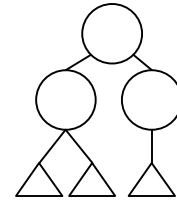
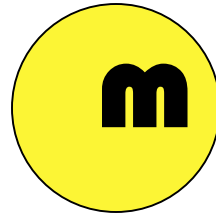
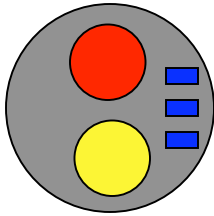


Digital Objects, Data Models, and Surrogates



Carl Lagoze
Computing and Information Science
Cornell University



Pathways Project



- NSF grant number IIS-0430906
- <http://www.infosci.cornell.edu/pathways/>
- PIs: Carl Lagoze, Sandy Payette, Herbert Van de Sompel, Simeon Warner
- Research Participants: Lyudmila Balakireva, Jeroen Bekaert, Xiaoming Liu, Chris Wilper, Zhiwu Xie



Lots of types of digital objects...

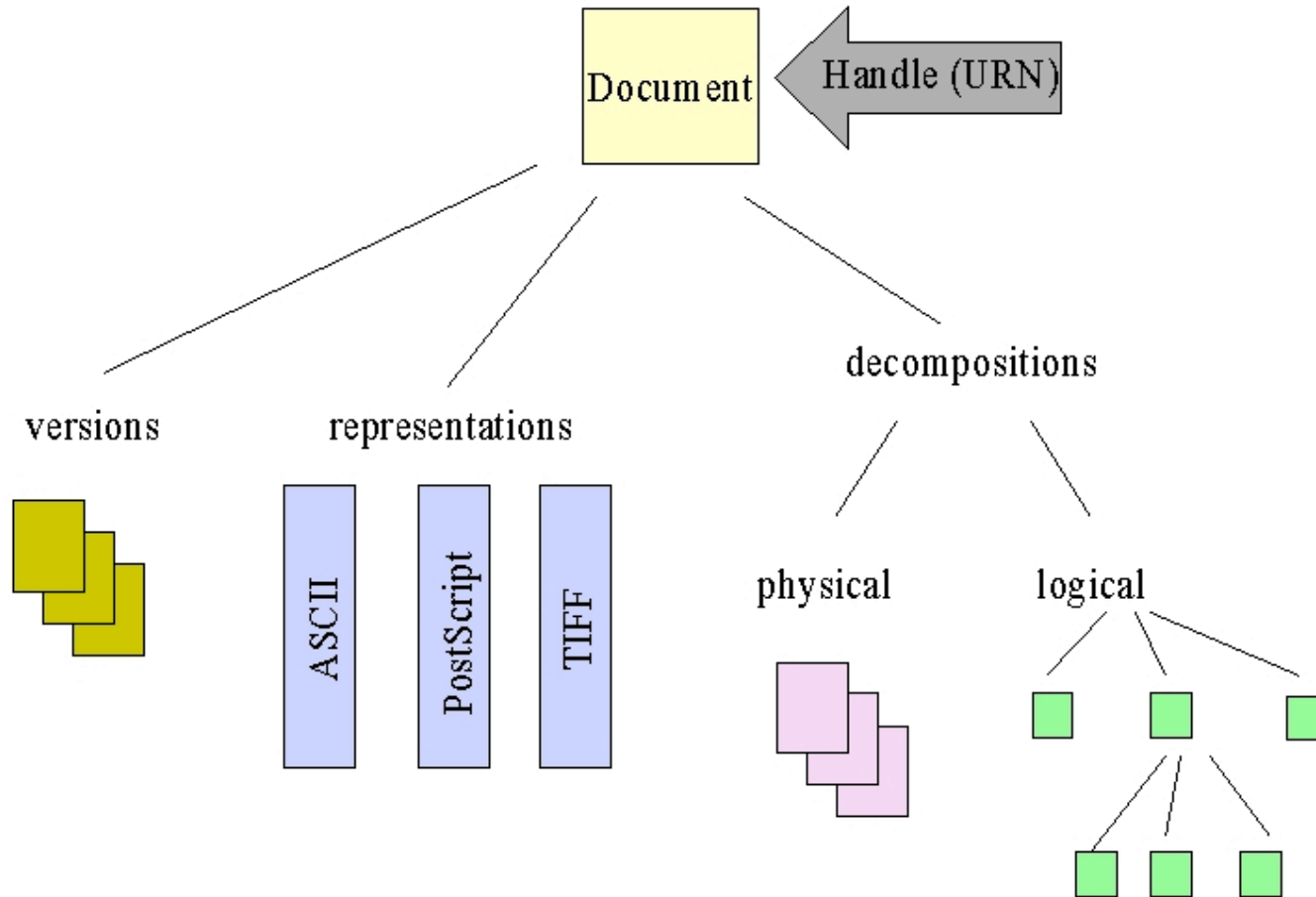
Lot's of data models....



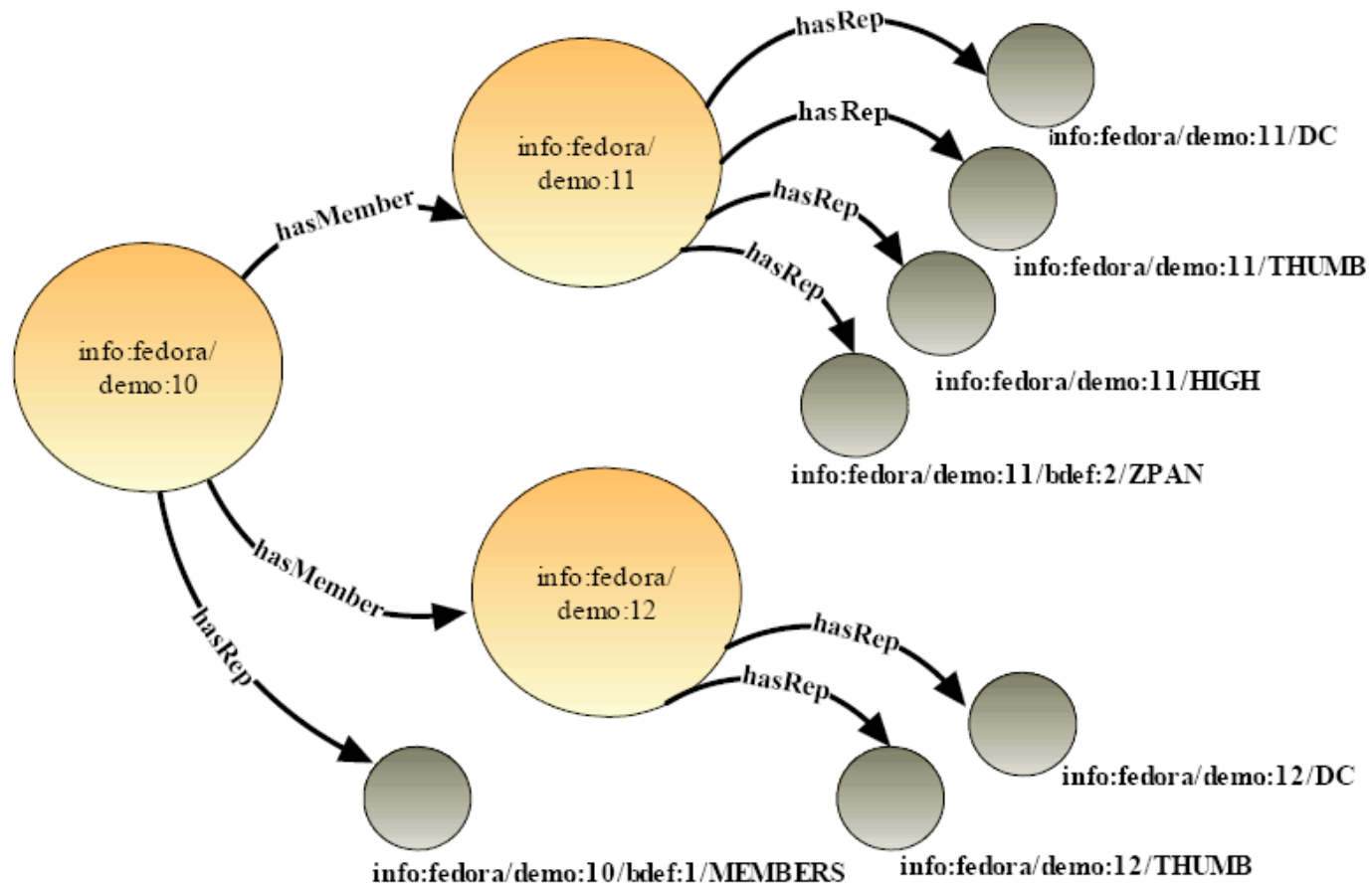
Carl Lagoze
Cornell Computing and Information Science



