

Project summary report: Promoting Reflexivity in Chinese Law and Policy: Feasibility of an Open-Access Research Portal

Authors: Teresa Chen, Edison Chen, Michael Cui, Chuhang Liu, Liane Ma, Lyla Miller, Danielle Park, and Alex Zhang

Affiliation: Bass Connections Building Open Access Portal for Chinese Law and Policy Group 2025-2026

I. Introduction & Background

The project aims to build an open-access portal of Chinese Law and Policy Research to promote 'reflexivity' for interdisciplinary scholars.

A. What does "reflexivity" mean in the context of interdisciplinary legal research?

Drawing on Bourdieu, we adapt the concept and framework of reflexivity to underscore the need for researchers to critically examine their positions, disciplinary training, and institutional affiliations. In Bourdieu's terms, "reflexive sociology" entails applying sociological tools to researchers themselves. Building on and expanding from this perspective, we propose designing a platform that makes research methods, theoretical choices, data sources, analytical processes, and limitations transparent, not only to authors but also to readers and researchers across disciplines. As Chinese studies is inherently an interdisciplinary/transdisciplinary field, we believe it is important to use this framework to help researchers reflect actively on the scope, limits, and justifications of their research and methods, and to help uncover "unthought categories of thought."

B. Why is an open-access portal, as opposed to subscription-based databases, the right format for this goal? How is it different from a full-text-based open-access portal?

We aim for an open-access portal that discloses each essential research component, rather than a traditional open-access portal that focuses on full-text dissemination. This allows us to comply with potential copyright and licensing restrictions while still effectively disseminating and promoting scholarships. For example, instead of sharing the full text, we categorize research methods, data sources, and limitations, and provide human-curated research summaries. This design helps us address several barriers identified in our legal feasibility study. Our goal is to provide free access to knowledge and information, but a subscription-based model would defeat that purpose.

C. Who is the primary intended audience: law students, social scientists, political scientists, or area studies scholars? What are their distinct research workflows?

Again, China studies are inherently diverse; our primary intended audiences are

researchers, policymakers, students, and the public interested in Chinese law and policy. By focusing on the fundamental research components, we provide accessible, structured information that can be adapted to researchers with distinct research workflows.

II. Problem Statement: Barriers to Access

A. Core Access Challenges

Access to Chinese law and policy resources remains fragmented and restricted for many reasons.

B. Key Barriers Identified

The poster/project identifies four major structural barriers:

1. Paywalled Databases

It is a reality that researchers worldwide face disparities in access to scholarships in China studies, including legal and policy scholarships. Unfortunately, this is not something we can fully overcome. However, our portal is designed to make the knowledge itself and the knowledge production process more visible and accessible in an informed way to researchers and other intended audiences across the globe. For information that is hidden behind the paywall, our project will provide meaningful leads by sharing the essential research components in compliance with copyright laws.

2. Incomplete and Fragmented Data

Our studies showed that there are a few issues with incomplete and fragmented data. One is that researchers do not always make their datasets and relevant data analytical records accessible. As a result, it is very difficult to know who created or annotated which datasets for what purpose and with what limitations, because this information is usually hidden in the articles themselves and is rarely indexed or captured in metadata regimes.

In this study, we examined comparable portals, such as the ICSPR (Institute for Social Science Research at the University of Michigan) data portal, which provides a detailed metadata scheme to describe each dataset. We hope to design a metadata regime similar to ICSPR to make datasets hidden in the scholarship more accessible and, hopefully, reusable by others. We plan to apply similar schemes to organize, capture, and disseminate information on primary sources of law, such as laws, regulations,

judicial interpretations, and other guidance materials analyzed in the scholarship. Through tagging and metadata, we hope to design a portal that allows users to search for these underlying research components, in addition to searching by author, article title, and other criteria.

Another issue related to incomplete and fragmented data is the restricted sharing of even primary legal sources in China. For example, stakeholders have expressed concerns about the incompleteness of China's court judgments online. Another related issue that has not been discussed enough in the literature is that Chinese laws are revised/repealed/enacted very frequently, and as a result, studies examining Chinese Securities law in 2006 may yield very different findings from studies discussing Chinese Securities law in 2020. Although we cannot fully fix the restricted data issue, we believe our portal will enable more robust cross-disciplinary and longitudinal analysis of China's primary legal sources, including cases, statutes, regulations, etc., by systematically indexing and tagging primary sources referenced in law and policy scholarship. It will also incorporate metadata on research methodologies, limitations, and scope, thereby enhancing transparency and facilitating more informed comparative research.

