Migrating Metadata Day 2

Digital Collections in RDF, PCDM & Fedora 4

Christina Harlow / cmh329@cornell.edu / @cm_harlow

http://github.com/cmh2166/ELAG16Metadata

Slides, Data, Scripts

github.com/cmh2166/elag16metadata

Exercises, Drawings, Notes

bit.ly/elag16metadata

Agenda

Day 1: 14:00-15:30

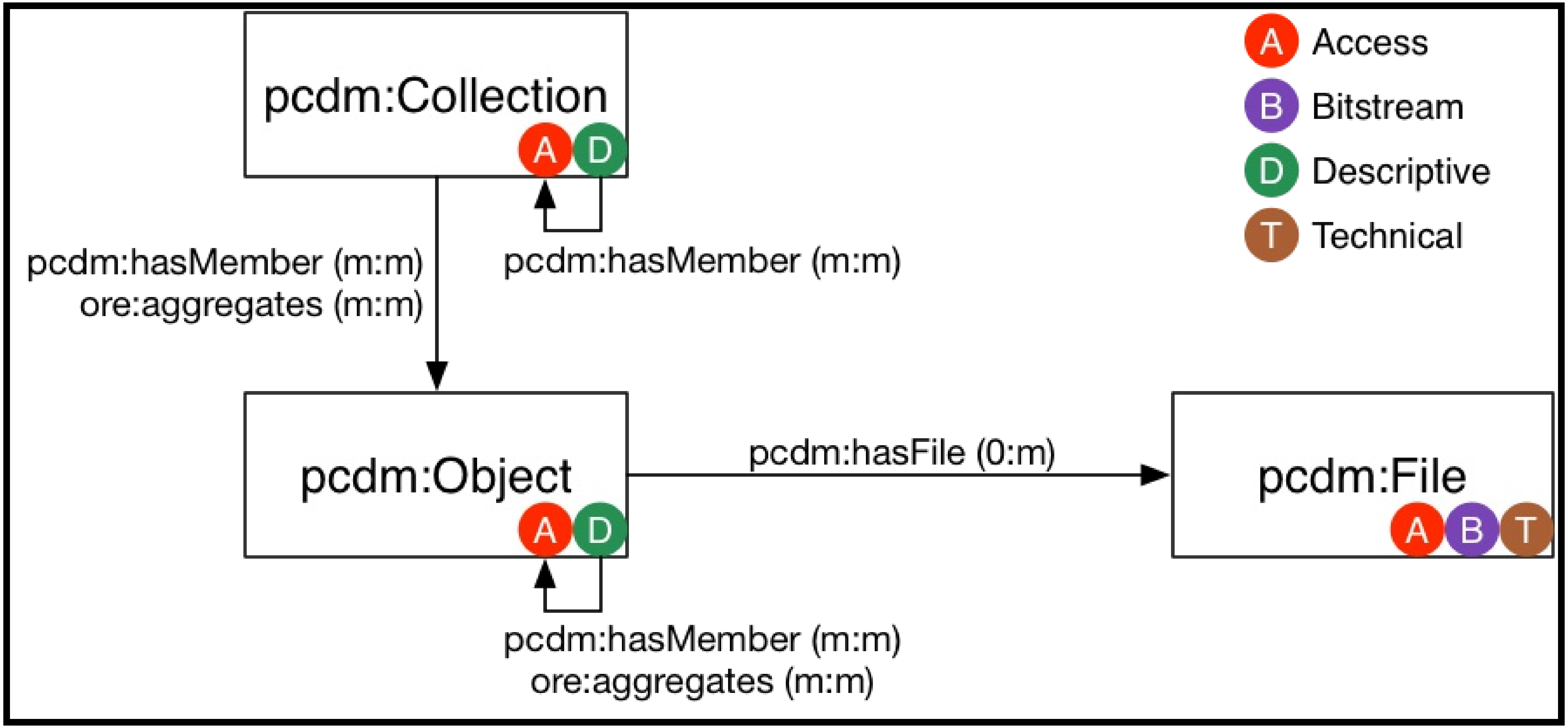
14:00-14:45	Existing Metadata Assessment
14:45-15:30	PCDM & Data Modeling

Day 2: 11:00-13:00

11:00-12:00	PCDM Modeling => Fedora 4
12:00-13:00	Playing with Fedora 4 & REST API

