Zellîj: A semantic pattern development and documentation system

George Bruseker (Takin.solutions)
CRM SIG, Luxembourg, 7/12/2022
What are we doing?

Semantic Data patterns database and documentation workflow system - a way of organizing semantic data patterns and sharing them, enabling semantic data to be generated and reused in a practical and sustainable way.

You get:

- the expressive flexibility of semantics
- the structural reliability of traditional data standards
- a way to consistently and effectively document all that
Background: SRDM

Person Reference Data Model

Author: George Bruseker, Nicola Carboni

Version: 1.0

Introduction

The person reference data model provides a list of standard descriptors (fields) that are typically present in the description of a person in cultural heritage data systems. The intention of this reference model is to provide a consolidated, high-level formal structure comprising the most commonly reused descriptors for a person entity and, further, to provide for these a set of semantic mapping to the CIDOC CRM. Moreover, each field is marked with respect to its potential functionality with regards to instance matching between overlapping datasets. This reference data model aims to serve a number of functions including:

1. to support the creation of semantic data entry forms that will create ‘born semantic’ data streams
2. to serve as a reference guide in designing eventually compatible data structures for feeding into larger aggregations of Cultural Heritage data
3. to guide mapping processes of extant data sources with common mapping patterns
4. to be adopted by institutions acting as aggregation hubs in order to create consistent re-expressions of extant reference information in a common CIDOC CRM-based paradigm
SRDM: Aims / Successes

- Gives a simple presentation of reference object in form understandable to domain expert but creating semantics for robustness of data
- Creates a guide for consistent application of semantics
  - To create new data OR
  - Map old data
- Continue to manage your data in its source form
- Solves the simple semantic problems consistently, so you can focus on the tricky ones

Ok, great, but...
New Challenges for Semantic Documentation

The success of the idea of Semantic Reference Data Models opens new challenges:

- Extension of models
- Revision/versioning of models
- Efficiency and Accuracy of Operation
- Full usability by all user types
- Adoption and proposals by other projects and broader community
- Attribution of authorship/provenance/funding support
- Authority
So... Zellīj
Zellīj: what, more specifically, is it?

A database and workflow that allows the documentation of semantic patterns:

- At different levels of granularity
- In a compositional manner
- With researcher appropriate labels and descriptions
- With developer appropriate description of the semantic pattern rdf notation
- With attribution of creation, funding, use
Who is it for?

Domain Specialist  Software Developer / Conceptual Modeller  Systems Implementer  Machines
Why are we doing it?

Building:

- Share Semantics Knowledge
- Increase Efficiency of Semantics Use
- Build Semantics Community
- Get benefits of Semantics without the Problems
- Create Sustainability

Towards:

- Share Research Data / Support New Knowledge Creation
System Software

Backend

Frontend
What does database consist of?

**Fields**: a named, ided, and intensionally defined, syntactically unique path from a domain node to a range node through $n$ triples, representing a semantically unique statement type, intended to represent one value/data point

**Collections**: a named, ided, and intensionally defined grouping of fields composing a pattern of mutually relevant statement types related to the same, intermediate subject (domain node), documented for the purpose of common reuse across entities in a semantic data project

**Models**: a named, ided and intensionally defined grouping of fields composing a pattern of mutually relevant statement types related to the same, central subject (domain node), documented for the purpose of common reuse across semantic data projects

**Category**: a named cluster of fields within a model grouped together according to their likelihood/capacity to respond to a similar field of inquiry
<table>
<thead>
<tr>
<th>Identifier</th>
<th>Field ID</th>
<th>Name</th>
<th>Description</th>
<th>Ontological Scope</th>
<th>CRM Path</th>
<th>Total_Turtle</th>
<th>Expected Value Type</th>
<th>Expected Resource Model</th>
<th>Expected Collection Model</th>
<th>Concept - Specified</th>
<th>Version</th>
</tr>
</thead>
<tbody>
<tr>
<td>FIE.1</td>
<td>00dba278</td>
<td>Identifier</td>
<td>This field is used to record an identifier attributed to the documented entity.</td>
<td>E1 CRM Entity</td>
<td>--&gt; P1 --&gt; E42[1_1] --&gt; P190 --&gt; rdf:literal</td>
<td><a href="http://www.srdm.org/object/E1">http://www.srdm.org/object/E1</a> a <a href="http://www.cidoc-crm.org/cidoc-crm/E1_CRM_Entity">http://www.cidoc-crm.org/cidoc-crm/E1_CRM_Entity</a>; crm:P1_is_identified_by <a href="https://www.srdm.org/name/fie_1_1">https://www.srdm.org/name/fie_1_1</a> . <a href="https://www.srdm.org/name/fie_1_1">https://www.srdm.org/name/fie_1_1</a> a crm:E42_Identifier; crm:P190_has_symbolic_content &quot;content&quot; .</td>
<td>String</td>
<td></td>
<td></td>
<td></td>
<td>1</td>
</tr>
</tbody>
</table>

