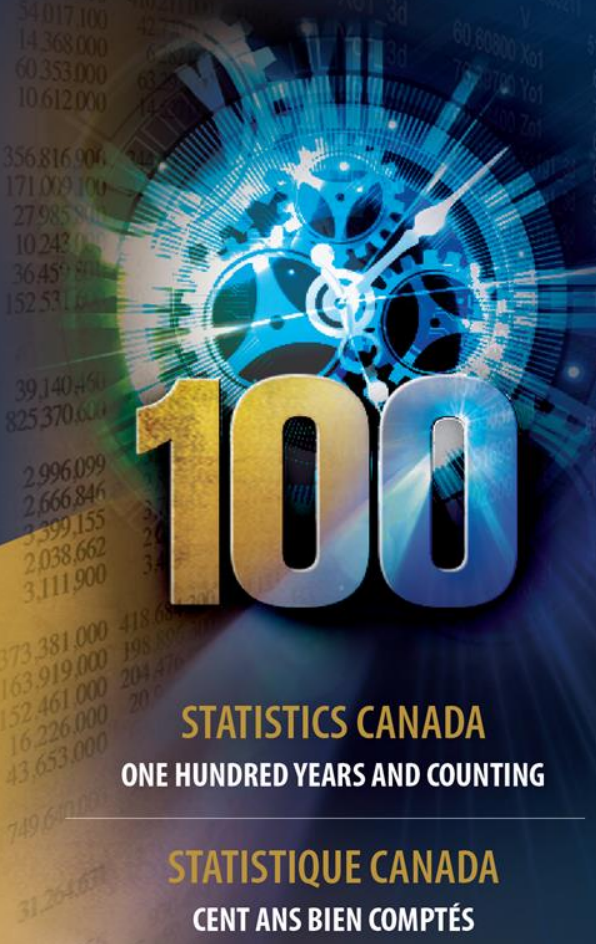
Open Projects at DEIL
exploring the potential of open data and open approaches

Data Exploration and Integration Lab (DEIL)
Statistics Canada



Open First Day
Ottawa, September 28, 2018

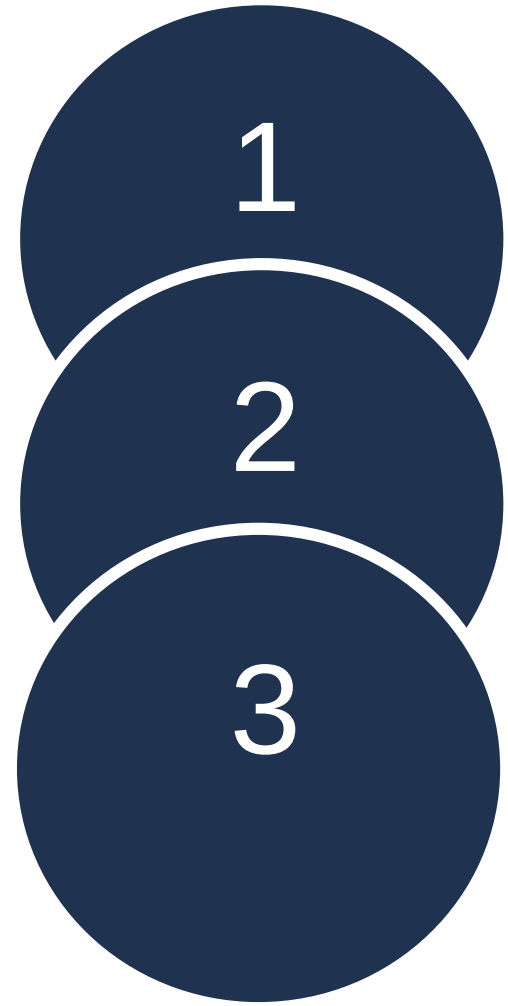
Context

- Statistics Canada has a history of producing open data but the environment is evolving
- NSOs are increasingly becoming consumers and stakeholders of open microdata; there are new producers of open microdata
- Crowdsourcing pilot project was a first step in exploring use of open data, open tools, open code
- And now we want more!

Take home messages

- Great potential for integration and use of open microdata from (municipal, provincial, federal) government sources

- Great potential for conducting this work in open spaces as an “open project”



Crowdsourcing pilot

Open microdata from government sources

***Open Projects* and collaborative data creation**

1

2

3

Crowdsourcing pilot

Open microdata from government sources

Open Projects and collaborative data creation

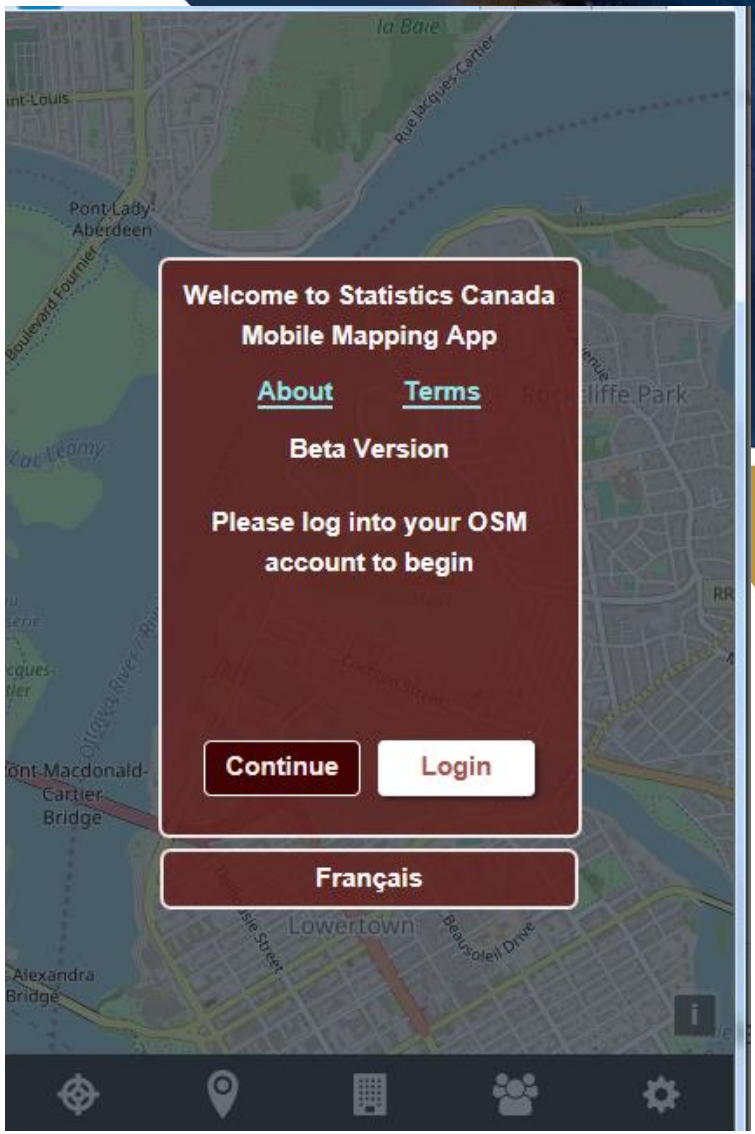
Crowdsourcing Pilot: *Overview*

- Collaborated with OSM and local municipalities
- Built open-source editor and mobile app
- Virtually complete map of pilot area (Ottawa-Gatineau) achieved

Crowdsourcing Pilot: *Mobile App*

- Created app to facilitate collecting data
- Used open source/code approach

CODAP on GitHub



Also, how much does a gram of weed cost in Ontario?

