

# Converting XML to RDF

Semantic Web Technologies  
Markus Plangg

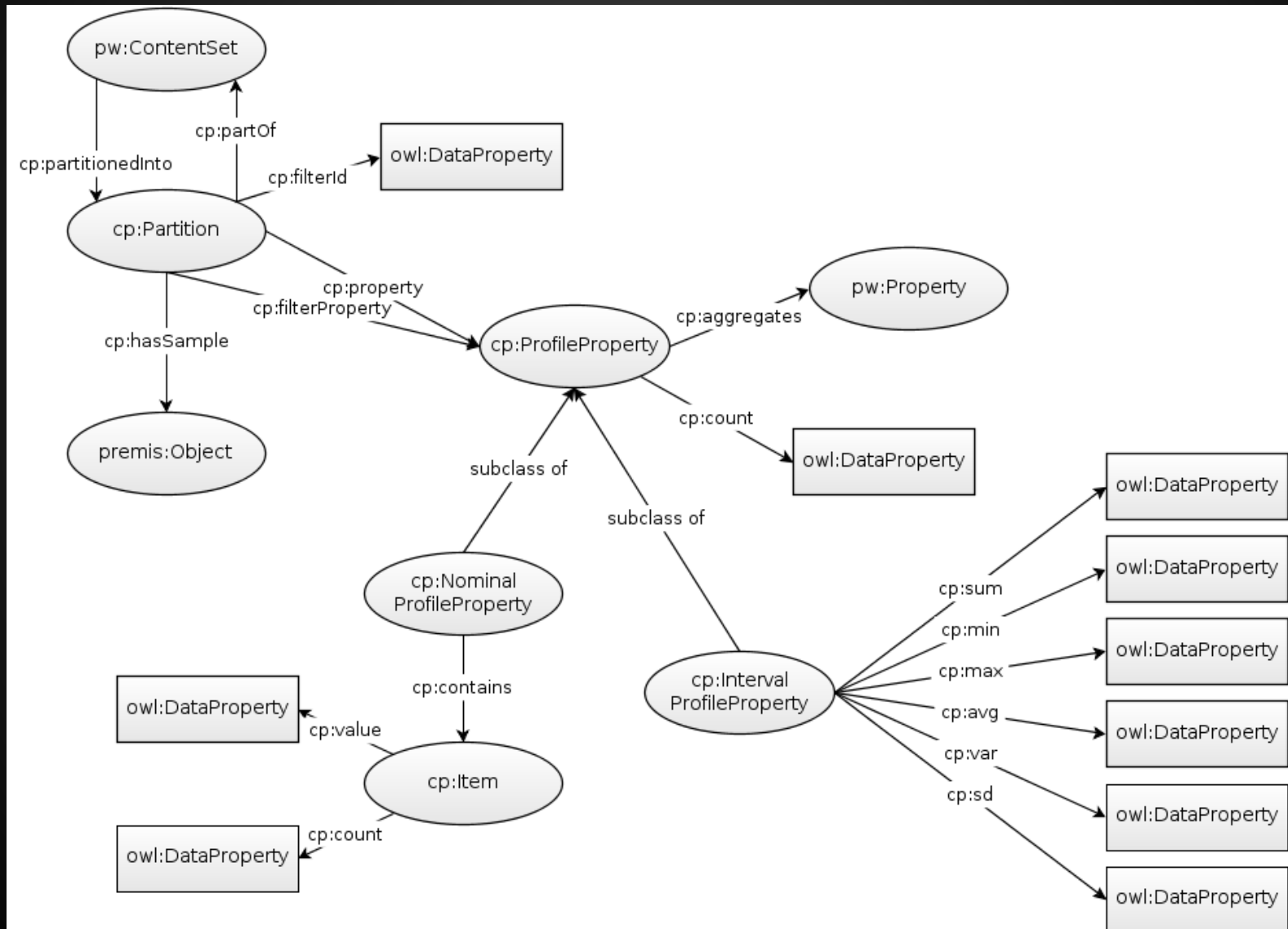
# Motivation

- XML widely available
  - Still fast growing
  - Output of third party tools
  - Established tool support
- 
- Adding the power of RDF
  - Integrating with RDF data

