Portland Common Data Model (PCDM):

Creating and Sharing Complex Digital Objects

Karen Estlund, Penn State University
Declan Fleming, University of California, San Diego
Mark A. Matienzo, Digital Public Library of America
Jon Stroop, Princeton University Library

Agenda

Purpose and History

Technical Details

Current Work

Future

Q & A

Why a common data model?

- Common language and concepts for:
  - simple to complex object structure
  - rights assertions and access control
  - technical metadata
- Allows for shared objects and coding efforts rather than reinventing for every instance
- Aggregates many minds, projects, and products around difficult concepts
- RDF approach allows for extensibility for the future and local customization

PCDM History

- Hydra implements basic RDF 4/2012
- UCSD complex object data modeling work published 10/2012
- Oregon RDF work - linked open data and vocabularies registry 2013
- Stanford RDF meeting 10/2013
- Hydra Connect 2014
- Use cases on Github - open process 2014 - (more than developers)
- Portland meeting 11/2014
- Portland Code4Lib PCDM Meeting 2/2015
    - concepts like “works,” “resource,” FRBR removed from the model
    - embraced it as an abstract object model not bibliographic

PCDM - More history

- Whiteboard modeling exercises 2015
- OR 2015
  - Open solicitation of input
  - refined ordering - still in progress but a working approach exists
- Still evolving
- “Portland” - where we met
- “Common Data Model”
  - Because it’s not just Hydra
    - Islandora, Europeana have same data needs
  - History of the naming
    - avoiding baggage
    - neutral model

Technical Details - Karen

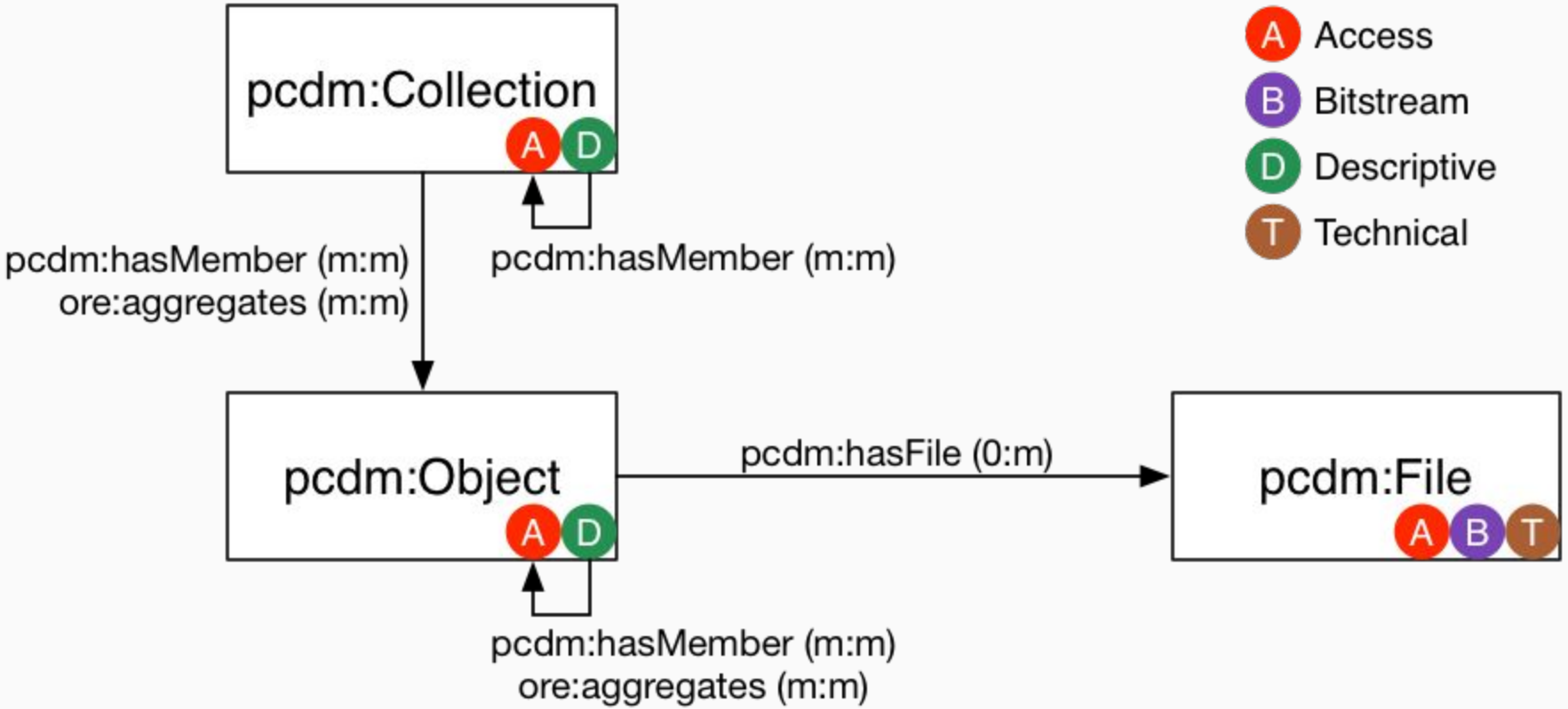
Things

- Collection
- Object
- File

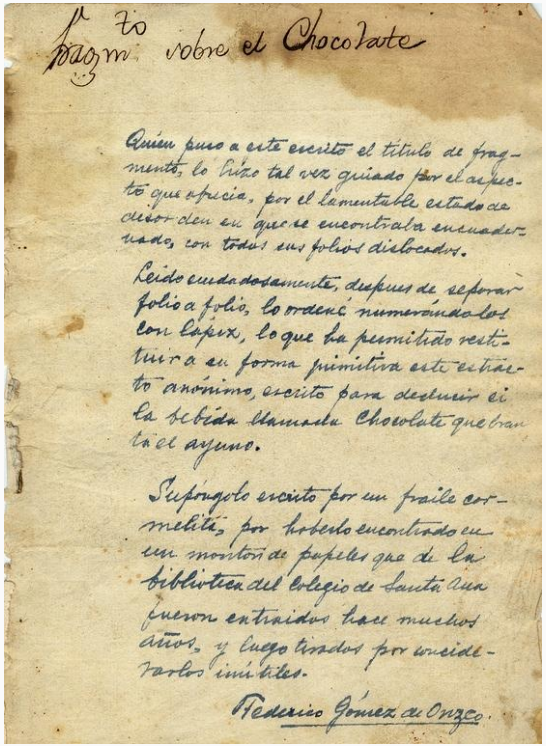
Relationships

- Has Member
- Has File
- Aggregates
- Has Related File

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



PCDM Domain Model



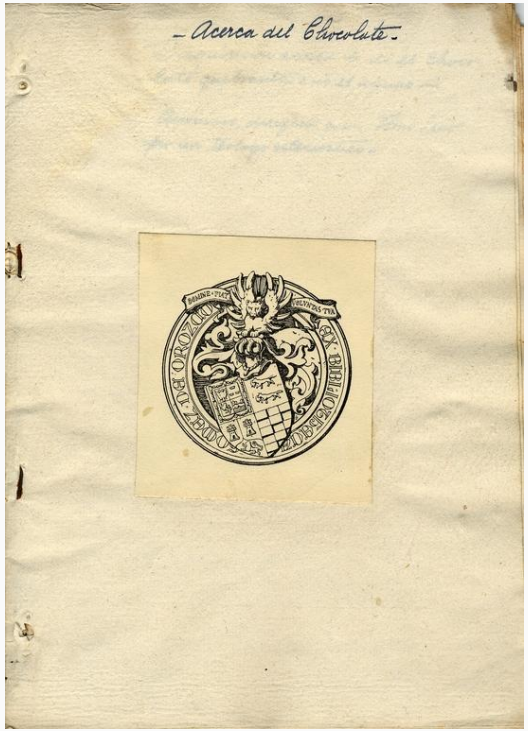
Fragmto sobre el Chocolate

Quien puso a este escrito el título de fragmento, lo hizo tal vez guiado por el aspecto que ofrecía, por el lamentable estado de desorden en que se encontraba encuadernado, con todos sus folios dislocados.

