Common MPD project pitfalls

Erki Saluveer
Positium

Twitter: @positium
www.positium.com
erki.saluveer@positium.com
Figure 5  Snapshot of countries using MBD analytics applications to support the Covid-19 response

#SOTEU

We propose to reduce emissions by at least 55% by 2030

Source: GSMA
Case: Official Travel Statistics, Estonia (2009-…)

Client: Central Bank of Estonia
(the official travel statistics provider in Estonia)

- Since 2009
- The longest-running official statistics time series based on big data in the world
- The only recurring official statistics based on big data in the world
- Quality checks done by statisticians at the bank
- Results compared to survey method:
  - 4x faster
  - 200x sample size
  - 12x country-level breakdown
  - 2.5x more cost-efficient
  - 100% less burden on tourists
Covid-19 first wave (2020)

On 12 March 2020 – the government declared an emergency situation in Estonia, several restrictions on population movement were imposed. Prime Minister Jüri Ratas, head of emergency situation, made a decision on 20.03.2020 that Statistics Estonia should carry out a task outside of its programme: use data from MNOs to build analytics tools regarding the spread of coronavirus.
Estonian Covid-19 pandemic dashboard

- Daily update on mobility index
- All three mobile network operators
- Estonian Statistics as quality assurance
- Live updates for Estonian Prime Minister and government institutions

https://mobility.positium.com/covid19/
FIVE PRINCIPLES TO ENSURE TRUST

Following the Fundamental Principles of Official Statistics

- **Necessity & Proportionality**
- **Professional Independence**
- **Privacy Protection**
- **Commitment to Quality**
- **International Comparability**

Guiding principles to maintain public trust in the use of mobile operator data for policy purposes (Data & Policy ISSN: 2632-3249)

https://tinyurl.com/mpdprinciplesarticle

Principles are interlinked

Maximizing efforts in adhering to one specific principle can **catalyse the downfall of other principles** – one principle might start to rule over other principles.
Principles are interlinked (1)

**Fit-for-purpose <-> international comparability**

**FIRST PITFALL**

Looking at MPD projects as isolated events where methodologies and best practices are chosen for a specific project, not for the field standards.

If the results are not comparable internationally, the value might degrade. Also, there is a big chance that isolated project will not be sustainable over longer time periods.
Principles are interlinked (2)

**Fit-for-purpose <-> data quality**

**SECOND PITFALL**

Failing to describe the differences between the reality model and the data model

- MPD is often fit for purpose, but it will only hold true if the principle of commitment to quality holds true
- Data quality = following scientific standards and methods
- There is one reality model, but with MPD projects there a number of different data models

Without quality aspects, the fit-for-purpose starts to degrade
Principles are interlinked (3)

Privacy $\leftrightarrow$ professional independence $\leftrightarrow$ data quality $\leftrightarrow$ fit-for-purpose

THIRD PITFALL

Letting the privacy aspects lead the course of the MPD projects and degrade data quality and through that fit-for-purpose

For example, the use of 24-hour maximum length of pseudonymous hash will have heavy impact on detection of usual environment and place of usual residence

Does heavy compromises on generic privacy approaches affect professional independence, data quality and fit-for-purpose?
How to prepare? Some tips

1. **Define the methodologies** that will serve the purpose of the project – this will generate an understanding of the impact of privacy preserving techniques and ability to use the best methodologies available
   - **Legal approach**: Do a data privacy impact assessment – is our proposed method proportional to the privacy risks? What are the safeguards in place (apart from aggregation at the source)?
   - **Technical approach**: Use computational privacy-preserving methods (a la multi-party computation)
   - **Institutional approach**: Instead of discussing with the PHD holder, work with a supplier of data science services to the PHD holder that holds the necessary credentials for work in transportation and can act in a transparent manner

2. **Define the best data model** that serves fit-for-purpose and quality principles at the beginning of the project

3. **Document the differences** between the reality model and the data model

4. Use above as an input for discussions with PHD holder
FIVE PRINCIPLES TO ENSURE TRUST

FOLLOWING THE FUNDAMENTAL PRINCIPLES OF OFFICIAL STATISTICS

- Necessity & Proportionality
- Professional Independence
- Privacy Protection
- Commitment to Quality
- International Comparability


Guiding principles to maintain public trust in the use of mobile operator data for policy purposes (Data & Policy ISSN: 2632-3249)
https://tinyurl.com/mpdprinciplesarticle
Always go back to fit-for-purpose!

Email: erki.saluveer@positium.com