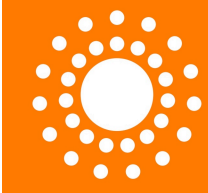




A DCMI constraint language for Description Set Profiles

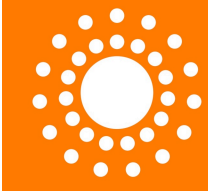
Thomas Baker
Dublin Core Metadata Initiative

2013-09-11: RDF Validation Workshop, Cambridge, MA



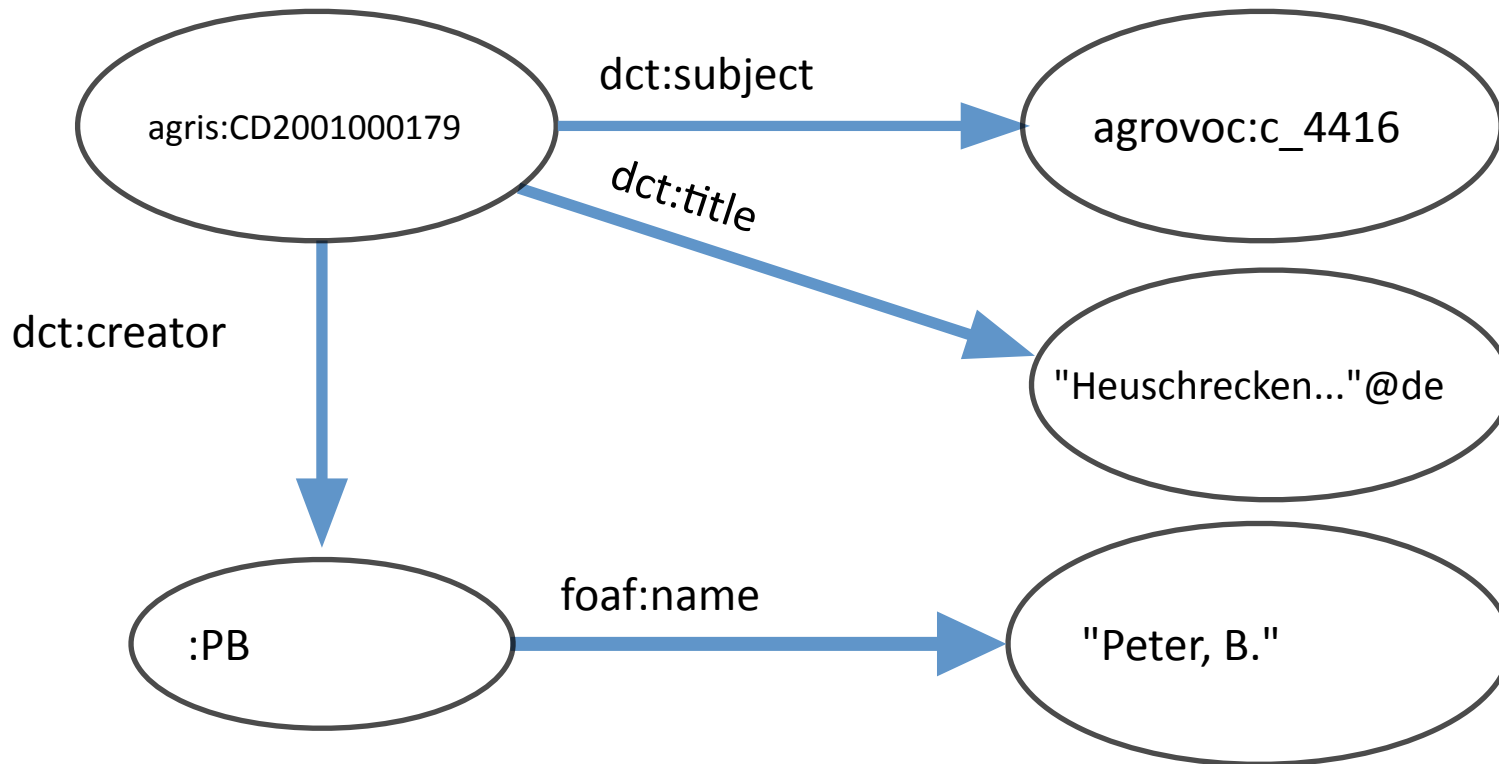
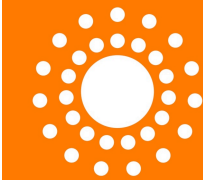
From Record Format to RDF vocabulary

- **1995** Dublin Core as “simple metadata record”
- **1999** Dublin Core as an RDF vocabulary
- **2000** “Application profiles” mix-and-match
- **2003-2007** DCMI Abstract Model
 - Basis for templating validatable record formats
 - Formats straightforwardly mappable to RDF triples

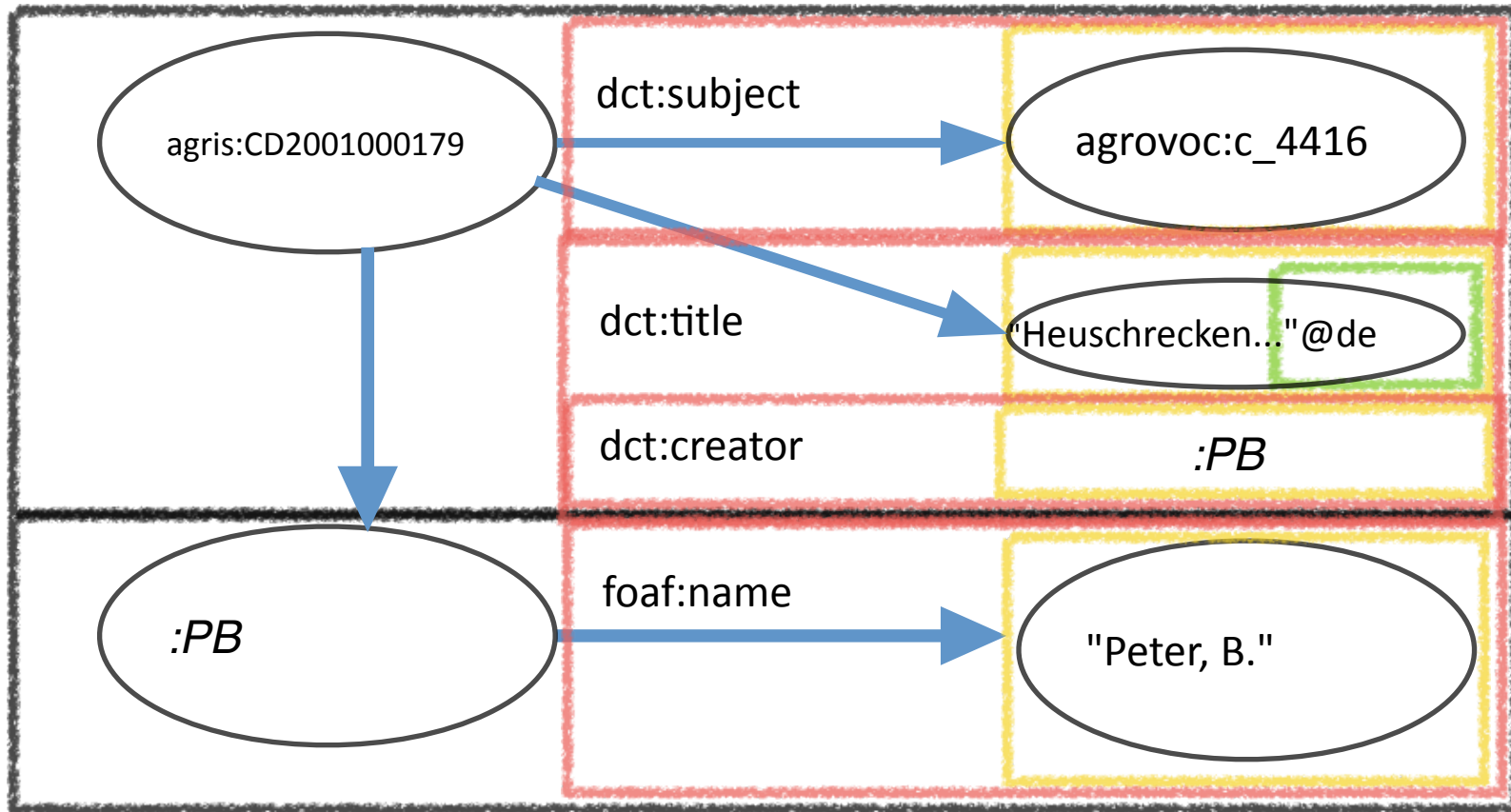
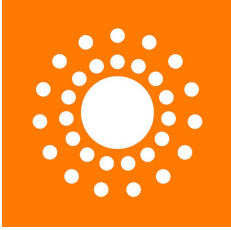


