

Future Work

This article introduces the framework of our digital humanities database of the *Basic Annals* of the *Shiji*. This open access online database consists of three parts: full-text, person, and location retrieval. In the future, we plan to refine the data with the help of historians and put it online for public use. Furthermore, we will add more functions for the word-based full-text and visualized retrieval. We hope this kind of database with rich information on persons and locations can be applied to more historical texts.

Acknowledgment. We are grateful for the comments and revision by the reviewers and editors. We also thank Liang Qiao and Ke Yang for their data annotation of the database. This work is supported in part by National Social Science Funds of China (No. 18BYY127), and China Scholarship Council.

Primary Sourcing: the Ten Thousand Rooms Platform for Digital Annotation of Primary Source Images

Nicholas Frisch*

East Asian Languages and Literatures, Yale University

*Corresponding author. Email: nicholas.frisch@yale.edu

doi:10.1017/jch.2020.25

Abstract

The Ten Thousand Rooms Project offers a modern digital solution to an age-old scholarly problem: a platform for annotating source texts for easy reference and retrieval, while facilitating easy access to original sources. Individuals or groups can upload scans of primary sources and annotate over the image, in multiple layers, toggling annotations on or off per research and reading needs. The project is hosted by Yale University.

Commentary and annotation anchor China's textual traditions. Since before Confucius, exegesis of canonical documents has been a bedrock of Sinitic literature: layering explanatory text over and alongside older texts, to make canons legible to contemporary readers. In both manuscript and print, scholars have struggled to efficiently represent accumulated layers of commentary and analysis. Marginalia, interlinear jottings, and annexed annotations crowd together on premodern pages.

In the twentieth century, modern typeset and punctuated editions have effaced much of this materiality of original primary sources. Researchers have lost a link to the original context of their primary sources: the configuration of texts and paratexts. Academic research is less likely to consider, and account for, the materiality of modes of scholarly exegesis that persisted until the last century.

The Ten Thousand Rooms Project (<https://tenthousandrooms.yale.edu/>) lets scholars collaboratively annotate images of premodern primary sources. Researchers upload images of their source—whether manuscript, print, bamboo, or bronze—and link

sections of the image to annotations which float in nearby windows. Multiple layers and versions annotation for a single section of a primary source can coexist: transcription, translation, commentary, and more.

How it Works

The basic unit of Ten Thousand Rooms is the image. Users create an account and a project, then upload and organize images of the primary source texts they wish to annotate. If you can scan or photograph it, you can upload it—from bamboo strips to bronzes to frescos. Images appear subdivided in sections according to user preference. A project page allows you to categorize and re-order images you've uploaded.

Inside the annotation viewer, users can view, navigate, and annotate their images. Users choose a customized section of an image which they wish to annotate and are offered customizable categories to classify the annotation: transcription, translation, commentary, or whatever else is useful. Annotations are typed and saved in a floating window. They appear as pale blue outlines layered over the image; click one, and the corresponding annotation will appear in a window alongside the image.

Technical Specifications

The Ten Thousand Rooms Project's core image annotation feature uses the IIIF standard (International Image Interoperability Framework) and the compatible Mirador image display and annotation platform. IIIF is the foundation of many university and museum efforts to make their holdings publicly accessible on a unified standard which allows for sharing and standardized metadata. Mirador, developed by Stanford University, is one of the major IIIF-compliant platforms used for displaying, manipulating, and annotating IIIF images. Ten Thousand Rooms uses a customized version of the IIIF standard's JSON metadata annotations to allow users to annotate images within the Mirador platforms, including multiple layers of linked annotation (e.g., "Translation," "Transcription," "Commentary") associated with a single area of the image.

Wrapped around the Mirador platform, we use a customized Drupal user interface where users can create, manage, and navigate projects. This interface includes an image uploader with versatile file re-organization and re-naming capabilities.

