GOVLAB

THE PROMISE AND CHALLENGES OF

"DATA COLLABORATIVES" ON MIGRATION

STEFAAN G. VERHULST

WORKSHOP
Big Data and Alternative Data Sources on Migration

ISPRA, November 30th, 2017

CHALLENGE

IMPROVING DISASTER RELIEF BY EXCHANGING MOBILE PHONE DATA

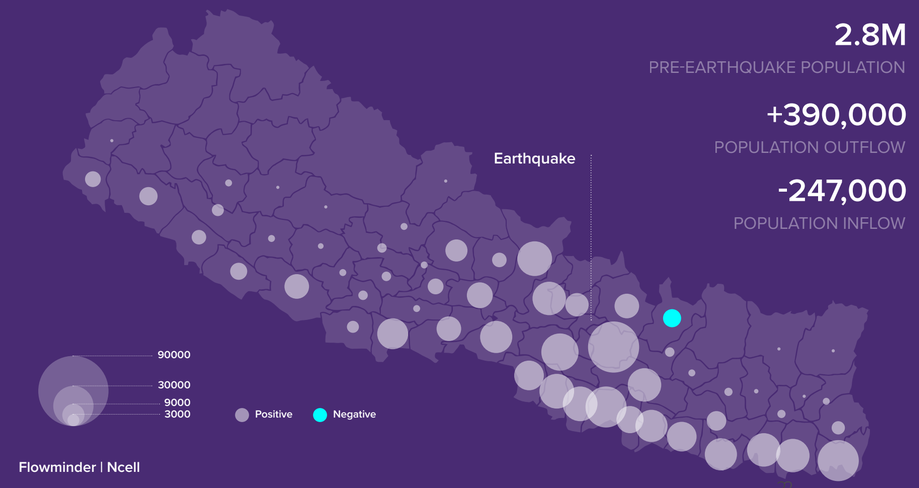
On April 25th, 2015, a violent earthquake hit Nepal—the worst of its kind since 1934. The damage left hundreds of thousands of people homeless and flattened entire villages. Ultimately, the Gorkha earthquake killed nearly 9,000 people and injured nearly 22,000.

Yet, the death toll could have been much worse.

DATA COLLABORATIVES

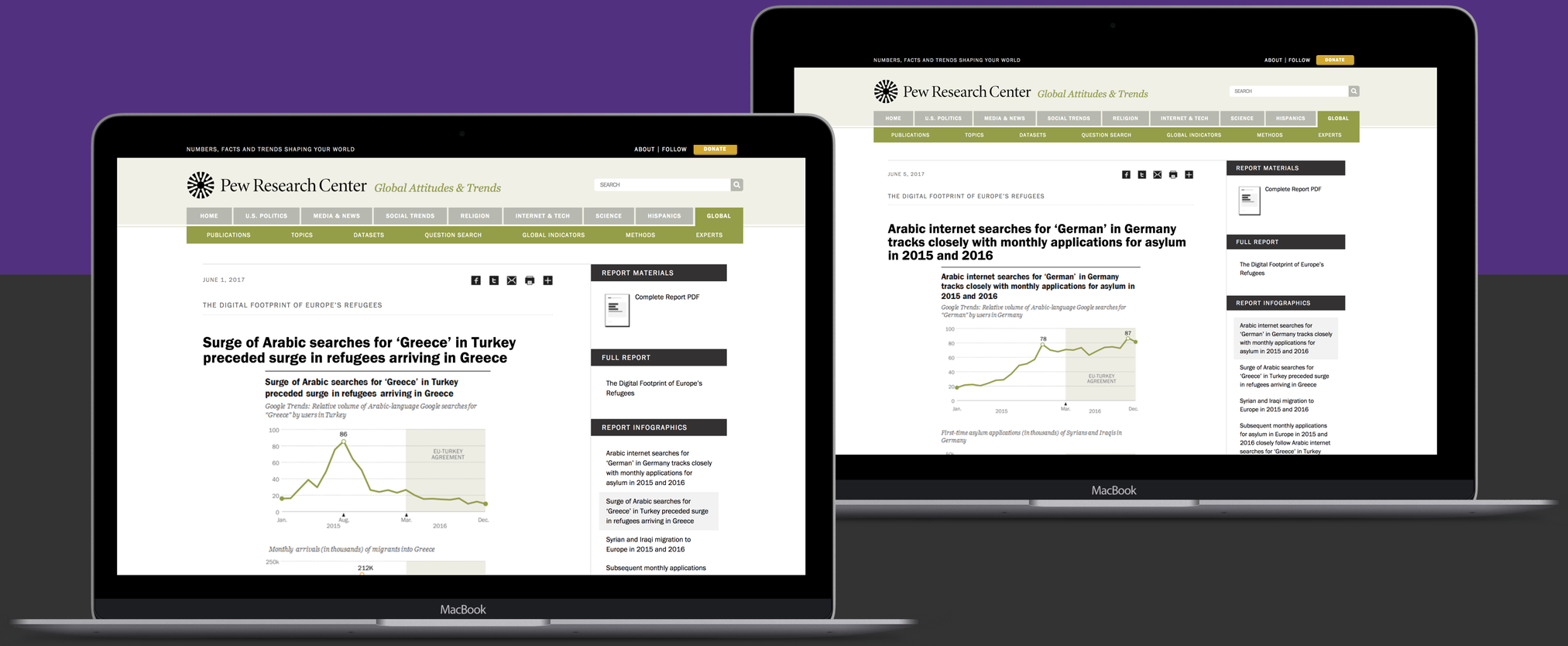
NCEL, Nepal's largest mobile operator, **shared anonymized mobile phone data** with the non-profit Swedish organization Flowminder. With this data, Flowminder mapped where and how people moved in the wake of the disaster, and shared this information with government and UN agencies to assist their relief efforts.

The Data Collaborative between NCEL and Flowminder allowed humanitarian organizations to better target aid to affected communities - saving hundreds if not thousands of lives.