# Content Profile<sup>[1]</sup> Data

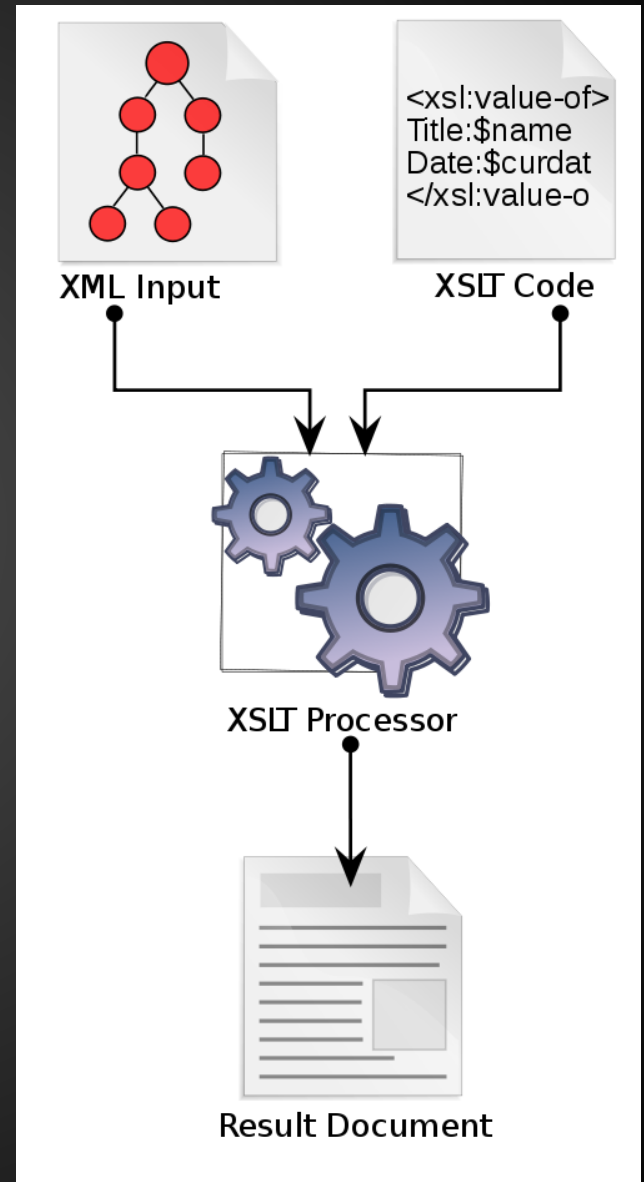
```
▼<profile version="0.1" collection="roda" date="Tue Jul 03 13:49:49 CEST 2012" count="264">
  ▼<partition count="264">
    ▼<filter id="76a0a02c-d26a-4514-8594-211189ed3782">
      ▼<parameters>
        ▼<parameter>
          <name>collection</name>
          <value>roda</value>
        </parameter>
      </parameters>
    </filter>
  ▼<properties>
    ▼<property id="mimetype" type="STRING" count="253">
      <item id="image/jpeg" value="108"/>
      <item id="image/tiff" value="104"/>
      <item id="text/xml" value="23"/>
      <item id="Unknown" value="11"/>
      <item id="application/octet-stream" value="8"/>
      <item id="application/pdf" value="8"/>
      <item id="application/vnd.oasis.opendocument.text" value="1"/>
      <item id="text/plain" value="1"/>
    </property>
    <property id="size" type="INTEGER" count="264" sum="6.22187353E8" min="352" max="1.29986092E8"
    avg="2356770.2765151514" var="7.852673758932184E13" sd="8861531.33433053"/>
    <property id="creating_application_name" type="STRING" count="123"/>
    <property id="lastmodified" type="DATE" count="250"/>
  </properties>
  ▼<representative-collection atMost="5">
    ▼<items>
      <item uid="http://roda.keep.pt/roda-core/get/roda:122/F0"/>
      <item uid="http://roda.keep.pt/roda-core/get/roda:112/F0"/>
      <item uid="http://roda.keep.pt/roda-core/get/roda:151/F0"/>
      <item uid="http://roda.keep.pt/roda-core/get/roda:154/F2"/>
      <item uid="http://roda.keep.pt/roda-core/get/roda:180/F10"/>
    </items>
  </representative-collection>
  ▼<elements>
    <element uid="http://roda.keep.pt/roda-core/get/roda:122/F0"/>
    <element uid="http://roda.keep.pt/roda-core/get/roda:112/F0"/>
    <element uid="http://roda.keep.pt/roda-core/get/roda:125/F0"/>
```

