Working Group meeting during P11; First complete draft for the outcome report for discussion (session participants).
September 2018	Presenting final report at P12; (co-chairs, members of the Working Group).

Adoption Plan

Adoption within CDMI as a protocol extension. The CDMI protocol includes aspects of QoS management and is currently being extended by the INDIGO-DataCloud (IDC) project. One aspect of this work is to support describing additional QoS and Data-Lifecycle attributes.

SNIA, the standards organisation supporting CDMI, has a standard procedure for accepting extensions to CDMI. With the adoption of this extension by IDC CDMI server, the extension becomes eligible for inclusion as an Appendix to the CDMI specification.

Adoption within EUDAT. Collaborative work between IDC and EUDAT has pointed towards places where a common approach would benefit both user-communities, with a standard vocabulary for describing desired Quality of Service being an obvious example. It is anticipated that EUDAT will adopt the vocabulary described by this Working Group within their services.

Adoption within EGI. The EGI project acts as a broker and expert within scientific disciplines within Europe. A common approach between IDC and EGI would bring benefits to both communities. It is anticipated that EGI will adopt the vocabulary described by this Working Group when drafting documents and matching services to user communities.

Final deliverables:

- A primer document describing the QoS and Data-lifecycle vocabulary.

- A set of defined terms that describe how the desired Quality of Service and Data-Lifecycle for storage may be described.

Initial Membership

Mikael Borg (co-chair),

Patrick Fuhrmann,

Marcus Hardt,

Paul Millar (chair),

Ville Tenhunen.