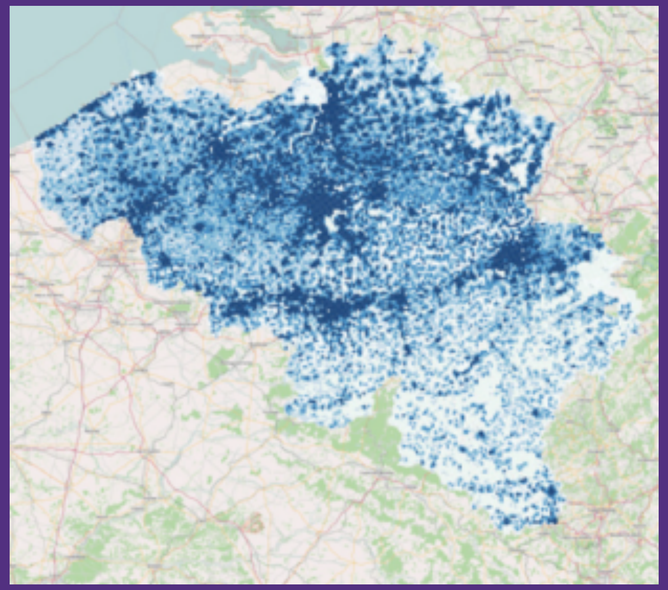
GOVLAB

GOOGLE SEARCHES AND REFUGEES

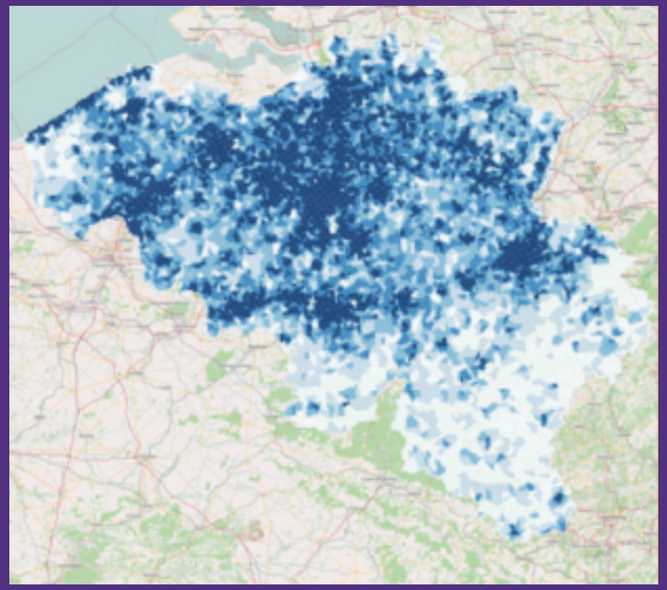


GOVLAB

POPULATION ESTIMATES THROUGH MOBILE PHONE DATA IN BELGIUM



BASED ON 2011 CENSUS

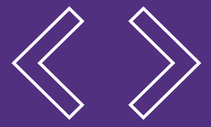


BASED ON MOBILE PHONE DATA

GOVLAB

PRIVATE DATA

NEW DATA SOURCES



Web Crawling
Web Scraping
Web Search Analysis

Social Media

Telecom Data

Sensor and
Geospatial Data

Commercial Transactions
Scanner Data
Credit Card Data

GOVLAB

ADVANCES IN AI, DATA AND COMPUTING SCIENCE

SENTIMENT ANALYSIS

NATURAL LANGUAGE PROCESSING

SCRAPING

CDR ANALYTICS

DATA MINING

MACHINE LEARNING

PREDICTIVE ANALYTICS

ETC.

GOVLAB

PRIVATE DATA FOR PUBLIC POLICY

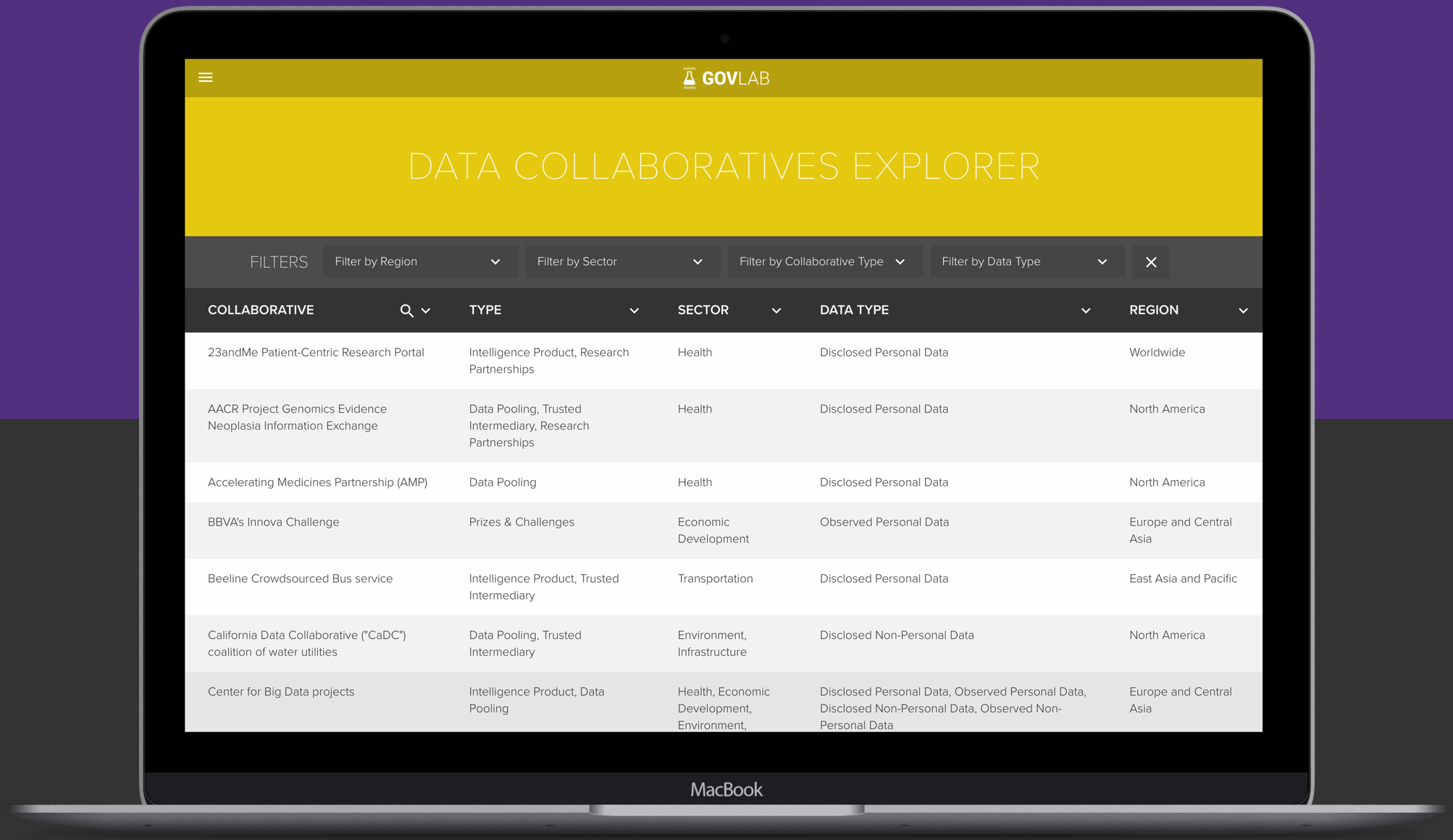
=

NEW PUBLIC PRIVATE PARTNERSHIPS FOR A DATA AGE

=

DATA COLLABORATIVES

GOVLAB



COLLABORATIVE	TYPE	SECTOR	DATA TYPE	REGION
23andMe Patient-Centric Research Portal	Intelligence Product, Research Partnerships	Health	Disclosed Personal Data	Worldwide
AACR Project Genomics Evidence Neoplasia Information Exchange	Data Pooling, Trusted Intermediary, Research Partnerships	Health	Disclosed Personal Data	North America
Accelerating Medicines Partnership (AMP)	Data Pooling	Health	Disclosed Personal Data	North America
BBVA's Innova Challenge	Prizes & Challenges	Economic Development	Observed Personal Data	Europe and Central Asia
Beeline Crowdsourced Bus service	Intelligence Product, Trusted Intermediary	Transportation	Disclosed Personal Data	East Asia and Pacific
California Data Collaborative ("CaDC") coalition of water utilities	Data Pooling, Trusted Intermediary	Environment, Infrastructure	Disclosed Non-Personal Data	North America
Center for Big Data projects	Intelligence Product, Data Pooling	Health, Economic Development, Environment,	Disclosed Personal Data, Observed Personal Data, Disclosed Non-Personal Data, Observed Non-Personal Data	Europe and Central Asia

GOVLAB

FIVE WAYS PRIVATE DATA CAN INFORM PUBLIC POLICY

SITUATIONAL AWARENESS AND RESPONSE

A greater understanding and ability to track conditions on the ground can improve interventions, including especially during emerging situations like crisis relief.