Dienst



Fedora

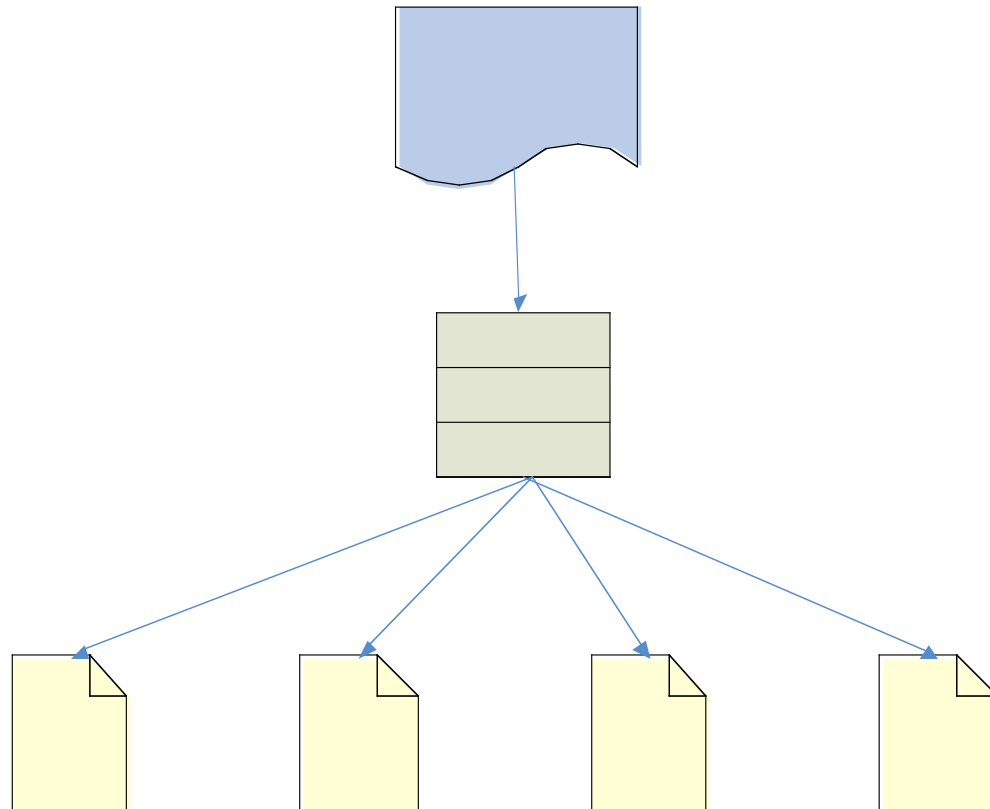


OAI-PMH

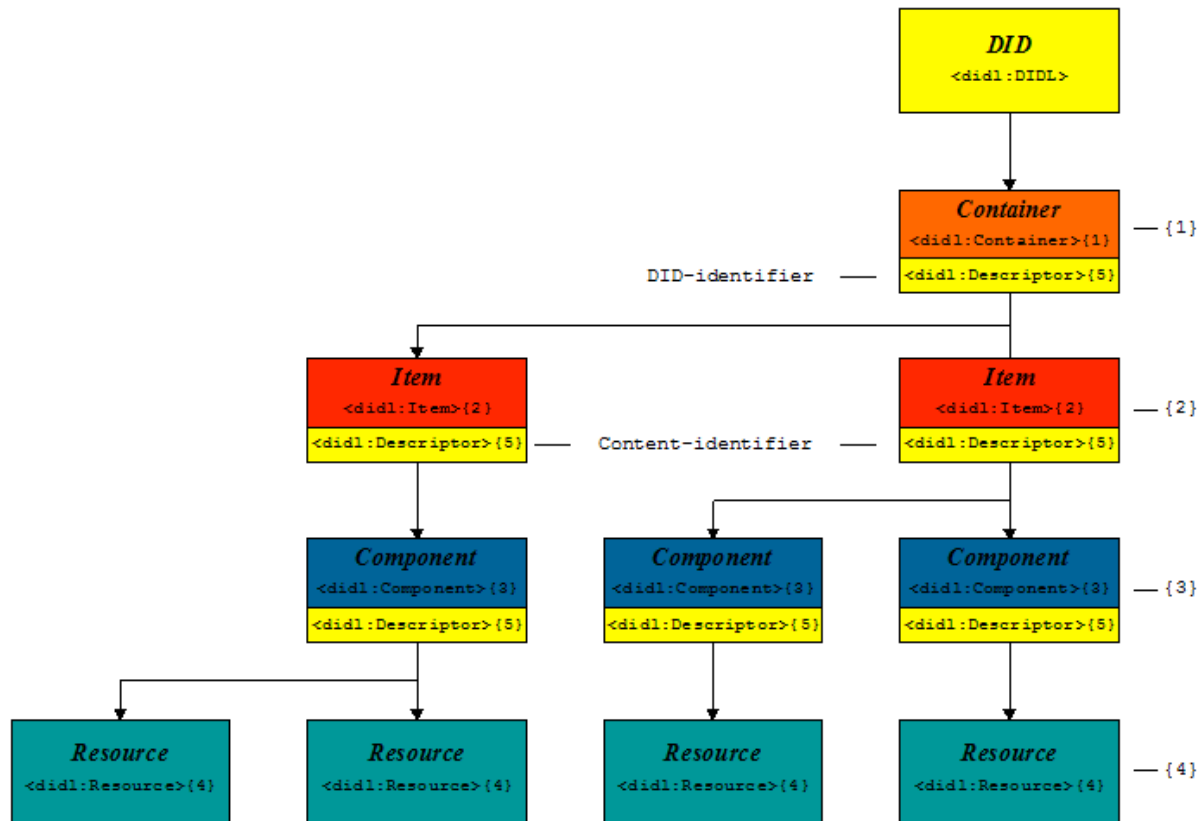
Resource
Abstract "content"

Item
*Available structured
data about the resource*

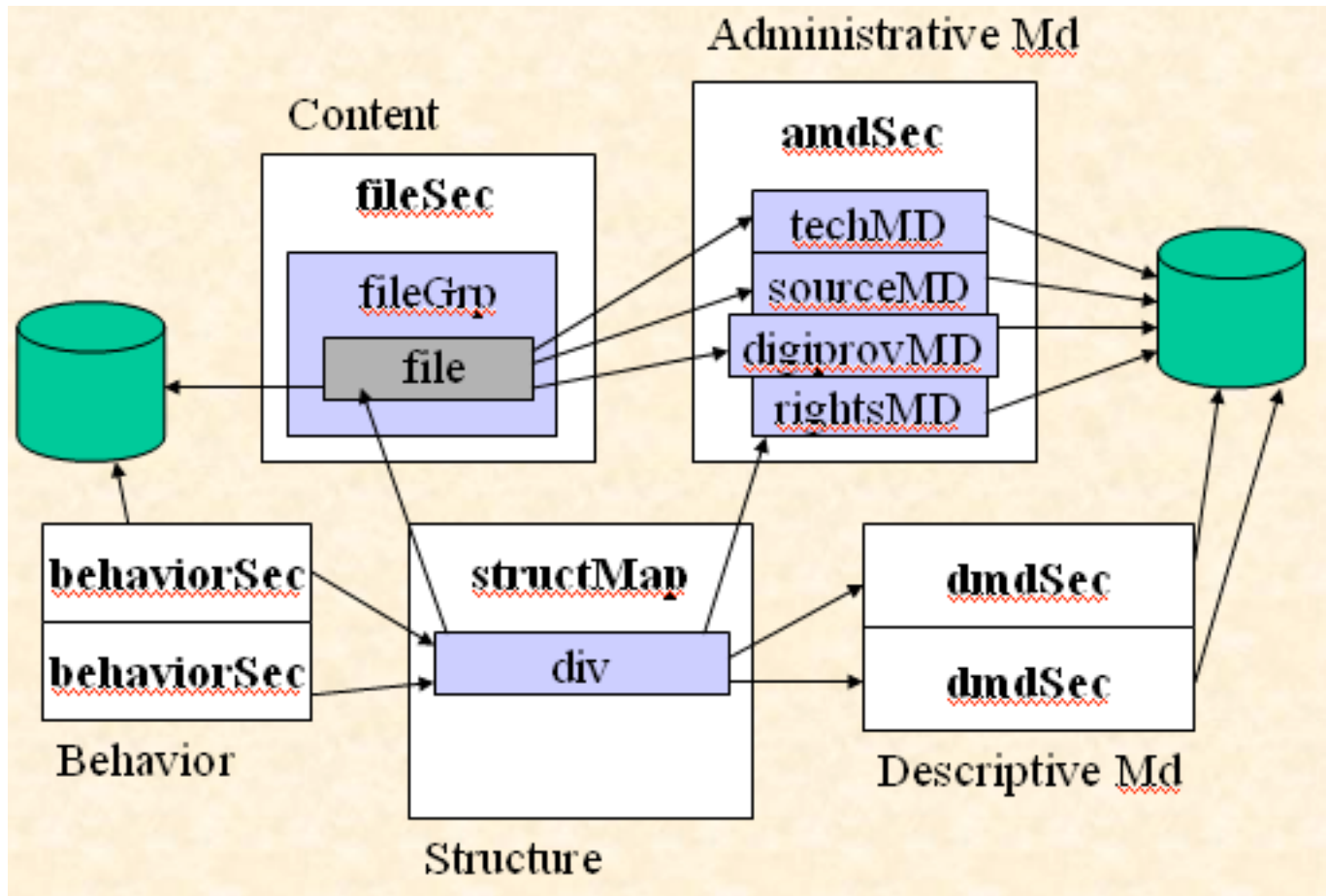
Record
*Disseminated structured
data about the resource*