PCDM & Data Modeling Review

PCDM Overview



https://github.com/duraspace/pcdm/wiki

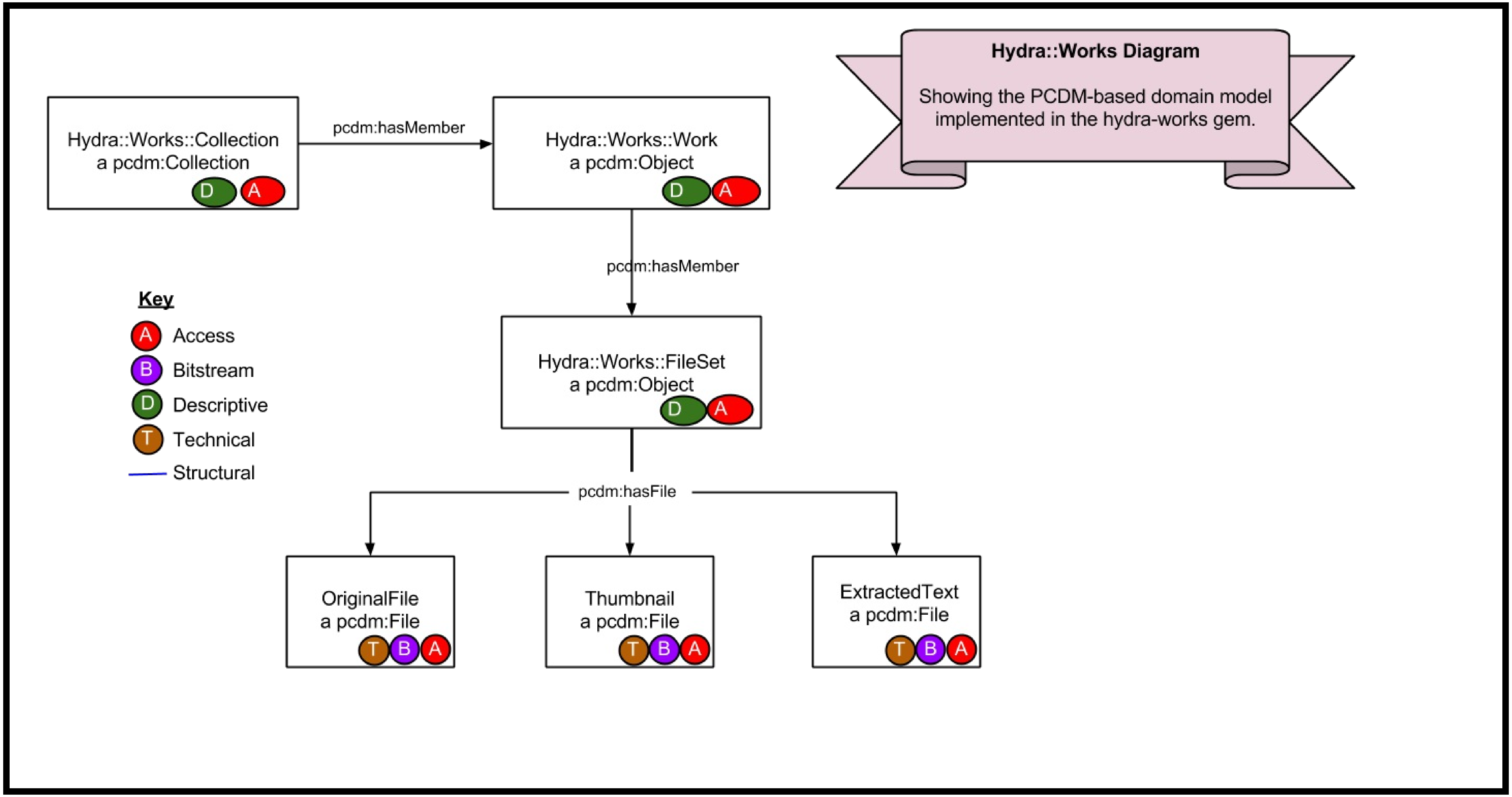
PCDM Core Classes

- **pcdm:Object**: An Object is an intellectual entity, sometimes called a "work", "digital object", etc...
- **pcdm:Collection**: A Collection is a group of resources...
- **pcdm:File**: A File is a sequence of binary data and is described by some accompanying metadata...

PCDM Properties

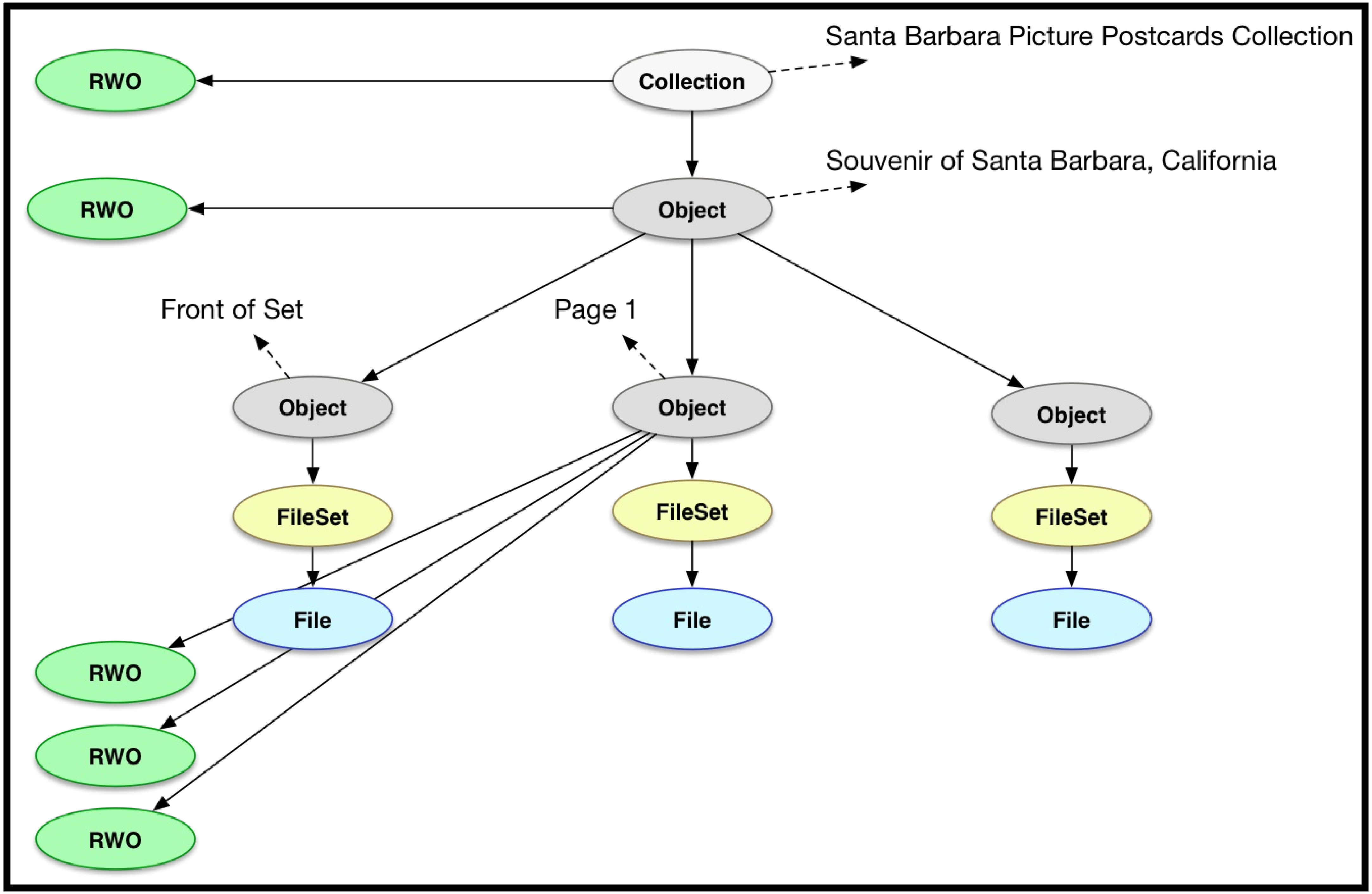
- **`pcdm:memberOf`**
  `Domain: ore:Aggregation | Range: ore:Aggregation`
- **`pcdm:hasMember`**
  `Domain: ore:Aggregation | Range: ore:Aggregation`
- **`pcdm:fileOf`**
  `Domain: pcdm:File | Range: pcdm:Object`
- **`pcdm:hasFile`**
  `Domain: pcdm:Object | Range: pcdm:File`
- **`pcdm:relatedObjectOf`**
  `Domain: pcdm:Object | Range: ore:Aggregation`
- **`pcdm:hasRelatedObject`**
  `Domain: ore:Aggregation | Range: pcdm:Object`