**Diagram:**

1. **http://www.srdm.org/object/E1**
   - rdf:type
   - crm:P1_is_identified_by
     - https://www.srdm.org/name/fie_1_1
   - crm:P190_has_symbolic_content
     - "content"
     - rdf:type
     - crm:E42_Identifier

2. **crm:E1_CRM_Entity**
   - crm:P1_is_identified_by
     - crm:E42_Identifier
     - rdfs:Literal
<table>
<thead>
<tr>
<th>Identifier</th>
<th>uuid</th>
<th>Name</th>
<th>Description</th>
<th>Ontological Scope</th>
<th>Collection_Fields</th>
<th>Authors</th>
<th>Funders</th>
<th>Users</th>
</tr>
</thead>
<tbody>
<tr>
<td>COL.1</td>
<td>1ac1aa</td>
<td>ID Attribution</td>
<td>A series of attributes used to indicate an identifier associated to a document object, its attributes and the provenance thereof.</td>
<td>E1 CRM Entity</td>
<td>[COL.1] ID Attribution,[FIE.1] Identifier,[COL.1] ID Attribution,[FIE.2] Identifier Type,[COL.1] ID Attribution,[FIE.3] Identifier Provider,[COL.1] ID Attribution,[FIE.4] Identifier Source</td>
<td>George Bruseker,Nicola Carboni</td>
<td>Swiss Art Research Institute</td>
<td>Swiss Art Research Institute</td>
</tr>
</tbody>
</table>

```xml
@prefix crm: <http://www.cidoc-crm.org/cidoc-crm/> .

  crm:P1_is_identified_by <https://www.srdm.org/name/fie_1_1> .

<https://www.srdm.org/name/fie_1_1> a crm:E42_Identifier ;
  crm:P190_has_symbolic_content "content" ;
  crm:P2_has_type <http://www.srdm.org/type/fie_2_1> ;
  crm:P37i_was_assigned_by <https://www.srdm.org/event/fie_3_1> .

<http://www.srdm.org/type/fie_2_1> a ns0:E55_Type .

<https://www.srdm.org/event/fie_3_1> a crm:E15_Identifier_Assignment ;
  crm:P14_carried_out_by <https://www.srdm.org/actor/fie_3_2> ;
  crm:P16_used_specific_object <https://www.srdm.org/conceptual_object/fie_4_1> .

<https://www.srdm.org/actor/fie_3_2> a ns0:E39_Actor .

<https://www.srdm.org/conceptual_object/fie_4_1> a ns0:E73_Information_Object .
```
MOD.1  Person

The person reference data model provides a list of standard descriptors (fields) that are typically present in the description of a person in cultural heritage data systems. The intention of this reference model is to provide a consolidated, high-level formal structure comprising the most commonly reused descriptors for a person entity and, further, to provide for these a set of semantic mapping to the CIDOC CRM. Moreover, each field is marked with respect to its potential functionality with regards to instance matching between overlapping datasets.