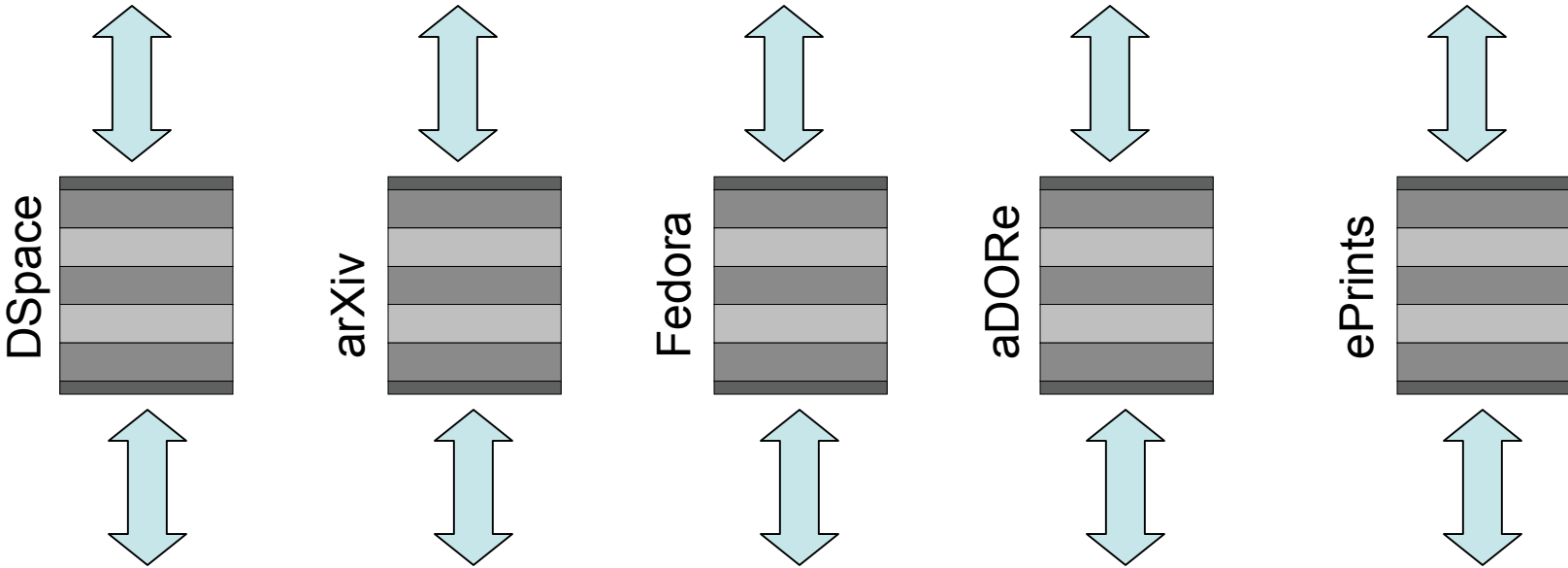
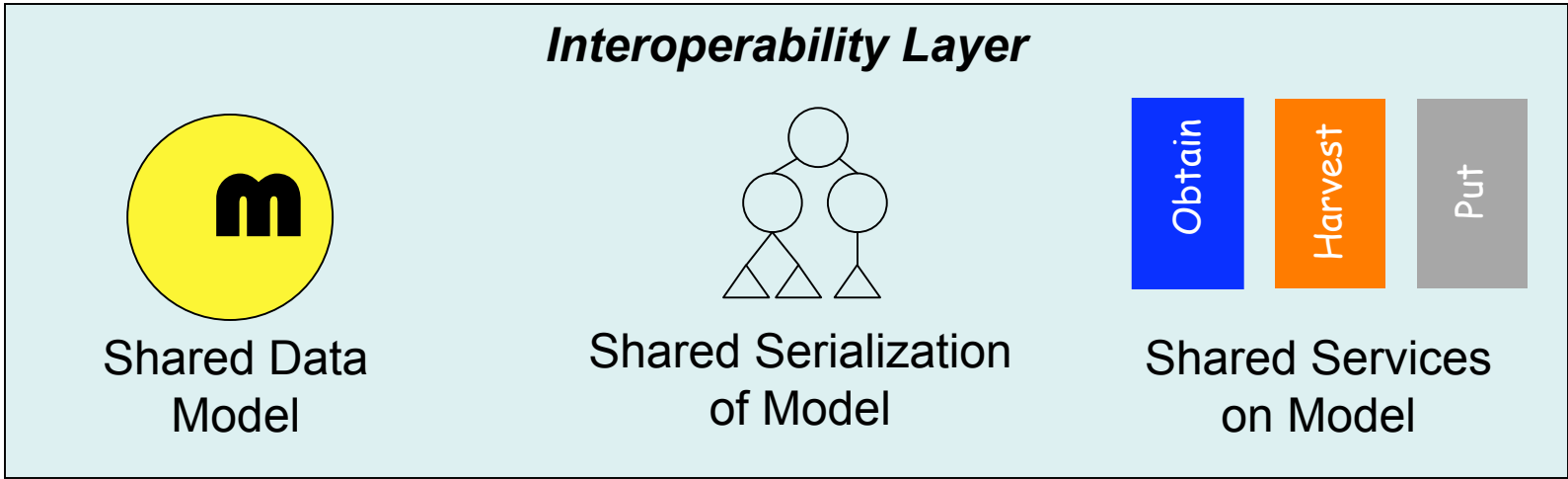


MPEG-21 DID



METS



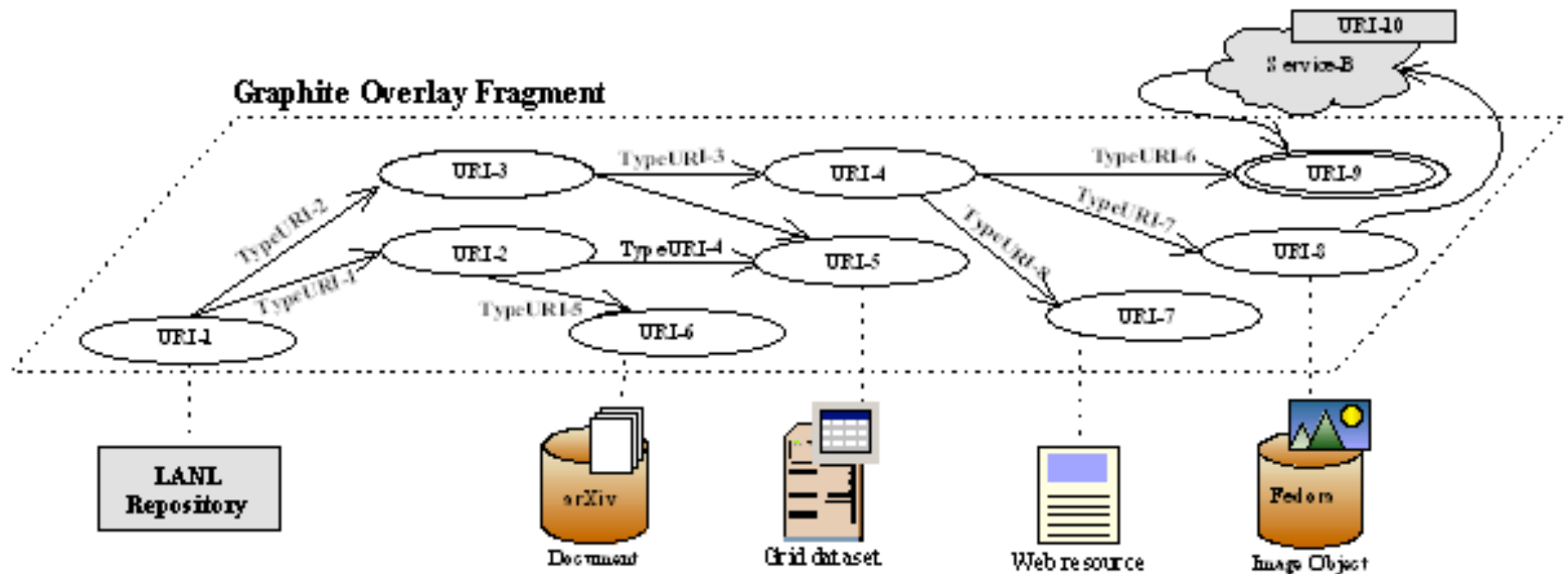


Individual Models and Interfaces

First pass: Graphite Model

Graph-based abstraction as common reduction across heterogeneous models (Payette and Erickson)