Crowdsourcing cannabis prices

- $7.37 per gram in ON
- Over 20k submissions
- Microdata available for download

[StatsCannabis - Demo]

1

Crowdsourcing pilot

2

Open microdata from government sources

3

Open Projects and collaborative data creation

Experimental Open Database of Buildings

"The ODB"

The ODB: *Details*

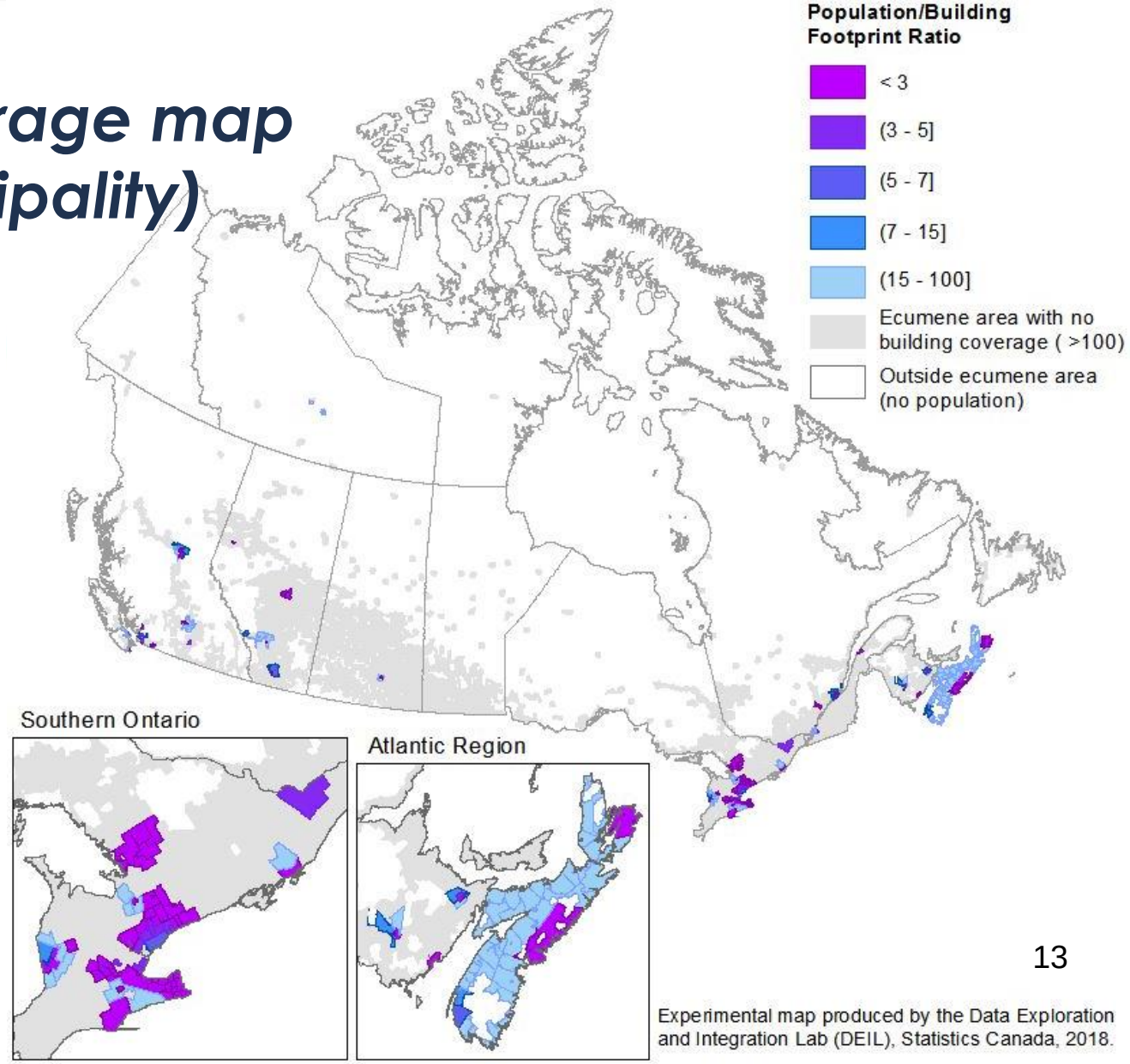
- 62 governmental datasets
- 4.3 million footprints (and counting)
- 100% PURE CANADIAN OPEN MICRODATA!

Open Data / Open License

means everyone can access and use the information

The ODB: *coverage map by CSD (municipality)*

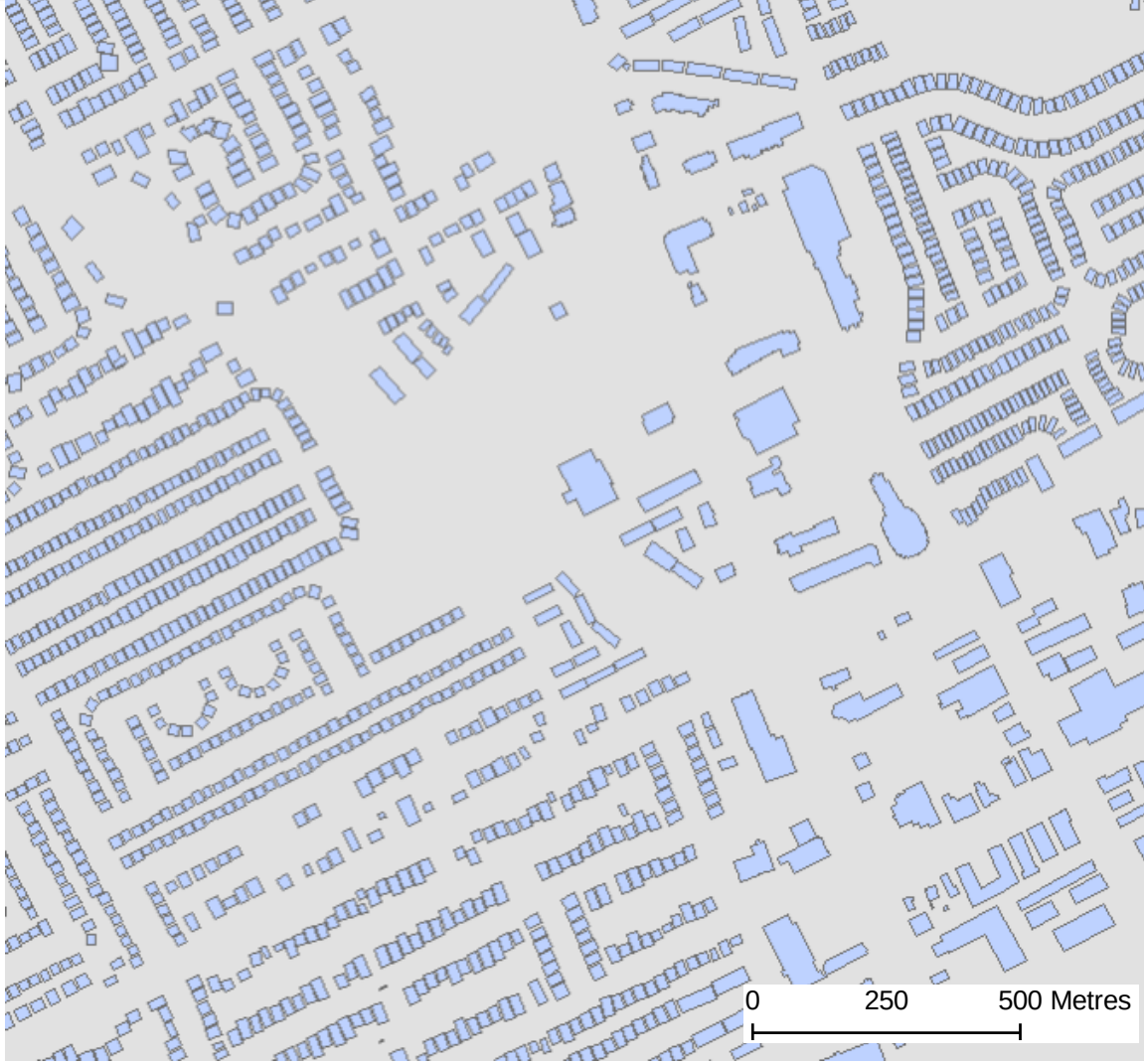
- 8 Provinces / Territories
- 40 CMA/CAs
- Urban and rural areas



Experimental map produced by the Data Exploration and Integration Lab (DEIL), Statistics Canada, 2018.

