Data Vocabularies Tool

The Finnish Interoperability Platform and tools
Riitta Alkula, Chief Specialist

DIGITAL AND POPULATION DATA SERVICES AGENCY

Digital and Population Data Services Agency's mission and impact objectives

Mission
(304/2019)

To promote the digitalisation of society,

ensure the availability of data

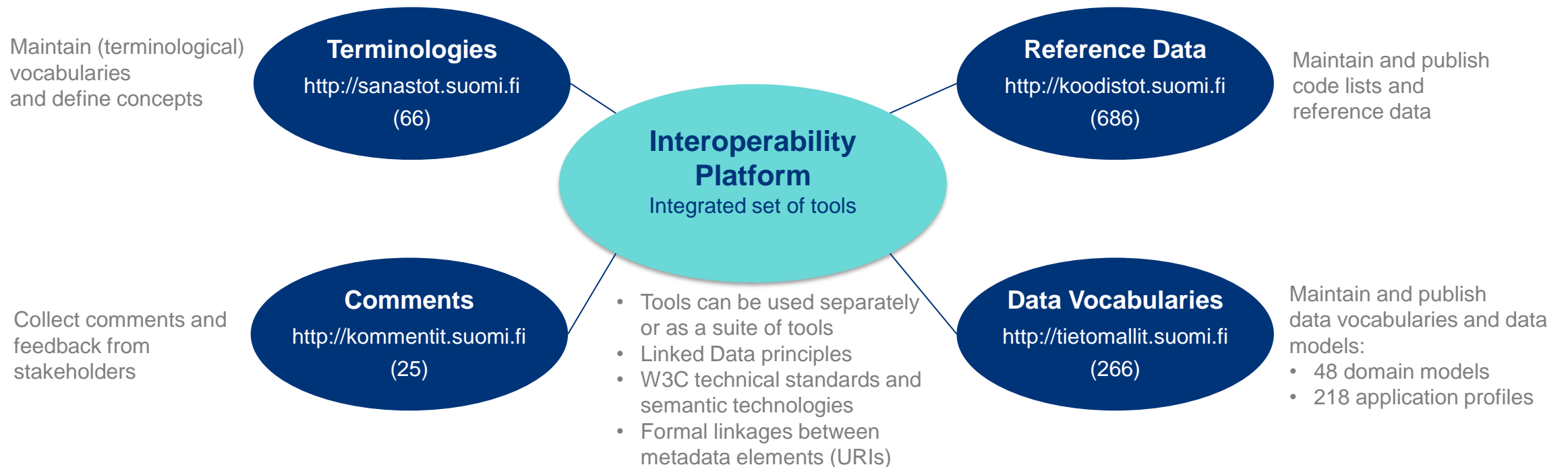and **offer** services for customers' life events.

Impact objectives

Digital services will increase the utilisation of the data and enhance the operation of public administration.

Trust and legal protection in society will improve.

Public administration services will be more customer-oriented and smoother for **citizens and companies** to use.