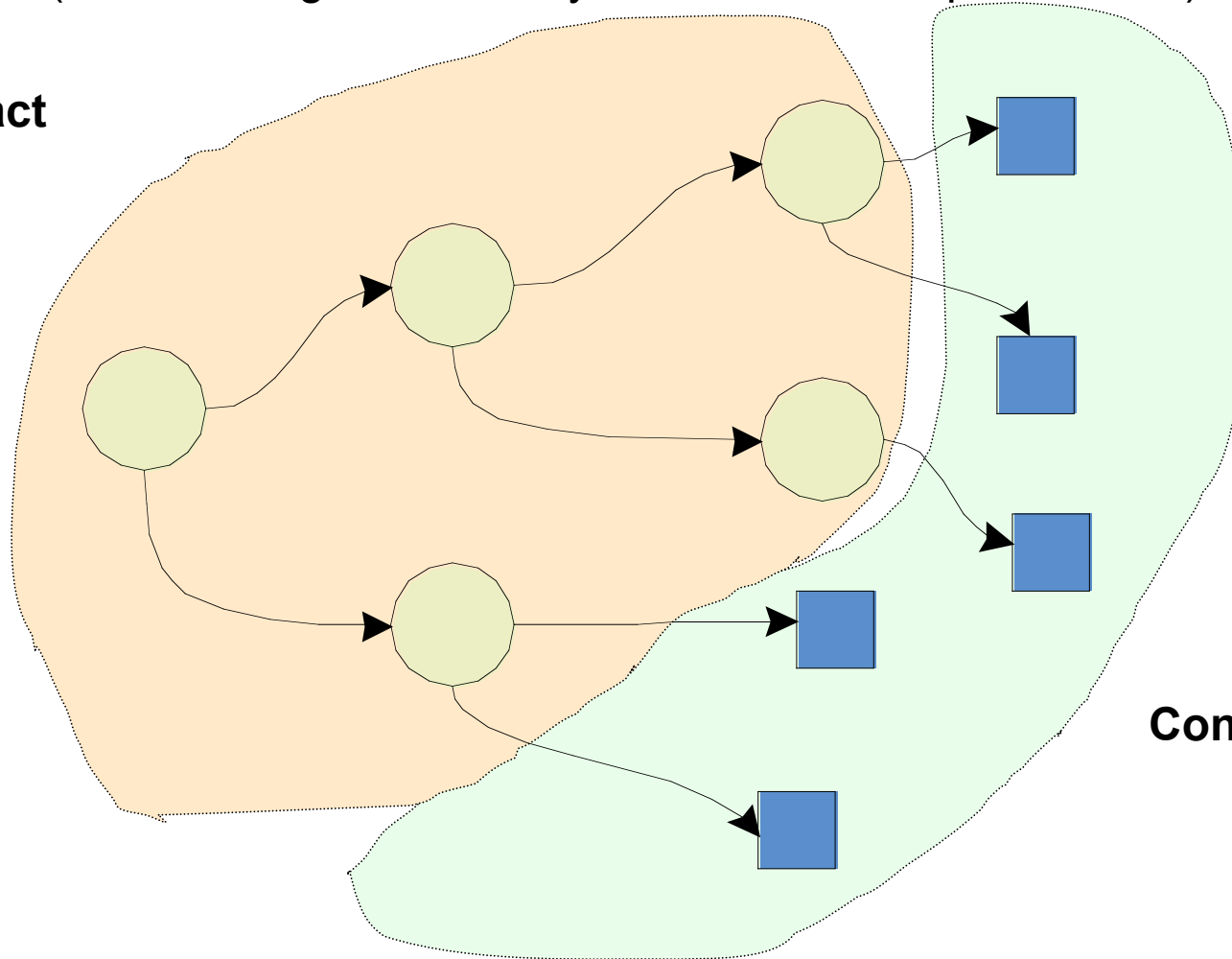
Basis for linking distributed identified resources (content, data, services)



Pathways Core: Graph-based Data Model

(Bekaert, Lagoze, Liu, Payette, Van de Sompel, Warner)

Abstract



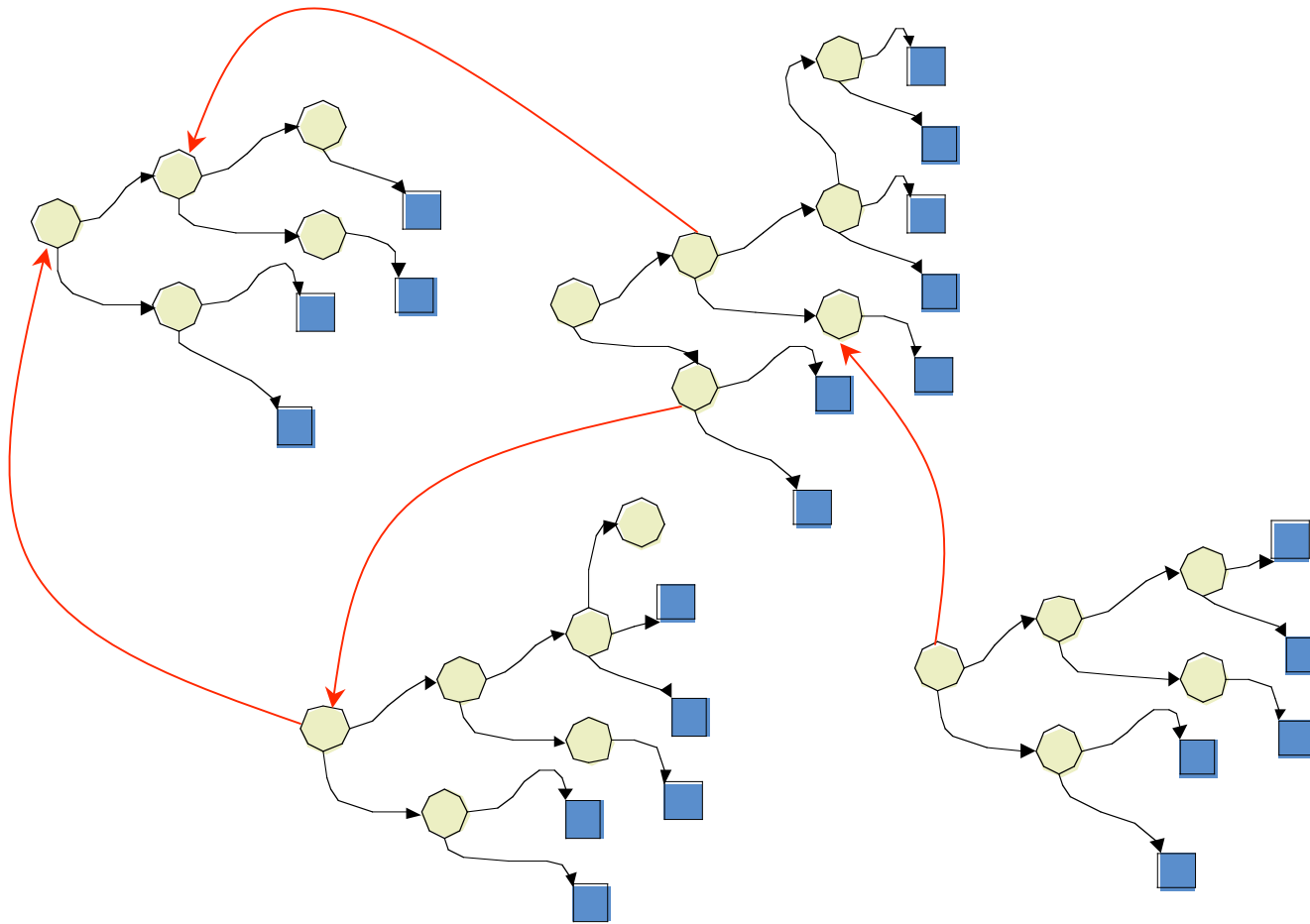
Concrete



Carl Lagoze
Cornell Computing and Information Science



Basis for a Network of Linked Objects



Carl Lagoze
Cornell Computing and Information Science



Why not just asset transfer?

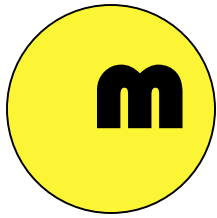
- Full transfer is only necessary for some applications – e.g., preservation mirroring
- In fact, in some cases it is **forbidden** and/or **undesirable**
- So, the infrastructure and model should accommodate but not be limited to transfer



By not committing to asset transfer...

