

The Open Cities AI Challenge: Building segmentation for resilience

Nick Jones

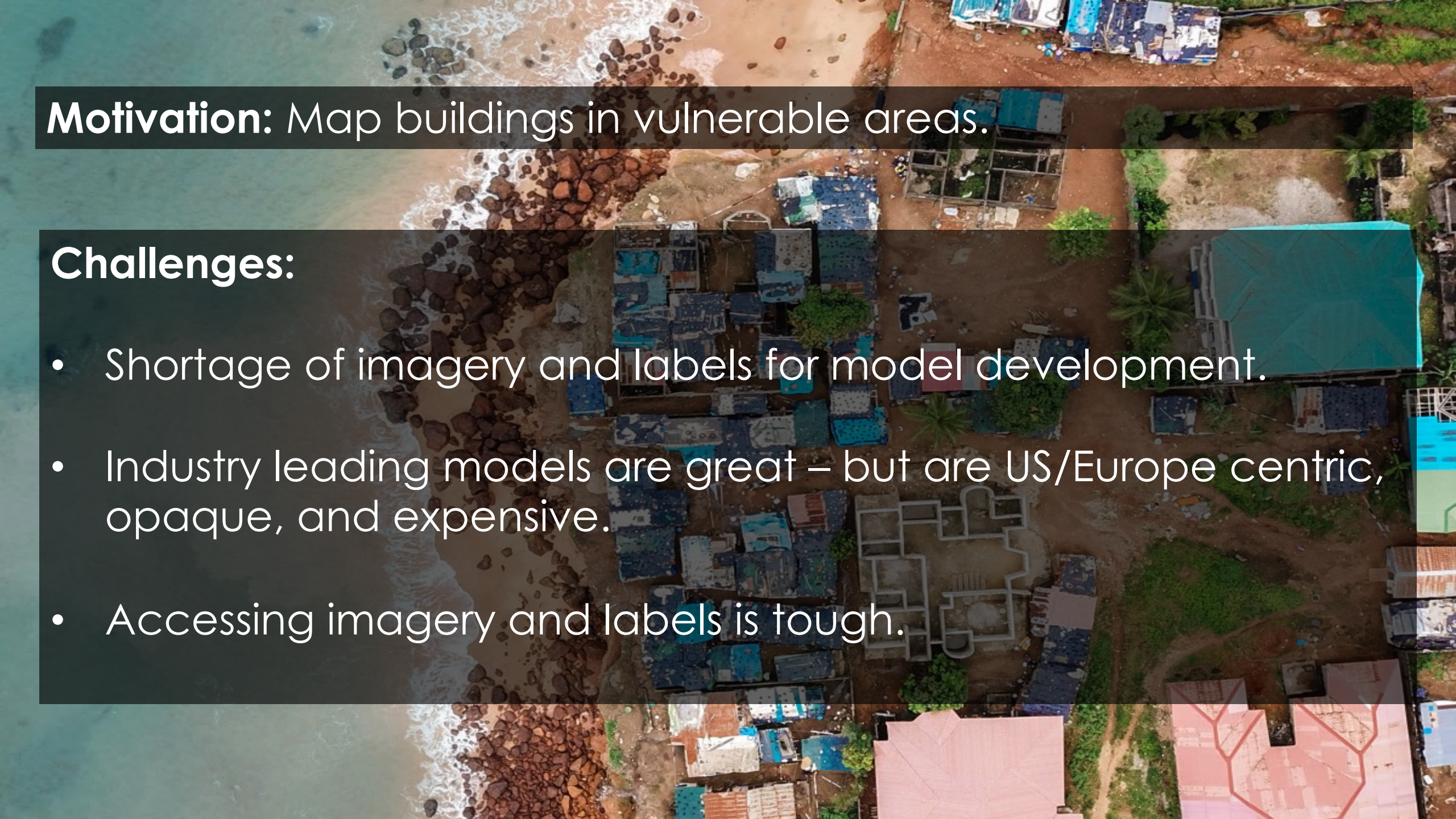
World Bank / GFDRR



<https://bit.ly/2O1W9ST>



Understanding and
quantifying
disaster risks requires
accessible, detailed,
up to date **data**

An aerial photograph of a coastal village. The image shows a rocky shoreline on the left with waves crashing against the rocks. The village consists of numerous small buildings, many with blue roofs, clustered together. Some buildings have white outlines overlaid on them, indicating they have been mapped or identified. The terrain is a mix of dirt paths and some greenery.

