OnTrackPH: Driving Data on Philippine Roads and the Path Forward

Stephanie Sy, CEO
Thinking Machines
stef@thinkingmachin.es

About Thinking Machines

Manila | Singapore | San Francisco

We use data to solve business problems

Founded in 2015 by Stanford Engineering graduates and tech veterans

KEY PARTNERS:

Google Cloud
Machine Learning Partner

unicef INNOVATION
AI Innovation Fund[1]

waze OUTSMARTING TRAFFIC, TOGETHER
Research Partner

1 Thinking Machines was selected by UNICEF's Innovation Fund as part of their AI cohort

Our Clients

	PRIVATE SECTOR	PUBLIC SECTOR
PHILIPPINES	Globe Ayala MERALCO LBCX	Department of Science and Technology MMDA Metro Manila Development Authority
INTERNATIONAL	MITSUI&CO. corepower YOGA	THE WORLD BANK IBRD • IDA | WORLD BANK GROUP ADB ASIAN DEVELOPMENT BANK Save the Children Teach For All

Agenda

1. **Transformational Power of Data**
   Case studies: OnTrackPH and LinkSight
2. **Insights from Working with PH Data**
   The data landscape and policy implications
3. **Q&A**
   Let’s keep the conversation going!

Agenda

1 **Transformational Power of Data**
Case studies: OnTrackPH and LinkSight

2 **Insights from Working with PH Data**
The data landscape and policy implications

3 **Q&A**
Let’s keep the conversation going!

Two case studies from the work we've done

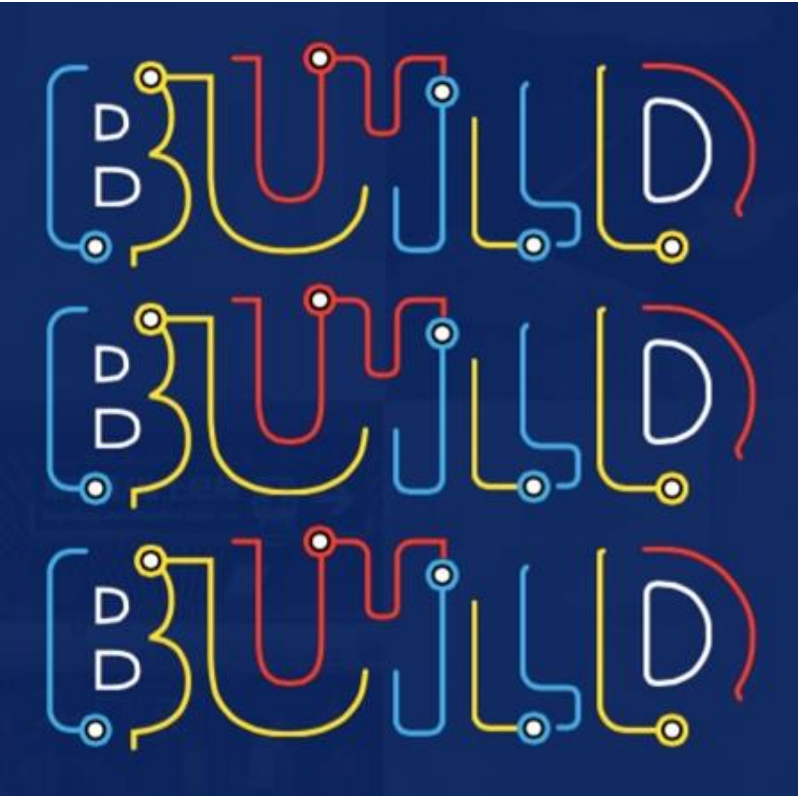
nTrackPH

A **software tool for matching records across data silos** in a fast, cost-effective, and scalable way

LinkSight

An **open-source repository of barangay-level Philippine data** that will enable users to add valuable context to their own data

Context: Why track transport infra projects?



