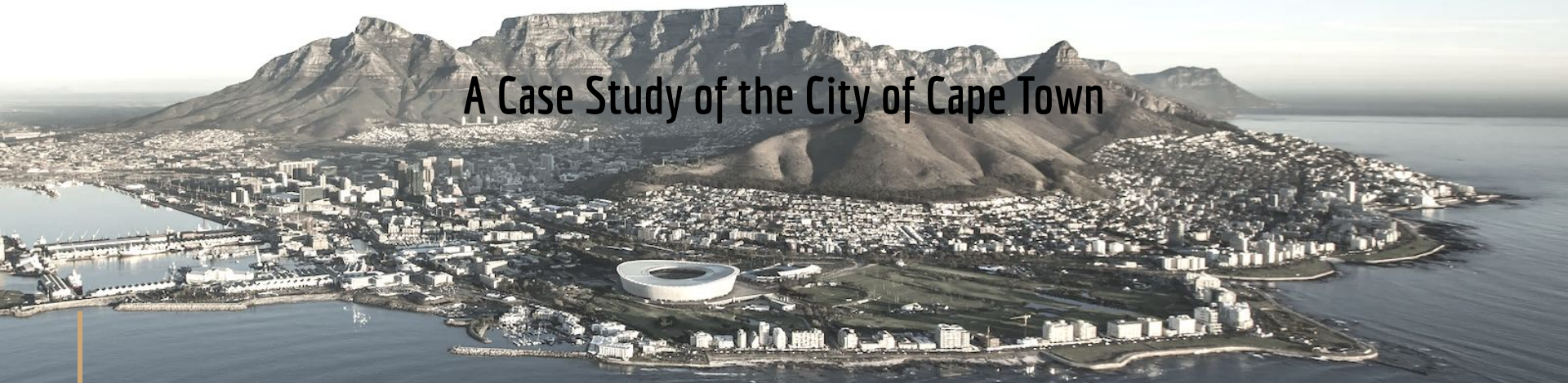
Aligning internal data capabilities with external research partnerships:

A Case Study of the City of Cape Town



Hugh Cole, Kelsey Jack, Derek Strong and Brendan Maughan-Brown

Outline

1. Motivation and background
2. Data use examples
3. Using administrative data for research
4. A framework for research collaboration
5. Work in progress

The authors

Hugh Cole: Director of Policy and Strategy, City of Cape Town

Kelsey Jack: Associate Professor, University of California - Santa Barbara, Co-Chair of J-PAL's Energy, Environment and Climate Change Sector

Derek Strong: Research Computing Associate, Center for Advanced Research Computing, University of Southern California

Brendan Maughan-Brown: Research advisor, J-PAL Africa, University of Cape Town

Motivation and background

A story of two perspectives:

1) **City of Cape Town**: Democratically elected local government for ~4 million residents of Cape Town, South Africa
    - Service provision responsibilities include electricity, water, sanitation, refuse, transportation, housing, emergency services, primary healthcare, environmental health, community development
    - Commitment to evidence-based policy-making and leveraging data for effective governance
    - 2016 restructuring laid the groundwork, including hiring of Hugh Cole

Motivation and background

A story of two perspectives:

2) **J-PAL Africa and UCSB**: Researchers engaged in collaborations with the City of Cape Town that relied on administrative data
    - Collaborations revealed both the strength of CCT's administrative data and areas for improvement
    - Also highlighted challenges sharing data with researchers, both in South Africa and internationally
    - These challenges also face data analysts and decision-makers *within* the municipal government

Motivation and background

Capitalizing on shared interests and goals:

1) Lower the time burden of identifying, standardizing and sharing datasets
2) Improve security, transparency and reciprocity of data sharing relationships
3) Identify opportunities for research -- by both internal and external researchers -- to contribute to policy

Data use examples

Three cases highlighted needs of both parties:

1) Impacts of pre-paid electricity metering (research collaboration)
2) Data use for planning and policy during Cape Town’s drought (policy)
3) Responding to and recovering from the COVID-19 pandemic (policy)

Data use examples

Three cases highlighted needs of both parties:

1) **Impacts of pre-paid electricity metering (research collaboration)**
2) Data use for planning and policy during Cape Town's drought (policy)
3) Responding to and recovering from the COVID-19 pandemic (policy)

Data use example: Research collaboration

Impacts of prepaid electricity metering

Question: How does prepaid electricity metering affect residential use and City costs and revenue, relative to postpaid metering (monthly billing)?

