



# The new ODF 1.2 Metadata Framework

**Svante Schubert**  
Software Engineer  
Sun Microsystems Inc.



# Why am I here?

- For introducing metadata feature & possibilities
  - > Member of the OASIS OpenDocument Format (ODF) TC
  - > Co-editor of the OASIS ODF Metadata Specification
  - > Member of Sun Microsystems StarOffice / OpenOffice.org team since 1999
  - > OpenOffice.org XML project co-lead

# Metadata Basics

- What is metadata?
  - > “Metadata is data about data”
- Why do I need metadata?
  - > Classify / 'Tag' your data
  - > Make data machine interchangeable
  - > Bridge the semantic gap
  - > ...

# Why using Metadata?

- What is metadata?
  - > Labels to identify/categorize your data
  - > Related data
- Why do I need metadata?
  - > Metadata makes your data interpretable by other applications

# Potential Interoperability Savings..

- *Information Integration in the Enterprise*, Philip A. Bernstein and Laura M. Haas, CACM, September 2008, pp. 72-79 (source for IT budgets (source for interoperability costs)
- Untapped Assets: The \$3 Trillion Value in US Enterprise Documents  
<http://www.brightplanet.com/images/stories/pdf/documentsvalue.pdf>
- NIST Strategic Planning Summary,  
[http://www.nist.gov/director/planning\\_summary\\_strategic\\_planning\\_study\\_results.html](http://www.nist.gov/director/planning_summary_strategic_planning_study_results.html)

# EU Public Key Interoperability Areas

- Income taxes: declaration / notifications
- Job search services by labour offices
- Personal documents (passport and driving licence)
- Car registration (new, used and imported cars)
- Application for building permission
- Declaration to the police (e.g. in case of theft)
- Public libraries (catalogues, search tools)
- Certificates (birth, marriage): request and delivery
- Enrolment in higher education / university

# EU Public Key Interoperability Areas (Businesses)

- Social contribution for employees
- Corporation tax: declaration, notification
- VAT: declaration, notification
- Registration of a new company
- Submission of data to statistical offices
- Customs declarations
- Environment-related permits (including reporting)
- Public procurement

# Do you know ODF?

- The OASIS / ISO standard for office documents (2005/06)
- The document format of many office applications

**OASIS** 

---

Open Document Format for Office  
Applications (OpenDocument) v1.0

OASIS Standard, 1 May 2005



International  
Organization for  
Standardization

**ISO/IEC 26300**



# Do you know ODF?

- Origin from OpenOffice.org's default format
- A zipped package of XML and related files (image, sound, user files)

**OASIS** 

---

Open Document Format for Office  
Applications (OpenDocument) v1.0

OASIS Standard, 1 May 2005

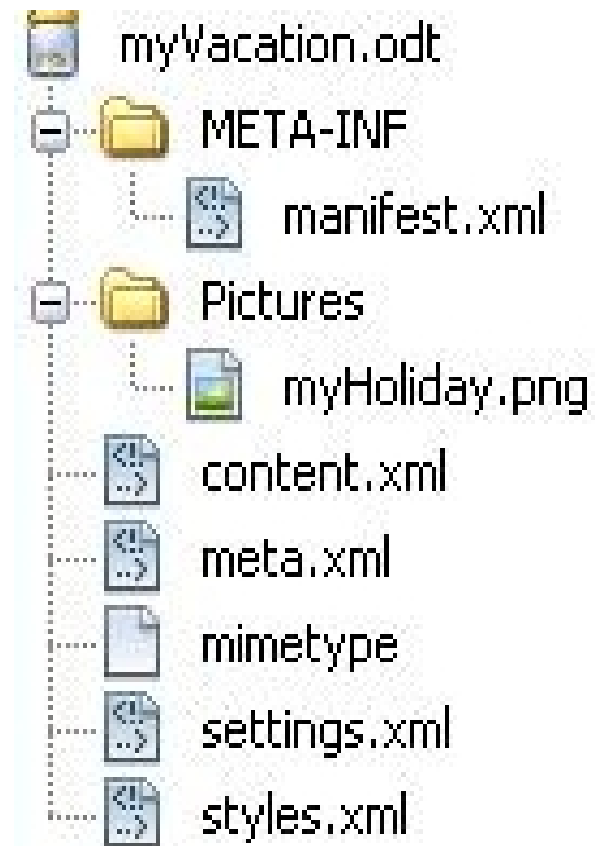


International  
Organization for  
Standardization

**ISO/IEC 26300**

# ODF Basics - Package & Files

- ISO standardized default content (as shown, but NOT Picture folder)
- Manifest as an Inventory / “table of content”
- Any user content...