Leído cuidadosamente, despues de separar folio a folio, lo ordené numerándolos con lápiz, lo que ha permitido restituir a su forma primitiva este extracto anónimo, escrito para deducir si la bebida llamada Chocolate quebranta el ayuno.

Supóngolo escrito por un fraile carmelita, por haberlo encontrado en un montón de papeles que de la biblioteca del Colegio de Santa Ana fueron extraidos hace muchos años, y luego tirados por considerarlos inútiles.

Federico Gómez de Orozco.

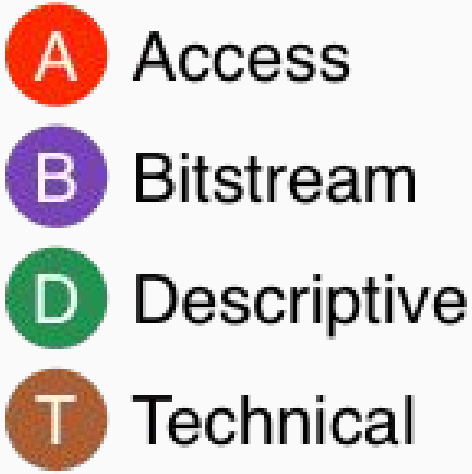


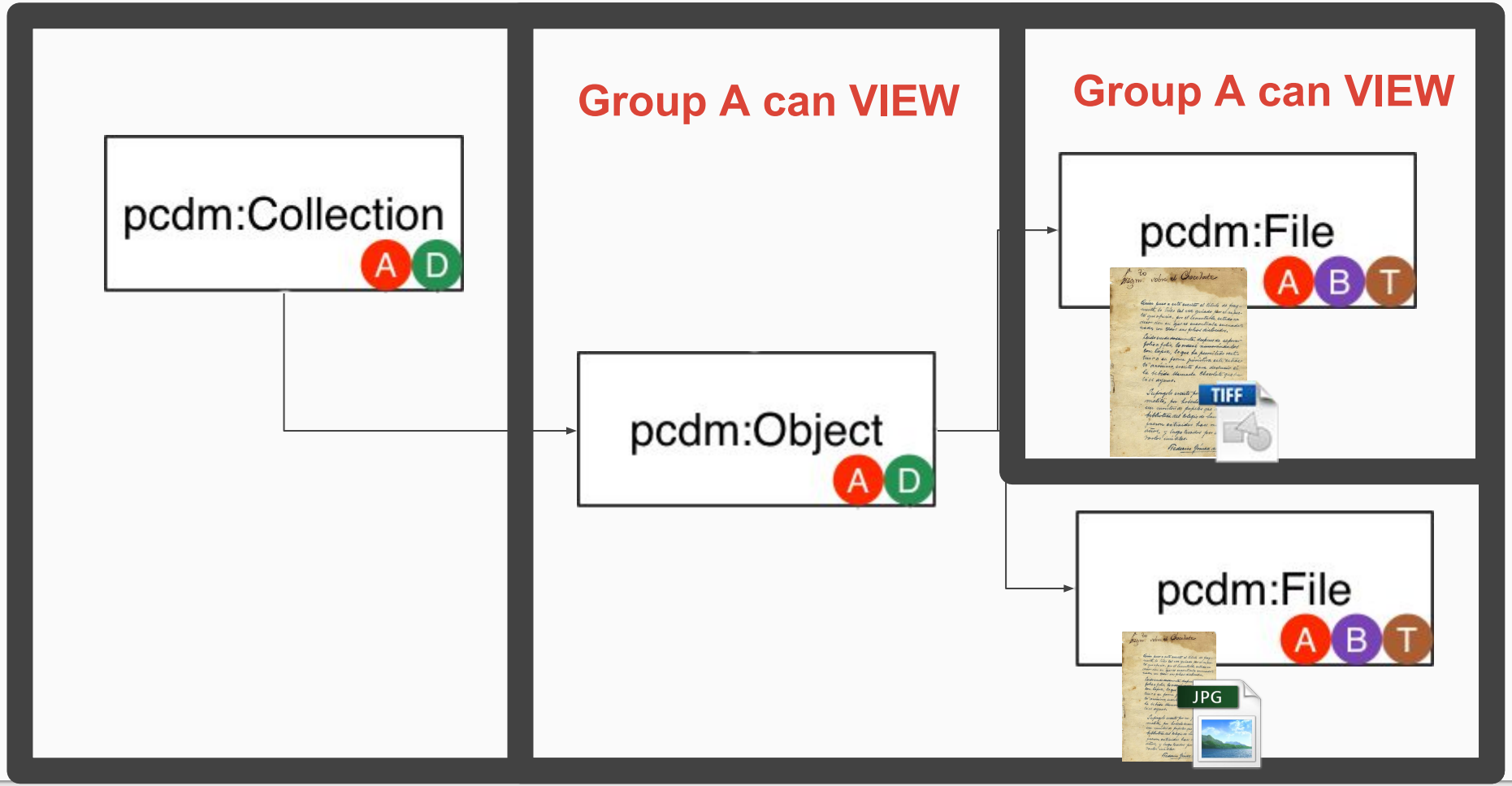
– Acerca del Chocolate –

Example: Acerca del chocolate. Mandeville Special Collections Library, UCSD

What Can go Where?

