

Chinese Antiquities in U.S. Museums: Online Collections Databases and Open Data Quality

Ximeng Deng



Supplementary Materials
Poster, paper, dataset, and code:
<https://tinyurl.com/ChineseAntiquities>

Key Sources:

- [1] Meyer, K. E., & Brysac, S. B. (2015). *The China collectors: America's century-long hunt for Asian art treasures*. Palgrave Macmillan Trade.
[2] Rujivacharakul, V. (2011). *Collecting China: the world, China, and a history of collecting*. University of Delaware Press.
[3] St. Clair, M. (2016). *The great Chinese art transfer: how so much of China's art came to America*. Fairleigh Dickinson University Press.
[4] Steuber, J., & Lai, G. (2014). *Collectors, collections & collecting the arts of China: histories & challenges*. University Press of Florida.
[5] Tomkins, C. (1970). *Merchants and masterpieces: the story of the Metropolitan Museum of Art* ([1st ed.]). E.P. Dutton.



UNIVERSITY OF
MARYLAND
COLLEGE OF INFORMATION

Abstract

Research purpose. Bringing museum information resources online benefits the public and researchers and, in this case, provides a practical alternative to in-person study or repatriation.

Problem. Despite decades of digitization, open access, and collaborative data sharing initiatives, few empirical studies have evaluated information retrieval effectiveness and open data quality.

Method. Web-based content analysis and a case study to review over 180,000 Chinese antiquities in 61 U.S. museums.

Key findings: Intertwined curatorial and cataloging practices; Challenges around the structural-flexibility tradeoff; Limited inter-institutional data interoperability.

Contribution: Digital curation of non-Western cultural objects, digital scholarships, and provenance research.

Introduction

Background. Ancient Chinese artifacts came into institutions abroad through **export, acultural and religious transmission, unregulated trade, and looting**. American museums now hold a significant portion of Chinese antiquities abroad.

Why digital collections matter. Modern museums serve as information service organizations; digitizing descriptive data and bringing museum information resources online have dramatically facilitated education and research despite physical distance (Marty, 2017).

What this poster reviews. The current state of digital museum collections — particularly the textual data, largely guided by the Categories for the Description of Works of Art (Beca & Harping, 2024) and the Cataloging Cultural Objects (Beca & Visual Resources Association, 2006).

Evaluation Criteria. This poster evaluates their public accessibility, their value as primary sources for digital scholarship, and their potential for interoperability within the museum community.

Research Questions

RQ1. Which U.S. museums and which of their curatorial departments have notable collections of Chinese antiquities?

RQ2. What is the current state of digitization and open access for these collections?

RQ3. How do database information systems, cataloging metadata, and controlled vocabularies affect their accessibility and usability?

