"OPPORTUNITIES OF DATA SHARING AND THE GLOBAL DATA ACCESS FRAMEWORK (DATA COMMONS)

30 January 2020

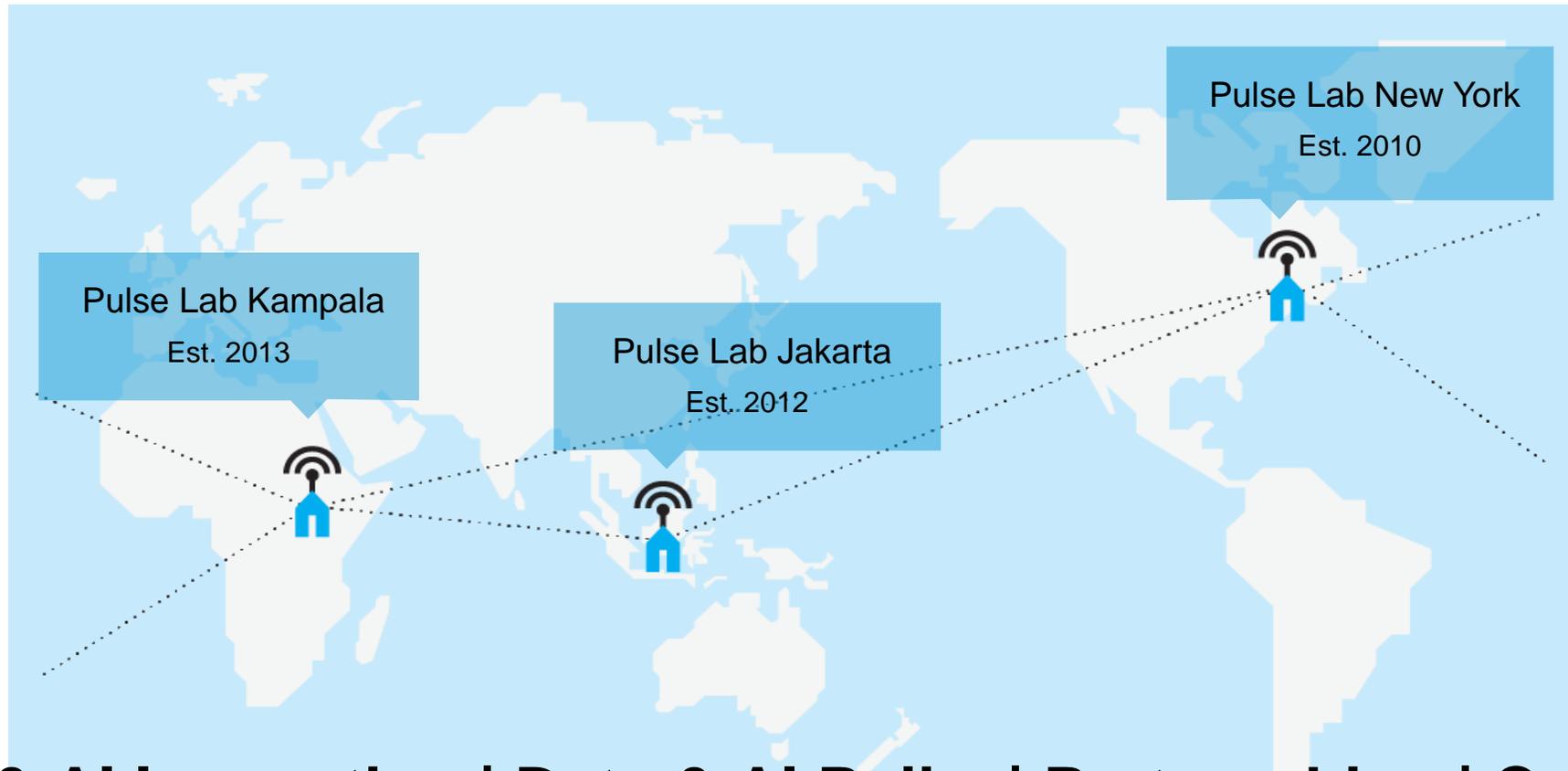**Mila Romanoff,**
Data Governance and Policy Lead, UN Global Pulse,
Executive Office of the UN Secretary-General

Global Pulse
An innovation initiative of the UN Secretary-General

Vision: A future in which big data is used responsibly for the public good

Mission: Accelerate discovery and adoption of data analytics for sustainable   development and humanitarian action

Pulse Lab New York
Est. 2010

Pulse Lab Kampala
Est. 2013

Pulse Lab Jakarta
Est. 2012

Data & AI Innovation | Data & AI Policy| Partnerships | Capacity

Data Markets
Data Philanthropy
Data Commons…

Real-Time Digital Data Sources that Reveal Behavior Change

What people say:

Online news

Social media

Retail advertising

Radio & Television

What people do:
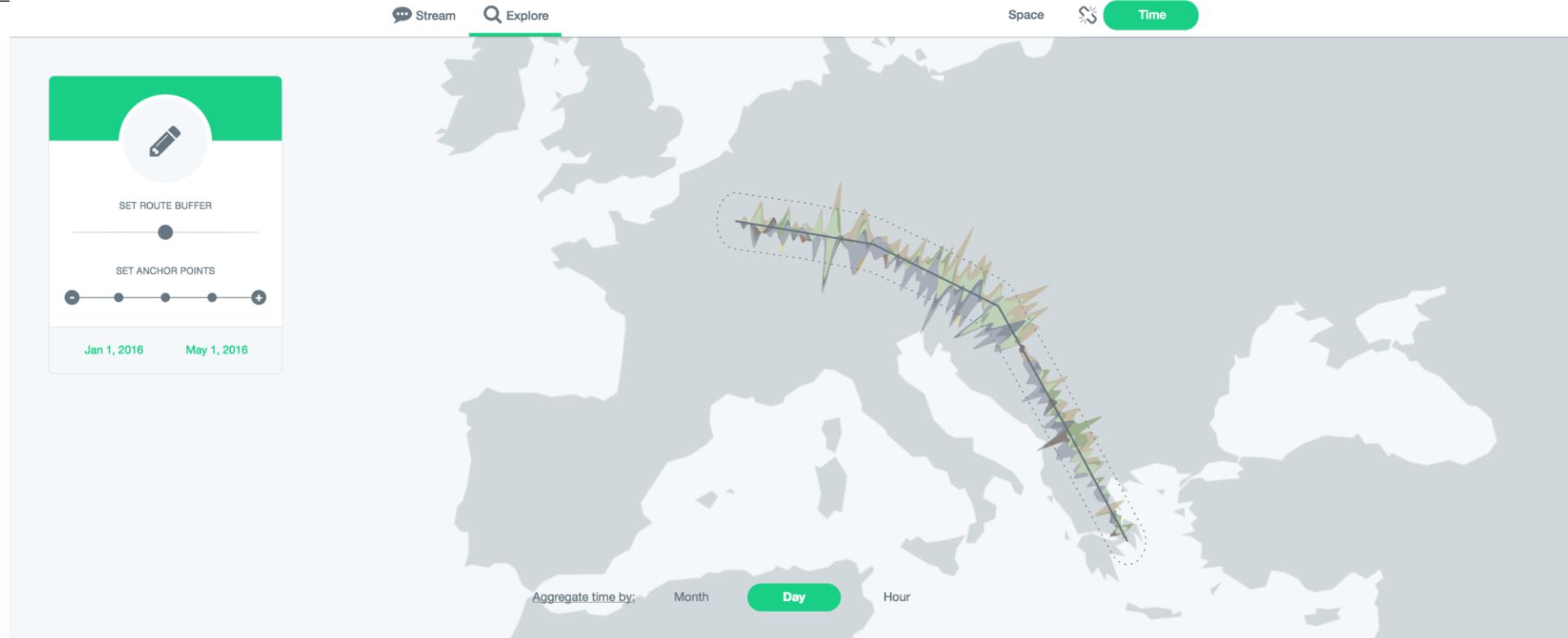
Online search

Web traffic

Mobile communications

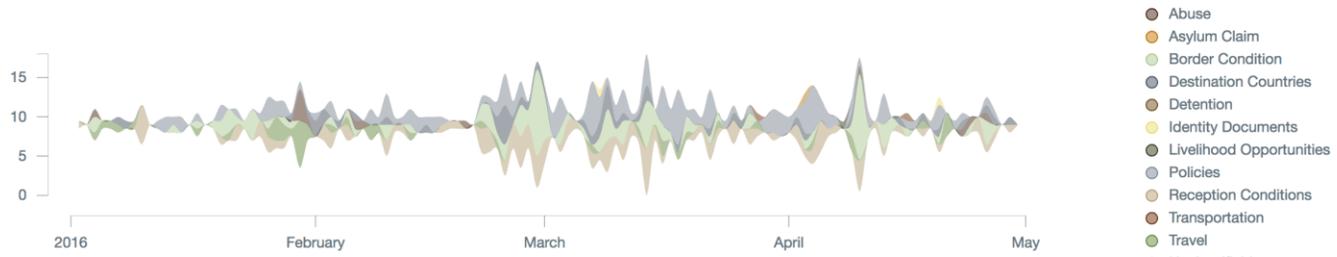Financial transactions

Postal traffic

Utility consumption

Emissions

App prototype developed for UNHCR for identifying, classifying, and quantifying xenophobic tweets about refugees in multiple languages along specific routes

