PCDM Update

Karen Estlund
Penn State University
kestlund@psu.edu

Esmé Cowles
Princeton University
kc16@princeton.edu

bit.ly/pcdm-hydra-vc
History

By Tony Webster from Portland, Oregon (Portland Oregon White Stag Sign) [CC BY 2.0 (http://creativecommons.org/licenses/by/2.0)], via Wikimedia Commons
PCDM History - What was Hydra Up To?

- Penn State implements basic RDF April 2012
- UCSD complex object data modeling work published October 2012
- Oregon RDF work - linked open data and vocabularies registry 2013
- Stanford RDF meeting October 2013
- Hydra Connect 2014
- Hydra Works use cases on Github - open process 2014
- Portland meeting, “Making Progress” November 2014

PCDM History: Islandora

Another Community at Work
PCDM - All Together Now!

- Portland Code4Lib PCDM Meeting
  February 2015
    - concepts like “works,” “resource,” FRBR removed from the model
    - embraced it as an abstract object model not bibliographic

- Whiteboard modeling exercises 2015

- Technical and Rights Metadata Work
  May - October 2015
- Refined Ordering - OR 2015
- Works Extension March 2016
- Initial Implementations
  Spring / Summer 2016
Community

pcdm.org
PCDM.org & GitHub

PCDM Ontologies

- PCDM Models Ontology
- PCDM Use Ontology
- PCDM Rights Ontology
- PCDM Works Ontology
- PCDM File Formats Ontology

PCDM Wiki

Ontology for the Portland Common Data Model

Namespace: http://pcdm.org/models#

Description: This model is intended to underlie a wide array of repository and DAMS applications.

Communication

- Project information see: https://github.com/duraspace/pcdm/wiki
- Email discussion see: https://groups.google.com/d/forum/pcdm
- IRC discussion see: #pcdm on irc.freenode.net

Developer resources

To generate HTML versions of the ontologies:

```bash
$ mvn xhtml:transform
```
Portland Common Data Model

Introduction

The Portland Common Data Model (PCDM) is a flexible, extensible domain model that is intended to underlie a wide array of repository and DAMS applications. The primary objective of this model is to establish a framework that developers of tools (e.g., Hydra-based engines, such as Sufia, Curate, Worthwhile, Avalon; Islandora; custom Fedora sites) can use for working with models in a general way, allowing adopters to easily use custom models with any tool. Given this interoperability goal, the initial work has been focused on structural metadata and access control, since these are the key actionable metadata.

To encourage adoption, this model must support the most complex use cases, which include rich hierarchies of inter-related collections and works, but also elegantly support the simplest use cases.
Site in development after Feedback at LDCX 2016

Ontology and Extensions

- PCDM Models Ontology
- PCDM Use Ontology
- PCDM Rights Ontology

More Information

The PCDM Wiki provides a fuller description of the core data model, diagrams, project documentation, and recent presentations.

How to get involved?

All contributions are welcome: use-cases, documentation, code, patches, bug reports, feature requests, etc. You do not need to be a programmer to speak up!

The RDF vocabularies are in the GitHub repository for PCDM, which also includes an issue queue that you’re welcome to contribute to. Anyone can open an issue to discuss changes to the model or documentation.

In addition to the GitHub repository, you can also interact with the community via our Google Group.
Model
Technical Details

**Things**
- Collection
- Object
- File

Proposed (2.0)
- File Set

**Relationships**
- Has Member
- Has File
- Aggregates
- Has Related File
- Has Related Object

[https://github.com/duraspace/pcdm/](https://github.com/duraspace/pcdm/)

Please see documentation for namespaces. Due to slide legibility not all namespaces are articulated.
PCDM Domain Model

pcdm:Collection

pcdm:hasMember (m:m)
ore:aggregates (m:m)

pcdm:Object

pcdm:hasMember (m:m)

pcdm:hasFile (0:m)

pcdm:File

A Access
B Bitstream
D Descriptive
T Technical
Acerca del chocolate

Acerca del chocolate. Mandeville Special Collections Library, UCSD
Acerca del chocolate
More than one Object

pcdm:Object

hasMember

pcdm:Object

hasMember

pcdm:Object

hasMember

pcdm:Object

hasMember

pcdm:File

pcdm:File

pcdm:File

Acerca del chocolate

Inscription

Book Plate
More with Objects!