PCDM Works



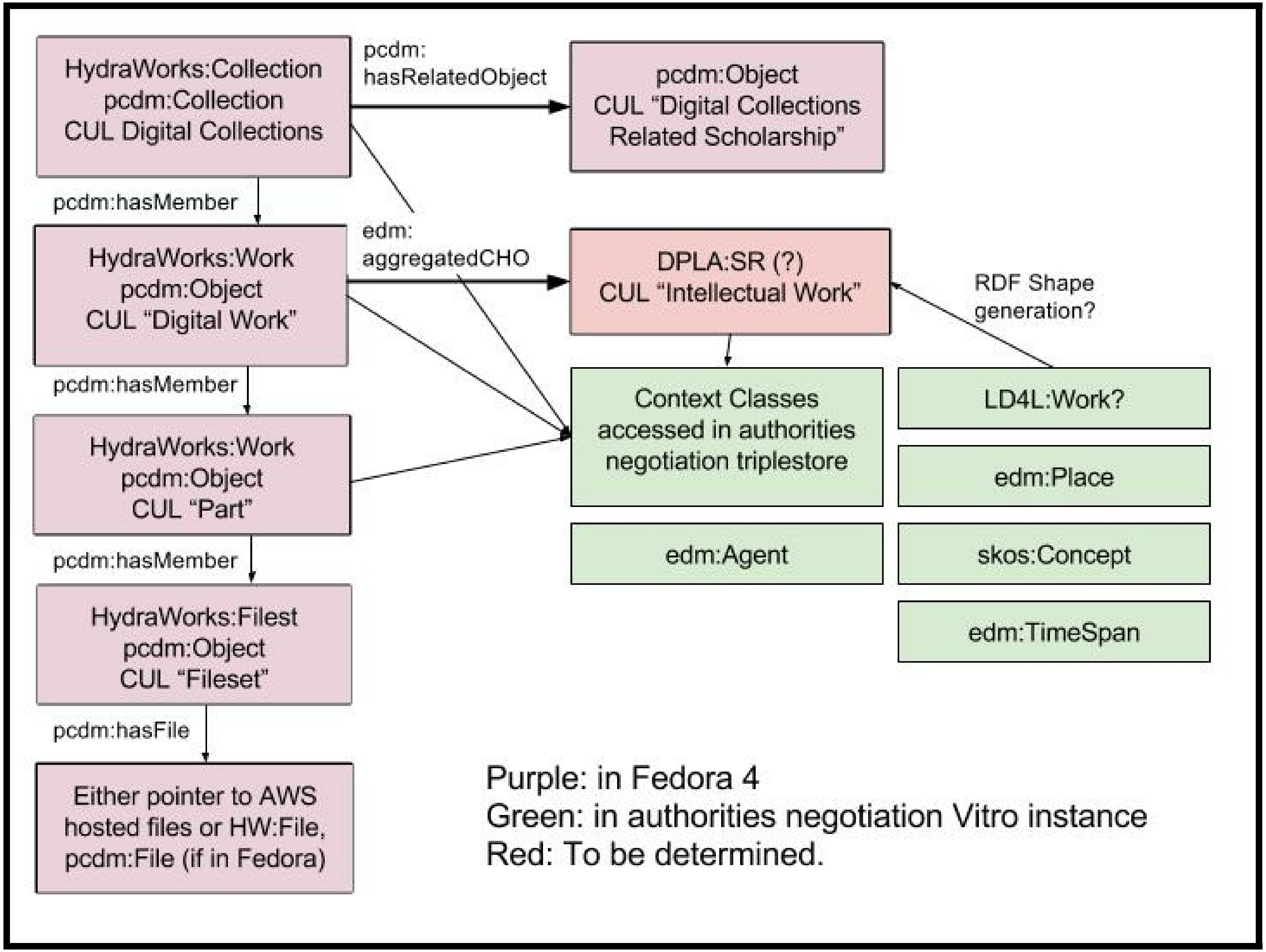
https://github.com/projecthydra/hydra-works

