United Nations | DESA Statistics Division

Non-traditional & Integrated Data Sources for SDGs Monitoring

Haoyi Chen, UNSD

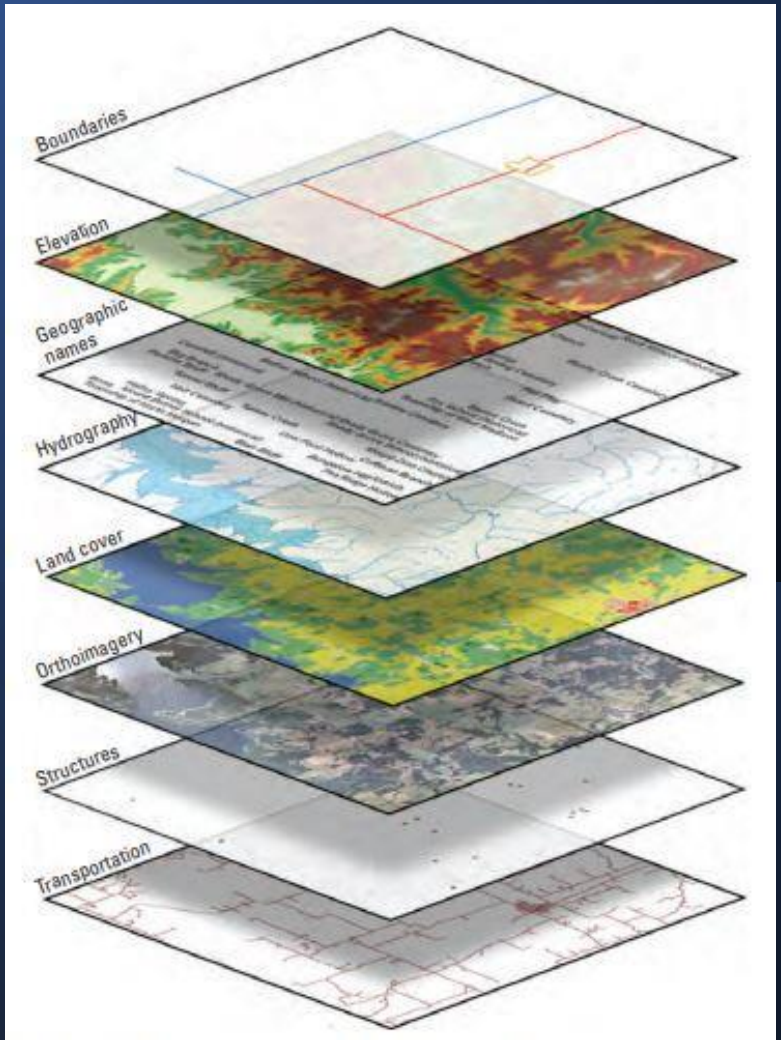
Uses of non-traditional data for SDGs monitoring

- A discussion around different types of data sources used for SDGs:
  - Geospatial information
  - Mobile phone positioning data
  - Social media data
  - Citizen data
- Data integration
- Data quality considerations

Source:

- IAEG-SDGs Wiki for Good Practices in Non-traditional Data Sources and Data Innovations
- UNESCAP Data Integration Community of Practice
- Academic research

Geospatial information (1)