REAL-TIME TWITTER FOOD INDEX
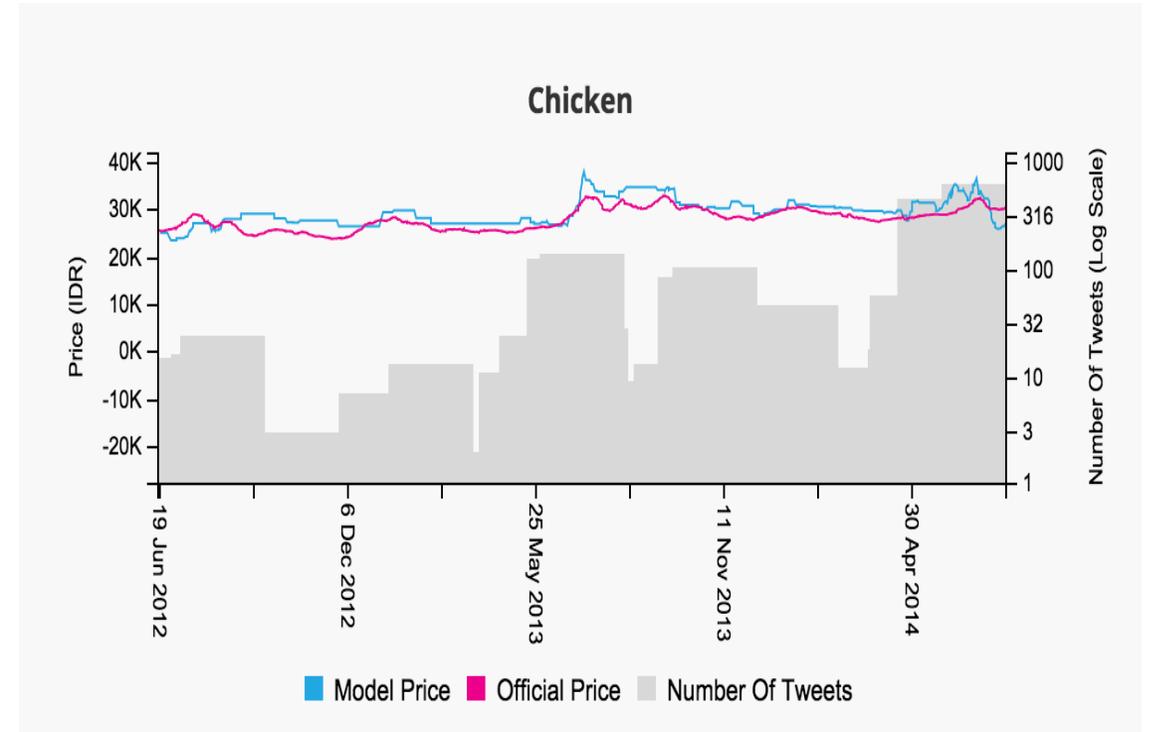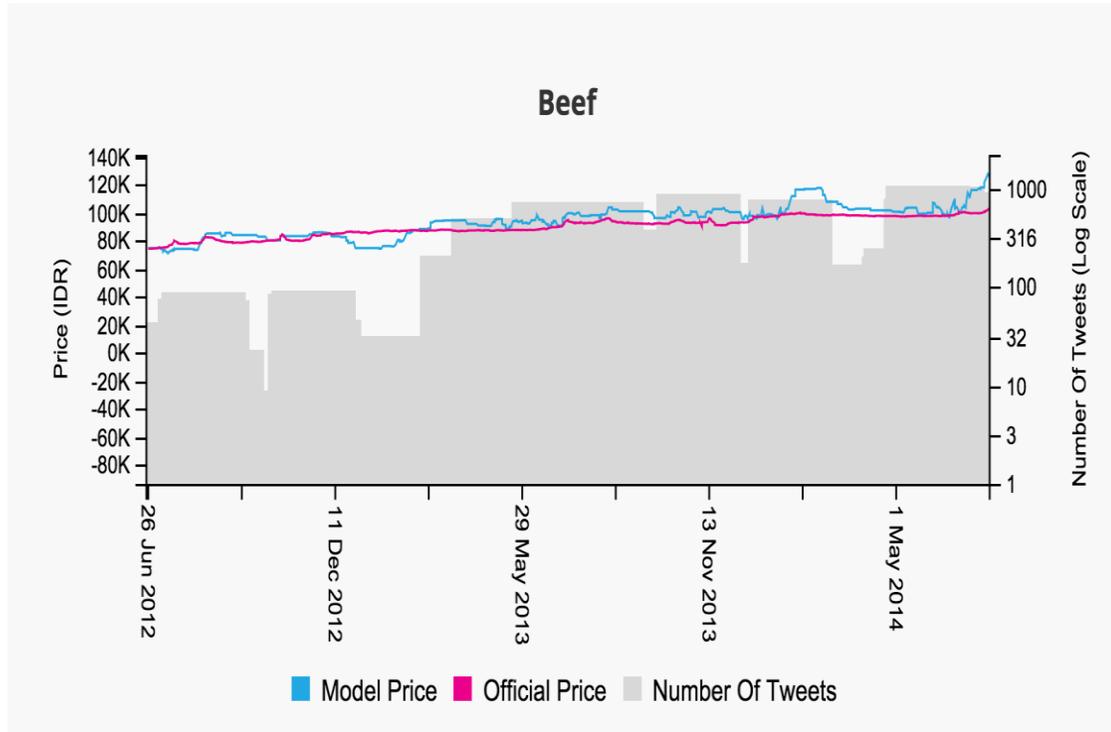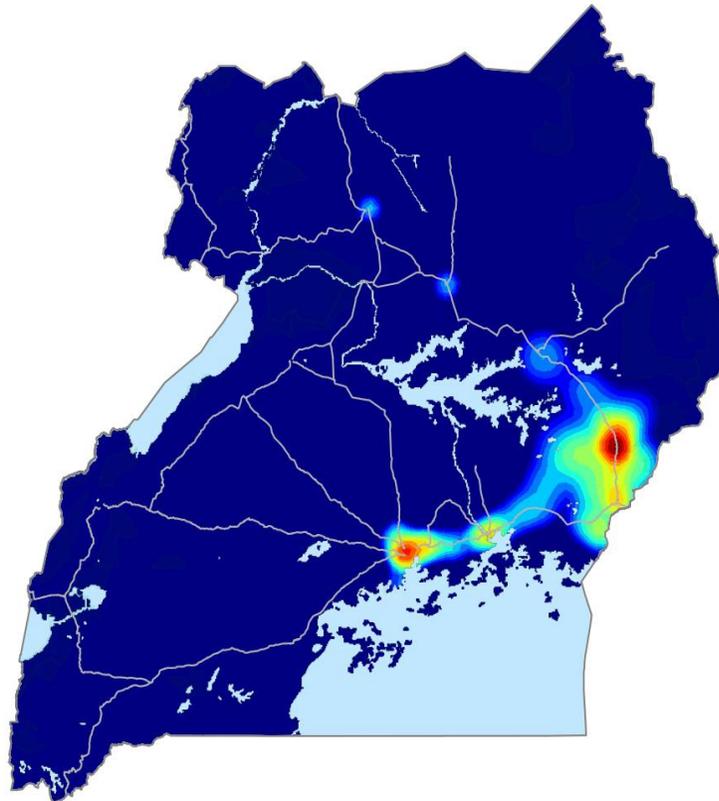