

# 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?

Establishing and Operating Data Trusts  
from Canadian Organizations and Initiatives

**Michael Brudno**  
Vector Institute

**Michael J. Schell**  
FSC, Health Data Research Network, Canada

**Kathleen Yang**  
Canadian Institute for Health Information

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

**Chris Loken**  
CompuS Ontario

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

**IJPDS**  
Issue 1, Vol 5

[Download/View PDF](#)

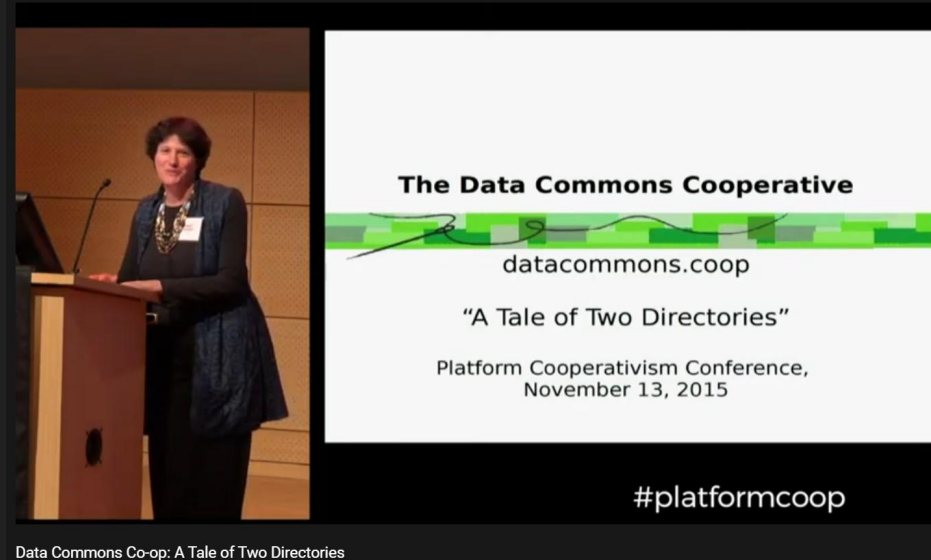
Published: Aug 24, 2020  
Updated: 2020-08-24  
DOI: <https://doi.org/10.23889/ijpds.v1i1.1353>

<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

cultivate.coop

## Data Commons Directory

From Cultivate Coop

This page details how to set up a [Data Commons](#) online directory, in the style of [find.coop](#). It is intended for developers who can contribute to the project (see [Data Commons Code Sprint](#)), modify it, and (ideally!) contribute to the project (see [Data Commons Code Sprint](#)).

These instructions are for Ubuntu. To use these instructions for other platforms, you could use VirtualBox (e.g. [Ubuntu on Windows](#)).

These instructions are based on:

- <http://bazaar.launchpad.net/~datacommons/datacommons/trunk/annotate/head%3A/doc/README>

### Contents

- 1 Prerequisites
- 2 The Directory Software
- 3 Setting up mysql
- 4 Searching
- 5 Committing changes
- 6 Schema
- 7 Staying up to date

### Prerequisites

Get ruby and rails, the language and web framework we use:

```
sudo apt-get install ruby rake rubygems ruby-dev libjs-jquery
sudo gem install --version 2.2.2 --no-rdoc --no-ri rails
sudo gem install --no-rdoc --no-ri rupaot acts_as_reportable ferret geokit
```

Get an interface to mysql, a database:

```
sudo apt-get install libmysqlclient-dev
sudo gem install --no-rdoc --no-ri mysql
```

Set up a mysql server on your machine, for testing:

```
sudo apt-get install mysql-server
```

Keep note of any username/password combination you set up. You'll need that later.

Get the tool needed to download our software (and, if you are feeling generous, to send back changes):

```
sudo apt-get install bzr
```

### The Directory Software

Pick a directory where you'd like to put the Data Commons Directory (DCD) software. Go there, and issue the following commands:

```
bzr branch lp:datacommons
```

You should now have a copy of the DCD in a directory called "datacommons".

Enter the "datacommons" directory:

```
cd datacommons
```

And configure the database:

```
cp config/database.yml.dist config/database.yml
```

You'll need to modify `database.yml` to match any username/password combination you set up for your mysql server.

```
gedit config/database.yml
```

# now find username and password entries and replace them.

Then create an empty database for the DCD.

```
rake db:create
rake db:reset
rake db:migrate
```

(do "rake -T" to see all options)

At this point, running a test server should work. Leave this running in a terminal:

```
script/serve
```

And then, from a browser, visit:

```
http://localhost:3000
```

### Setting up mysql

If the rake commands above (rake db:) don't work, one reason might have to do with your mysql setup. One possibility is that you're not running as root:

```
mysql -u root -p
```

Create a new user and grant privileges:

etc.