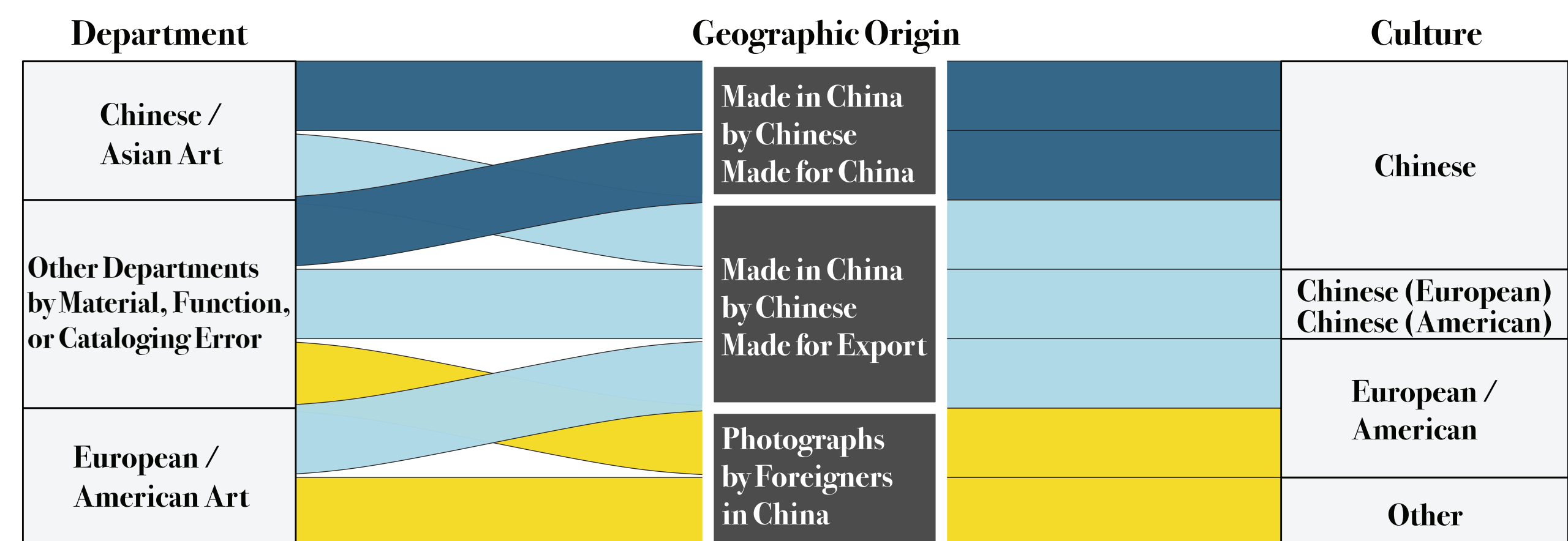
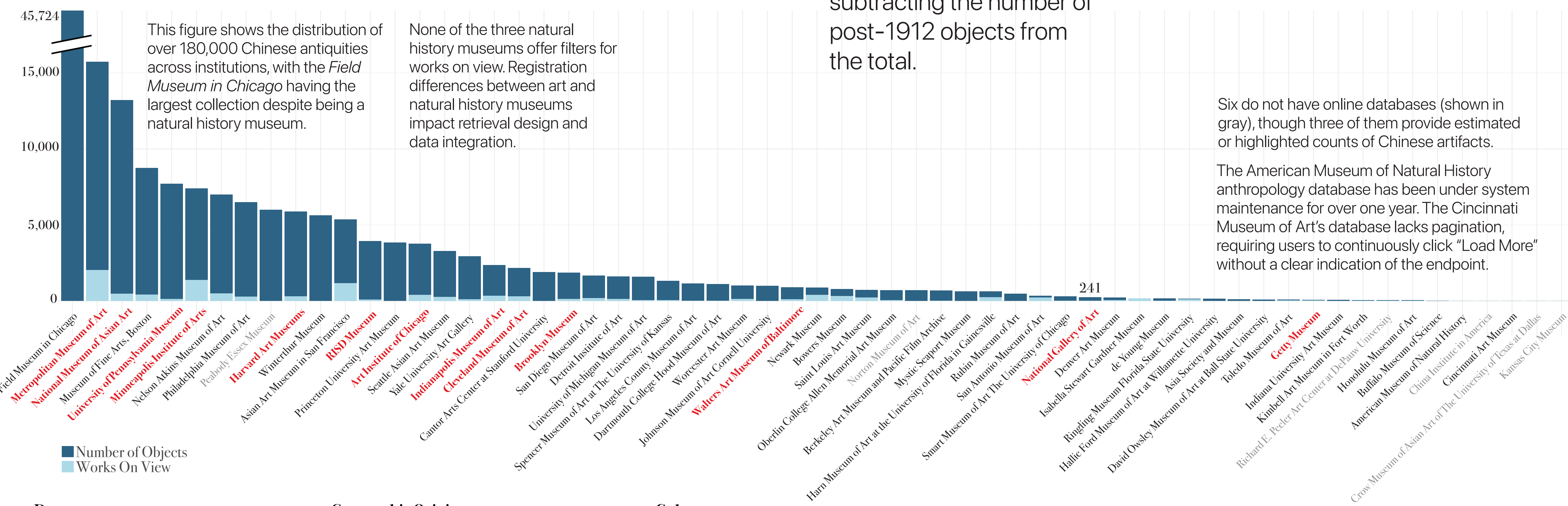
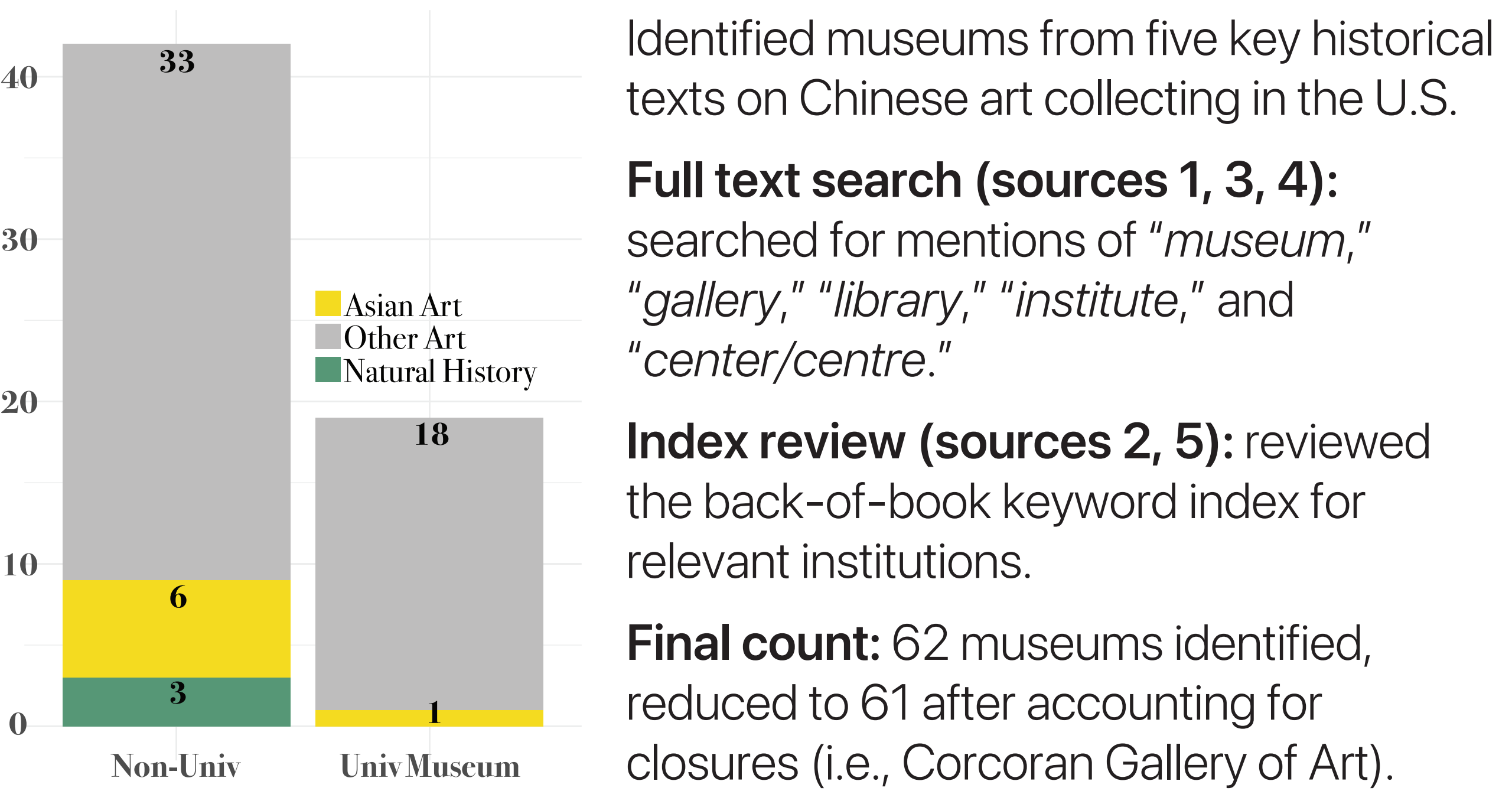
RQ4. To what extent is open data on Chinese antiquities in U.S. museums prepared to support digital scholarship, including quantitative analysis?