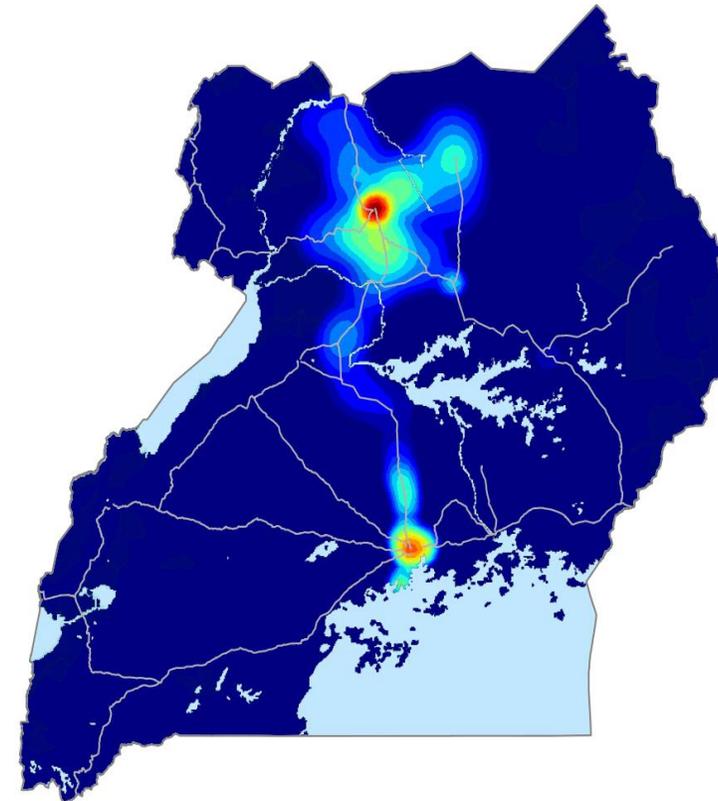
Population movements from areas with outbreaks predict the spread of disease in Uganda
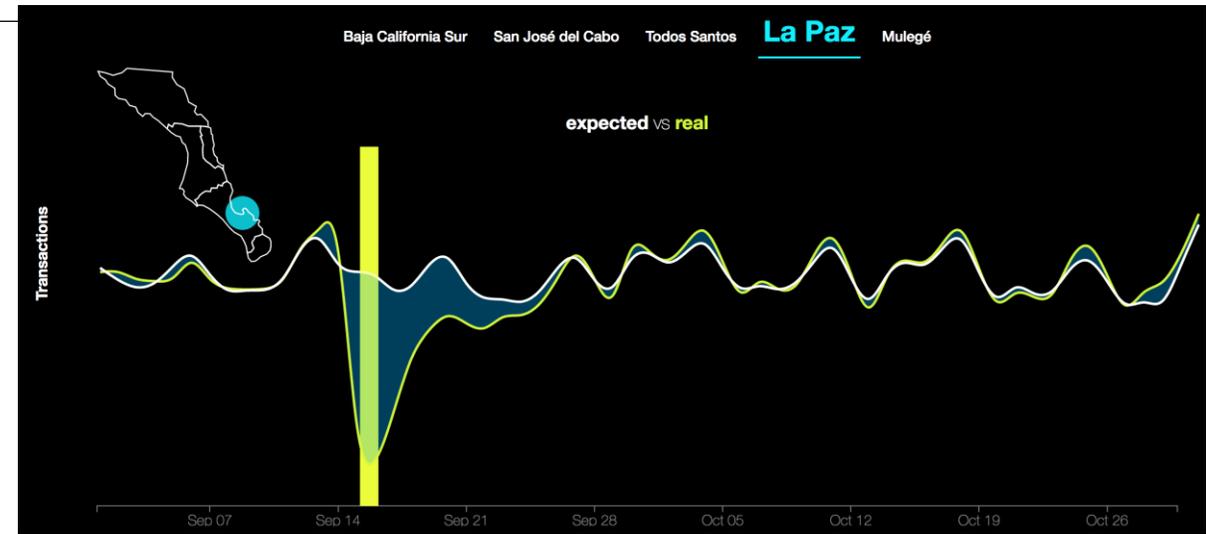

