Identifier Service developed in NSO Finland including GSIM-based information objects

ModernStats World Workshop 2019
The Identifier Service: vision

The Identifier Service will deliver global persistent identifiers for the needs of statistical data production in Statistics Finland.

Organisations managing their own PIDs

Organisations using StatFIN PIDs

Thereby The Identifier Service enables interoperability at national and international level. In addition, it supports intelligent automation of processes and trackability.
Governance model of the Identifier Service

**Identifier Owner of the Identifier Service**
In charge of developing the functionalities of the Identifier System further. In charge of proper guidance.

**Register Identifiable Artefacts in the Identifier Service**

**Identifier Service**
Includes objects added into the information model

**Information model TIMO (GSIM-based)**

**Operator**
In charge of the technical maintenance of the Identifier Services as well as adding new object types into the Identifier Service by request.

**Architecture group**
Develops the information model further taking into account the needs in data production.

**Production System**

Identifiers for Business and Processes:
- (Statistical Program, Process Step, Rule...)

Identifiers for Data Structures:
- (Data Set, Logical record, Referencial Metadata Set...)

Identifiers for Content Metadata:
- (Unit Type, Universe, Variable...)

Identifiers for Data Collection:
- (Data provider, Collection round...)

Identifiers for Geospatial objects?
- (Geographies...)

Identifiers for Data Dissemination:
- (Product...)

**Information model TIMO (GSIM-based)**
Content of the Identifier Service for statistical data production

Information model TIMO (GSIM-based)
Principles of the Identifier Service

- Delivers identifiers for specified information objects only
- Delivers only identifiers that are globally unique and persistent in time
- There is no limit for the number of identifiers
- Delivers only those types of identifiers that are specified
- Delivers identifiers only to registered production systems
- Only operator can add new information object types as well as production systems requesting identifiers
- The location of the identified object can be generated / registered
- It is possible to keep a record of the "old" existing identifiers, not generated by this service
The Identifier Service consists of:

- Offering identifiers for specified objects
- Keeping a record of the existing identifiers
- Offering the location of the identified object
- Access rights
- Identifying the type of identifiable artefacts
- Offering identifiers for specified objects
- Keeping a record of the existing identifiers

User Interface for the Administrator

Service for administration

Registering/Identifying Service

Resolving
Logical information model
Implemented in Azure environment

- Scalability
- Usability across different systems
- Possibility to open the API outside Statistics Finland if needed
How will the persistent identifiers look like?

- The system created can, in principle, deliver any kind of identifiers.
- In the near future, both URI and URN identifiers will be used. For exceptional cases also DOI can be obtained.
- URN’s for published objects are planned to be linked to the URN Service of the National Library of Finland.

Prefix, an example

URN:NBN:fi:stat/subsection/identifier

Suffix, has to parts

The elements included in the information model, i.e. Population, Rule, Concept etc.

Identifier for a certain information object, typically a surrogate.
Projects piloting the Identifier Service 2019 in Statistics Finland

- Quality reports (SIMS) for the producers of Official Statistics in Finland
- GSIM based metadata system METSY: Core metadata objects (from GSIM green and yellow part)
- Web-project: main objects that will be published on our website
- Linked Open Data: Geospatial classifications

All the necessary changes needed based on this piloting will be implemented in the Identifier Service during the year 2019

→ Q1/2020 Identifier Service in production in Statistics Finland
Challenges and open questions

• Resolving in more complex cases (for instance, several locations) as well as creating the connection to the URN Service of National Library
• How to ensure that the model behind, GSIM-based TIMO v0.5, will react to specific needs and at the same time stay stable enough? (TIMO is more of a logical than conceptual model)
• To balance with the need to create identifiers for objects outside the model = outside statistical data production → one solution could be to have logically separate service with separate governance model but still the same physical service?
• The governance of common services is a challenge, even within our NSO. What would be required if the Identifier Service would be a CSPA service?
Any questions? Contact us:

Essi Kaukonen essi.kaukonen@stat.fi
Daniel Davis daniel.davis@stat.fi
Mikko Saloila mikko.saloila@stat.fi
Toni Räikkönen toni.raikkonen@stat.fi
Aura Pasila aura.pasila@stat.fi
Jennika Leino jennika.leino@stat.fi

If you are interested, please, contact us!