The ODB: *example of the data*

- Ex: Footprints for Richmond Hill, Toronto
- Quality seems high, buildings are tightly knit
- Working to make ODB available for download



FID	Shape *	Longitude	Latitude	CSDUID	CSDNAME	Data_prov	Build_ID	Shape_Leng	Shape_Area
0	Polygon ZM	-79.726934	44.282285	3543021	Essa	Barrie	35430210000000	33.565106	63.971128
1	Polygon ZM	-79.699185	44.281962	3543017	Innisfil	Barrie	35430170000000	72.635583	269.749831
2	Polygon ZM	-79.726066	44.282264	3543021	Essa	Barrie	35430210000000	73.317978	278.524173

1

Crowdsourcing pilot

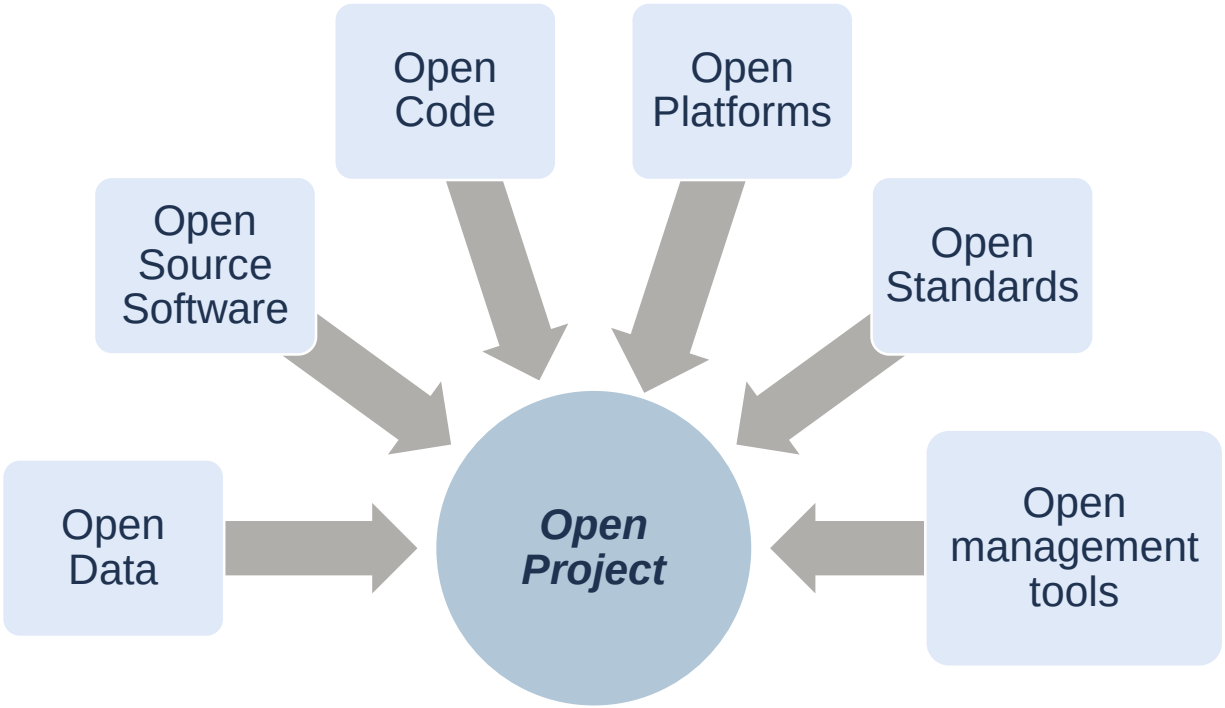
2

Open microdata from government sources

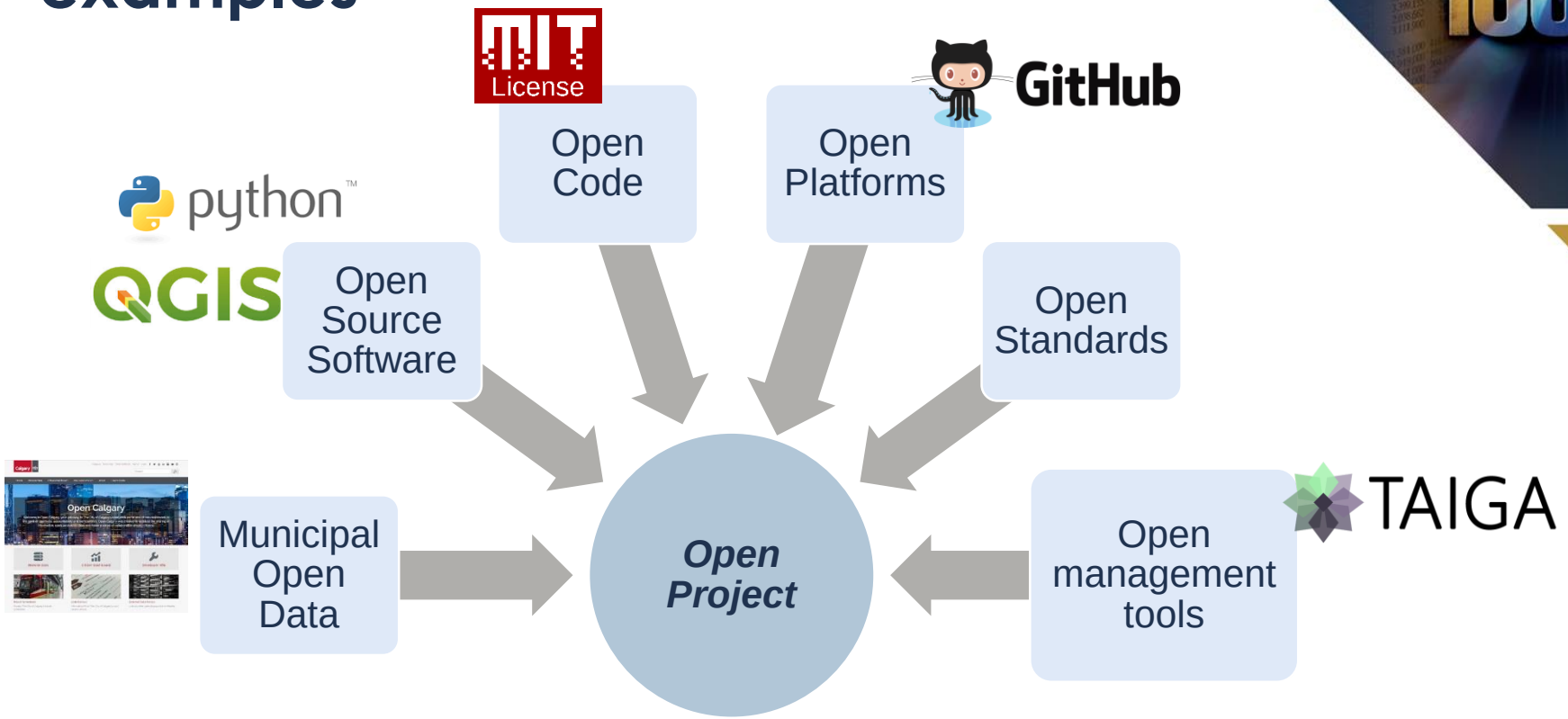
3

Open Projects and collaborative data creation

The *Open Project* concept



The *Open Project* concept: examples



Data development, an example: Open Businesses Repository "The OBR"