...and what can we say about different types of things





Access Control - pcdm:Collection, pcdm:Object, and pcdm:File

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

2013-11-29T13:56:46-04:00

d3b07384d113edec49eaa6238ad5ff00

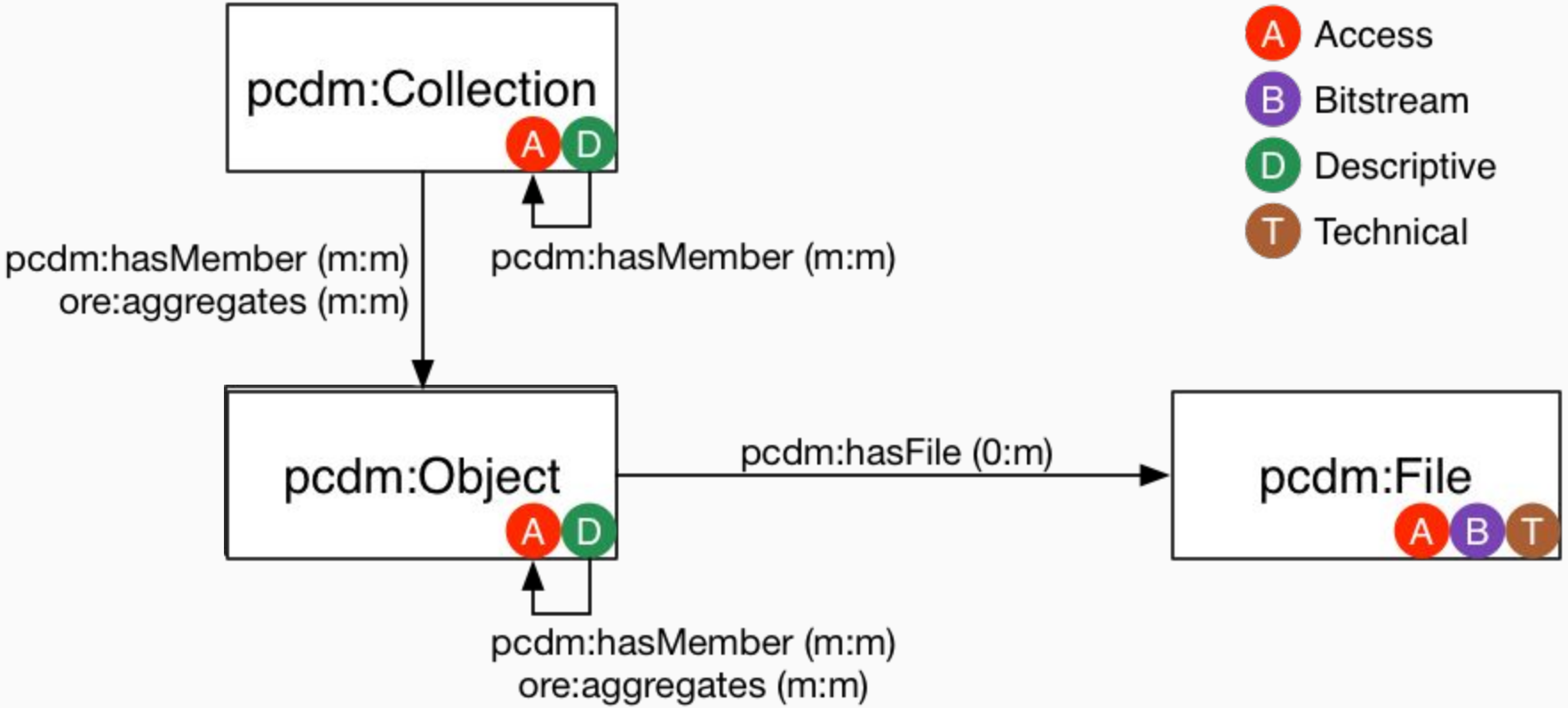
Intermediate File

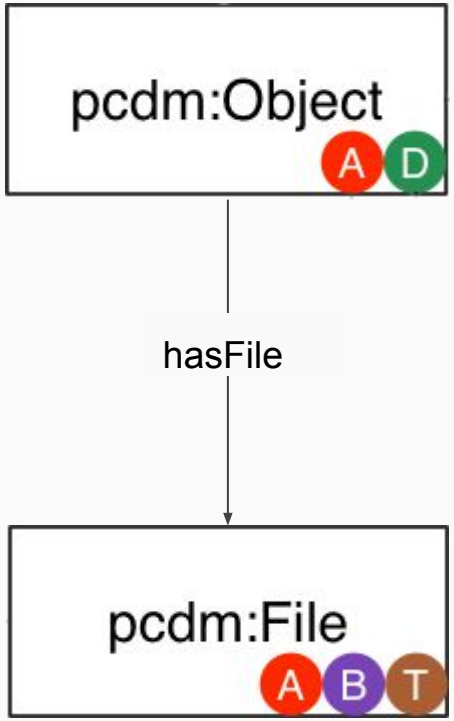
527 KB

20775-bb3482101x-1-1-3.jpg

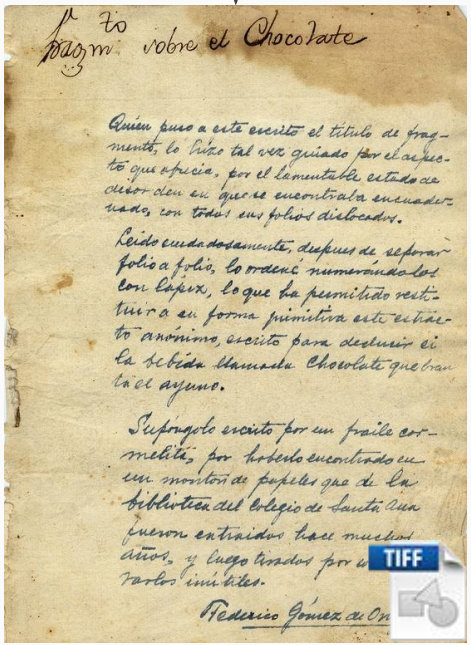
0000E32F
90990902

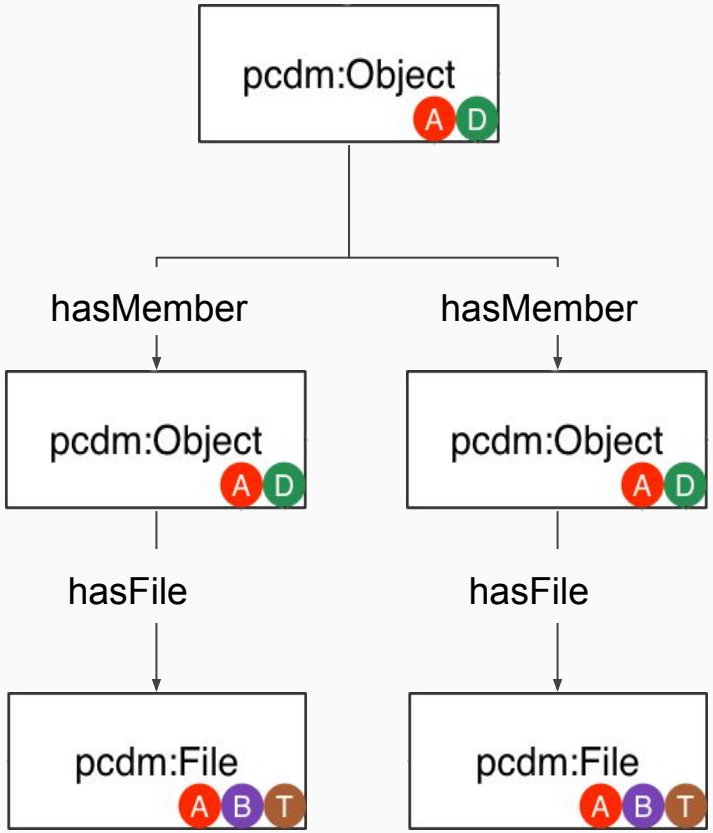
image/tiff


pcdm:Collection
A
D
pcdm:hasMember (m:m)
ore:aggregates (m:m)
pcdm:hasMember (m:m)
pcdm:Object
A
D
pcdm:hasFile (0:m)
pcdm:File
A
B
T
pcdm:hasMember (m:m)
ore:aggregates (m:m)
A Access
B Bitstream
D Descriptive
T Technical