The Build, Build, Build program plans to spend over PHP 8 trillion on 75 projects

~98% of this budget is for roads, bridges, and other transport infrastructure

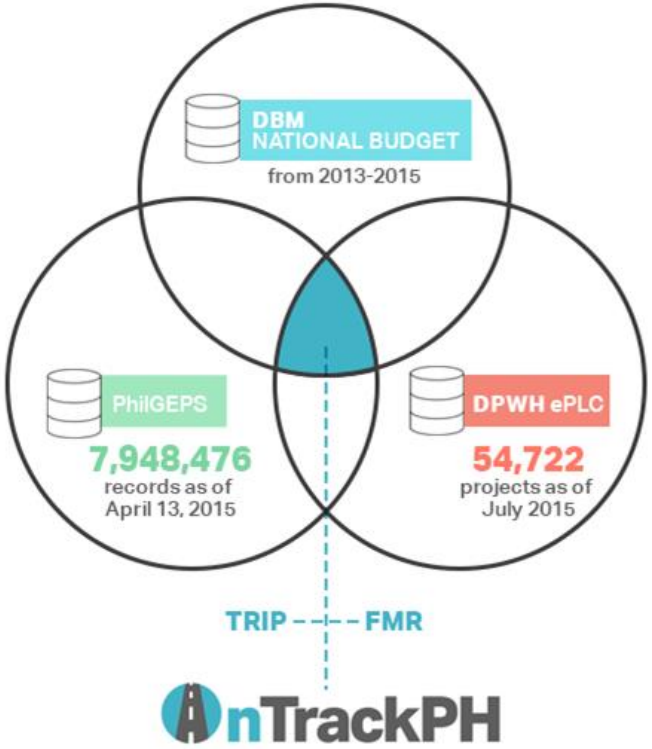
SOURCE: build.gov.ph, Rappler

The Challenge

Difficult to monitor and evaluate spending on infrastructure because:

- Data on projects is siloed across different databases
- Identical projects had no common identifiers

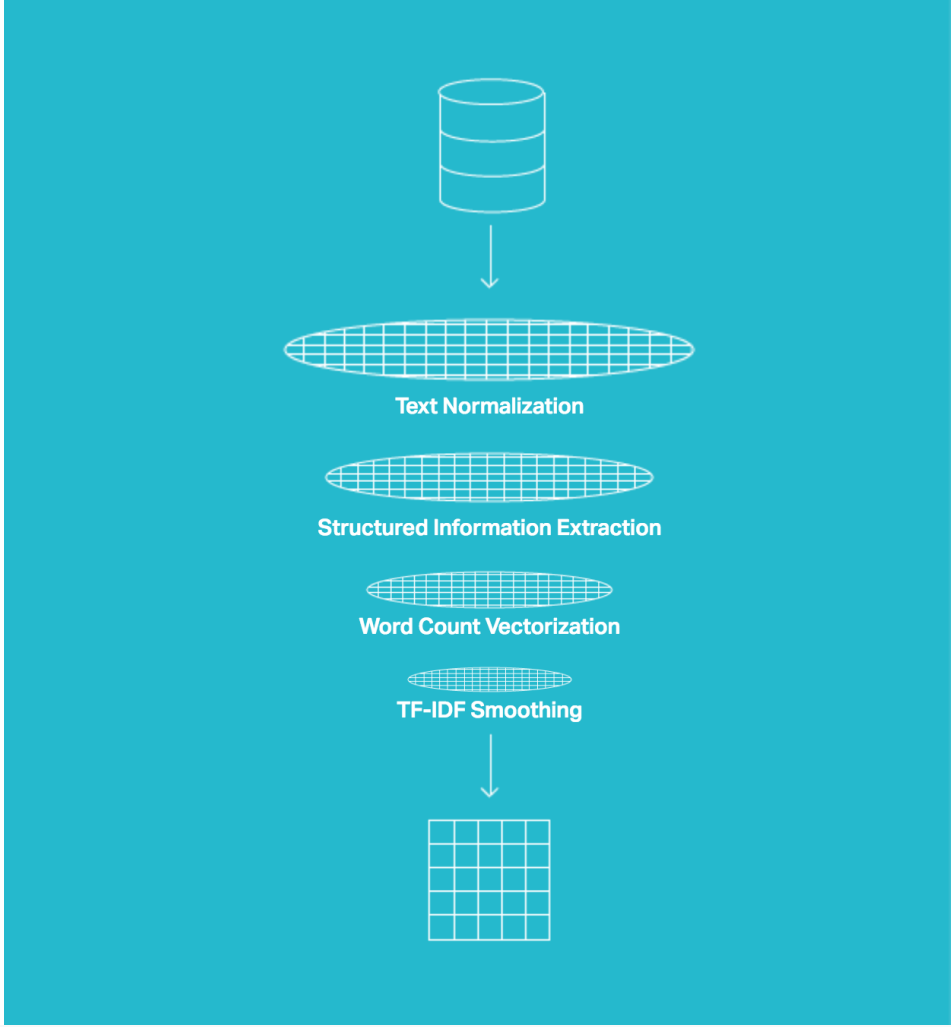
This led to **low transparency** in monitoring the end-to-end status of projects.



