Pluralist Data

Antranig Basman

Raising the Floor - International

Starting from Luke Church's final slide:

Governance [and AI] design is starting from Data

Data is a fundamentally alien phenomenon

What is the material that we should be governing?

and Where?

The problematic notion of “data collection”

A centralising activity before decision-making (inference, AI) is named “data collection”

- implies that “the data” (considered a definitive, centralised material) is moved from outside some boundary to inside it

- after the data has been collected, it loses any connection with the community to which it is referred or by which it is owned

- typically impossible to even trace what use has been made of the data, let alone express ownership or governance over it - and express consent over it

- data is certainly “the new oil” - given we do not mean to treat it in an extractive way

“Essential Requirements for Establishing and Operating Data Trusts”

Many good considerations -

“... Data traceability so that data trusts can fully execute on ... consent withdrawal, bias monitoring, audits, and regulatory agency review”

“Secure and auditable computing environments”

“Public engagement that goes beyond informational transparency and into activities like co-design and deep involvement of data subjects in governance”

But from certain points of view puzzling gaps -

Description is remarkably high-level and abstract - how does this correspond to everyday activities of real communities who simply “have data”?

How are these communities helped in their everyday work by this abstract description? How can they work with the data they have, consistent with their ways of working, at a price they can afford?


hing and Operating Data Trusts

een Canadian Organizations and Initiatives

Andrea Smith
Vector Institute

Michael Brudno
HPC4Health; University Health Network; Hospital for Sick Children; University of Toronto, Department of Computer Science

Brian Courtney
Sunnybrook Research Institute

Chris Loken
Compute Ontario

Michael J. Schull
ICES; Health Data Research Network Canada

Adrian Thorogood
McGill University, Centre of Genomics and Policy; Global Alliance for Genomics and Health

Kathleen Yang
Canadian Institute for Health Information

isms or approaches to
f practical guidance

nd the Vector Institute
nadian
health data. The
f data trusts.
nd co-authors identified

s were identified. The
egal authority to collect,
accountable governing
comprehensive data
age, access, disclosure
ongoing public and

IJPDS

Issue 1, Vol 5

Download/View PDF

Published: Aug 24, 2020

Updated: 2020-08-24

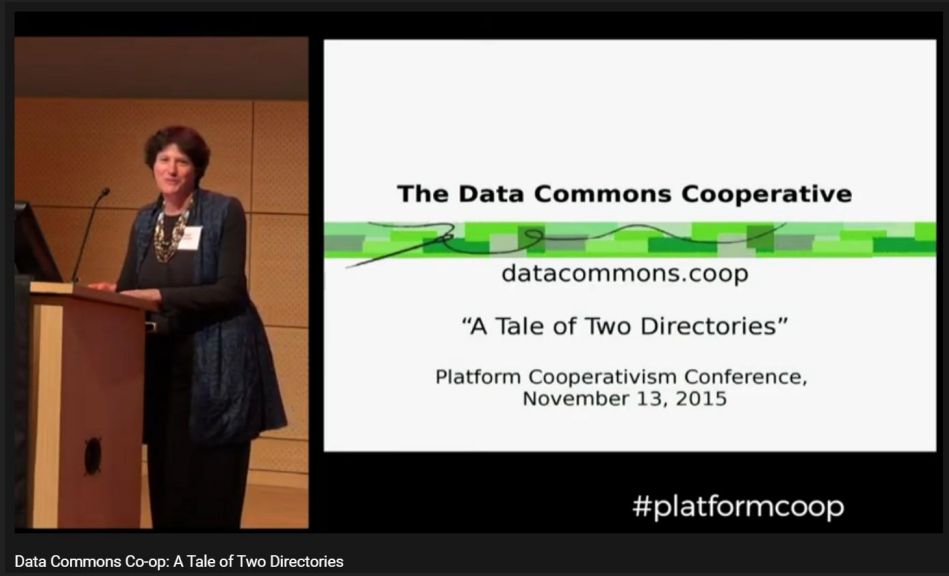
DOI: https://doi.org/10.23889/ijpds.v5i1.1353

https://ijpds.org/article/view/1353

Noemi Giszpenc at Platform Cooperativism Conference, 2015

“rather than a centralisation of information what the data Commons is aiming for is a liberation of information”

“When a grassroots group finds out about a new co-op in their area … and they update their list … a change digest will flow … these organisations will be able to select and patch that new information into their own directories”



