# DHSI Workflow, Timeline, and To-Do

# Workflow Timeline

# March to Mid-April: Days 1 & 2

# Mid-April to Mid-May: Days 3 & 4

# Mid-May to June: Day 5

# To Do:

# (by March 24th)

# Coursepacks

# To Do:

# (by April 15th)

# Carolyn:

* Slides for Introduction to metadata (Day 1)
* Slides on schemas/ontologies (Day 2)
* Exercises on schemas/ontologies (Day 2)

**Sean:**

* Slides(?) for OpenRefine Demo (Day 1)
* Slides on crosswalking/transforming metadata (Day 2)
* Exercises on crosswalking/transforming metadata (Day 2)

**Carolyn/Sean**

* Exercises for working with OpenRefine (Day 1)
* Sean to rewrite description for XSLT session
* Research available datasets, what types of metadata transformations people in DH want to do (we could talk to James Lee; Carolyn could also put the question out on the DHSI listserv or Twitter)
* Think about possible course e-exhibit to show on last day

# ----------------------------------------------------------------------------------------------------------------------------------

# Daily Breakdown

# Monday: 10:15 to noon, 1:30 to 4

# TWR: 9 to noon, 1:30 to 4

# Friday: 9 to noon, 12:15-1:15 (lunch/course e-exhibits)

# Title:

Beyond TEI: Metadata for Digital Humanities

# Summary:

High-quality metadata is essential for the description, discovery, and preservation of DH projects. While TEI is the most used metadata standard in DH, there is so much more to learn and explore! This course will introduce metadata schemas and standards such as Dublin Core, VRA, controlled vocabularies, and linked data and RDF. We will also discuss ontologies, ethics of standardization, data management, and digital preservation. Hands-on work with participants' own datasets will be given to practice metadata/data cleaning with OpenRefine, creating custom schemas, and linking to external authorities. Students need no prior experience with metadata or programming.

# Schedule:

*Day 1*

Morning/Early Afternoon – Introduction to metadata (Carolyn/Sean). Here we will discuss types of metadata, standards and schemas, ontologies, controlled vocabularies, ethics, and editorial policy. Participants will also share their project goals, what kinds of documents or data they are working with, and what schemas they are considering.

Morning Outline:

* Start with introductions from Carolyn/Sean
* Participants introduce themselves and their projects, what kinds of documents/ data they are working with and what schemas they are considering…
* Ask about downloading OpenRefine / technology problems

Needed for Session:

* Slides for Introduction to metadata (Carolyn)

Readings:

* Something from “Introduction to Metadata” (by the Getty, Carolyn requested 3/7 through ILL)

Afternoon – Hands-on time (Carolyn/Sean). An introduction to the OpenRefine tool for cleaning metadata will be provided and participants will be given project-based exercises to learn OpenRefine’s basic functionality when working with humanities data.

Afternoon Outline:

* OpenRefine Demo (Sean)
* Free time for students to do exercises using OpenRefine (Carolyn/Sean float)

Needed for Session:

* Slides(?) for OpenRefine Demo (Sean)
* OpenRefine Exercises (Carolyn/Sean to create)

Readings:

* Something from “Using OpenRefine” by Ruben V. (ebook available through Serial Solutions)
* Others?

*Day 2*

Morning – Choosing schemas and ontologies (Carolyn). We will discuss how to choose schemas when working with common DH software such as Omeka, Drupal, and Neatline, as well as how to create custom schemas. Exercises will be provided to give participants experience working with different schemas as well as to see how their data changes depending on the schema used.