Classic Example: Postcards



https://github.com/hybox/models/blob/master/notes/usecase2.md

Cornell’s PCDM…so far

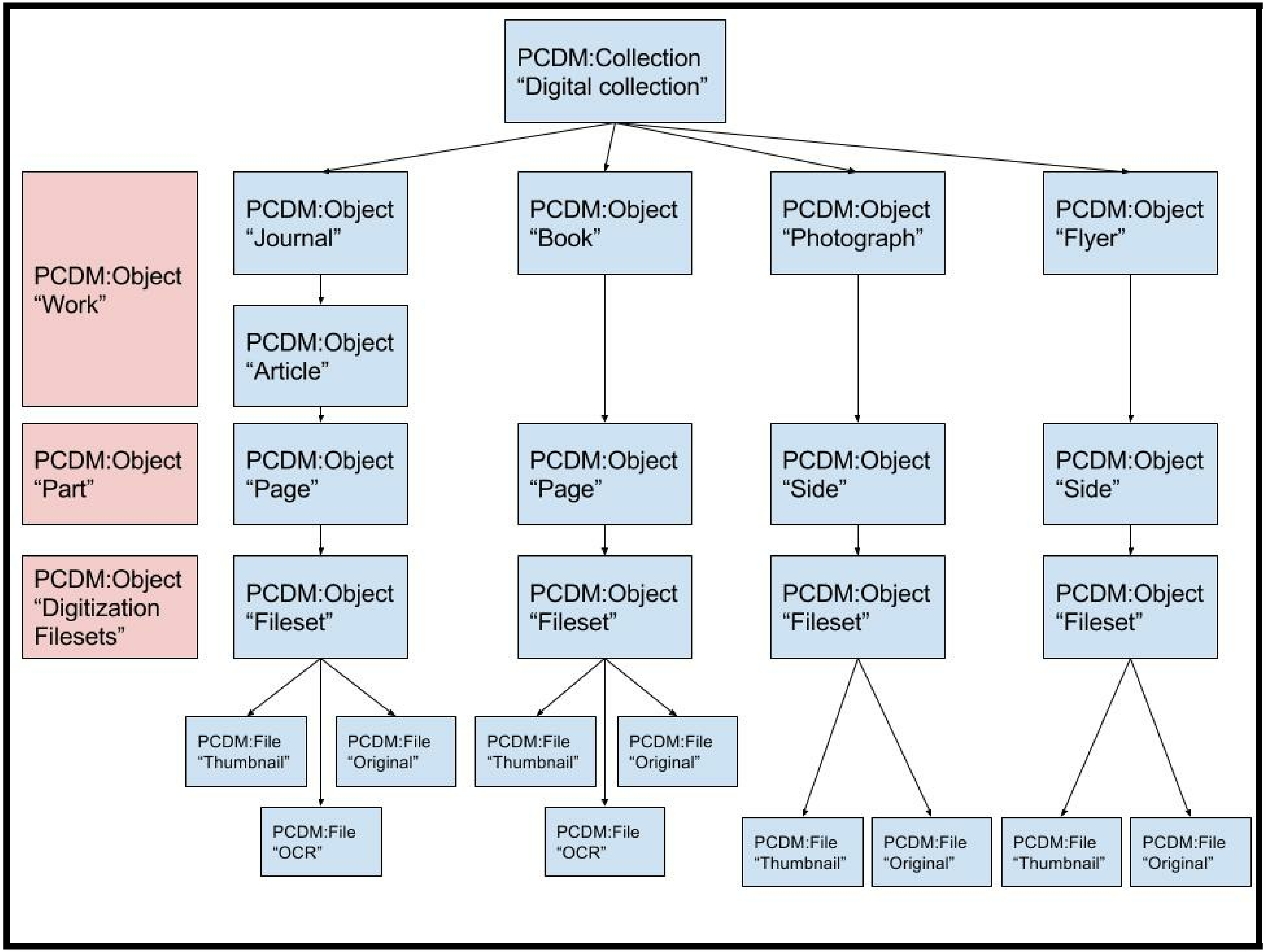


Purple: in Fedora 4
Green: in authorities negotiation Vitro instance
Red: To be determined.

Cornell's PCDM…so far

Bringing it all together...

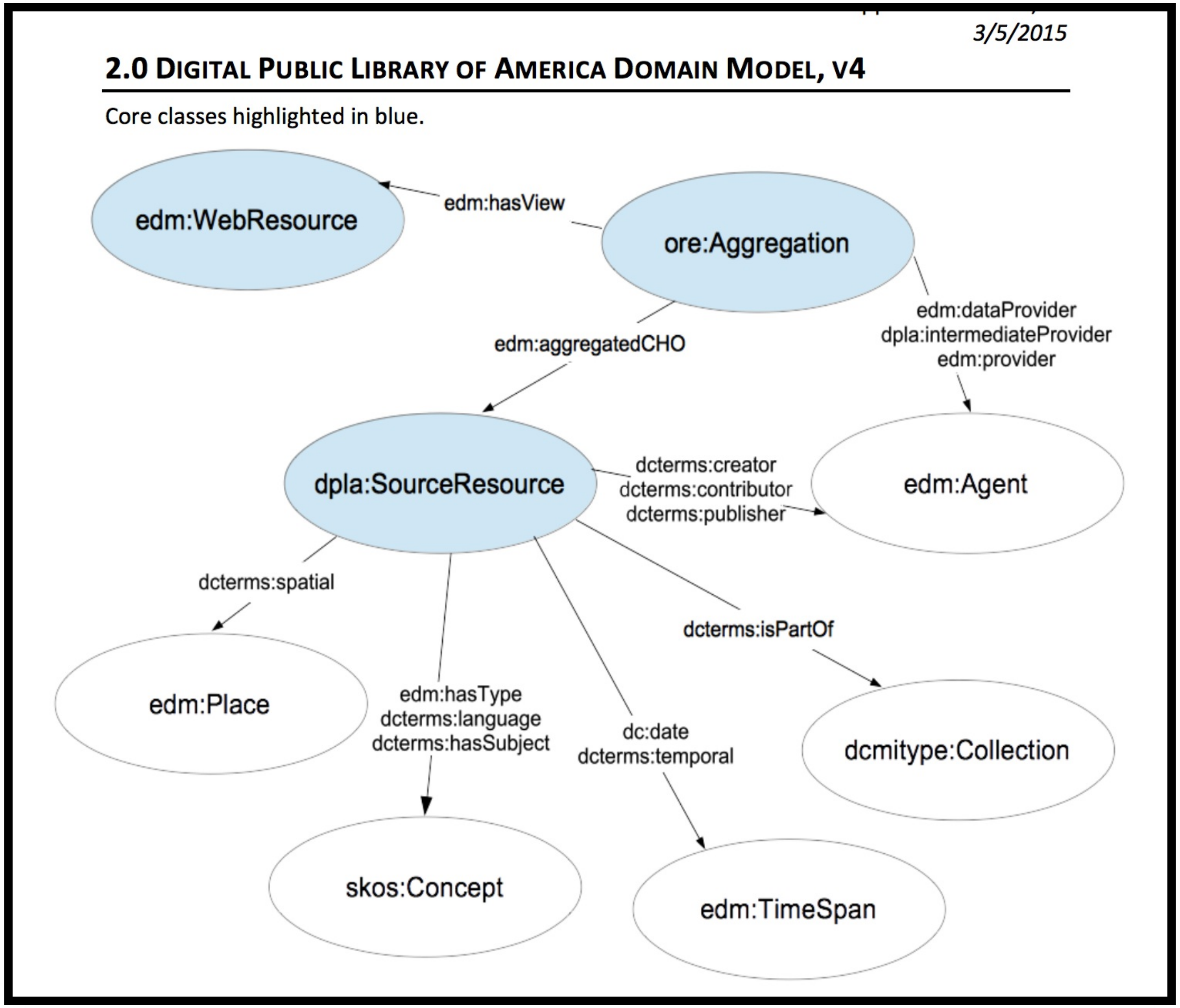
Filling in the Details: Classes



Filling in the Details: Properties

- Becomes Easier to Determine with Clearer Classes
- Be Aware of Predicate Domains & Ranges
- DCMI Publishing Guidelines Helpful Example
- This supports Machine Inference, Not Really for Validation Purposes

”Context Classes” and/or RWOs



3/5/2015

2.0 Digital Public Library of America Domain Model, v4

Core classes highlighted in blue.

https://dp.la/info/wp-content/uploads/2015/03/MAPv4.pdf

Preliminary Metadata Application Profile

Here is a start at merging Models & Metadata Application Profiles we started to pull together yesterday

Google Drive ELAG Folder, Merged Map file

Where we stopped yesterday...