PUBLIC SERVICE DESIGN AND DELIVERY

Access to previously inaccessible datasets can enable more accurate modelling of public service design and guide service delivery in a targeted, evidence-based manner.

KNOWLEDGE CREATION AND TRANSFER

Bringing more and more diverse datasets to bear can fill knowledge gaps and ensure that those responsible for solving problems have the most useful information at hand.

PREDICTION AND FORECASTING

New predictive capabilities enabled by access to datasets can help institutions be more proactive, putting in place mechanisms based on sound evidence that mitigate problems or avert crises before they occur.

IMPACT ASSESSMENT AND EVALUATION

Access to additional datasets can help institutions monitor and evaluate the real-world impacts of policies and inform iteration.

GOVLAB

SIX TYPES OF DATA COLLABORATIVES

DATA COOPERATIVES
OR POOLING

PRIZES
& CHALLENGES

RESEARCH
PARTNERSHIPS

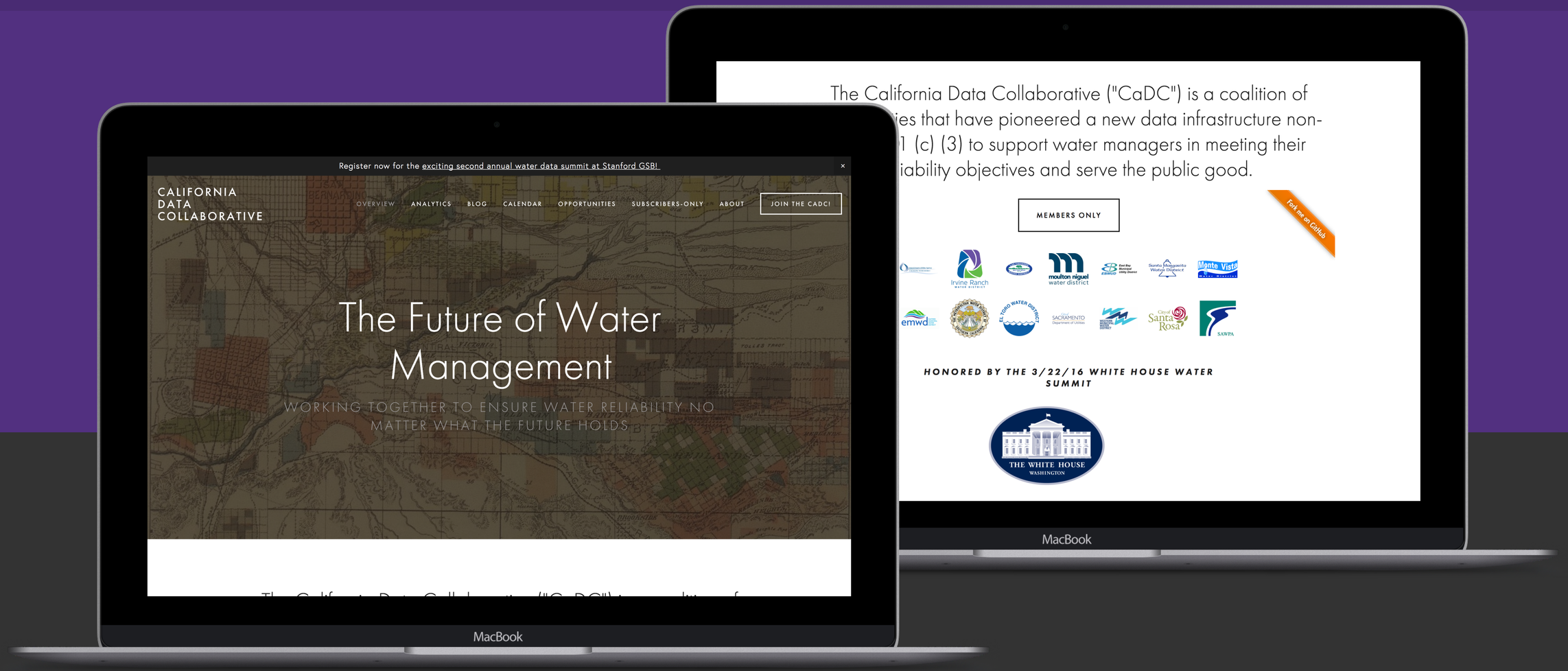
INTELLIGENCE
PRODUCTS

APPLICATION PROGRAMMING
INTERFACES (APIS)