Motivation: Map buildings in vulnerable areas.

Challenges:

- Shortage of imagery and labels for model development.
- Industry leading models are great – but are US/Europe centric, opaque, and expensive.
- Accessing imagery and labels is tough.

OPEN CITIES AFRICA

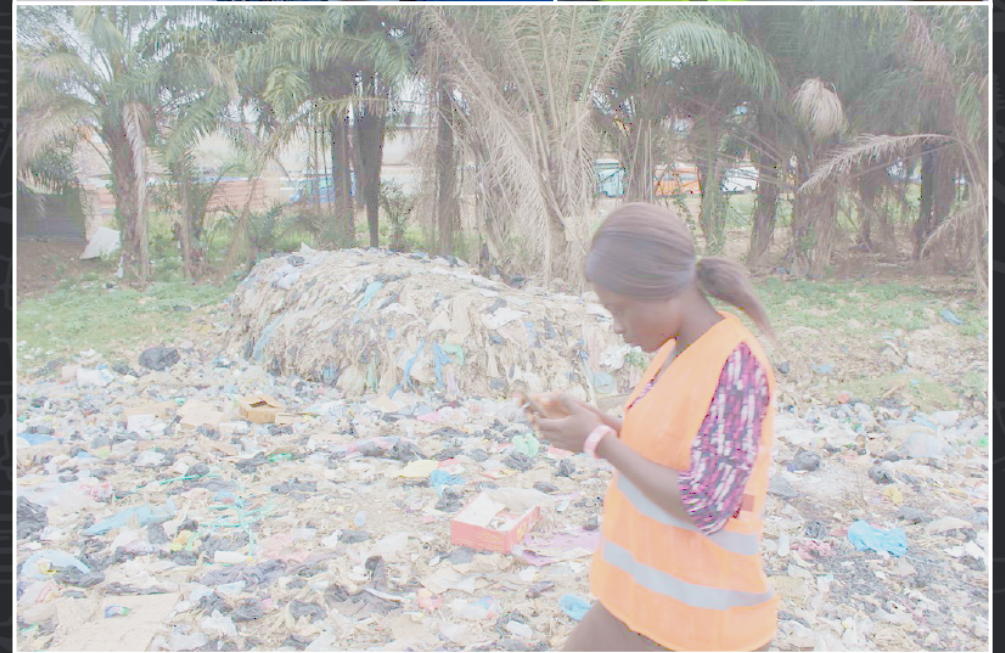
- Community mapping initiative: **12 cities** simultaneously
- **500,000** features mapped