# Content Profile Ontology



# XSLT<sup>[2]</sup>

- Transformation of XML documents
- Declarative
- Rule based



# Generic transformation<sup>[3]</sup>

- Generic XSLT
- Can be used for all XML data

```
<?xml version="1.0" encoding="UTF-8"?>
  <!DOCTYPE rdf:RDF [
    <!ENTITY rdf "http://www.w3.org/1999/02/22-rdf-syntax-ns#">
    <!ENTITY ns "http://scafe-project.eu/pw/content-profiles#">
  ]>
<xsl:stylesheet xmlns:xsl="http://www.w3.org/1999/XSL/Transform"
  version="2.0" xmlns:rdf="&rdf;">

  <xsl:strip-space elements="*" />
  <xsl:output indent="yes" />

  <xsl:template match="/">
    <rdf:RDF xmlns:rdf='&rdf;' xmlns:ns="&ns;">
      <xsl:call-template name="element" />
    </rdf:RDF>
  </xsl:template>

  <xsl:template match="*" name="element">
    <xsl:variable name="separate-descriptions" select="*[count(@*!*
```

# Generic transformation

```
<xsl:template match="*" name="element">
  <xsl:variable name="separate-descriptions" select="*[count(@*|*)>0 and count(text())=0]" />
  <rdf:Description rdf:nodeID="{generate-id()}" xmlns="&ns;">
    <xsl:for-each select="@*">
      <xsl:attribute name="{local-name()}" namespace="&ns;">
        <xsl:value-of select="." />
      </xsl:attribute>
    </xsl:for-each>

    <xsl:for-each select="$separate-descriptions">
      <xsl:element name="{local-name()}">
        <xsl:attribute name="rdf:nodeID" select="generate-id()" />
      </xsl:element>
    </xsl:for-each>

```

# Generic transformation

```
<xsl:for-each select="* except $separate-descriptions">
  <xsl:element name="{local-name()}">
    <xsl:choose>
      <xsl:when test="count(*)>0">
        <xsl:attribute name="rdf:parseType">Literal</xsl:attribute>
        <xsl:copy-of select="*|text()" copy-namespaces="no" />
      </xsl:when>
      <xsl:otherwise>
        <xsl:value-of select="." />
      </xsl:otherwise>
    </xsl:choose>
  </xsl:element>
</xsl:for-each>
```