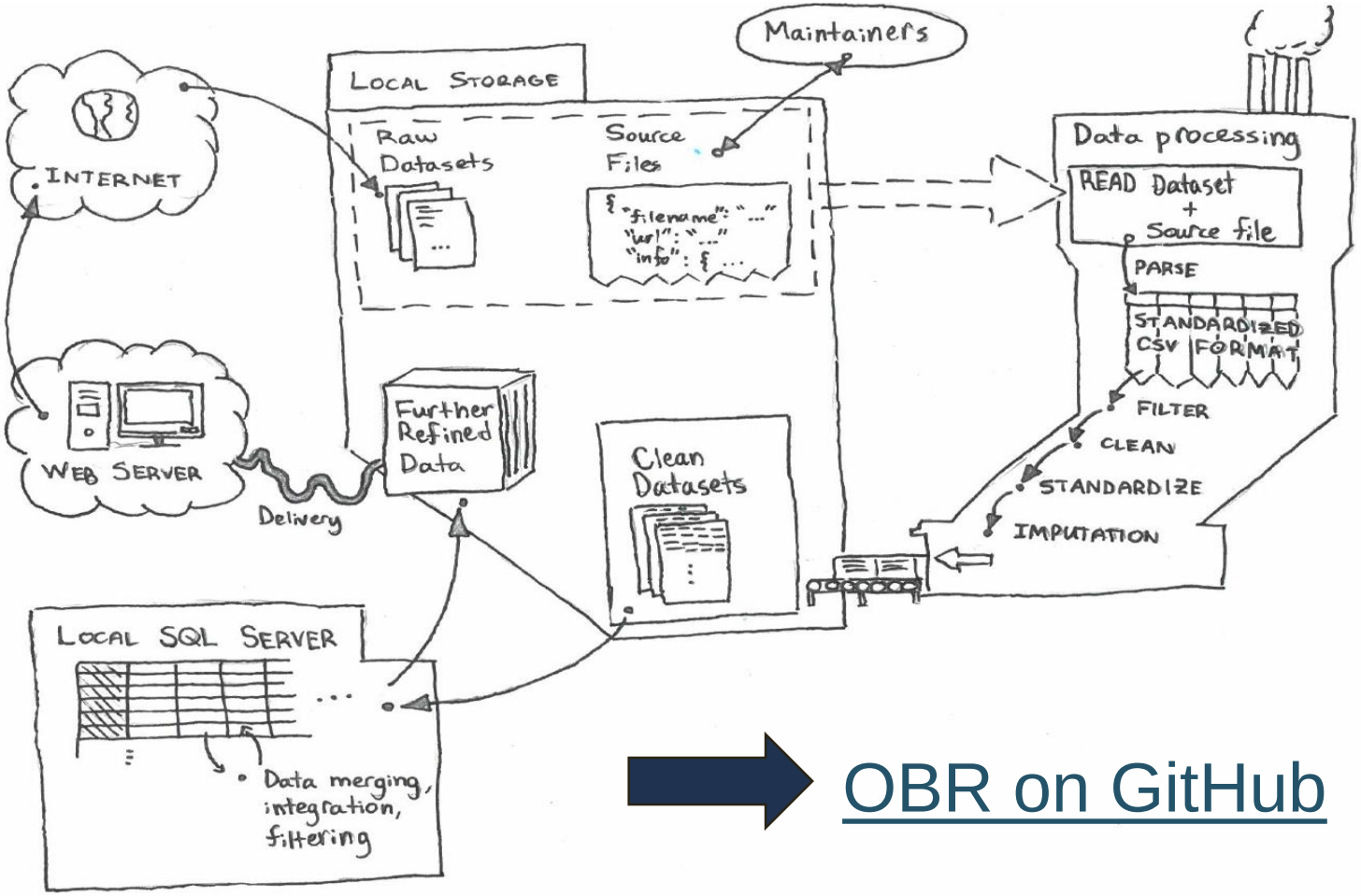
Collaborative data creation

- Data are becoming a basic infrastructure
- Partner across entities (private, public, community) for data creation
- Ex: OpenAddresses
- *Open Project* approach is an enabler
- Ex: Collaboration with Microsoft

The ORB: *Details*

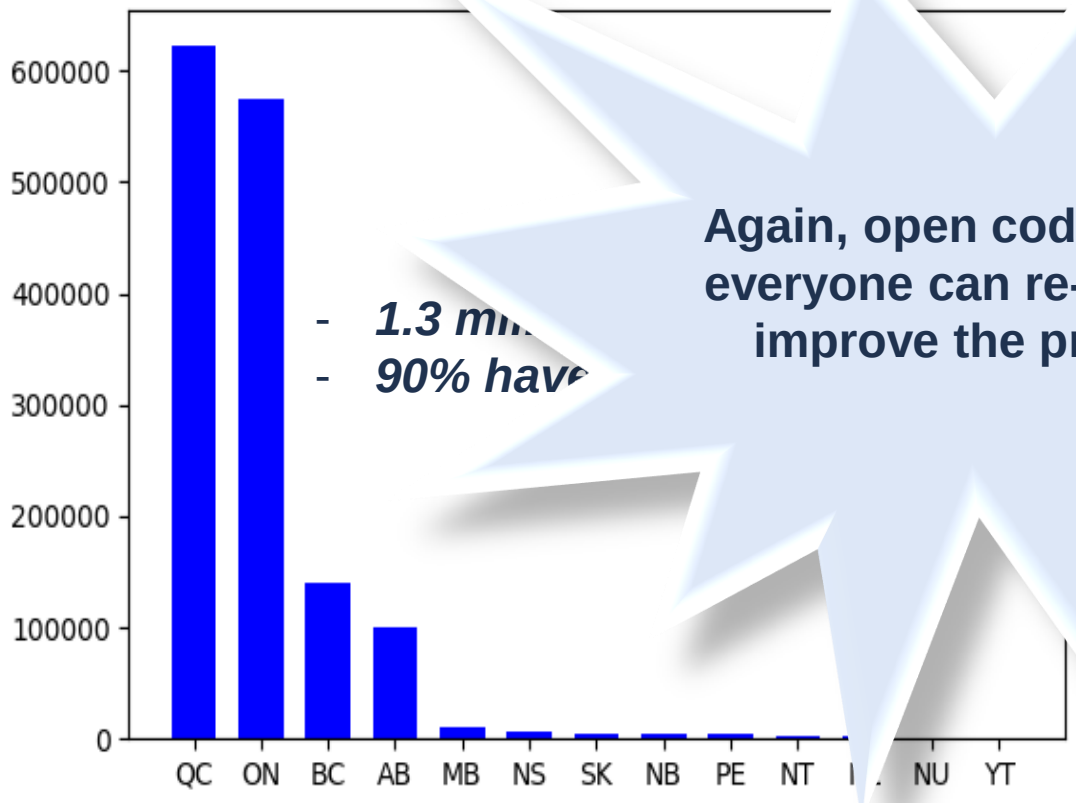
- Exploratory and experimental open project
  (Note: work in progress)
- OBR is an open software solution ≠ database
- Government/authoritative sources of open microdata on businesses
- OBR script (on GitHub) compiles, processes, integrates

The OBR



[OBR on GitHub]