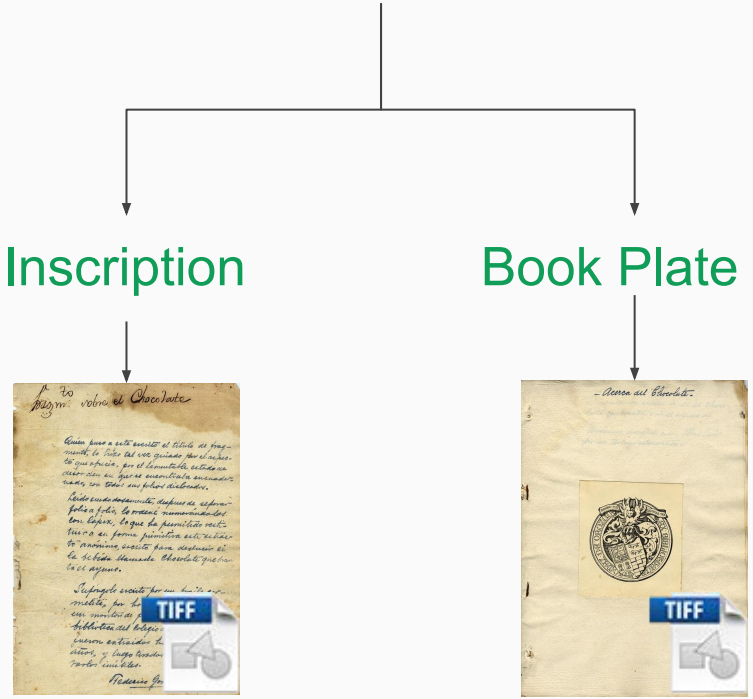


Acerca del chocolate

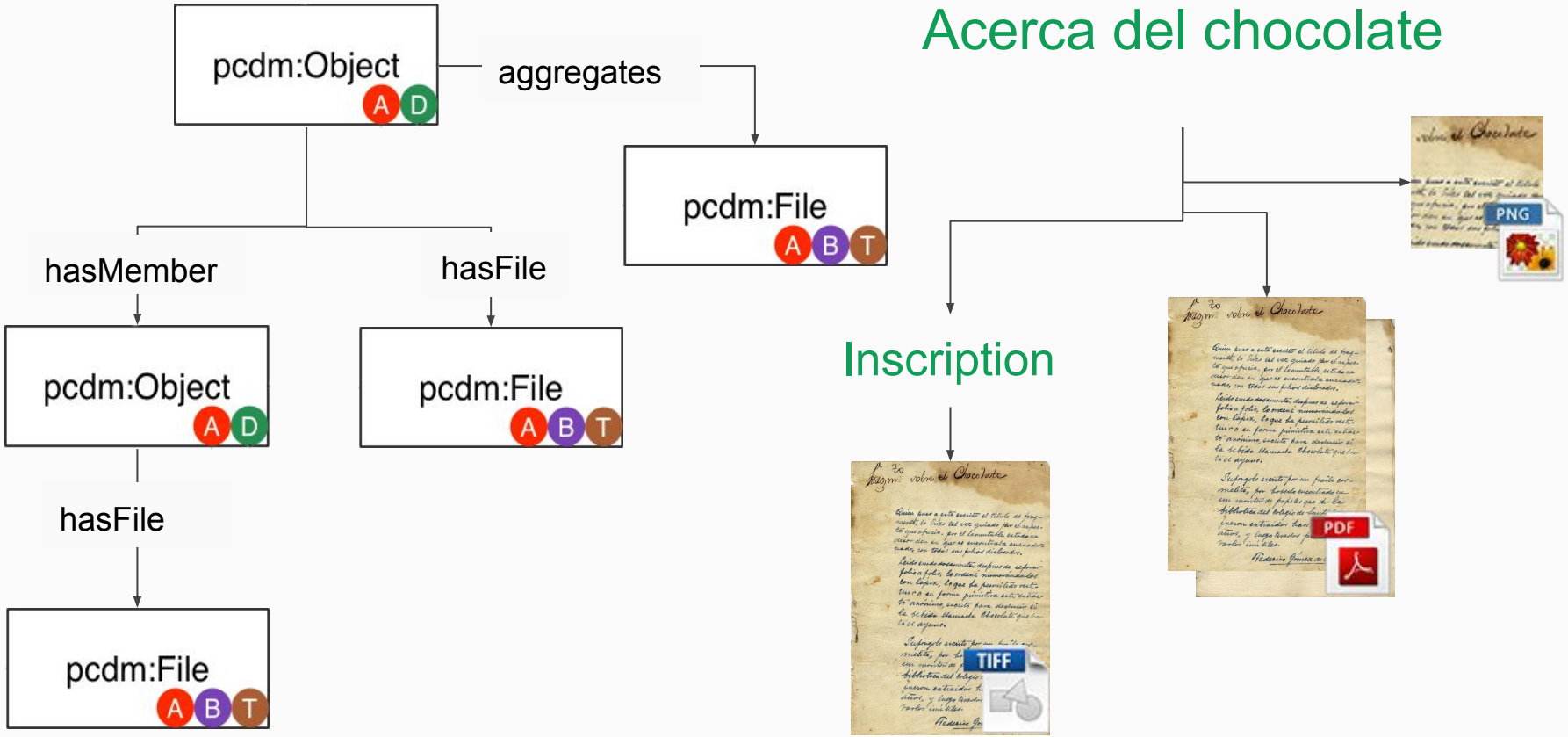


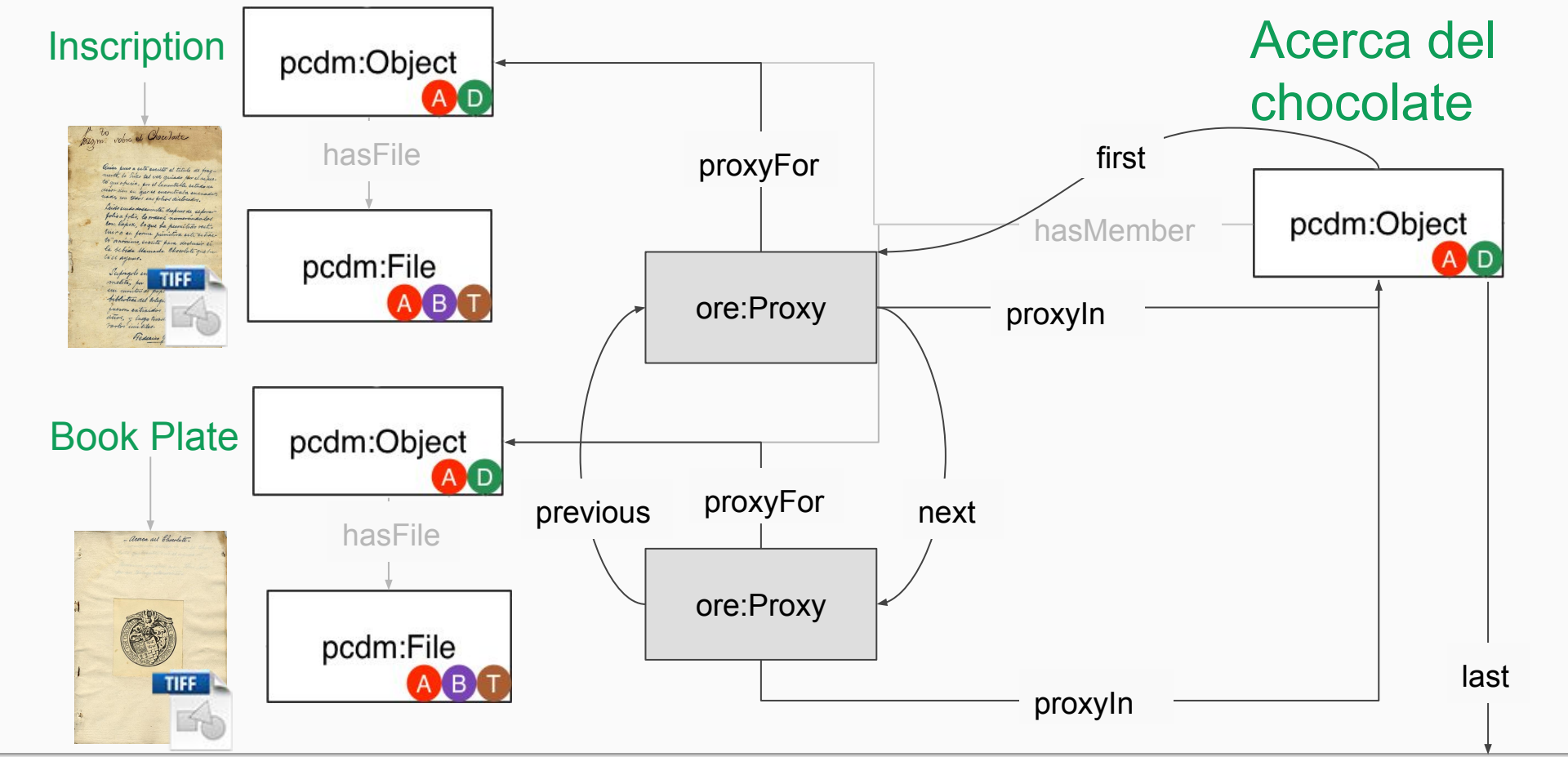


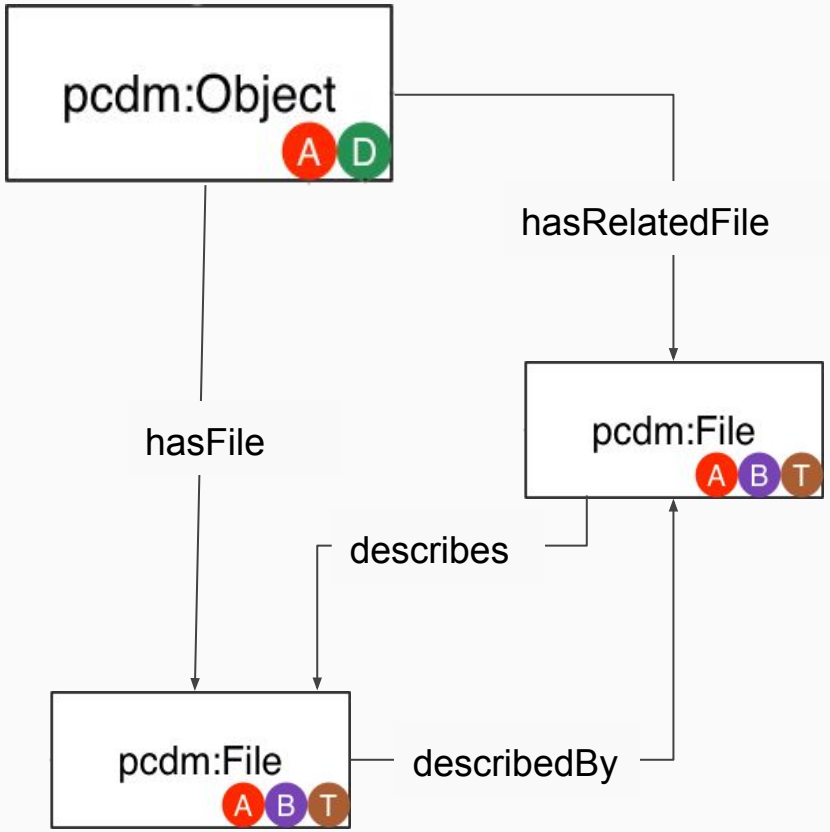
Acerca del chocolate



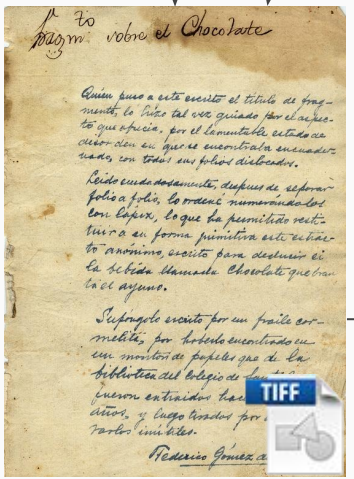
More than one Object


pcdm:Object
A D
aggregates
pcdm:File
A B T
hasMember
hasFile
pcdm:Object
A D
