Wikimedia for Faculty, Students, and Us

Background

NB: "Wikimedia" will be used here as shorthand to refer to any of the projects administered by the Wikimedia Foundation. All are mainly edited by volunteers, despite their official oversight and server space. The three most common for Harvard Library purposes are Wikipedia (encyclopedic articles), Wikimedia Commons (high-quality reusable images), and Wikidata (structured Linked Open Data).

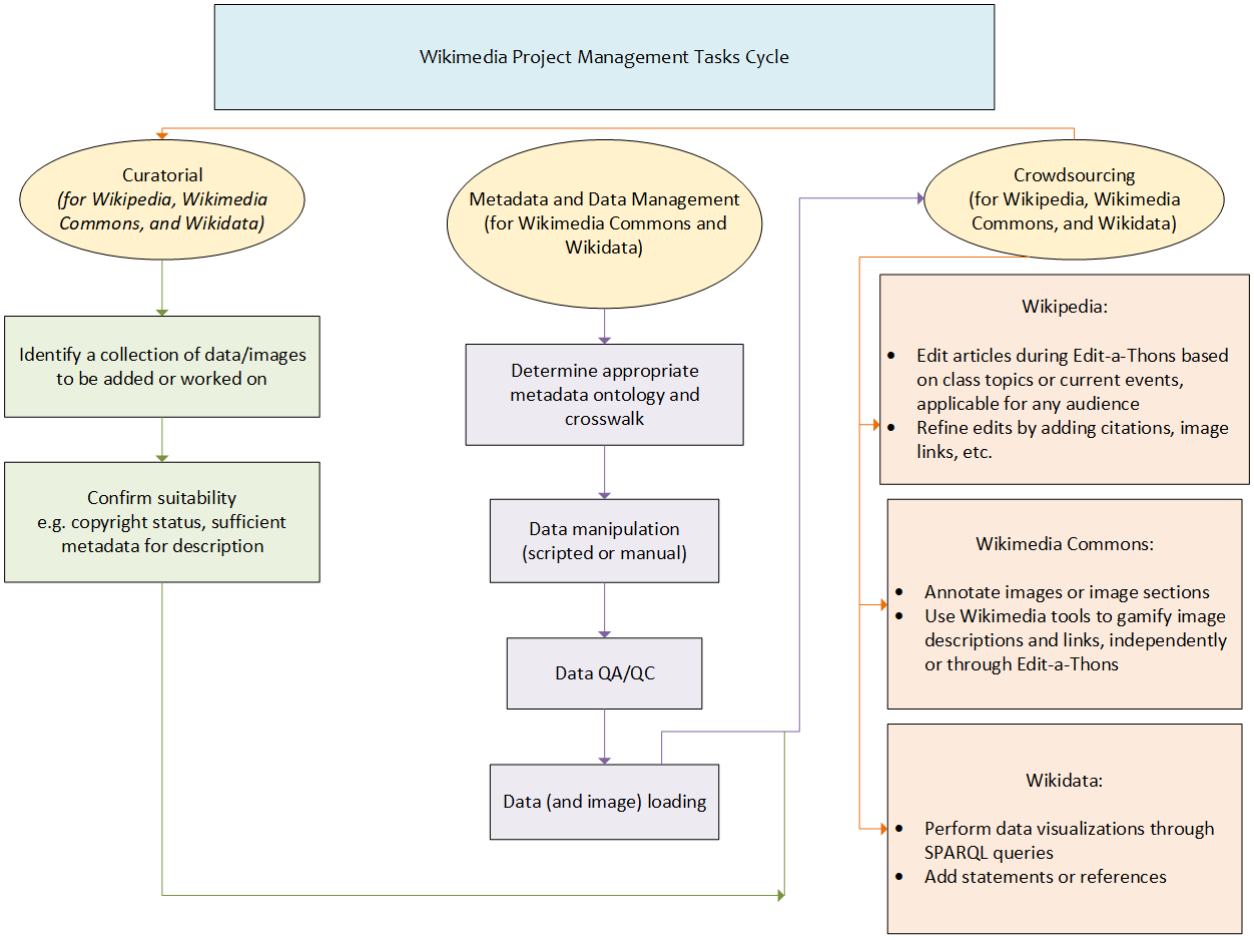
Wikimedia projects are appropriate for a wide variety of projects involving collaborative digital writing assignments, digital collections curation, data analysis, and any combination of these, as well as others not yet foreseen. As we enter a second phase of Harvard Library Remote Work for the 2020-21 academic year, Wikimedia will be a necessary open platform for students, faculty, and information professionals to use for online teaching and learning. Three use cases that have already been requested by students and faculty will be highlighted in this document, along with examples, workflows, and documentation.

1. Editing Wikipedia by adding information to articles missing information or citations, adding Harvard Library materials as external links for well-developed articles, creating new articles when appropriate, etc. — Case Study: Black Lives Matter Edit-a-thon
2. Importing Harvard Library Digital Collections objects into Wikimedia Commons for increased Web visibility and image annotation — Case Study: American Currency Collection
3. Performing data analysis and visualizations on Wikidata items
    a. Current Events — Case Study: COVID-19 topic graphs
    b. Diversity, Inclusion, Belonging, and Anti-Racism — Case Study: Representation of African Archaeology Sites across Wikimedia
    c. Optionally also adding Harvard Library records (even if not digitized) to Wikidata for analysis — Case Study: Cartographic Notebooks of Erwin Raisz
    d. Digital Humanities — Case Study: Map from the Survey of Scottish Witchcraft data
    e. Climate Change — Case Study: Multilingual Structured Climate Research Data in Wikidata

While these projects may look and sound very different from one another, they actually have more in common than not. When it comes to devising and managing projects, they have largely similar workflows and steps, as agreed and documented by the Wikimedia Remote Work Project Team.