- Avoid embedding IP issues into the core interoperability layer
- Accommodate “service-tuned” asset transfer
- Allow “live” references, rather than static copies



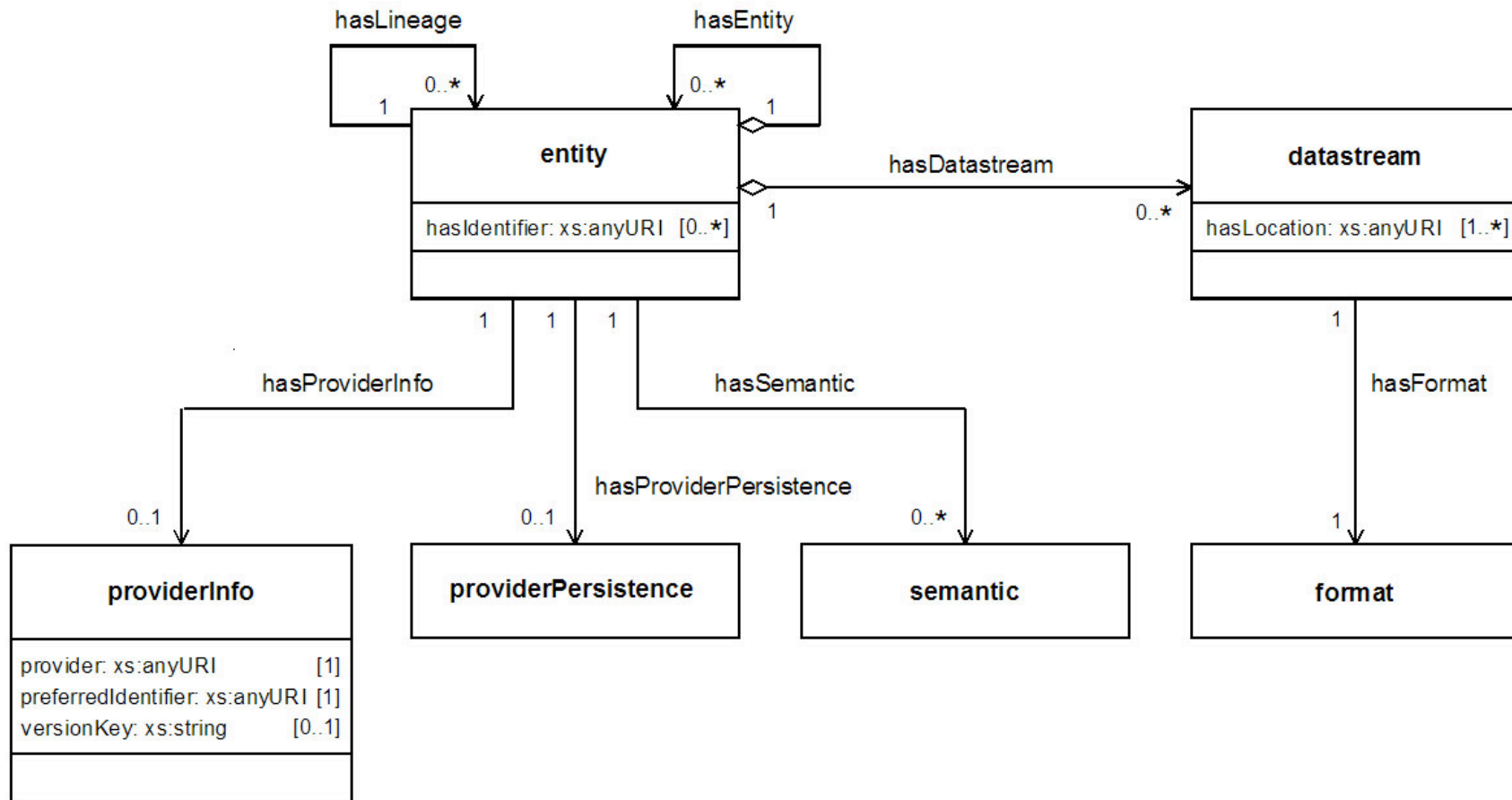


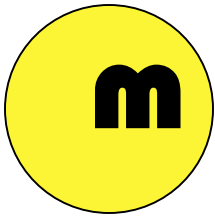
Model Core Requirement

- **Identity** independent of specific schemes
- **Lineage** relationships among objects
 - evidence of workflow for evidential citation
- **Semantics** associated with entities
 - facilitate service mapping
- **Recursion** for n-levels of entity containment
- Link to **concrete representation**
- Assertion of **persistence** levels