# Existing Metadata Standards

## Resource Description Framework (RDF)

- Resources
  - > Unique identification by unicode enabled URI - Internationalized Resource Identifier (IRI)
  - > Described by RDF statements

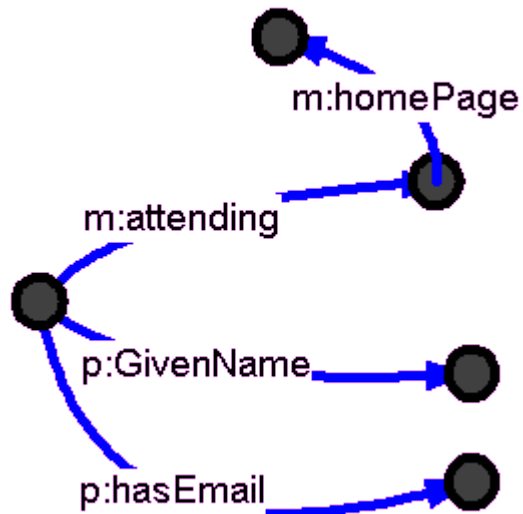
# Existing Metadata Standards

## Resource Description Framework (RDF)

- RDF Statements -  
Triple: subject + predicate + object
  - > <http://dbpedia.org/resource/Beijing>  
<http://dbpedia.org/ontology/PopulatedPlace#populationTotal>  
“17430000” .

# Existing Metadata Standards

RDF graphs...





# Existing Metadata Standards

## Web Ontology Language (OWL)

- Ontology (from Greek)
  - > Onto (being)
  - > Logia (written/spoken discourse)
- Description of entities/concepts and their relations
  - > Like OO language using classes, properties, etc.
- ODF 1.2 includes OWL package description
  - > OWL Classes of pkg:Document, pkg:File, odf:Element

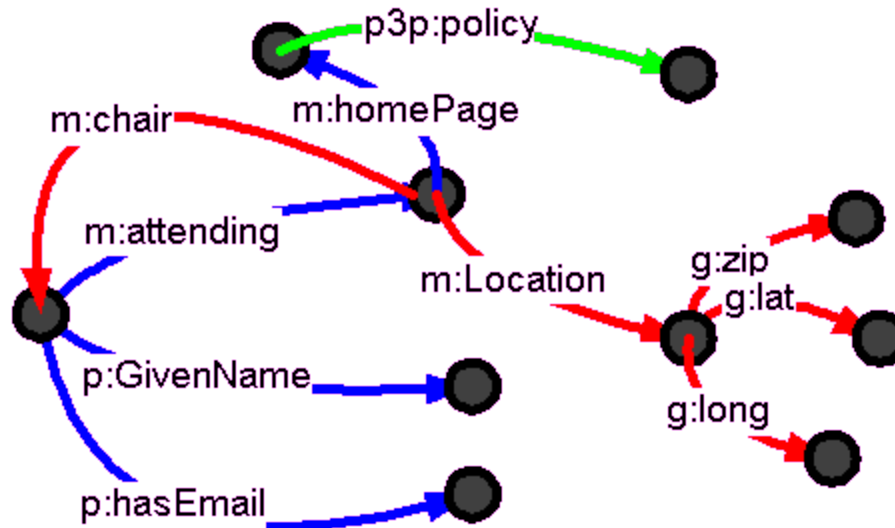
# Describing ODF content with RDF

- pkg:Document
- pkg:Element
- pkg:File
- pkg:hasPart
- pkg:idref
- pkg:MetadataFile
- pkg:mimeType
- pkg:path
- xhtml:about
- xhtml:content
- xhtml:datatype
- odf:ContentFile
- odf:Element
- odf:prefix
- odf:StylesFile
- odf:suffix



# Using RDF/XML standard in ODF

- How connect graph from RDF/XML to ODF Package and ODF XML files?
  - > RDF/XML not compatible to XPATH
  - > Data has to be mapped to Lingua Franca of applications



# Describing ODF with RDF

## Metadata Files in the Package

Meta data manifest  
“manifest.rdf”

Content file  
e.g. “/content.xml”

RDF file  
e.g. “/meta/data.rdf”

Content file  
e.g. “/styles.xml”

RDF file  
e.g. “/meta/cita.rdf”

# Metadata Model of ODF 1.2

## Metadata Files in the Package

- Content files (e.g. content.xml, styles.xml)
  - > About 50 ODF elements with xml:id attribute
- Metadata manifest (manifest.rdf)
  - > Heart of metadata model
  - > Lookup index for application interested in metadata
- User RDF/XML files
  - > Metadata file possibly from an office extension