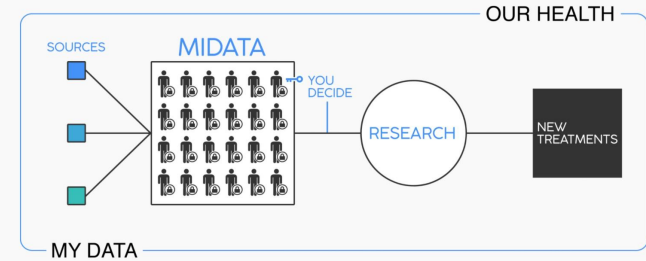
# Some Data Commons Tools

## 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



## Mission

MIDATA shows how data can be used for the common good, while at the same time ensuring the citizens' control over their personal data.

The MIDATA model is designed for international application: MIDATA Switzerland supports the foundation of regional or national MIDATA cooperatives that share the data platform infrastructure.

Owners of a data account at MIDATA may actively contribute to medical research and clinical studies by granting selective access to their personal data. They may become members of the cooperative and thereby control the cooperative.

BECOME A MEMBER

DATA ACCOUNT

# 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))

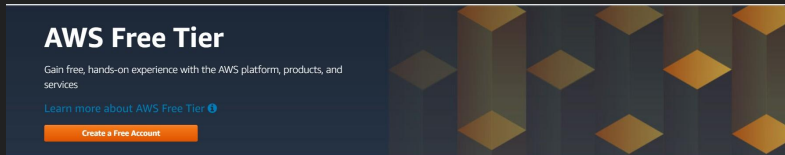


# Rot the Log of Capitalism

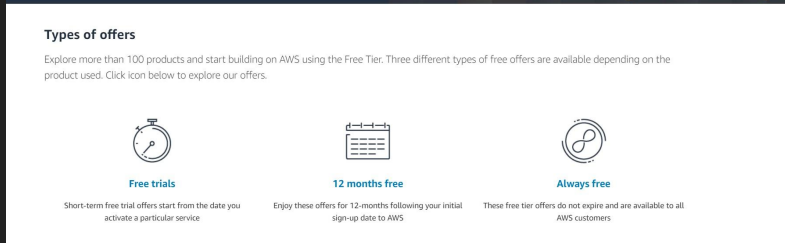


# “Free” stuff on the table

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

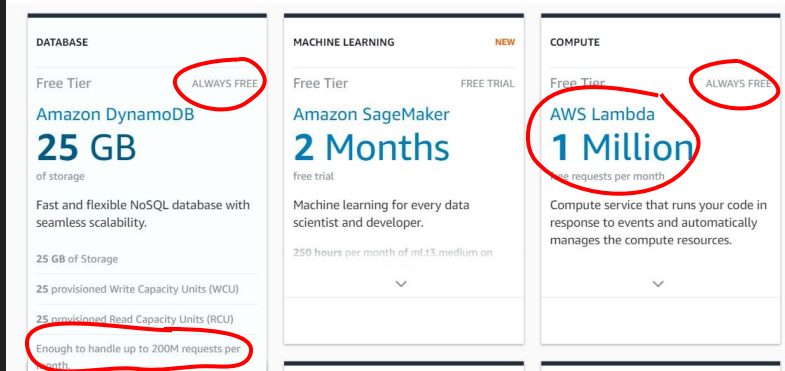


**AWS Free Tier**  
Gain free, hands-on experience with the AWS platform, products, and services.  
[Learn more about AWS Free Tier](#)  
[Create a Free Account](#)



**Types of offers**  
Explore more than 100 products and start building on AWS using the Free Tier. Three different types of free offers are available depending on the product used. Click icon below to explore our offers.

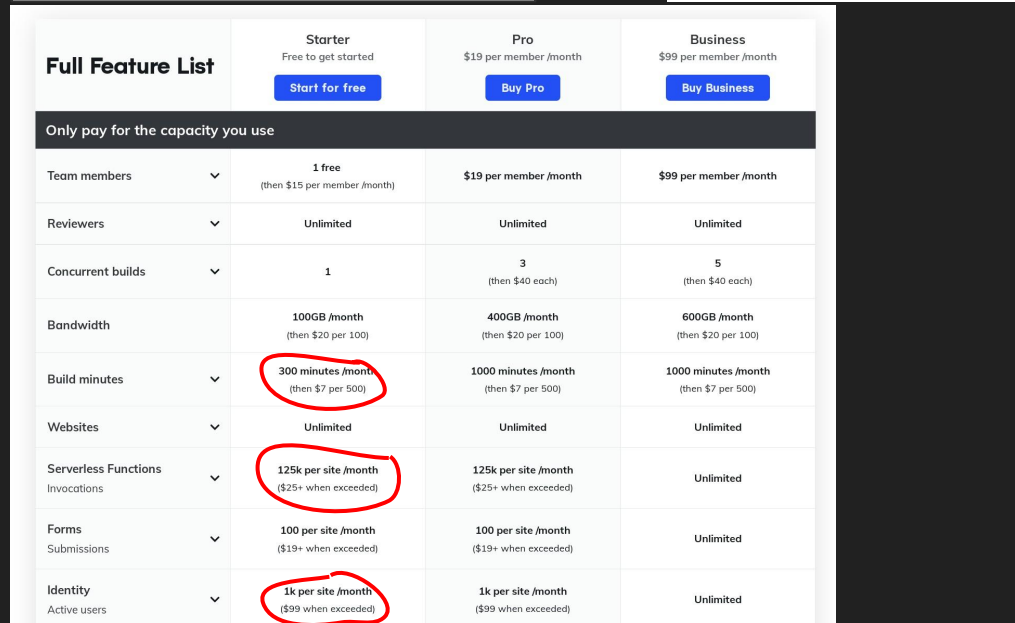
- Free trials** (Short-term free trial offers start from the date you activate a particular service)
- 12 months free** (Enjoy these offers for 12-months following your initial sign-up date to AWS)
- Always free** (These free tier offers do not expire and are available to all AWS customers)