Movement out from Mbale — Tue Jan 01

Movement out from Gulu — Tue Jan 01

We can now quantify economic impact and track recovery in real-time
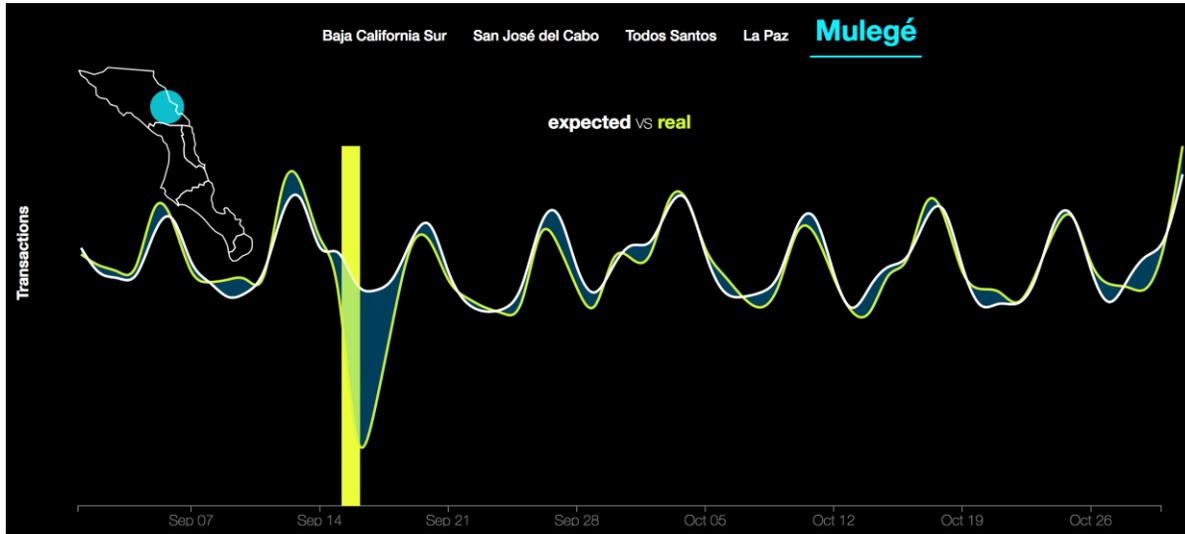
Understanding Perceptions to a Refugee Influx Through Analysis of Radio Content

Uganda currently ranks among the top three refugee hosting nations in the world. A mass exodus of refugees from South Sudan in 2016 doubled the population to over one million, making Uganda the leading refugee hosting country in Africa. Uganda has one of the most favorable refugee protection environments in the world, providing refugees with freedom of movement, the right to work, and access to social services through a generous asylum policy.

Pulse Lab Kampala was tasked by the UN in Uganda with unearthing the public perceptions with regard to the interactions between refugee and hosting communities. The Lab and partners use the Radio Content Analysis Tool to take the pulse and analyse public discussions aired by local radio stations to support the Government to sustain its refugee open-door policy.

2 ZERO HUNGER

3 GOOD HEALTH AND WELL-BEING

11 SUSTAINABLE CITIES AND COMMUNITIES

15 LIFE ON LAND

There are several existing initiatives aiming to increase access to data for the sustainable development goals

Data sharing

UN Big Data Platform stores data and shares expertise

CGIAR platform for Big Data convenes partners to share data

Following strict data protection protocols, enables **data use for** social good on premises and under supervision of MA data scientists

Hosts **datathon** where people can share their data to a specific problem

Insight sharing

Predicts risk of disasters and **sharing warnings** with citizens in advance

NASA open data portal publicly shares NASA datasets, APIS, visualizations

Using satellite and telco data to estimate hazard prone areas and improve pre-disaster planning

Hosts a **collaborative data sharing platform**, Magic Box, to inform disaster response decisions

Creates **disaster maps** using satellite imagery and population location for crisis response

Sends **algorithms to data providers** to allow them to **share insights** without sharing data

Analyses IOT data from mobile phones to help end extreme poverty

Protocols, governance, tools…

Convenes **network for more open use of data** (Data4SDGs)

European Commission Business-to-Government data task force identifies best practice data sharing processes

Generates simple **data sharing licenses**

GovLab 100 questions initiative recommends **clarifying the problem to be solved** using AI

Offers **different licensing models** adapted to the variety of data users (e.g., NGO…)

Provides **tools, expertise and other capabilities** needed by non-profit, civic and government organizations to help solve the world's most pressing challenges

Despite various data collaborative initiatives, there are challenges in sharing data at scale

Although numerous data sharing initiatives exist globally, there are **significant gaps in the ecosystem**

We have a long way to go **to ensure that data sharing and data consumption is at the aspired scale** – where all stakeholders (incl. governments, NGOs, entrepreneurs, citizens) stand to gain