Bridging two mindsets

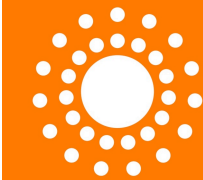
- Oriented to ***Record Formats***
 - Bounded sets of fields to be “filled in” with information
- Oriented to ***Graphs***
 - Unbounded webs of information connected by statements



Let's take a graph...



...view it in terms of a “record”...



agris:CD2001000179

dct:subject

agrovoc:c_4416

dct:title

"Heuschrecken"

de

dct:creator

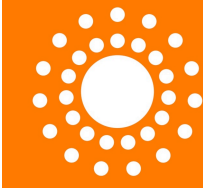
:PB

:PB

foaf:name

"Peter, B."

...focusing on components that can be validated...



Description Set

Description

Described Resource URI

Property URI

Value URI

Property URI

Value String

Lang

Property URI

Value ID

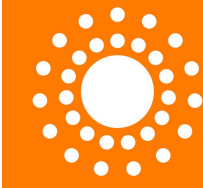
Description

Value ID

Property URI

Value String

..and abstract those components to a generalized model.



Description Set ←

Description ←

Non-literal ←

Property URI

Value URI

Value String

Lang

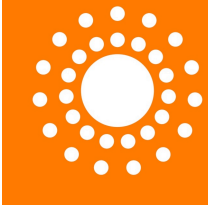
Vocabulary Encoding
Scheme URI

Literal ←

Value String

Lang

...using grouping constructs with no equivalent in RDF.



Property URI

Described Resource URI

```
<?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:subject rdf:resource="http://aims.fao.org/aos/agrovoc/c_4416" />
    <dcterms:creator rdf:nodeID="PB" />
  </rdf:Description>
  <rdf:Description rdf:nodeID="PB">
    <foaf:name>Peter, </my:name>
  </rdf:Description>
</rdf:RDF>
```

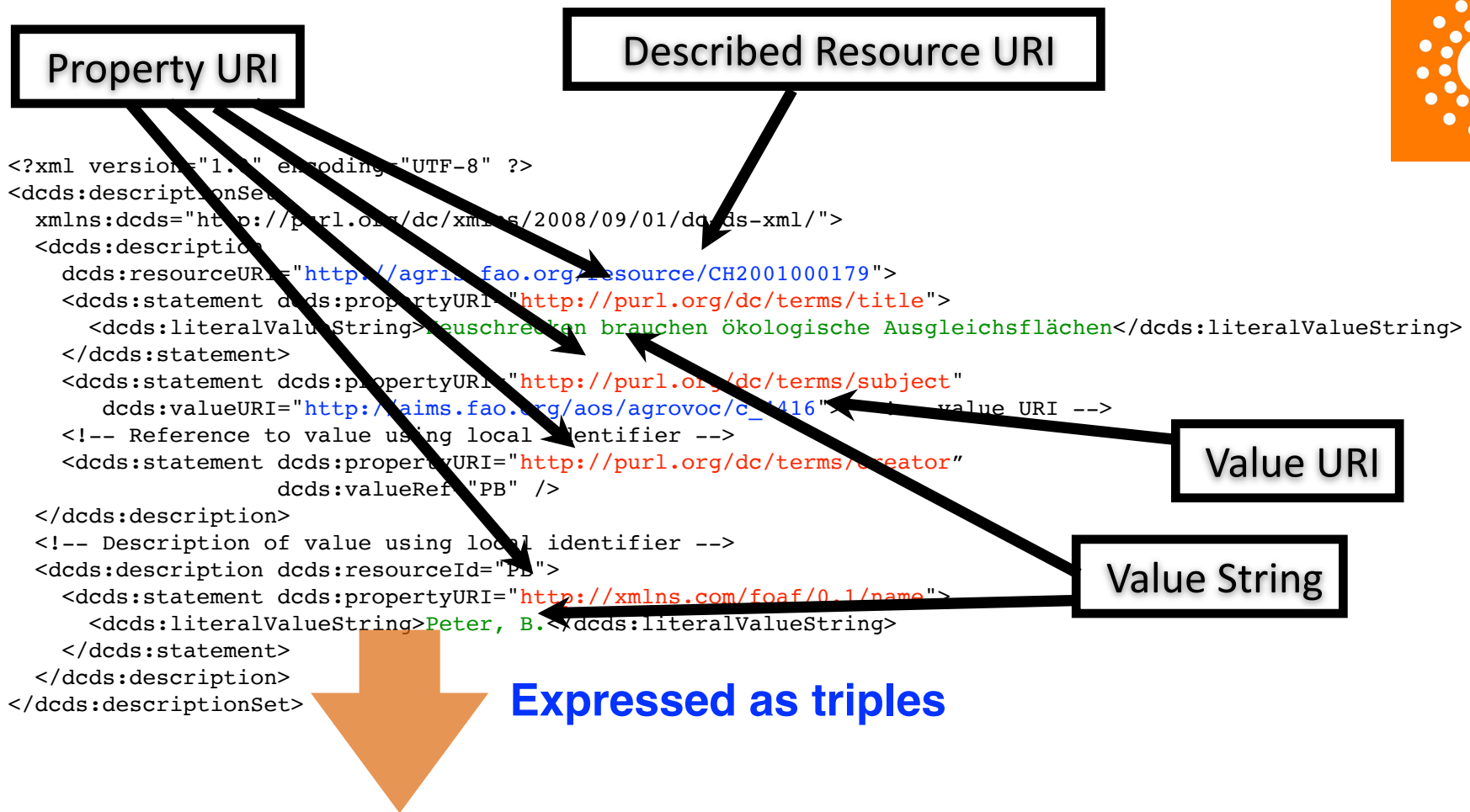
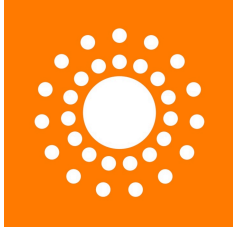
Value URI