# Metadata Model of ODF 1.2

## Metadata Files in the Package

Content files  
e.g. "/content.xml"

```
<table:table xml:id="someID">
    ... Hospital Doctor Duty List ...
</table:table>
```

Meta data manifest  
"/manifest.rdf"

```
<odf:Element rdf:about="content.xml#someID"/>
```

RDF files  
e.g. "/meta/data.rdf"

```
<odf:Element rdf:about="content.xml#someID">
  <ex:workingHoursOf>
    <med:Doctors rdf:about="John.Example@somewhere.org">
      <med:fieldName xml:lang="en">Neurologist</med:fieldN.
```

# Metadata Model of ODF 1.2

## In Content Metadata

- If metadata is equal text/visual data
- Reason: No data duplication
- Used by ODF 5 elements:
  - > Bookmark start - <text:bookmark-start>
  - > Heading - <text:h>
  - > Metadata text - <text:meta>
  - > Paragraph - <text:p>
  - > Table cell - <table:table-cell>

# Metadata Model of ODF 1.2

## In Content Metadata

Content files  
e.g. "/content.xml"

```
<text:p>The doctor's name was  
  <text:meta xhtml:about="John.Example@somewhere.org"  
            xhtml:property="http://xmlns.com/foaf/0.1/name">  
    Dr. J. Example</text:meta>
```

RDF files  
e.g. "/meta/data.rdf"

```
<med:Doctor rdf:about="John.Example@somewhere.org">  
  <med:hasPatient>  
    <med:Patient rdf:about="http://hospital-DB/patients/ID007">
```

# The RDF Application as Field User

## Metadata Text Field

- Content of field “text:meta-field” based on metadata
  - > Appears within paragraph
  - > Holds any paragraph content  
(E.g. Citation “According to [2]”)

```
<text:p>According to  
  <text:meta-field xml:id="someID">  
    <text:style text:style-name="s1">[2]
```

# Metadata support in OOo 3.1

- Support of metadata in the ODF 1.2 package
- API for metadata extension developers
  - > Providing OOo API to access metadata
  - > Wrapping existing opensource tool:  
Redland RDF Library – librdf
- Possibility of a generic metadata extension
  - > Import / Create your own RDF vocabulary
  - > Relate vocabulary to ODF content by GUI



# Metadata support in OOO 3.1

- Metadata on different Documents in a Package
  - > Multiple documents in an ODF package
  - > Metadata relative to document
  - > Metadata concept reusable by ODF package implementors

# Authoring with Metadata in OOo 3.1

- Scenarios with Embedded Objects
  - > Insert Embedded Object
  - > Extract Embedded Object
  - > Remove Embedded Object

# Authoring with Metadata in OOo 3.1

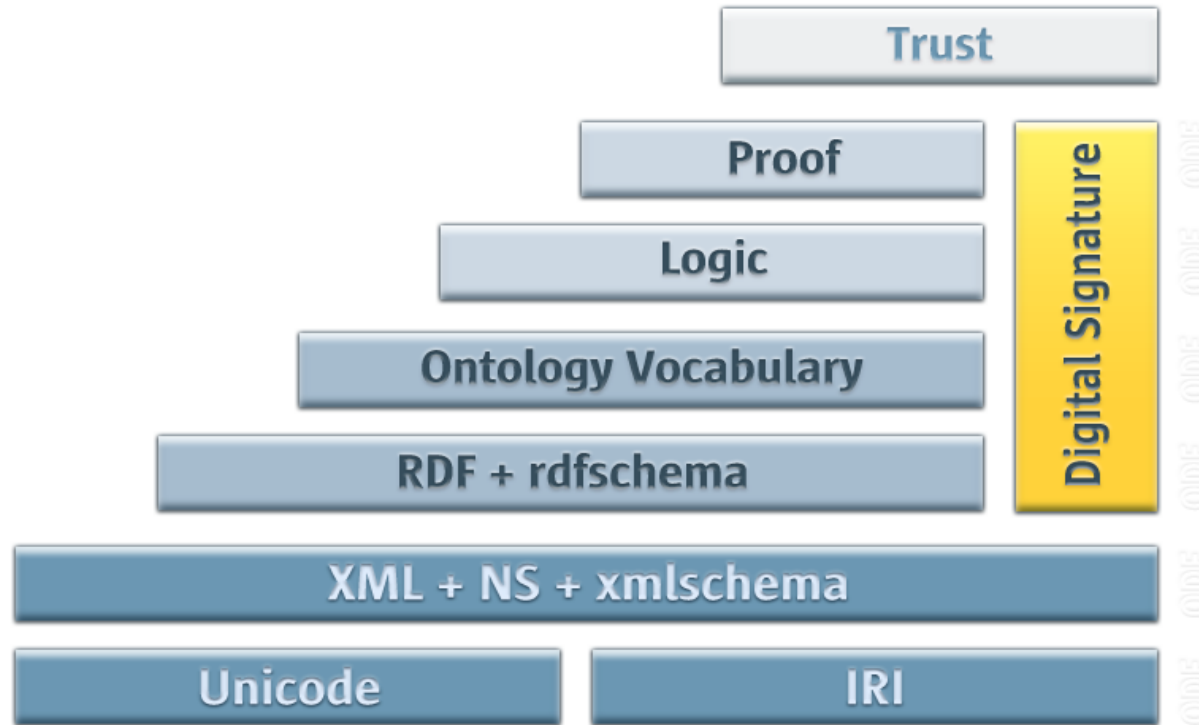
- Handling of metadata on editing a text document
  - > Deletion of an entity
  - > Copy of an entity
  - > Cut of an entity
  - > Paste of an entity
  - > Edit of entity content
  - > Split of entity
  - > Merge of two entities

