Quality of Service
&
Data LifeCycle

Paul Millar

paul.millar@desy.de

RDA Plenary 7 meeting
Tokyo, Japan.
2016-02-28 – 2016-03-03
https://rd-alliance.org/plenary-meetings/rda-seventh-plenary-meeting.html

INDIGO-DataCloud is co-founded by the
Horizon 2020Framework Programme

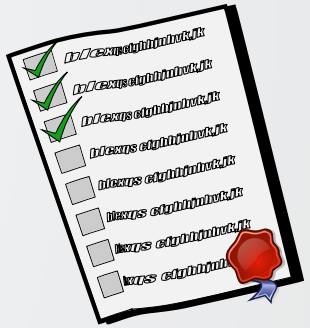
Plans for today

- A quick introduction

  In case anyone new is here.

- Plans for today

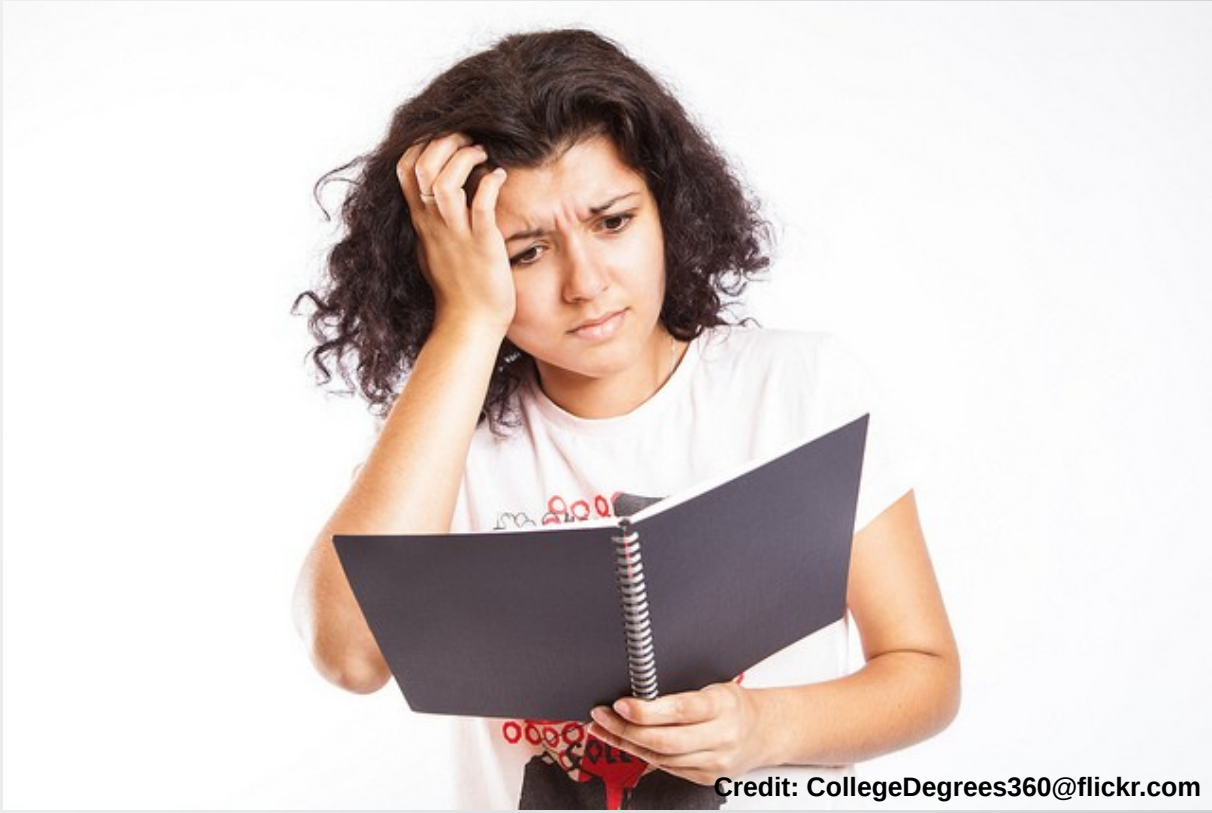
  What we want to achieve.



- Next steps

  Plans for after RDA Plenary 7 is over.

QoS/DataLC: a quick introduction



Credit: CollegeDegrees360@flickr.com

QoS: provisioning

INDIGO - DataCloud

- **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

INDIGO - DataCloud

- 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: aggregating

INDIGO - DataCloud

- Requirements may be difficult to achieve

  Research communities may have requirements that are hard to satisfy

- Enabling federated storage

  Provide an aggregate service, based on multiple services.

- May be a manual or automatic process

  Could have an agent that can commission storage, as needed.

- A common vocabulary:

  Facilitate understanding of how such an aggregate system will behave.

QoS: optimising

INDIGO - DataCloud

- 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.

Data-LifeCycle

INDIGO - DataCloud

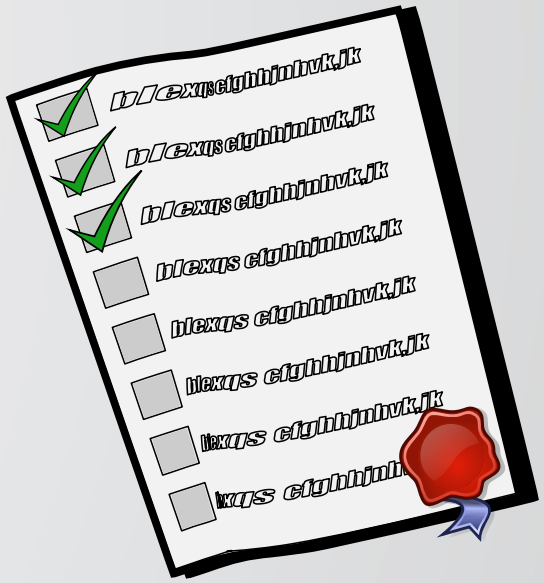
- QoS is about **time-invariant** quality

  Not the measurable reality, but the promise

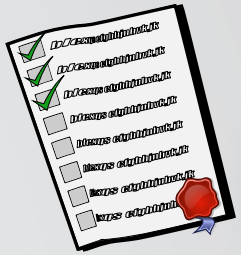
- Data-LC are **time-dependent** transitions:
  - Accept/Reject during online analysis,
  - Scientific review (e.g., peer-reviewed journeys),
  - Public embargo (supporting members),
  - Hot → Cool → Cold data transitions: QoS,
  - Archiving / Deleting data.
- Hand over responsibility:

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

Plans for today



Plans for today



INDIGO - DataCloud

- Come up with concrete proposal for how this group will operate:
  - Meetings: frequency and method?
  - Procedure: how to we agree on things?
  - Goals: what are we going to do?
  - Timelines: when are we going to do them?
- Put together a first draft of the case statement,
- Start collecting existing QoS / DLC examples:
  - SRM, CDMI, ...
- Maybe start defining terms ...

Next steps

Next steps

INDIGO - DataCloud

- Put case statement on mailing list,
- Complete the RDA WG formation process,
- Start regular meetings and get the wheels in motion.

INDIGO - DataCloud

Backup slides