Collaboration: A randomized phase in of meter replacements

- Close coordination of operations, research design and data flow
- Iterative process that worked around CCT logistical constraints

Data use example: Research collaboration

Impacts of prepaid electricity metering

Data needs: Electricity data from multiple sources

- Billing data from SAP system, Vending data from PoS system, GIS data on properties, contractor data on installations

Data challenges: Complicated data flow requiring in-person interactions to transfer data, use of administrative data both for design and for study outcomes, and multi-step process for linking across data sources

Data use example: Research collaboration

Impacts of prepaid electricity metering

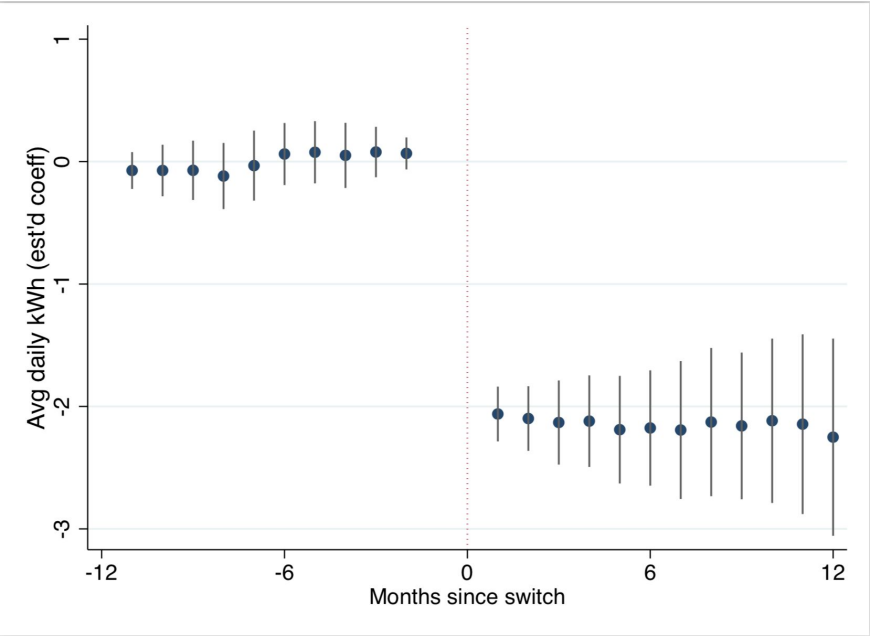
Institutional set up:

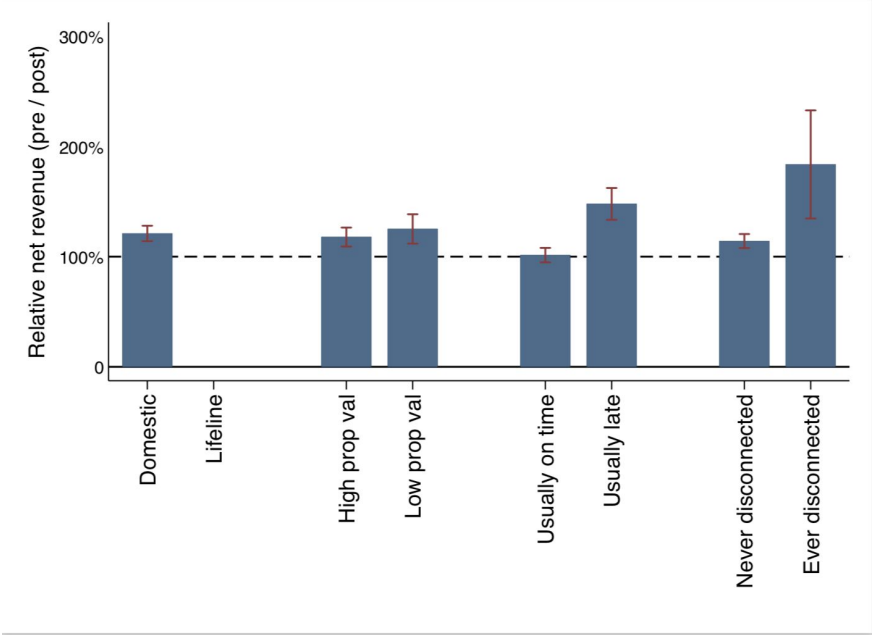
- Built on an existing research relationship between CCT's electricity department and a PhD student at UCT, Grant Smith
- Existing Data Use Agreement with CCT, which was modified to include Kelsey Jack