A Research Tool

For scholars tackling new texts or extending a prior project, Ten Thousand Rooms offers a digital workspace for organizing texts as they choose. Scholars can make key decisions about research design as they go along, adding, annotating, and organizing images as necessary. Annotation targets on the images can be as large or small as research requires; customized polygons are available. If a scholar wishes to draw attention to a precise area of an image—for instance, a single line of text—and layer transcription, transliteration, translation, commentary, and scholarly analysis all on the same polygon or pinpoint, she can do so. Within an image, annotation targets can be as sharply or loosely defined as the scholar wishes. Commentary layers can be customized; if, for example, a group of researchers wishes to keep their individual contributions distinct, individual custom categories can distinguish them. Images, annotations, and categories can be added as needed as a research project

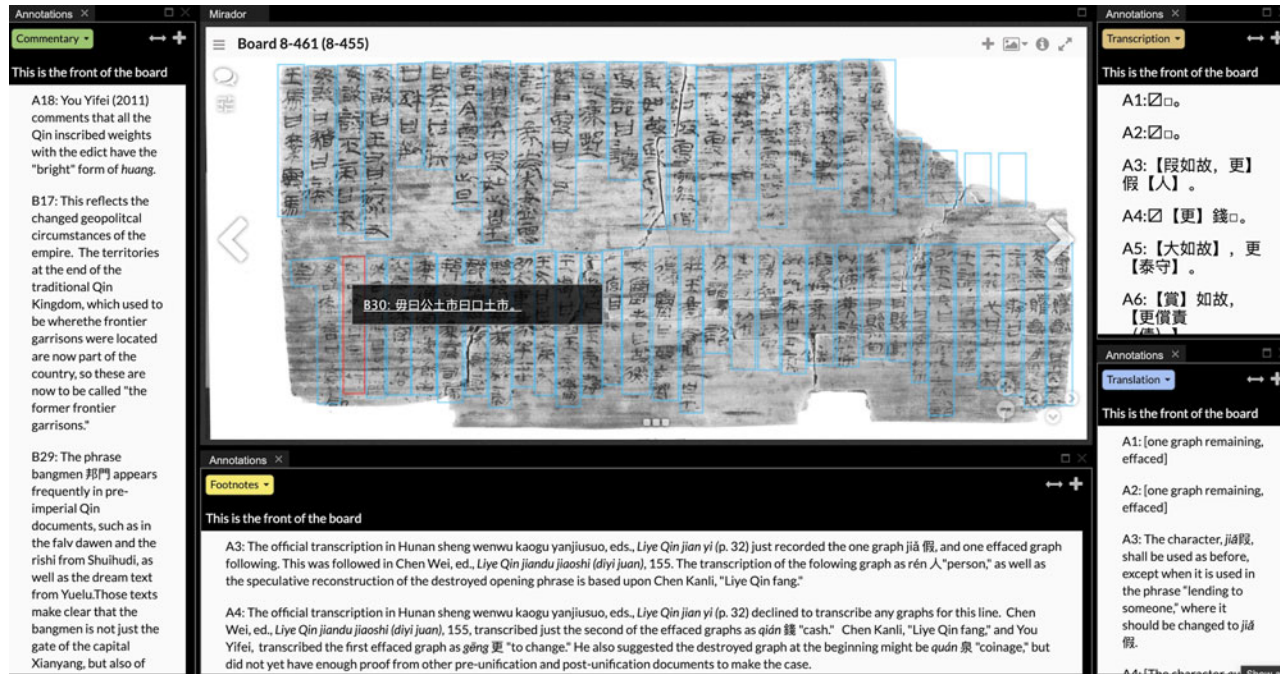
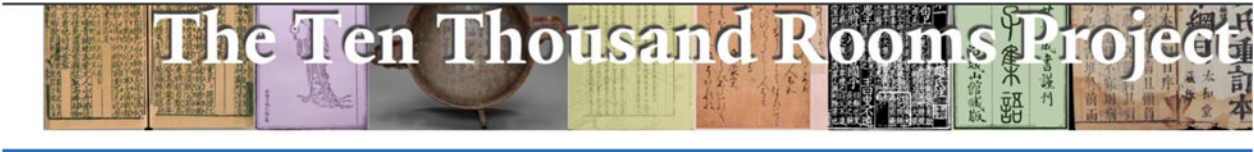


