Unified Migration Data Analytical System (UMAS)

Secretariat of the State Commission on Migration Issues
Public Service Development Agency
Ministry of Justice of Georgia

Bangkok, Thailand
5-8 February, 2019
Briefly about UMAS, Georgia’s migration management and data sources

Main talking points:
1. System workflow and architecture;
2. ETL, data quality assessment;
3. Person identification;
4. Business application;
5. Coordination.
Progress from 2014 to present

- Project concept developed
- Legal framework adjusted
- Project team staffed
- Data sources secured
- System trial version
- Test analytical reports
- Operation in real settings
- New data sources added
- Development of business processes
- Team staffing ongoing
- Ministry of Internal Affairs;
- State Security Service of Georgia;
- **Public Service Development Agency**;
- National Agency of Public Registry;
- Ministry of Finance;
- Ministry of Foreign Affairs;
- Ministry of Education and Science.
1. System workflow and architecture
Simplified workflow

- First 4 steps take place in the System, using ready soft + specific scripts;
- 5th step is implemented outside the system (BI tool).
Components of the System

Hadoop eco system

Scala-Java

Spark

HDFS
Hive
Livy

Directories
Translation module
Version control
System administration module

Transformer
Process manager
Identification
Report generator

NiFi
Talend
Tableau

Hadoop ecosystem
2. Data cleaning and quality assessments
Data Cleaning throughout ETL

- Deleting additional symbols;
- Cleaning from duplicates;
- Deleting incorrect entries;
- Systematizing date formats;
- Systematize coding;
- Correcting entries where possible after analyzing the relevant business processes.

!! *Cleaning, analyzing and correcting data points typically used for person identification is of particular importance.*
Data Cleaning throughout ETL

A **data passport** is prepared for each data source:

- Every data point is analysed separately;
- abnormalities studied and cross-checked with the business process;
- recommendations and questions noted if relevant.

*Data quality assessment needed to understand and interpret the results and to work with the data sources in the future where possible.*
3. Person identification
Data needed for person identification

Varies across the data sources:

- Name in English
- Last name in English
- Name in Georgian
- Last name in Georgian
- Date of birth
- Identity number
- Passport number
- Citizenship
- Passport issue date
- Place of birth
- Tax identity number
Person identification process

- Groups of Similar persons
- Identified Groups of similar individuals
- Anonymized individuals In the database with different IDs
- New anonymized individuals

Steps:
1. Grouping
2. Identification
3. Grouping using different method
4. Identification
5. Anonymised individuals
4. Business application
Access to different layers of the System
Application

- UMAS is Primarily a ‘B2B’ system, focusing on the needs of the state agencies involved in migration management.

- The System is intended as a business intelligence tool, improvement and production of the migration statistics is not the primary goal, but rather a positive by-product.

- It is an analytical system and cannot be used for administrative or operative purposes.

- Users can apply using special request form and receive a report with access to anonymous individual data.
5. Coordination
Coordination on many levels is important because of:
1. Regular data exchange;
2. Report request and generation;
3. Understanding the details of business process behind data;
4. Interpretation of the results;
5. Maintain political support;
Nino Ghvinadze
Data Analyst
Secretariat of the State Commission on Migration Issues (SCMI)
Public Service Development Agency
Ministry of Justice of Georgia

n.ghvinadze@sdagov.ge
scmi-secretariat@sdagov.ge
www.migration.commission.ge