Scope of Identified Museums

Included. Early collecting museums acquired pieces through donations and purchases from collectors and dealers specializing in Chinese art.

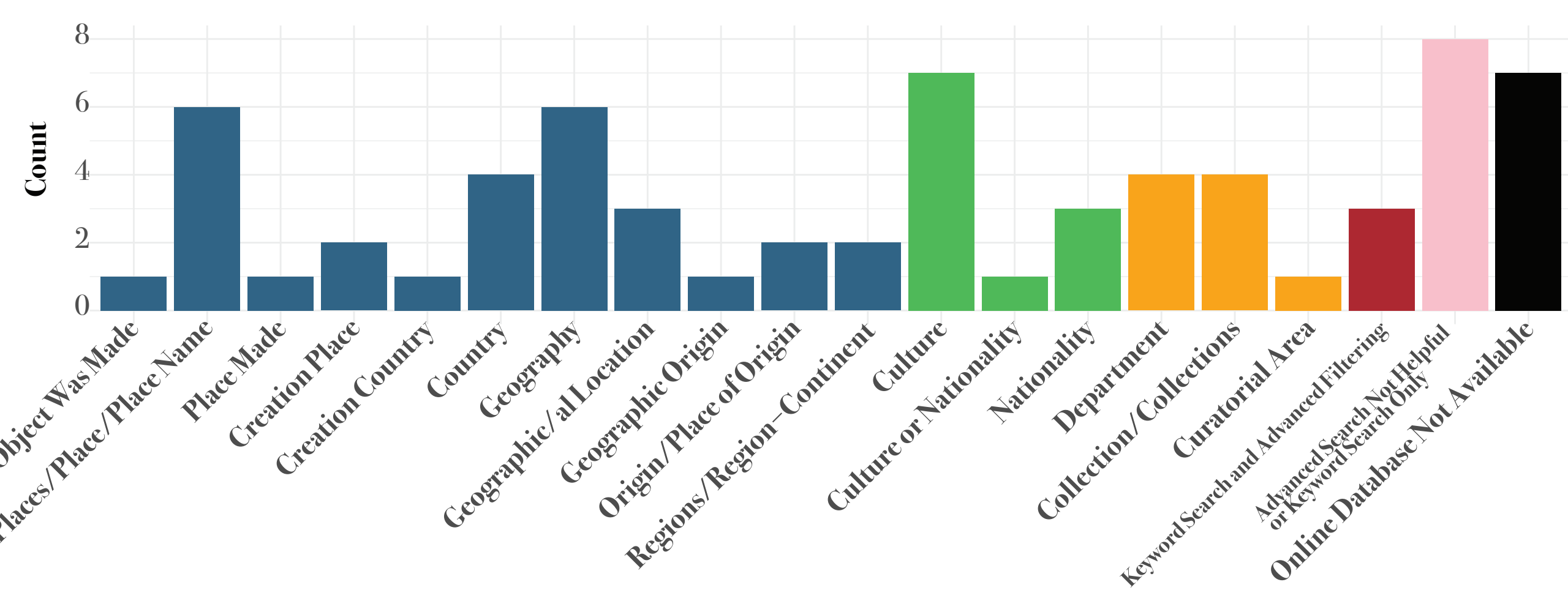
Excluded. Institutions primarily involved in more recent acquisitions without documented historical provenance, as this study does not focus on the contemporary American secondary market.

Museum Identification



This figure illustrates how Chinese antiquities are classified and cataloged across museum databases, showing the intertwined relationships between geographic origin, cultural attribution, and curatorial department.