TRUSTED
INTERMEDIARY

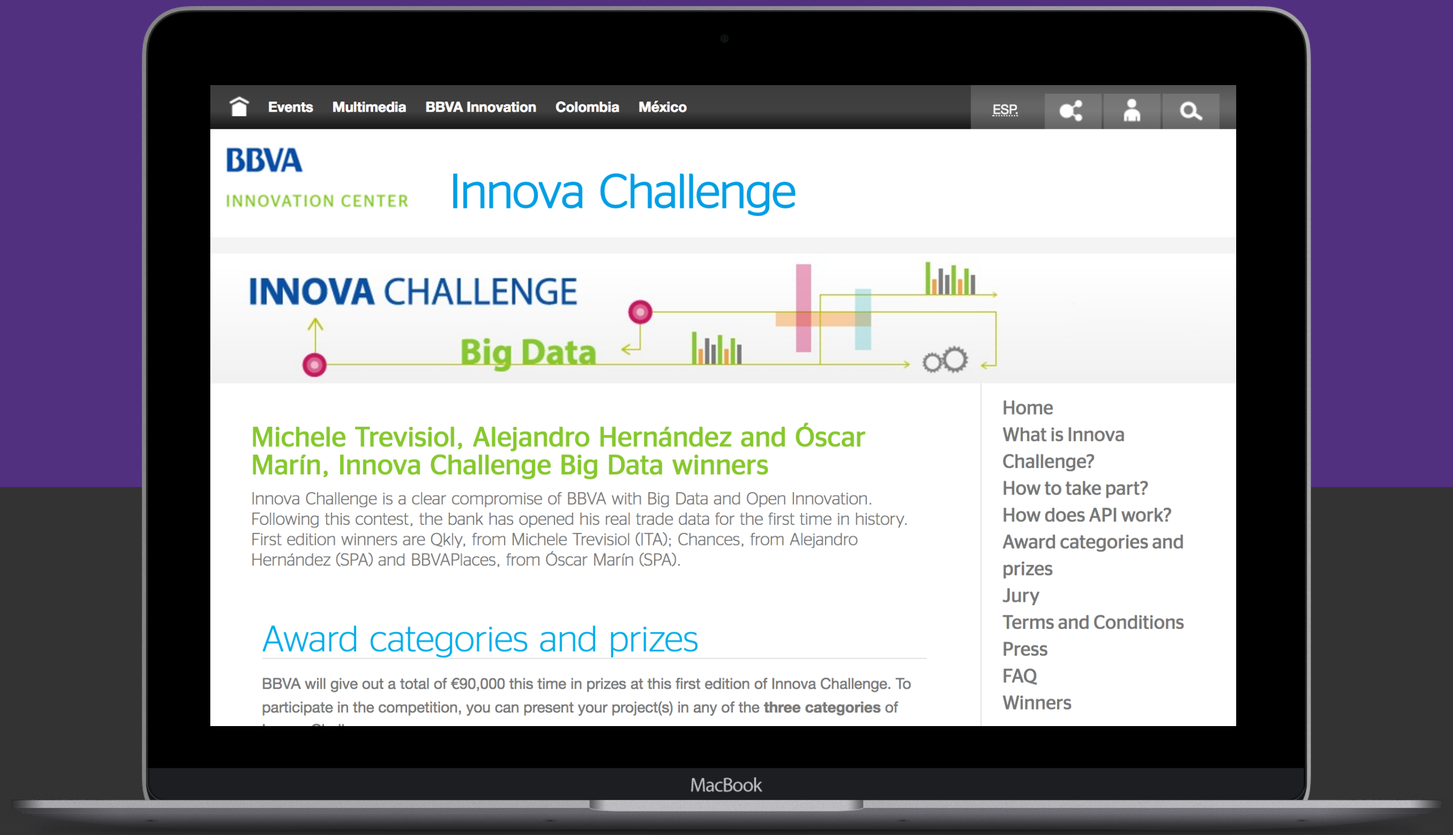
GOVLAB

DATA POOLING



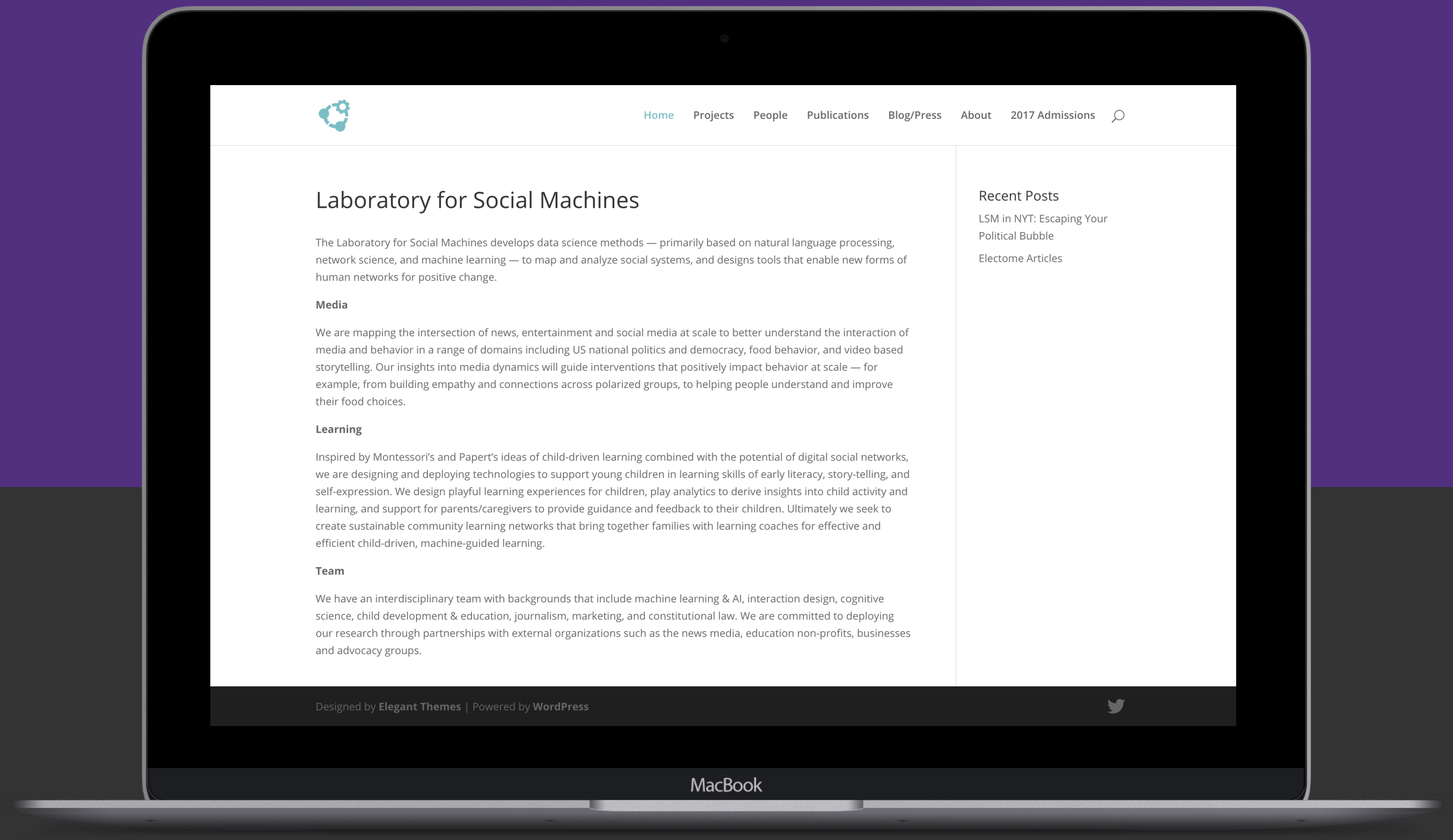
GOVLAB

PRIZES AND CHALLENGES



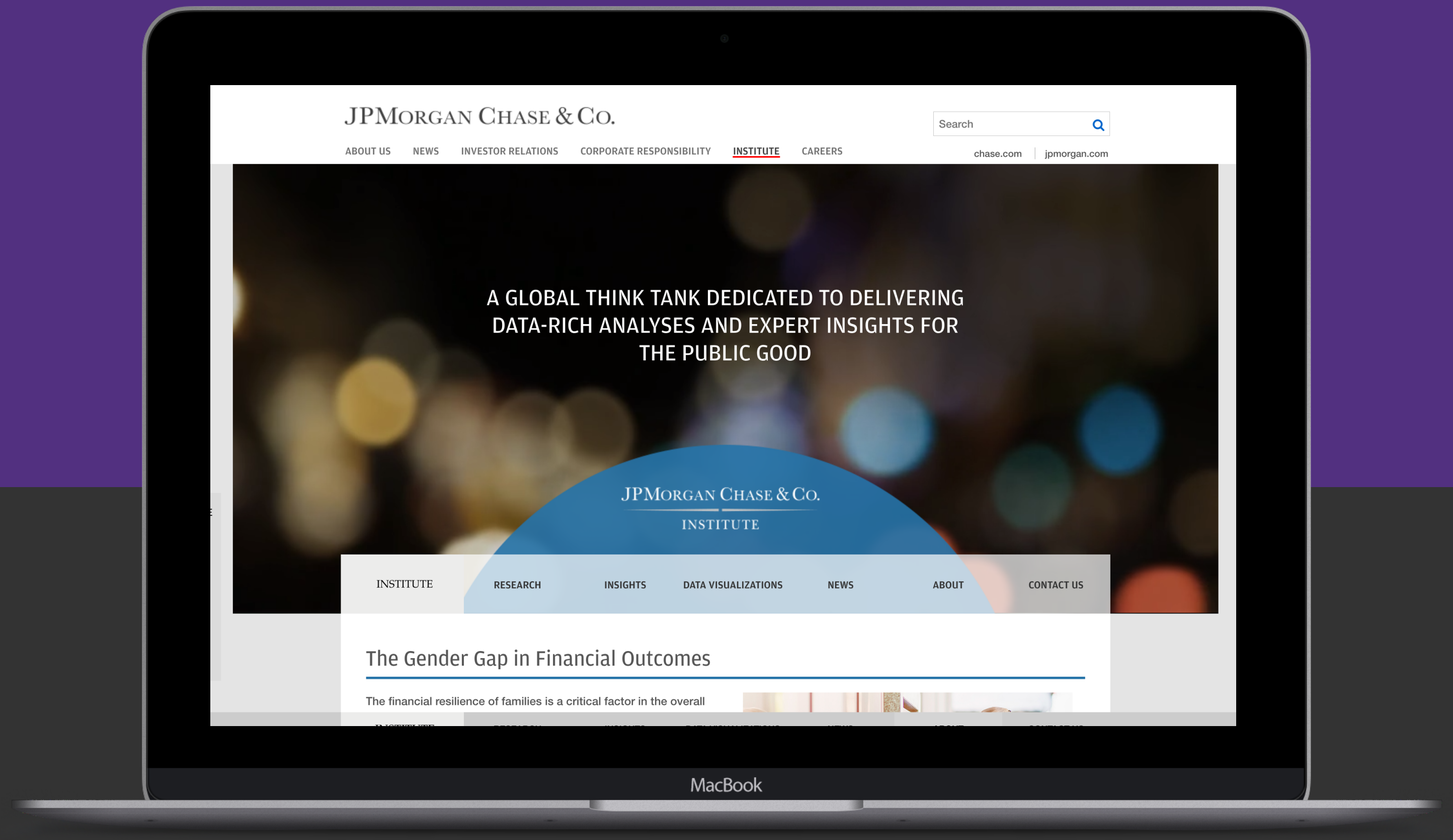
GOVLAB

RESEARCH PARTNERSHIPS



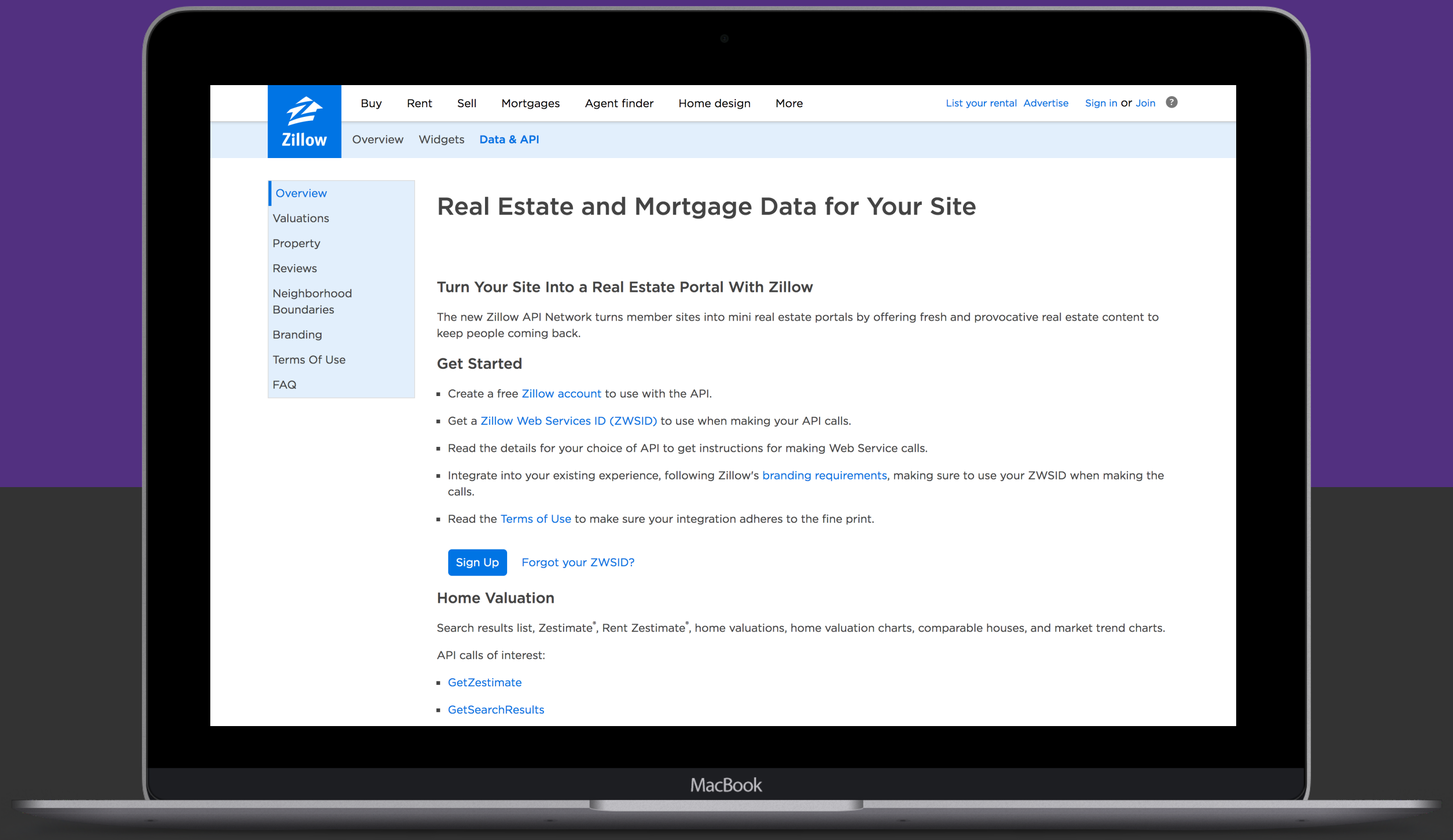
GOVLAB

INTELLIGENCE PRODUCTS



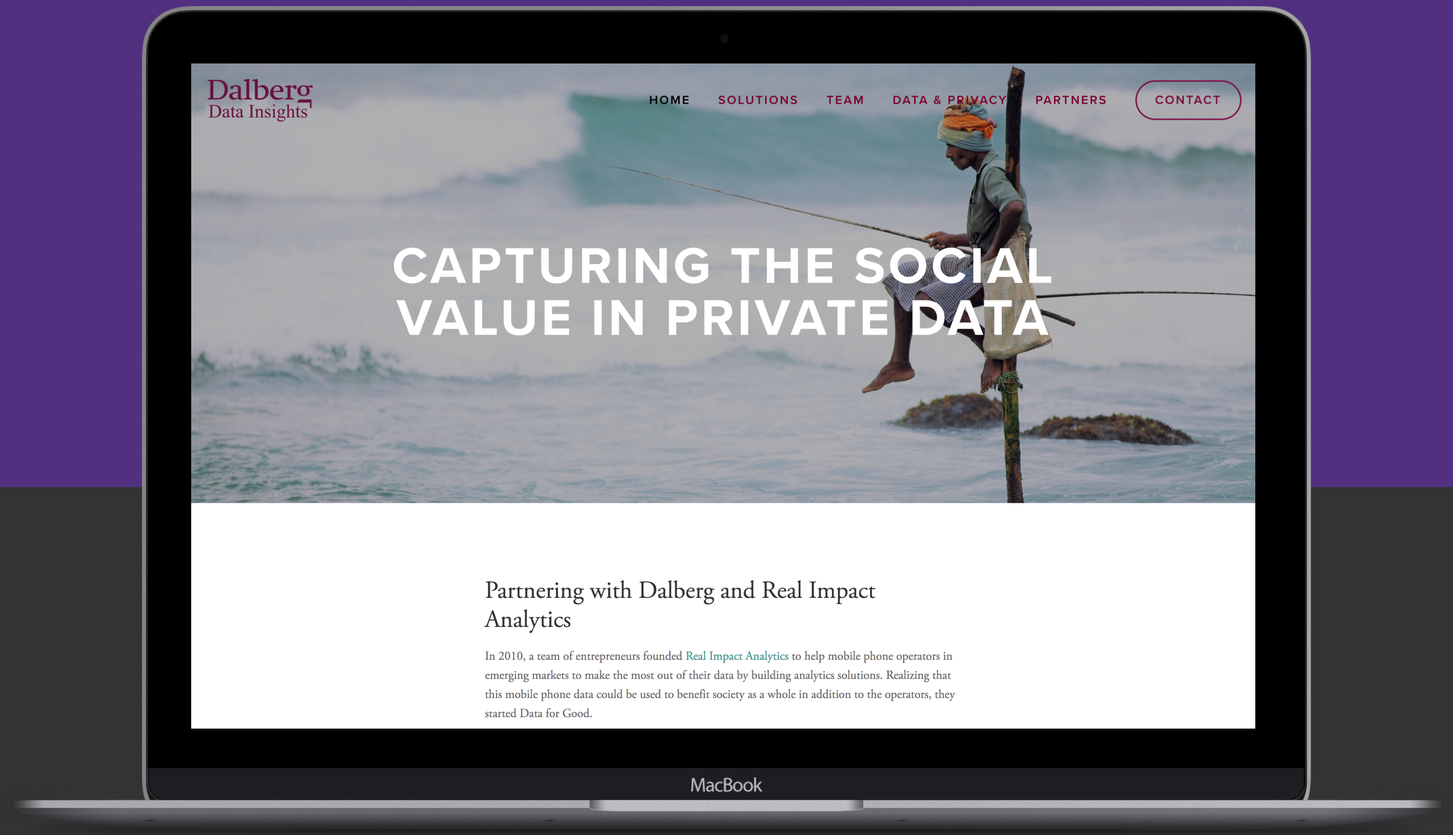
GOVLAB

API


Zillow
Buy
Rent
Sell
Mortgages
Agent finder
Home design
More
List your rental
Advertise
Sign in or Join
Overview
Widgets
Data & API
Overview
Valuations
Property
Reviews
Neighborhood Boundaries
Branding
Terms Of Use
FAQ
Real Estate and Mortgage Data for Your Site
Turn Your Site Into a Real Estate Portal With Zillow
The new Zillow API Network turns member sites into mini real estate portals by offering fresh and provocative real estate content to keep people coming back.
Get Started
Create a free Zillow account to use with the API.
Get a Zillow Web Services ID (ZWSID) to use when making your API calls.
Read the details for your choice of API to get instructions for making Web Service calls.
Integrate into your existing experience, following Zillow's branding requirements, making sure to use your ZWSID when making the calls.
Read the Terms of Use to make sure your integration adheres to the fine print.
Sign Up
Forgot your ZWSID?
Home Valuation
Search results list, Zestimate®, Rent Zestimate®, home valuations, home valuation charts, comparable houses, and market trend charts.
API calls of interest:
GetZestimate
GetSearchResults
MacBook