**Table:**

<table>
<thead>
<tr>
<th>Identification</th>
<th>Name</th>
<th>Description</th>
<th>Ontological Scope</th>
<th>Model_Fields</th>
<th>Authors</th>
<th>Funders</th>
<th>Users</th>
<th>Version</th>
<th>Status</th>
</tr>
</thead>
<tbody>
<tr>
<td>MOD.1</td>
<td>Person</td>
<td>The person reference data model provides a list of standard descriptors (fields) that are typically present in the description of a person in cultural heritage data systems. The intention of this reference model is to provide a consolidated, high-level formal structure comprising the most commonly reused descriptors for a person entity and, further, to provide for these a set of semantic mapping to the CIDOC CRM. Moreover, each field is marked with respect to its potential functionality with regards to instance matching between overlapping datasets.</td>
<td>E21 Person</td>
<td>[MOD.1] Person [FIE.1] Identifier,[MOD.1] Person [FIE.2] Identifier Type,[MOD.1] Person [FIE.3] Identifier Provider,[MOD.1] Person [FIE.4] Identifier Source,[MOD.1] Person [FIE.5] Name,[MOD.1] Person [FIE.6] Name Language,[MOD.1] Person [FIE.7] Name Part,[MOD.1] Person [FIE.8] Name Part Type,[MOD.1] Person [FIE.10] Alternative Name,[MOD.1] Person [FIE.11] Alternative Name Type,[MOD.1] Person [FIE.12] Alternative Name Language,[MOD.1] Person [FIE.13] Alternative Name Use Period - ...</td>
<td>George Bruseker, Nicola Carboni</td>
<td>Swiss Art Research Infrastructure</td>
<td>Swiss Art Research Infrastructure</td>
<td>1.0</td>
<td>Released</td>
</tr>
</tbody>
</table>

...
## Automated Documentation

<table>
<thead>
<tr>
<th>Identifier</th>
<th>Name</th>
<th>Function / Description</th>
<th>CRM Path</th>
</tr>
</thead>
<tbody>
<tr>
<td>[The_188]</td>
<td>&quot;Group Description&quot;</td>
<td>&quot;This field is used to record a description in free text of the documented group.&quot;</td>
<td>[(\rightarrow P129) (\rightarrow E33(188,1)) (\rightarrow P190) (\rightarrow rdfs:Literal)]</td>
</tr>
<tr>
<td>[The_190]</td>
<td>&quot;Group Description Type&quot;</td>
<td>&quot;This field is used to record the type of description given of the documented group.&quot;</td>
<td>[(\rightarrow P129) (\rightarrow E33(188,1)) (\rightarrow P2) (\rightarrow E55(190,1))]</td>
</tr>
<tr>
<td>[The_191]</td>
<td>&quot;Group Description Author&quot;</td>
<td>&quot;This field is used to record the author of the description given of the documented group.&quot;</td>
<td>[(\rightarrow P129) (\rightarrow E33(188,1)) (\rightarrow P94i) (\rightarrow E65(191,1)) (\rightarrow P14) (\rightarrow E39(191,2))]</td>
</tr>
<tr>
<td>[The_192]</td>
<td>&quot;Group Description Earliest Date&quot;</td>
<td>&quot;This field is used to record the earliest possible date for the creation of the description of the documented group.&quot;</td>
<td>[(\rightarrow P129) (\rightarrow E33(188,1)) (\rightarrow P94i) (\rightarrow E65(191,1)) (\rightarrow P4) (\rightarrow E52(191,2)) (\rightarrow P82a) (\rightarrow) (xsd:dateTime)]</td>
</tr>
<tr>
<td>[The_193]</td>
<td>&quot;Group Description Latest Date&quot;</td>
<td>&quot;This field is used to record the latest possible date for the creation of the description of the documented group.&quot;</td>
<td>[(\rightarrow P129) (\rightarrow E33(188,1)) (\rightarrow P94i) (\rightarrow E65(191,1)) (\rightarrow P4) (\rightarrow E52(191,2)) (\rightarrow P82b) (\rightarrow) (xsd:dateTime)]</td>
</tr>
</tbody>
</table>