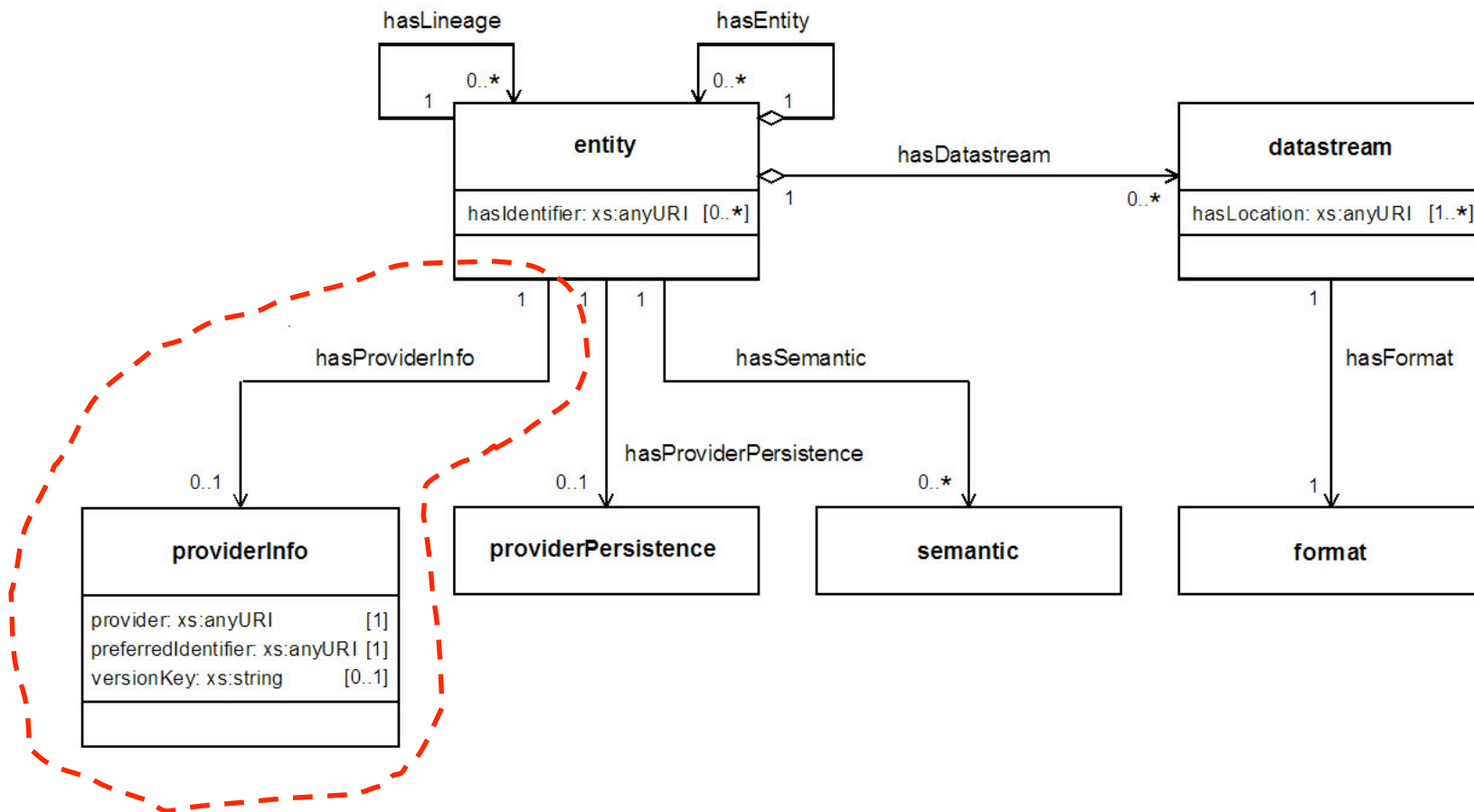


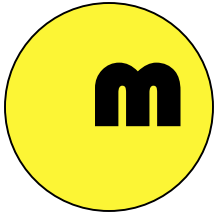
Data Model



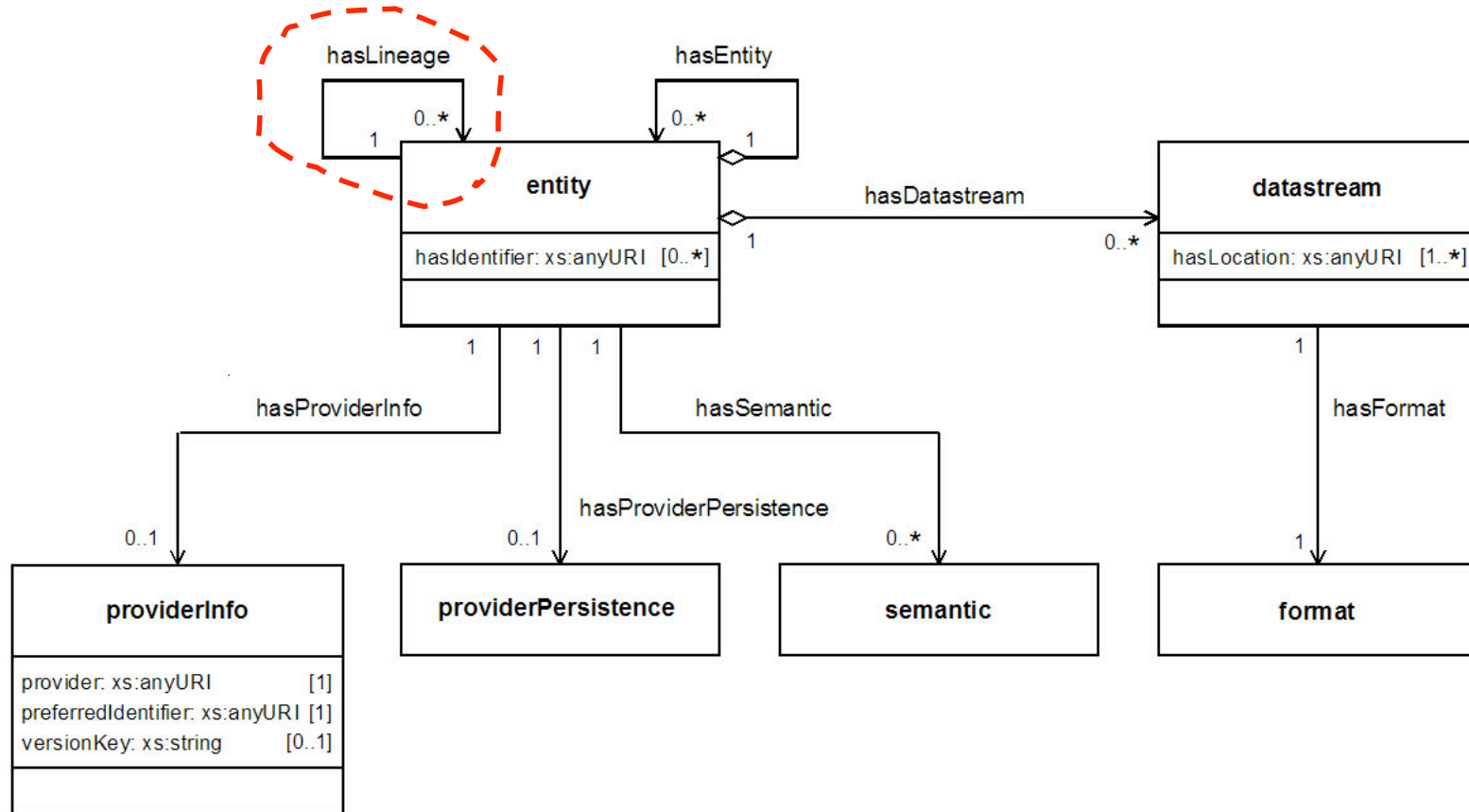


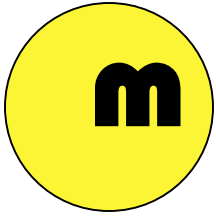
Identity



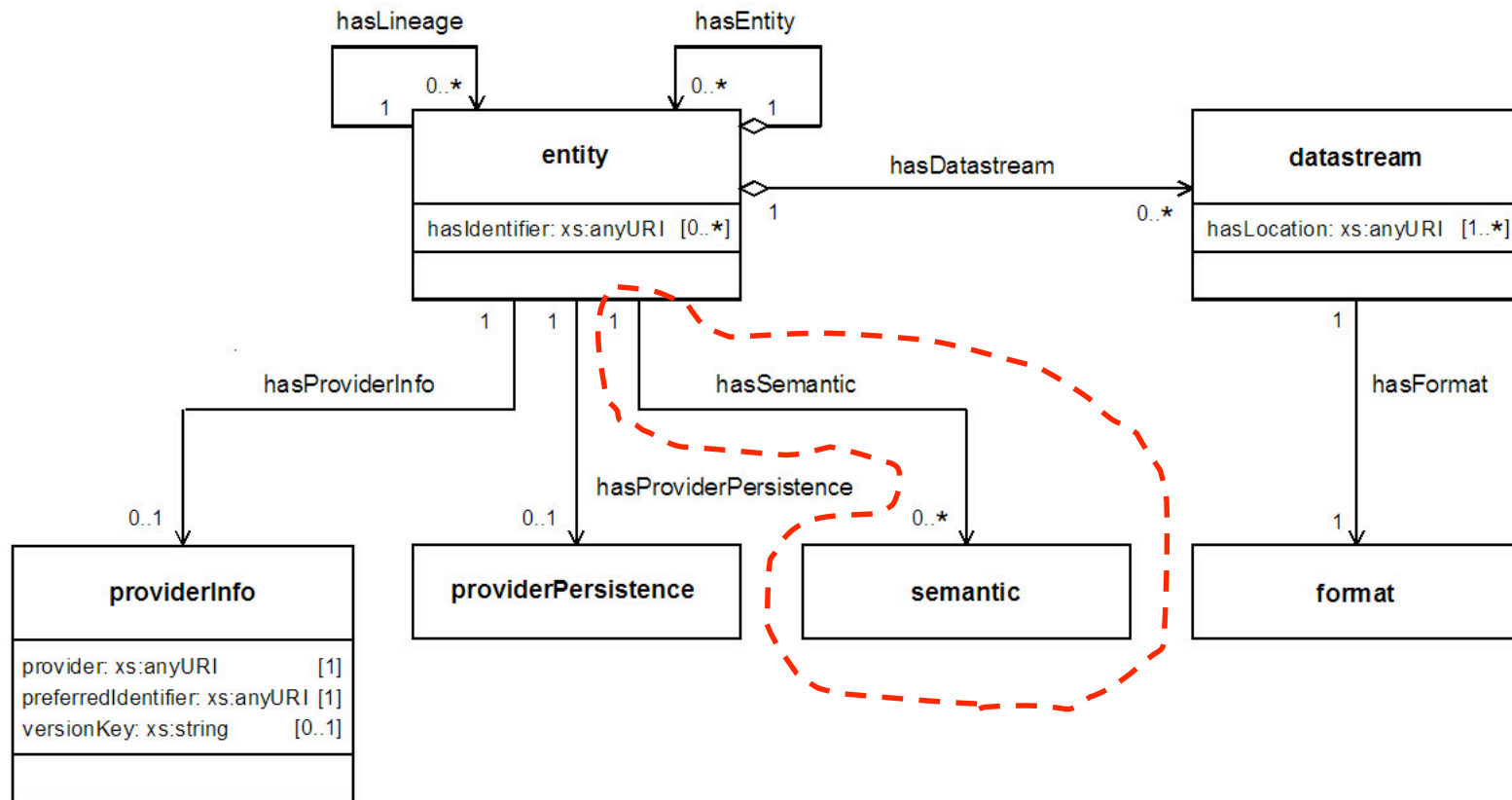


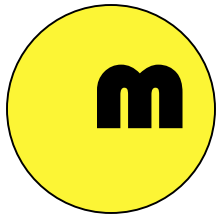
Lineage Relationships



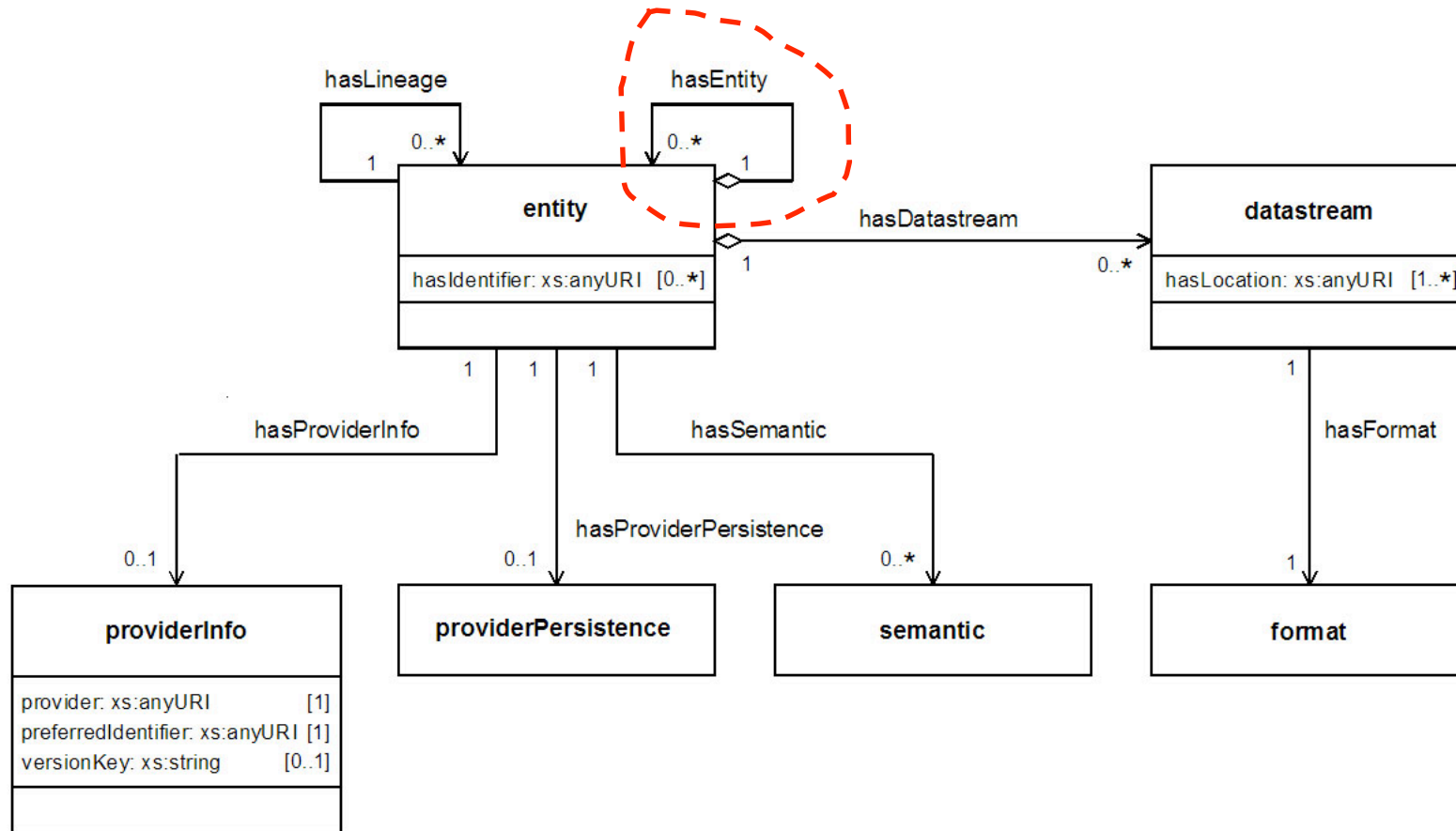


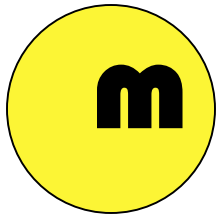
Semantics



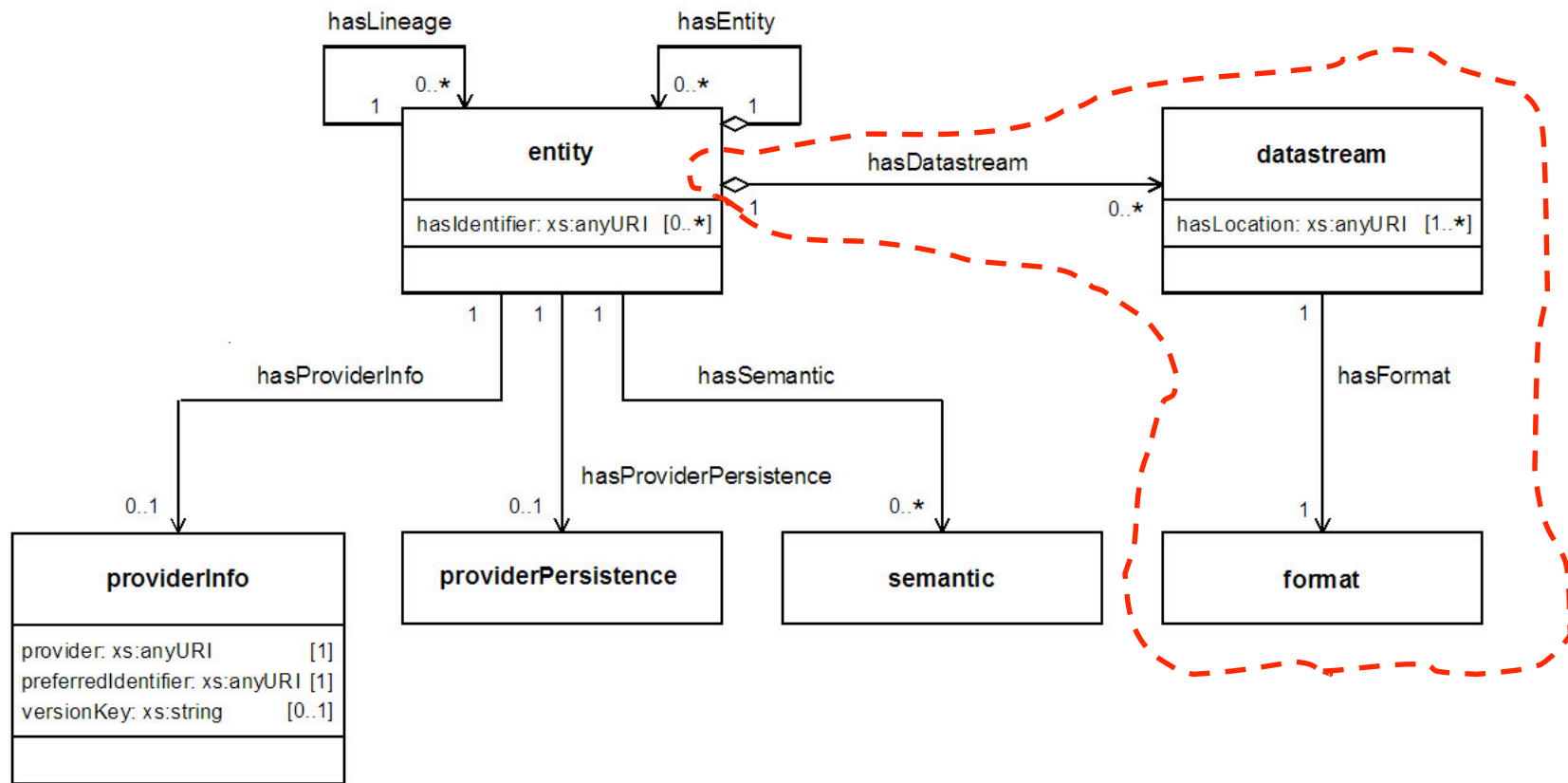


Recursion



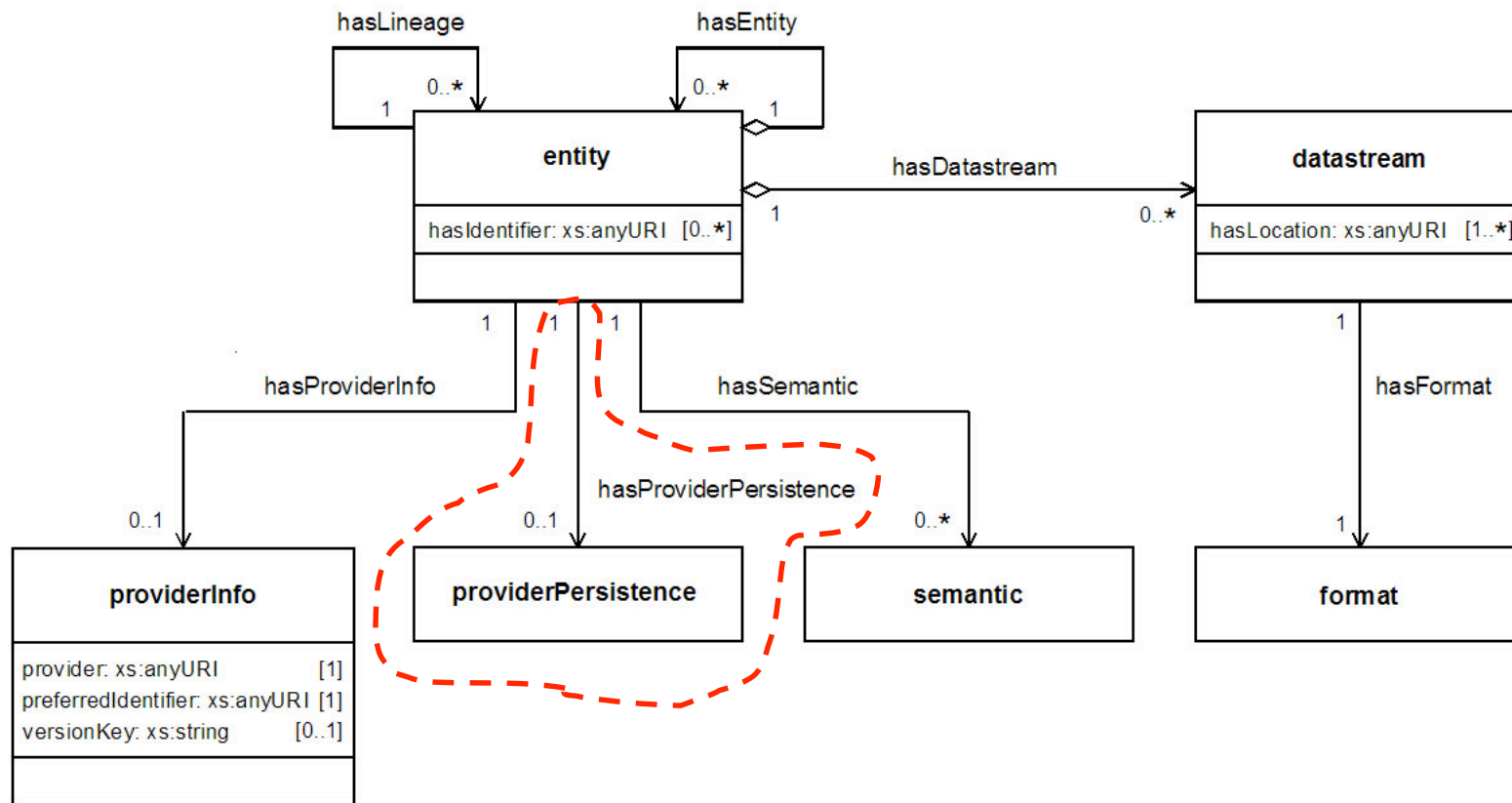


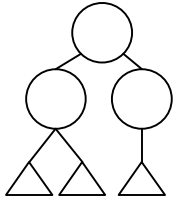
Concrete Representation





Persistence Guarantees

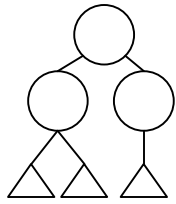




Serializing the Data Model

- **Surrogate** is a serialized and transportable representation of a digital object according to the model
- Accessed via **obtain** and **harvest**
- Deposited via **put**
- Prototype serializes via RDF/XML
 - Serializations in other formats possible – e.g., DIDL, METS

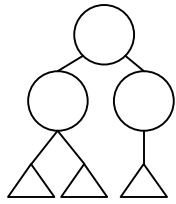




Surrogates \leftrightarrow Identity

- **providerInfo** records obtain information