DIGITAL SKILLS

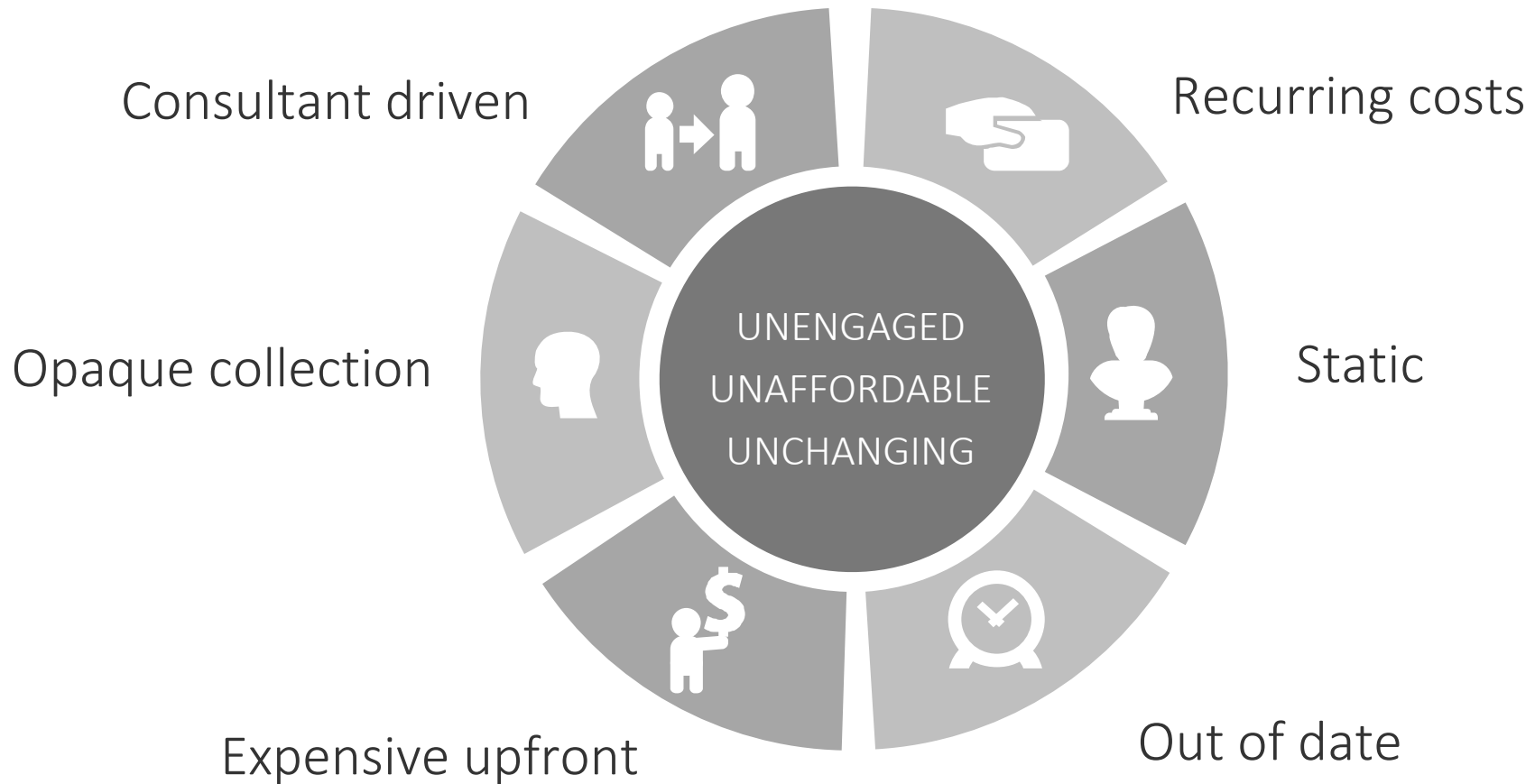
- **500 people** trained on digital cartography
- At **41% female participation rate**