Generated by ZeSS, © 2022.
SARI USE Case

**Fields:** 202 distinct base patterns

**Collections:** 43 distinct grouping patterns

**Models:** 8 fully documented standard semantic reference data models
## Fields Breakdown

<table>
<thead>
<tr>
<th>Ontological Scope</th>
<th>Count</th>
<th>Grouping</th>
</tr>
</thead>
<tbody>
<tr>
<td>E21 Person</td>
<td>21</td>
<td>Actor</td>
</tr>
<tr>
<td>E74 Group</td>
<td>15</td>
<td>Actor</td>
</tr>
<tr>
<td>E39 Actor</td>
<td>17</td>
<td>Actor</td>
</tr>
<tr>
<td>D1 Digital Object</td>
<td>13</td>
<td>Conceptual Object</td>
</tr>
<tr>
<td>F24 Publication Expression</td>
<td>11</td>
<td>Conceptual Object</td>
</tr>
<tr>
<td>E89 Propositional Object</td>
<td>3</td>
<td>Conceptual Object</td>
</tr>
<tr>
<td>E2 Temporal Entity</td>
<td>11</td>
<td>Event</td>
</tr>
<tr>
<td>E7 Activity</td>
<td>9</td>
<td>Event</td>
</tr>
<tr>
<td>E4 Period</td>
<td>3</td>
<td>Event</td>
</tr>
<tr>
<td>E1 CRM Entity</td>
<td>38</td>
<td>General</td>
</tr>
<tr>
<td>E72 Legal Object</td>
<td>3</td>
<td>General</td>
</tr>
<tr>
<td>E18 Physical Thing</td>
<td>20</td>
<td>Physical Thing</td>
</tr>
<tr>
<td>E24 Physical Man-Made Thing</td>
<td>24</td>
<td>Physical Thing</td>
</tr>
<tr>
<td>E53 Place</td>
<td>14</td>
<td>Place</td>
</tr>
</tbody>
</table>

### Count

- **Place**: 6.9%
- **Actor**: 26.2%
- **Conceptual Object**: 13.4%
- **Physical Thing**: 21.8%
- **General**: 20.3%
- **Event**: 11.4%
## Collections Breakdown

<table>
<thead>
<tr>
<th>Ontological Scope</th>
<th>Count</th>
<th>Grouping</th>
</tr>
</thead>
<tbody>
<tr>
<td>E21 Person</td>
<td>4</td>
<td>Actor</td>
</tr>
<tr>
<td>E74 Group</td>
<td>3</td>
<td>Actor</td>
</tr>
<tr>
<td>E39 Actor</td>
<td>4</td>
<td>Actor</td>
</tr>
<tr>
<td>D1 Digital Object</td>
<td>3</td>
<td>Conceptual Object</td>
</tr>
<tr>
<td>F24 Publication</td>
<td>2</td>
<td>Conceptual Object</td>
</tr>
<tr>
<td>E2 Temporal Entity</td>
<td>4</td>
<td>Event</td>
</tr>
<tr>
<td>E7 Activity</td>
<td>1</td>
<td>Event</td>
</tr>
<tr>
<td>E4 Period</td>
<td>1</td>
<td>Event</td>
</tr>
<tr>
<td>E1 CRM Entity</td>
<td>6</td>
<td>General</td>
</tr>
<tr>
<td>E72 Legal Object</td>
<td>1</td>
<td>General</td>
</tr>
<tr>
<td>E24 Physical</td>
<td>5</td>
<td>Physical Thing</td>
</tr>
<tr>
<td>Man-Made Thing</td>
<td></td>
<td></td>
</tr>
<tr>
<td>E18 Physical Thing</td>
<td>3</td>
<td>Physical Thing</td>
</tr>
<tr>
<td>E53 Place</td>
<td>2</td>
<td>Place</td>
</tr>
</tbody>
</table>

### Count

- **Place**: 5.1%
- **Actor**: 28.2%
- **Physical Thing**: 20.5%
- **Conceptual Object**: 12.8%
- **General**: 17.9%
- **Event**: 15.4%
High level Documentation Architecture

SARI inherits

Pharos  Hertziana Semantic  Semantic Census
High level Documentation Architecture

SARI Models

Pharos

Pharos

Census

Census

inherits

curates

SARI?
What have we achieved?

- A data model for a reusable database of patterns
- A database of basic, compositional patterns (fields, collections)
- A database with a set of standard models built compositionally (8)
- A workflow for generating consistent documentation
What is there still to do?

- Make whole platform open source
- Extend with Canned Sparql
- Extend with SHACL
- Extend with auto build out of system templates for Arches / Research Space / Wisski etc.
- Connect to Ontology production systems like OntoMe
THANK YOU