Data Commons Co-op: A Tale of Two Directories

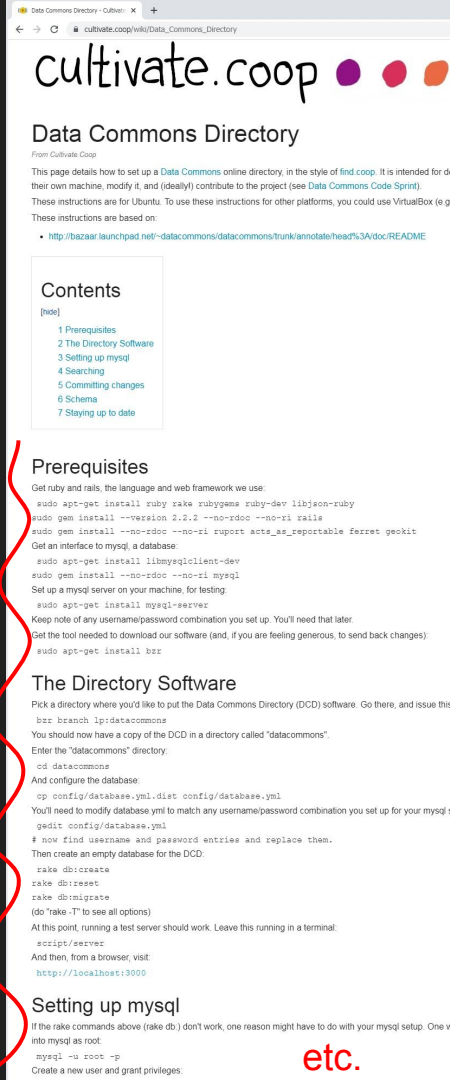
https://www.youtube.com/watch?v=D2vWX8aYmwA

Work was done on this, but in a form that is hard to take advantage of

Some Data Commons Tools

cultivate.coop

- A long list of highly technical installation instructions
- a system that requires specialised expertise to administrate
- running costs likely a minimum of $20/month
- still quite a specialised tool, working with data in a relatively fixed schema for particular kinds of communities
- “Stone Soup” software is still running at find.coop but lightly maintained and limited geographical area



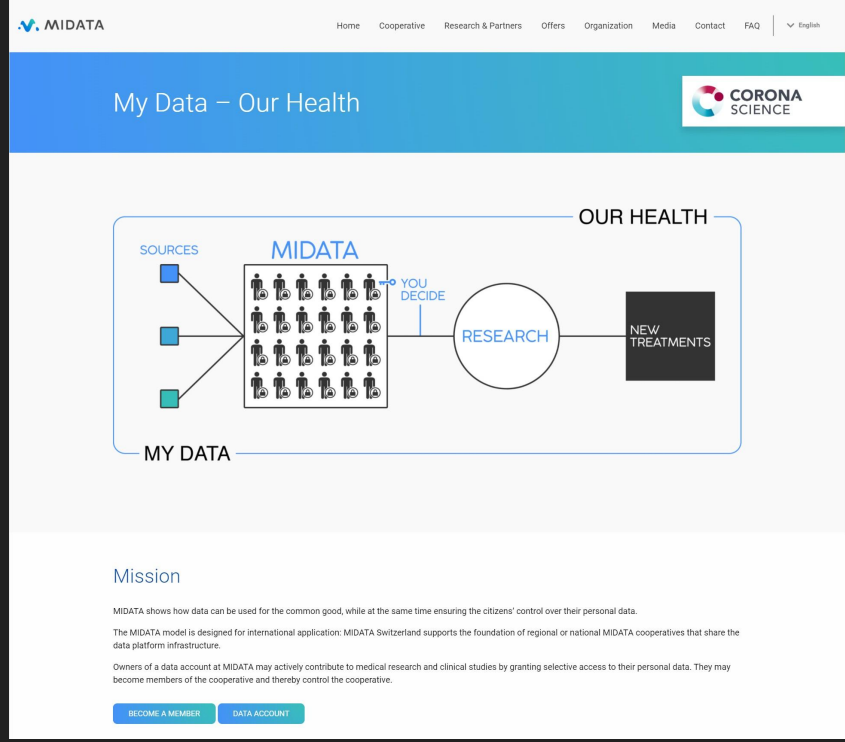
Some Data Commons Tools

MIDATA

- Again hard for a community to appropriate and “make their own” (complex install, administration, running costs) although like find.coop is open source

- Unclear how citizens will come to entrust their data to this, and what its proposition really is

- Relatively narrow focus and inflexible schema



Some ideas

- Help communities work with relatively unstructured data that is already in some format they understand (example: simple tabular data - equivalent to CSVs)

    - “Go where people are” - work with their existing tools and working practices

    - Facilitate building communities without centres or boundaries

- Produce an infrastructure that makes it clear how communities can meet their own running costs for owning their own infrastructure