Building off the metadata review for your objects:

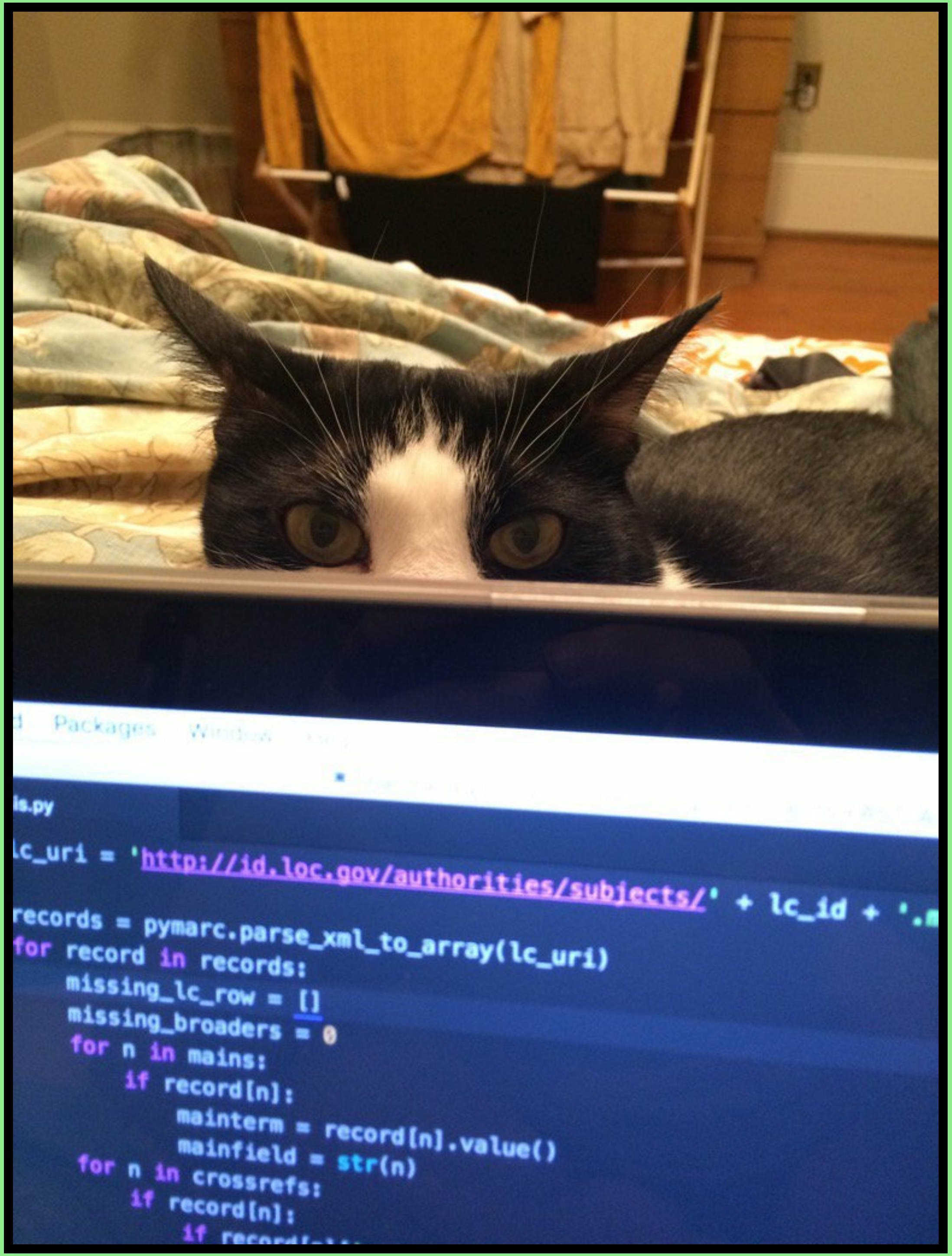
1. Group Your Entities into Possible PCDM Classes;
2. Map fields to relationships between Objects;
3. Try Drawing Model in Google Drawings;
4. What Fits? What Doesn't?
5. Start thinking about properties: Linked Open Vocabs can help! http://lov.okfn.org/

Right now…

Building off the metadata review for your objects:

1. Create your group's PCDM Model in Google Drawings;
2. Map fields to relationships between Objects;
3. Review mapped properties in merged MAP
4. What Fits? What Doesn't?

And… Fedora 4?