3. The "Leapfrogging" Problem

While conducting feasibility studies, we identified several methodological deficiencies. For example, the rapid growth of empirical and computational legal studies in China has been characterized by what Tang and Liu (2020) call a "leapfrogging" phenomenon: the field has jumped directly from doctrinal scholarship to data-driven research without developing the necessary intermediate methodological infrastructure. Such infrastructure is vital for providing context on research questions, methodologies, and data collection processes, helping other scholars understand how results are generated. For example, when computational techniques are used, details about the algorithms and statistical models should be clearly documented as part of the research output.

Additionally, human judgment remains crucial even in computational legal and humanities research; therefore, the interpretive frameworks and theoretical foundations guiding the analysis must also be explicitly stated. These requirements highlight the importance of intentionally and systematically documenting the research design, methodological choices, and analytical processes. This effort will be especially important in the age of generative AI, as the methodological opacity and the black-box problem are likely to intensify due to the nature of generative AI tools when applied to scholarly research. For example, as cross-border field research opportunities in China diminish, researchers may rely on online studies, such as distributing surveys and collecting data virtually, which are vulnerable to AI-generated responses or the creation of synthetic data. Therefore, careful review of research methods and data analysis is essential for filtering, validating, and sharing trustworthy scholarly knowledge.

4. Geopolitical and Institutional Constraints

Access is limited not only by market forces but also by deliberate political controls on information flows. Our research identified several geopolitical and institutional constraints, including financial constraints in higher education more broadly, federal funding cuts, and increasingly stringent immigration and border-control laws and regulations. Truex (2024) described how scholars studying China find themselves “caught in the middle,” viewed as potentially suspicious by both the Chinese and American governments. Others have found that all social science researchers interviewed had shifted their research trajectories in response to the geopolitical situation and concerns about the securitization of academic exchanges (Baum and Ma, 2023). The loss of funding and financial cuts will also directly affect researchers' ability to access high-quality materials, including scholarships. While our project alone cannot overcome these barriers, our portal, if established, will help bridge some gaps.

III. Research Method

A. Mixed-Methods Approach

The project employs a mixed-methods design integrating legal analysis, platform benchmarking, and AI evaluation.

B. Phase 1: Scoping & Framework Development

1. Define portal scope and identify key theoretical frameworks, research methods, and data sources. Map primary legal sources and datasets to assess their accessibility and reliability.

More specifically, we restrict the project's initial scope to Chinese law and policy scholarship. We retrieved the first batch of scholarship articles (about 140) using Keywords, subject areas, and authors to conduct the feasibility study. We intended the batch to include a variety of research methods, including qualitative, quantitative, doctrinal, and historical, and to cover many subject areas, including law, policy, economics, and sociology, because we believe China studies is inherently interdisciplinary and transdisciplinary. For each article, we coded it across 8 major areas (basic article metadata, accessibility, content summary, research questions and findings, research methods, theoretical framework, legal sources and policy documents, and datasets). These coding and variables prepare us for the portal building and feature testing underway.

C. Phase 2: Legal Feasibility Analysis: focusing on US copyright and AI law

We focus on US copyright law analysis, more specifically, the four-factor fair use analysis, licensing regimes, and open access framework as applied to scholarship on China. We studied the US copyright law statutory section (17 USC 107), reviewed secondary sources from copyright offices and other scholars, researched US case law interpreting the statutory section, and discussed the contract override principle with US copyright specialists. We reviewed the types of open-access licenses (via Creative Commons) and the implications of commercial licenses. We also reviewed the journal websites and identified each journal's open-access policy.

D. Phase 3: Platform Benchmarking: Benchmark existing academic portals to evaluate their architecture and metadata standards.

Before building the prototypes, we reviewed comparable platforms, such as library catalog systems, the Civil Rights Clearinghouse Project, the Harvard China Online Project, and ICPSR's data portal, to identify features that could improve our platform's usability. We then developed two prototypes to test these features, with particular attention to preservation, transferability, and a user-friendly UX.