Value String

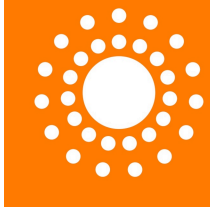
**Abstract Model
components embedded
in application syntaxes**

Expressed as triples

| Subject | Predicate | Object |
|--------------------|-------------|----------------------|
| agris:CD2001000179 | dct:subject | agrovoc:c_4416k |
| agris:CD2001000179 | dct:title | "Heuschrecken..."@de |
| agris:CD2001000179 | dct:creator | :PB |
| :PBS | foaf:name | "Peter, B." |

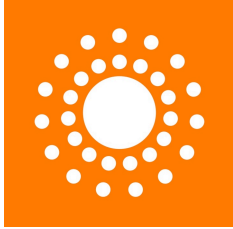


| Subject | Predicate | Object |
|--------------------|-------------|----------------------|
| agris:CD2001000179 | dct:subject | agrovoc:c_4416k |
| agris:CD2001000179 | dct:title | "Heuschrecken..."@de |
| agris:CD2001000179 | dct:creator | :PB |
| :PBS | foaf:name | "Peter, B." |



2008 *draft* Description Set Profile Constraint Language

- Constraints commonly required for validation:
 - *minOccurs/maxOccurs*
 - *resourceClass* (what is the Description about?)
 - *standalone* (describe Person without describing Book?)
 - *valueURI* (mandate a particular controlled vocabulary)



Templates for Description Sets

Constraints on Templates

```
Description Set [template]
  Description [template]
    Statement [template]
      Property [constraint]
        <http://purl.org/dc/terms/subject>
      VocabularyEncodingSchemeURI [constraint]
        <http://aims.fao.org/aos/agrovoc>

    Statement [template]
      Property [constraint]
        <http://purl.org/dc/terms/title>
      MinOccurs [constraint]
        1
      MaxOccurs [constraint]
        1

    Statement [template]
      Property [constraint]
        <http://purl.org/dc/terms/creator>

  Description [template]
    Resource Class [constraint]
      <http://xmlns.com/foaf/0.1/Person>
    Statement [template]
      Property [constraint]
        <http://xmlns.com/foaf/0.1/name>
```

- Data using this Description Set Profile describes:
 - a Resource with *exactly one* [DC] *title*,
 - the [DC] *subject* of which is a *URI from AGROVOC*.
 - Authors, i.e. members of the [FOAF] *class Person* that have [FOAF] *names*.



Templates and Constraints

- **Description Template** constraints: Identifier, Standalone, Minimum occurrence, Maximum occurrence, Resource class membership
- **Statement Template** constraints: Minimum occurrence, Maximum occurrence, Type
 - Property constraints: Property list, Sub-property
 - **Literal Statement Template** constraints:
 - Literal value constraints: Literal list, Literal language, Literal language list, Syntax Encoding Scheme, Syntax Encoding Scheme list
 - **Non-literal Statement Template** constraints:
 - Non-literal value constraints: Description template reference, Class membership
 - Value URI constraints: Value URI occurrence, Value URI list
 - Vocabulary encoding scheme (VES) constraints: VES occurrence, VES list
 - Value string constraints: Minimum occurrence, Maximum occurrence (plus all other constraints that apply to literal values - see above)

Record

Resource A

```
{
  "items" : [
    {
      "id": "http://bibfra.me/work/u2",
      "title" : "Bluebeard",
      "titleRemainder" : "a novel",
      "type" : "Book"
      "classification" : "025.3 M465f 2008",
      "author" : "http://bibfra.me/auth/person/kurt\_vonnegut",
      "subject": [ "http://bibfra.me/auth/subject/foo",
                   "http://bibfra.me/auth/subject/bar" ],
      "language": "english",
      "summary": "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.",
      "hasInstance": "http://bibfra.me/instance/u2-1"
    },
    {
      "id": "http://bibfra.me/instance/u2-1",
      "type": "HardcoverBook",
      "edition": "1st",
      "volumes": "1",
      "pages": "300",
      "date" : "1987-09-02",
      "publisher" : "http://bibfra.me/auth/org/delacorte\_press",
      "ISBN": "9780385295901",
      "image": "http://ecx.images-amazon.com/images/I/51vbo8zwJlL.\_SL160\_.jpg"
    }
  ]
}
```

Resource B

```
},
{
  "id": "http://bibfra.me/instance/u2-1",
  "type": "HardcoverBook",
  "edition": "1st",
  "volumes": "1",
  "pages": "300",
  "date" : "1987-09-02",
  "publisher" : "http://bibfra.me/auth/org/delacorte\_press",
  "ISBN": "9780385295901",
  "image": "http://ecx.images-amazon.com/images/I/51vbo8zwJlL.\_SL160\_.jpg"
}
]
```

Description Set

Templating Descriptions

```
"items" : [
  {
```

```
    "id": "http://bibfra.me/work/u2",
    "title" : "Bluebeard",
    "titleRemainder" : "a novel",
    "type" : "Book"
    "classification" : "025.3 M465f 2008",
    "author" : "http://bibfra.me/auth/person/kurt_vonnegut",
    "subject": [ "http://bibfra.me/auth/subject/foo",
                  "http://bibfra.me/auth/subject/bar" ],
    "language": "english",
    "summary": "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.",
```

| | | |
|----------------|--------|---|
| id | URI | "must be http://bibfra.me/... " |
| title | String | "must be one" |
| titleRemainder | String | |
| type | String | from controlled vocabulary |
| classification | String | |
| author | URI | |
| subject | URI | |
| lang | String | |
| summary | String | |
| hasInstance | URI | "one and only one" |

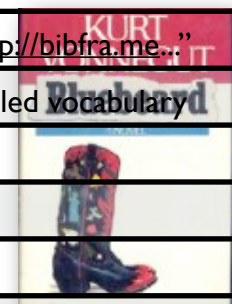
"hasInstance": "<http://bibfra.me/instance/u2-1>"

```
},
```