GOVLAB

TRUSTED INTERMEDIARY



GOVLAB

MOTIVATIONS TO SHARE: THE SIX Rs BEHIND CORPORATE DATA SHARING

RECIPROCITY

REVENUE

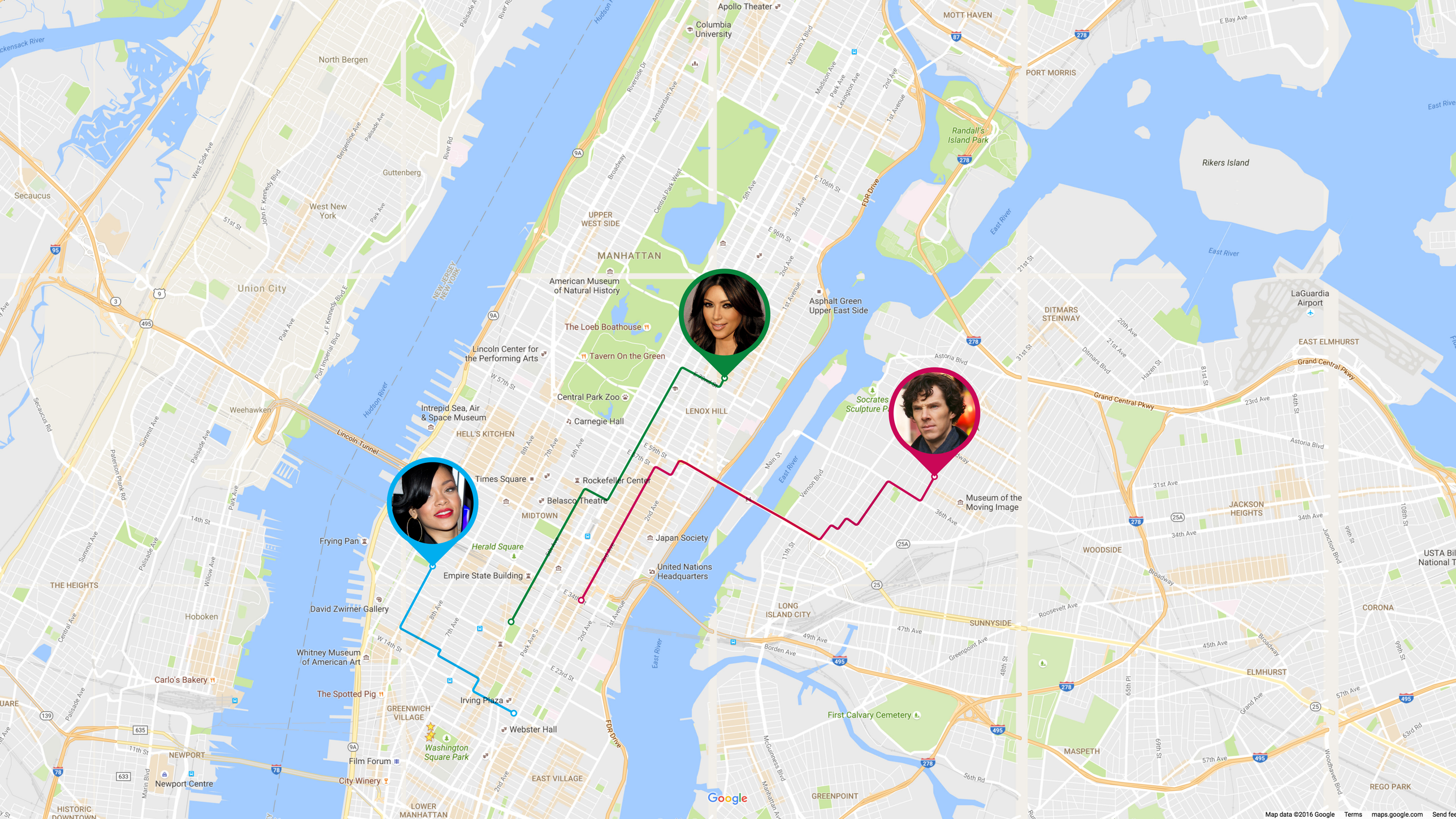
RESEARCH & INSIGHTS

REGULATORY COMPLIANCE

REPUTATION & RETAINMENT OF TALENT

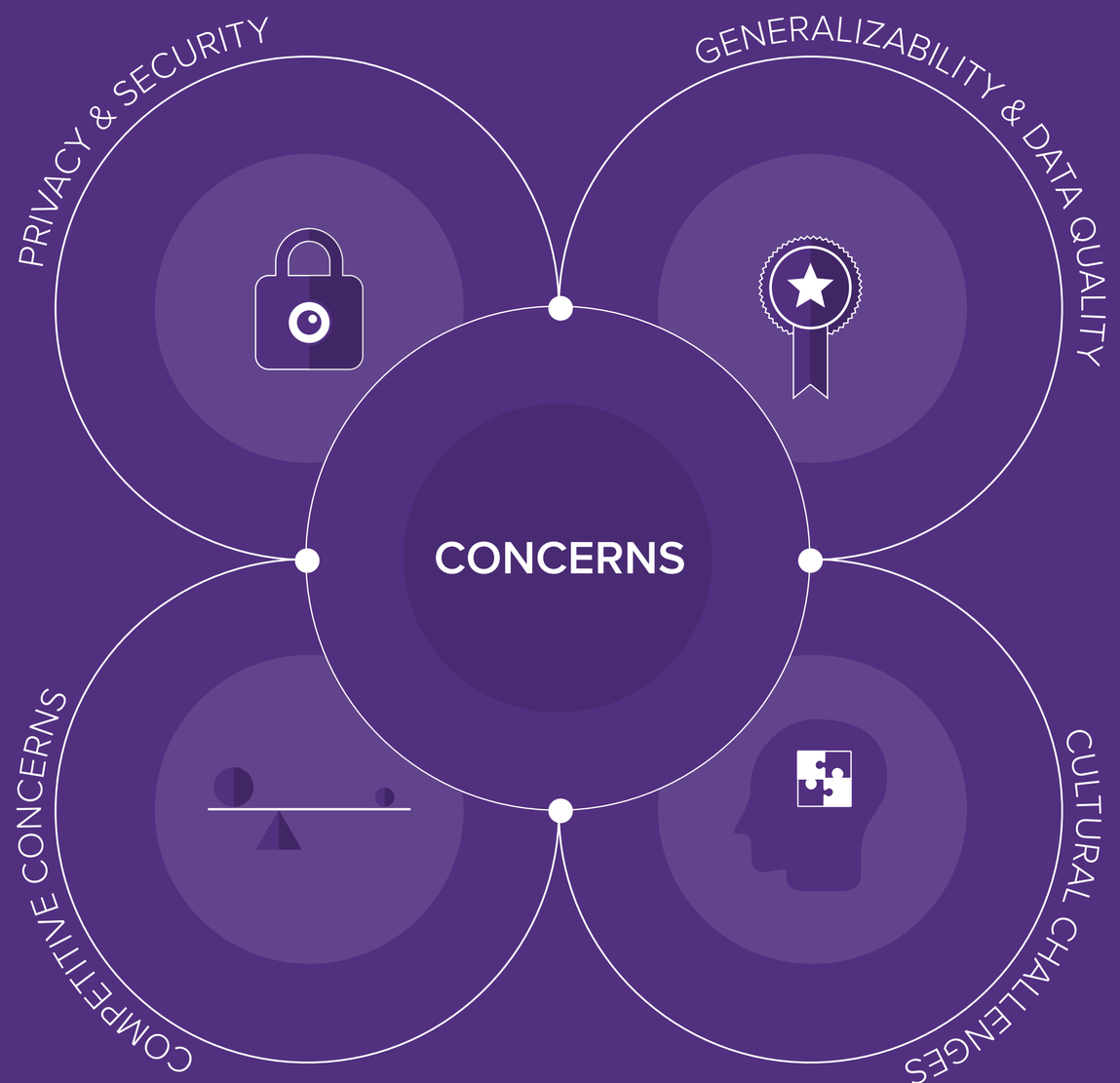
RESPONSIBILITY

TLC
TRIP DATA


Apollo Theater
Columbia University
MOTT HAVEN
PORT MORRIS
North Bergen
Guttenberg
West New York
Secaucus
Union City
Weehawken
Randall's Island Park
Rikers Island
East River
UPPER WEST SIDE
MANHATTAN
American Museum of Natural History
The Loeb Boathouse
Lincoln Center for the Performing Arts
Tavern On the Green
Central Park Zoo
Carnegie Hall
LENOX HILL
Asphalt Green Upper East Side
Intrepid Sea, Air & Space Museum
HELL'S KITCHEN
Times Square
Rockefeller Center
Belasco Theatre
MIDTOWN
Frying Pan
Herald Square
Empire State Building
Japan Society
United Nations Headquarters
Museum of the Moving Image
Socrates Sculpture P
DITMARS STEINWAY
LaGuardia Airport
EAST ELMHURST
JACKSON HEIGHTS
WOODSIDE
LONG ISLAND CITY
SUNNYSIDE
CORONA
ELMHURST
MASPETH
REGO PARK
GREENPOINT
First Calvary Cemetery
THE HEIGHTS
Hoboken
David Zwirner Gallery
Whitney Museum of American Art
Carlo's Bakery
The Spotted Pig
GREENWICH VILLAGE
Irving Plaza
Webster Hall
Washington Square Park
Film Forum
City Winery
EAST VILLAGE
LOWER MANHATTAN
NEWPORT
Newport Centre
HISTORIC DOWNTOWN
Google
Map data ©2016 Google
Terms
maps.google.com
Send fe