- Lack of strong evidence base to support investment
- Asymmetrical supply and demand – the marketplace is not yet established
- Lack of shared vision and cohesive implementation mechanisms
- Lack of capacity to adapt and integrate to insights
- No common approach to consumer privacy, data protection, and risk mitigation associated with data use (ethics and human rights)
- Technological advancements may have outpaced regulatory frameworks

UN Secretary-General High-Level Panel on Digital Cooperation Recommendation 1b): Digital Public Goods

We recommend that a broad, multi-stakeholder alliance, involving the UN, create a platform for sharing digital public goods, engaging talent and pooling data sets, in a manner that respects privacy, in areas related to attaining the SDGs.

digital COOPERATION
UN SECRETARY-GENERAL'S HIGH-LEVEL PANEL

"As a global community, we face questions about security, equity, and human rights in a digital age. We need greater cooperation to tackle these challenges and mitigate risks."

—UN Secretary-General António Guterres

#digitalcooperation

The Global Data Access Framework (GDA-F) aims to enable data sharing and to deploy AI to help achieve the Sustainable Development Goals (SDGs)

Key objectives

- Enable data sharing across public and private sector organizations

- Capitalizing on the immense volume of data available and using AI to tackle the world's greatest challenges

- Detect, present, and help scale-up use cases for AI to achieve the 17 SDGs

- The use of AI for Sustainable Development Goals will allow us to:
  - **Monitor** progress towards the achievement of the SDGs
  - **Simulate** implications
  - **Predict** outcomes of measures taken
  - **Make recommendations** for policymakers

Our vision and objectives

VISION

This initiative aspires to set in motion a global movement to significantly scale-up responsible access to data, empowering an unprecedented number of public, private and social sector actors to use data for public good

Objectives of this initiative:

- Scale-up existing and future data-collaborative initiatives and use cases

- Enable and deliver at-scale data sharing through automation and AI

- Ensure privacy, security and access management

- Assure responsible access and usage of data for public good

- Build assets and enablers that can serve all SDGs

Global Data Commons: Harnessing AI for the Public Good

- A **multi-stakeholder, action-driven, collaborative** exercise
- Lead by **The Future Society, UN Global Pulse, AI Commons and McKinsey's Noble Intelligence initiative**
- Started at the **World Government Summit in Dubai**, brought to **AI for Good in Geneva** and the **UNGA**
- **Over 100 institutions** participated so far

3 workshops with over 100+ organizations have helped develop our thinking

1st Workshop:

World Government Summit, UAE

Setting the context

2nd Workshop:

AI for Good, Geneva

Defining the scope of the problem

3rd Workshop:

Around UNGA, New York

Defining reference cases and path ahead

Next steps

- **Conceptualized the GDC** to help support SDGs
- **Received 40+ position papers** from partners on need for GDC and roadmap to develop it
- Developed a structured **use case driven approach** and a preliminary 3-year plan for achieving GDC
- Collected participant feedback on potential **challenges to implementation**

- Defined **preliminary technology requirements** for high level architecture and data requirements
- Defined **preliminary approach to governance** including potential business incentives and legal considerations for data sharing
- Identified **existing data lakes and initiatives** that GDC could potential tap into

- Agreed on **key principles** for GDC establishment
- Identified **priority SDGs** or data sets to prioritize for **initial 'reference case'** implementation
- Prioritized **technical capabilities** required for the reference case
- Defined **working model** to progress the GDC after the workshop

- Develop **value proposition** and outline **business / operating model**
- Establish **reference cases and partnerships** required
- Detail **technical reference architecture** plan
- Identify **financing needs** and define funding mechanisms

The position papers discuss several challenges that prevent the implementation of a GDA-F

