



Department of Homeland Security Metadata Center of Excellence

The DRM XML Schema

Michael C. Daconta, Metadata Program Manager
Andy Hoskinson, OMB FEA PMO
Joseph M. Chiusano, Booz Allen Hamilton
June 13, 2005

MCOE

METADATA CENTER OF EXCELLENCE





- ▶ **What is the DRM XML Schema?**
- ▶ **Purpose of the DRM XML Schema**
- ▶ **DRM XML Schema Structure**
- ▶ **DRM XML Instance Examples**
- ▶ **Schema Design Features, Issues, & Tradeoffs**
- ▶ **Conclusion**



What is the DRM XML Schema?



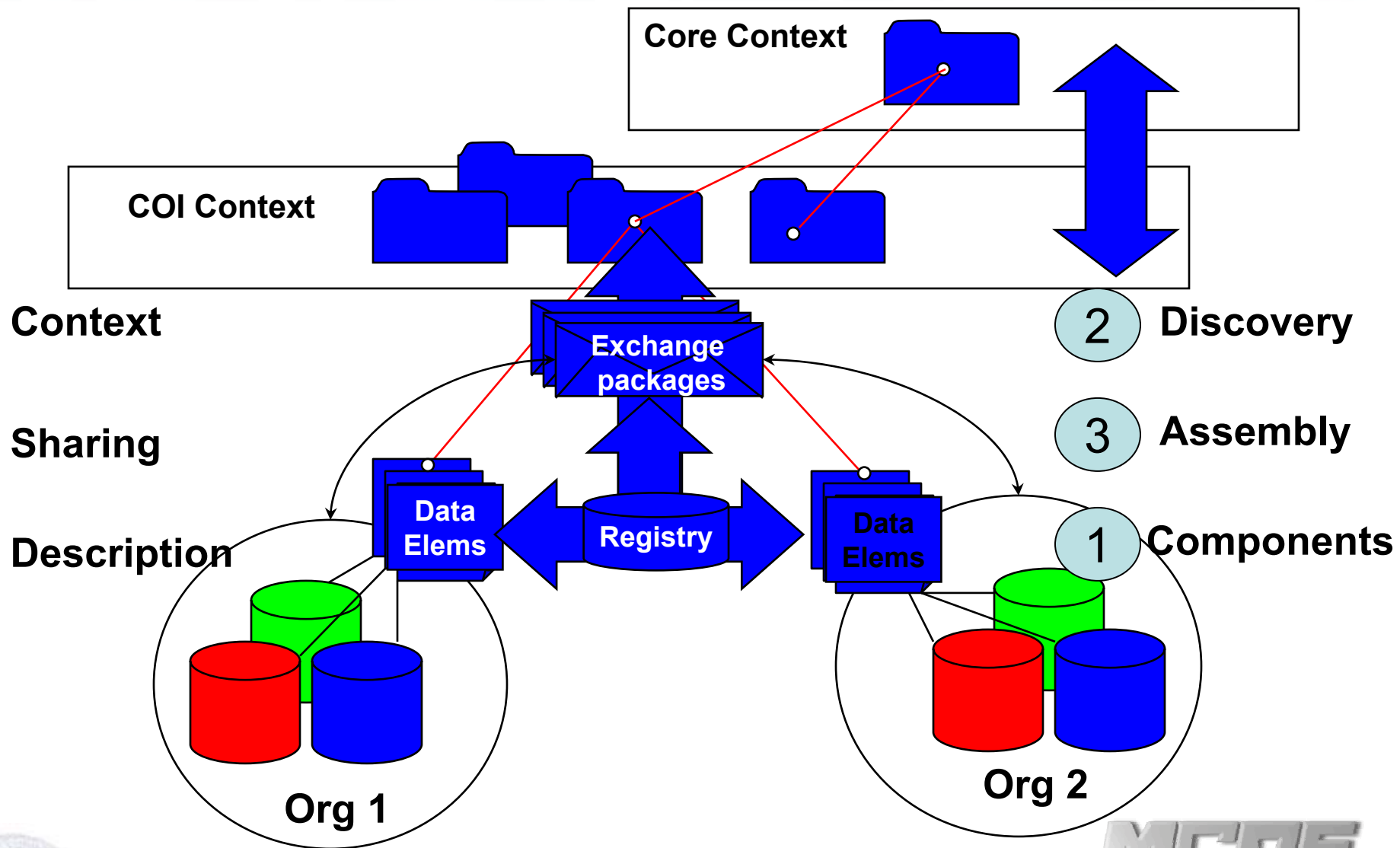
- ▶ The DRM XML Schema is a W3C XML Schema that serves as an abstract metamodel for the DRM
- ▶ It represents all 3 of the DRM's major standardization areas (Data Description, Data Sharing, Data Context)
- ▶ Federal agencies will create XML instances (documents) based on the DRM XML Schema that contain information for the agency pertaining to the 3 DRM areas
- ▶ Development of the DRM XML Schema at an early stage is enabling us to ensure that we are capturing the correct information for the DRM, in the correct format – as soon as possible



Purpose of the DRM XML Schema

- ▶ The DRM XML Schema will:
 - Support the DRM's primary use case of facilitation of interagency information sharing
 - Facilitate the inventory, cataloging, and discovery of information holdings as required by law and policy (OMB Circular A-130, Management of Federal Information Resources)
 - Support harmonization across the federal government of data artifacts, and establishment of authoritative data sources
 - Provide an open and well-documented standard to enable the organization and categorization of government information, in ways that are searchable, and interoperable, across agencies

DRM Primary Use Case: Interagency Information Sharing



DRM XML Schema Structure

- ▶ The DRM XML Schema has 3 major sections:

Section	Description
DataDescription	Provides a standard means for agencies to describe their data and data sources clearly, concisely, and unambiguously
DataSharing	Provides a standard means for describing interagency data exchanges and data sharing capabilities
DataContext	Provides a standard means for representing taxonomies that an agencies use to categorize their data

DRM XML Schema Structure



► “DataDescription” section – major subsections:

Subsection	Description
StructuredData	Data described via the E-R (Entity-Relationship) or class model
UnstructuredData	Data that is not described according to an E-R model, but is rather of a more free-form format, such as multimedia files or unstructured text
SemiStructuredData	Data that has characteristics of both structured and unstructured data
DataSources	Databases, systems, applications, or other electronic entities that produce and/or house data



DRM XML Schema Structure



► “DataSharing” section – major subsections:

Subsection	Description
ExchangePackages	Descriptions of specific information exchanges between organizations
DataAccessPoints	Endpoints (network or otherwise) providing an interface for querying data sources



DRM XML Schema Structure



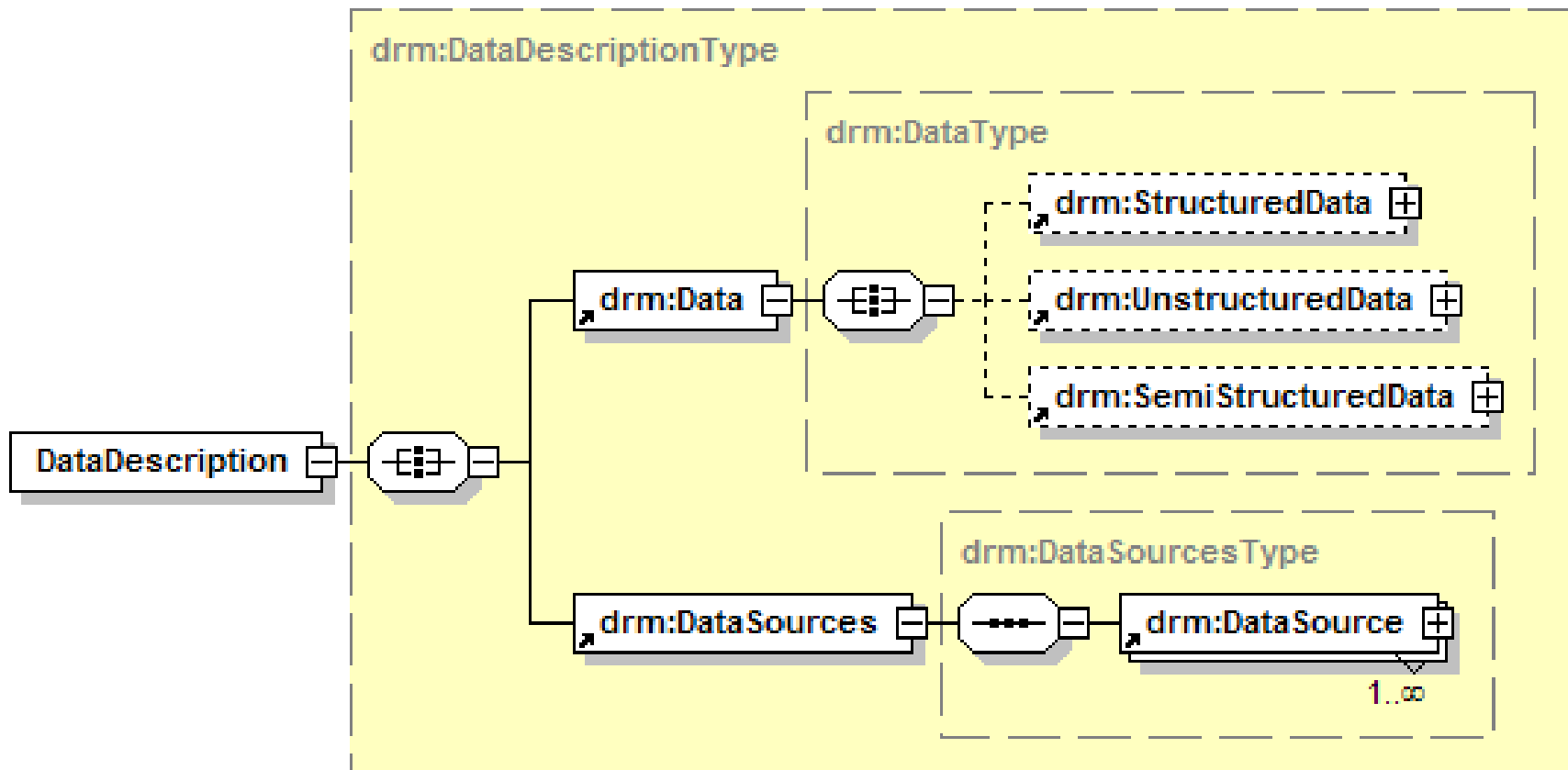
- ▶ “DataContext” section – major subsections:

Subsection	Description
Taxonomies	Hierarchical information models that define the scope of a knowledge domain
FEAMapping	Enables mapping of taxonomy nodes in a DRM instance to FEA reference model categories
DataStandards	Open standards that apply to the DRM submission. May also point to an XSLT stylesheet that transforms a document conforming to the particular standard to DRM XML syntax.