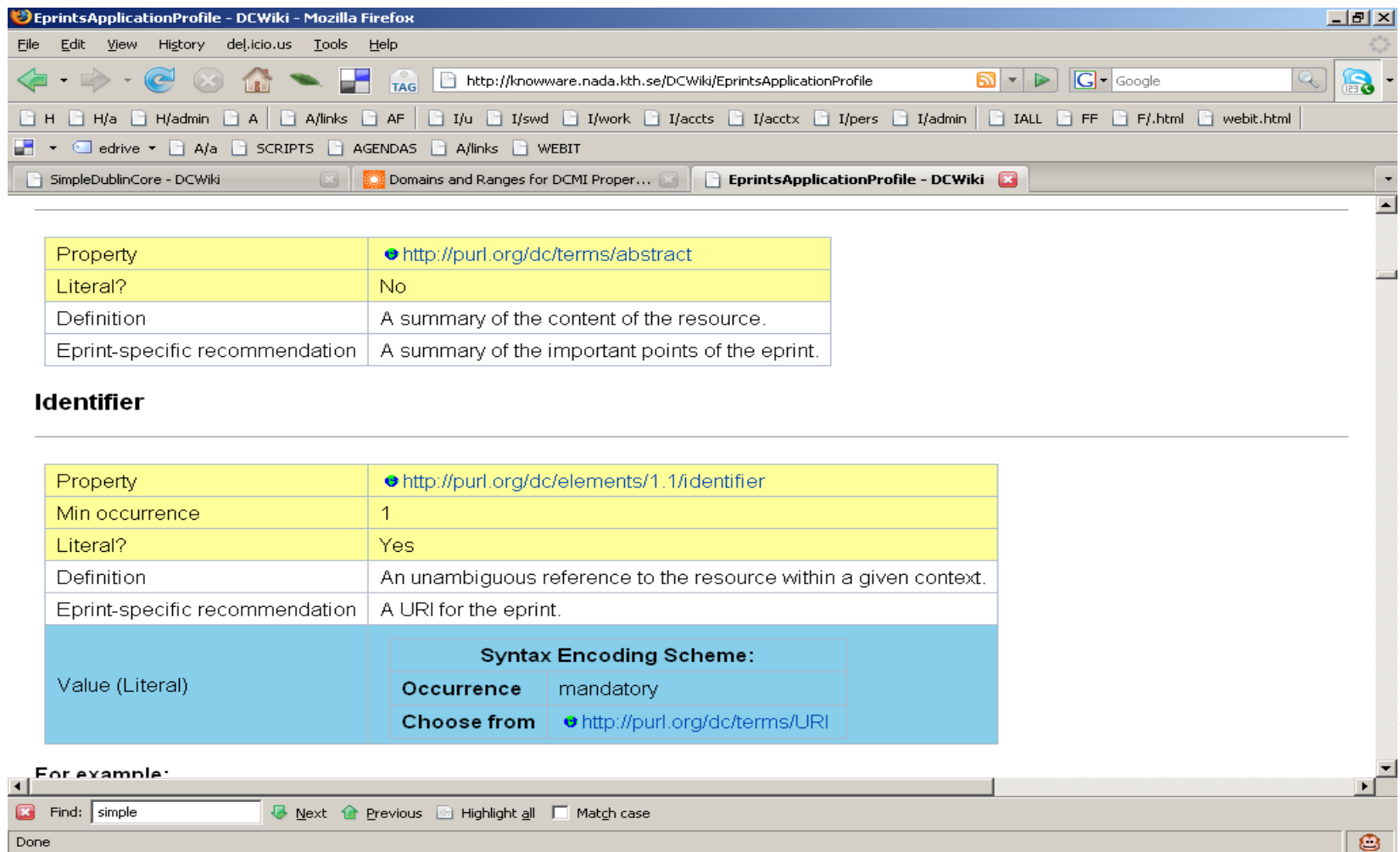
```
{
```

```
    "id": "http://bibfra.me/instance/u2-1",
    "type": "HardcoverBook",
    "edition": "1st",
    "volumes": "1",
    "pages": "300",
    "date" : "1987-09-02",
    "publisher" : "http://bibfra.me/auth/org/delacorte_press",
    "ISBN": "9780385295901",
    "image": "http://ecx.images-amazon.com/images/I/51vbo8zwJlL_SL160 .jpg"
```

| | | |
|-----------|--------|---|
| id | URI | "must be http://bibfra.me/... " |
| type | String | from controlled vocabulary |
| edition | String | |
| volumes | String | integer |
| pages | String | integer |
| date | Date | use Bibframe date-time format |
| publisher | URI | |
| ISBN | String | must be valid ISBN |
| image | URI | |



Show the user readable/editable document...



EprintsApplicationProfile - DCWiki - Mozilla Firefox

File Edit View History del.icio.us Tools Help

http://knowware.nada.kth.se/DCWiki/EprintsApplicationProfile

Google

H H/a H/admin A A/links AF I/u I/swd I/work I/accts I/acctx I/pers I/admin IALL FF F/.html webit.html

edrive A/a SCRIPTS AGENDAS A/links WEBIT

SimpleDublinCore - DCWiki Domains and Ranges for DCMI Proper... EprintsApplicationProfile - DCWiki

| | |
|--------------------------------|---|
| Property | http://purl.org/dc/terms/abstract |
| Literal? | No |
| Definition | A summary of the content of the resource. |
| Eprint-specific recommendation | A summary of the important points of the eprint. |

Identifier

| | | | | | | | |
|--------------------------------|--|-------------------------|--|------------|-----------|-------------|---|
| Property | http://purl.org/dc/elements/1.1/identifier | | | | | | |
| Min occurrence | 1 | | | | | | |
| Literal? | Yes | | | | | | |
| Definition | An unambiguous reference to the resource within a given context. | | | | | | |
| Eprint-specific recommendation | A URI for the eprint. | | | | | | |
| Value (Literal) | <table border="1"><tr><td colspan="2">Syntax Encoding Scheme:</td></tr><tr><td>Occurrence</td><td>mandatory</td></tr><tr><td>Choose from</td><td>http://purl.org/dc/terms/URI</td></tr></table> | Syntax Encoding Scheme: | | Occurrence | mandatory | Choose from | http://purl.org/dc/terms/URI |
| Syntax Encoding Scheme: | | | | | | | |
| Occurrence | mandatory | | | | | | |
| Choose from | http://purl.org/dc/terms/URI | | | | | | |

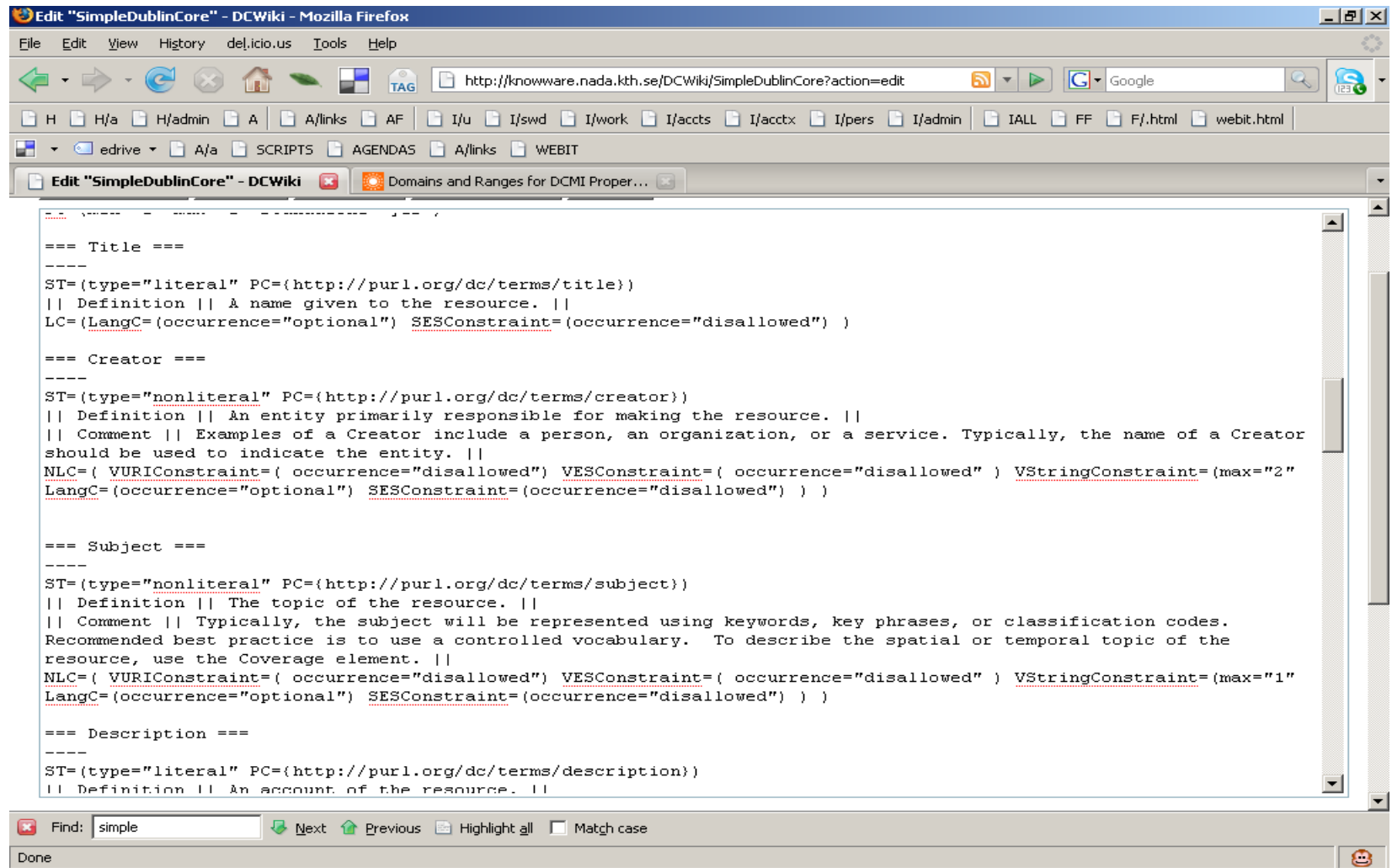
For example:

Find: simple Next Previous Highlight all Match case

Done

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

...with embedded constraints...



```
==== Title ====
----
ST=(type="literal" PC={http://purl.org/dc/terms/title})
|| Definition || A name given to the resource. ||
LC=(LangC={occurrence="optional"} SESConstraint={occurrence="disallowed"} )

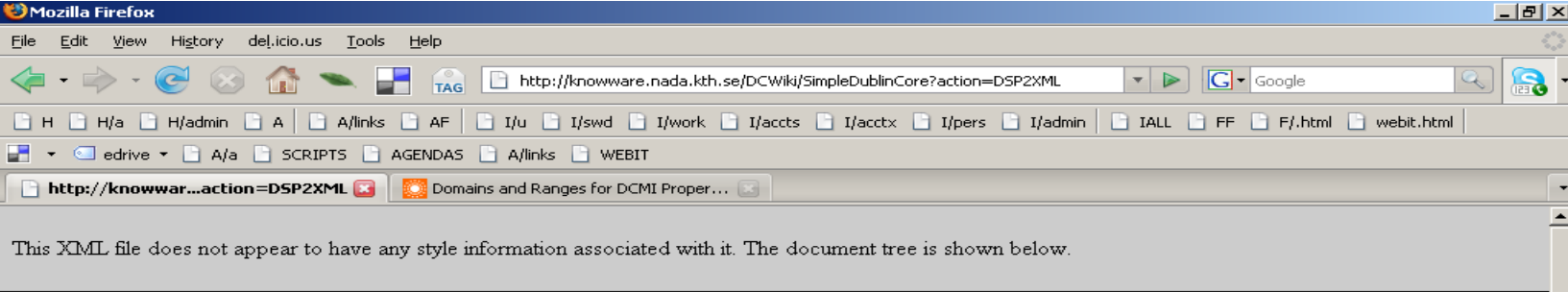
==== Creator ====
----
ST=(type="nonliteral" PC={http://purl.org/dc/terms/creator})
|| Definition || An entity primarily responsible for making the resource. ||
|| Comment || Examples of a Creator include a person, an organization, or a service. Typically, the name of a Creator
should be used to indicate the entity. ||
NLC=( VURIConstraint={ occurrence="disallowed"} VESConstraint={ occurrence="disallowed"} VStringConstraint={max="2"
LangC={occurrence="optional"} SESConstraint={occurrence="disallowed"} ) )

==== Subject ====
----
ST=(type="nonliteral" PC={http://purl.org/dc/terms/subject})
|| Definition || The topic of the resource. ||
|| Comment || Typically, the subject will be represented using keywords, key phrases, or classification codes.
Recommended best practice is to use a controlled vocabulary. To describe the spatial or temporal topic of the
resource, use the Coverage element. ||
NLC=( VURIConstraint={ occurrence="disallowed"} VESConstraint={ occurrence="disallowed"} VStringConstraint={max="1"
LangC={occurrence="optional"} SESConstraint={occurrence="disallowed"} ) )

==== Description ====
----
ST=(type="literal" PC={http://purl.org/dc/terms/description})
|| Definition || An account of the resource. ||
```

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

...extractable to a generic representation...



The screenshot shows a Mozilla Firefox browser window with the address bar displaying `http://knowware.nada.kth.se/DCWiki/SimpleDublinCore?action=DSP2XML`. The browser's status bar indicates that the XML file does not have any style information associated with it. The document tree is displayed below the status bar, showing a hierarchical structure of XML elements. The tree starts with a root element `<DescriptionSetTemplate>`, which contains two `<DescriptionTemplate>` elements. The first `<DescriptionTemplate>` has a `maxOccur="1"` and `minOccur="1"` attribute, and contains a `<StatementTemplate type="literal">` element. This `<StatementTemplate>` contains a `<Property>` element with the value `http://purl.org/dc/terms/title`, followed by a `<LiteralConstraint>` element. The `<LiteralConstraint>` element contains a `<SyntaxEncodingSchemeOccurrence>` element with the value `disallowed`, and a `<LanguageOccurrence>` element with the value `optional`. The second `<DescriptionTemplate>` has a `maxOccur="1"` and `minOccur="1"` attribute, and contains a `<StatementTemplate type="nonliteral">` element. This `<StatementTemplate>` contains a `<Property>` element with the value `http://purl.org/dc/terms/creator`, followed by a `<NonliteralConstraint>` element. The `<NonliteralConstraint>` element contains a `<ValueURIOccurrence>` element with the value `disallowed`, a `<VocabularyEncodingSchemeOccurrence>` element with the value `disallowed`, and a `<ValueStringConstraint maxOccur="1">` element. The `<ValueStringConstraint>` element contains a `<SyntaxEncodingSchemeOccurrence>` element with the value `disallowed`, and a `<LanguageOccurrence>` element with the value `optional`. The third `<DescriptionTemplate>` has a `maxOccur="1"` and `minOccur="1"` attribute, and contains a `<StatementTemplate type="nonliteral">` element. This `<StatementTemplate>` contains a `<Property>` element with the value `http://purl.org/dc/terms/subject`.

```
- <DescriptionSetTemplate>
- <DescriptionTemplate maxOccur="1" minOccur="1">
- <StatementTemplate type="literal">
  <Property>http://purl.org/dc/terms/title</Property>
  <LiteralConstraint>
    <SyntaxEncodingSchemeOccurrence>disallowed</SyntaxEncodingSchemeOccurrence>
    <LanguageOccurrence>optional</LanguageOccurrence>
  </LiteralConstraint>
</StatementTemplate>
- <StatementTemplate type="nonliteral">
  <Property>http://purl.org/dc/terms/creator</Property>
  <NonliteralConstraint>
    <ValueURIOccurrence>disallowed</ValueURIOccurrence>
    <VocabularyEncodingSchemeOccurrence>disallowed</VocabularyEncodingSchemeOccurrence>
    <ValueStringConstraint maxOccur="1">
      <SyntaxEncodingSchemeOccurrence>disallowed</SyntaxEncodingSchemeOccurrence>
      <LanguageOccurrence>optional</LanguageOccurrence>
    </ValueStringConstraint>
  </NonliteralConstraint>
</StatementTemplate>
- <StatementTemplate type="nonliteral">
  <Property>http://purl.org/dc/terms/subject</Property>
```

