Quality of Service

&

Data Life Cycle Definitions

Mikael Borg
mikael.borg@nbis.se

RDA Plenary 9 meeting
Barcelona, Spain
2017-04-05 – 2017-04-07
https://www.rd-alliance.org/plenaries/rda-ninth-plenary-meeting-barcelona

Agenda

- Introductions
- Overview
- How the research data infrastructure could benefit from QoS and DataLC definitions? Case: The Project Mildred (Ville Tenhonen)
- First technical implementation of QoS in storage in the INDIGO-DataCloud project (Patrick Fuhrmann)
- Discussion: case statement, initial work and how to move forward

999.9
FINE
GOLD
19154
88

QoS: provisioning

- **Expectations** researchers have:

  Integrity of service, Performance of service, …

- **Promises** that service providers make:

  Ideally matches requirements

- The two **one-to-many** problem:
  - Storage provider talking with many research communities
  - Research communities talking with many storage providers

- A common vocabulary:

  Facilitates communication and reduces likelihood of misunderstanding

QoS: brokering

- Research communities likely not experts in technology

  Deciding between options requires considerable background knowledge

- Organisations exist to help
  - Requirement-capture, identifying available resource providers, …
  - Currently a rather ad-hoc process.
- Brokering could become automated

  MANY (communities) to ONE (vocabulary) to MANY (storage providers)

- A common vocabulary:

  Reduce complexity, simplifying the decision process

QoS: optimising

- Limited financial resources

  In the end, storage cost money and needs to be funded.

- Can we differentiate storage requirements?

  For example, “hot” data and “cold” data

- Different kinds of data can have different QoS requirements

  Store “cold” data on cheaper hardware, so that “hot” data can be stored on more expensive hardware.

- A common vocabulary:
  - Provides research communities with the ability to describe what their data needs in a dynamic and segmented fashion.
  - Reduces a barrier in storage procurement.

Examples

- Performance (bandwidth, latency)
- Replicated storage.
- Geographic constraints (e.g. *“can only be stored within Europe”*)
- Scrubbing frequency (integrity checks)
- Deletion standards (e.g. *“disks must be physically destroyed”*)
- …

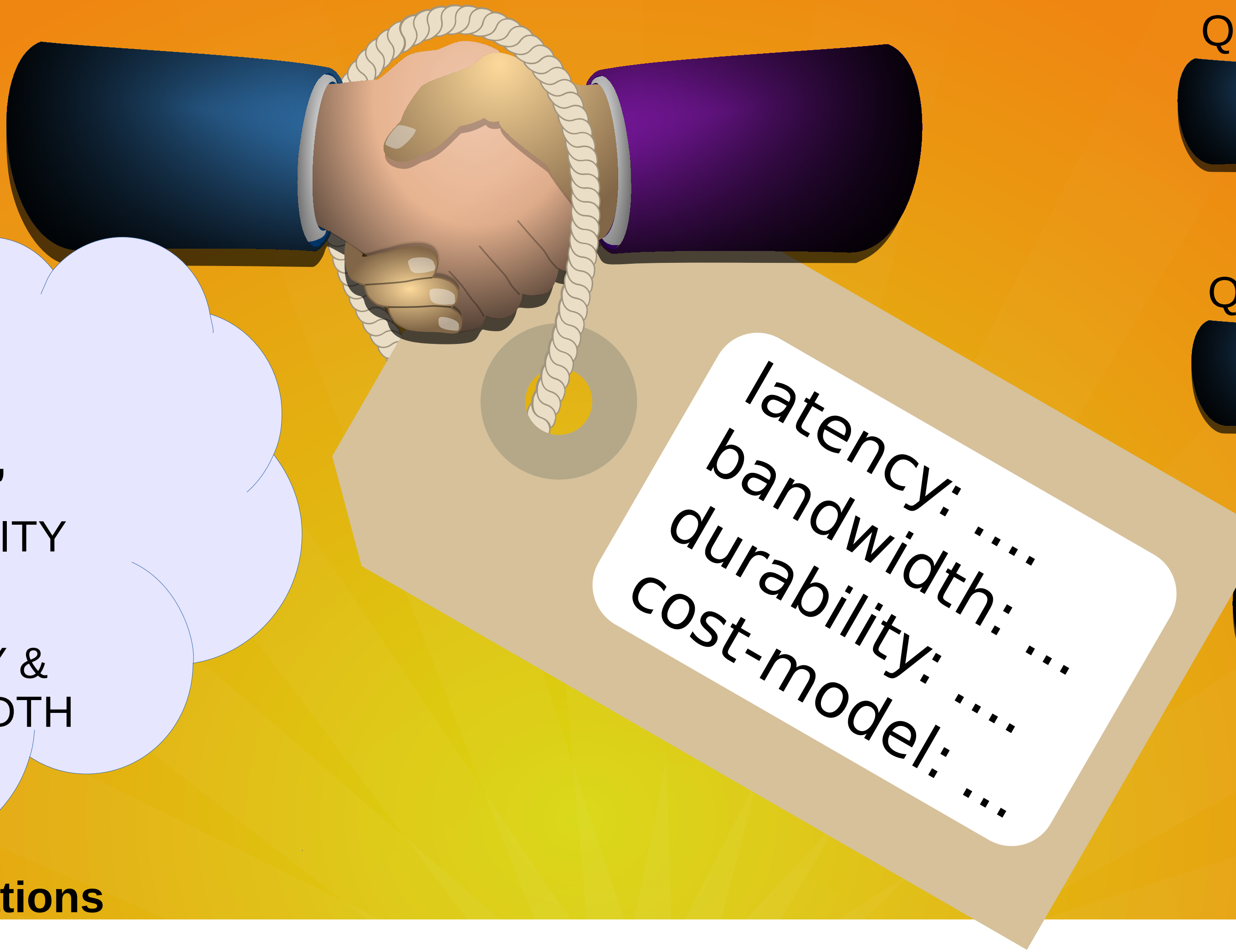
Choosing QoS

“SCRATCH”
(latency)

“ARCHIVAL”
DURABILITY

“FAST”
LATENCY &
BANDWIDTH

User expectations



QoS #1: **SCRATCH**

QoS #2: **SCRACH, FAST**

QoS #3: **ARCHIVAL**

QoS #4:

Data-LifeCycle

- QoS is about **time-invariant** quality

  Not the measurable reality, but the promise

- Data-LC are **time-dependent** transitions:
  - Accept/Reject during online analysis,
  - Scientific review (e.g., peer-reviewed journeys),
  - Public embargo (supporting members),
  - Hot → Cool → Cold data transitions: QoS,
  - Archiving / Deleting data.

- Hand over responsibility:

  Automation is possible, but only if the desired behaviour can be described.

Work so far

- WG initiated by Paul Millar
- BoF sessions at RDA plenaries 6 and 7
- WG case statement submitted to RDA (Mar -17)
  - available at RDA website
- initial QoS definitions created (Paul Millar)
  - SKOS (Simple Knowledge Organization System)
- access to semantic web technology platform PoolParty via ANDS (thanks!)

Related work

- Practical policies WG (concluded)
- Data Foundations and Terminology IG
- National Data Service IG

Next steps

- review case statement
- plan work up to next plenary
  - expand QoS definitions
  - engage more stakeholders?
  - regular WG meetings

Backup slides

Case statement:
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.