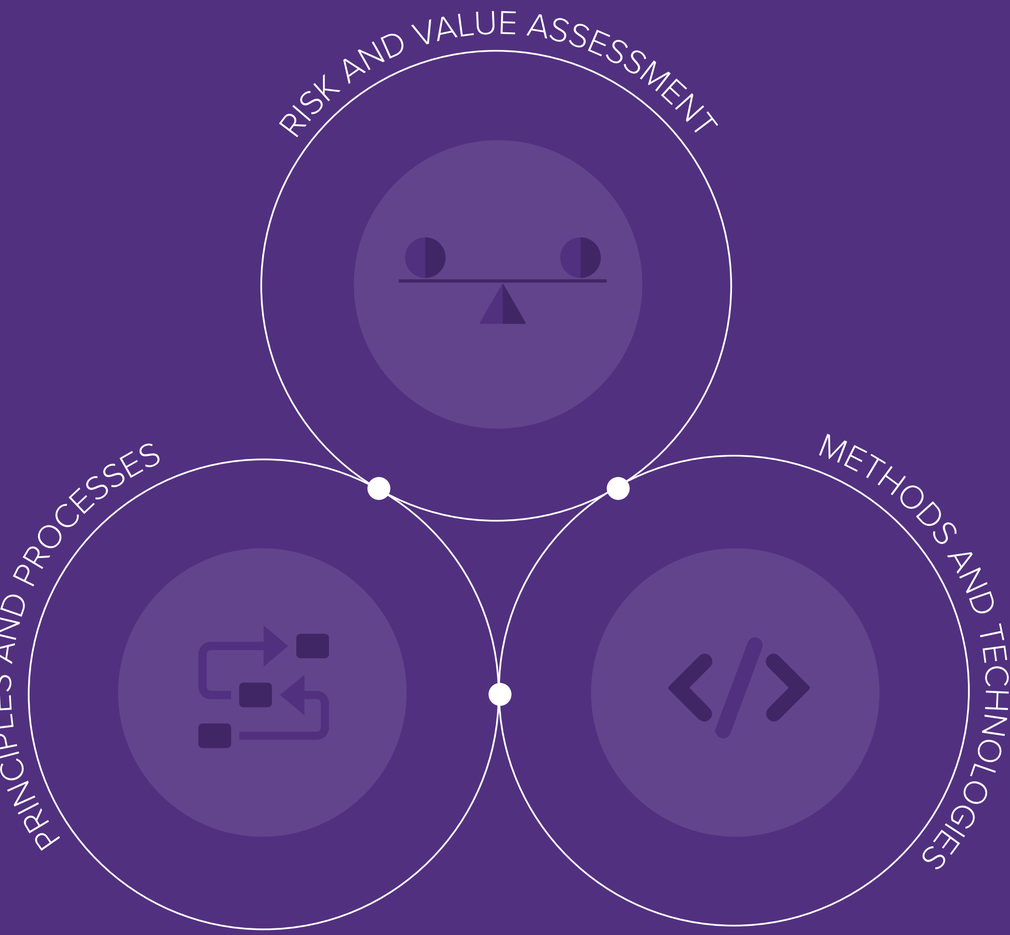
NEW YORK DOLLS Gentlemen
AMISH
EQUINOX
ONE WAY
POLICE
NYPD
5269

GOVLAB


PRIVACY & SECURITY
GENERALIZABILITY & DATA QUALITY
CONCERNS
COMPETITIVE CONCERNS
CULTURAL CHALLENGES

GOVLAB

TOWARD DATA RESPONSIBILITY



GOVLAB

TOWARD DATA RESPONSIBILITY

VALUE AND RISK ASSESSMENT

GOVLAB

Collecting inaccurate, old or “dirty” data affecting data quality and ability draw meaningful insights

Unauthorized or intrusive data collection potentially leading to privacy harms;

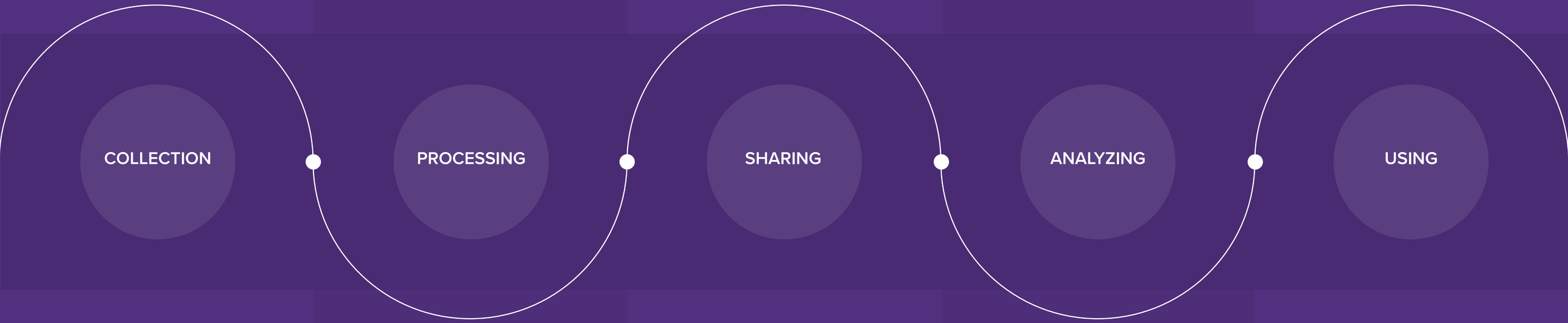
Insufficient, outdated or inflexible security provisions creating the potential for data vulnerabilities or breaches;

Lack of interoperable cultural and institutional norms and expectations, creating a difficult environment to collaborate toward mutual benefit;

Lack of data stewardship at both ends to ensure the responsible use of personally identifiable information as it travels across use cases and sectors;

Poor problem definition or research design, potentially leading to data being analyzed in a way that does not add value toward the ultimate objective

When data is ultimately put to use, risks emerge especially from collaborative organizations using shared data controversially or incongruously in relation to the original objective for its collection and/ or the original consent provided by the data subject (if any). Such risks can have negative results like the misinterpretation of data, the re-identification of individual data subjects, and decisional interference



Incomplete or non-representative sampling of the universe – e.g., ignoring “data invisibles,” or population segments with a limited data footprint – potentially leading to non-inclusive or unrepresentative approaches or interventions.

Aggregation and correlation of incompatible datasets can create ‘apples and oranges’ scenarios where the eventual sharing and analysis of commingled datasets are doomed for failure.

Improper or unauthorized access to shared data as it passes between entities, whether by unsanctioned actors inside or outside of collaborating organizations;

Conflicting legal jurisdictions and different levels of security within collaborating entities, making the eventual congruous data use difficult.

Inaccurate data modeling or use the of biased algorithms, which, like dirty data at the Collection stage, can lead to confidence in fundamentally flawed insights.

Additionally, at the Using stage, many of the risks from previous stages could yield true, identifiable harms for the first time – e.g., a negatively impactful policy decision being made based on faulty data from the Collection stage.

GOVLAB

TOWARD DATA RESPONSIBILITY

PRINCIPLES AND PROCESSES

GOVLAB

UPDATING FIPPS/IRB?

THE SIGNAL CODE

The Signal Code articulates five human rights to information during crisis:

- THE RIGHT TO INFORMATION
- THE RIGHT TO PROTECTION FROM HARM
- THE RIGHT TO DATA SECURITY AND PRIVACY
- THE RIGHT TO DATA AGENCY
- THE RIGHT TO REDRESS AND RECTIFICATION

EVOLVING THE IRB

Building Robust Review for Industry Research

Molly Jackman
Facebook

Lauri Kanerva
Facebook

GOVLAB

TOWARD DATA RESPONSIBILITY

METHODS AND TECHNOLOGIES

GOVLAB

NEW METHODS AND TECHNOLOGIES

DECISION TREES

Differential Privacy?

GOVLAB

REALIZING THE POTENTIAL OF PRIVATE DATA 4 PUBLIC GOOD

STEWARDS

METHODS

EVIDENCE

MOVEMENT

THANK YOU

stefaan@thegovlab.org

datacollaboratives.org