# Generic transformation

```
<rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
  xmlns:ns="http://scape-project.eu/pw/content-profiles#">
  <rdf:Description xmlns="http://scape-project.eu/pw/content-profiles#" rdf:nodeID="d1">
    <profile rdf:nodeID="d1e1"/>
  </rdf:Description>
  <rdf:Description xmlns="http://scape-project.eu/pw/content-profiles#" rdf:nodeID="d1e1"
    ns:version="0.1"
    ns:collection="roda"
    ns:date="Tue Jul 03 13:49:49 CEST 2012"
    ns:count="264">
    <partition rdf:nodeID="d1e2"/>
  </rdf:Description>
  <rdf:Description xmlns="http://scape-project.eu/pw/content-profiles#" rdf:nodeID="d1e2"
    ns:count="264">
    <filter rdf:nodeID="d1e3"/>
    <properties rdf:nodeID="d1e10"/>
    <representative-collection rdf:nodeID="d1e144"/>
    <elements rdf:nodeID="d1e154"/>
  </rdf:Description>
  <rdf:Description xmlns="http://scape-project.eu/pw/content-profiles#" rdf:nodeID="d1e3"
    ns:id="76a0a02c-d26a-4514-8594-211189ed3782">
    <parameters rdf:nodeID="d1e4"/>
  </rdf:Description>
  <rdf:Description xmlns="http://scape-project.eu/pw/content-profiles#" rdf:nodeID="d1e4">
    <parameter rdf:nodeID="d1e5"/>
  </rdf:Description>
  <rdf:Description xmlns="http://scape-project.eu/pw/content-profiles#" rdf:nodeID="d1e5">
    <nameasrdf>collection</nameasrdf>
    <valueasrdf>roda</valueasrdf>
  </rdf:Description>
</rdf:RDF>
```

# Custom Transformation

```
<!-- RDF root -->
<xsl:template match="/">
  <rdf:RDF xmlns:rdf='&rdf;' xmlns:ns="&cp;">
    <xsl:apply-templates />
  </rdf:RDF>
</xsl:template>

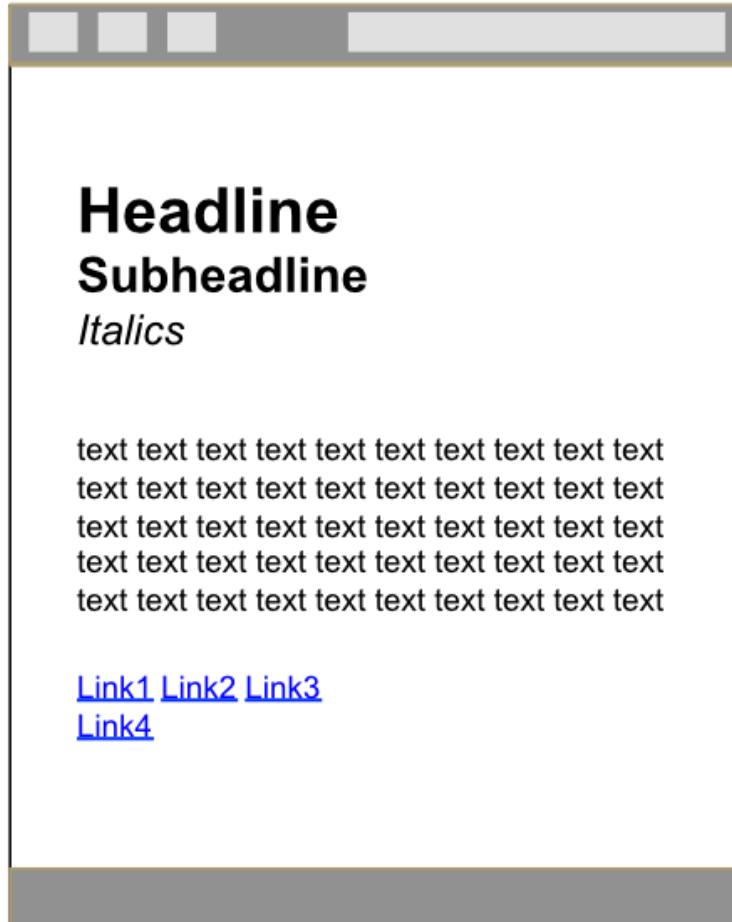
<!-- Profile -->
<xsl:template match="/profile">
  <rdf:description about="{@collection}">
    <rdf:type rdf:resource="&pw;ContentSet" />
  </rdf:description>
  <xsl:apply-templates />
</xsl:template>

<!-- Partition -->
<xsl:template match="/profile/partition">
  <rdf:description>
    <rdf:type rdf:resource="&cp;Partition" />
    <cp:filterId>
      <xsl:value-of select="filter/@id" />
    </cp:filterId>
    <xsl:apply-templates select="representative-collection/items/item" />
    <xsl:for-each select="properties/property">
      <cp:contains rdf:resource="{generate-id()}" />
    </xsl:for-each>
  </rdf:description>
  <xsl:apply-templates select="properties" />
  <xsl:apply-templates select="elements" />
</xsl:template>
```