The OBR: *preliminary results*



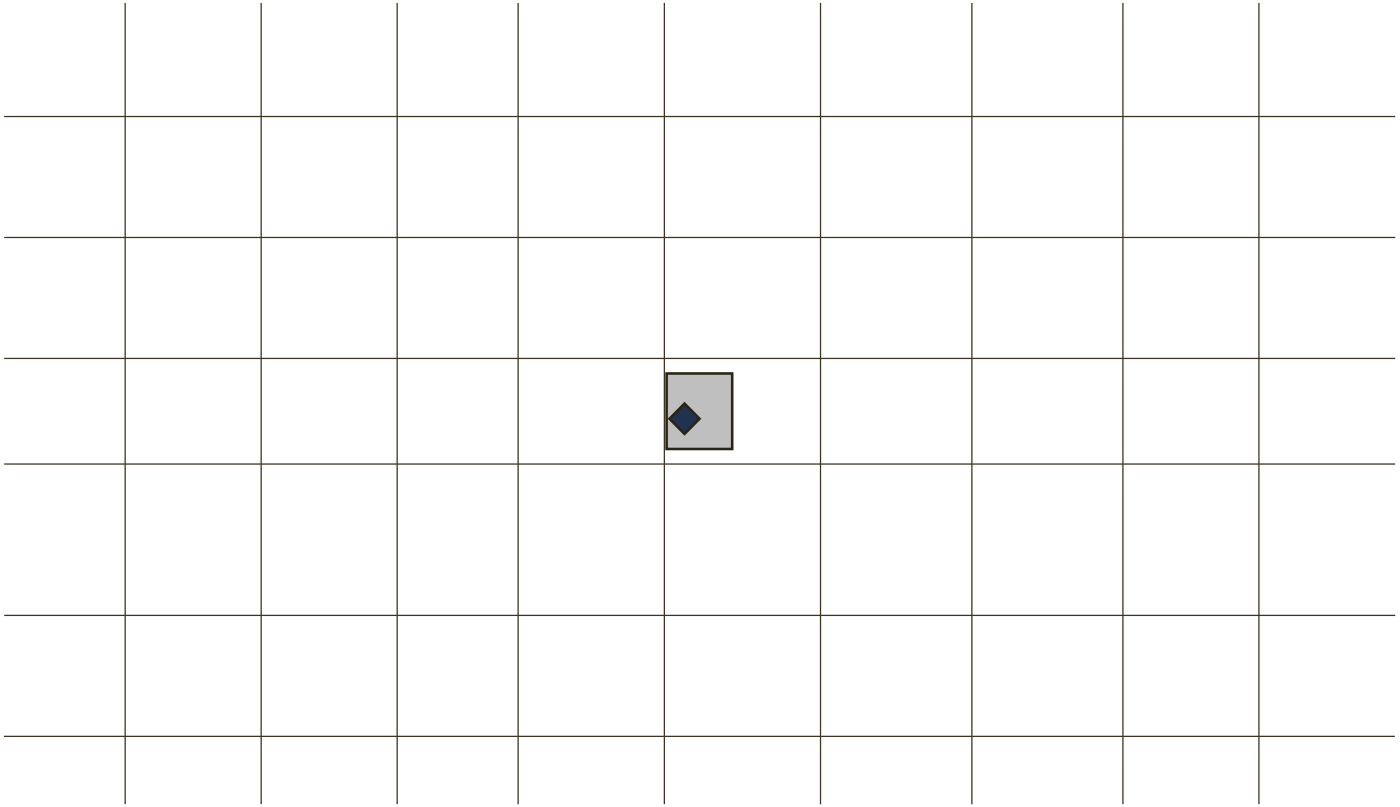
- *1.3 mi*
- *90% have*

Again, open code means everyone can re-use and improve the product

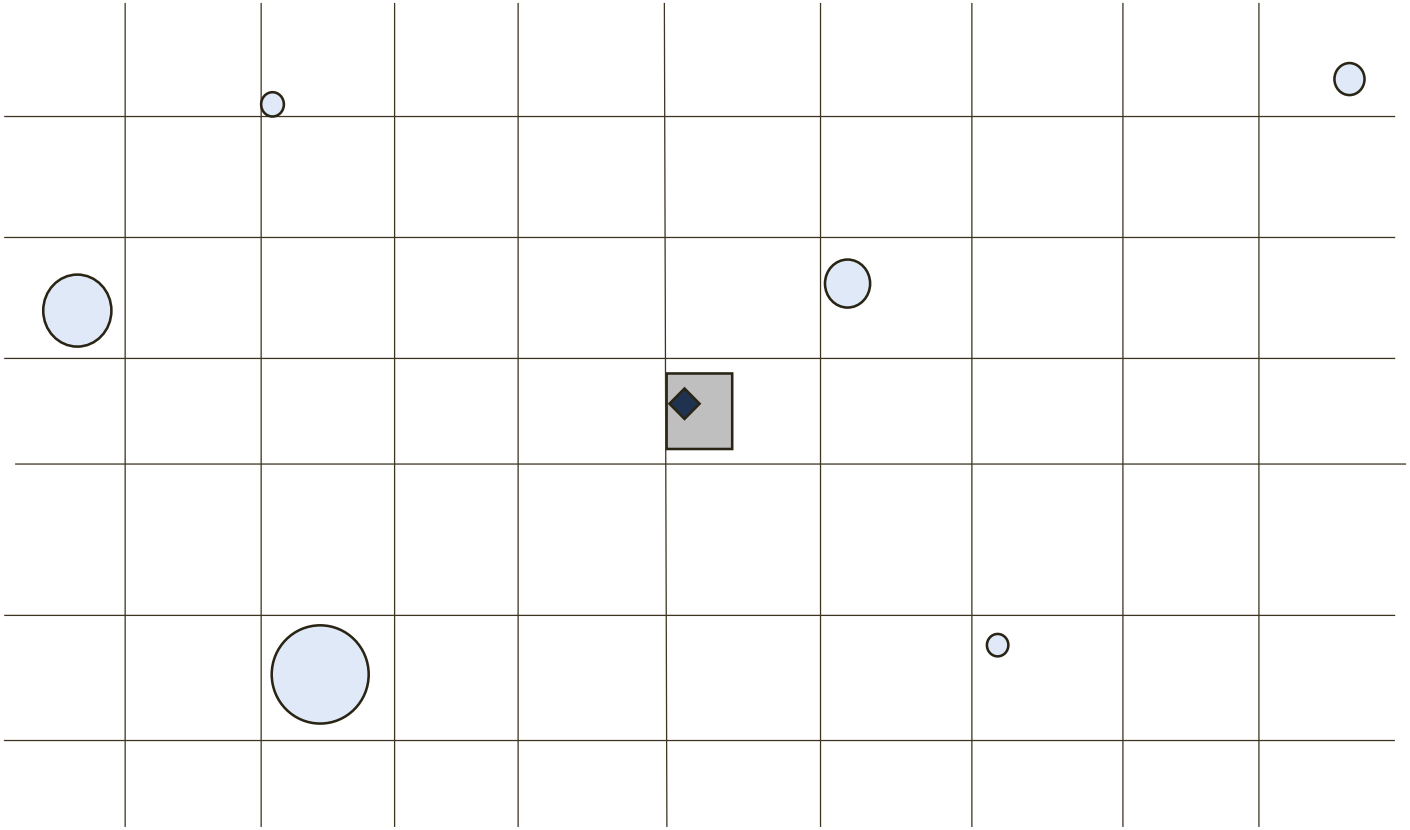
Data integration and analysis, an example:

Proximity measures

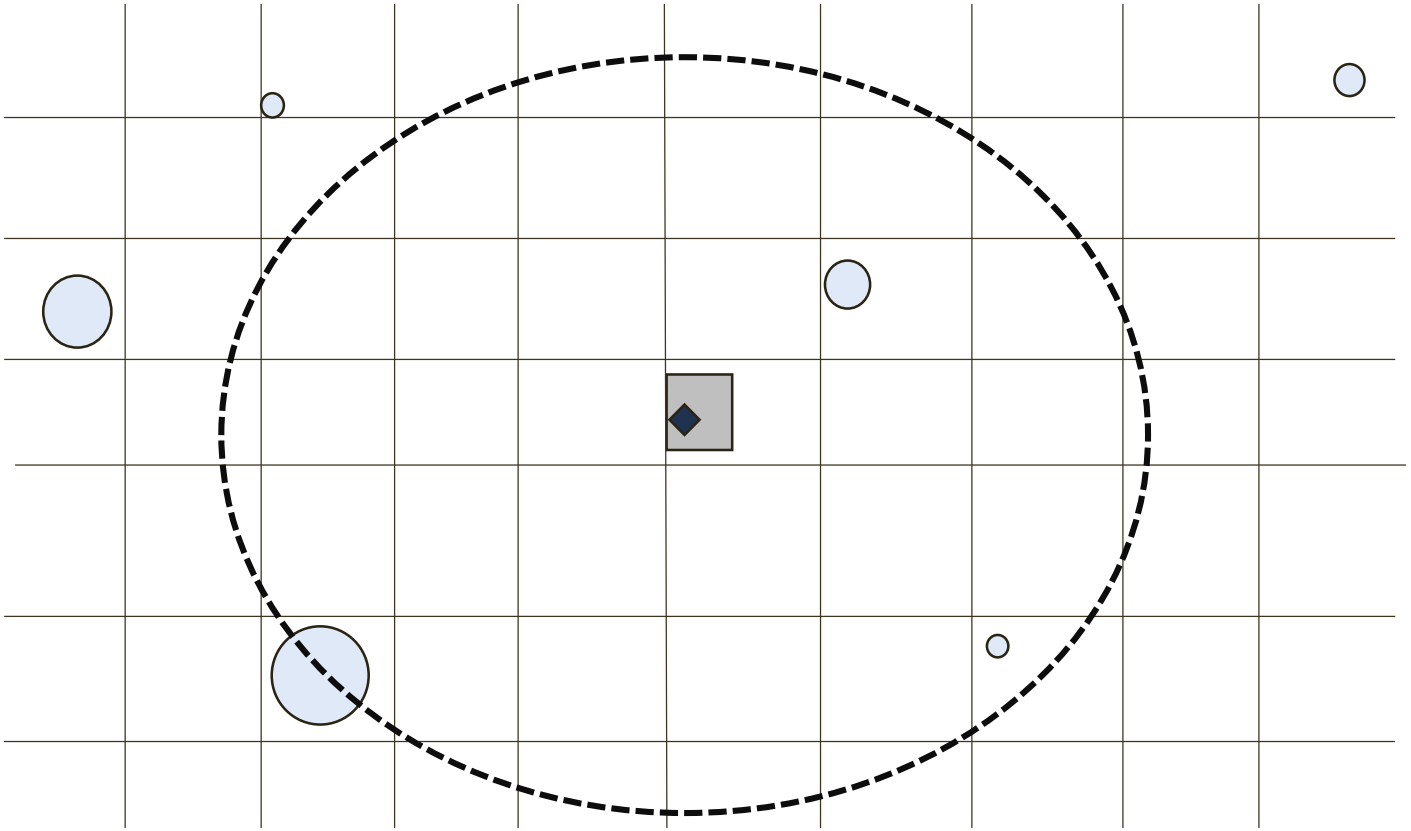
What can we do with the data?
Example - Proximity Measures



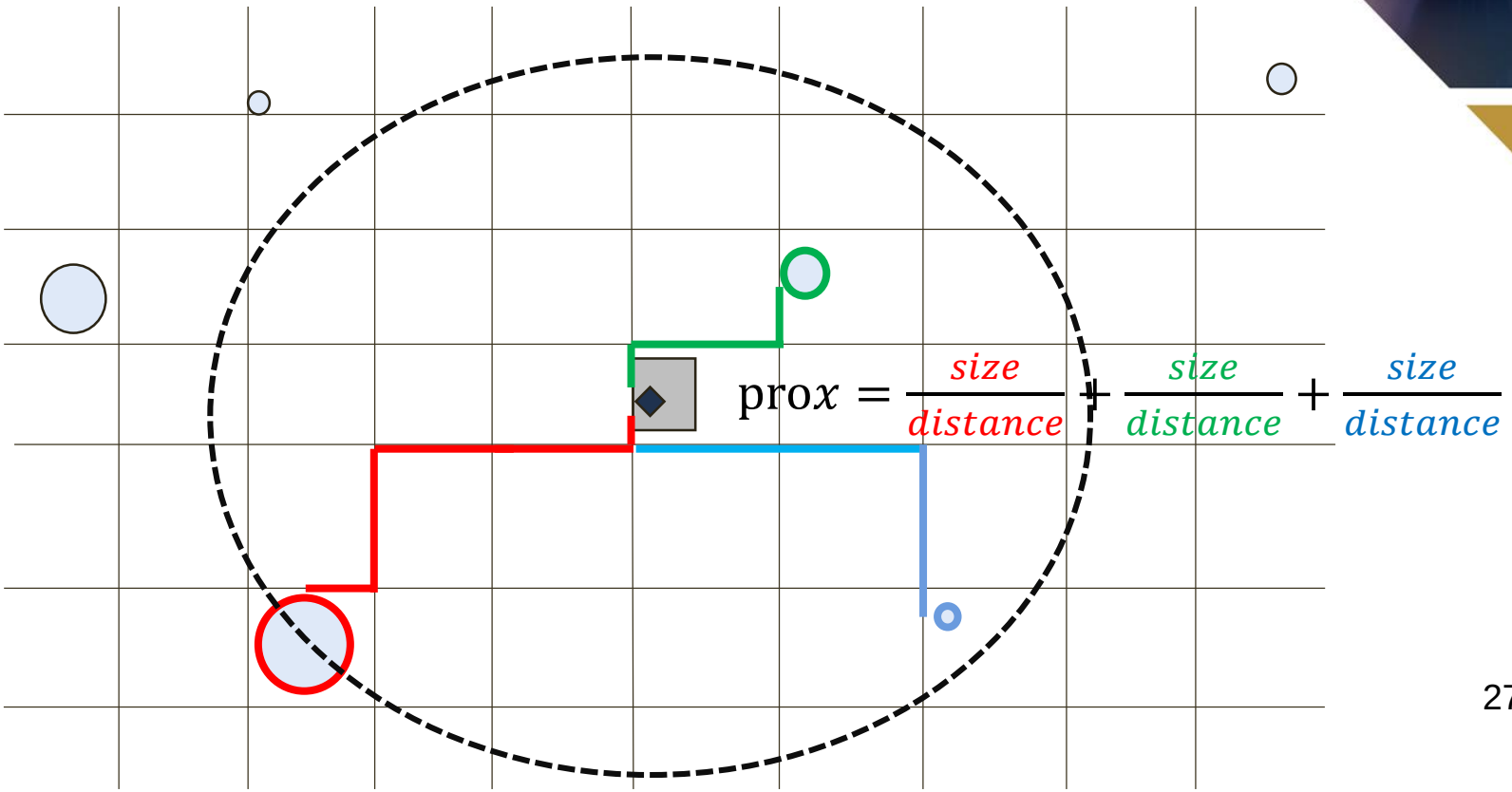
What can we do with the data?
Example - Proximity Measures