# Scenarios to add Metadata to ODF

- ODF Template with Metadata (e.g. formular)
- RDF application (e.g. Citation Extension) offers some GUI
- Explicitly by the user
  - > Similar to spell-checker but text pattern base (User have to agree on metadata)
  - > Generic Metadata Extension:  
User choose some RDF Vocabulary and relates vocabulary to ODF content by GUI

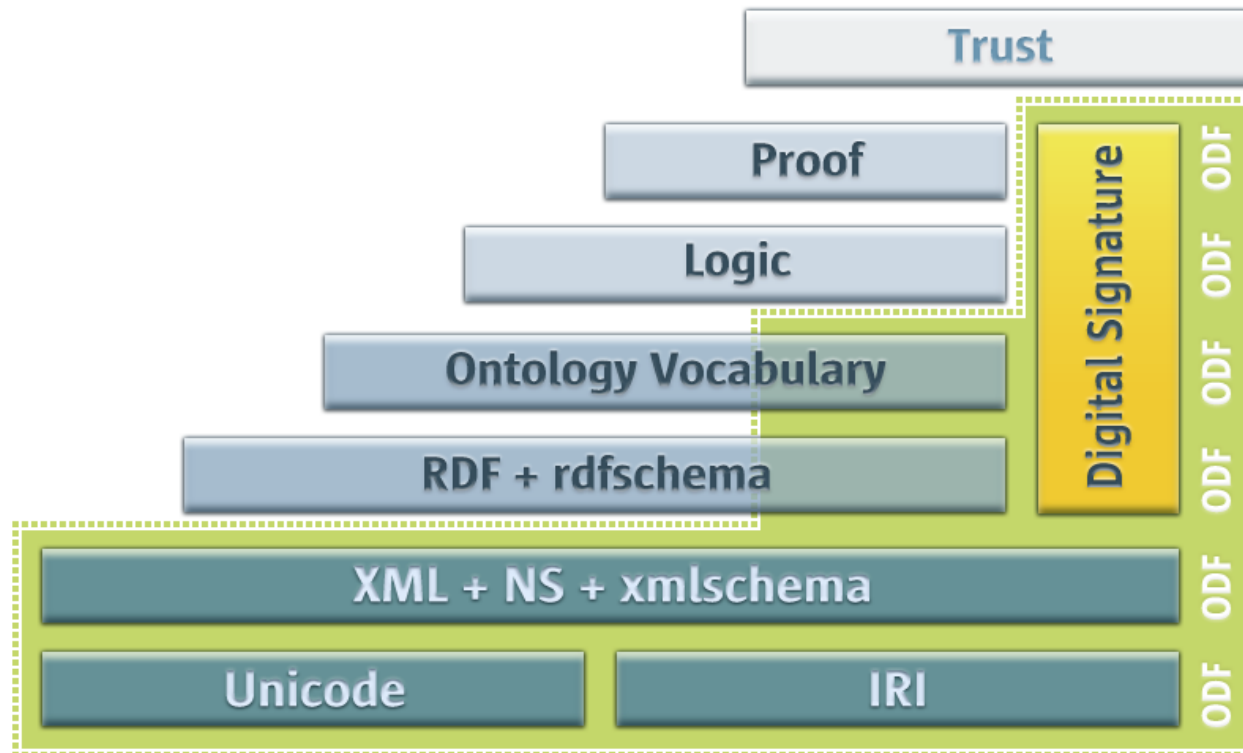
# Existing Metadata Standards

## Semantic Web Architecture



# Existing Metadata Standards

## ODF in the Semantic Web



# More Information

- Download latest ODF specification:
  - > Latest Draft: <http://www.oasis-open.org/committees/download.php/29446/08-09-24-ODF-Metadata-Change-Request.odt>
  - > ODF Metadata Examples: <http://www.oasis-open.org/committees/download.php/26846/08-01-16-ODF-Metadata-Examples-Committee-Draft.odt> (old)
  - > Upcoming adaptations: <http://lists.oasis-open.org/archives/office/200811/msg00006.html>
- OOo API (not final):
  - > <http://api.openoffice.org/docs/common/ref/com/sun/star/rdf/module-ix.html>



# Thank you – Questions & Answers

**Svante Schubert**

**Svante.Schubert@sun.com**





# The new ODF 1.2 Metadata Framework

**Svante Schubert**  
Software Engineer  
Sun Microsystems Inc.



## Why am I here?

- For introducing metadata feature & possibilities
  - > Member of the OASIS OpenDocument Format (ODF) TC
  - > Co-editor of the OASIS ODF Metadata Specification
  - > Member of Sun Microsystems StarOffice / OpenOffice.org team since 1999
  - > OpenOffice.org XML project co-lead

2

Page 2

- Co-lead of OpenOffice XML project
- Responsible for the XML based filters (added new XSLT2 support)

## Metadata Basics

- What is metadata?
  - > “Metadata is data about data”
- Why do I need metadata?
  - > Classify / 'Tag' your data
  - > Make data machine interchangeable
  - > Bridge the semantic gap
  - > ...

1) <http://en.wikipedia.org/wiki/Metadata>

3

Page 3

The semantic gap characterizes the difference between two descriptions of an object by different linguistic representations,

Make metadata interchangeable with other applications  
Reuse the existing Semantic Web Technology Stack

## Why using Metadata?

- What is metadata?
  - > Labels to identify/categorize your data
  - > Related data
- Why do I need metadata?
  - > Metadata makes your data interpretable by other applications

## Potential Interoperability Savings..

- *Information Integration in the Enterprise*, Philip A. Bernstein and Laura M. Haas, CACM, September 2008, pp. 72-79 (source for IT budgets (source for interoperability costs)