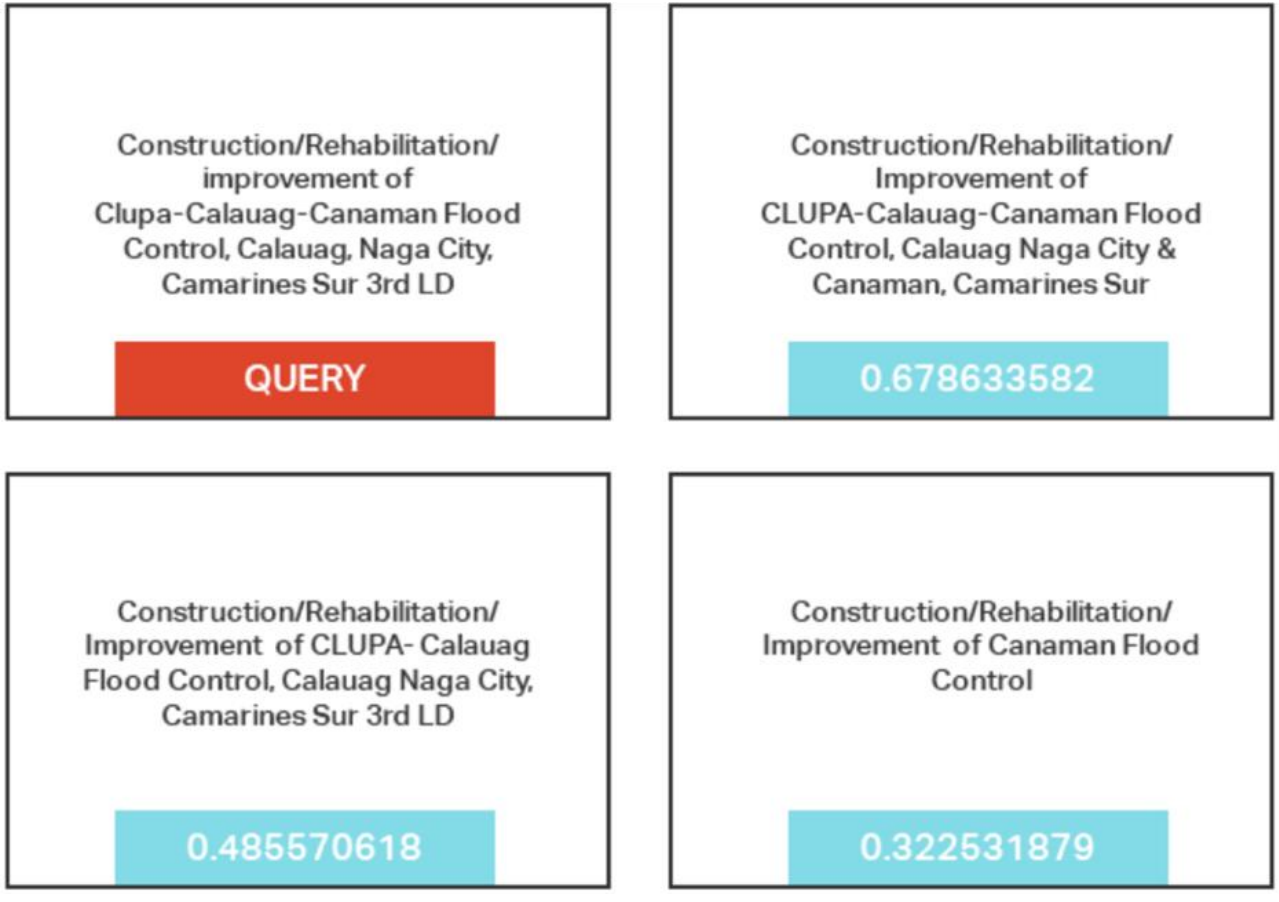
The Solution: Part 1

OnTrackPH, a multi-step “sieve approach” algorithm

- Starts by tagging the most precise matches
- Scores the remaining ambiguous matches with bag-of-words vectorization and cosine scoring
- Combines natural language processing to boost match accuracy



