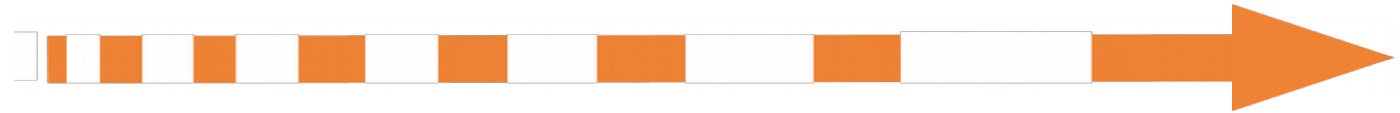


Storage Service Defn WG Introduction & Update



Data Management for extreme scale computing



Paul Millar paul.millar@desy.de



eXtreme DataCloud is co-funded by the Horizon2020
Framework Program – Grant Agreement 777367
Copyright © Members of the XDC Collaboration, 2017-2020

Agenda

- ✂ Introduction round: who's who.
- ✂ Update on our current status (this talk).
- ✂ Plan for entering real-world data: *Milan Daneček, Ville Tenhunen.*
- ✂ SKOS vs Ontologies: where we are and where we are going:
Nick Car.
- ✂ Plan for adoption as EC ICT specification: *Raphael Ritz.*
- ✂ Links with related activities – the broader picture (everyone here).
- ✂ Our next steps.

Quick overview of WG

Storage provisioning for scientists

✗ **Expectations** researchers have:

Integrity of service, Performance of service, ...

✗ **Promises** that institutes make:

(ideally, this should match requirements)

✗ There are two **one-to-many** connections:

- Institutes talking with (potentially) many research communities.
- Research communities talking with (potentially) many institutes.

✗ A common vocabulary/ontology:

Facilitates communication and reduces likelihood of misunderstanding.

Storage provisioning by institutes

✗ **Expectations** institutes have:

Integrity of service, Performance of service, ...

✗ **Promises** that vendors / cloud-providers make:

(ideally, this should match requirements)

✗ There are two **one-to-many** connections:

- Vendors / cloud-providers talking with (potentially) many institutes.
- Institutes talking with (potentially) many vendors / cloud-providers.

✗ A common vocabulary/ontology:

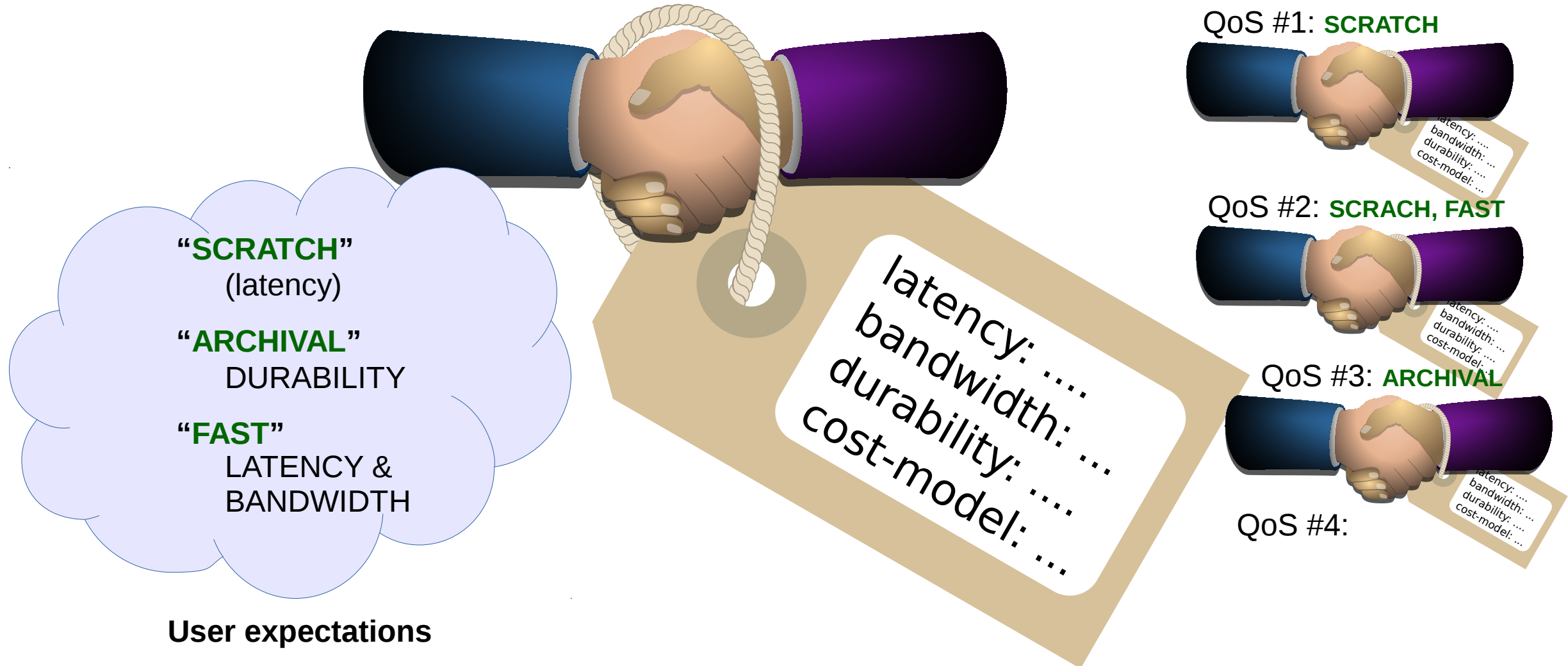
Facilitates communication and reduces likelihood of misunderstanding.

Examples of aspects of storage

... more than just media

- ✗ Performance (bandwidth, latency).
- ✗ Replicated storage (e.g., “*2 copies in different buildings*”).
- ✗ Geographical constraints (e.g., “*only store within Europe*”).
- ✗ Scrubbing (integrity checking) frequency,
- ✗ Maximum lifetime of data (e.g., HPC-connected storage),
- ✗ Deletion standards (e.g., “*disks must be physically destroyed*”).
- ✗ Are there backups? (how often? how are they handled?)
- ✗ Protection against physical access to storage hardware?
- ✗ ...

Choosing QoS



Quick updates

New co-chair: Nicolas Car

- ✂ Senior Experimental Scientist at CSIRO
(formally at Geoscience Australia)
- ✂ Our resident Ontology expert
- ✂ Member of the W3C Data Exchange WG



XDC is the new INDIGO

- ✗ **INDIGO-DataCloud** (EU / Horizon 2020-funded project) officially ended 2017-09-30.
- ✗ **eXtreme DataCloud** (XDC) is one of the successor projects
 - XDC is from 2017-11-01 to 2020-01-31,
 - XDC focus is on data storage; the other project (DEEP) has computing as the focus.
- ✗ XDC's WP4 has strong involvement on Quality of Service
 - ... which will require a common definition of a storage service.

W3C's Data Exchange WG (DataX)



- ✂ Nick has raised semantic Storage Service Description with DataX.
DataX consensus is that this lies outside the working-group's remit; however, ...
- ✂ DataX are expanding the DCAT Distribution semantics to include...
 - indicators for extension; i.e., we can add a bridge to our ontology,
 - how datasets may be obtained from data services.
- ✂ DataX will produce a DCAT→ schema.org mapping.
This may help us make a similar mapping for our ontology.

Thanks for listening!