pcdm:File
A B T
hasFile
pcdm:File
A B T
Acerca del chocolate
Inscription
PNG
PDF
TIFF

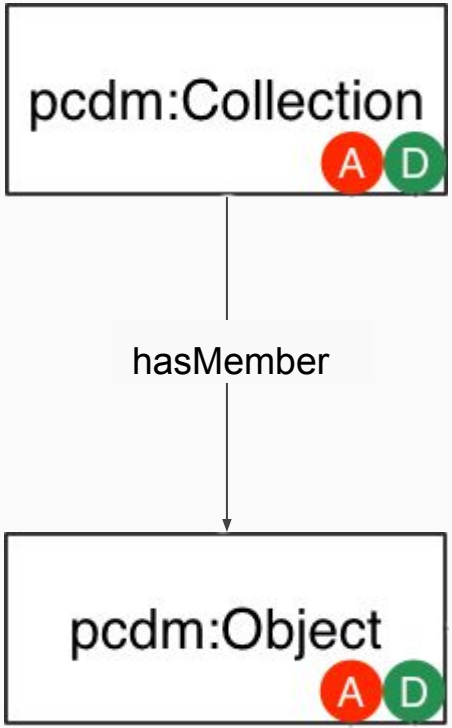

Inscription
pcdm:Object
A D
hasFile
pcdm:File
A B T
proxyFor
Acerca del chocolate
first
hasMember
pcdm:Object
A D
ore:Proxy
proxyIn
Book Plate
pcdm:Object
A D
hasFile
pcdm:File
A B T
previous
proxyFor
next
ore:Proxy
proxyIn
last