Figure 1. Annotations Within the Mirador Interface



[Home](#) [Explore the Site](#) [News](#) [Projects](#) [Site Forum](#) [Terms of Use](#) [Help](#)

HOME » [ROBERT ENO'S TRANSLATIONS OF THE ANALECTS AND MENCIOUS](#)

Robert Eno's Translations of the Analects and Mencius

[View](#) [Edit](#) [Group](#) [Members](#) [Revisions](#) [Forum](#) [Export](#)

 [The Analects of Confucius: An Online Teaching Translation \(2015\)](#)

 [Mencius: Translation, Notes, and Commentary](#)

[Subscribe to group](#)

Search Annotations

User Menu

[My account](#)

[My projects](#)

[Add project](#)

[My Image Collections](#)

[Manage users](#)

[Log out](#)

Figure 2. An Individual Project's Homepage

The screenshot displays a digital interface for 'The Analects of Confucius: An Online Teaching Translation (2015)'. On the left, a sidebar lists chapters from 00 to 09, with Chapter 05 selected. The main area shows a grid of Chinese characters in a traditional layout. A red box highlights the characters '毋我' (wú wǒ) in the second row, second column. Below this, a blue box highlights the characters '子畏於匡' (zǐ wèi yú kuāng) in the third row, second column. To the right of the grid, there are several annotation panels. The top panel, titled 'Textual Commentary', provides context for the highlighted characters, mentioning the Han history *Shiji* and the sage King Wen. The bottom panel, titled 'Translation', provides a modern English translation of the highlighted text. The interface also includes a 'Transcription' panel and a 'Textual Commentary' panel for the same page.

Figure 3. Digitally Enhanced Exegesis: Annotations of The Analects

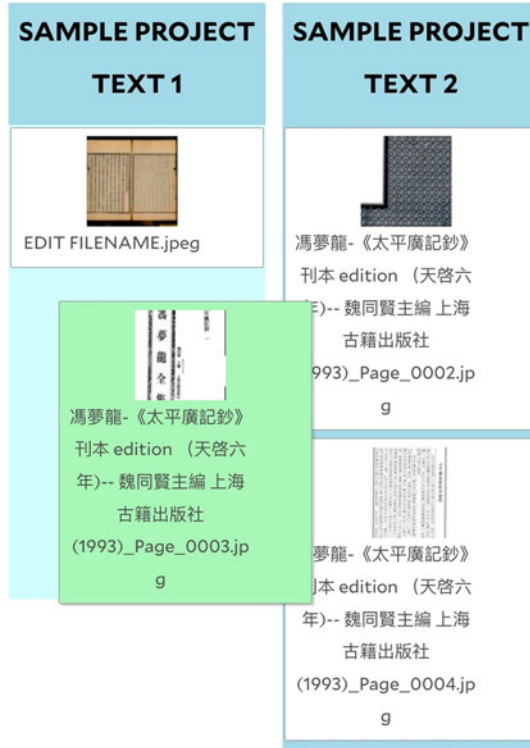


Figure 4. The Image Organization Interface, for Customizing the Sequence and Categorization of Bulk Image Uploads with Thumbnails and Filenames Visible