- {provider, id, version} in surrogate makes it an evidential record of the digital object, provided by a specific service, in specific version





Surrogate \leftrightarrow Persistence

- **hasProviderPersistence** expresses commitment of provider to persistence of entity handle
- Commitment can vary: e.g., transient, handle persists, versions persist, object is stable...



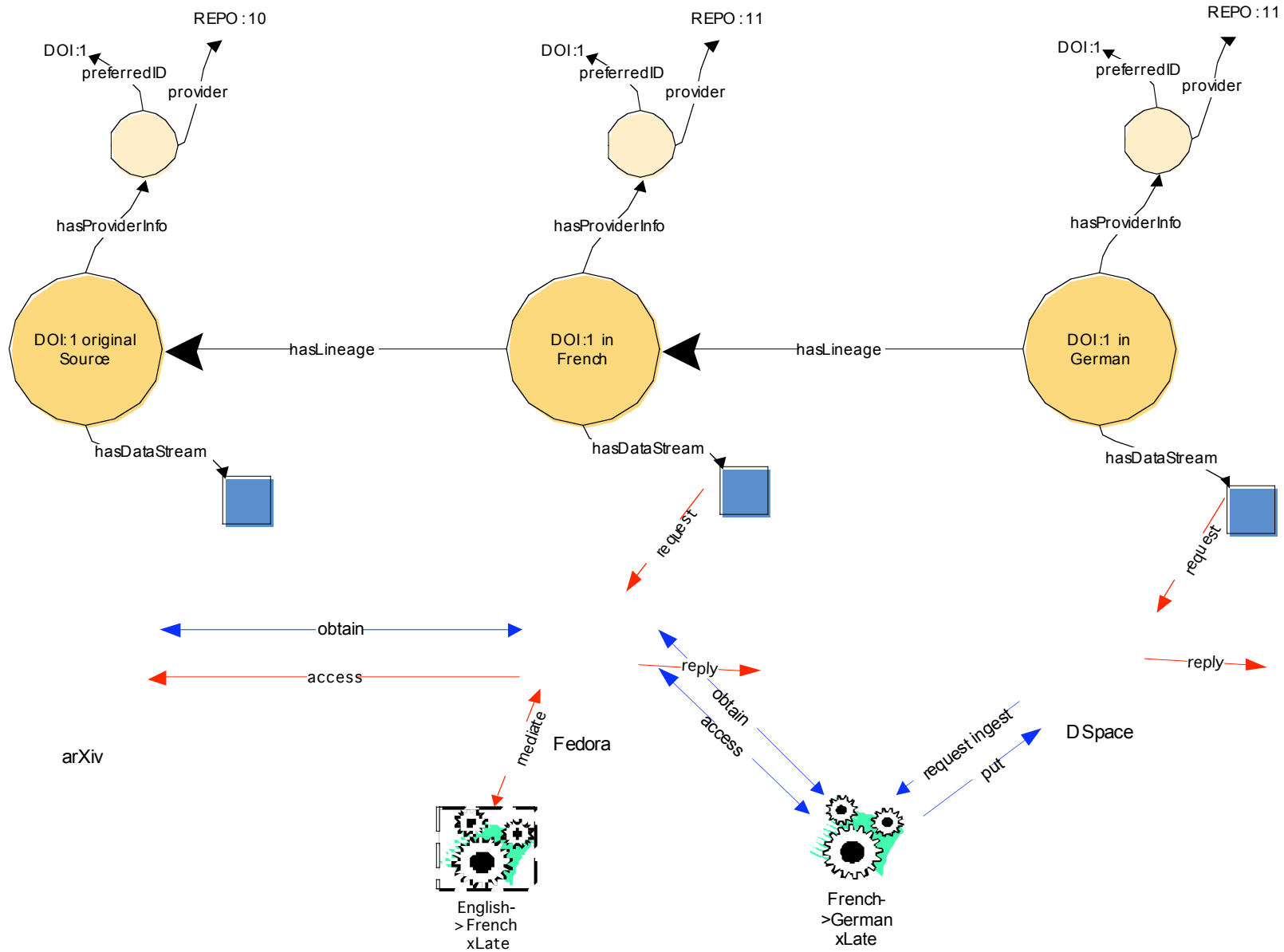
RDF/XML Representation of the Model

```
<?xml version="1.0" encoding="UTF-8"?>
<rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#" xmlns:core="info:pathways/core#">
  <core:entity
    rdf:about="info:pathways/entity/info%3Aid%2Fpubs.fedora-cis.org%3Apathways/info%3Ais%2Fri%2Fpub-payette-1">
    <core:hasIdentifier>info:is/ri/pub-payette-1 </core:hasIdentifier>
    <core:hasProviderPersistence rdf:resource="info:pathways/persistence/persistent"/>
    <core:hasSemantic rdf:resource="info:pathways/semantic/journal-article"/>
    <core:hasProviderInfo>
      <core:providerInfo>
        <core:provider>info:sid/pubs.fedora-cis.org:pathways </core:provider>
        <core:preferredIdentifier>info:is/ri/pub-payette-1 </core:preferredIdentifier>