Sample Match

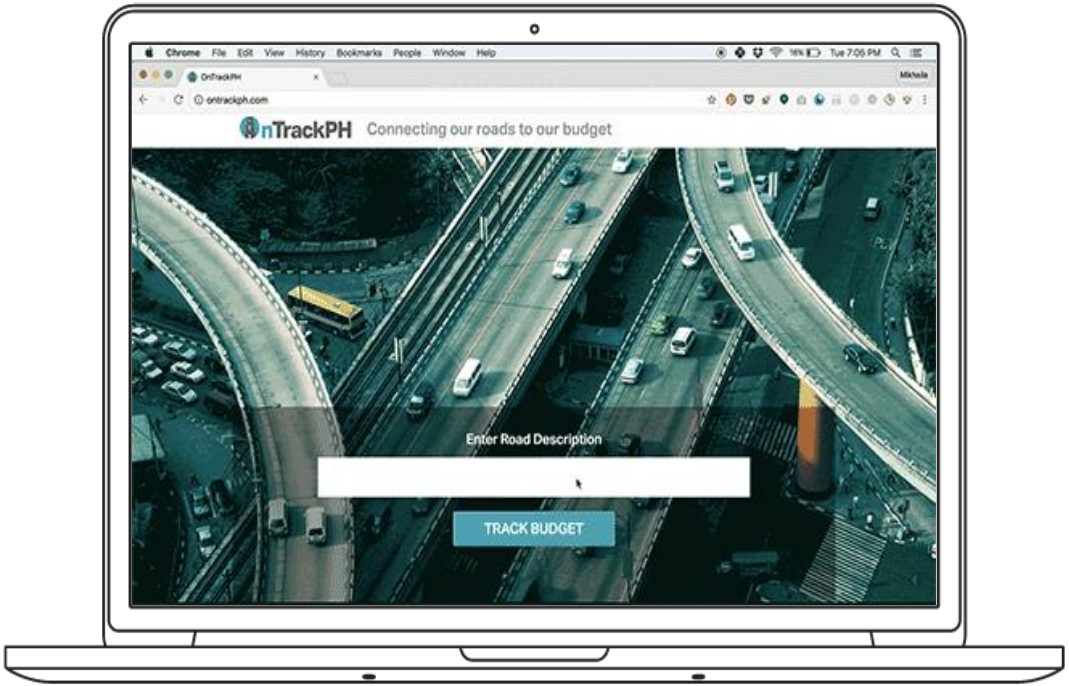


Thinking Machines
Data Science

The Result

TRIP FMR to ePLC & PhilGEPS	Manual Matching	OnTrackPH
Time	**3** months	**15** minutes
Manpower	**5** people	**1** person (or none)

The Result: OnTrackPH Web Visualization



However, **funding** is not the only determinant of **project success**

Project sites should also account for the following:

- **Historical sites** that cannot be demolished
- Number of **displaced communities**
- Other **geospatial data**

NGOs are critical partners for this purpose

Next Challenge: Bringing Monitoring to the NGO Sector

Many NGOs and MSMEs have difficulty accessing and utilizing geospatial third-party data:

- Data is often **hard to find, hard to access, or incomplete**
- Data is **hard to use** and may require major data processing capabilities

GRUPO KALINANGAN

Case example: **Grupo Kalinangan**

- A coalition of volunteer heritage advocates who used geospatial data to map over 30,000 heritage structures and historical sites
- Faced the challenge of accessing and wrangling 3rd party data to integrate flood/environmental damage, infrastructure development, etc. into their map

unicef

LinkSight: An Evolution of OnTrackPH

Integration of public data sources

The Solution: Part 2

LinkSight, an open-source location intelligence tool for Philippine data

1. **Users upload datasets** with barangay, city/municipality, and province information
2. Users select **new datasets to be combined** with their original data (population, disaster risk, barangay competitiveness)
3. An **integrated dataset is generated**

