Non-traditional data sources in the context of the SDG data and statistics

Pilot projects at UN ESCWA

Contents

- Introduction
- Income and poverty – price indices
  - Web scraping
  - Barcode data
- Socio-economic situation of refugees, IDPs and host communities
  - Call Detail Records
  - Facebook Ad Platform
  - Global Database of Events (GDELT)
  - Twitter
  - Lessons learned
- Geospatial technique
  - Google Earth
  - Geospatial information in population and housing censuses

Introduction: a case for non-traditional data

- **Resource constraints** – traditional surveys are demanding for human and financial resources;
- **Statistics in crisis** – non-traditional data sources may provide needed information/data at the time of crisis, e.g. war and conflicts, COVID-19, natural disasters;
- **Opportunities brought by technology** – additional data colleacted through non-traditional sources may complement existing official statistics;
- **Time considerations** – fast data collection and estimates in specific situations.
- **Complementarity** – alternative sources do not necessarily allow to get data for SDG indicators, but provide an expanded view.

Income and poverty – price indices

- Consumer Price Index (CPI);
- International Comparison Programme (ICP).

Two approaches experimented by ESCWA:
- Web scraping – use of "robots" browsing through websites of distribution companies;
- Bar codes – utilizing data stored in the systems of large distribution companies.

contact: Majed Skaini   skaini@un.org

Use of web scraping

- High frequency of data;

- Availability of data on specific websites;

- Sources available to "robots" (not restricted);

- Sustainability of sources.

Behavioral and Humanitarian Data
on Refugees and Host Comunities in Lebanon

- Call Detail Records (Touch, Alfa)
- Facebook Ad Platform
- Global Database of Events, Language and Tone (GDELT)
- Twitter

Partner: Data-Pop Alliance (DPA)

contact: Fouad Mrad   mrad@un.org

UNITED NATIONS
ESCWA

Shared Prosperity **Dignified Life**

Call Detail Records

- Social behaviour
- Mobility patterns

- Data4Refugees (D4R) Challenge:
  - fine grained mobility data
  - Health
  - education
  - unemploymentB
  - safety
  - social integration

- Population distribution
- Population pyramid
- Population mobility
- Economic activity
- Border movement
- Security

Facebook Advertising Data

- 80% usage (of the used social platforms) in Lebanon

- Gender
- Education
- Current location
- Home location

- Flow of refugees

- Population pyramid
- Refugee distribution
- Economic activities
- Education

GDELT

- Supported by Google Jigsaw

- 3 sentiment dimensionsGender
  - Polarity
  - Magnitude
  - Score

- Open source database
- Google Cloud Platform – Big Query

- Refugees
- Economic activities
- Employment
- Poverty
- Humanitarian aid
- Incidents
- Shelter

Twitter

- Used, but not the most popular platform in Lebanon

- Scraping for tweets based on specific
  - Keywords
  - Topics
  - Locations
  - Hashtags

  - Refugees
  - Economic activities
  - Employment
  - Poverty
  - Incidents
  - Shelter
  - Demand of return

Additional sources

- Data linking
- Data validation

- VASyR - Vulnerability Assessment of Syrian Refugees in Lebanon
- CAS – LFS – Labour Force and Living Conditions Survey by the Central Administration for Statistics

Lesson learned

- Enabling conditions crucial to project success
  - Council for the Orientation for Development and Ethics (CODE)
- Privacy-conscientious design
  - Safe research with strong privacy considerations
  - On-site data processing
  - geographic and temporal data aggregation
- Accessing Call Detail Records
  - Maintain close communication between researchers and Telecom Operators
  - Utilize readily made tools
  - Potential to be gained from integrating census or survey data to individual CDRs

Geospatial techniques

- Google Earth project
  - Introduction of GIS systems
  - Remote sensing
  - Training

- Population and housing censuses
  - Automated geo-referencing of data
  - Potential for spatial analysis / small area statistics
  - Use of geospatial information in management of census operations

For further inquiries, please do not hesitate to contact ESCWA:

Fouad Mrad	mrad@un.org
Majed Skaini	skaini@un.org
Ismail Lubbad	lubbad@un.org
Wafa Aboul Hosn	aboulhosn@un.org
Juraj Riecan	riecan@un.org

UNITED NATIONS
ESCWA

Shared Prosperity **Dignified Life**