Attribute (search strategy)	Fallback availability	Recall	Precision
Geographic Origin	N = 29	High	Good
Culture / Nationality	N = 11	Lower	High
Department / Curatorial Area	N = 9	Lower	Lower



Web-based Content Analysis

Methods. Searched museum online collection interfaces and documented results.

Systematically reviewed each museum's cataloging of ancient Chinese artifacts dated before 1912, the year the Republic of China was proclaimed, and the Qing Dynasty ended. The number was calculated as subtracting the number of post-1912 objects from the total.

Findings. Museum databases vary widely in how they support the search of Chinese antiquities. Institution-specific information organization is common, but this also raises critical questions about the definitions and curatorial rationales of concepts like “*Chinese Artifacts*” or “*Objects from China*.”

Web search ≠ easy search. The quality of search results depend on users' familiarity with database structures, data content standards, and retrieval tactics (e.g., in keyword-only systems, typing * or toggling “works on view” to return all objects).

Metadata & Vocabulary

Finding 1. Inconsistent cataloging specificity and the lack of linked hierarchy in vocabulary lead to search results that exclude objects described with more specific terms.

Attribute	Level of specificity	Limitation	Good practice example
Geographic Origin	China, Province, Archaeological site	In cases dropdowns allow only one selection, and free-text searches require guessing the “right” term, making it easy to miss relevant objects.	Harvard Art Museums — searching “Place: China” automatically includes subordinate regions.
Culture / Nationality	Chinese, Ethnic group, Dynastic and reign period		Cetty Museum — “Cultures” field supports multi-select from a controlled list.

Finding 2. Uncertainty (e.g., “*China or Japan*,” “*Korea? or China?*” “*Probably China*”) is stored as free text. **Flexibility** is necessary, but current database information systems lack a design model that supports both descriptive nuance and structured indexing.

Finding 3. Other metadata fields including **artwork type, styles, subjects, and classifications**, do not necessarily differentiate facets, maintain a hierarchy, or prevent overlap. Filters are not exclusive, nor are the values within each filter.

Finding 4. Objects with range-based **date displays** that span 1912 often result in exclusions from both pre- and post-1912 time-based searches.