- Untapped Assets: The \$3 Trillion Value in US Enterprise Documents  
<http://www.brightplanet.com/images/stories/pdf/documentsvalue.pdf>
- NIST Strategic Planning Summary,  
[http://www.nist.gov/director/planning\\_summary\\_strategic\\_planning\\_study\\_results.html](http://www.nist.gov/director/planning_summary_strategic_planning_study_results.html)

<http://europa.eu.int/idabc/3761>

RDF start by Government..

Common spread RDF authoring application (even Xforms enabled)

National Institute of Standards and Technology

## EU Public Key Interoperability Areas

- Income taxes: declaration / notifications
- Job search services by labour offices
- Personal documents (passport and driving licence)
- Car registration (new, used and imported cars)
- Application for building permission
- Declaration to the police (e.g. in case of theft)
- Public libraries (catalogues, search tools)
- Certificates (birth, marriage): request and delivery
- Enrolment in higher education / university

6

Page 6

<http://europa.eu.int/idabc/3761>

RDF top down approach  
start with Government..

Common spread RDF  
authoring application  
(even Xforms enabled)

## EU Public Key Interoperability Areas (Businesses)

- Social contribution for employees
- Corporation tax: declaration, notification
- VAT: declaration, notification
- Registration of a new company
- Submission of data to statistical offices
- Customs declarations
- Environment-related permits (including reporting)
- Public procurement

7

[http://europa.eu.int/idabc/  
3761](http://europa.eu.int/idabc/3761) <sup>Page 7</sup>

## Do you know ODF?

- The OASIS / ISO standard for office documents (2005/06)
- The document format of many office applications

**OASIS**

Open Document Format for Office  
Applications (OpenDocument) v1.0

OASIS Standard, 1 May 2005



International  
Organization for  
Standardization

**ISO/IEC 26300**



## Do you know ODF?

- Origin from OpenOffice.org's default format
- A zipped package of XML and related files (image, sound, user files)

**OASIS**

Open Document Format for Office  
Applications (OpenDocument) v1.0

OASIS Standard, 1 May 2005

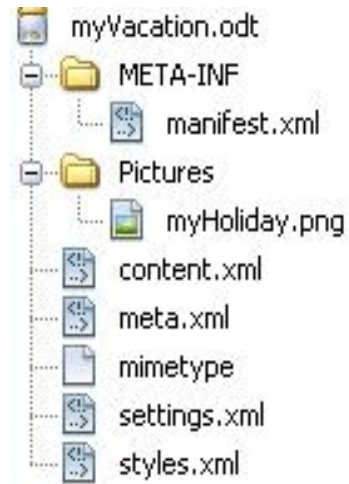


International  
Organization for  
Standardization

**ISO/IEC 26300**

## ODF Basics - Package & Files

- ISO standardized default content (as shown, but NOT Picture folder)
- Manifest as an Inventory / "table of content"
- Any user content...



