

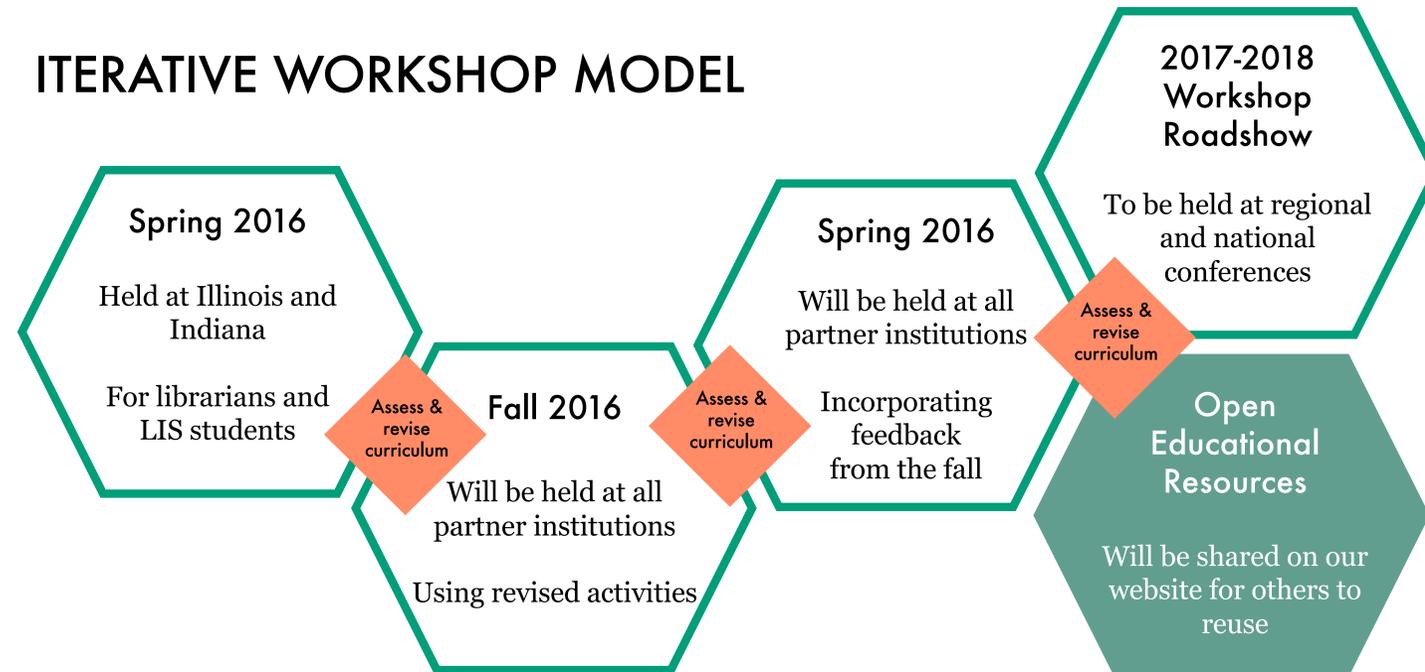
Digging Deeper, Reaching Further: Libraries Empowering Users to Mine the HTDL Resources

PROJECT GOALS

This 3-year project aims to:

- Arm academic librarians with practical skills in text mining methods and knowledge of trends in digital scholarship.
- Situate the academic library as a space for encountering and engaging with “big data” tools and methodologies vis-à-vis the HTRC.
- Train librarians in digital scholarship skills in a way that recognizes the decentralized and fluid nature of digital humanities research support in libraries.

ITERATIVE WORKSHOP MODEL



PROJECT PARTNERS

- University of Illinois
- Indiana University
- Lafayette College
- Northwestern University
- University of North Carolina at Chapel Hill

THE HTRC

The HathiTrust Research Center is the research arm of the HathiTrust. It facilitates scholarly research using the HathiTrust Digital Library by providing mechanisms for researchers to study HathiTrust content using computational tools for text analysis.

INSTRUCTIONAL DESIGN

Outcome-oriented

- Learning outcomes written for each module
- Emphasize concepts

Modular

- Scaffolded instruction
- Translatable into OER
- Partner institutions can remix and reorganize content

Reusable

- Applicable to distinct audiences & unique lab set-ups
- In pilot workshops, test scenarios

Hands-on

- Using HTRC web-based tools
- Basic command line and Python skills
- Building skills in OpenRefine

CURRICULUM

Module 1: Introduction

- Intro to text mining and to the HathiTrust

Module 2: Gathering Textual Data

- Two lessons, one for beginners and one more advanced
- How to create textual datasets/HTRC worksets
- *Hands-on*: build a collection of text to analyze

Module 3: Managing Textual Data

- Intro to data curation
- How to prepare a textual dataset for analysis
- *Hands-on*: run script to remove HTML tags, or clean spreadsheet in OpenRefine

Module 4: Performing Text Analysis

- Two lessons, one for beginners and one more advanced
- How to run a text analysis algorithm
- *Hands-on*: run algorithm in HTRC portal, or run a script to study HTRC Extracted Features files

Module 5: Visualizing Text Data with HT+BW

- How to use HathiTrust + Bookworm to visualize HTDL data and metadata
- *Hands-on*: track trends in how words were used over time

ASSESSMENT

- Administer assessment survey after each workshop
- IRB-approved with goal of sharing results in the future

Sample questions:

- What was the most interesting thing you learned?
- What could be improved or added?
- How effective was this workshop in teaching you about the HTRC?
- How do you see yourself applying what you learned?