REGIONAL NETWORK

- Online training platform with over **100 participants**

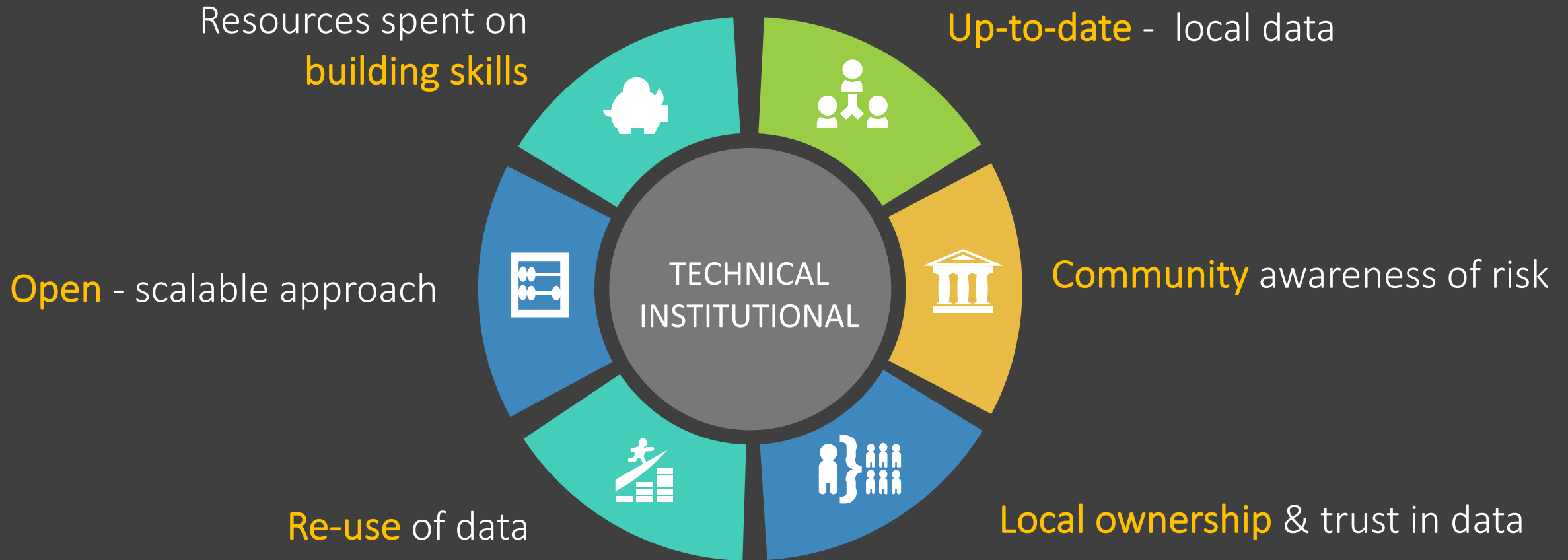


TRADITIONAL APPROACH TO DATA COLLECTION



DISRUPTIVE TECHNOLOGY AND HUMAN CAPITAL

COLLABORATIVE APPROACH TO GEOSPATIAL DATA COLLECTION



Geospatial Market growth from
US\$ **9.8 billion** in 2017 to
US\$ **17.5 billion** by 2023

Geospatial data and services
generate US\$ **400 billion**
in global revenue

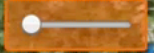


AI-Mapped Buildings in Tanzania

Data from Commission for Lands, Revolutionary Govt of Zanzibar & WeRobotics (source, CC BY 4.0)