Inscription

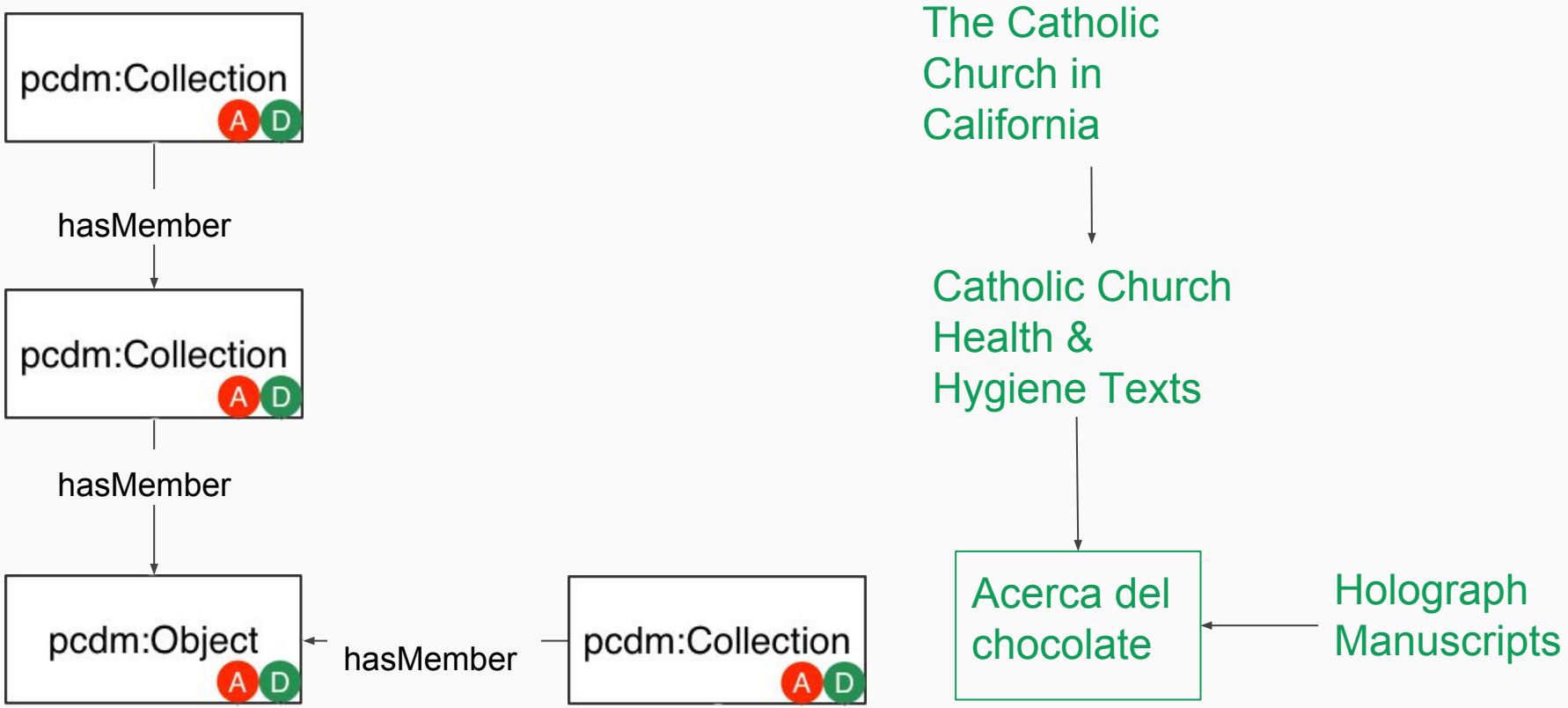


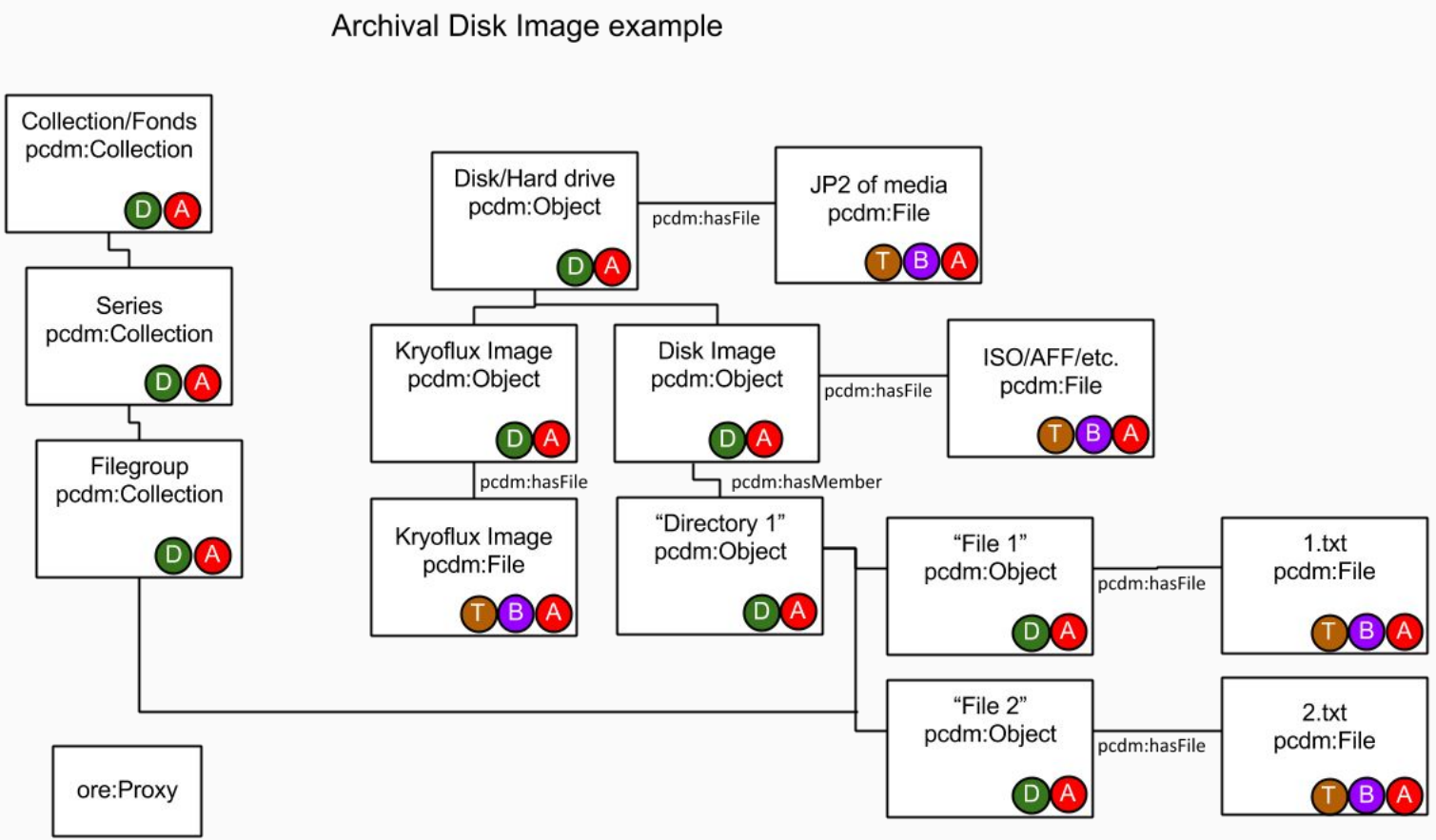

```
- <xml>
  - <identification>
    - <identity format="Plain text"
      mimetype="text/plain" toolname="FITS"
      toolversion="0.8.x">
        <tool toolname="Jhove"
        toolversion="1.5"/>
        <tool toolname="file utility"
        toolversion="5.04"/>
      </identity>
    </identification>
</xml>
```



18th Century Catholic Church Medicinal Texts

Acerca del chocolate

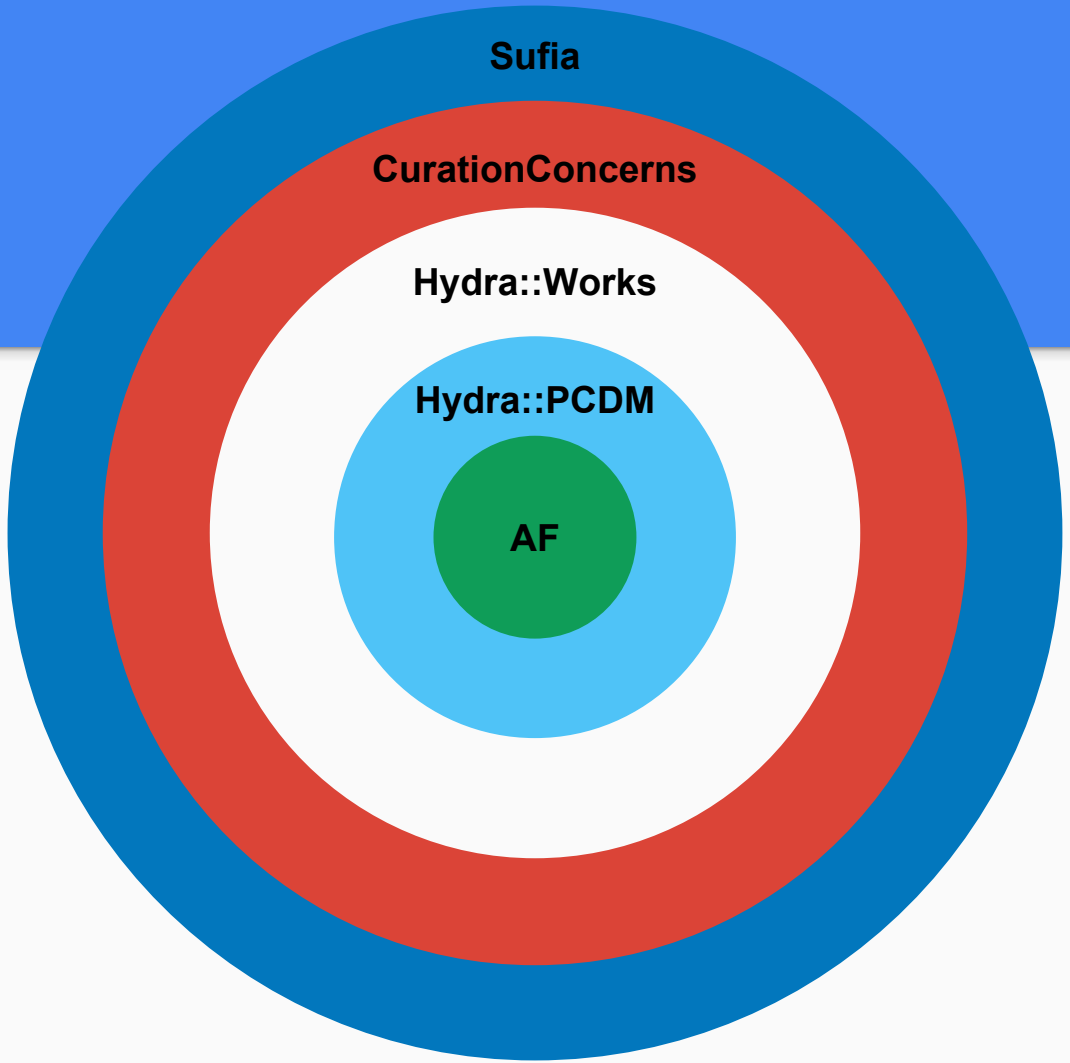

pcdm:Collection
A D
hasMember
pcdm:Collection
A D
hasMember
pcdm:Object
A D
hasMember
pcdm:Collection
A D
The Catholic Church in California
Catholic Church Health & Hygiene Texts
Acerca del chocolate
Holograph Manuscripts



https://wiki.duraspace.org/display/FF/PCDM+Mappings+-+Reference+Diagrams+for+Comment

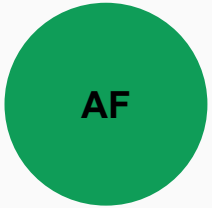
Reference Diagram from Use Case Exercise

