Guaranteed and semantically interoperable public government data published and shared as Linked Open Data

Jakub Klímek, Martin Nečaský, Petr Křemen
2021 legislation update in Czechia

Obligation to **publish** data as open data by default

Obligation to **consume** open data
Public Data Space (PDS)


Guaranteed Accessibility for public administration
PDS: regular open data not enough

https://data.gov.cz

Interoperability?
Reusability?
EU DIRECTIVE 2019/1024
Article 2 (15)

‘formal open standard’ means a standard which has been laid down in written form, detailing specifications for the requirements on how to ensure software interoperability.

In Czech, FOSes are called Otevřené formální normy (OFN) and they are coordinated by MINISTRY OF THE INTERIOR OF THE CZECH REPUBLIC.
FOS - what it contains

1. Conceptual data model
   - data format independent
1. **Conceptual data model**
   - data format independent

2. **Specification**
   - detailed description of classes, attributes and relations
   - reused code-lists

<table>
<thead>
<tr>
<th>RDF Class:</th>
<th>dcat:Catalog</th>
</tr>
</thead>
<tbody>
<tr>
<td>Definition:</td>
<td>A curated collection of metadata about resources (e.g., datasets and data services in the context of a data catalog)</td>
</tr>
<tr>
<td>Sub-class of:</td>
<td>dcat:Dataset</td>
</tr>
<tr>
<td>Usage note:</td>
<td>A Web-based data catalog is typically represented as a single instance of this class.</td>
</tr>
<tr>
<td>See also:</td>
<td>§ 6.5 Class: Catalog Record, § 6.6 Class: Dataset</td>
</tr>
</tbody>
</table>
FOS - what it contains

1. Conceptual data model
   ○ data format independent

2. Specification
   ○ detailed description of classes, attributes and relations
   ○ reused code-lists

3. Examples and data schemas
   ○ JSON(-LD) + JSON Schema
   ○ possibly CSV + CSV on the Web
   ○ possibly XML + XML Schema

Project: Development of data policies in the field of improving the quality and interoperability of public administration data. Reg. number: CZ.03.4.74/0.0/0.0/15_025/0013983
FOS - what it contains

1. Conceptual data model
   - data format independent

2. Specification
   - detailed description of classes, attributes and relations
   - reused code-lists

3. Examples and data schemas
   - JSON(-LD) + JSON Schema
   - possibly CSV + CSV on the Web
   - possibly XML + XML Schema

4. Predefined metadata
   - DCAT-AP
   - for National Open Data Catalog

---

**Číselník pro sporty**

Ministry of the Interior of the Czech Republic


Dataset theme
Government and public sector

Spatial coverage
Česká republika

Documentation
Show documentation

Frequency
irregular

Contact point
Kontakt pro otevřená data

Specification
OFN: Číslníky

---

FOS for codelists

---

FOS for codelists of sports
FOS - why is it useful

Formal open standard (FOS) function

- Obviously: *syntactic* interoperability among various publishers

- But also: protection of interoperability
  - When system supplier changes
  - When system operator changes

- Moreover: faster Linked Data adoption
  - JSON in fact JSON-LD => LD
  - CSV + CSVW => LD

- Compliance with FOS => LD

---

*Interoperability*

*Reusability*
Public Data Space recap

What we have so far in Public Data Space (PDS)

- Open data
- **Findability** in National Open Data Catalog
  - and therefore data.europa.eu
- Guaranteed *accessibility* for public administration
- **Syntactic interoperability + reusability**
  - JSON-LD => Linked Data
  - Formal Open Standards (FOSes)

What about *semantic interoperability*?
Semantic mapping of FOSes

Data structures in FOSes can be mapped to multiple semantic definitions

{
  "typ": "Turistický cíl",
  "iri": "https://www.spilberk.cz/",
  "název": {
    "cs": "Hrad Špilberk",
    "en": "Špilberk Castle"
  }
},

Context 1: Common Linked Data vocabularies

@prefix schema: <http://schema.org/> .
@prefix dcterms: <http://purl.org/dc/terms/> .
<https://www.spilberk.cz/> a schema:TouristAttraction ,
  dcterms:title "Hrad Špilberk"@cs , "Špilberk Castle"@en .

Context 2: SGoV - semantic government vocabulary

@prefix cíle: <https://slovník.gov.cz/datový/turistické-cíle/pojem/> .
@prefix věci: <https://slovník.gov.cz/generický/věci/pojem/> .
<https://www.spilberk.cz/> a cíle:turistický-cíl  ;
  věci:název "Hrad Špilberk"@cs , "Špilberk Castle"@en .
Semantic government vocabulary (SGoV)

- ecosystem of layered linked ontologies based on Czech legislation
  - e.g. one per Act
  - e.g. one per specific domain
  - ...