Outcome: Jack, B.K. and G. Smith (2020) "Charging ahead: Prepaid electricity metering in South Africa" *American Economic Journal: Applied Economics*, 12(2).

Data use example: Research collaboration

Impacts of prepaid electricity metering





Data use examples

Three cases highlighted needs of both parties:

1) Impacts of pre-paid electricity metering (research collaboration)
2) **Data use for planning and policy during Cape Town's drought (policy)**
3) Responding to and recovering from the COVID-19 pandemic (policy)

Data use example: Policy

Policy challenge: Historic drought in Cape Town threatened CCT's water supply, led to threat of "Day Zero"

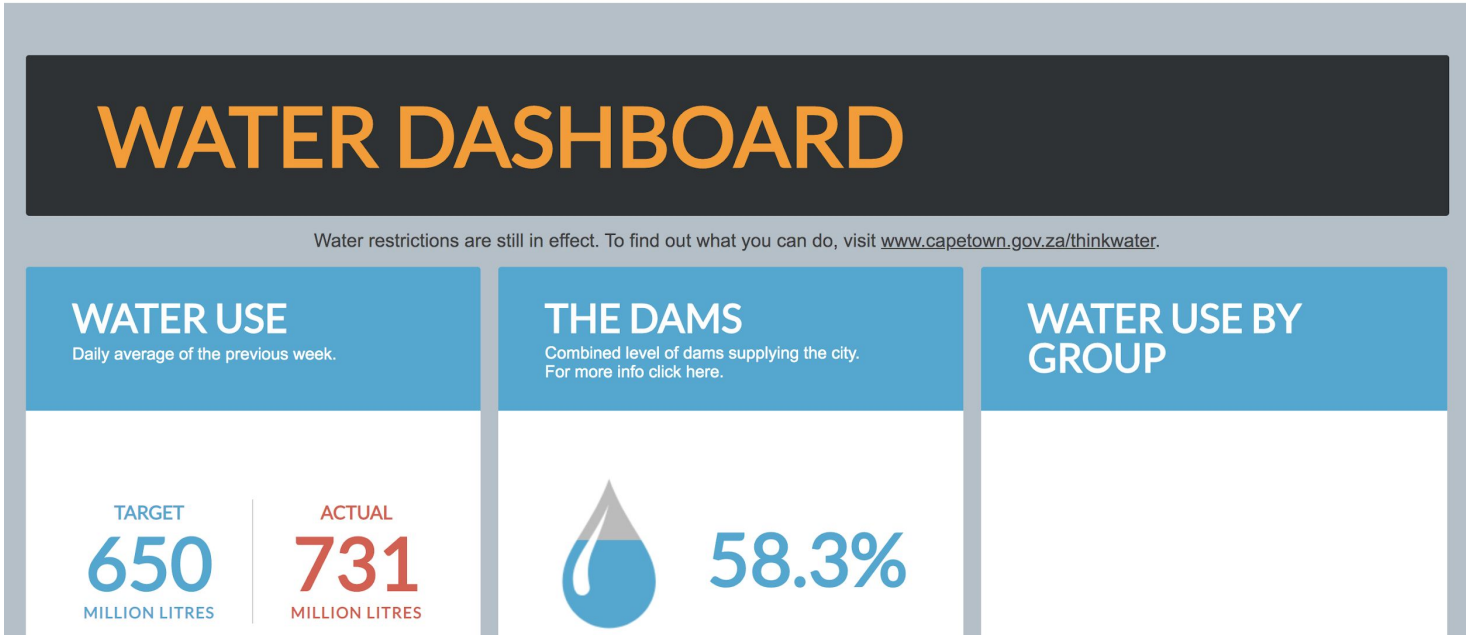
Data needs: Internal use for communication, behavior change and water management (including pressure reductions, infrastructure upgrades)

