Application Profiles since DC-2000

Thomas Baker
Dublin Core Metadata Initiative

Taming the Graph: Profiles over Linked Data
DC-2017, Washington-DC
27 October 2017
This talk

• “Application Profiles”
  – Single biggest topic of discussion and research in Dublin Core community since 1990s!

• This talk includes
  – Ideas that did catch on
  – Ideas that did not catch on
  – Topics that are still unresolved

• Topics covered later today are out of scope:
  – BIBFRAME profiles
  – ShEx, SHACL, JSON-LD...
From Record Format to RDF Data

• **1995** Dublin Core as “simple metadata record”
• **1999** Dublin Core as an RDF vocabulary
• **2000** Application Profiles (“mix-and-match”)
• **2003-2008** Description Set Profiles
  – Validatable record formats
  – Straightforwardly mappable to RDF triples
Metadata in the 2000s

• From *Static Record Formats*
  – Bounded sets of fields to be “filled in” with information

• Towards *Recombinant Graph Data*
  – Unbounded webs of information connected by statements
Bridging two mindsets

• Oriented to *Record Formats*
  – Bounded sets of fields to be “filled in” with information

• Oriented to *Graphs*
  – Unbounded webs of information connected by statements
Let’s take a graph...
...view it in terms of a “record”...
...focusing on components that can be validated...
..and abstract those components to a generalized model.
Broad humor and bitter irony collide in this fictional autobiographical novel of Rabo Karabekian, who, at age seventy-one, wants to be left alone on his Long Island estate with the secret he has locked inside his potato barn. But then a voluptuous young widow badgers Rabo into telling his life story—and Vonnegut in turn tells us the plain, heart-hammering truth about man’s careless fancy to create or destroy what he loves.
Broad humor and bitter irony collide in this fictional autobiography of Rabo Karabekian, who, at age seventy-one, wants to be left alone on his Long Island estate with the secret he has locked inside his potato barn. But then a voluptuous young widow badgers Rabo into telling his life story—and Vonnegut in turn tells us the plain, heart-hammering truth about man's careless fancy to create or destroy what he loves.
2007: Singapore Framework (revised)
Design a profile on specific requirements
Your model of “things in the world”

Application Profile

- Functional Requirements
- Domain Model
- Description Set Profile
- Data Format
- Usage Guidelines
  - annotate

Domain Standards

- Community Domain Model
- RDF Vocabularies
- Constraint Language
- Compatible Syntaxes

Foundation Standards

- RDFS
- RDF
Construct a profile using RDF vocabularies

Application Profile

- Functional Requirements
- Domain Model
- Description Set Profile
- Data Format

Domain Standards

- Community Domain Model
- RDF Vocabularies
- Constraint Language
- Compatible Syntaxes

Foundation Standards

- RDFS
- RDF

annotate
With a language for expressing constraints

Application Profile

- Functional Requirements
- Domain Model
- Description Set Profile
- Data Format
- Usage Guidelines
- compatible Syntaxes

Domain Standards

- Community Domain Model
- RDF Vocabularies
- Constraint Language

Foundation Standards

- RDFS
- RDF

RDF Vocabularies

RDF Vocabularies

annotate
Use compatible concrete syntaxes

- Application Profile
  - Functional Requirements
  - Domain Model
  - Description Set Profile
  - Data Format
  - Usage Guidelines
  - annotate

- Domain Standards
  - Community Domain Model
  - RDF vocabularies
  - Constraint Language
  - Compatible Syntaxes

- Foundation Standards
  - RDFS
  - RDF
Graph data is based on RDF vocabularies.
Record formats are documents

<?xml version="1.0" ?>

- <metadata>
  - <idinfo>
    - <citation>
      - <citeinfo>
        <origin>Canadian Ice Service (CIS)</origin>
        <pubdate>20080324</pubdate>
      </citeinfo>
    </citation>
  - <abstract>The SIGRID-3 vector archive format is one of the World Meteorological Organization (WMO) standards for archiving digital ice charts in the Global Digital Sea Ice Data Bank (GDSIDB). The WMO ice chart archiving formats are the Sea Ice Grid (SIGRID) format developed in 1981 and formalized in 1989 and its successor SIGRID-2. The Canadian Ice Service digital Ice Analysis charts (Regional, Daily and Image) are encoded in SIGRID-3 and have two main components: the shapefile containing the Ice Analysis ice information (ice polygons and related attributes) and the metadata describing the Ice Analysis data under the SIGRID-3 format. The
Thousands of record formats...
Good profiles make good graph data

Profile A
Profile B
Profile C
Good triples can be merged coherently.
Good triples can be merged coherently
Profiles get lost or forgotten...

Profile A

Profile B

Profile C
The data remains
Meaning is based on published vocabularies
2006: Extracting triples

- Profiles as targets for “triplifying” non-RDF data
- Representing non-RDF data in RDF for the purposes of merging
<?xml version="1.0" encoding="UTF-8" ?>
<rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
   xmlns:dcterms="http://purl.org/dc/terms/"
   xmlns:foaf="http://xmlns.com/foaf/0.1/" >
  <rdf:Description rdf:about="http://agris.fao.org/resource/CH2001000179">
    <dcterms:title>Heuschrecken brauchen ökologische Ausgleichsflächen</dcterms:title>
    <dcterms:creator rdf:nodeID="PB" />
  </rdf:Description>
<rdf:Description rdf:nodeID="PB">
  <foaf:name>Peter, B.</foaf:name>
</rdf:Description>
</rdf:RDF>

---

### Express as triples

<table>
<thead>
<tr>
<th>Subject</th>
<th>Predicate</th>
<th>Object</th>
</tr>
</thead>
<tbody>
<tr>
<td>agris:CD2001000179</td>
<td>dct:subject</td>
<td>agrovoc:c_4416k</td>
</tr>
<tr>
<td>agris:CD2001000179</td>
<td>dct:title</td>
<td>&quot;Heuschrecken...&quot;@de</td>
</tr>
<tr>
<td>agris:CD2001000179</td>
<td>dct:creator</td>
<td>:PB</td>
</tr>
<tr>
<td>:PBS</td>
<td>foaf:name</td>
<td>&quot;Peter, B.&quot;</td>
</tr>
</tbody>
</table>

---

Abstract Model components embedded in application syntaxes
Heuschrecken brauchen ökologische Ausgleichsflächen

Subject | Predicate | Object
--- | --- | ---
agris:CD2001000179 | dct:title | "Heuschrecken..."@de
agris:CD2001000179 | dct:creator | :PB
:PBS | foaf:name | "Peter, B."
2008: Embed DSPs in wiki documents?

http://dublincore.org/documents/2008/10/06/dsp-wiki-syntax/DescriptionSetProfile-dist.zip
...in machine-readable syntax

http://dublincore.org/documents/2008/10/06/dsp-wiki-syntax/DescriptionSetProfile-dist.zip
...and extract to configure an editor?

http://dublincore.org/documents/2008/10/06/dsp-wiki-syntax/DescriptionSetProfile-dist.zip
Profile Requirements

• Authored in an idiom usable by normal people
• Constraints reusable for data automatic validation
Profile Use Cases

• Community consensus on metadata models
• Validating metadata
• Constructive feedback to implementers
• Building user interfaces
• Many types of profile?