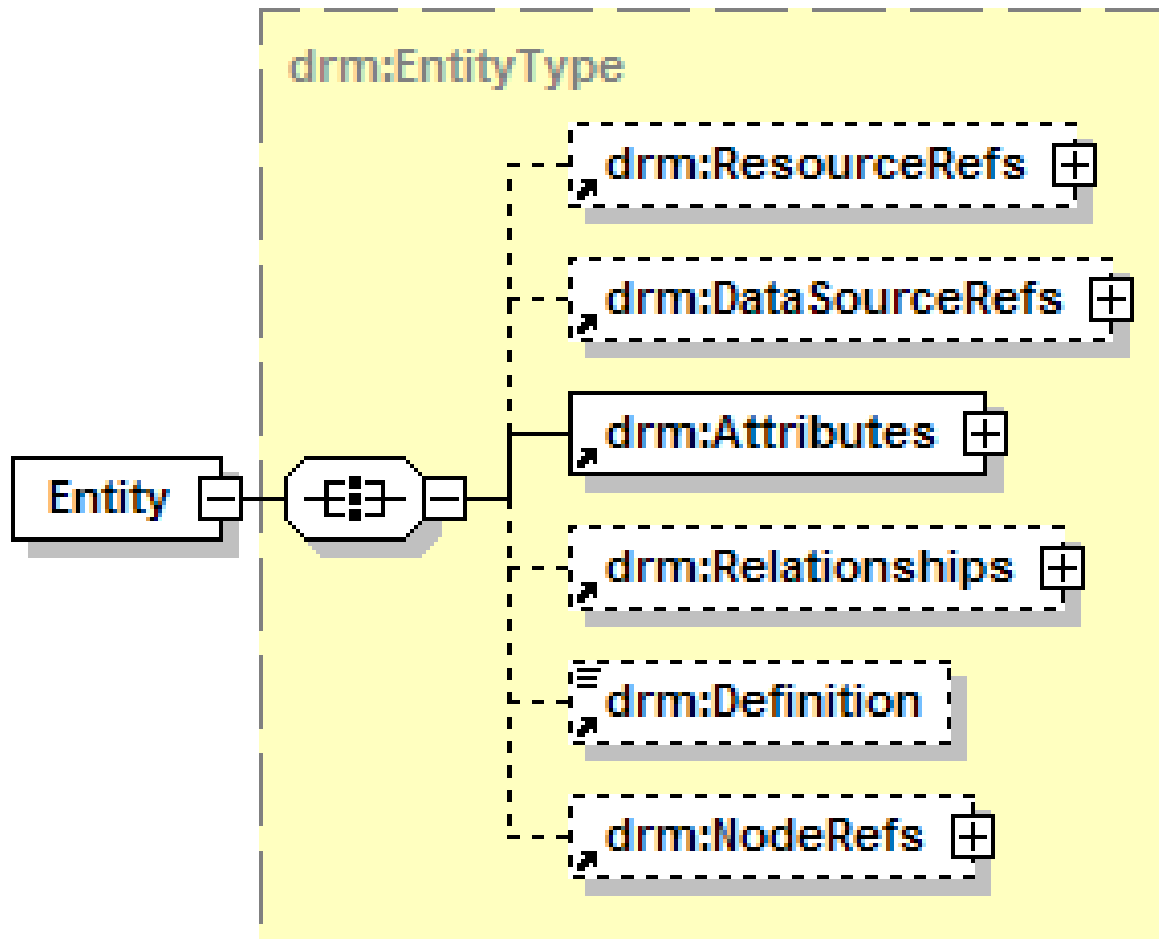
- ▶ Schema examples follow



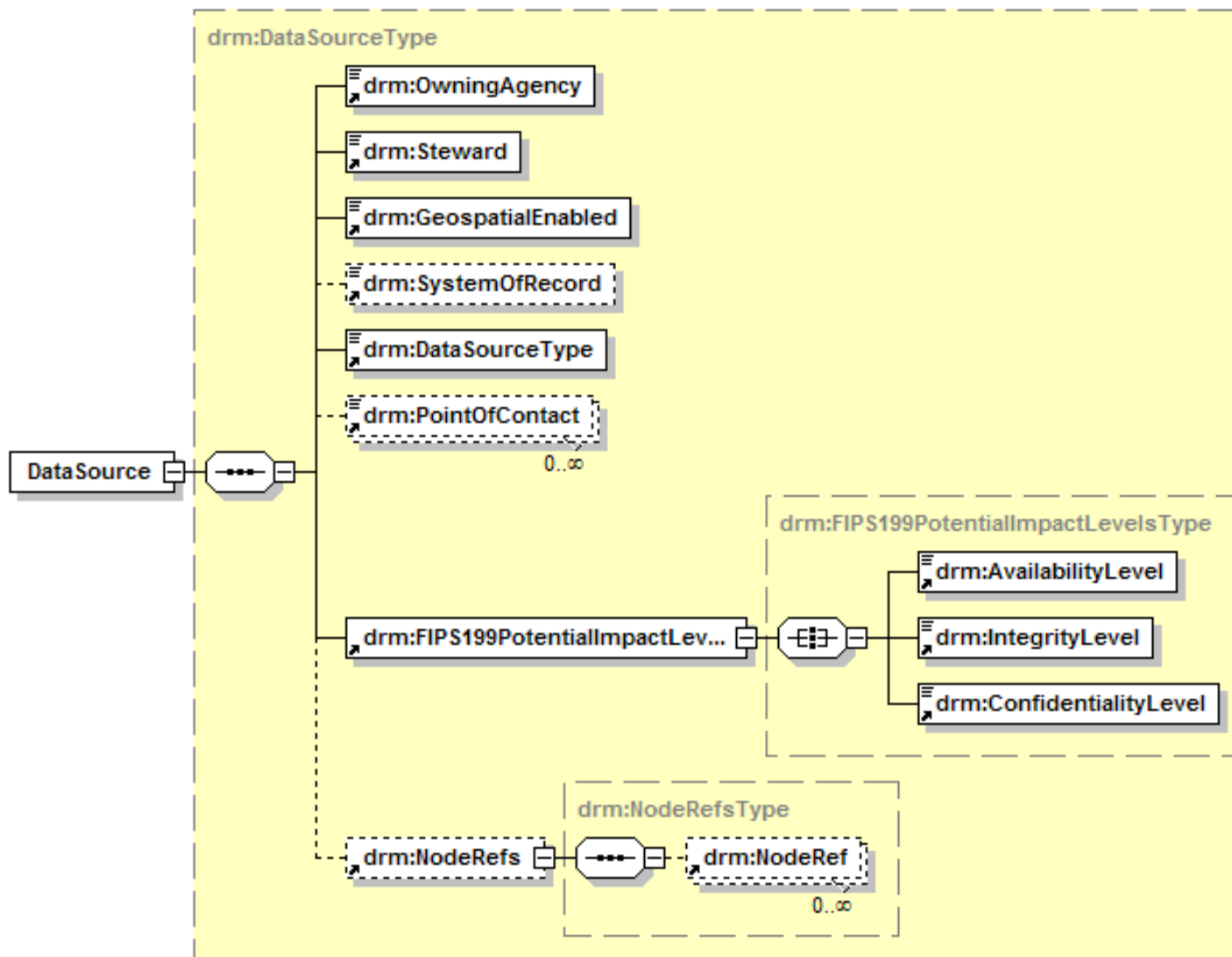
“DataDescription” Section



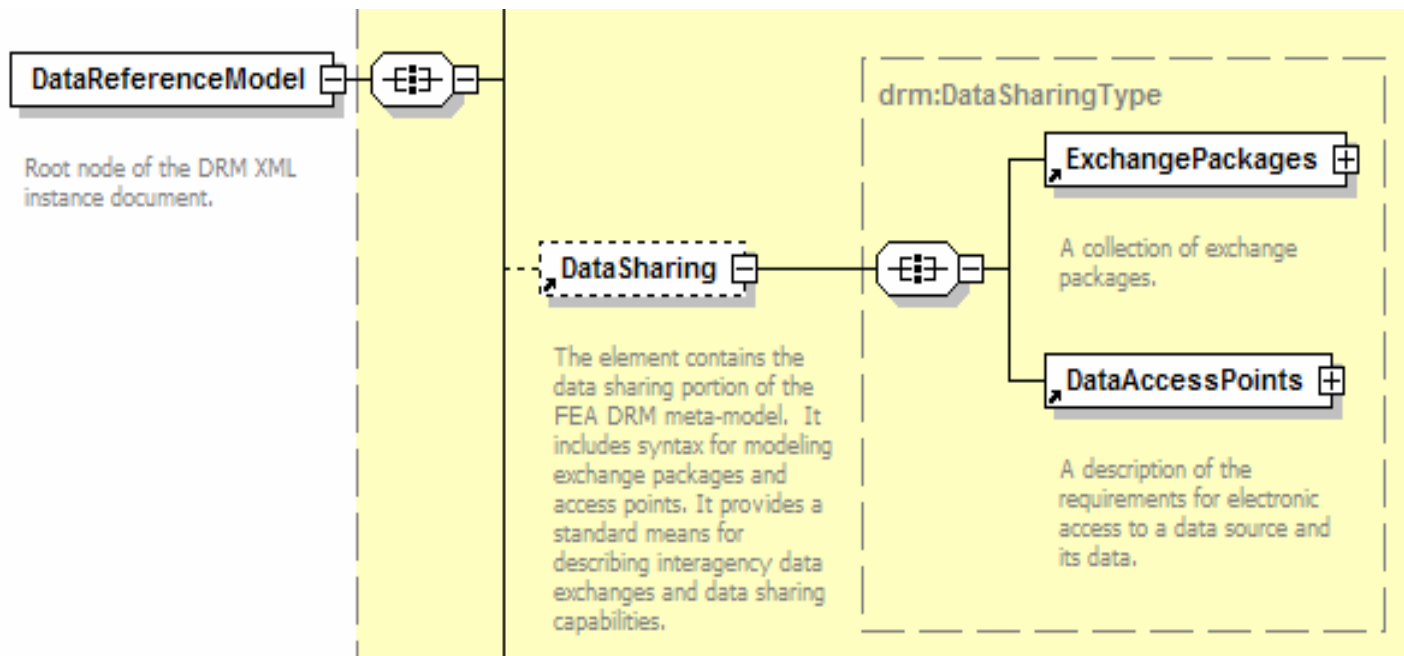
“Entity” Section



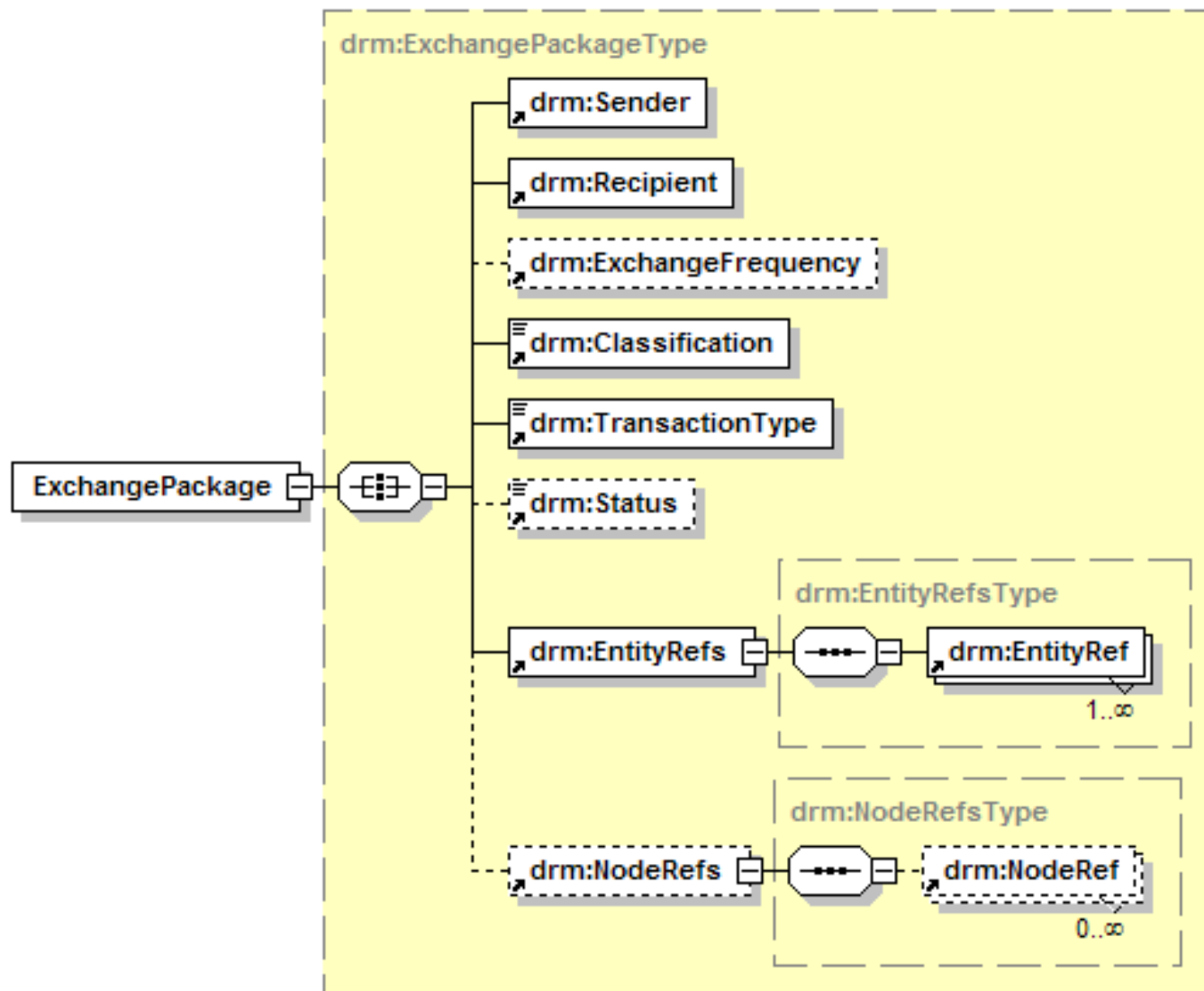
“DataSource” Section



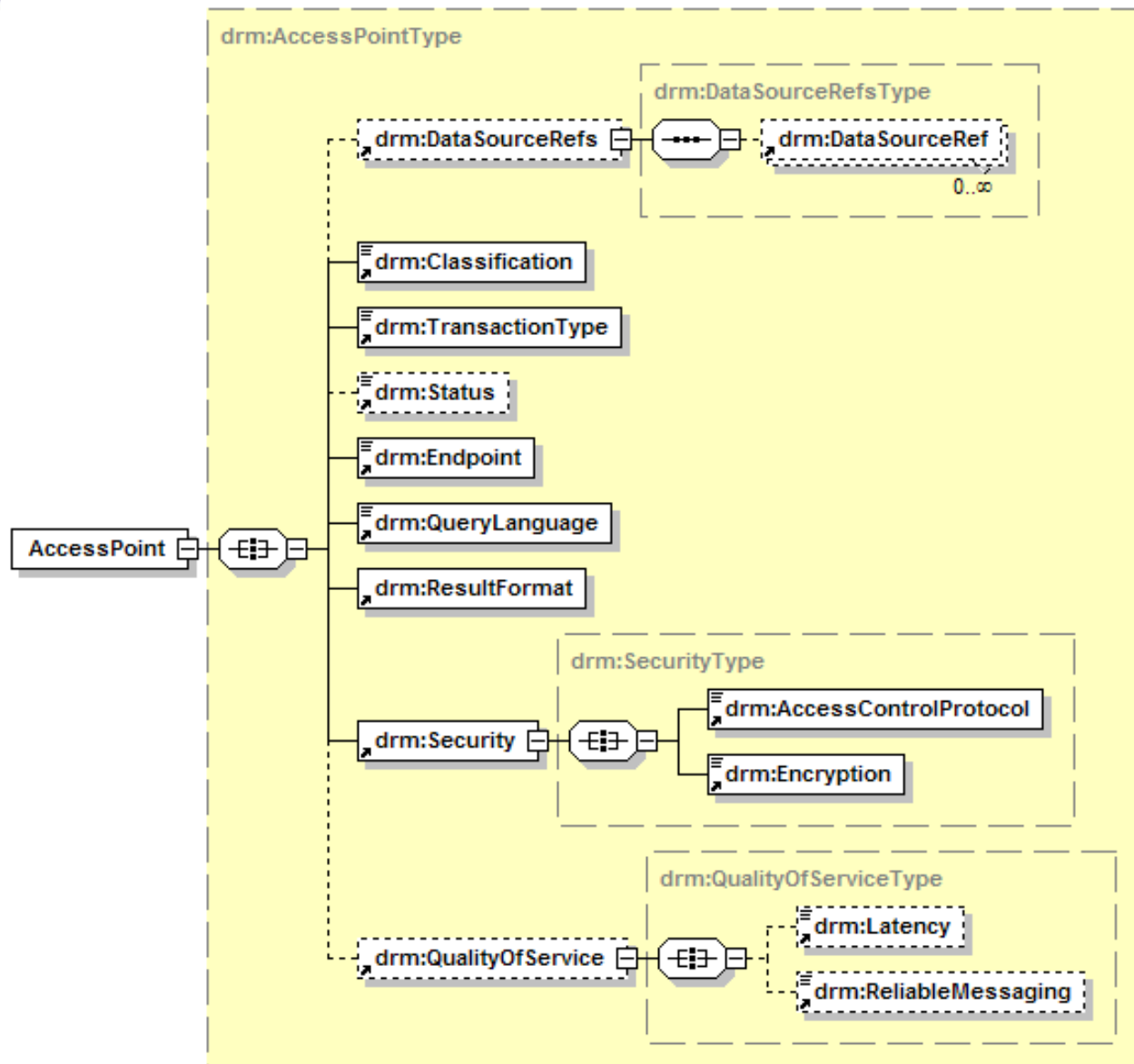
“DataSharing” Section



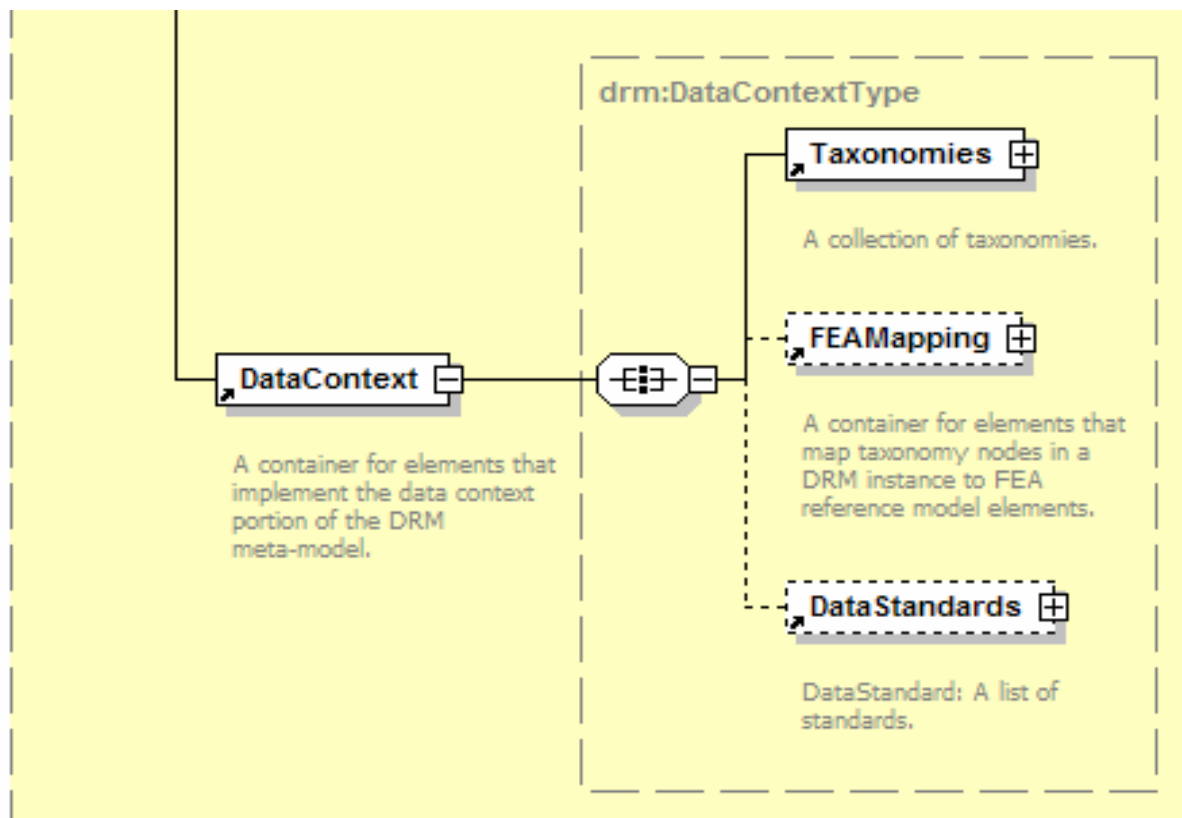
“ExchangePackage” Section



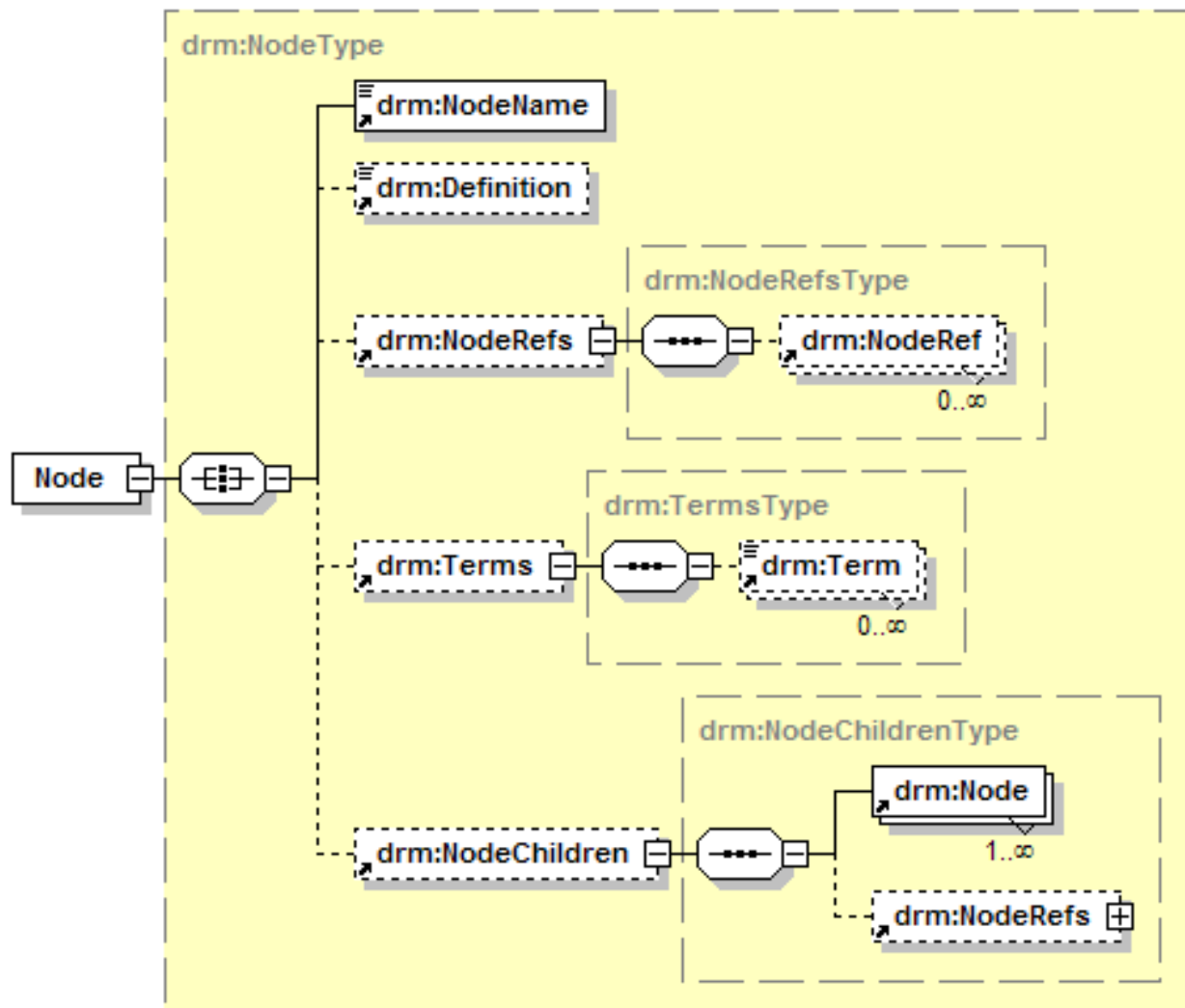
“AccessPoint” Section



“DataContext” Section



“Node” Section