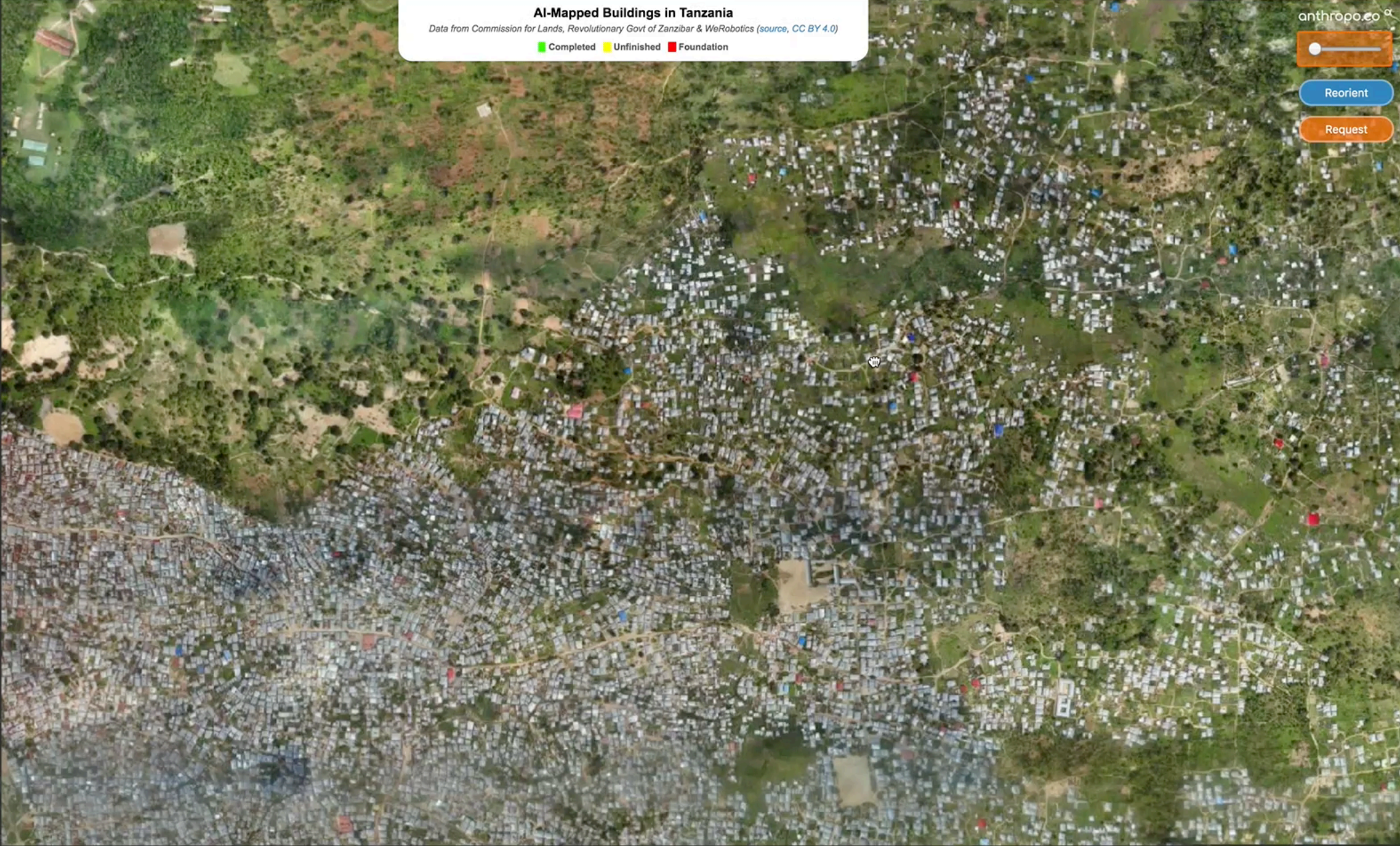
■ Completed ■ Unfinished ■ Foundation

anthropo.co



Reorient

Request



Open Cities Africa dataset

A novel labelled dataset of drone imagery and building footprints.

Collected via participatory mapping:

- Open Cities Africa
- Dar Ramani Huria
- Zanzibar Mapping Initiative.