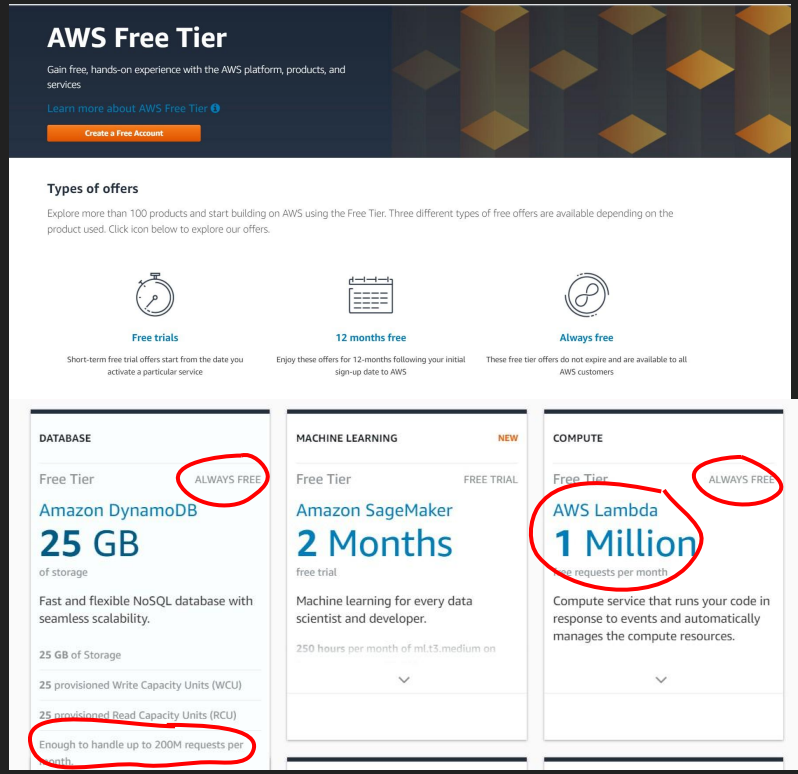
- Create an open “welcoming” interface that shows directly and immediately how to contribute and experience data (like Wikipedia)

- Exploit architectures already created by large corporations to help with the incarnations of these problems faced by technical people (auditable updates, linking and forking, tracing provenance - (like git/GitHub)

Rot the Log of Capitalism

“Free” stuff on the table

The very largest scales of capitalist technology now leave “table scraps” that are quite substantial - for those equipped to exploit them



netlify &

GitHub

Full Feature List	Starter Free to get started Start for free	Pro $19 per member /month Buy Pro	Business $99 per member /month Buy Business
Only pay for the capacity you use			
Team members	1 free (then $15 per member /month)	$19 per member /month	$99 per member /month
Reviewers	Unlimited	Unlimited	Unlimited
Concurrent builds	1	3 (then $40 each)	5 (then $40 each)
Bandwidth	100GB /month (then $20 per 100)	400GB /month (then $20 per 100)	600GB /month (then $20 per 100)
Build minutes	300 minutes /month (then $7 per 500)	1000 minutes /month (then $7 per 500)	1000 minutes /month (then $7 per 500)
Websites	Unlimited	Unlimited	Unlimited
Serverless Functions Invocations	125k per site /month ($25+ when exceeded)	125k per site /month ($25+ when exceeded)	Unlimited
Forms Submissions	100 per site /month ($19+ when exceeded)	100 per site /month ($19+ when exceeded)	Unlimited
Identity Active users	1k per site /month ($99 when exceeded)	1k per site /month ($99 when exceeded)	Unlimited

An Objection

“If the corporations consider that their free tier is being exploited, they will try to shut you down”

My response - This would be a great problem to have.

Another objection

“The boundaries of the free tier are fragile and constantly shifting - it might be cheaper in the long run to set people up paying regularly for more stable service”

In any case, keeping “close” to the architecture of the free tier is likely to minimise costs, and setting up support networks to help communities manage their costs will be essential

Some of my writing

“What Lies in the Path of the Revolution” (PPIG 2018)
https://ppig.org/files/2018-PPIG-29th-basman.pdf
- Description of what it would take for communities to be able to take ownership of their software, its architecture and its relationships

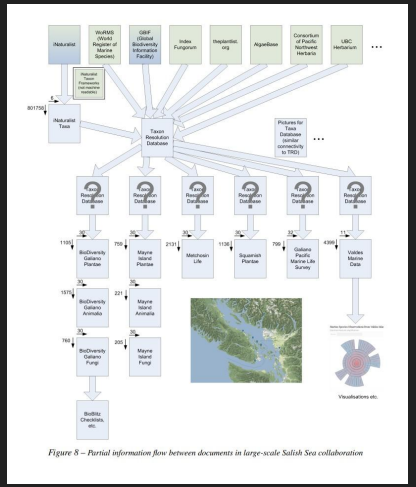


Figure 8 – Partial information flow between documents in large-scale Salish Sea collaboration

“The Naturalist’s Friend” (PPIG 2019)
https://www.ppig.org/files/2019-PPIG-30th-basman.pdf
- Case study and blueprint for pluralist data tools and infrastructure

Some of our work

Project WeCount (IDRC, 2019-)

Demonstration of basic libraries supporting pluralist data architecture - nightly job fetches latest data from a public feed, merges it with locally collected and synthetic data, and displays it together with retained provenance