GRUPO KALINANGAN

Case example: **Grupo Kalinangan**

- With LinkSight, Grupo Kalinangan can:
  - Identify priority heritage sites at risk of flooding or other environmental hazards
  - Identify target barangays who fit demographic criteria to become heritage partners

unicef

LinkSight: An Evolution of OnTrackPH

Integration of public data sources

The Result

Democratized geospatial data access for various organizations

- **Fast and simple data integration**, without the need for complex analytics
- **Crowdsourcing of geospatial data** from private sector partners to expand access

Social sector use cases include:

- Leveraging crowdsourced calamity data to **optimize relief operations**
- Identification of lower SEC **communities to target for livelihood programs** using mapped demographic and competitiveness data

The Big Message: Data + Usability = Insight

- **Bad raw data** makes it impossible to make good decisions
- **Good raw data** isn't enough. It needs usability design so people can pull insights from it.

Agenda

1 **Transformational Power of Data**
Case studies: OnTrackPH and LinkSight

2 **Insights from Working with PH Data**
The data landscape and policy implications

3 **Q&A**
Let’s keep the conversation going!

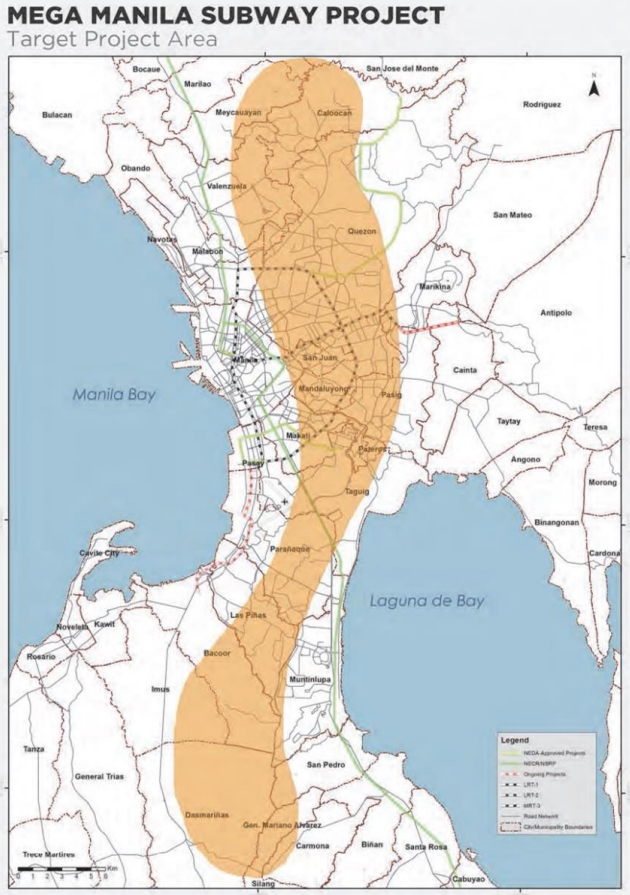
Four challenges in today's public data environment

- Data Availability and Accuracy
- Lack of Digitization and Processing
- Data Capabilities
- Data Ethics

Four challenges in today's public data environment

Data Availability and Accuracy

- "What gets measured gets managed"
  - **Global indicators/indices**
  - **Government priorities**
- Local/global research may be anchored on inaccurate data because:
  - Relevant metrics were captured in **dated, one-off studies**
  - Different government institutions report **different values for the same metric**



JICA's Metro Manila Dream Plan sparked conversations and launched projects to accelerate transport development

