eSciDoc Content Models using CMA and ECM

Frank Schwichtenberg

Fedora-UK&I&EU Meeting 2009, Oxford
Services of the eSciDoc Infrastructure

Policy Decision Point
Object Manager

- Store, manage, and retrieve basic objects like
  - Contexts
  - Containers
  - Items

- Object lifecycle methods (workflow)
- Convenient shortcut methods and administrative search
- Versioning
- Multiple, arbitrary metadata records
- Audit trails and PREMIS events
- Support for object relations and multiple ontologies
- Interface to the underlying repository system
Object Patterns

- Even though eSciDoc covers many disciplines, we need only very few generic **Object Patterns**
  - Describing the basic layout and structure of objects

- **Item**
  - an entity which consists of metadata and different representations of the object content (e.g., PDF, XML, and MS Word files)

- **Container**
  - an aggregation of Content Objects
eSciDoc Object Hierarchy

Organizational Unit

Context (administrative container)

Container

Container

Container

Items

Items

Items
Content Models

- Content Objects are further restricted by **Content Models**
  - Each concrete instance following one of the Object Patterns refers a Content Model with additional constraints

- Content models are user-definable
  - rule-based language? (e.g. Schematron)
  - content models may be discipline-specific

- Examples of such content models are images, videos, translations, and transcriptions.

- Discussions with Fedora and State and University Library (Denmark)
  - Fedora Content Model Architecture
  - "Enhanced Content Models for Fedora" by Asger Blekinge-Rasmussen
Using CMA and ECM

- Awaiting CMA, just a dummy CM Object in eSciDoc

- With CMA an eSciDoc CM which is mapped to Fedora CM is needed
  - eSciDoc Items are mapped to several Fedora objects
    → eSciDoc CM maps to several Fedora CMs

- Resources with sections in eSciDoc
  - Build in and user-defined relations

- Datastreams (incl. RELS-EXT) in Fedora
Using CMA and ECM

- eSciDoc defines a special resource: Content Model
- Proprietary syntax
- Well defined mapping to rules/restrictions in Fedora CMs
Sections of an eSciDoc Item

- Item
  - Properties
  - Metadata Records
- Components
  - SomeMODS
- Resources
  - MyRightsMD
- Components
- Resources
Mandatory Metadata Record

- eSciDoc

```xml
<escidocContentModel:md-record name="escidoc"
  schema-href="http://escidoc.mpg.de/metadata.../0.1/theSchema.xsd"/>
```

- CMA + ECM

```xml
<dsTypeModel ID="escidoc">
  <form MIME="text/xml"/>
  <extensions name="SCHEMA">
    <schema:schema type="xsd" datastream="escidoc_xsd"/>
  </extensions>
...
</dsTypeModel>
```
Sections of an eSciDoc Item

- Item
  - Properties
  - Metadata Records
  - Components
  - Resources
  - Version History
  - Relations
  - Some behavior (from Content Model)
• SDEF

```xml
<fmm:MethodMap name="Fedora MethodMap for Sdef">
  <fmm:Method operationName="mdt"/>
</fmm:MethodMap>
...
<foxml:datastream CONTROL_GROUP="R" ID="xslt">
  <foxml:datastreamVersion ID="xslt.0" MIMETYPE="text/xml">
    <foxml:contentLocation REF="http://localhost/.../mdt.xsl" TYPE="URL"/>
  </foxml:datastreamVersion>
</foxml:datastream>
```

• SDEP

```xml
<wsdl:operation name="trans">
  <http:operation
      location="http://localhost/.../SaxonServlet?\ style=(xslt)\&source=(escidoc)"/>
  ...
```
Sections of an eSciDoc Item

- Item
  - Properties
  - Metadata Records
  - Components
- Resources
  - Full Scan
  - Web Resolution
  - Relations
  - Resources Behavior (from Content Model)
Allowed Component Types

<!-- allowed component types; here just one -->
<rdf:subClassOf>
  <owl:Restriction>
    <owl:onProperty>
      rdf:resource="http://escidoc.de/...relations/component"/>
    <owl:allValuesFrom>
      <owl:Class>
        <owl:unionOf rdf:parseType="Collection">
          <!-- check for evaluation of unionOf in ECM -->
          <owl:Class rdf:resource="...component_FULLSIZE"/>
        </owl:unionOf>
      </owl:Class>
      <owl:allValuesFrom>
        <owl:Restriction>
      </owl:Restriction>
    </owl:Class>
  </owl:Restriction>
</rdf:subClassOf>
Mandatory Component Types

<!-- mandatory component -->
<rdfs:subClassOf>
  <owl:Restriction>
    <owl:onProperty rdf:resource="http://escidoc.de/...relations/component"/>
    <owl:someValuesFrom rdf:resource="...component_FULLSIZE"/>
  </owl:Restriction>
</rdfs:subClassOf>
Conclusion

- Open issues
  - Optional metadata records (resp. datastreams)
  - How far goes ECM?
    - OWL DL
  - Several CMs for one Fedora object (issue or advantage?)

- ECM is very important
- It is easy to map self-defined restrictions to Fedora Content Models
Thank you!

Questions?

Frank Schwichtenberg
frank.schwichtenberg@fiz-karlsruhe.de

http://www.escidoc.org