Different influences on which schema you,choose (structure of data, temporal considerations, how that can be important, impact on workflow, interoperability), example of machine-readability

-data looks different when you export than it does in the application

-streaming into different contexts, mashups, remixing, applying to different contexts

Morning Outline:

* Lecture on schemas (Carolyn)
* Free time for students to work on exercises (Carolyn/Sean float)

Needed for Session:

* Slides on schemas/ontologies (Carolyn)
* Exercises on schemas/ontologies (Carolyn)

Afternoon – Curatorial decisions, ontology discussion and examples (Carolyn to rewrite)

Exercises: HoR cleaning, What’s on the Menu Remapping, Crime ontology mapping

Afternoon Outline:

* Lecture on crosswalking and transforming metadata (Sean)
* Free time for students to work on exercises (Carolyn/Sean float)

Needed for Session:

* Slides on crosswalking/transforming metadata (Sean)
* Exercises on crosswalking/transforming metadata (Sean)

*Day 3*

Linked Data and RDF

Morning/Early Afternoon – Linked Data and RDF (Carolyn/Sean). An introduction to linked data, the Semantic Web, RDF, SPARQL, APIs, and Linked Data Fragments will be provided.

Application of things discussed earlier in class.

-Build on “don’t build an API” discussion from DCMI

-Create RDF with OpenRefine, talk about standard, choosing predicates

-Export RDF and see how unreadable It is

-What do they do with the RDF? Is it useful? Is it worth doing for your project?

Morning Outline:

* Lecture on Linked Data, RDF, SPARQL (talk about, but not deep dive), APIs, Linked Data Fragments, etc (Carolyn/Sean to split?)
* Free time for students to work on exercises (Carolyn/Sean float)

Needed for Session:

* Slides on Linked Data, RDF, etc (Carolyn/Sean)
* Exercises on Linked Data, RDF (Carolyn)

Afternoon – Catch-up from morning if needed. Hands-on time with participants. (Carolyn/Sean)

Afternoon Outline:

* Demo on how to use OpenRefine for creating linked data (Sean?)
* Free time for students to work on exercises (Carolyn/Sean float)

Needed for Session:

* Slides (?) for demo on how to use OpenRefine for creating linked data (Sean?)
* Exercises on creating linked data with Open Refine (Carolyn/Sean?)

*Day 4*

TEI and OpenRefine

Morning/Early Afternoon – TEI and OpenRefine (Carolyn/Sean). Participants will learn how to use OpenRefine for preparing, cleaning, and transforming hierarchical TEI metadata.

-Need some TEI

-use Python for this

-take TEI, extract names, reconcile names in OpenRefine against LC, put them back

-CH to research uris in TEI for names, need a TEI dataset with names

Morning Outline:

* Demo/slides(?) on using TEI and OpenRefine (Carolyn/Sean to split?)
* Exercises on using TEI and OpenRefine (Carolyn/Sean float)

Needed for Session:

* Slides (?) for demo on how to use OpenRefine for creating linked data (Carolyn/Sean?)
* Free time for students to work on exercises (Carolyn/Sean float)

Afternoon – Introduce e-exhibit project. Hands-on time with participants (Carolyn/Sean).

Ontology of crime (is there a way to make it look cool?)

-Maybe something with morgue data for an exercise?

Afternoon Outline:

* Introduce e-exhibit project (Carolyn/Sean?)
* Free time for students to work on exercises (Carolyn/Sean float)

Needed for Session:

* E-Exhibit project ideas (give a couple ideas to students and let them decide?) (Carolyn/Sean)

*Day 5*

Data management and digital preservation (Carolyn/Sean). We will discuss best practices for creating metadata and file systems that can be migrated and preserved, including PREMIS, METS, and BagIt. Long-term hosted storage solutions such as institutional repositories and digital consortiums, will be examined. Additional hands-on time.

Morning Outline:

* Lecture on data management and digital preservation (Carolyn)
* Free time for students to work on e-exhibit, own projects, earlier exercises, etc. (Carolyn/Sean float)

Needed for Session:

* Slides on data management and digital preservation (Carolyn)

More Notes:

-something on JSON

-lean heavily on House of Refuge (RDF, application)

-HoR, curatorial decisions about metadata, historical dataset, bias (race, parentage, crimes committed)

-NYPL WHat's on the Menu for discussion of Dublin Core v custom schema, ontologies