Towards a smoothly functioning Finland

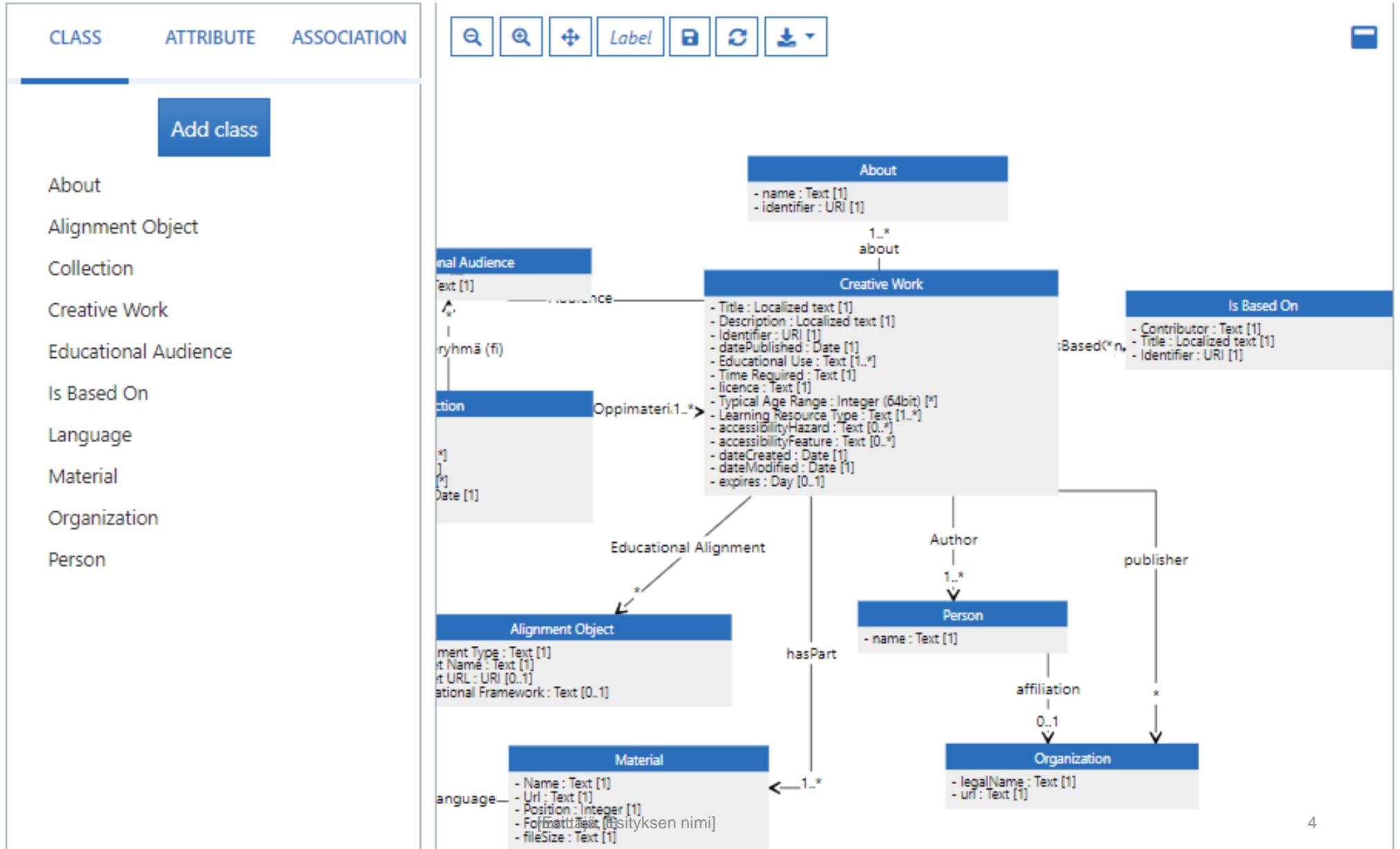# Interoperability platform and its tools

Maintain (terminological) vocabularies and define concepts

Terminologies

http://sanastot.suomi.fi

(66)

Reference Data

http://koodistot.suomi.fi

(686)

Maintain and publish code lists and reference data

Interoperability Platform

Integrated set of tools

Collect comments and feedback from stakeholders

Comments

http://kommentit.suomi.fi

(25)

- Tools can be used separately or as a suite of tools
- Linked Data principles
- W3C technical standards and semantic technologies
- Formal linkages between metadata elements (URIs)

Data Vocabularies

http://tietomallit.suomi.fi

(266)

Maintain and publish data vocabularies and data models:
- 48 domain models
- 218 application profiles

Metadata as human and machine-readable data web – turning in future into a smart data catalog.

