# Amended Proposal

# Team Teaching Approach/Justification:

# Initially, this course was proposed with a sole instructor. However, after receiving feedback on the abbreviated workshop version of the course at DHSI16, Carolyn felt the course offering should be revised and a second instructor added. Specifically, workshop participants requested more hands-on time with the metadata cleaning tool OpenRefine and enhanced, customized functionality (ex. TEI transformation, XSLT, and RDF). With the added emphasis on OpenRefine, two instructors will greatly help with troubleshooting technical problems and keep the course on schedule. Carolyn and Sean also have complimentary skills; Carolyn is an expert in metadata and experienced with DH projects and historical datasets, while Sean is a developer with strong technical skills in scripting, application customization and development, and SPARQL. Carolyn and Sean will split course sections as described below. The initial proposal is included at the end of the document.

# 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). 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.

Afternoon – Hands-on time (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.

*Day 2*

Morning – Choosing schemas and ontologies (Carolyn). We will discuss how to choose schemas and ontologies 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.

Afternoon – Crosswalking and transforming metadata (Sean). We will discuss the basics of XSLT as well as GUI-based tools such as MarcEdit. XSLT stylesheets for common DH metadata transformations will be provided and participants will have hands-on time with transformation exercises.

*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, and Linked Data Fragments will be provided. Participants will learn how to use OpenRefine for creating linked data.

Afternoon – Hands-on time with participants. (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.

Afternoon – Hands-on time with participants (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.

**Initial Proposal**

# 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. Here we will discuss types of metadata, standards and schemas, ontologies, and controlled vocabularies.

Afternoon – Metadata standardization, ethics, and editorial policy. We will discuss editorial decisions needed when working with metadata, specifically related to searching and indexing. How can we maintain the authenticity of historical or literary data while creating consistent, standardized data need for DH applications such as databases, digital editions, and data visualizations?

*Day 2*

Morning – Choosing schemas and ontologies. We will discuss how to choose schemas and ontologies, as well as how to create custom schemas. An introduction to the OpenRefine tool for cleaning metadata will be provided.

Afternoon – Hands-on time. Participants will share their project goals, what kinds of documents or data they are working with, and what schemas they are considering. Participants will also have time to work on metadata cleaning with OpenRefine.

*Day 3*

Linked Data and RDF

Morning/Early Afternoon – Linked Data and RDF. An introduction to linked data, the Semantic Web, RDF, and SPARQL will be provided. Participants will learn how to use OpenRefine for creating linked data.

Afternoon – Hands-on time with participants.

*Day 4*

Morning – Data management and digital preservation. 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.

Afternoon – Hands-on time with participants.

*Day 5*

Course wrap-up as well as additional hands-on time. Discussion of what the future holds for metadata as well as the newest developments in standards, schemas, and tools.