PCDM in Hydra



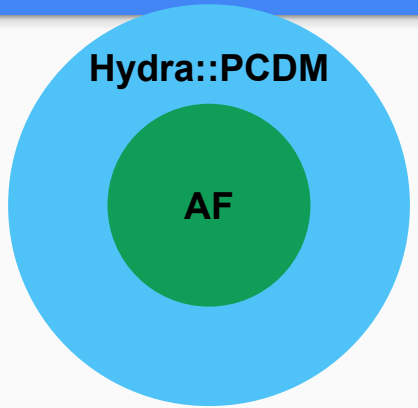
hydra

PCDM in Hydra



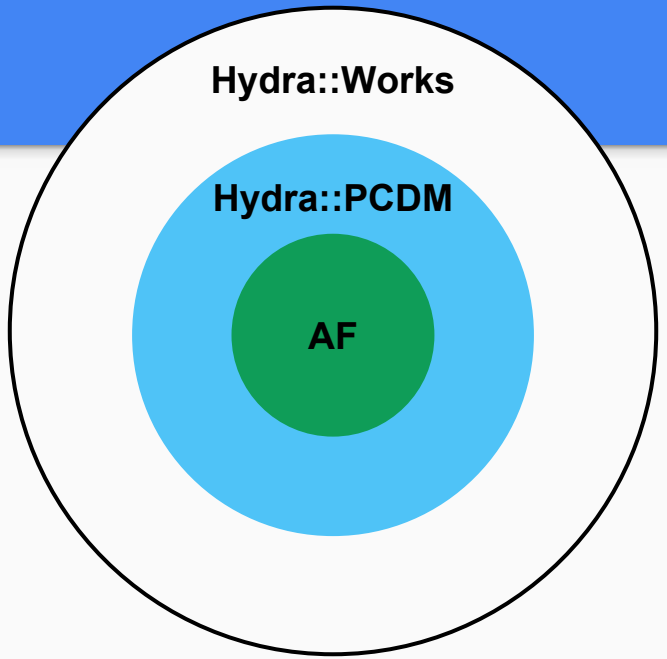
Work with Fedora in a way that feels like Ruby on Rails

PCDM in Hydra



Build models based on PCDM objects and relationships

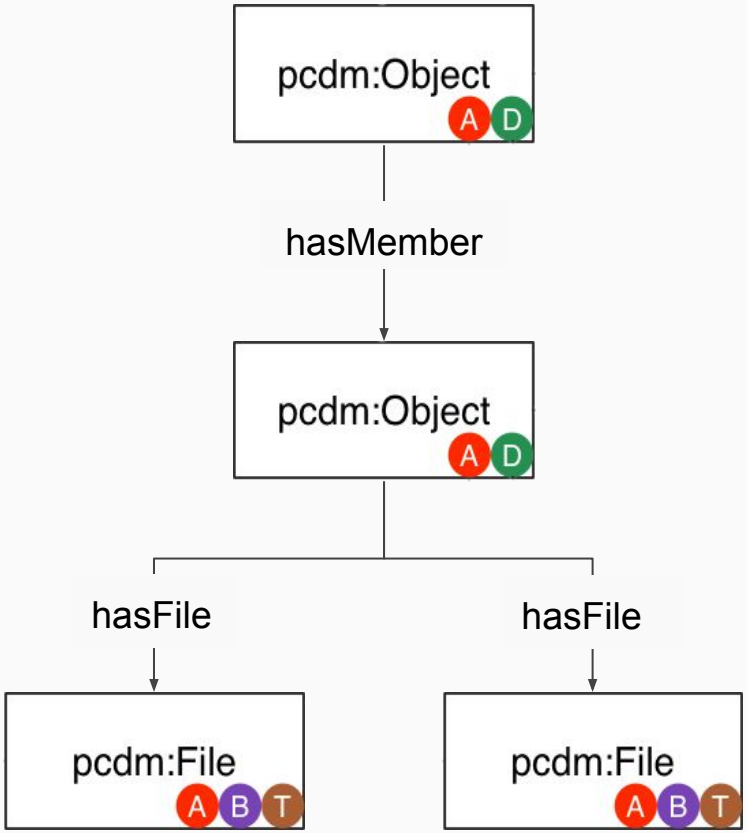
PCDM in Hydra

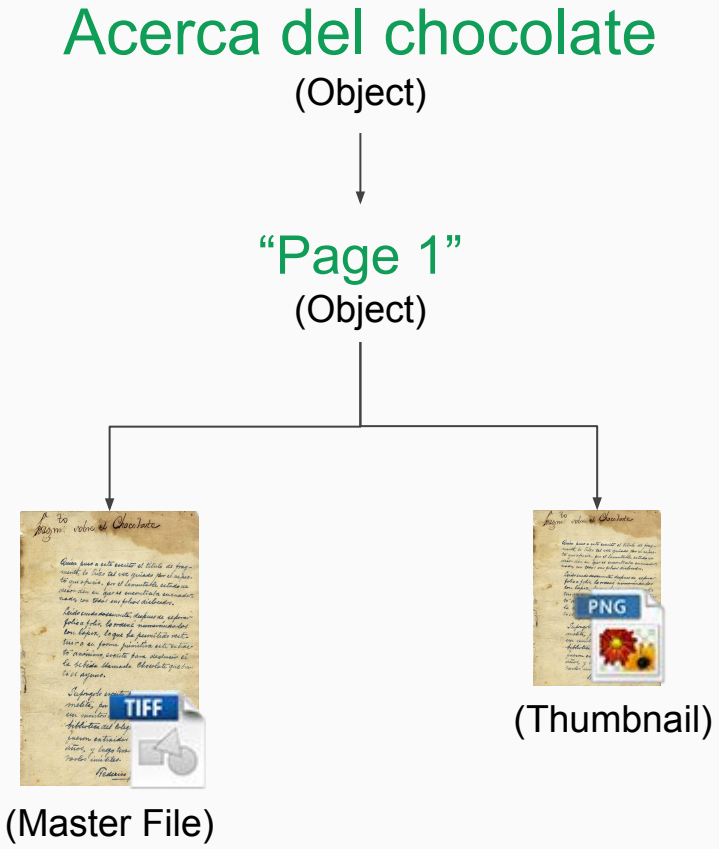


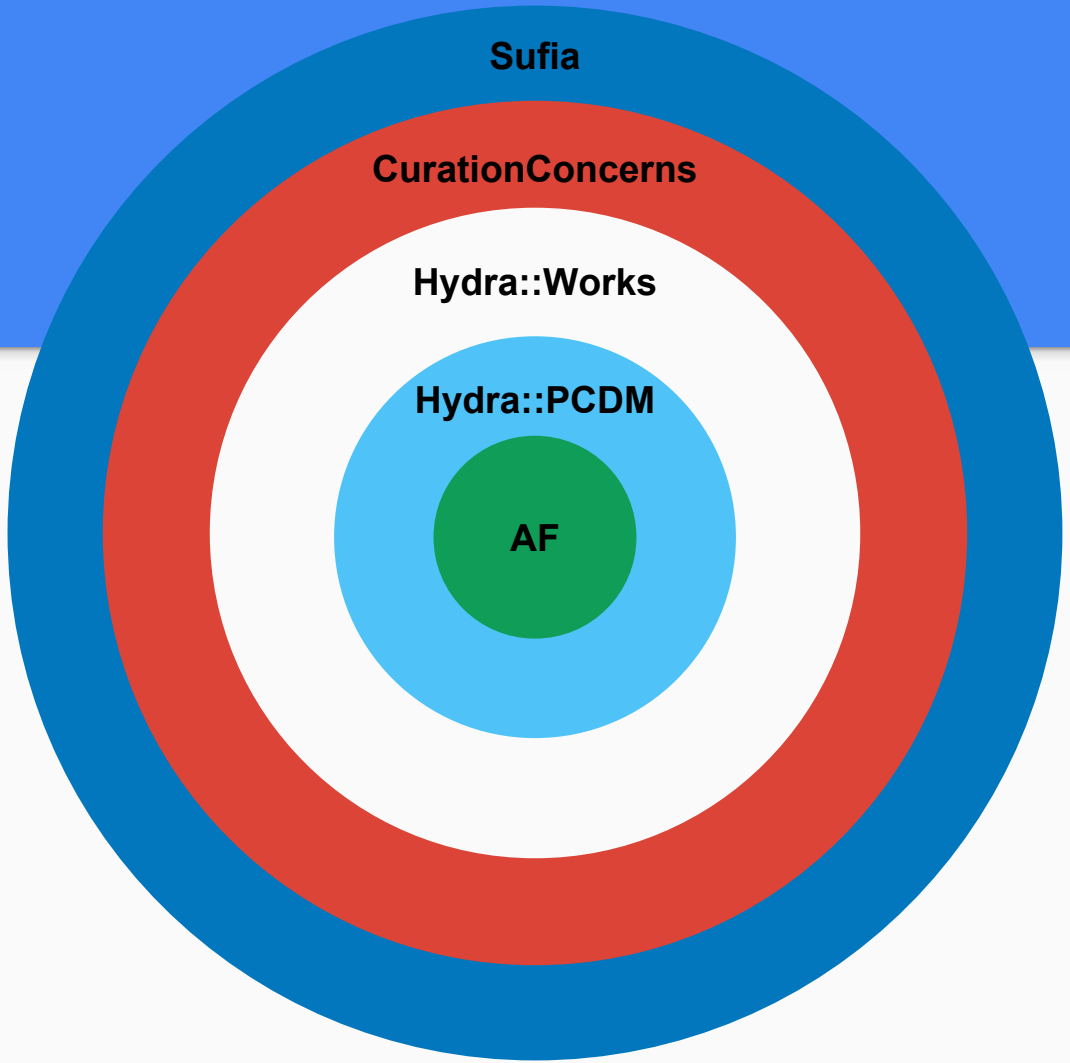
- `Work` Objects
- `FileSet` Objects
- Group `FileSets` to create `Works`