10

## Existing Metadata Standards

### Resource Description Framework (RDF)

- Resources
  - > Unique identification by unicode enabled URI - Internationalized Resource Identifier (IRI)
  - > Described by RDF statements

### Internationalized Resource Identifiers (IRIs)

Page 11

Internationalized Resource Identifiers (IRIs) are a new protocol element, a complement to URIs [RFC2396]. An IRI is a sequence of characters from the Universal Character Set (Unicode/ISO10646). There is a mapping from IRIs to URIs, which means that IRIs can be used instead of URIs where appropriate to identify resources.

(<http://www.w3.org/International/O-URL-and-ident.html>)

Web Philosophy - Localisation vs. Semantic Philosophy – Identificaiton

<http://sun/employees/svante> ex:isPresenting  
<http://2007.xmlconference.org/public/schedule/speaker/579>

## Existing Metadata Standards

### Resource Description Framework (RDF)

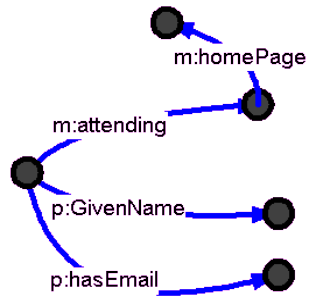
- RDF Statements -  
Triple: subject + predicate + object
  - > <http://dbpedia.org/resource/Beijing>  
<http://dbpedia.org/ontology/PopulatedPlace#populationTotal>  
"17430000" .

Based on <http://dbpedia.org/page/Beijing> searched by <http://sindice.com>

12

# Existing Metadata Standards

RDF graphs...

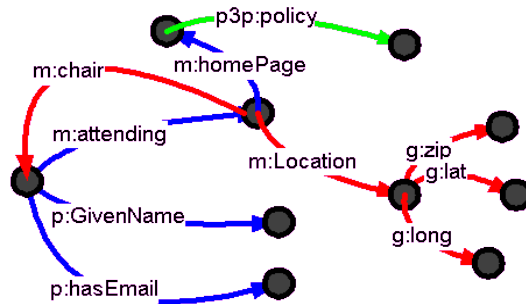


Based on <http://www.w3.org/2000/Talks/1206-xml2k-tbl/slide6-0.html>

13

# Existing Metadata Standards

...superimpose



Based on <http://www.w3.org/2000/Talks/1206-xml2k-tbl/slide7-0.html>

14

## Existing Metadata Standards

### Web Ontology Language (OWL)

- Ontology (from Greek)
  - > Onto (being)
  - > Logia (written/spoken discourse)
- Description of entities/concepts and their relations
  - > Like OO language using classes, properties, etc.
- ODF 1.2 includes OWL package description
  - > OWL Classes of pkg:Document, pkg:File, odf:Element

15

OWL instead of WOL like the name of the <sup>Page 15</sup> OWL from Winnie Poh. Symbol of Wisdom, nice marketing.. ;-)

## Describing ODF content with RDF

- pkg:Document
- pkg:Element
- pkg:File
- pkg:hasPart
- pkg:idref
- pkg:MetadataFile
- pkg:mimeType
- pkg:path
- xhtml:about
- xhtml:content
- xhtml:datatype
- odf:ContentFile
- odf:Element
- odf:prefix
- odf:StylesFile
- odf:suffix

16

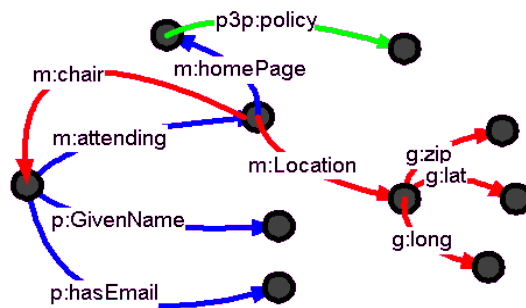
Page 16

Our own RDF Vocabulary to describe our ODF document(s) in the ODF Package.



