

Data Modeling 101: What, How, and Why it's more than PCDM

Monday, October 3rd, 9 AM - 12:30 PM

Location: [Boston Public Library Central Location, Rabb Lecture Hall](#)

Public (read-only) link to this document: <http://bit.ly/HC16DataModelWkshop>

! These workshop details could still change before the workshop. Links to other resources will be added as they are ready for public consumption. Please feel free to add comments with questions or feedback to this document.

Session Details

Abstract

This will be a half-day, hands-on workshop covering data modeling primarily in RDF. We hope to bring a diverse group of Hydra community members together to learn, discuss, and build out examples that will inform Hydra community best practices for data modeling. This modeling work will be taught in the context of helping Hydra and Fedora development, metadata, and interoperability efforts. We will discuss how model uses a number of standards, and demo the different ways to represent models. We will compare and contrast data modeling with metadata standards/profiles. We will walk through modeling efforts around PCDM and its place in our work and community - but this workshop will not focus on PCDM alone or even primarily (this is not a PCDM or RDF workshop). We want this workshop to bring together, develop and engage a larger corps of data modelers in the Hydrasphere.

Target Audience

All. Audience experience and expertise diversity is welcome, as modeling is at the intersection of many systems, workflows, applications, and work areas. Audience experience level aimed at beginning - intermediate.

Facilitators

Julie Hardesty, Christina Harlow, Tom Johnson, Mark Matienzo

Communication Channels

For any discussions, questions, comments or other before, during and after the Data Modeling Workshop, check out the [#hc2016-datamodeling](#) channel on the [Project Hydra Slack](#).

Workshop Resources

Links will be added / opened up as the resources become available for public sharing.

- [Link to this Workshop's Google Drive Folder \(containing all related resources\)](#)

- [Link to the Session's Shared Notes](#)
- [Link to Slides](#)
- Example resource: <http://bit.ly/HC16DataModelExample>
- [Link to Resources for Moving Forward with Data Modeling](#)

Required & Recommended Workshop Preparation

Participant Requirements

Participants are strongly recommended to bring laptops with internet connection and web browser (not Internet Explorer). We are working to have a space with whiteboards or chalkboards to support breakout group work (so be prepared to take part and move around).

Recommended Preparatory Readings

- Strongly Recommended: Chapter 1 (esp. 1.2-1.4) and, if time, 2.1, 3.1 of [Conceptual Modeling of Information Systems](#) (Don't worry if not everything sinks in; use this reading to just get an idea of intellectual context for the materials)
- Optional: Helpful to review for RDF(S) at high level
 - [RDF 1.1 Primer](#)
 - [RDF Schema Spec](#)
- Optional: Browse [Ontology Design Patterns](#)

Workshop Schedule

9 - 9:15 AM	Workshop Introduction
9:15 - 10:15 AM	Introduction to Data Modeling - How Models work with Systems
10:15 - 10:30 AM	15 minute break
10:30 - 10:45 AM	Introduction to Data Modeling - How to express data models
10:45 - 11 AM	Break out groups: Data Modeling Examples - Full Group Breakout Setup
11 - 11:45 AM	Break out groups: Data Modeling Examples - Breakouts
11:45 - Noon	15 minute break
Noon - 12:15 PM	Break out groups: Data Modeling Examples - Full Group Breakout Recaps
12:15 - 12:30 PM	Build follow-up, Community, Share Resources Going Forward

Workshop Outline / Topics

Introduction - 15 minutes - Mark

- Our expectations of you
 - Follow the [Project Hydra Code of Conduct](#)
 - Follow [Recurse Center \(Hacker School\) Rules](#), please
- This is an informal session, so ask questions whenever, and also reach out to co-leaders
- Recap the abstract information:
 - Help build data modelling capacity
 - Help build best practices for data models
- Workshop facilitators' goals:
 - Share motivations for data modeling as part of the application development process.
 - Equip participants with knowledge needed to instigate this work at their own institutions and participate in broader community discussion.
 - Demonstrate modeling practices and pitfalls.
 - Provide context for standards work in the Hydra community:
 - Interoperability of “structural model”;
 - How modeling fits in with other metadata standards;
 - Models as an application artifact.
- However, this is not a:
 - LD/RDF workshop
 - Though we're happy to answer RDF-specific questions as they come up
 - Metadata standards workshop
 - Though we hope to clarify the role of interactions between them, data modeling, and application design
 - PCDM workshop
 - Though we will be touching on PCDM later as an exemplar
 - But other options exist for learning more about these topics.
- Audience's goals:
 - Write on whiteboard, shared notes, somewhere to revisit.