DATABASE	MACHINE LEARNING	COMPUTE
<b>Free Tier</b> <b>ALWAYS FREE</b> <b>Amazon DynamoDB</b> <b>25 GB</b> of storage Fast and flexible NoSQL database with seamless scalability. 25 GB of Storage 25 provisioned Write Capacity Units (WCU) 25 provisioned Read Capacity Units (RCU) <b>Enough to handle up to 200M requests per month</b>	<b>Free Tier</b> <b>FREE TRIAL</b> <b>Amazon SageMaker</b> <b>2 Months</b> free trial Machine learning for every data scientist and developer. 250 hours per month of ml.t3.medium on	<b>Free Tier</b> <b>ALWAYS FREE</b> <b>AWS Lambda</b> <b>1 Million</b> requests per month Compute service that runs your code in response to events and automatically manages the compute resources.



**netlify** & **GitHub**



Full Feature List	Starter	Pro	Business
	Free to get started <a href="#">Start for free</a>	\$19 per member /month <a href="#">Buy Pro</a>	\$99 per member /month <a href="#">Buy Business</a>
<b>Only pay for the capacity you use</b>			
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	<b>300 minutes /month</b> (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	<b>125k per site /month</b> (\$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	<b>1k per site /month</b> (\$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



“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

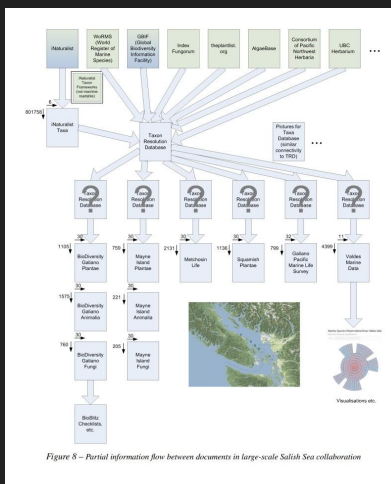


Figure 8 – Partial information flow between documents in large-scale Salish Sea collaboration

# 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

**We Count** + show preferences

ABOUT INITIATIVES EVENTS BADGES NEWS RESOURCES Search...

The data in this map is a synthesis of three sources -

- Live data from the Ontario Data Catalogue, most recently fetched from URL <https://data.ontario.ca/dataset/8ba...locations.csv>, on 30/11/2021, 23:01:48,
- Accessibility data collected by the WeCount project most recently on 2nd September 2020
- Synthetic accessibility data generated at random for live testing centres not covered by the WeCount collection

This map is produced only for demonstration of the WeCount data architecture, to learn about the location of COVID-19 test centres in Ontario, please visit the [Government of Ontario's live site](#).

Search by postal code or city

[Back to locations](#)

**Markham Stouffville Hospital**  
Monday: 09:00-18:00, Tuesday: 09:00-18:00,  
Wednesday: 09:00-18:00, Thursday: 09:00-18:00, Friday: 09:00-18:00  
381 Church Street, Markham, ON L3P 7P3  
null

**1210 Castlemore Ave - Dynacare**  
Monday: 17:00-20:30, Tuesday: 17:00-20:30,  
Wednesday: 17:00-20:30, Thursday: 17:00-20:30, Friday: 17:00-20:30  
1210 Castlemore Avenue, Markham, ON L6E 0H7  
null

**Carefirst Markham Office**  
Monday: closed, Tuesday: 09:30-14:30,  
Wednesday: 09:30-14:30, Thursday: closed,  
Friday: closed  
4461 Highway 7 East, Markham, ON L3R 1M1  
416-646-5108

**105 Gibson Centre**  
Monday: closed, Tuesday: 09:30-14:30,  
Wednesday: 09:30-14:30, Thursday: closed

**Markham Stouffville Hospital**

UI elements are sourced from verified data

**Hours:** Monday: 09:00-18:00, Tuesday: 09:00-18:00, Wednesday: 09:00-18:00, Thursday: 09:00-18:00, Friday: 09:00-18:00

**Location**  
381 Church Street, Markham, ON L3P 7P3  
+1 416-586-5054  
<https://www.msh.on.ca/>

**ACCESSIBILITY FILTERS** 0

- ACCESSIBLE ENTRANCES
- ACCESSIBLE WASHROOMS
- ACCESSIBLE PARKING
- INDIVIDUAL SERVICE
- WAIT ACCOMMODATIONS

Reset Filters  Apply Filters

UI elements are sourced from 2020 verified data

**Accessibility features**

- Accessible Parking
- Individual Service
- Wait Accommodations