# Custom transformation

```
<rdf:RDF xmlns:ns="http://scape-project.eu/pw/content-profiles#"
  xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
  xmlns:cp="http://scape-project.eu/pw/content-profiles#">
  <rdf:description about="roda">
    <rdf:type rdf:resource="http://scape-project.eu/pw/vocab#ContentSet"/>
  </rdf:description>
  <rdf:description>
    <rdf:type rdf:resource="http://scape-project.eu/pw/content-profiles#Partition"/>
    <cp:filterId>76a0a02c-d26a-4514-8594-211189ed3782</cp:filterId>
    <rdf:hasSample rdf:resource="http://roda.keep.pt/roda-core/get/roda:122/F0"/>
    <rdf:hasSample rdf:resource="http://roda.keep.pt/roda-core/get/roda:112/F0"/>
    <!-- ... -->
    <rdf:hasSample rdf:resource="http://roda.keep.pt/roda-core/get/roda:180/F10"/>
    <cp:contains rdf:resource="d1e22"/>
    <cp:contains rdf:resource="d1e44"/>
    <!-- ... -->
    <cp:contains rdf:resource="d1e309"/>
  </rdf:description>
  <rdf:description rdf:about="d1e22">
    <cp:aggregates rdf:resource="http://scape-project.eu/pw/vocab/format_version"/>
    <cp:count>140</cp:count>
    <rdf:type rdf:resource="http://scape-project.eu/pw/content-profiles#NominalProperty"/>
    <cp:contains rdf:resource="d1e24"/>
    <cp:contains rdf:resource="d1e26"/>
    <!-- ... -->
    <cp:contains rdf:resource="d1e41"/>
  </rdf:description>
  <rdf:description rdf:about="d1e24">
    <rdf:type rdf:resource="http://scape-project.eu/pw/content-profiles#Item"/>
    <cp:value>Unknown</cp:value>
    <cp:count>124</cp:count>
  </rdf:description>
  <rdf:description rdf:about="d1e26">
```

# RDFa



A mockup of a web browser window with a grey header and footer. The main content area is white and contains the following text:

**Headline**  
**Subheadline**  
*Italics*

text text text text text text text text text text  
text text text text text text text text text text  
text text text text text text text text text text  
text text text text text text text text text text  
text text text text text text text text text text

[Link1](#) [Link2](#) [Link3](#)  
[Link4](#)



A mockup of a web browser window with a grey header and footer. The main content area is white and contains the following text:

**Title**  
**Author**  
*Publication Date*

article content article content article content  
article content article content article content  
article content article content article content  
article content article content article content  
article content article content article content

[Tag1](#) [Tag2](#) [Tag3](#)  
[Copyright License](#)

# RDFa 1.0

- Specified for XHTML
- First proposed in Feb 2004<sup>[4]</sup>
- W3C Working Draft in Oct 2007<sup>[5]</sup>
- Recommendation in Oct 2008<sup>[6]</sup>

# RDFa 1.1

- RDFa Core 1.1
  - XHTML+RDFa 1.1
  - HTML+RDFa 1.1
  - RDFa Lite 1.1
- 
- Working Draft in April 2010<sup>[7]</sup>
  - Recommendation June 2012<sup>[8]</sup>

# RDFa Lite 1.1

- Subset of RDFa Core
- Simple, easy to learn
- For simple data markup for properties