Find: simple Next Previous Highlight all Match case
Done

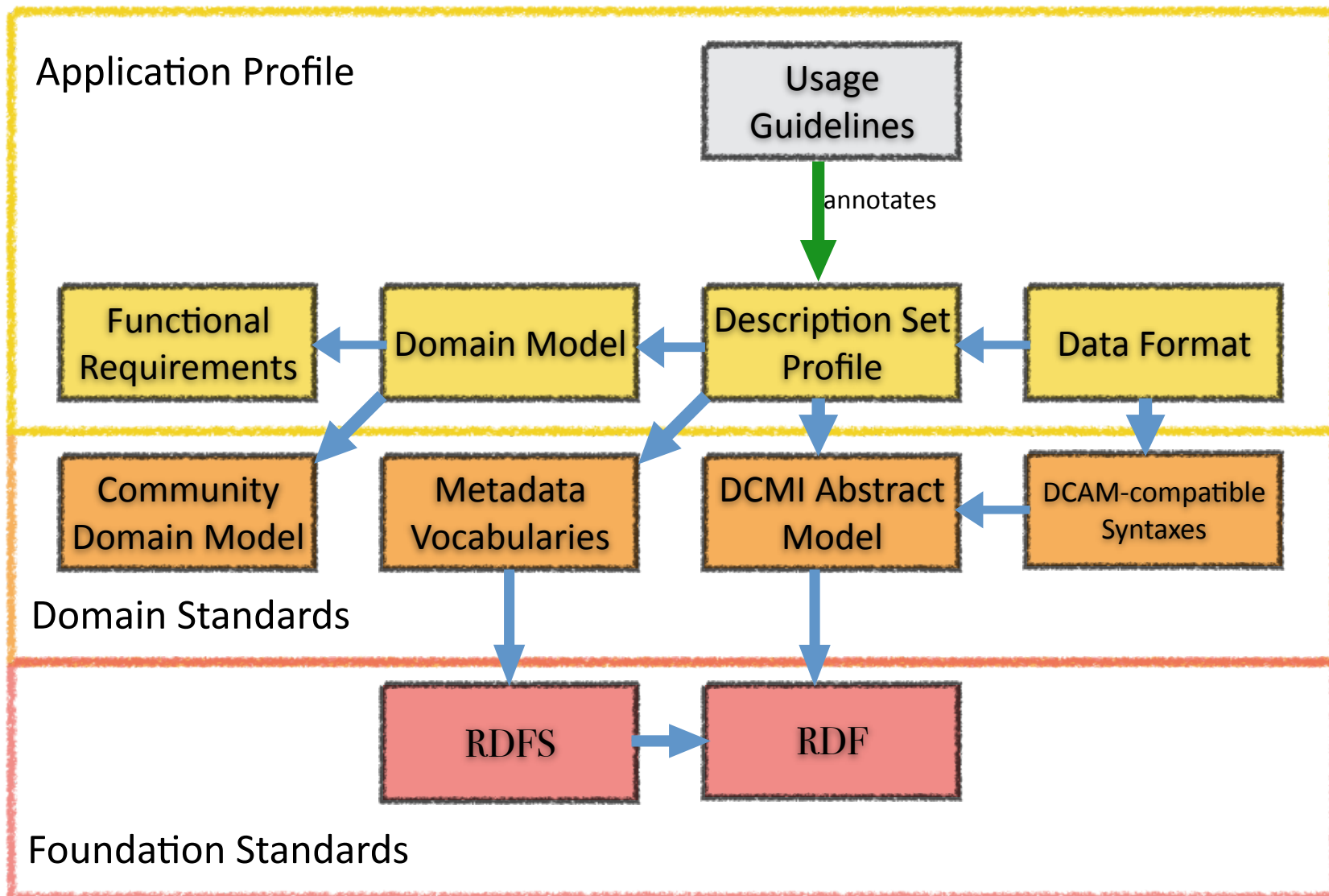
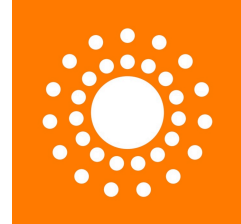
<http://dublincore.org/documents/2008/10/06/dsp-wiki-syntax/DescriptionSetProfile-dist.zip>

...usable to generate editor, validator, etc.

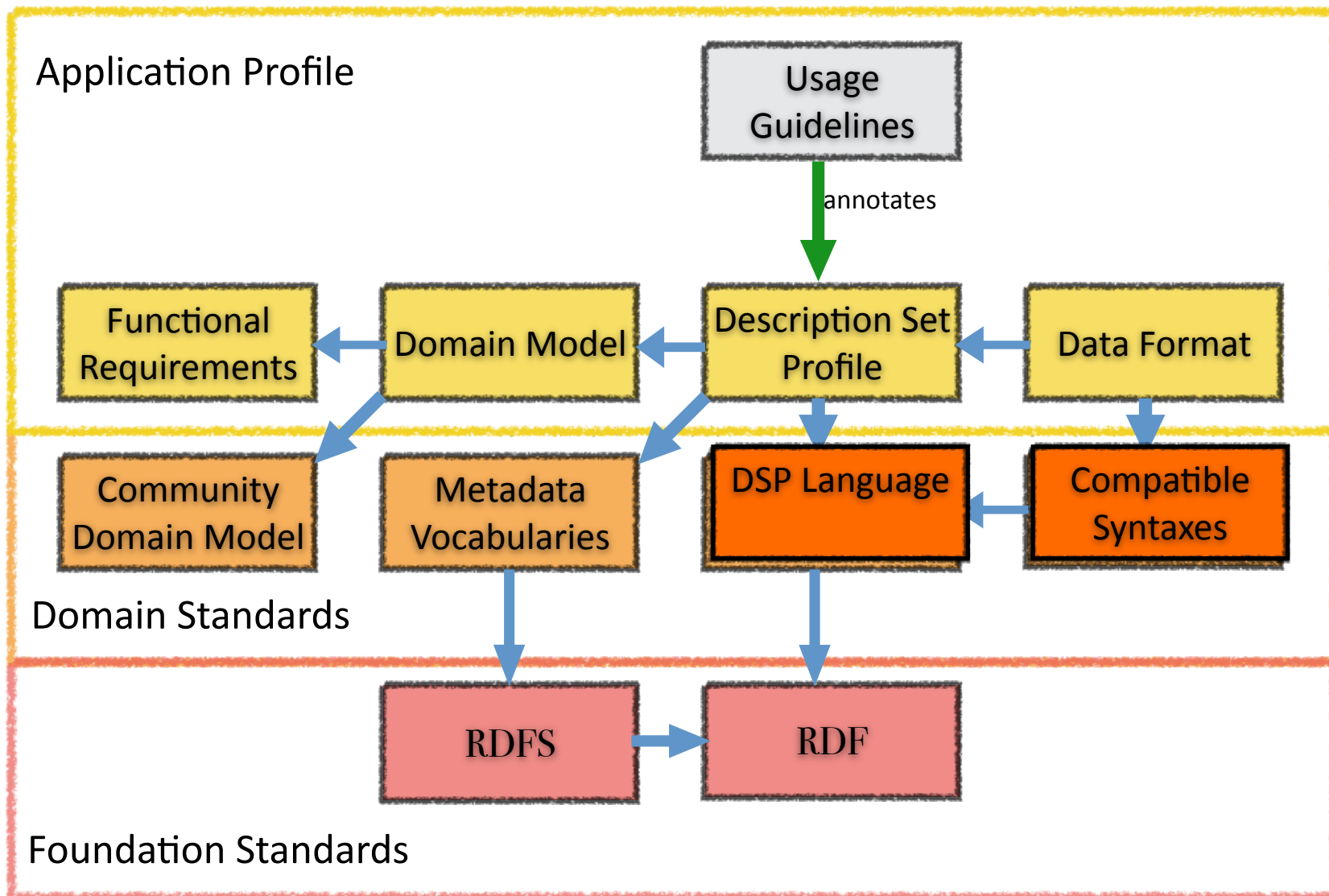
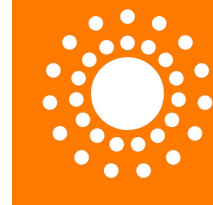
The screenshot shows a web browser window with a URL bar containing `http://www.dn.se/`. The page has a light blue background and contains several form fields for creating a Description Set Profile. Each field is accompanied by a green '+' button and a red '-' button. The fields are as follows:

- Title:** A text box containing 'DSP Presentation' and a dropdown menu set to 'English'.
- Creator:** A text box containing 'Mikael Nilsson' and a dropdown menu.
- Subject:** A text box containing 'application profiles' and a dropdown menu set to 'English'.
- Subject:** A text box containing 'description set profiles' and a dropdown menu set to 'English'.
- Subject:** A text box containing 'urce description framework' and a dropdown menu set to 'English'.

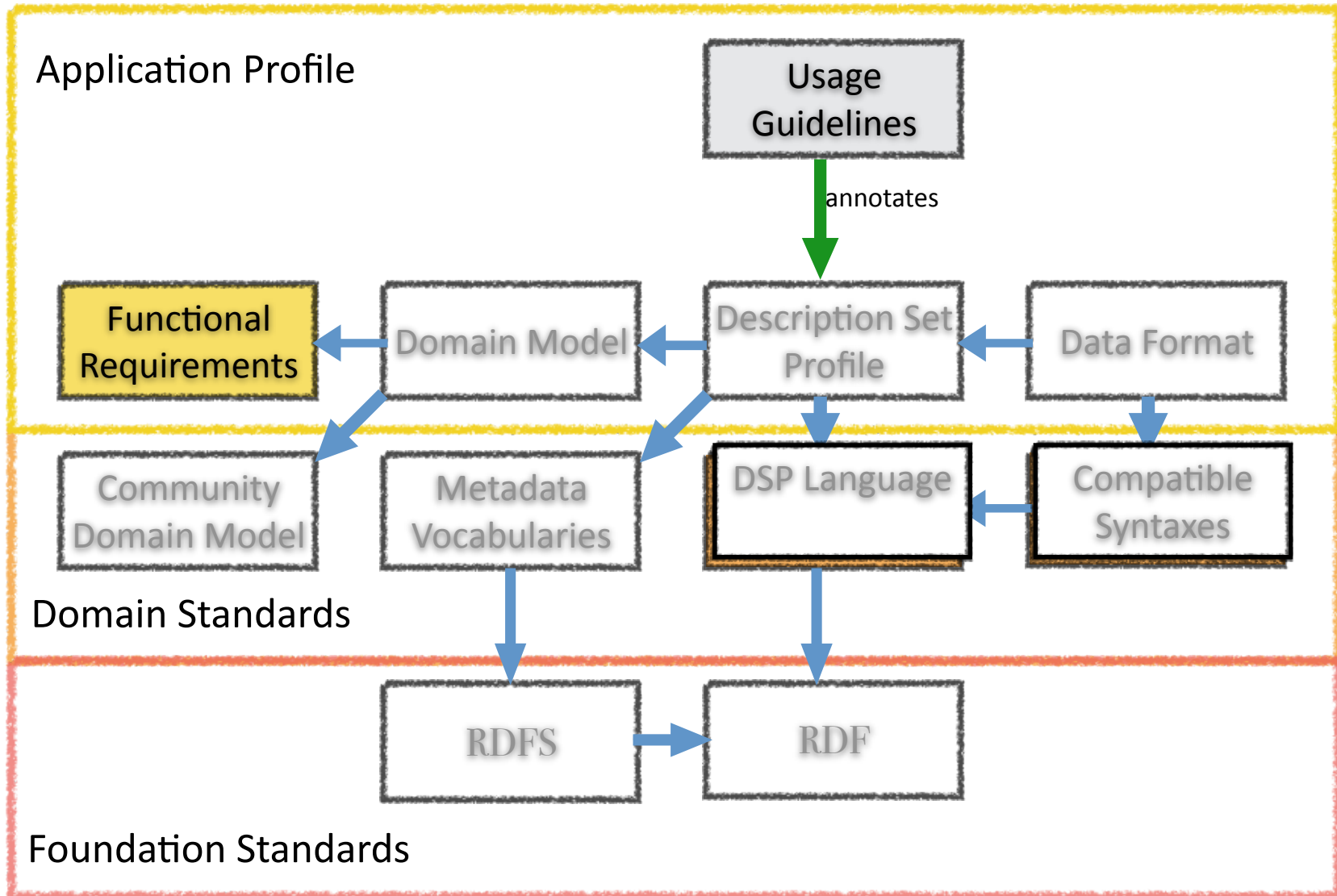
Singapore Framework for Dublin Core Application Profiles (2007)



An updated Description Set Profile templating language?