Data challenges: Real-time data sharing across departments and between contractors and CCT, geo-referencing and communicating data to the public

Outcome: Massive decline in water use allowed CCT to avoid Day Zero

Data use example: Policy

CITY OF CAPE TOWN
ISIXEKO SASEKAPA
STAD KAAPSTAD



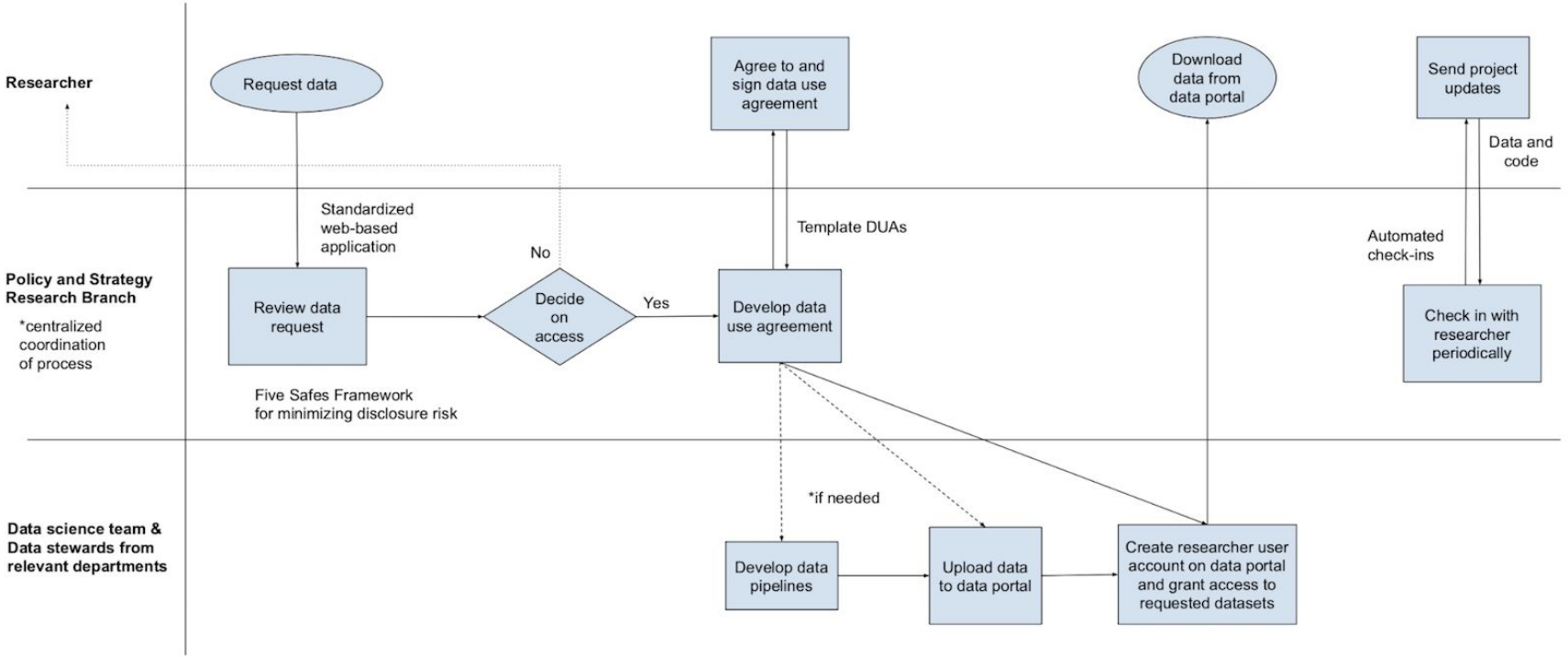
Using administrative data for research

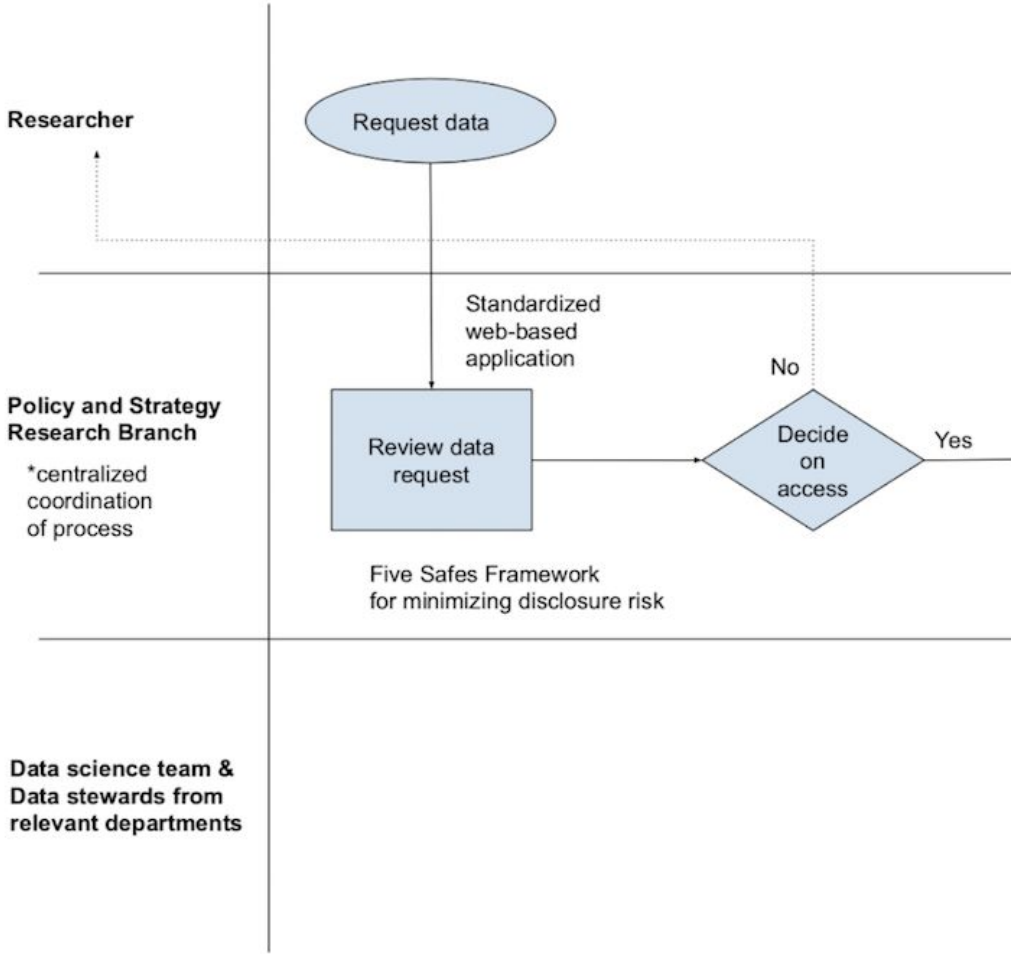
Goal: Make administrative datasets more accessible for both internal and external research