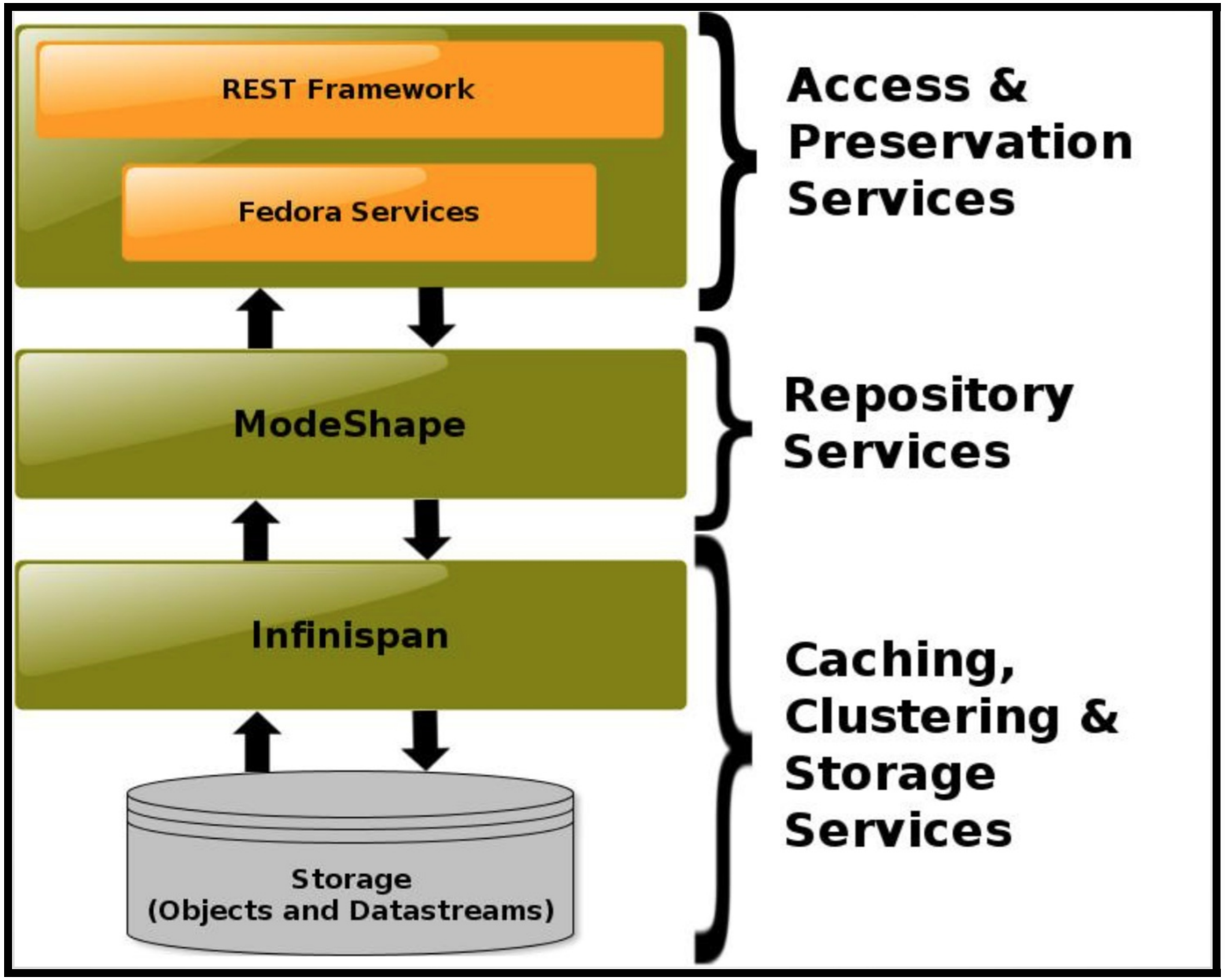
```
lc_uri = 'http://id.loc.gov/authorities/subjects/' + lc_id + '.m

records = pymarc.parse_xml_to_array(lc_uri)
for record in records:
    missing_lc_row = []
    missing_broaders = 0
    for n in mains:
        if record[n]:
            mainterm = record[n].value()
            mainfield = str(n)
    for n in crossrefs:
        if record[n]:
```

Fedora 4

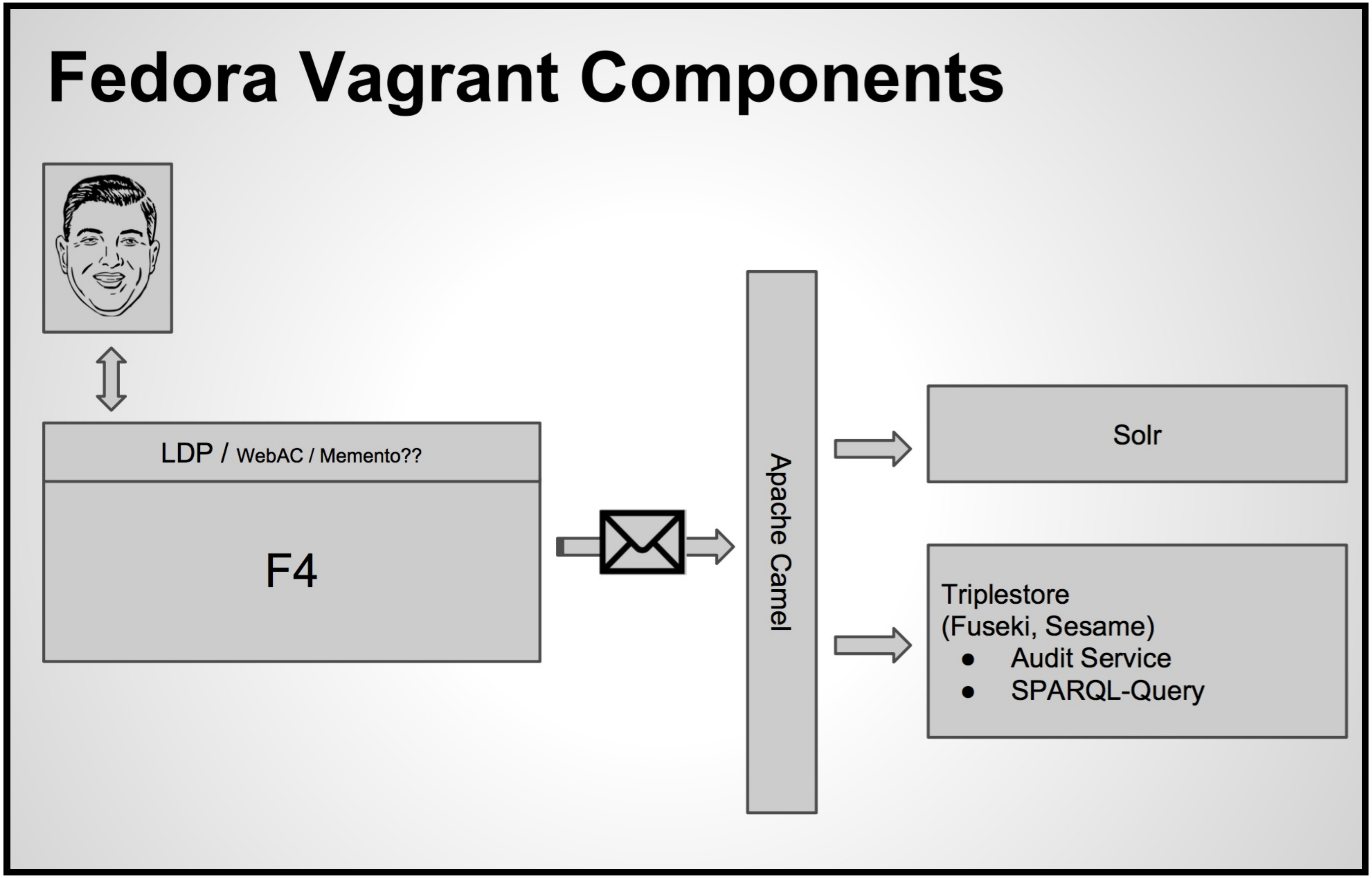
1. Digital Object Store & Platform
2. RDF-based & Objects are Atomic Resources w/semantic connections
3. Flexible, extensible content modeling (PCDM entirely optional)
4. Offers Architectural Differences from Fedora 3
5. Above Focused on Descriptive Metadata + Fedora 4

Fedora 4 Stack



https://wiki.duraspace.org/display/FF/Fedora+Repository+Home

Fedora 4 Vagrant Stack



Fedora Vagrant Components

LDP / WebAC / Memento??