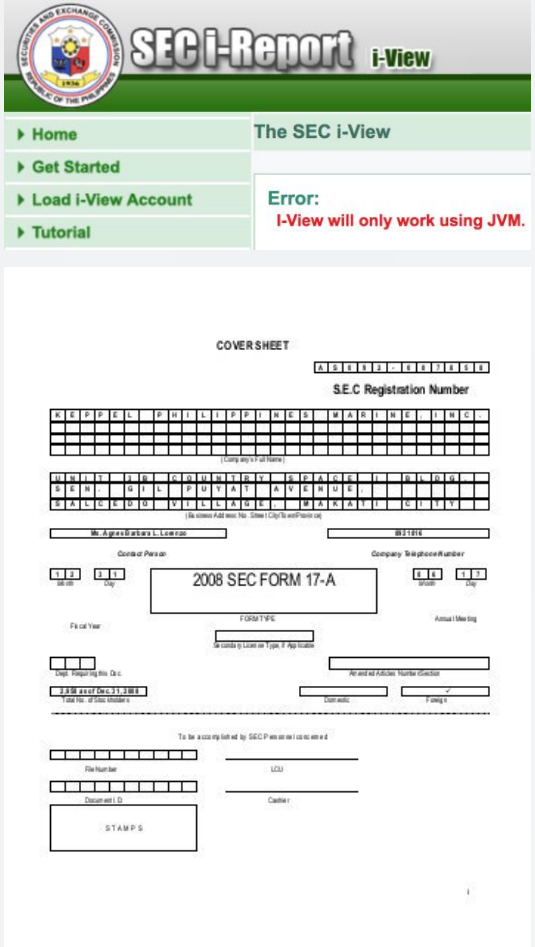
SOURCE: Japan International Cooperation Agency

Four challenges in today's public data environment

Lack of Digitization and Processing

While almost all government agencies collect information, a lot of data is not readily usable because:

- Data is left on **hard-copy documents** and not encoded
- Information is uploaded as **image or PDF files**, which can be challenging to run analytics on
- Raw data is uploaded, with **major data cleaning required** before processing



SEC's i-View allows the public to view corporate financial submissions, but presents these as image files on a challenging interface

SOURCE: SEC i-View

Four challenges in today's public data environment

Data Capabilities

- Many government agencies are limited by their analysts' capabilities to collect, store, or analyze data
- This results in several challenges, including:
  - Difficulty in knowledge transfer across different government institutions
  - Limited insight generation, resulting in ineffective or inefficient decision-making

The MMDA TEC has limited capabilities in using video data to count different types of traffic moving through an intersection, so the team does this manually using field counters.

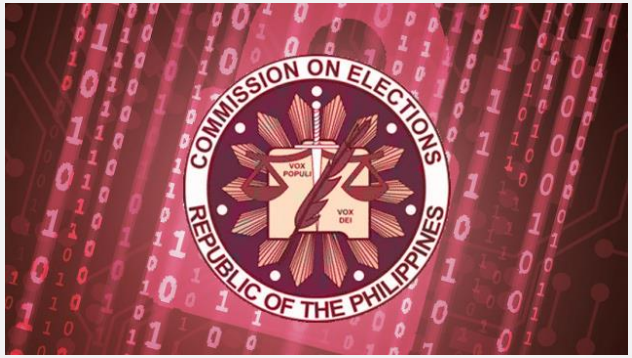
Thinking Machines
Data Science

Four challenges in today's public data environment

Data Ethics

- **Technological advancements have far outpaced data regulation** worldwide
- This raises key questions on the ethical way to gather and use data:
  - What data should we collect? What **shouldn't we collect**?
  - What **regulatory restrictions** should be imposed to enforce responsible data usage?

China's Police Department has used facial recognition AI technology to capture criminals at Jacky Cheung concerts



Hacking group LulzSec Pilipinas accessed and leaked the COMELEC's voting records database

SOURCE: Slate (image from STR/AFP/Getty Images); Rappler (image from Rappler)

The Big Message: investments in data will catalyze effective growth in the Philippines

3 priorities when investing in data-driven projects

- **Collection**: initiatives that capture data on:
  - **Priority issues** (traffic, poverty, health, etc.)
  - **Marginalized communities** and **critical advocacies** (indigenous peoples, environment)
- **Collaboration**: initiatives that facilitate data sharing among institutions (public, private, NGOs)
- **Capability**: training and infrastructure that enable accurate and properly formatted data collection and processing

Thinking Machines
Data Science

Agenda

1 **Transformational Power of Data**
Case studies: OnTrackPH and LinkSight

2 **Insights from Working with PH Data**
The data landscape and policy implications

3 **Q&A**
Let’s keep the conversation going!

Questions?

Read more data stories on our blog
stories.thinkingmachin.es

Follow Us
/thinkdatasci
@thinkdatasci