Acerca del chocolate

[Deed of Gift]

Deed of Gift

[Image of chocolate]

“Inscription”

[pdm:Object hasRelated Object]

[pdm:Object hasFile]

[pdm:File]

[pdm:Object hasFile]

[pdm:File]

[Deed of Gift]

[pdm:Object hasFile]

[pdm:File]

[pdm:Object]

[pdm:File]

[pdm:Object hasFile]

[pdm:File]

[pdm:Object]
Collections

- pcdm:Collection
  - hasMember
    - pcdm:Object

18th Century Catholic Church Medicinal Texts

- Acerca del chocolate
Optional Ordering Extension
Acerca del chocolate

Inscripción

Book Plate

Ordering
What Can go Where?

...and what can we say about different types of things
**Access Control** - `pcdm:Collection`, `pcdm:Object`, and `pcdm:File`

Group A can VIEW

- `pcdm:Collection`
  - `hasMember`:
    - `pcdm:Object`
      - `hasFile`:
        - `pcdm:File`
          - `hasFile`:
            - `pcdm:File`
## Web ACLs

<table>
<thead>
<tr>
<th>Property</th>
<th>Range</th>
<th>Usage</th>
<th>Obligation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Agent (acl:agent)</td>
<td>foaf:Agent</td>
<td>Individual user this ACL applies to.</td>
<td>min 0, max unbounded</td>
</tr>
<tr>
<td>Agent Class (acl:agentClass)</td>
<td>rdfs:Class</td>
<td>Class of users this ACL applies to.</td>
<td>min 0, max unbounded</td>
</tr>
<tr>
<td>Mode (acl:mode)</td>
<td>rdfs:Class</td>
<td>Actions permitted by this ACL (e.g., acl:Read, acl:Write, hydra:Discover, etc.).</td>
<td>min 1, max unbounded</td>
</tr>
<tr>
<td>Resource (acl:accessTo)</td>
<td>gen:InformationResource</td>
<td>Individual resource this ACL applies to.</td>
<td>min 0, max unbounded</td>
</tr>
<tr>
<td>Resource Class (acl:accessToClass)</td>
<td>rdfs:Class</td>
<td>Class of resources this ACL applies to.</td>
<td>min 0, max unbounded</td>
</tr>
</tbody>
</table>
Acerca del chocolate: Manuscript on chocolate

Catholic Church--Mexico--History--18th century
Cocoa--Therapeutic use--Early works to 1800

ca. 1730

Mandeville Special Collections Library, University of California, San Diego, La Jolla, 92093-0175

https://libraries.ucsd.edu/ark:/20775/bb3482101x
Technical Metadata and Bitstream: pcdm:File

Intermediate File

20775-bb3482101x-1-1-3.jpg

527 KB

0000E32F
90990902

d3b07384d113edec49eaa6238ad5ff00

2013-11-29T13:56:46-04:00
Technical Metadata Recommendations

- ebucore:filename
- ebucore:fileSize
- rdfs:label
- ebucore:dateCreated
- premis:hasMessageDigest
- premishash:md5

- rdf:type
- ebucore:hasMimeType
- ebucore:dateModified
- pronom:puid
- sweetjpl:hasByteOrder

https://wiki.duraspace.org/display/hydra/Technical+Metadata+Application+Profile
File Use Vocabulary

- Original File
- Thumbnail Image
- Extracted Text
- Preservation Master File
- Intermediate File
- Service File
- Transcript

https://wiki.duraspace.org/display/hydra/File+Use+Vocabulary
https://github.com/duraspace/pcdm/blob/master/pcdm-ext/use.rdf
Rights Metadata Recommendation

Minimum:

- edm:rights
- dct:rightsHolder

Recommended:

- dc:rights
- marcrel:cpc (claimant)
- premisHasCopyrightStatus
- pcdmrts:rightsOverride
- pcdmrts:rightsOverrideExpiration

https://wiki.duraspace.org/display/hydra/Rights+Metadata+Recommendation
Works Extension

- **Work**: A work or intellectual entity, such as a book, film, dissertation, etc. Works have member FileSets representing their physical structure (e.g., pages in a book), and related TopRanges representing their logical structure or structures (e.g., sections and chapters in a book).