        <core:versionKey>2006-04-18T01:01:59.797 </core:versionKey>
      </core:providerInfo>
    </core:hasProviderInfo>
    <core:hasLineage>
      <core:entity
        rdf:about="info:pathways/entity/info%3Aid%2Fnull/info%3Ais%2Fri%2Fpub-payette-1">
        <core:hasProviderPersistence rdf:resource="info:pathways/persistence/persistent"/>
        <core:hasProviderInfo>
          <core:providerInfo>
            <core:provider>info:sid/null </core:provider>
            <core:preferredIdentifier>info:is/ri/pub-payette-1 </core:preferredIdentifier>
            <core:versionKey>?? </core:versionKey>
          </core:providerInfo>
        </core:hasProviderInfo>
      </core:entity>
    </core:hasLineage>
    <core:hasDatastream>
      <core:datastream>
        <core:hasFormat rdf:resource="info:pathways/fmt/pronom/638"/>
        <core:hasLocation>http://memex.cs.cornell.edu:8100/fedora/get/cis-pubs:59/RELS-EXT </core:hasLocation>
      </core:datastream>
    </core:hasDatastream>
  </core:entity>
</rdf:RDF>
```



Carl Lagoze
Cornell Computing and Information Science





Carl Lagoze
Cornell Computing and Information Science