The Wikimedia Project Management Tasks Cycle

The following diagram maps out the Project Team's conceptual and practical framework for completing projects through Wikimedia platforms, beginning with the curatorial decision to highlight a specific collection and ending with the opportunity for non-library professionals, including faculty and students, to get involved and use the project for teaching and learning. The work in the middle to make these projects possible comes from a wide variety of library staff.



We will now provide more detail for these tasks.

Curatorial

Scenario 1: Someone who has a topic focus that is not well-represented in Wikipedia identifies an editing project to fill in missing information.

Steps:

1. Determine notability standards are met for the missing information to be added.
2. Gather information and high-quality citations. See if the topic already exists in Wikidata or Wikimedia Commons, both of which can be used to add information to Wikipedia. For more instructions on Wikipedia editing, see the Project Team's documentation: Wikipedia:GLAM/HarvardLibrary/Articles-To-Work-On
3. If applicable, seek a Metadata and Data Management intervention. Otherwise, move to Crowdsourcing.

Scenario 2: A collection manager identifies a collection of interest for importing into Wikimedia Commons and/or Wikidata. This could be an image collection, but it could also be a non-image dataset.

Steps:

1. Create a preservation and data management plan.
    a. For items being digitized for the first time, this may include physical conservation, in which case preservation employees must be consulted.
    b. For existing image or data files, this will require determining how and where the files are stored, how and where they can be accessed, and how and where they can be copied or transferred for project production and subsequent safe storage.
    c. If it has not been done already, this also requires a full item-level inventory of the collection.
2. Identify existing metadata that describes the collection and add to it as needed.
    a. Whether MARC records, XML or JSON metadata, Access databases, or otherwise, all metadata must be consolidated and assessed.
    b. Basic access metadata (so that each file can be distinguished and identified) must be complete before any other work can be done.
    c. Sources of existing metadata and additional information should be identified as possible.
3. Determine eligibility of the collection.
    a. Be sure that all works to be loaded conform with the Wikidata notability or Wikimedia Commons project scope standards.

b. Be sure that there are no restrictions on loading, such as images that are still in copyright.
c. In the case of images, determine the appropriate rights statement and find an analogous Wikimedia Commons template, e.g. PD-old-70 or any of the PD-USGov license tags. Consult copyright advisers within the library as necessary.

Metadata and Data Management

Information professionals — catalogers, metadata librarians, other technical services staff, and other staff who are knowledgeable about open data standards — should be brought in to work on the data transfer process for loading collections into Wikidata and/or Wikimedia Commons. Depending on the expertise and resources of the collection manager, metadata and data staff may need to perform these tasks as well, at least in part.

Steps:

1. Information professionals should assess the existing metadata.
    a. Determine if the existing metadata is sufficient. If it isn't, they should determine what's missing and prepare workflows for how to add necessary information.
    b. Determine how granular the metadata and collection loading should be based on the collection. For example:
        i. One collection of data → one Wikidata entity at collection level, e.g. Gazetteer of Australia
        ii. One collection of data broken down into many items → one collection-level Wikidata entity plus a Wikidata entity for each individual item, e.g. Erwin Raisz Notebooks containing Notebook 9 [Erwin Raisz notebooks] and others
        iii. One collection of images → one image category with individual Wikimedia Commons images, e.g. American Currency
        iv. One collection of images → one image category that has a corresponding Wikidata entity and contains individual Wikimedia Commons images, e.g. Atlantic Neptune
        v. One collection of images → Wikimedia Commons images that may be part of one or multiple image categories, but each have their own corresponding Wikidata item, e.g. Harvard Art Museums works in Sum of All Paintings WikiProject

2. Information professionals should develop the appropriate metadata crosswalk for the collection to be loaded into Wikidata and/or Wikimedia Commons, as determined in step 1.
3. Information professionals should set up scripts and API calls for data gathering and manipulation.
4. Information professionals should help with data cleaning and remediation after conversion processes from step 3.
5. The files and metadata should now be ready to load.
6. Information professionals can help with determining what following collection enrichment needs to be done, and how much of it is appropriate for the Crowdsourcing phase.

For an excellent overview of how this process looks in practice, see the GitLab documentation for the Harvard Library American Currency Wikimedia Commons project.

Crowdsourcing

When the project planning and information gathering are done — meaning either that the Wikipedia editing assignment and materials are ready, or that the project files are loaded into Wikimedia Commons and/or Wikidata with viable metadata, as the case may be — the project can be enhanced with crowdsourced help. This help could come from library staff who are being redeployed through SMAT assignments, but it could also come from classes that are searching for collaborative projects over the course of a remote fall semester. Tasks include:

- Writing new Wikipedia articles
- Editing and updating existing Wikipedia articles
- Loading additional images to new images categories
- Creating image categories for images that should be united but haven't been
- Adding individual images or image categories to Wikipedia pages or Wikidata items
- Adding Wikidata items to image categories they describe
- Creating Wikidata-powered infoboxes on Wikipedia pages
- Performing data visualizations on Wikidata through SPARQL queries
- Adding reference statements to Wikidata items so that they can be better queried
- Annotating Wikimedia Commons images with structured data so that they can be searched

These and many more ideas can be discussed with the Project Team, who can assist with implementation as well.