Introduction to Data Modeling - 1 hour 15 minutes

- How Models work with Systems - 1 hour
 - Motivating Modeling - Tom (40 minutes)
 - What do systems “know”?
 - Modeling for:
 - design feedback
 - engineering documentation
 - dissemination, interchange & interoperability
 - consistency, validation, process
 - Data Models - Tom

- Multiple meanings of “data model”
 - “Conceptual Model/Schema” (as discussed here) vs. “Abstract Syntax” (e.g. RDF Graph)
 - Why we’re focusing mostly on RDF as opposed to other models
 - Acknowledging semantic modeling biases; contrast with other approaches to conceptual modeling.
 - Models in a context - Christina (20 minutes)
 - Big question: what does data modeling buy you in ...
 - Preservation context
 - Models versus Serializations
 - Database Structures
 - Data Exchange Protocols
 - Application Logic
 - Metadata Standards
- How to express data models (high level overviews) - 15 minutes - Christina - will look to co-leaders to ad-libs
 - So we’ve learned a bit about Models, we’ve seen some of the syntax we’re about to review, but it’s good to know how this work gets represented. This is a quick and dirty rundown of what you will see often - or probably have already seen - and it will help lead us into the hands on portion of this workshop:
 - RDF (dump if not enough time)
 - XML (dump if not enough time)
 - UML (dump if not enough time)
 - Java (other lang?) classes (dump if not enough time)
 - Good ol’ whiteboards :-) (i.e. diagramming)

Break out groups: Data Modeling Examples 1 hour 15 minutes

- Full Group: - 15 minutes - Christina
 - High level models discussion (for sake of knowing who wants to join what)
 - PCDM
 - Europeana / DPLA
 - IIF
 - Get specific groups of use cases and have folks model them
 - Have this work stored somewhere all can
 - Each leader helps with a group
 - Breakout group example: book
- Breakout Groups: - 45 minutes
 - PCDM - Julie
 - Hydra stack (PCDM through Fedora, Gems) leading to...
 - LDP’s interactions, expectations
 - ORE’s basis
 - Europeana / DPLA - Tom
 - Motivations

- Descriptive model focused on aggregation
 - Not complete for everyone, but basis for building profiles
 - Not designed to be exhaustive
 - These models are more concerned with description than system functions.
 - Talk about how these two concerns tend to co-exist in the Hydra community
- IIIF - Mark
 - Highlight what kinds of objects exist in the model's domain & what objects it chooses not to know about.
 - Manifests in context of use
- Full Group Again: Christina - 15 minutes
 - Break out report back + Discussion:
 - Go through how you would have these models possibly interact in a system
 - Highlight differences, similarities in approaches
 - Highlight areas of possible modeling reuse
 - Address questions like: where do the models we made work & where could they cause tension in our current workflows/processes?

Build follow-up, community, Resources going forward - 15 minutes

- Maintain momentum (ask participants) - Mark + Julie
 - Hydra MIG work - Julie
 - What communication channels to use for this ongoing work
 - Document our workshop but other spaces to watch
 - How to continue to support data modeling development in our communities
- Data Modeling Tools + Resources (share as a doc or other)
 - List of Tools
 - Links to other models in the cultural heritage community
 - Places to see issues unresolved for modeling questions
 - Where Hydra community could use data modelers (or data modeling work is already happening)

General Notes About this Session

- This is not an RDF or Linked Data tutorial, although it's focused heavily on RDF modeling (as opposed to relational/entity-relationship, other).
- This is not a PCDM tutorial, though we will use PCDM in some places of the workshop as an example to work with/learn from.
- We will try to spend some time on "how to data model", along with mention of some tools, resources, exemplars of documentation, etc.
- We want to focus on best practices around integration across systems, using standards.
- We want to mention [data modeling design patterns](#) to use and pitfalls to avoid.
- We want to compare and contrast metadata standards versus data modeling.

- We want to demonstrate, involve participants in experiencing, how a data model might be expressed in different ways (RDF, XML, UML, Java classes) and how the model underlies any individual expressions.
- One big objective for this workshop is to begin developing corps of data modelers in the HydraSphere as well as contextualize this community in other/larger communities.
- Another big objective for this workshop is to create a thread of discussion and forum for ongoing discussion for this “corps of data modelers in the HydraSphere”.