Objectives of the Interoperability Platform

- To provide a common set of tools, guidelines and principles for the Finnish public sector in one place
- Better visibility for content providers
- Efficient reuse of once created content
- Both human and machine readable content
    - When a data model is created with the Data Vocabularies tool, it is provided with a unique identifier (URI)
    - Machine readable and identifiable concepts, data models, reference data etc. can be referenced elsewhere
    - For example, classes and properties created in data models can be based on existing concepts in terminologies
- So common concept definitions and general level "reusable" data models promote interoperability between information systems
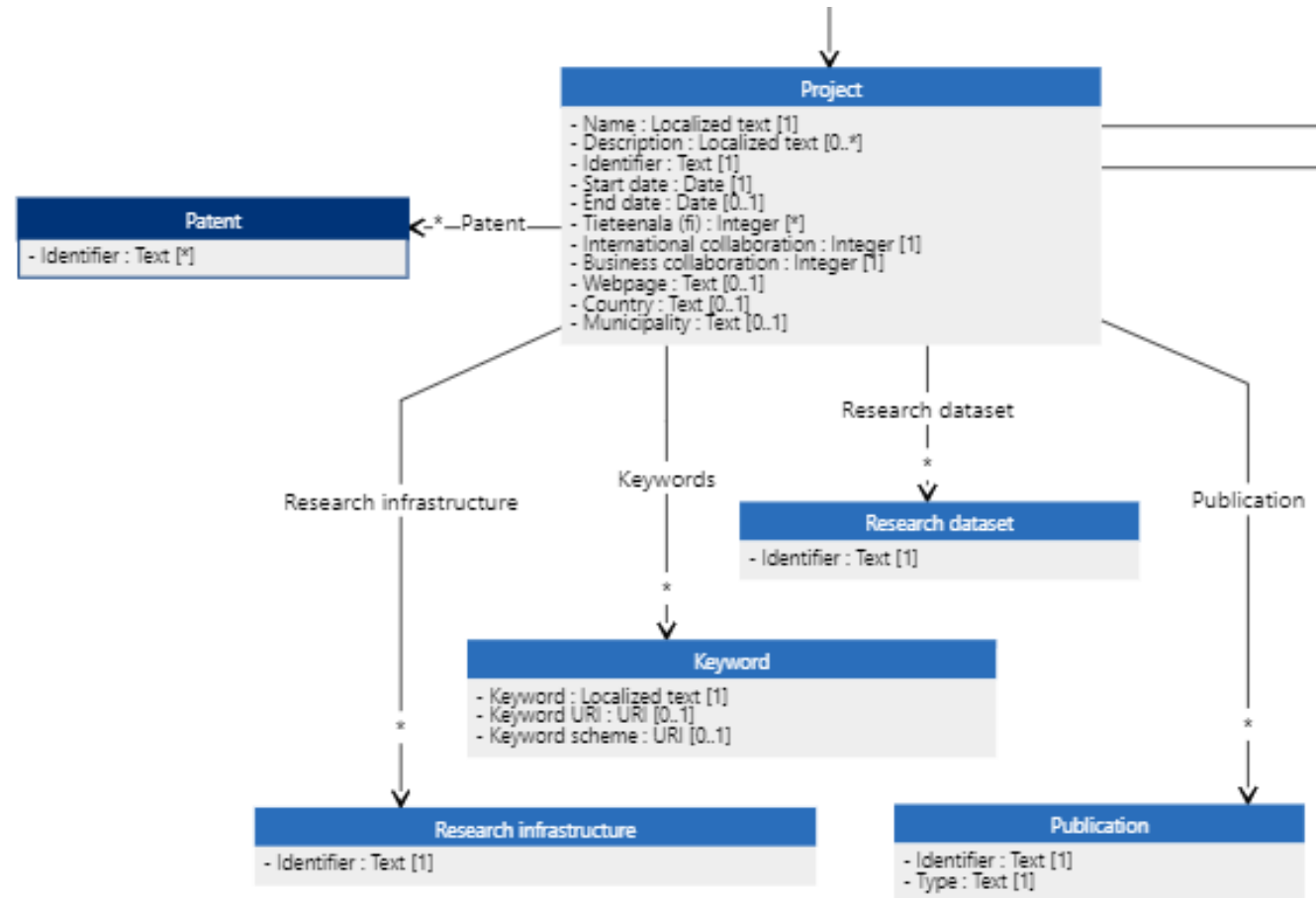
Content creation and editing process

DVV adds an organisation on the Interop. Platform

↓

Main user gives access rights to the editors (modellers)

↓

Editors create data models

↓

Editor creates/adds classes, attributes and associations

↓

Editor adds links to other resources

↓

Human and machine-readable descriptions are published

↓

The model is edited/modified, when needed

↓

In the end, the model is retired or replaced

DVV creates a new user organisation on the platform. DVV gives access rights to the main user of the organisation.