DRM XML Instance Example: Entity



```
<drm:Entity rdf:id="BusinessArea" drm:name="FEA BRM Business Area">
  <drm:ResourceRefs>
    <drm:ResourceRef drm:type="partOf" rdf:idref="resource04" />
  </drm:ResourceRefs>
  <drm:DataSourceRefs>
    <!-- links this entity to a data source defined in the <drm:DataSources> section -->
    <drm:DataSourceRef drm:authoritativeSource="false" rdf:idref="data_source01"/>
  </drm:DataSourceRefs>
  <drm:Attributes>
    <drm:Attribute drm:keyType="primary key" drm:name="businessAreaID"
                  drm:datatype="xsd:integer"/>
    <drm:Attribute drm:name="BusinessAreaName" drm:datatype="xsd:string"/>
    <drm:Attribute drm:name="BusinessAreaDefinitionText" drm:datatype="xsd:string"/>
  </drm:Attributes>
  <drm:Relationships>
    <drm:Relationship drm:cardinality="1" drm:key="businessAreaID"
                    drm:name="Business Area to Business Line Association" >
      <drm:RelationshipTarget drm:key="refbusinessAreaID" drm:cardinality="unbounded"
                            rdf:idref="BusinessLine"/>
    </drm:Relationship>
  </drm:Relationships>
  <drm:NodeRefs>
    <!-- links this entity to a node defined in the <drm:Taxonomies> section -->
    <drm:NodeRef drm:type="partOf" rdf:idref="node06"/>
  </drm:NodeRefs>
</drm:Entity>
```



DRM XML Instance Example: DataSource

```
<drm:DataSource rdf:id="data_asset01" drm:href="https://www.feams.gov/"
    drm:name="Federal Enterprise Architecture Management System">
  <drm:OwningAgency>Office of Management and Budget</drm:OwningAgency>
  <drm:Steward>Federal Enterprise Architecture Program Management Office</drm:Steward>
  <drm:GeospatialEnabled>>false</drm:GeospatialEnabled>
  <drm:SystemOfRecord>>false</drm:SystemOfRecord>
  <drm:DataSourceType>Web service</drm:DataSourceType>
  <drm:PointOfContact>support@feapmo.gov</drm:PointOfContact>
  <drm:FIPS199PotentialImpactLevels>
    <drm:AvailabilityLevel>moderate</drm:AvailabilityLevel>
    <drm:IntegrityLevel>moderate</drm:IntegrityLevel>
    <drm:ConfidentialityLevel>moderate</drm:ConfidentialityLevel>
  </drm:FIPS199PotentialImpactLevels>
</drm:DataSource>
```