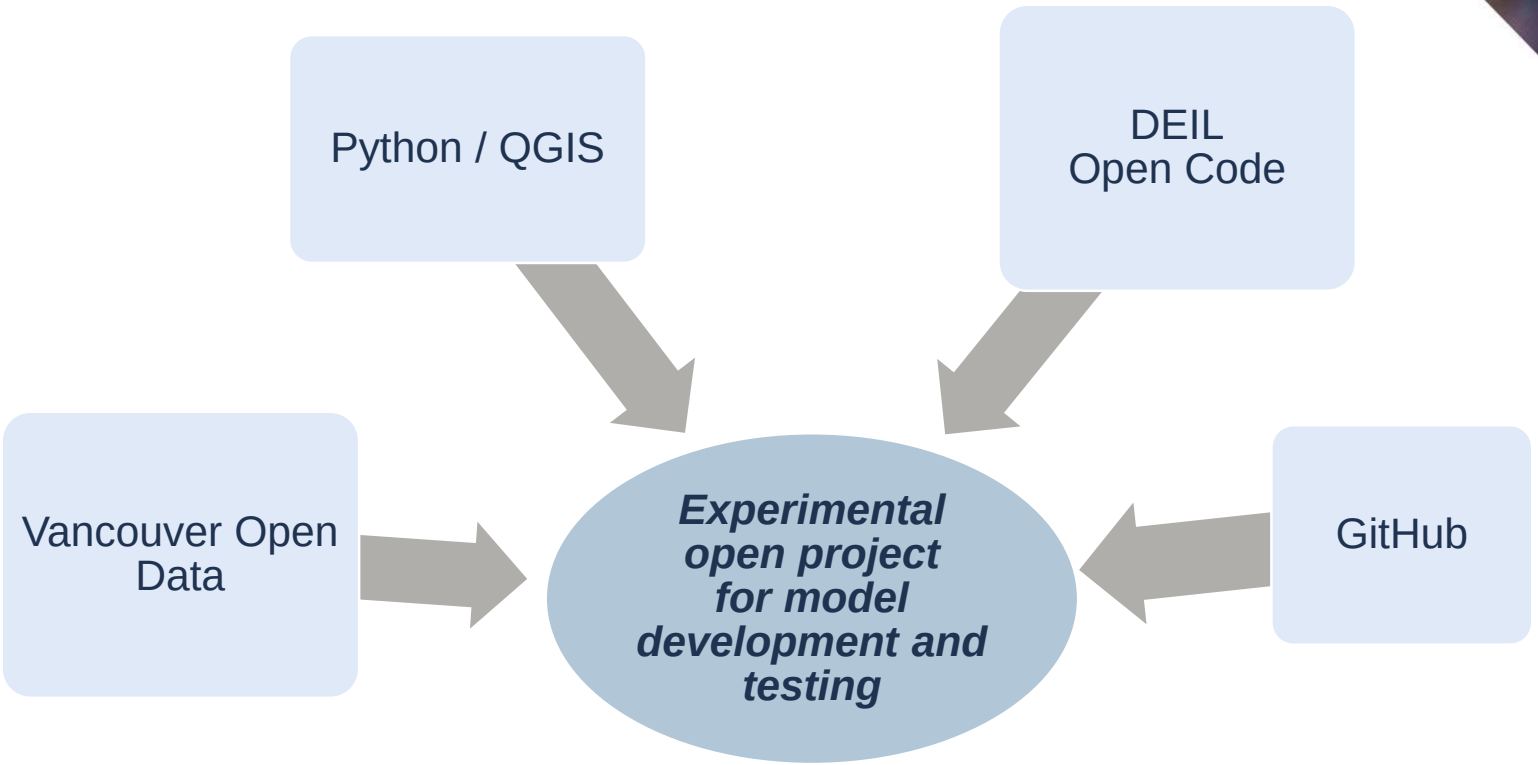
What can we do with the data?
Example - Proximity Measures



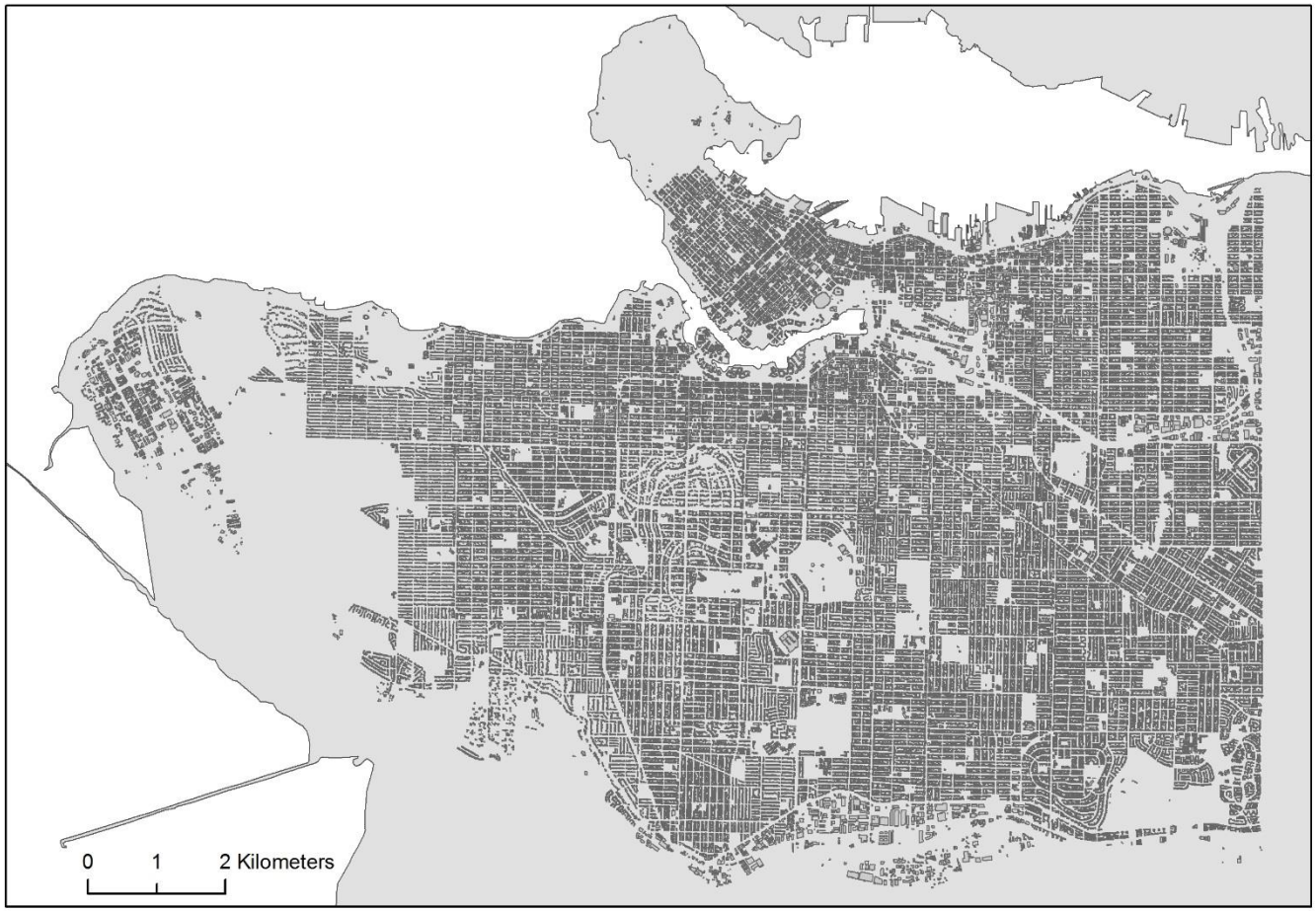
What can we do with the data?
Example - Proximity Measures