This also allows for shared code around::

- Characterization (JHOVE, DROID, etc.)
- Make Derivatives
- Virus Checking
- Text Extraction


pcdm:Object
A D
hasMember
pcdm:Object
A D
hasFile
hasFile
pcdm:File
A B T
pcdm:File
A B T

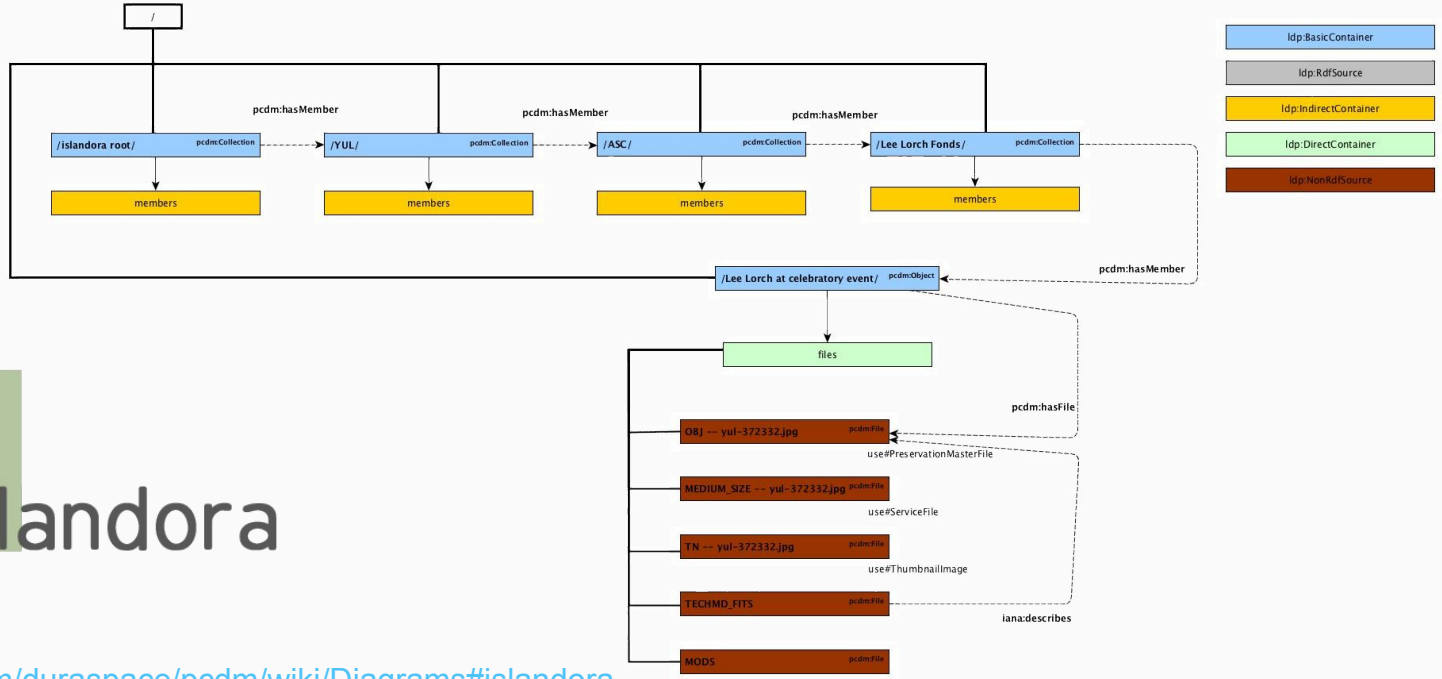

Acerca del chocolate
(Object)
“Page 1”
(Object)
TIFF
(Master File)
PNG
(Thumbnail)



PCDM in Hydra

- Custom Models
- Routing for CRUD
- File Auditing / Versioning
- Single-Use links
- Upload File Sets
- Access Controls
- Leases and Embargoes
- Proxy Deposit
- User Dashboard & Profiles
- Featured Works & Researchers
- Contact Form
- Activity Streams
- Upload via Cloud Providers

PCDM in Islandora



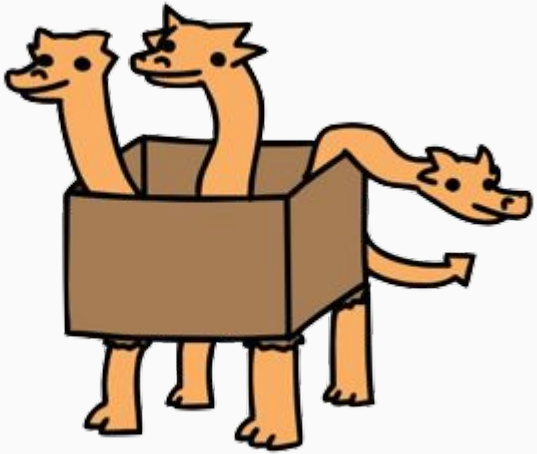
https://github.com/duraspace/pcdm/wiki/Diagrams#islandora

The Future of PCDM

- Within Hydra, Islandora, and Fedora communities
- Interoperability across other communities
- Governance, engagement, and maintenance

The future of PCDM:
Within Hydra/Islandora/Fedora communities

- Hydra in a Box



- Use and development of PCDM recommendations & extensions
  - Rights metadata
  - Technical metadata
    - File format genre ontology
    - File use vocabulary

The future of PCDM: Interoperability across communities

- Alignment with/mappings to additional community models
  - International Image Interoperability Framework
  - Europeana Data Model
  - DPLA Metadata Application Profile
  - RightsStatements.org
- Expansion into W3C standards development processes
  - Linked Data Platform
  - WebAccessControl

The future of PCDM: Engagement, governance, and maintenance

- Community engagement beyond the Hydra, Islandora, and Fedora 4 communities
- Determining best venues for sustained community engagement
- Vocabulary management infrastructure
- Identifying the best long-term home for PCDM

How to learn more and get involved

- Visit the PCDM Wiki: https://github.com/duraspace/pcdm/wiki
- Join the PCDM Google Group: https://groups.google.com/d/forum/pcdm
- Participate in working groups:
  - Hydra Metadata Interest Group
  - Islandora Metadata Interest Group

- **Questions for attendees**
  - Where should we go next?
  - What do your communities need?
  - How we can get you more involved?
  - What won't work for you?

Thank you!

Karen Estlund
Penn State University

Declan Fleming
University of California, San Diego

Mark A. Matienzo
Digital Public Library of America

Jon Stroop
Princeton University Library