E. Phase 4: AI Evaluation: Test AI-generated summaries and assess their accuracy and interdisciplinary value.

To scale the project with realistic estimates of financial and human resources, we explored incorporating AI into our workflow for both content creation and portal design. For example, we started using Duke ChatGPT to generate AI summaries with predefined prompts and are currently validating them. We also explored website-building tools with Vibe coding features to mock up website designs.

IV. Legal Feasibility Findings

A. Three-Tier Licensing Framework

The team developed a tiered classification of content based on licensing restrictions:

Content Access Tiers

Tier	Category	Description
Restricted	Commercial databases	License terms contractually override fair use; content cannot be redistributed.
Conditional	Academic publishers (CUP, OUP)	Rights reserved, but fair use is not fully abrogated; case-

		by-case analysis is required.
Open	Creative Commons-licensed content	Full-text reposting is permitted with attribution.

B. Key Takeaways from US Case Law

We focus on U.S. copyright law analysis, specifically the four-factor fair use analysis, licensing regimes, and the open access framework as applied to scholarship on China. **More specifically, we try to answer the question: are we going to run into any legal barriers if we make publicly available basic article metadata, human-curated research summaries, information about specific research method components, and human-curated data descriptions?**

Overall, our assessment is positive: our open-access portal's aggregation of abstracts, summaries, citations, and dataset descriptions is likely protected under the U.S. fair use doctrine because it serves a nonprofit, educational, and highly transformative purpose by helping users discover and understand scholarship without replacing the original works.

Under the current US copyright law, providing this type of information is generally permissible; we are not reproducing full-text articles; instead, we are transforming them into structured metadata, summaries, and methodological annotations. Our project has a nonprofit, non-commercial mission, and we primarily work with scholarly and legal materials. The findings also remind us that when we curate our own summaries and data descriptions, we should deliberately add new meaning and utility, making them transformative. Market analysis is essential to our project: it is important to note that we do not intend to replace the original articles; instead, we aim to complement and enhance researchers' ability to appreciate the original article, and, in some way, create a distinct market through research method tagging and by providing structured research summaries and data descriptions.

Because we also aim to incorporate AI tools into our workflow, we examined the current US copyright and AI framework and sought to identify the key legal considerations for using AI tools to analyze variables in the scholarship, create research summaries, and, in the future, potentially use copyrightable materials to fine-tune an AI product, such as an AI chatbot. This is a very murky legal area right now. We researched this area via commercial legal databases, the US Copyright Office's case law database, and GW's Copyright case databases; however, there is no definitive answer to any of these questions yet. The current status is that the US appears to continue to use the traditional four-factor analysis to weigh these types of new legal cases. The law does not distinguish significantly based on authorship, so the analysis still turns on transformative use, market impact, commercial use, etc.

On the other hand, purely AI-generated content is unlikely to be copyrightable. The

good news is that we are not planning to distribute any such content. Using proprietary materials to train AI tools can be highly problematic, which guides us in designing a workflow to curate research summaries of articles behind paywalls.

Here is a selected list of cases we consulted in drafting the memo on this topic:

American Society for Testing & Materials v. Public Resources Org (2018), where the Court of Appeals in the D.C. Circuit upheld the educational purpose of collecting and aggregating research to make it more broadly accessible as a paradigmatic fair use purpose. A more recent case with similar facts was decided in April 2026, in which the U.S. Court of Appeals for the Third Circuit ruled in favor of UpCodes, which hosted a database of building codes and standards, holding that UpCodes' use is transformative because it "achieves a distinct objective of making the law freely accessible and educating the public on the content of binding laws" (Am. Soc'y for Testing & Materials v. UpCodes, Inc., No. 24-2965, slip op. at ____ (3d Cir. Apr. 7, 2026)). These two cases will be in our favor on both educational and transformative use grounds.

In New York Times Co. v. Roxbury Data Interface, Inc., 434 F. Supp. 217 (D.N.J. 1977), the court held that the defendants' use of data from the plaintiffs' index to create their own index was protected by the "fair use doctrine". In that case, the court found that the defendant did not copy the index in full; instead, it extracted names and compiled its own index. Furthermore, the defendant's name index may have only a slight impact on the market. This case supports our approach. We do not reproduce the full text of others' works; instead, we extract selected information and generate original research summaries. As in Roxbury Data, our use is transformative and limited in scope, and it is unlikely to substitute for or significantly affect the market for the underlying materials. Similarly, a US Supreme Court case from 1991 would also help us justify our project by holding that citations and dataset information are uncopyrightable, see Feist Publications v. Rural Telephone Service, 499 US 340 (1991).