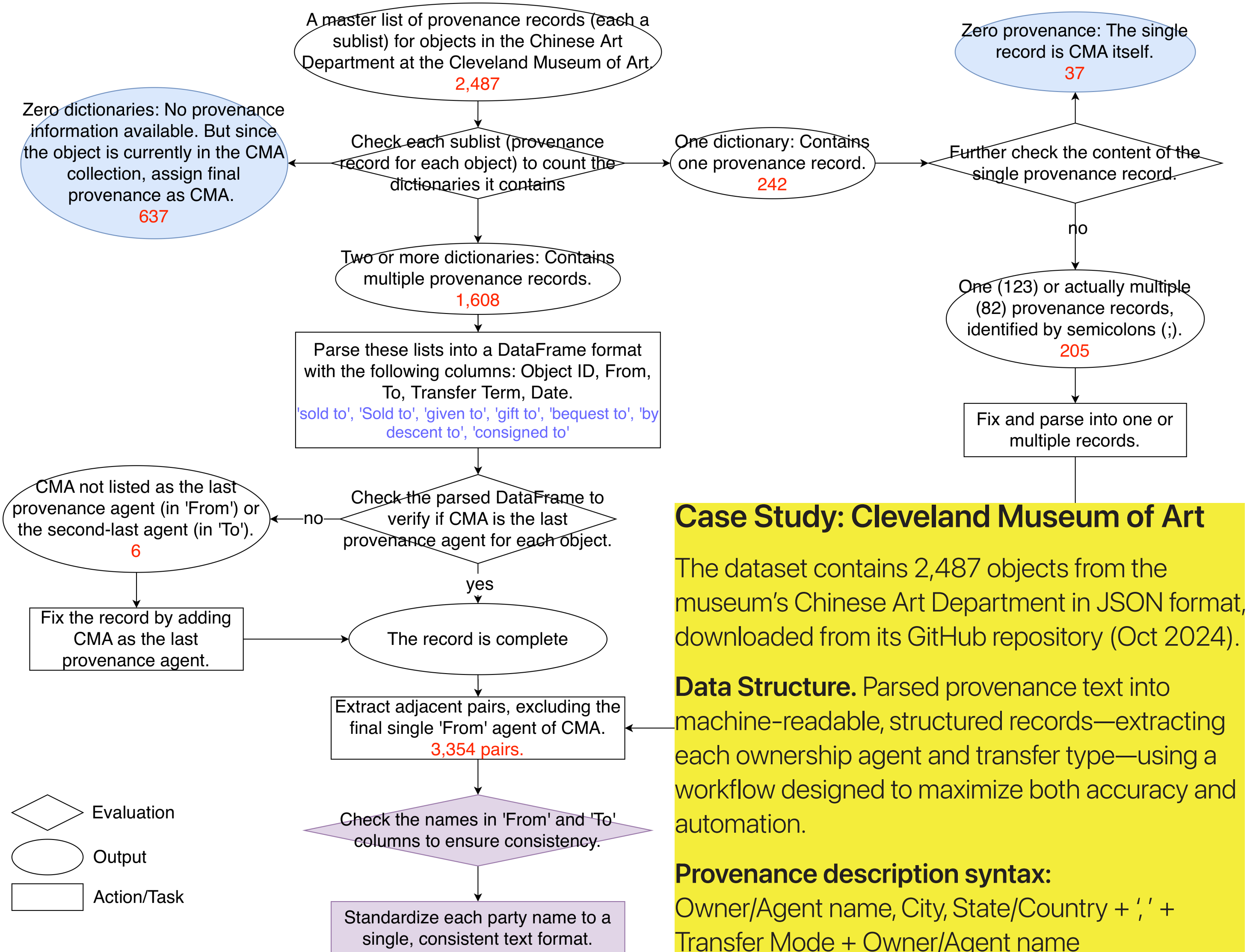
Open data is an important interim solution because it makes inconsistent vocabularies visible and allows external researchers to analyze, link, and correct them.

Open Data

Thirteen museums provide open access to their collections data with documentation websites (shown in red). **Eleven** host GitHub repositories, **nine** have released APIs, and **twelve** offer JSON or CSV files.

Although this approach meets only the third level of the five-star cumulative system standard for linked open data (W3C, 2013), it remains practical for bulk downloads, individual and offline analysis, and application development.