F4

Apache Camel

Solr

Triplestore
(Fuseki, Sesame)

- Audit Service
- SPARQL-Query

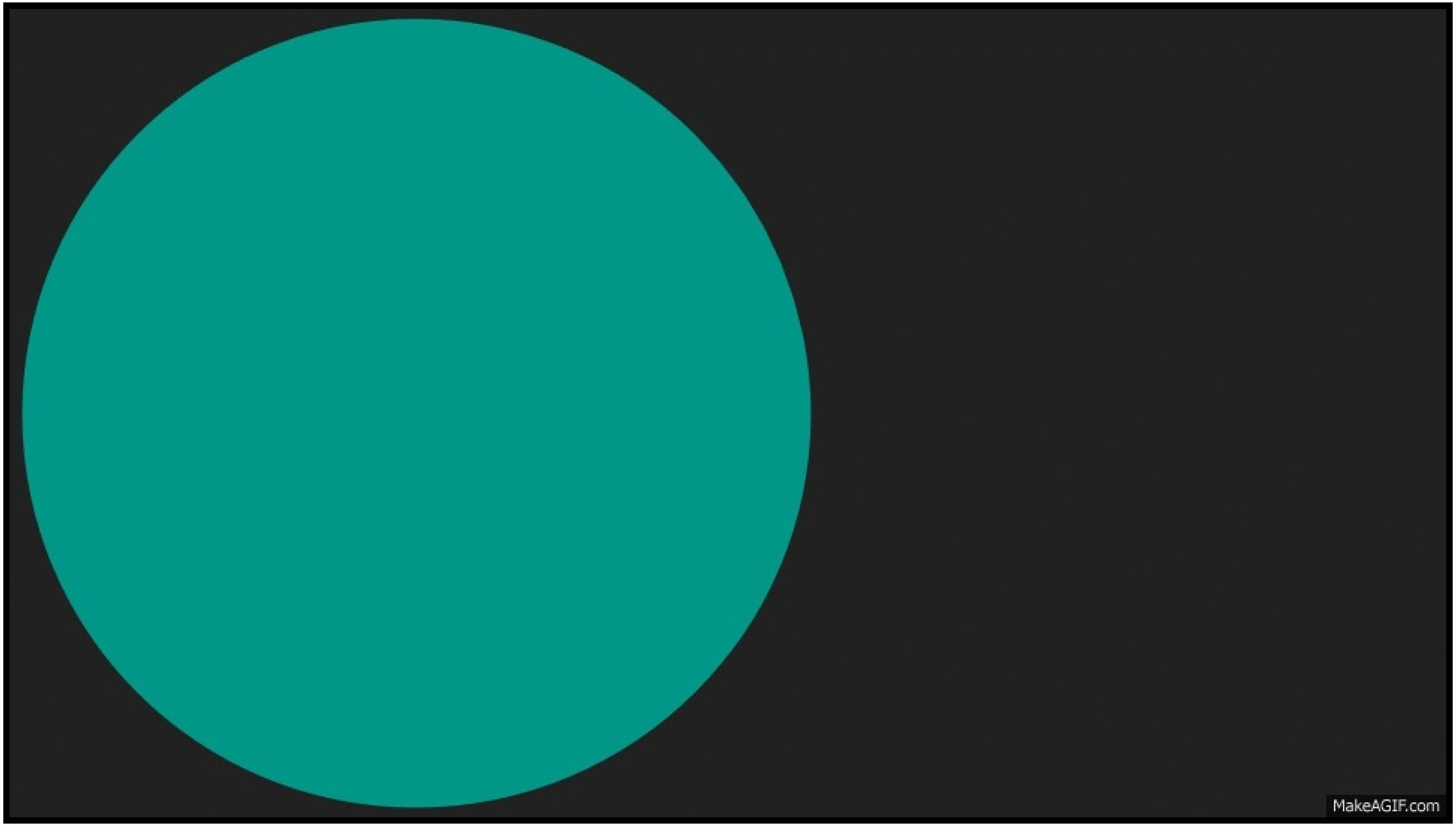
Thanks to Andrew Wood's LDCX Slides

Fedora 4 & RDF

- All RDF statements in Fedora 4 must have Fedora Resource as Subject
- Fedora resource? RDF Resource Managed by Fedora
- You could keep XML metadata as related file...
- ... But this limits granular update requests

But Wait, Where is Hydra?

Hydra ”Gobstopper”



Credit to Mike Giarlo, Jon Stroop, Matt Zumwalt

ActiveFedora + Hydra Context Overview

- **Fedora 4:** Object Repository
- **ActiveFedora:** Manage Fedora Objects as Ruby Objects
- **Hydra:** Set of DAM repository components
- **PCDM-Works:** Sets behaviors of PCDM objects in Hydra
- **Curation Concerns:** Gem that broadens Hydra functionalities for working with PCDM, including an object editor
- **Sufia 7:** Tailors all of the above for self-deposit institutional repositories, not much in way of metadata

Functionally, what does this mean for metadata

All of this to say, with Hydra in particular, application modeling concerns may/may not intersect with PCDM (meta)data modeling.

LDP (Linked Data Platform)

- W3C Recommendation from 2015
- Set of rules for HTTP operations on web resources
- Provides architecture for read-write Linked Data on the web

LDP Definition for Humans from Adam Wead

An LDP server is a web server that lets you manage relationships in RDF by creating URLs you can perform CRUD operation on:

- http://my_url/some_resource
- http://my_url/some_resource/children
- http://my_url/some_resource/children/child_resource

http://islandora.ca/sites/default/files/images/It%E2%80%99s%20Dangerous%20to%20Go%20it%20Alo

LDP Containers

LDP has different types of containers for handling resources:

- Basic Container (default in Fedora)
- Direct Container
- Indirect Container

And Now All of Us!

We'll walk through starting up a Fedora Vagrant VM & creating objects, metadata in Fedora 4.