Open Cities Africa dataset

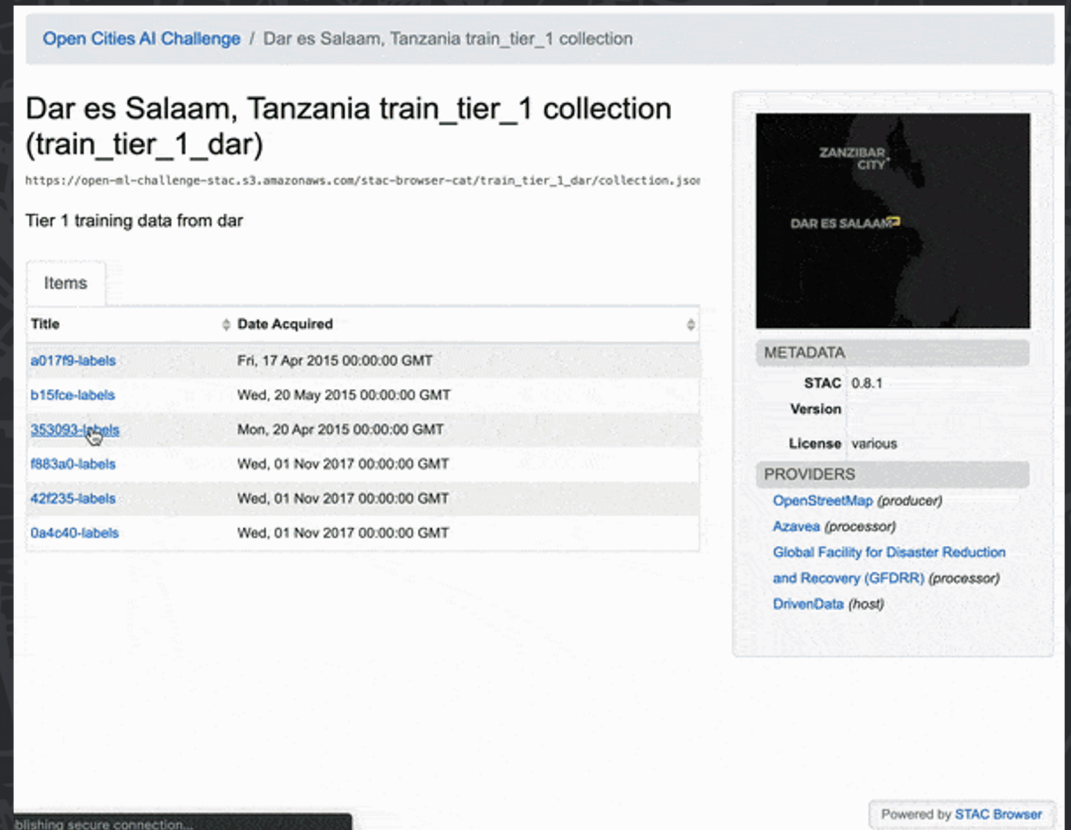
City/Region	Data class	Scene count	AOI area (sq km)	Building count	Average building size (sq m)	Building ratio (% area covered)
Accra	train_tier_1	4	7.86	33,585	84.84	0.36
Dar es Salaam	train_tier_1	6	42.90	121,171	99.20	0.28
Dar es Salaam	train_tier_2	31	223.28	571,047	94.16	0.24
Ngaoundéré	train_tier_2	2	12.54	15,792	81.05	0.10
Kampala	train_tier_1	1	1.14	4,056	53.14	0.19
Kinshasa	train_tier_2	2	1.01	2,357	71.29	0.17
Mahe Island	train_tier_2	4	19.40	7,313	206.48	0.08
Monrovia	train_tier_1	4	2.90	6,947	150.71	0.36
Niamey	train_tier_1	1	0.68	634	47.43	0.04
Niamey	train_tier_2	2	2.46	7,444	62.76	0.19
Pointe-Noire	train_tier_1	2	1.87	8,731	72.73	0.34
Zanzibar	train_tier_1	13	102.61	13,407	120.83	0.02
Total		72	418.65	792,484	95.43	0.18

Open-source standards for ease of access

- Images: Cloud-Optimized GeoTiff (COG)
- Dataset: Spatio-Temporal Asset Catalogue (STAC)

```
{
  "type": "Feature",
  "stac_version": "0.8.1",
  "id": "a42435-labels",
  "properties": {
    "label:description": "Geojson building labels for
scene a42435",
    "area": "acc",
    "label:type": "vector",
    "label:properties": [
      "building"
    ],
    "label:overviews": [
      {
        "property_key": [
          "building"
        ],
        "counts": [
          {
            "name": "yes",
            "count": 6647
          }
        ]
      }
    ]
  }
}
```

(1) Example metadata JSON for training labels.