## Using RDF/XML standard in ODF

- How connect graph from RDF/XML to ODF Package and ODF XML files?
  - > RDF/XML not compatible to XPATH
  - > Data has to be mapped to Lingua Franca of applications

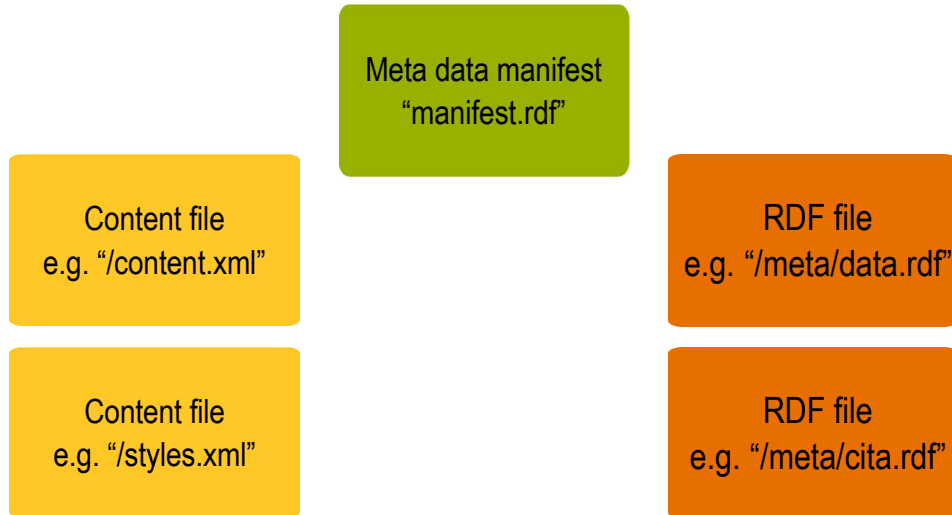


Based on <http://www.w3.org/2000/Talks/1206-xml2k-tbl/slide7-0.html>

17

# Describing ODF with RDF

## Metadata Files in the Package



18

## Metadata Model of ODF 1.2

### Metadata Files in the Package

- Content files (e.g. content.xml, styles.xml)
  - > About 50 ODF elements with xml:id attribute
- Metadata manifest (manifest.rdf)
  - > Heart of metadata model
  - > Lookup index for application interested in metadata
- User RDF/XML files
  - > Metadata file possibly from an office extension

# Metadata Model of ODF 1.2

## Metadata Files in the Package

Content files  
e.g. "/content.xml"

```
<table:table xml:id="someID">
  ... Hospital Doctor Duty List ...
</table:table>
```

Meta data manifest  
"/manifest.rdf"

```
<odf:Element rdf:about="content.xml#someID"/>
```

RDF files  
e.g. "/meta/data.rdf"

```
<odf:Element rdf:about="./content.xml#someID">
  <ex:workingHoursOf>
    <med:Doctors rdf:about="John.Example@somewhere.org">
      <med:fieldName xml:lang="en">Neurologist</med:fieldN.
```

## Metadata Model of ODF 1.2

### In Content Metadata

- If metadata is equal text/visual data
- Reason: No data duplication
- Used by ODF 5 elements:
  - > Bookmark start - <text:bookmark-start>
  - > Heading - <text:h>
  - > Metadata text - <text:meta>
  - > Paragraph - <text:p>
  - > Table cell - <table:table-cell>

21

# Metadata Model of ODF 1.2

## In Content Metadata

Content files  
e.g. "/content.xml"

```
<text:p>The doctor's name was  
<text:meta xhtml:about="John.Example@somewhere.org"  
xhtml:property="http://xmlns.com/foaf/0.1/name">  
Dr. J. Example</text:meta>
```

RDF files  
e.g. "/meta/data.rdf"

```
<med:Doctor rdf:about="John.Example@somewhere.org">  
<med:hasPatient>  
<med:Patient rdf:about="http://hospital-DB/patients/ID007">
```

## The RDF Application as Field User

### Metadata Text Field

- Content of field “text:meta-field” based on metadata
  - > Appears within paragraph
  - > Holds any paragraph content  
(E.g. Citation “According to [2]”)

```
<text:p>According to  
  <text:meta-field xml:id="someID">  
    <text:style text:style-name="s1">[2]
```

23

## Metadata support in OOO 3.1

- Support of metadata in the ODF 1.2 package
- API for metadata extension developers
  - > Providing OOO API to access metadata
  - > Wrapping existing opensource tool:  
Redland RDF Library – librdf
- Possibility of a generic metadata extension
  - > Import / Create your own RDF vocabulary
  - > Relate vocabulary to ODF content by GUI



## Metadata support in OoO 3.1

- Metadata on different Documents in a Package
  - > Multiple documents in an ODF package
  - > Metadata relative to document
  - > Metadata concept reusable by ODF package implementors