## Example

```
<p vocab="http://schema.org/" typeof="Person">  
  My name is  
  <span property="name">Manu Sporny</span>  
  and you can give me a ring via  
  <span property="telephone">1-800-555-0199</span>  
  or visit  
  <a property="url" href="http://manu.sporny.org/">my homepage</a>.  
</p>
```

# RDFa Lite 1.1

- Prefixes
- Identify resources

## Example

```
<p vocab="http://schema.org/" prefix="ov: http://open.vocab.org/terms/" resource="#manu" typeof="Person">
  My name is
  <span property="name">Manu Sporny</span>
  and you can give me a ring via
  <span property="telephone">1-800-555-0199</span>.
  
  My favorite animal is the <span property="ov:preferredAnimal">Liger</span>.
</p>
```



# RDFa Core 1.1

## Example

This document is licensed under a  
<a prefix="cc: http://creativecommons.org/ns#" rel="cc:license" href="http://creativecommons.org/licenses/by-nc-nd/3.0/">  
Creative Commons License  
</a>.

# RDFa Core 1.1

## Example

```
<html
  xmlns="http://www.w3.org/1999/xhtml"
  prefix="cal: http://www.w3.org/2002/12/cal/ical#"
>
<head><title>Jo's Friends and Family Blog</title></head>
<body>
  <p>
    I'm holding
    <span property="cal:summary">
      one last summer Barbecue
    </span>,
    on September 16th at 4pm.
  </p>
</body>
</html>
```

# RDFa Core 1.1

## Example

```
<html
  xmlns="http://www.w3.org/1999/xhtml"
  prefix="biblio: http://example.org/
         dc: http://purl.org/dc/elements/1.1/"
  >
  <head>
    <title>Books by Marco Pierre White</title>
  </head>
  <body>
    I think White's book
    '<span about="urn:ISBN:0091808189" typeof="biblio:book"
      property="dc:title">
      Canteen Cuisine
    </span>'
    is well worth getting since although it's quite advanced stuff, he
    makes it pretty easy to follow. You might also like
    <span about="urn:ISBN:1596913614" typeof="biblio:book"
      property="dc:description">
      White's autobiography
    </span>.
  </body>
</html>
```

## Example

```
<html xmlns="http://www.w3.org/1999/xhtml"
      prefix="rdfa: http://www.w3.org/ns/rdfa#">
  <head>
    ...
  </head>
  <body>
    <p>This is an example to defining the standard RDF and
      Dublin Core prefixes
    </p>

    <p typeof="">
      The "<span property="rdfa:prefix">rdf</span>" prefix can
      be used for the URI:
      "<span property="rdfa:uri">http://www.w3.org/1999/02/22-rdf-syntax-ns#</span>".</p>

    <p typeof="">
      The "<span property="rdfa:prefix">rdfs</span>" prefix can
      be used for the URI:
      "<span property="rdfa:uri">http://www.w3.org/2000/01/rdf-schema#</span>".</p>

    <p typeof="">
      The "<span property="rdfa:prefix">dc</span>" prefix can
      be used for the URI:
      "<span property="rdfa:uri">http://dublincore.org/documents/dcmi-terms/</span>".</p>
  </body>
</html>
```

# Conclusion

- Generic XSLT
- Custom XSLT
- RDFa

# References

[1] <https://github.com/peshkira/c3po/>

[2] <http://www.w3.org/TR/xslt20/>

[3] A standard transformation from XML to RDF via XSLT, F. Breitling, *Astronomical Notes*

[4] <http://www.w3.org/MarkUp/2004/02/xhtml-rdf.html>

[5] <http://www.w3.org/TR/2007/WD-rdfa-syntax-20071018/>

[6] <http://www.w3.org/TR/2008/REC-rdfa-syntax-20081014/>

[7] <http://www.w3.org/TR/2010/WD-rdfa-core-20100422/>

[8] <http://www.w3.org/TR/2012/REC-rdfa-core-20120607/>

# Questions



# PW Ontology

