Roadmap to Mobile Phone Data

Building preparedness towards use of mobile data

Siim Esko
Positium
Estonia

@positium
www.positium.com
<table>
<thead>
<tr>
<th>Data</th>
<th>Infrastructure</th>
<th>Skills</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mobile operators</td>
<td>Data centre</td>
<td>Staff, consultants, universities</td>
</tr>
<tr>
<td>Legal</td>
<td>Technology</td>
<td>Funding</td>
</tr>
<tr>
<td>Privacy</td>
<td>Mature big data tech</td>
<td>Low-hanging fruit International co-funding</td>
</tr>
<tr>
<td>preservation</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Legal supervision</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Processing Location

Depends on:
• Legal basis -> What is possible?
• Method -> Is centralized processing necessary for validity?
• Potential Future Scope -> What is efficient?
Components of the system – open source

Data extraction → Raw data → Transfer data → Data importer and formatter

Prepared data → Core processes → Core data model → Domain-specific processes

Aggregated data → Reference data & calibration → Extrapolating to general population → Domain specific resulting indicators

Dissemination (API) → Domain aggregated data → API / Applications
Desired aspects of the system

- Modular architecture
- Transitional databases (for data mining and QA)
- Metadata generation throughout the process
- Parallel processing (Hadoop?) according to processing time requirements
- API for dissemination
- Data transmission standards (SDMX, JSON, XML, ...)

- White box / Grey box / Black box
- Maintenance agreements (SLA)
Required Skills

• Tools for running processing algorithms (e.g. Java, Python, R)

• Supporting tools for QA of the data and visual analyses (e.g. GIS tools such as Esri, MapInfo, QGIS, R, and Excel, Tableau, Qlik, etc.).

• Skills of developing software for the parallel processing of big data – data processing software, ETL (Extract, Transform and Load) developer with Python, Java, PIG, Hive, Spark, SQL, etc.

• Skillset for PostgreSQL/PostGIS, Oracle or any other database system (including GIS data) is required.

• System architect and system operators

• Statistical data processing methodologists with big data experience.
Desired Timeline

- **Interest: seminars**
- **Feasibility**: 6-12 months
- **Decision**: 6-12 months
- **Implement**: 6-12 months
- **Continuous updates**
Cost of the system

- Seminars, workshops
- Pilot Project / feasibility assessment
- Implementation cost
- Maintenance (SLA)
Cost of the system
Cost of the system
Funding for Mobile Data Analytics comes from Value

- Population statistics
- Everyday commuting
- Transportation planning
- Urban planning
- De facto population
- Tourism statistics
- Epidemiology
- Economics
- Natural disasters, proximity of population to risks
- Safety and Security
- ICT statistics
- Spatial marketing
- Scientific research in many areas
Homework

• What are the first priority projects in your country for the use of mobile phone data?
  • Personal opinion – Gathered at the workshop
  • Consolidated NSO opinion – To do after returning back to office