

Non-traditional data sources for official statistics

Pilot projects at UN ESCA



UNITED NATIONS

الأمم المتحدة

ESCWA

Shared Prosperity **Dignified Life**



Contents

- Introduction
- Price statistics
 - Web scraping
 - Barcode data
- Statistics on refugees and host communities
 - Call Detail Records
 - Facebook Ad Platform
 - Global Database of Events (GDELT)
 - Twitter
- Lesson learned

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 collected through non-traditional sources may complement existing official statistics;
- **Time considerations** – fast data collection and estimates in specific situations.

Price statistics

- 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 Communities 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

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
 - Refugees
 - Economic activities
 - Employment
 - Poverty
 - Humanitarian aid
 - Incidents
 - Shelter
- 3 sentiment dimensions
 - Polarity
 - Magnitude
 - Score
- Gender
- Open source database
- Google Cloud Platform – Big Query

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-conscious 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

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

**Fouad Mrad
Majed Skaini
Juraj Riecan**

**mrad@un.org
skaini@un.org
riecon@un.org**