Start Your Engines

`Find your VM and ’vagrant up’`

Fedora 4 Graphical User Interface

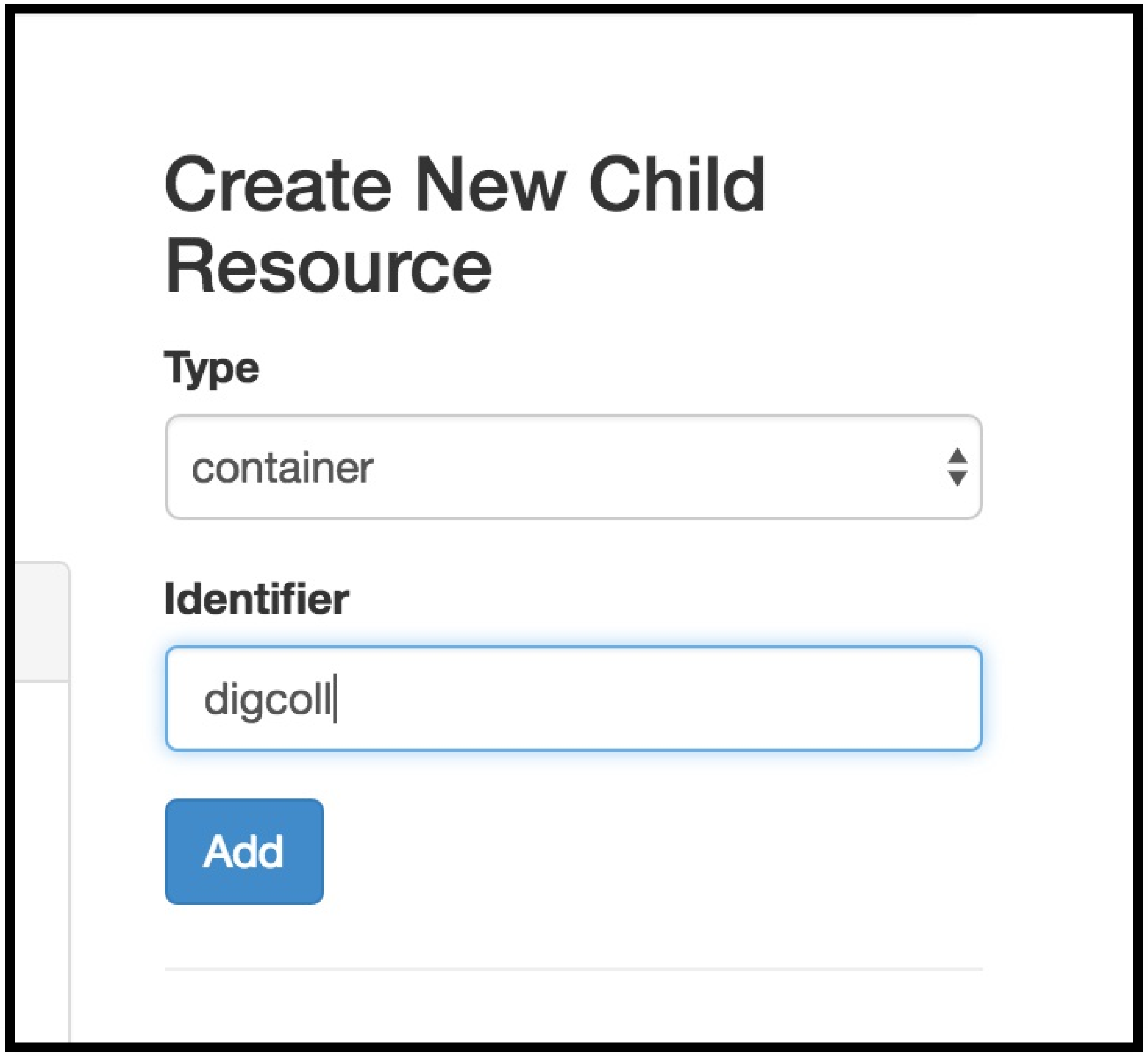
`http://localhost:8080/fcrepo/rest/`

`username: fedoraAdmin`

`password: secret3`

Create Collection Container

Use Fedora 4 GUI to first create new child Container:



Create New Child Resource

Type

container

Identifier

digcoll

Type Container as PCDM:Collection

Then, in page for container you just created, enter following into Update box to type it as a PCDM:Collection:

```
PREFIX pcdm: <http://pcdm.org/models#>
INSERT {
  <> rdf:type pcdm:Collection
        }
WHERE {}
```

Add PCDM:Collection Metadata

Go the page for the Collection you just created, and you can add more metadata if you like:

```
PREFIX dcterms: <http://purl.org/dc/terms/>
PREFIX pcdm: <http://pcdm.org/models#>
INSERT {
  <> dcterms:title
       "Huntington Collection"@en ;
    dcterms:publisher <http://id.loc.gov/authorities/names/n85179829> ;
    dcterms:date "2016"^^xs:date .
}
WHERE {}
```

Create Some PCDM:Objects

Under that Collection, build out a few Objects by iterating the above process and changing the metadata values.

Try Creating non-PCDM Context Class Instance

For example, a foaf:Agent instance that you can then add metadata to (name, email address, etc.). Use the properties here as a guideline:

https://github.com/hybox/models/blob/master/models/agent

Demo the REST API for Updates

You can use curl, Python requests, or a web-based tool like Postman.

1. Get RDF data about a resource:
2. curl -H "Accept: text/turtle" "http://fedoraAdmin:secret3@localhost:8080/fcrepo/rest
3. Create a new Fedora LDP container:
4. curl -X POST -H "Slug: newBook" "http://fedoraAdmin:secret3@localhost:8080/fcrepo/rest

REST API for Batch Metadata Work

```
curl -X PATCH -H "Content-Type: application/sparql-update"
  --data-binary "@body.rdf"
  "http://fedoraAdmin:secret3@localhost:8080/fcrepo/rest/digcoll/newBook

Request Body: (i.e. body.rdf)

PREFIX dc: <http: purl.org="" dc="" elements="" 1.1="">
INSERT {
  <> dc:title "New Book" .
}
WHERE {}
</http:>
```

Open Questions

1. Strings to Things
2. Linking Repository Objects to Descriptive Entities (i.e. BIBFRAME)
3. Application <=> Data Infrastructure
4. Validation, Inference, Rules Oh My!

Thank You!

Christina Harlow / @cm_harlow / cmh329@cornell.edu