Main user gives access and editor rights to appropriate tools for the users of the organisation

Editors create data models: either general data compoment libraries for reuse, or application profiles for context spesific descriptions

Editors create content (resources) within the data model. Data is stored as RDF triples. The tool gives persistent URI addresses automatically.

Resources are linked to other resources. Associations are added between classes, attributes linked to code sets etc. when needed

Editor gives status codes to express the maturity of the model: *Hidden* for internal review, *Draft* for candidate, *Valid* for stable content

When the data model meets the end of its life cycle, it gets a status code *Superseded* if replaced with another data model, *Retired* if not in use anymore.

Data VocabulariesTool in more detail

- The Data Vocabularies Tool is a browser-based application for managing and publishing data vocabularies (data models)
- **The aim is to give easy to use interface for business users. From the logical level descriptions, technical descriptions are created "under the hood".**
- The tool contains data component libraries, which are data specifications for harmonising information, used jointly by different actors. (Ref. Core Vocabularies)
- Based on those data component libraries, application specific data models, i.e. application profiles, are created.
- The tool facilitates linking between information contents and subsequently the linking of information to international or other (preferably official) standards the actors wish to use. The tool creates automatically the URI identifiers for the resources.
- Internally, metadata descriptions are based on RDF, SKOS, OWL, and SHACL.
- From internal descriptions, various output formats (serialisations) can be generated, including RDF, JSON-LD, Open API, Turtle, and XML formats.
- The contents are visualized in UML-like graphic presentation.

The Tool is intended for

- information architects and modellers
- business users in an organisation to visualise their data resources used in information exchange between organisations
- anyone searching for information in the data repositories of public administration
- administrators of registries and data repositories who wish to make their data machine-readable
- system vendors building new systems that use data and data models from the information systems and applications of public administration.

The Tool is used for

- data modelling that aims at semantic interoperability necessary for modern digital services
- describing information contents in a such way that the meaning of the information can be defined using terminologies and transmitted in a machine-readable format, using permanent identifiers
- creating and managing information content as open linked data.

Specifications

- Structure based on linked-data standards (RDF) and vocabularies (RDFS, OWL, SHACL, SKOS, Dublin Core)

- Aligned from the start with the EU EIF/ISA linked data principles and linked data best practices, aware of the ISO 21838 family (top-level ontology) development, importing OWL ontologies under development

- Support for enabling in-tool inferencing and validation under consideration (already supported by the Jena backend)

- Multiple serialisations: JSON-LD, RDF/XML, Turtle, etc.

- Opening SPARQL endpoint(s) are under consideration

"Re-use when possible, mint when necessary"

Our principles and as well as the slogan above are based on the process presented in the [Cookbook for Translating Relational Data Models to RDF Schemas](#):

1. Research existing concepts, terms, codes and data models and their usage and maximise re-use of them.

2. When new concepts can be seen as spesialisations of existing concepts, create narrower concept, or sub class and sub properties as appropriate.

3. Where new concepts are required, create them following commonly agreed best practice in terms of naming conventions etc.

4. Publish descriptions within a highly stable environment designed to be persistent and identify them with URI addresses. (URIs given by the system automatically.)

More information

- Information in English: https://wiki.dvv.fi/display/YTIJD/Finnish+interoperability+project%2C+platform+and+tools
- The tools:
  - https://sanastot.suomi.fi/ (the Terminologies Tool)
  - https://koodistot.suomi.fi/ (the Reference Data Tool)
  - https://tietomallit.suomi.fi/ (the Data Vocabularies Tool)
- Presentation videos: https://youtu.be/eeonIC5mPeM (in English), https://youtu.be/ERLe8kFmbx0 (in Finnish), https://youtu.be/3x_ryInrpnM (in Swedish)