1 Access to quality data — 30

2 Technical challenges — 26

3 Legal challenges — 23

4 Political/ regulatory challenges — 22

5 Social challenges — 21

6 Business/ commercial challenges — 11
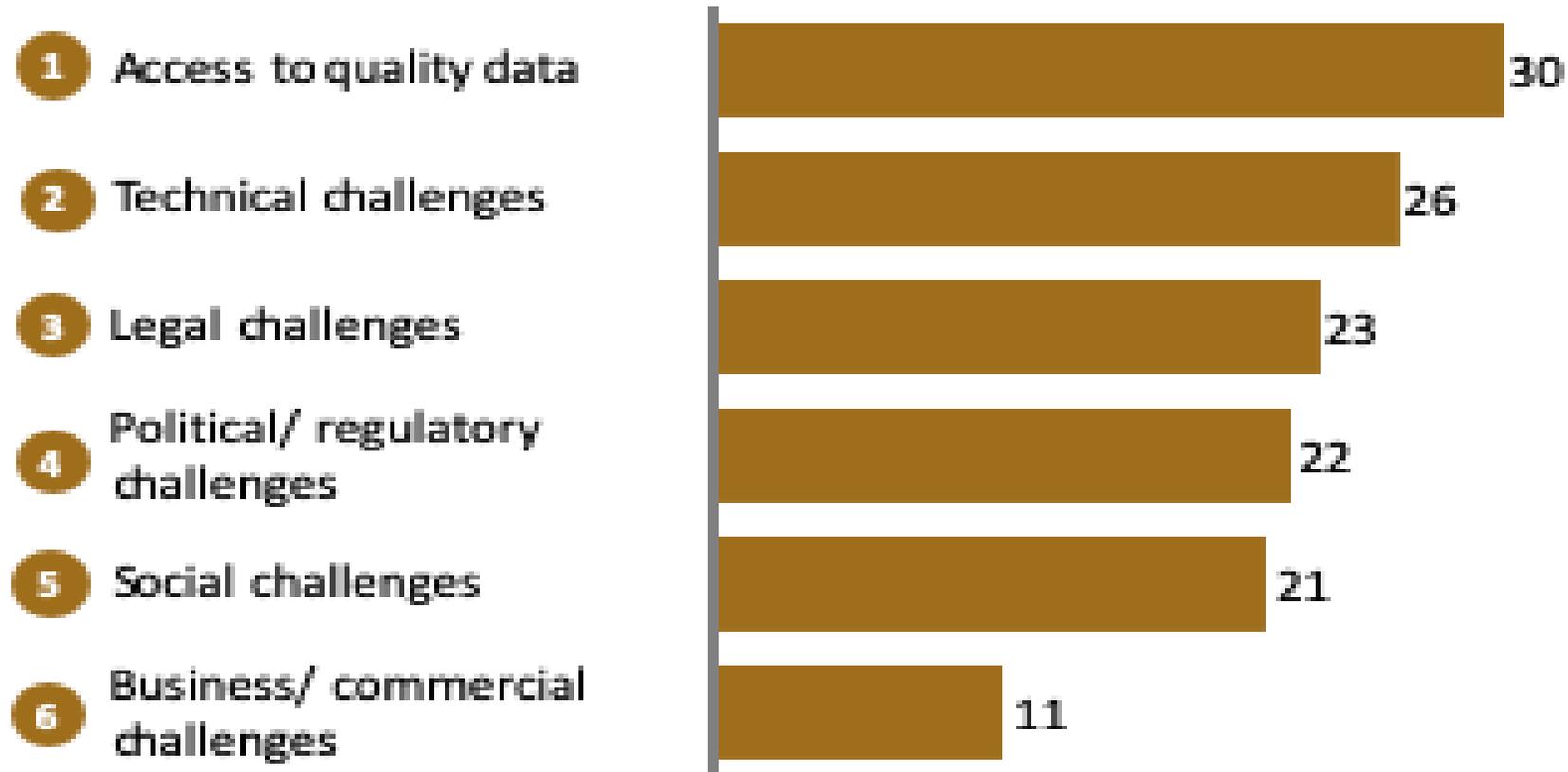
We are now working on 4 major areas

A
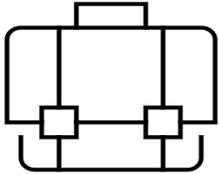
Collaboration and project mobilization

Onboard partners and manage collaboration

Oversee delivery, milestones and timelines

B
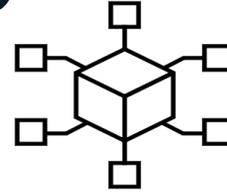
Develop reference cases

Define reference case requirements

Onboard partners (pilot entities, users, etc)

Oversee/coordinate implementation

C

Technology and governance protocol

Prioritize business and tech capabilities

Design solution space

Implement solution

D

Funding and resourcing

Identify sources of funding and resourcing (in kind contributions)

Acquire and oversee contributions

Smart Dubai, ajid Al Futtaim Retail, and Dubai Health authorities are collaborating to share data to track and promote healthy eating habits

Preliminary　　**Illustrative**

Reference case title: Monitoring & promoting healthy eating habits in Dubai　**SDG focus: Good health and Wellbeing (3)**

Demand-side scaling actors...

Accelerators

World Health Organization　　International Diabetes Federation

- GPSDD　　GSMA　　DIAL SDSN

End-users

Local gov't[1]	Hospitals	Insurers	Fitness provider

AL ZAHRA HOSPITAL DUBAI　　Daman
DEPARTMENT OF HEALTH　　AMERICAN HOSPITAL DUBAI　　SAADA　　DUBAI 30X30
IMPERIAL COLLEGE LONDON DIABETES CENTRE A Mubadala Company

Beneficiaries

- Citizens (at-risk of diabetes and other eating related diseases)
- Path and time to impact - tbd
- Info via currently used Carrefour app, phone network (Etisalat/Du), BTL activities at end users stated above, social media channels

Problem statement:
How can retailers' POS data and health sector's patient data be leveraged to predict incidence of diabetes at a particular area?

🇦🇪 **Dubai, UAE**

Pilot entities

Demand　　　　**Supply**

DUBAI HEALTH AUTHORITY　　SMART DUBAI

MAJID AL FUTTAIM

Carrefour

Data Collaboration Accelerators

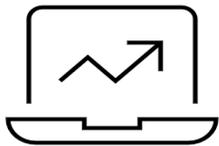| Protocols |
|---|
| Assets |
| Capabilities |

Supply-side scaling actors...