DRM XML Instance Example: Taxonomy

```
<drm:Taxonomy rdf:id="federal_irm">
  <drm:Nodes>
    <drm:Node rdf:id="node15" drm:nodeType="collection">
      <drm:NodeName>Strategic Planning</drm:NodeName>
      <drm:NodeChildren>
        <drm:Node rdf:id="node100" drm:nodeType="collection">
          <drm:NodeName>Information Policy</drm:NodeName>
          <drm:NodeChildren>
            <drm:Node rdf:id="node101" drm:nodeType="collection">
              <drm:NodeName>Clinger-Cohen Act</drm:NodeName>
            </drm:Node>
            <drm:Node rdf:id="node102" drm:nodeType="collection">
              <drm:NodeName>E-Government Strategy</drm:NodeName>
            </drm:Node>
            <drm:Node rdf:id="node103" drm:nodeType="collection">
              <drm:NodeName>Presidential Initiatives</drm:NodeName>
            </drm:Node>
          </drm:NodeChildren>
        </drm:Node>
        <drm:Node rdf:id="node100" drm:nodeType="collection">
          . . . .
        </drm:Node>
      </drm:NodeChildren>
    </drm:Node>
  </drm:Nodes>
</drm:Taxonomy>
```



- ▶ Several features have been included in the DRM XML Schema with efficiency and effectiveness in mind:
 - **Leveraging of existing vocabularies:** Use of Dublin Core metadata for resources, RDF identifiers
 - **External references:** Enable agencies to reference existing data artifacts that comply with one of a pre-specified set of open standards, in lieu of providing detailed information about those data artifacts in the DRM format
 - **Document interlinking:** For cases in which associations are made between DRM components, enables agencies to refer to information in existing (i.e. previously submitted) external DRM instances in a DRM XML document, rather than re-specifying the same information



Design Feature: Existing Vocabularies

- ▶ Dublin Core Version 1.1 is used to describe unstructured data (resources), as well as for submission metadata
 - **dc>Title**
 - **dc:Identifier**
 - **dc>Date**
 - **dc:Creator**
 - **dc:Format**
 - **dc:Description**
 - **dc:Source**
 - **dc:Type**
 - **dc:Publisher**
 - **dc:Contributor**
 - **dc:Language**
 - **dc:Relation**
 - **dc:Coverage**
 - **dc:Rights**
 - **dc:Subject**
- ▶ **rdf:id** is used for providing unique identifiers for components within a DRM XML instance, as well as for references to other components (inline or inter-document)



Design Feature: External References

► Current external references:

Section	External Reference Element	Potential Standard(s)
DataDescription	ExternalEntitiesRefs	<ul style="list-style-type: none">▪ OMG XML Metadata Interchange (XMI)▪ W3C XML Schema
	ExternalResourcesRefs	<ul style="list-style-type: none">▪ W3C Resource Description Framework (RDF)▪ RSS
DataSharing	ExternalExchange-PackageRef	<ul style="list-style-type: none">▪ W3C Web Services Description Language (WSDL)
DataContext	ExternalTaxonomyRefs	<ul style="list-style-type: none">▪ W3C Web Ontology Language (OWL)▪ W3C Simple Knowledge Organization System (SKOS)

Design Feature: Document Interlinking

- ▶ Examples of associations to which document interlinking can apply:

From	To*	What Association Represents
Entity	Resource	A resource (e.g. spreadsheet, document) that contains one or more instances of an entity
Entity	DataSource	A data source that contains one or more instances of an entity, and which may be an authoritative data source for that entity
Entity	Node	A taxonomy node with which instances of an entity are associated (i.e. by which it is categorized)
ExchangePackage	Entity	An entity that is represented in an information exchange
Node	Node	A node that is a parent or child of another node in a taxonomy

*“To” values are externally defined

Design Feature: Document Interlinking

- ▶ Example: Associating an entity with an externally defined resource and data source



Design Issues & Tradeoffs

- ▶ A “Red Team” recently convened to discuss design issues and tradeoffs
 - Comprised of a subset of the DRM Working Group
- ▶ The Red Team worked through a series of design issues
 - More are pending



Design Issues & Tradeoffs

- ▶ The following is a sample of issues and resolutions from the Red Team

Issue	Resolution
Need federated registries to which DRM XML instances are registered	Address in DRM Data Management Strategy.
Entities vs. attributes – can an attribute have attributes?	No. If something has attributes, it cannot be an attribute. However, there may be cases in which one agency may represent a concept as an entity while another may represent it as an attribute, per their business case. If the 2 agencies participate in a Common Operating Picture, the entity and attribute representations must be harmonized.
Need “submission metadata” in the DRM XML Schema.	Added submission metadata – e.g. submission date, submission time, submitting agency, etc.

Design Issues & Tradeoffs

► Issues and resolutions (cont'd):

Issue	Resolution
How can globally unique identifiers for components within DRM XML instances be ensured?	It will be the responsibility of agencies to ensure that their identifiers are globally unique for their agency. Government-wide global uniqueness is a related issue that needs to be addressed.
Need a <Definition> element for both entities and attributes, in addition to taxonomy nodes.	Added <Definition> element for entities and attributes. This element is optional for both entities and attributes, unless it is known that the entity and/or attribute will be harmonized within a Community of Interest (COI) at a later time – in which case it is mandatory. Submitters also have the option of referring to an inline or external taxonomy node containing a term rather than providing a definition directly with an entity or attribute.

Conclusion

- ▶ The DRM XML Schema is a metamodel that enables consistent description of structure, exchange, access and categorization
- ▶ This draft release of the XML Schema implements best practices in modeling, linking and modularity
- ▶ Agencies will populate the schema differently to support different use cases
- ▶ The Red Team is working through design issues and will post resolutions to the public wiki



Conclusion

- ▶ We need your support!
 - Please have your technical staff review the DRM XML Schema and provide comments!
 - Vendors, work with us to support it in your products!
 - Together ... we can achieve significant gains in data description, sharing and discovery!

- ▶ All are invited to provide comments on this XML schema at:

<http://colab.cim3.net/cgi-bin/wiki.pl?DataReferenceModel>



Questions?



MCCOE

METADATA CENTER OF EXCELLENCE