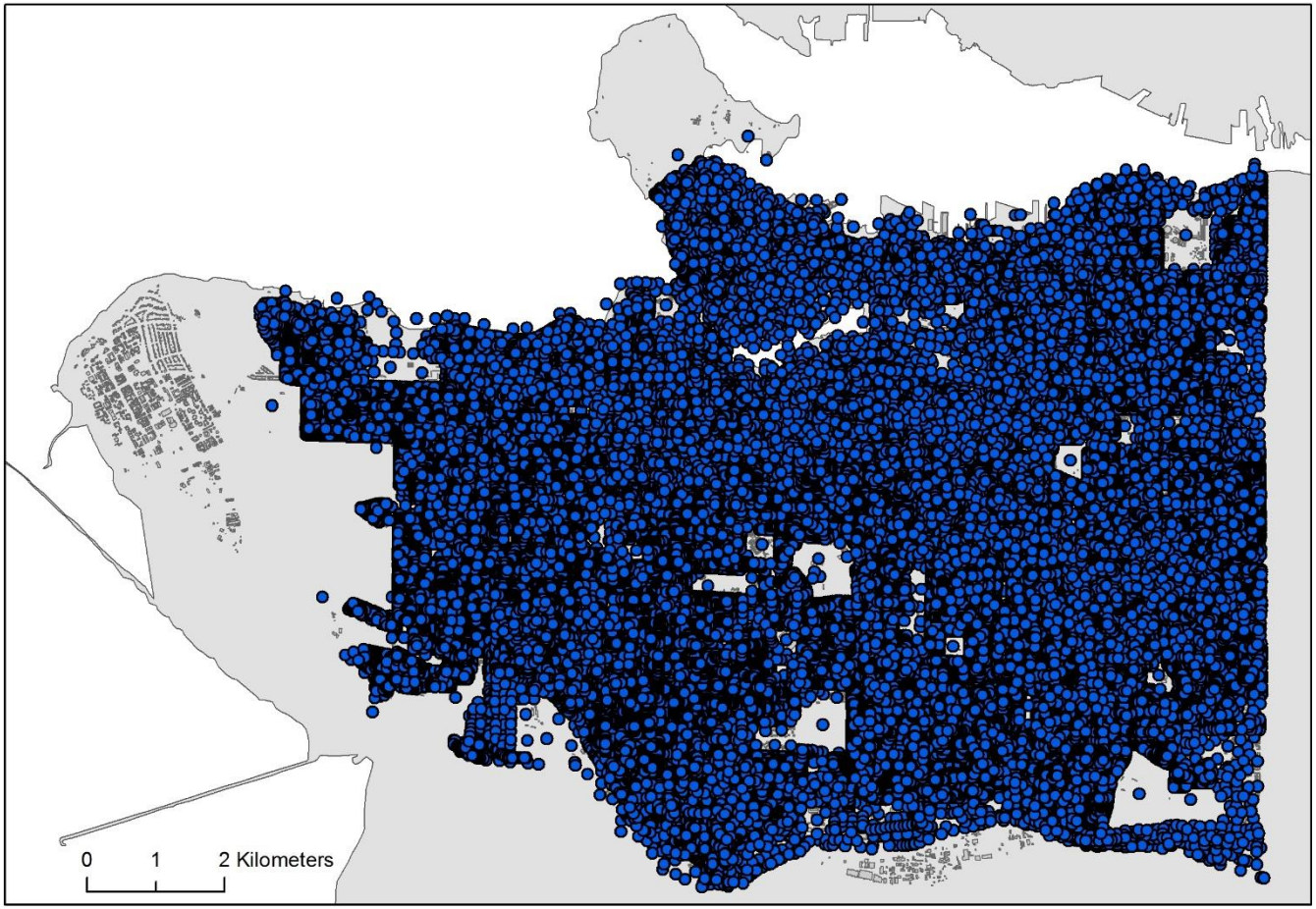
Is it possible with an open project approach?



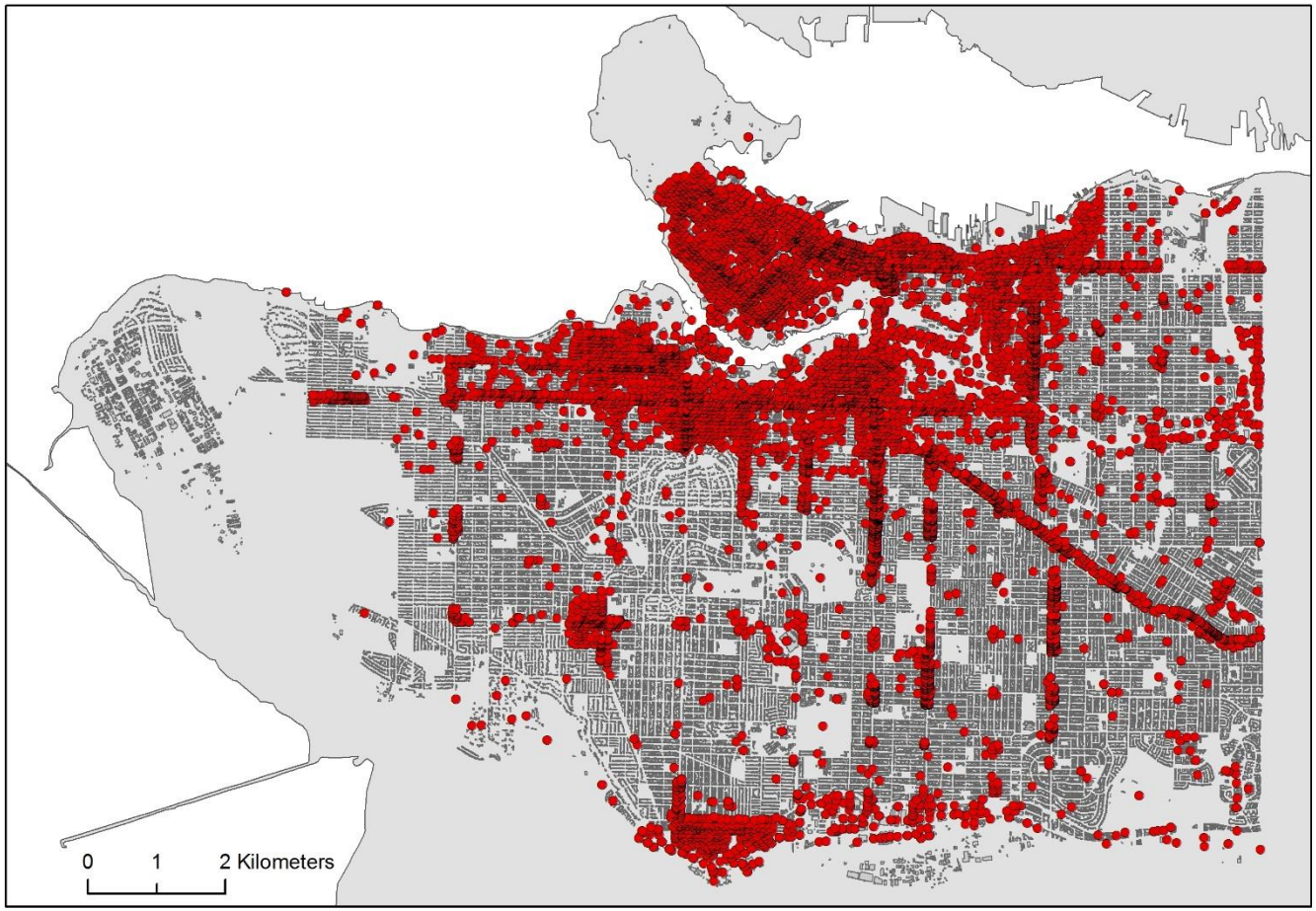
Example of Vancouver Buildings



Example of Vancouver Addresses



Example of Vancouver Businesses



Example of Vancouver Buildings + addresses

Example of Vancouver Buildings+addreses+businesses

Example of Vancouver Buildings+addresses+food businesses

Challenges

- Coverage of open microdata
- Governance and licensing
- Heterogeneity of formats, concepts, definitions

Take home messages

- Different options for use and integration of open data
- Great opportunities for integrating and using open microdata from local/prov./fed. government sources
  - Buildings, businesses, addresses and much more
- Great potential for conducting this work in open spaces as an “open project”
  - You can contribute, develop, re-use!

THANK YOU!

For more information,

alessandro.alasia@canada.ca
jean.lemoullec@canada.ca
haaris.jafri@canada.ca

MERCI!

Pour de plus amples renseignements,

alessandro.alasia@canada.ca
jean.lemoullec@canada.ca
haaris.jafri@canada.ca

#StatCan100