## Authoring with Metadata in OOO 3.1

- Scenarios with Embedded Objects
  - > Insert Embedded Object
  - > Extract Embedded Object
  - > Remove Embedded Object

## Authoring with Metadata in OOo 3.1

- Handling of metadata on editing a text document
  - > Deletion of an entity
  - > Copy of an entity
  - > Cut of an entity
  - > Paste of an entity
  - > Edit of entity content
  - > Split of entity
  - > Merge of two entities

## Scenarios to add Metadata to ODF

- ODF Template with Metadata (e.g. formular)
- RDF application (e.g. Citation Extension) offers some GUI
- Explicitly by the user
  - > Similar to spell-checker but text pattern base (User have to agree on metadata)
  - > Generic Metadata Extension: User choose some RDF Vocabulary and relates vocabulary to ODF content by GUI

28

1) ONCE TEMPLATE PREPARED FOR THE USER (FORMULAR) (PLUGIN CREATOR) Page 28

NOTE: ODF 1.2 has XForms support

EU has many of those scenarios (no User Knowledge)

2) Plugin offers GUI (mostly domain specific plugin)

2a) Explicitly by the user (domain level) (PLUGIN CREATOR / OOo USER)

2b) PATTERN MATCHING (PLUGIN CREATOR / OOo USER)

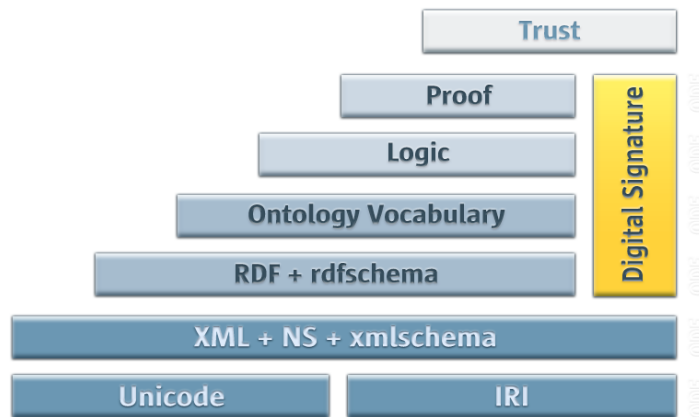
2c) Explicitly by the user (Generic / RDF level) (OOo USER)

Choose an RDF vocabulary (from server)

3) By a different application using the OOo 3.1 API (THE PROGRAMMER)

# Existing Metadata Standards

## Semantic Web Architecture

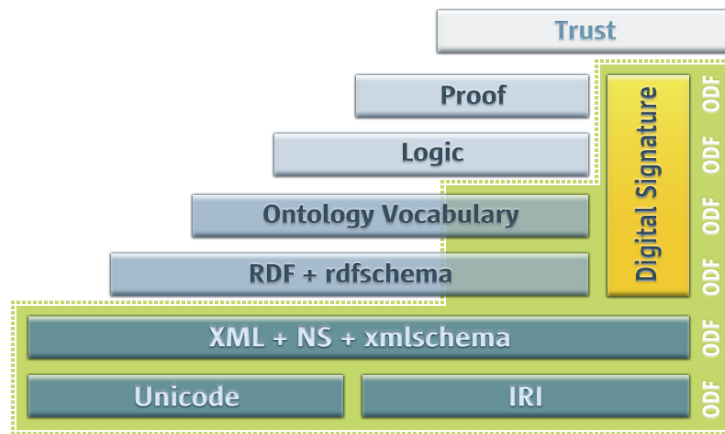


Based on <http://www.w3.org/2000/Talks/1206-xml2k-tbl/slide10-0.html>

29

# Existing Metadata Standards

## ODF in the Semantic Web



Based on <http://www.w3.org/2000/Talks/1206-xml2k-tbl/slide10-0.html>

30

## More Information

- Download latest ODF specification:
  - > Latest Draft: <http://www.oasis-open.org/committees/download.php/29446/08-09-24-ODF-Metadata-Change-Request.odt>
  - > ODF Metadata Examples: <http://www.oasis-open.org/committees/download.php/26846/08-01-16-ODF-Metadata-Examples-Committee-Draft.odt> (old)
  - > Upcoming adaptations: <http://lists.oasis-open.org/archives/office/200811/msg00006.html>
- OOo API (not final):
  - > <http://api.openoffice.org/docs/common/ref/com/sun/star/rdf/module-ix.html>



## **Thank you – Questions & Answers**

**Svante Schubert**  
Svante.Schubert@sun.com