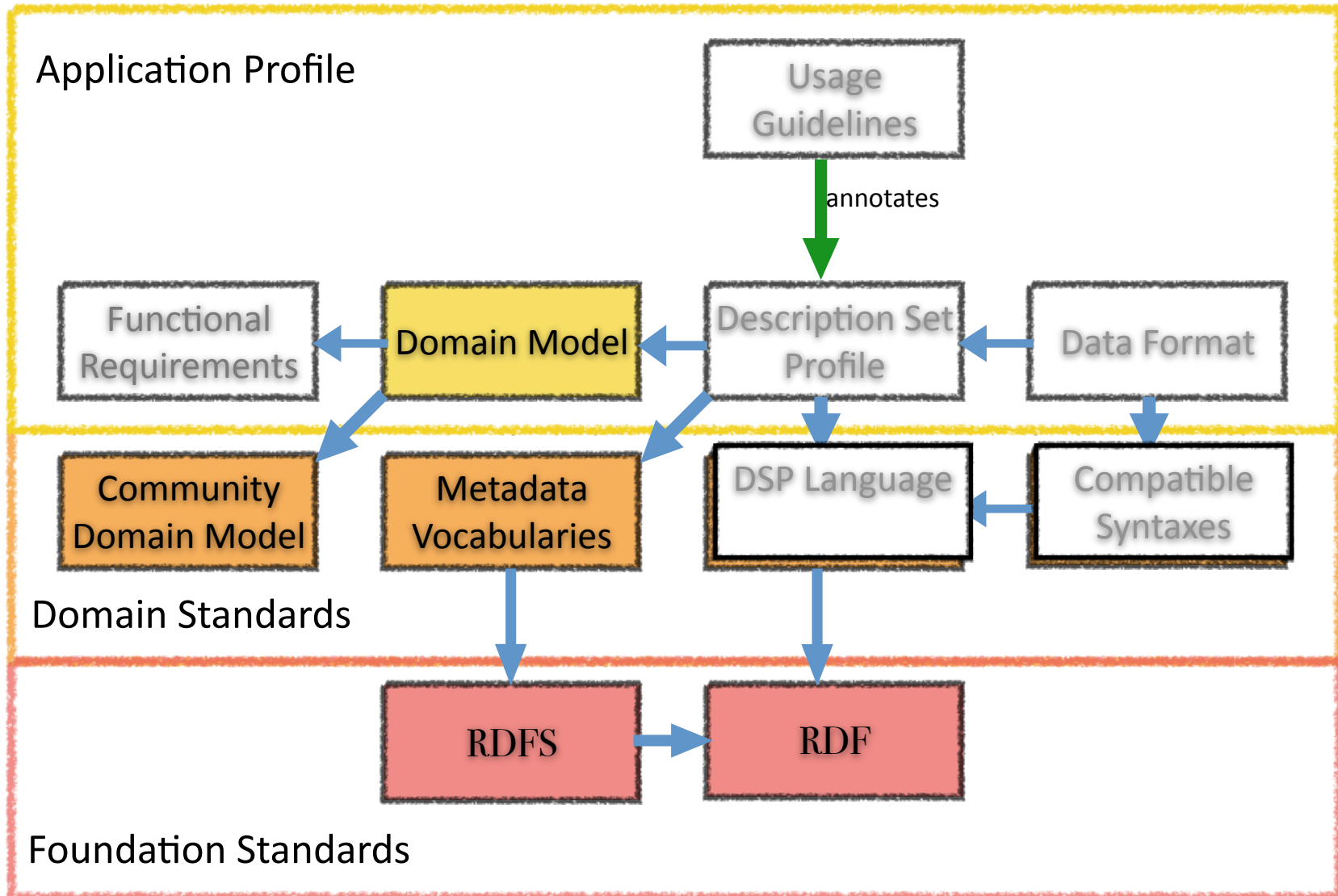
Base a Description Set Profile on specific requirements!





Model “reality” with a light touch!

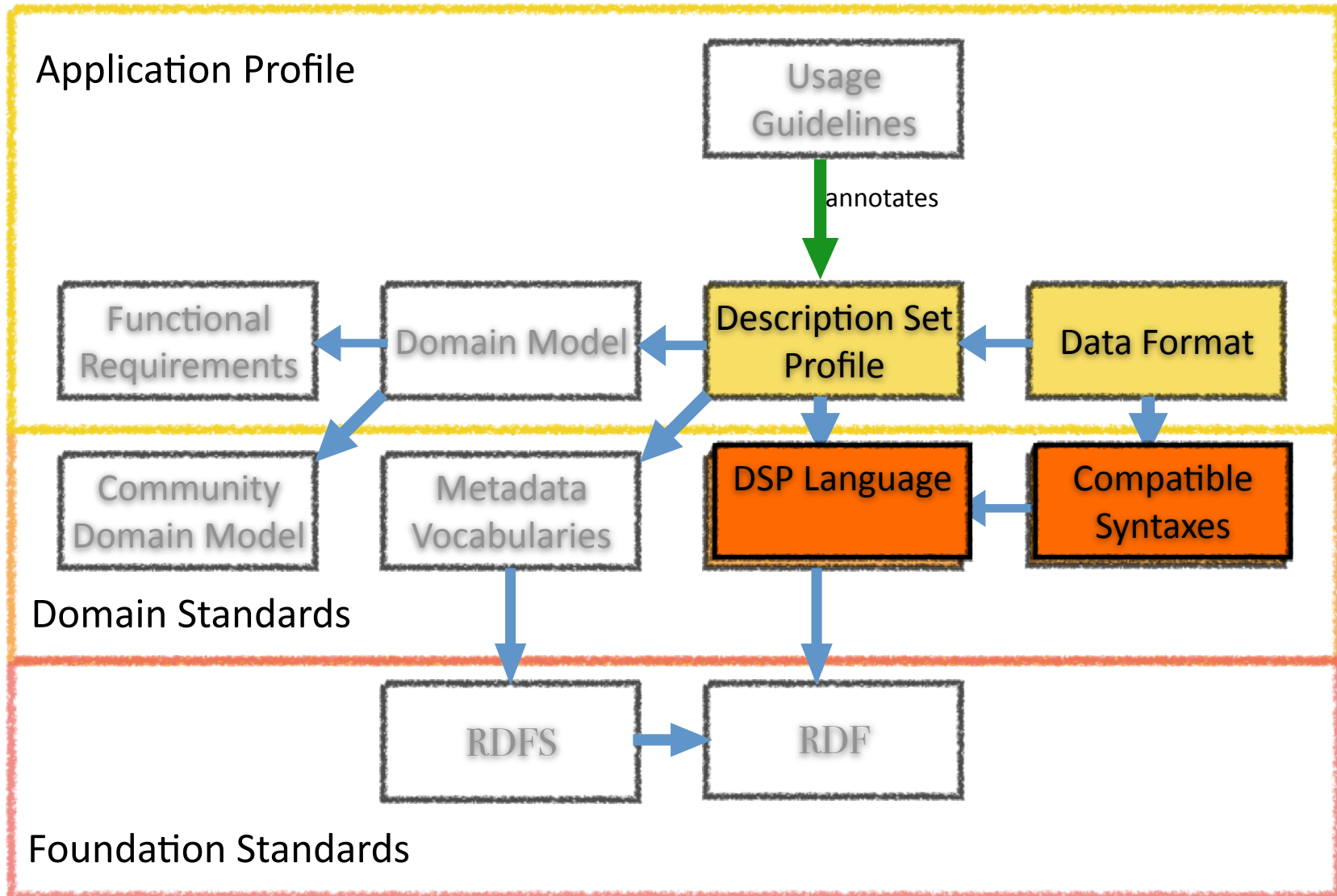
Principle of Minimal Semantic Commitment (using DC and similar vocabularies)





Constrain the *data* (*not* the vocabularies)!

For consistency and quality control





DSP design philosophy

- Design data for peaceful coexistence
- Pragmatic control of data quality
 - Go easy on sweeping ontological generalizations.
 - Constrain data produced.
 - Expect unexpected uses.
- Simple models have bigger audiences



DSP Requirements

- Authored in an idiom usable by normal people
- Constraints reusable for data automatic validation



September 4 discussion at DC-2013, Lisbon

- SPARQL a good candidate as a rules language for expressing constraints
- But users may need generic constraints language:
 - for authoring Application Profiles
 - for understanding Application Profiles
- Something like DC-DSP...?
 - using a syntax straightforwardly convertible into SPARQL queries?
 - as user-facing terminology for documenting Application Profiles?



tom@tombaker.org