Data Structure & Value for Provenance

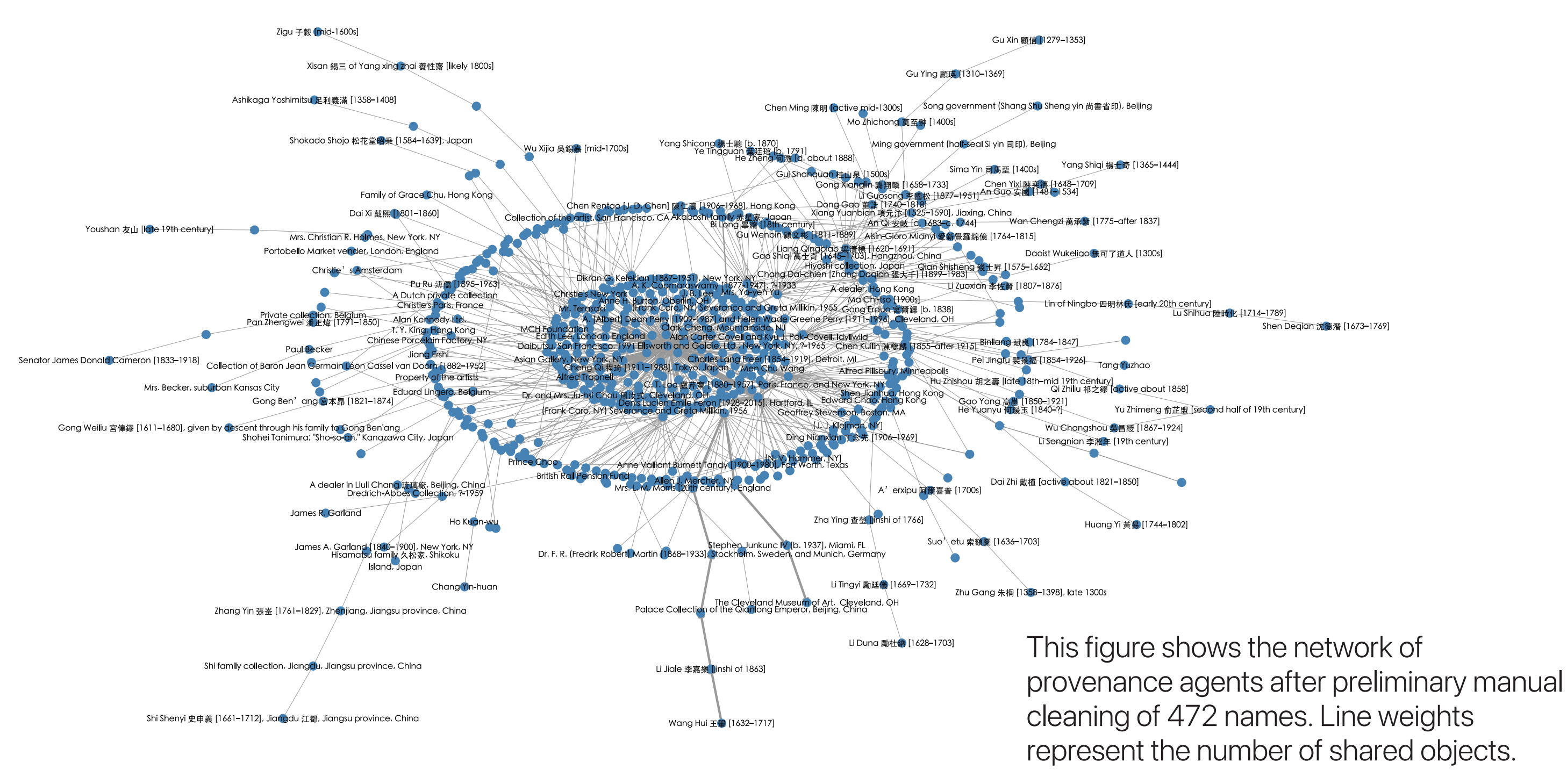


Case Study: Cleveland Museum of Art

The dataset contains 2,487 objects from the museum's Chinese Art Department in JSON format, downloaded from its GitHub repository (Oct 2024).

Data Structure. Parsed provenance text into machine-readable, structured records—extracting each ownership agent and transfer type—using a workflow designed to maximize both accuracy and automation.

Provenance description syntax: Owner/Agent name, City, State/Country + ‘,’ + Transfer Mode + Owner/Agent name



This figure shows the network of provenance agents after preliminary manual cleaning of 472 names. Line weights represent the number of shared objects.

Finding	Example	Limitation
Inferred Transfers	Between “Mr.” and “Mrs.,” parents and descendants, or individuals and their companies.	Disrupt the expected overlap continuity between consecutive records.
Ambiguous Transfer Terms	“Acquired in...”, “Purchased with funds from...”	Unclear ownership chain requires manual review.
Duplicate Names (1,311 listed, but fewer are truly unique)	Inconsistency in spacing, capitalization, punctuation (e.g., commas, periods, parentheses, brackets, en dashes); Inconsistency naming in middle names, full names, initials, women, couples, honorifics (e.g., “Dr.”), non-English names, and corporate names; Dates of ownership should appear only in the “date” field; Auction information such as lot numbers, should be recorded consistently, and placed in its designated field.	Inflates number of “unique” agents and creates noise.