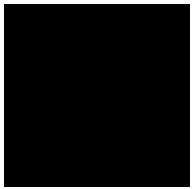
Data Commons Workshop

23 October 2020
Organized by Harvard University

Data Commons from a Library Perspective

Philipp Conzett
Senior Research Librarian
University Library
UiT The Arctic University of Norway

ORCID 0000-0002-6754-7911
@PhilippConzett



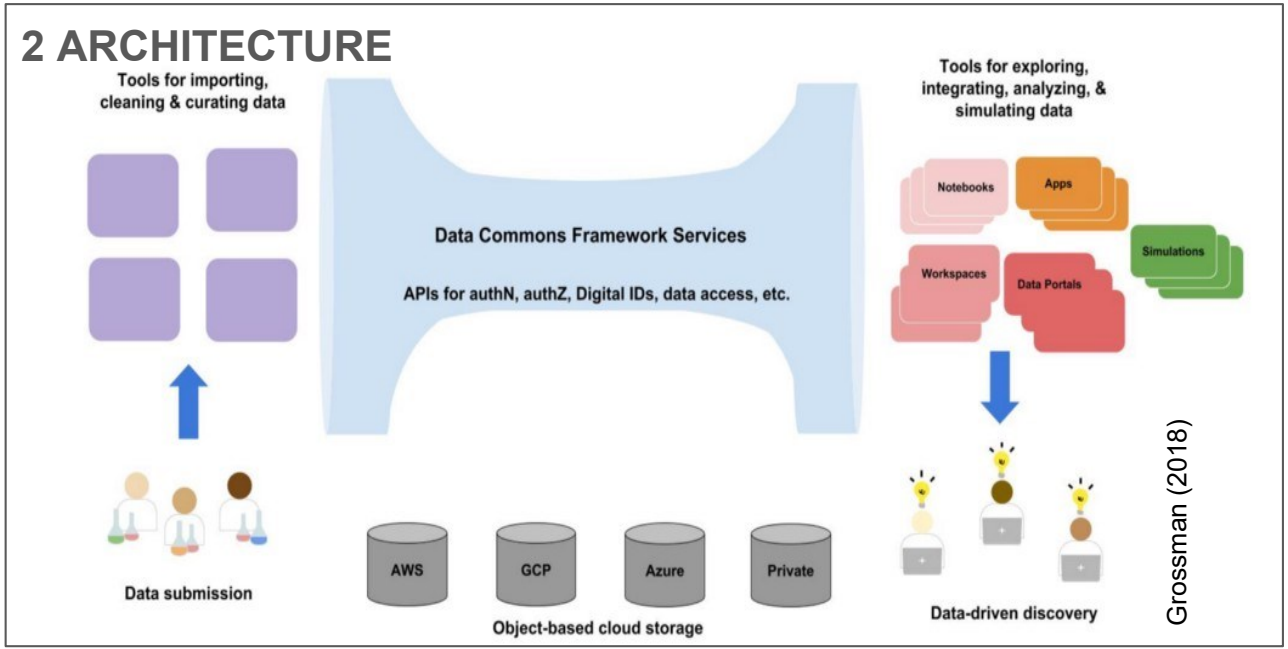
Thanks for inviting me!

1. What does a data commons mean to me?

Common view on data commons

1 DEFINITION “Data commons collocate data, storage, and computing infrastructure with core services and commonly used tools and applications for managing, analyzing, and sharing data to create an interoperable resource for the research community.” (Grossman et al., 2016,10)

2 ARCHITECTURE



3 STAKEHOLDERS

- data commons service providers
- data contributors
- data users

Things I miss in commons views on data commons?

1. The long tail of research: Researchers dealing with big data are often part of a domain-specific data commons, whereas researchers from the long tail of research often lack this kind of data infrastructure support and thus are to a larger degree dependent on support services provided by the library. (See also The e-IRG Task Force on the Long Tail of Data, 2016.)

2. Human support services: Maybe included in Grossman's "data commons service providers", but should be addressed more explicitly; cf. data stewards, data librarians, data curators… Therefore, it's good to see practices being included on today's agenda.

2. What technologies, practices, or standards are we (/ am I) involved in that could be part of a data commons solution?

Technologies

- **Storage, computing and transfer:** Local, national (cf. Sigma2), international (cf. GÉANT)
- **PID systems:** DOI, ORCID, ROR, …
- **Authentication:** National authentication service for HE institutions (Feide), …
- **Data management planning:** DMP tools by NSD and Sigma2?
- **Data and project management:** Office365, OSF, R, RSpace, ...
- **Data sharing and publishing:** DataverseNO, the Tromsø Repository of Language and Linguistics (TROLLing)
- **Data discovery:** B2FIND, BASE, DataCite Search, Google Dataset Search, Open Polar (https://site.uit.no/open-polar/), ...
- **Registries:** FAIRsharing, re3data, … (the standards used in DataverseNO and TROLLing are registered in FAIRsharing)
- **Outreach and compliance:** Current Research Information System in Norway (Cristin), Humanities Commons, ...
- **International data commons:** CLARIN, DARIAH, ELIXIR, SSHOC, EOSC, …
- ...

Standards

- Standards used for secure storage at UiT
- Standards used in Dataverse
  - Metadata standards: Dublin Core, DDI, …
  - API: OAI-PMH, SWORD, …
  - ...
- Standards used in national authentication service (Feide)
  - OAuth, ...
- ...

Practices

RDM skills, training, and outreach:

- Locally at UiT: RDM training and outreach coordinated by the UiT Library
- National efforts and initiatives are coordinated by the Norwegian RDA node
- Internationally, by participating in relevant IGs, WGs and other venus (CLARIN, LIBER, RDA, …)

Data curation:

- DataverseNO Curator Network

National coordination in general:

- Norwegian Directorate for ICT and Joint Services in Higher Education & Research

What is my vision for next steps related to implementing a data commons for my institution?

Some concrete steps...

- Implement RSpace and OSF Institutions
- Integrate DataverseNO with (a) DMP tool(s) and a data policy manager
- Integrate data curation and publishing with review of publications (cf. DataverseNO Plus grant proposal)
- Integrate DataverseNO and TROLLing with EOSC (European Open Science Cloud)
- Provide support for Domain Data Protocols (cf. Science Europe) once they are in place

In parallel: A more unitary approach

Getting established a **vision and plan** for a data commons at UiT

- through **researcher engagement**;
- in **collaboration with other units** at UiT (ICT, Research Administration, …);
- **endorsed** and supported by UiT **leadership**;
- supporting **all types of data**, from long-tail to big data;
- **aligned** with local, national and global needs and requirements, and with national, global, and domain-specific data commons;
- and always focusing on **co-located human research support** as close to the researcher as possible.

Thank you for listening!

ORCID 0000-0002-6754-7911

@PhilippConzett

References

Grossman, Robert L., Allison Heath, Mark Murphy, Maria Patterson, and Walt Wells. ‘A Case for Data Commons: Toward Data Science as a Service’. *Computing in Science Engineering* 18, no. 5 (2016): 10–20. https://doi.org/10.1109/MCSE.2016.92.

Grossman, Robert. ‘A Proposed End-To-End Principle for Data Commons’. Medium, 6 July 2018. https://medium.com/@rgrossman1/a-proposed-end-to-end-principle-for-data-commons-5872f2fa8a47.

The e-IRG Task Force on the Long Tail of Data. ‘Long Tail of Data’. E-IRG Task Force Document. The Hague: e-IRG, 2016. http://e-irg.eu/documents/10920/238968/LongTailOfData2016.pdf.