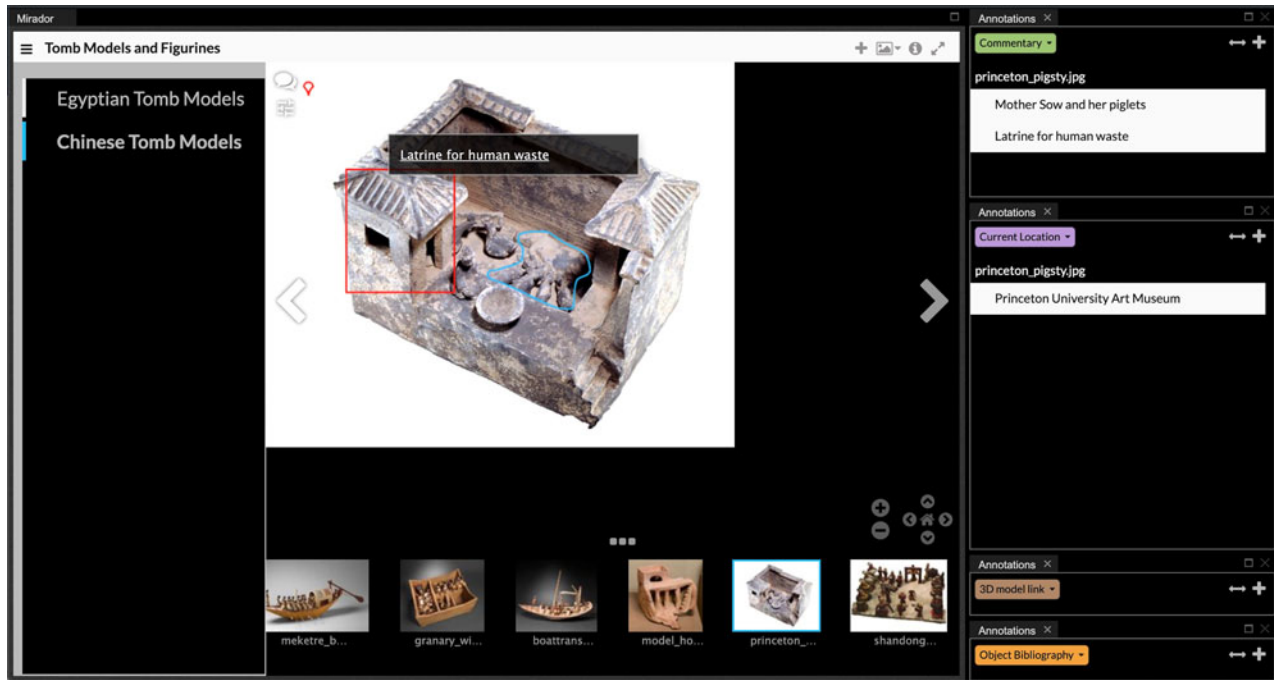


Figure 5. Beyond Text: Applications of the Ten Thousand Rooms Project for Art History and Archaeology

progresses, giving the scholar flexibility in developing the direction of a project. In effect, Ten Thousand Rooms provides the digital equivalent of unrestricted margin space on a paper document; infinite file folders and as many colored pens and highlighters as the scholar requires. Plain-text annotations can be exported for use in other settings.

A Teaching Tool

A variety of flexible user settings facilitate the use of Ten Thousand Rooms as a pedagogical tool. Creators of a project can designate collaborators and create primary-source-based assignments for students to experience and annotate texts directly. A seminar or symposium devoted to decoding a complex text can be conducted in real-time, with students around a table; or, it can unfold over time with consecutive contributions from designated users.

The annotation of images has applications beyond textual studies; students of art history, or scholars researching documents combining text and image, can combine their expertise with others in other disciplines. Students at undergraduate and graduate levels can work together, making respective contributions (e.g., direct transcription versus translation) and improving each other's work in the process.

A Publishing Tool

Ten Thousand Rooms affords scholars an opportunity to supplement, or even re-imagine, the final product of much scholarly research: the academic monograph. The platform helps researchers "show their work," and allows other to examine and improve it. Information that cannot fit easily within the main text or footnotes of an academic volume can be presented for the wider scholarly community.

Looking Ahead

After initial support from the Mellon Foundation, Ten Thousand Rooms has an ongoing commitment for support and development from Yale University. The platform hopes to introduce multi-target annotation (allowing a single annotation to correspond with multiple sections of image) soon. We intend to improve interoperability with other digital projects, introducing features including an API for both images and text. With images, an API would involve the ability to directly import and manipulate IIIF-standard images from other databases. For text, we hope to introduce a more sophisticated export of plain-text generated by projects, allowing different categories of annotation to be exported and treated separately and offering possibilities for text structuring and markup (e.g., with TEI Text Encoding Initiative Standards) and integration of geographic information with annotation text. The ultimate objective is to develop a tool that bridges the gap between images of original primary source documents and the wider world of sophisticated textual and geographical tools revolutionizing the study of Chinese history.