- **FileSet**: A group of related Files, typically a single master File and its derivatives.

- **TopRange**: A logical ordering of the component parts of a Work, corresponding to a IIIF Range with the "top" viewing hint. Has member Ranges that represent the logical structure, such as chapters within a book, scenes in a film, etc.

- **Range**: A section of a Work, corresponding to a IIIF Range. Has member FileSets representing the physical parts of the Work are part of the section (e.g., which pages are in a chapter).
Hydra::Works Diagram

Showing the PCDM-based domain model implemented in the hydra-works gem.

Key
- A: Access
- B: Bitstream
- D: Descriptive
- T: Technical
- Structural

Hydra::Works::Collection a pcdm:Collection

Hydra::Works::Work a pcdm:Object

Hydra::Works::FileSet a pcdm:Object

OriginalFile a pcdm:File

Thumbnail a pcdm:File

ExtractedText a pcdm:File
Reference Diagram from Use Case Exercise

Archival Disk Image example

https://wiki.duraspace.org/display/FF/PCDM+Mappings+-+Reference+Diagrams+for+Comment
Building PCDM into the Hydra stack

- Building a new Hydra stack for PCDM
- Community realignment around collaborative development
- Doing things the Hydra Way™
- Current status:
  - Curation Concerns 1.0
  - Sufia 7.0 beta 4
ActiveFedora

- (Still) the foundation of the Hydra stack
- Makes Fedora resources look like ActiveRecord
- Added support for ORE Proxy ordering
Hydra::PCDM

- Implementation of core PCDM classes and relationships
- Intentionally simple and unopinionated to allow other Ruby apps to use PCDM without having to buy into the rest of the stack
Hydra::Works

● Repository toolkit
  ○ Characterization, derivatives, collections, virus scanning
  ○ Services for uploading files, versioning, etc.

● Extends PCDM with a few more classes
  ○ Work: pcdm:Object representing whole or part of digital object
  ○ FileSet: pcdm:Object representing a File and its derivatives
  ○ see also: PCDM 2.0

● Good place to start for
  ○ gems and tools
  ○ radically different apps
CurationConcerns

- Rails engine with controllers and views
  - Provides a usable application out of the box
  - Generators for creating work models
  - Functionality in modules to make it easier to override/customize
- Default functionality
  - Works with member FileSets
  - Works with member Works
  - UI and workflow for uploading files, FITS characterization, generating derivatives, etc.
  - Drag-and-drop ordering
- Good place to start for digital collections apps
Sufia

- Adds institutional repository functionality
  - Self-deposit workflow, proxy deposit, social features
  - Dashboards, statistics, reporting
  - Now extending CurationConcerns

- Default functionality
  - GenericWork with member FileSets

- Good place to start for institutional repositories
Evolving Data Model

- **Do we need a Work class?**
  - Hydra::Works has a Work class and uses FileSets for parts
  - Moving towards using pcdm:Object for works and their parts

- **FileSet: logically-grouped sets of Files**
  - e.g., a source File and its derivatives — not a part/component/page

- **Physical Structure vs. Logical Structure**
  - (Top)Range to represent overlays like tables of contents
  - Also aligns with IIIF
PCDM 2.0

PCDM 2.0

A Collection is a group of resources, such as:
- Resources managed by an institution or unit
- Resources received from a donor
- Resources related to a topic
- Resources selected by a user
- etc.

Collections have descriptive metadata, access metadata, and may link to Objects and/or Collections. By default, member Objects and Collections are an unordered set, but can be ordered using the ORE Proxy class.

Properties:
- pcdm:hasMember < ore:aggregates: Links to an Object that is a member of this Collection, or a child Collection. Note on transitivity: hasMember is not defined as transitive, but applications may treat it as transitive as local needs dictate.
- pcdm:hasRelatedObject < ore:aggregates: Links to an Object that is related to the collection, but not a member of it. Typically used for documentation, thumbnails, etc. Can also be used to link to AlternateOrder to encode multiple orders.

https://github.com/duraspace/pcdm/wiki/PCDM-2.0
PCDM at Open Repositories

- Well-attended workshop
- Interest from Islandora community
- Research data
- Questions about recent modeling discussion topics
Questions!