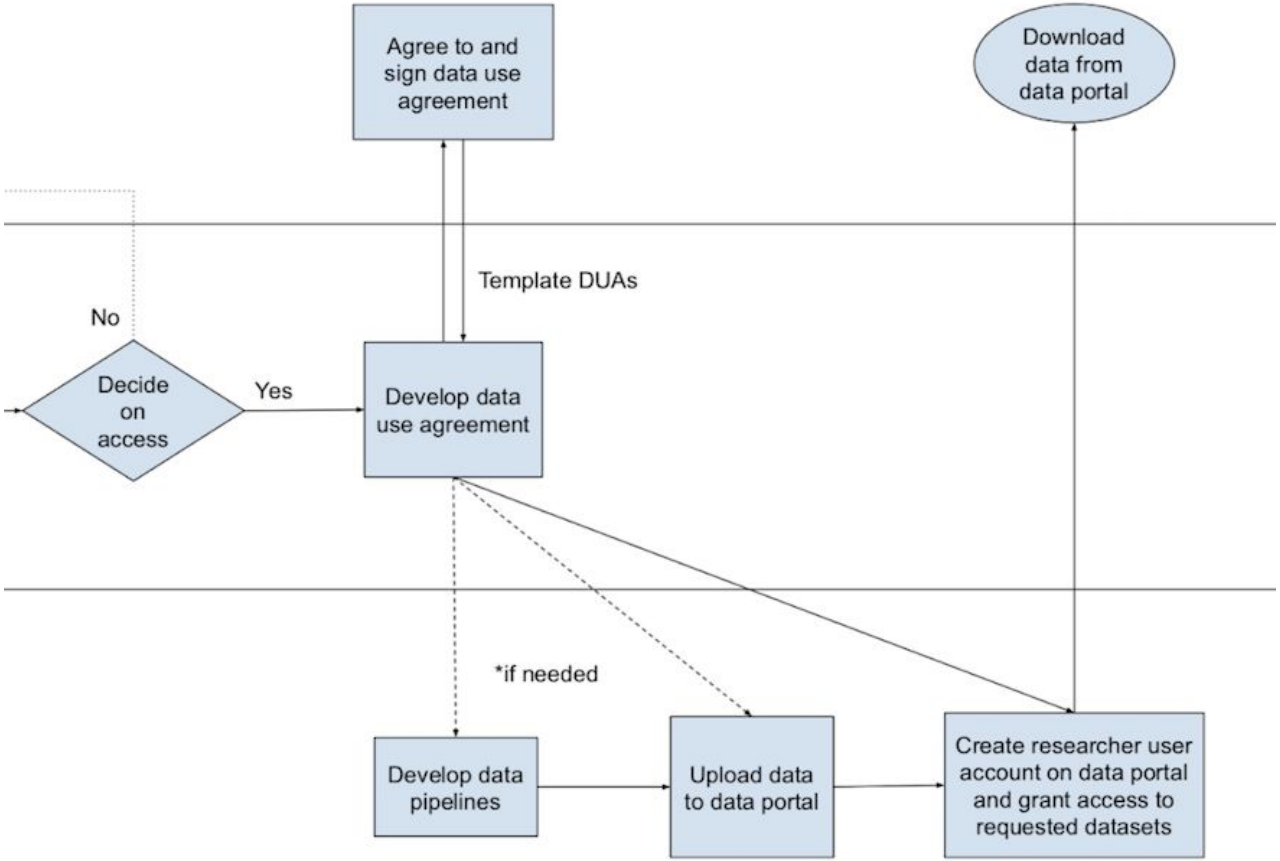
Status quo challenges:

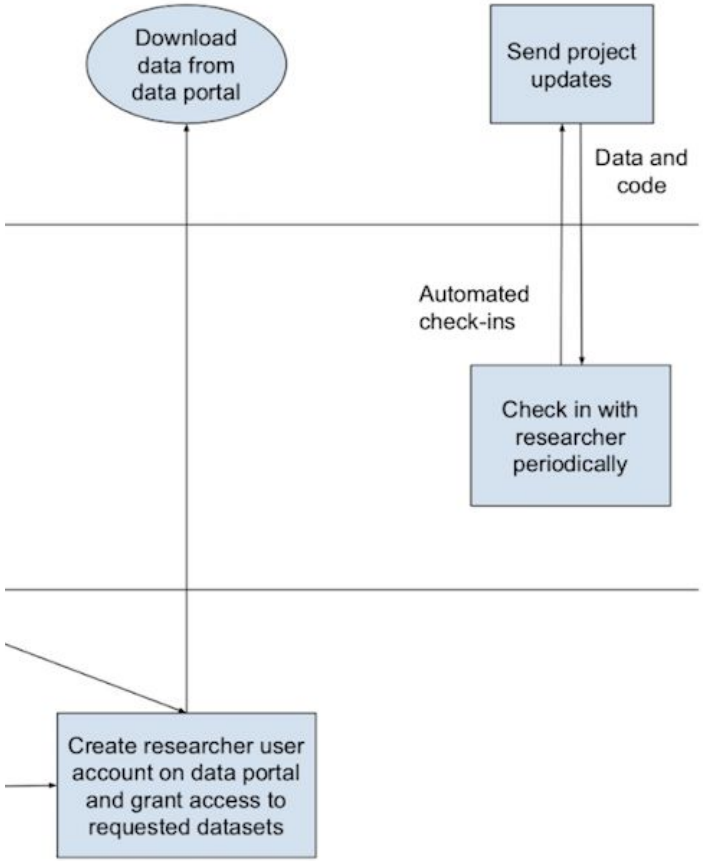
- Variety of sources generating different types of data stored in different formats and managed by different people
  - Electricity and water, billing, transportation, GIS, etc.
- Ownership of data → data stewards
- Make data FAIR (findable, accessible, interoperable, and re-usable)
- Maintain anonymity and security of restricted-use data

A framework for research collaboration




Researcher
Request data
Standardized
web-based
application
No
Policy and Strategy
Research Branch
*centralized
coordination
of process
Review data
request
Decide
on
access
Yes
Five Safes Framework
for minimizing disclosure risk
Data science team &
Data stewards from
relevant departments


Agree to and sign data use agreement
Download data from data portal
Template DUAs
No
Decide on access
Yes
Develop data use agreement
*if needed
Develop data pipelines
Upload data to data portal
Create researcher user account on data portal and grant access to requested datasets


Download data from data portal
Send project updates
Data and code
Automated check-ins
Check in with researcher periodically
Create researcher user account on data portal and grant access to requested datasets

A framework for research collaboration

Solutions to the status quo challenges:

- Policy foundation drafted and accepted by CCT: *Data Strategy* and *Research Framework*
  - *Data Strategy* lays out data management process, approaches to lowering the costs of sharing and combining data across sources, creates specific roles within CCT
  - *Research Framework* clarifies procedures for sharing data with external partners, updates research management practices, stresses reciprocal exchange of value and prioritizes research that will inform policy

A framework for research collaboration

Solutions to the status quo challenges:

- Investment in people: Chief Data Officer, Organizational Policy and Planning, Data science team, many others
- Investment in technology and infrastructure
  - Data sharing platform → CKAN, downloading/uploading, metadata
  - Data APIs to populate platform, reduce time burden for data stewards
- Data sharing process: Revised workflow and SOPs, data inventory (over 1000 research-relevant datasets), searchable metadata

Work in progress

Implementation of the *Data Strategy* and *Research Framework* are underway in Cape Town

Data sharing platform being used on a limited basis

- Uploads by data stewards and downloads by researchers

COVID-19 has both slowed things down and highlighted the importance of remote data sharing

- Data use example #3: Nascent data sharing platform has been used heavily for internal purposes

Comments? Questions?

Hugh Cole: Hugh.Cole@capetown.gov.za

Kelsey Jack: kelseyjack@ucsb.edu

Derek Strong: dstrong@protonmail.com

Brendan Maughan-Brown: bmaughan@povertyactionlab.org