The examination of evolving US copyright law also reminded us to carefully consider how to design our portal, including whether to provide abstracts and how to use AI tools to enhance it. For example, in Harper & Row v. Nation Enterprises, 471 US 539 (1985), the Court, when weighing the factor on the amount and substance of the portion used, focused on the qualitative amount taken from the original work and found the excerpted portion was the "heart of the work." This analysis is important because it shows that it may be risky to decide where to place each article's abstract on our open-access portal. In ProCD v. Zeidenberg, 86 F. 3d 1447 (7th Cir. 1996), the Court held that contracts can override fair use protection. This case leads us to incorporate the review of the current journal policies on their respective websites into our assessment workflow.

We also examined the potential legal barriers to using AI tools to generate research summaries from open-access and CC-licensed works, as well as from proprietary

works. We found that the courts are currently split on this issue. For example, in *Kadrey v. Meta Platforms, Inc.*, the court held that Meta's copying and use of the entire book to train its large language models constituted a fair, transformative use because (1) the resulting tool could not be used to access any meaningful amount of the original work and (2) the copying was "reasonable necessary" for the transformative purpose of teaching and training AI models.

A similar case, *Bartz v. Anthropic PBC*, illustrates the nuance of training large language AI models on copyrighted material. The ruling was mixed: the court found that training LLMs on copyrighted works is legal under the fair use doctrine, but that AI platforms cannot use pirated materials in their training. Using legally purchased digital books to train LLMs does not infringe copyright, as it constitutes a "format change" and was not done with the intention of infringing the owners' copyright rights, thus being considered "transformative." With respect to our portal, this reasoning supports the view that using legally obtained articles to generate non-substitutive summaries would similarly qualify as transformative use.

Thomson Reuters v. ROSS Intelligence, on the other hand, underscores the risks associated with using AI. The good news is that our portal is quite distinct from Ross in terms of function, purpose, and operation.

As noted, we will need to continue to monitor the development of case law in this fast-evolving area.

V. Platform Design & Mock-Ups

A. Design Philosophy

The portal mock-ups were developed in ****Figma**** (Mock-Up 1) and ****Vercel**** (Mock-Up 2), reflecting different design philosophies and user experience priorities.

The WordPress portal mock-up in Figma is designed as a user-centered, research-oriented platform that prioritizes clarity, usability, and efficient information retrieval for legal scholars. Its structure emphasizes a clean, intuitive interface with a prominent search function, simplified navigation hierarchies, and clearly defined content categories to support goal-driven research and reduce cognitive load. The Portal also prioritizes portability and long-term sustainability. From a technical perspective, the design follows the Harvard Digital China Web tech stack to ensure stability and transparency. The system integrates a modular, self-hosted architecture that can support AI functions.

The CLPP Vercel mock-up showcases a minimalist, high-performance digital database that emphasizes precise search, easy access to resources, and a user-friendly design. The portal is also sustainable and transferable, using cloud storage and open-source tools to enable long-term flexibility and low-cost maintenance.

B. Core Portal Features (intended)

1. Cross-source search with unified metadata
2. Abstract and summary displays (fair-use-compliant)
3. AI-generated interdisciplinary summaries
4. Licensing status indicators by document
5. Researcher workflow integrations (to be investigated further in the future)

VI. Significance & Broader Implications

We hope to promote more participation in Chinese law and policy research through our portal. The democratization of access to knowledge and the production process would help ensure equitable decision-making for researchers and policymakers, supporting a more informed decision-making process.

- A. For AI and law: The project's AI evaluation findings have broader implications for AI-assisted legal research. What guardrails are needed for AI tools working with foreign (Chinese) legal systems?

Our current findings also highlight a few new areas for further research. For example, how to incorporate AI ethically and responsibly into the legal and policy research workflow, balancing the needs for accuracy, reliability, and efficiency. Furthermore, our findings show that we need to expand our research to other jurisdictions where scholarship in Chinese law and policy is widely published, such as the UK, Canada, the EU, and China.

- B. For open-access movements: How does this project contribute to ongoing debates about the open-access mandate in legal education and legal scholarship?

We believe this project, focusing on reflexivity, represents a new era of the open access movement, beyond passive full-text distribution.