- semantic definitions of
  - concepts
  - relationships

- mappings among concepts from different domains
  - including ISA^2 Core Vocabularies and other LD vocabularies

- based on UFO-A, UFO-B, UFO-MLT
  - + OntoUML as modeling language
  - inspired by Giancarlo Guizzardi
SGoV example: Vehicles

Czech Act 56/2001 on the conditions of operation for vehicles on roads

COUNCIL DIRECTIVE 1999/37/EC on the registration documents for vehicles

REGULATION (EU) No 2013/168 OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL on the approval and market surveillance of two- or three-wheel vehicles and quadricycles

REGULATION (EU) 2018/858 OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL on the approval and market surveillance of motor vehicles and their trailers, and of systems, components and separate technical units intended for such vehicles
SGoV example: Vehicles: EU reg. 1

REGULATION (EU) No 2013/168 OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL on the approval and market surveillance of two- or three-wheel vehicles and quadricycles

Electric engine
- maximum continuous rated power

Combustion engine
- maximum net power
- engine capacity

Engine

Manufacturer
- is responsible for a type of vehicle
- has manufacturer

Type of vehicle
- has vehicle type

L-category vehicle

Three-wheel moped
Two-wheel motorcycle with side-car
Light two-wheel powered vehicle

Two-wheel motorcycle
Powered tricycle
Light quadricycle

Base vehicle
Incomplete vehicle
Completed vehicle
Complete vehicle

Heavy quadricycle
REGULATION (EU) 2018/858 OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL on the approval and market surveillance of motor vehicles and their trailers, and of systems, components and separate technical units intended for such vehicles
COUNCIL DIRECTIVE 1999/37/EC on the registration documents for vehicles

SGoV example: Vehicles: EU directive

EU reg. 1

EU reg. 2
SGoV example: Vehicles: Czech act

Czech Act 56/2001 on the conditions of operation for vehicles on roads
SGoV tools and workflow

1. Look-up/Create relevant vocabularies
   - Control Panel
   - Vocabulary stubs
   - Thesaurus

2. Create new concepts or extend existing ones
   - TermIt

---

Project: Development of data policies in the field of improving the quality and interoperability of public administration data. Reg. number: CZ.03.4.74/0.0/0.0/15_025/0013983
SGoV tools and workflow

1. Look-up/Create relevant vocabularies
2. Create new concepts or extend existing ones
3. Connect concepts by semantic links (e.g. specializations)

Vocabulary stubs | Thesaurus | Conceptual Model

Project: Development of data policies in the field of improving the quality and interoperability of public administration data. Reg. number: CZ.03.4.74/0.0/0.0/15_025/0013983
SGoV tools and workflow

1. Look-up/Create relevant vocabularies

2. Create new concepts or extend existing ones

3. Connect concepts by semantic links (e.g. specializations)

Vocabulary stubs | Thesaurus | Conceptual Model | Semantic vocabulary

4. Publish vocabularies as linked open data

- IRI dereference
- SPARQL endpoint

Project: Development of data policies in the field of improving the quality and interoperability of public administration data. Reg. number: CZ.03.4.74/0.0/0.0/15_025/0013983
SGoV tools and workflow

1. Look-up/Create relevant vocabularies
2. Create new concepts or extend existing ones
3. Connect concepts by semantic links (e.g. specializations)
4. Publish vocabularies as linked open data
5. Generate a data schema and its documentation
6. Fill-in the form to create/edit data
7. Generate a form according to the data schema
8. Work in progress...

Vocabulary stubs | Thesaurus | Conceptual Model | Semantic vocabulary | Data schema | Form | Data

Control Panel | OntoGrapher | FOC | Data Manager

Project: Development of data policies in the field of improving the quality and interoperability of public administration data. Reg. number: CZ.03.4.74/0.0/0.0/15_025/0013983
Public Data Space (PDS) architecture

- Data in one of FOS formats (XML, JSON, CSV, …)
- PDS storage
- Open-source, reusable ETL-like tool
- Guaranteed accessibility
- Regular open data
- Data in all FOS formats (LOD, JSON, CSV, XML, …)
Current state of PDS

- Public Data Space (PDS) architecture is done
- Allocated finances from NextGenerationEU
- Ongoing preparations for implementation

- First datasets to populate PDS will be codelists
- FOS for codelists based on SKOS
Vision of extension to EU context

Czech data about vehicle registrations

```json
{
  "typ": "Vozidlo",
  "iri": "https://vozidla.cz/vin/YBHNASD123458791",
  "název": {
    "cs": "Moje nové elektroauto",
    "en": "My new electric vehicle"
  }
}
```

Automatic, semantically interoperable transformations

SEMIC Vehicle registration evidence (SDG)

Czech vehicles act (SGoV)

CZ <-> EU legislation mapping (SGoV)

EU legislation based SEMIC Vocabulary mapping
Guaranteed and semantically interoperable public government data published and shared as Linked Open Data

Questions?

Jakub Klímek, Martin Nečaský, Petr Křemen