Source: U.S. Geological survey

- Most widely used in SDGs monitoring
- Data integration is required

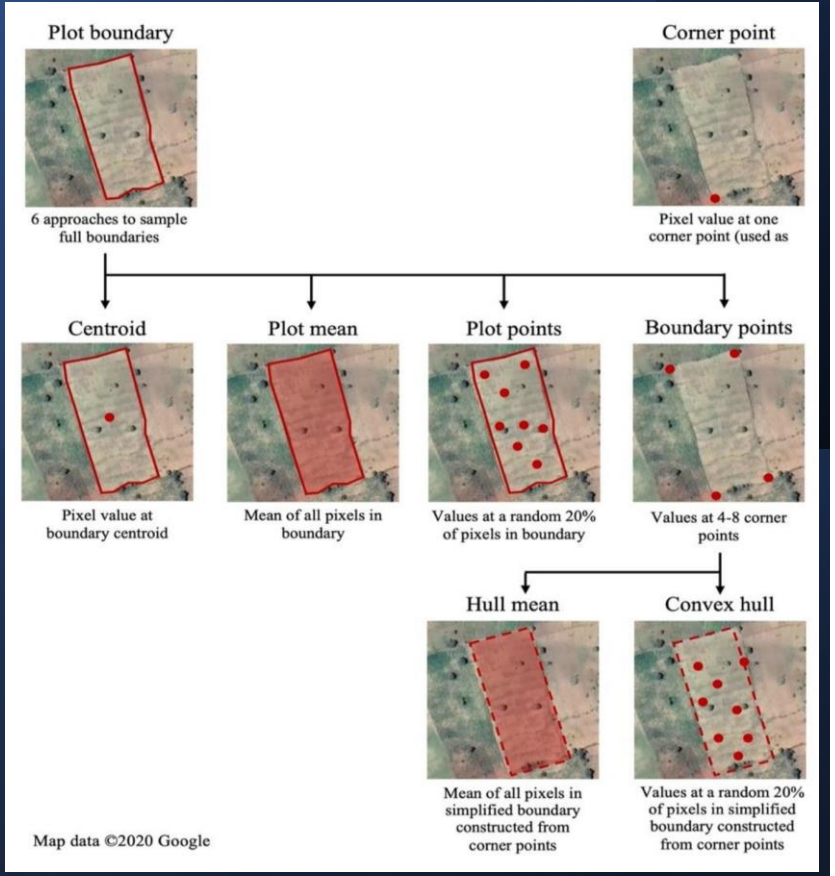
Traditional data under NSOs

- Geographic Information System (GIS) to establish boundaries
- Geocoded population information (people)

- Satellite images (remote sensing)
- Locations of key infrastructures (hospitals, banks, etc.)
- Sensors (e.g., to monitor air pollution)

Non-traditional data sources

Geospatial information (2)



Source: Azzari et al. 2021

❖Variations in data availability

- Availability of governmental sources
- Financial and technical capability to procure non-governmental sources

❖Data quality considerations

- Comparability and Consistency: Key consideration for data integration
- Relevance: Do the data have adequately high resolution?
- Accuracy: Images → Meaning
  - Study by Azzari et al. 2021: Use household surveys to support satellite-based crop type mapping

Mobile phone positioning

- Mobile phone signals ~ People
  - Population counts
  - Population movements
- Fine time intervals (hours, days, months)
  - Relevance for specific topics
  - Cot et al. 2021 study on social distancing
  - Improved timeliness for topics of population change and mobility (internal migration)
- Data integration → Data quality
  - Accuracy: Benchmarking using traditional statistics
  - Consistency & Comparability: Use multiple sources of non-traditional data (e.g., Tu et al. 2017 study combines mobile phone positioning and social media check-ins data in China)

Social media data

- ❖Social (media) listening, usually qualitative information
  - About the users (gender, age, etc.)
  - About their sentiments, opinions, etc.
- ❖Challenges in analysis
  - Multiple media: texts, photos, videos
  - Multiple postings per user
  - Required sophisticated analytical tools (Natural language processing, Sentiment analysis, etc.)
  - Example:
    - Pristiyono et al. 2021's sentiment analysis of COVID-19 vaccines in Indonesia on Twitter
    - Ondrikova et al. 2023's analysis of Google search terms to predict norovirus spread
- ❖Data quality consideration
  - Availability using web-scrapping or scanner tools
  - Data concerns are intertwined with concerns about the analytical strategies

Citizen data

- Diverse in types and methodological orientations
- Fill data gaps for SDGs
  - Specific SDG indicators:
    - Ghana was the first country to report SDG indicator 14.1.1b for marine plastic using citizen data
    - Foundation for Free Press (FLIP) Collaborate with DANE to support efforts in producing estimates for SDG indicator 16.10.1
  - Leaving no one behind:
    - Community-driven data that identifies population groups that are left behind
- Implementation of the Copenhagen Framework on Citizen Data to support work on citizen data
  - Data quality
  - Qualitative data
  - Sufficient and meaningful engagement of citizens in the DVC
  - Trust building

To recap

- Large variations in non-traditional data types and usages
- Opportunities and challenges:
  - Filling data gap and making data more inclusive
  - Data access
  - A large variety of data types
  - Quality and comparability (data integration can help?!)
  - Data analysis
  - Data integration

United Nations | DESA Statistics Division

Thank you!

chen9@un.org