The screenshot displays the STAC browser interface for the 'Dar es Salaam, Tanzania train_tier_1 collection'. The page title is 'Dar es Salaam, Tanzania train_tier_1 collection (train_tier_1_dar)'. Below the title, there is a URL: https://open-ml-challenge-stac.s3.amazonaws.com/stac-browser-cat/train_tier_1_dar/collection.json. The main content area is titled 'Tier 1 training data from dar' and contains a table with the following data:

Title	Date Acquired
a017f9-labels	Fri, 17 Apr 2015 00:00:00 GMT
b15fce-labels	Wed, 20 May 2015 00:00:00 GMT
353093-labels	Mon, 20 Apr 2015 00:00:00 GMT
f883a0-labels	Wed, 01 Nov 2017 00:00:00 GMT
42f235-labels	Wed, 01 Nov 2017 00:00:00 GMT
0a4c40-labels	Wed, 01 Nov 2017 00:00:00 GMT

On the right side of the interface, there is a map showing the location of Dar es Salaam, Tanzania. Below the map, there is a 'METADATA' section with the following information:

- STAC: 0.8.1
- Version: 0.8.1
- License: various

Below the metadata, there is a 'PROVIDERS' section with the following information:

- OpenStreetMap (producer)
- Azavea (processor)
- Global Facility for Disaster Reduction and Recovery (GFDRR) (processor)
- DrivenData (host)





At the bottom right of the page, there is a 'Powered by STAC Browser' logo.

(2) STAC browser to visualize dataset online.

Challenges & opportunities

(1) Images of diverse locations, resolutions, capture conditions:

TIER 1 SAMPLE

Area (abbreviation)	Scene ID	Thumbnail	Resolution	Pixel width x height
Accra (acc)	665946		3 cm	84466 x 150147
Kampala (kam)	4e7c7f		4 cm	39270 x 40024
Pointe-Noire (ptn)	f49f31		20 cm	6605 x 4185
Zanzibar (znz)	aee7fd		7 cm	40551 x 40592

(2) Labels of varying quality:

Tier 1:



Tier 2:



Two-track contest

(1) Semantic segmentation track:

- Pixel-wise binary target (building or no building)
- Evaluated on Jaccard index (IoU)
- \$12,000 for top three entries

(2) Responsible AI track:

Notebooks, slides, block, essays, product mock-ups...
Evaluated by judging panel, \$3k for top 3 submissions

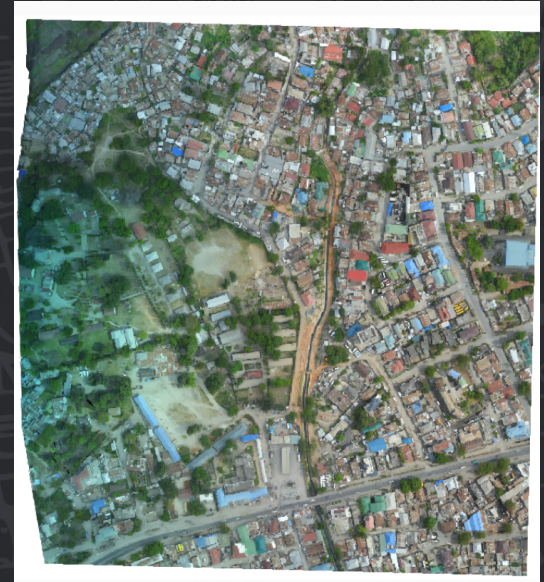


FIGURE: Sample test image and prediction chip

Join the challenge!

Competition live until
March 16, 2020

[www.drivendata.org/
competitions/60](http://www.drivendata.org/competitions/60)



Presented by:



Thanks! Dave Luo, Vivien Deparday, Axel Blanc, Caroline Gevaert, Cristiano Giovandi, Grace Doherty, Ran Goldblatt, Steven Roubinyi.