Data providers

Carrefour

- UAE Ministry of Health & Prevention

Analytics providers

Carrefour　　SMART DUBAI

- UPenn Data Analytics lab
- Universities

Platform/tech host

SMART DUBAI　　Microsoft

- IaaS, SaaS or PaaS?

Regulators/ compliance bodies

GOVERNMENT OF UNITED ARAB EMIRATES　　GOVERNMENT OF DUBAI

- + tbd

1. + local govs of 25 countries Carrefour will provide data for

3 GOOD HEALTH AND WELL-BEING

Pulse Lab Kampala and Uganda Ministry of Health are collaborating to share data to track outbreaks of infectious diseases

Preliminary   **Illustrative**

Reference case title: Predicting outbreaks of infectious diseases   **SDG focus: Good health and Wellbeing (3)**

Potential Demand-side scaling actors

Accelerators

World Health Organization   THE WORLD BANK   UNDP

2030 VISION   unicef   GSMA

DIAL   ECA, ECOWAS

GLOBAL PARTNERSHIP FOR SUSTAINABLE DEVELOPMENT DATA

End-users
- Ministry of Health
- Local policymakers
- Hospitals/health centers

Beneficiaries
- Citizens (at-risk of infection)
- Path and time to impact – 2 years
- Info only accessible to policymakers and health centers now; could be transmitted through apps/hospital communications

Problem statement:
How can real-time mobility data and predictive analytics help resource-constrained national health systems shift from merely responding to outbreaks to preventing them?

Uganda

Pilot entities

Supply

PULSE LAB KAMPALA

Demand
- Ministry of Health
- Local policymakers
- Hospitals/ health centers

Data Collaboration Accelerators

Protocols

Assets

Capabilities

Potential Supply-side scaling actors

Data providers

PULSE LAB KAMPALA
- MTN
- Airtel

Analytics providers

esri   Dalberg
- Analytics provider? AA required?

Platform/tech host

Microsoft
- DHIS2 open software database used by MOH
- Same software as many other countries; good for scaling

Regulators/ compliance bodies
- Local govt
- NITA
- Data Protection Authority (if established)

1.   + local govs of 25 countries Carrefour will provide data for

UN Global Pulse and BBVA are collaborating to share data to track communities' economic resilience to disasters

Preliminary   **Illustrative**

Reference case title: Tracking economic resilience to disasters   **SDG focus: Decent work and economic growth (8)**

Potential demand-side scaling actors

Accelerators

THE WORLD BANK   FUNDACIÓN Carlos Slim   UNDP Empowered lives. Resilient nations.

- ECLAC
- GPSDD

GSMA   SDSN

End-users
- Government
- Local policymakers
- Businesses
- Banks

Beneficiaries
- Citizens (recovering from disaster)
- Path and time to impact - tbd
- Info access could be provided through ATMs, stores, banks—any points of financial activity

---

Problem statement:
How can we use financial data to track consumption patterns to measure, predict, and strengthen communities' economic resilience to disasters?

🇲🇽 **Mexico**

Pilot entities

Supply

GLOBAL PULSE   BBVA

Demand
- Local policymakers
- Businesses
- Banks

THE WORLD BANK

Data Collaboration Accelerators

Protocols

Assets

Capabilities

Potential supply-side scaling actors

Data providers

BBVA   Telefónica
mastercard   Twitter   +Insurance Companies

Analytics providers
- tbd

Platform/tech host

Microsoft

Regulators/ compliance bodies
- Local govt
- Data Protection Authorities

---

1.   + local govs of 25 countries Carrefour will provide data for

We have started defining a capability model and will continue detailing the technical solution

A Data governance | Capabilities to manage data | **A1** Data ownership/stewardship | **A2** Data cataloguing | **A3** Data dictionary

B Access management | Granular control of access to datasets and data models for complex group of GDC stakeholders

C Data security | Security to audit, monitor and protect data and manage privacy as per policies

D Data publishing | **E** Data storage, processing and engineering | **F** Data consumption

D1 **Data publishing and policy management**
Publishing of data repositories (structured, semi-structured, unstructured), and usage policy setting

D2 **Data ingestion**
Exchange methods to enable movement of data between different environments, data sources and repositories

E1 **Data repository**
Layer for raw, Structured, semi-structures and unstructured data, with potential for distributed/federated data management

E2 **Data processing quality**
Components that enable the integration and transformation of data, executing complex queries, manipulating data at scale

E3 **Data modeling**
Custom engineering of data modules catering to specific data types and use-case families to make them consumable at scale

E4 **Master data management (MDM)**
Capabilities enabling creation of common references for critical master data scattered across systems

F1 **Workspace/Sandbox**
Ability to create workspaces allowing for data access, upload and joining, modeling and analytics

F2 **Visualization**
Self-service BI tools for dashboards/visualization

F3 **AI Library**
Open library of AI algorithms and code to support use cases

E Protocols/Processes | Capabilities to manage | **G1** Digital contracts | **G3** Workflows | **G5** Licensing | **G7** Partnerships
| | | **G2** Monetization / commerce support | **G4** Digitized processes for publishing and access | **G6** Privacy management for providers |

F Services and capability building | Building capabilities through trainings and services management

THANK YOU!

romanoff@unglobalpulse.org

www.unglobalpulse.org/privacy-and-data-protection