2023

Legal Literacies for Text Data Mining – Cross-Border ("LLTDM-X"): White Paper

Rachael G. Samberg

Timothy Vollmer

Thomas Padilla

Legal Literacies for Text Data Mining – Cross-Border ("LLTDM-X"): White Paper

Summary

Legal Literacies for Text Data Mining - Cross-Border ("LLTDM-X") is a Level 1 Advancement Grant project addressing legal and ethical issues faced by U.S. digital humanities (DH) practitioners whose text data mining (TDM) research and practice intersects with foreign-held or - licensed content, or involves international cooperations. LLTDM-X is a collaboration between the University of California Berkeley Library and Internet Archive, and builds upon the previous NEH-sponsored institute, Building Legal Literacies for Text Data Mining (Building LLTDM). That institute provided guidance and strategies to DH TDM researchers on navigating legal literacies for text data mining (including copyright, contracts, privacy, and ethics) within a U.S. context.

A common challenge highlighted during Building LLTDM was the fact that TDM practitioners encounter numerous and complex legal problems in cross-border TDM research. These occur when: (i) the materials practitioners want to mine are housed in a foreign jurisdiction, or are otherwise subject to foreign database licensing or laws; (ii) the human subjects they are studying or who created the underlying content reside in another country; or, (iii) the colleagues with whom they are collaborating reside abroad, yielding uncertainty about which country's laws, agreements, and policies apply.

We designed LLTDM-X to identify and better understand the cross-border issues that DH TDM practitioners face, with the aim of using these issues to inform prospective research and education. We also hoped that LLTDM-X would yield preliminary guidance to benefit researchers in the meantime, as instructional materials are being developed. In early 2023, we hosted a series of three online round tables with U.S.-based cross-border TDM practitioners ("Practitioners"), and law and ethics experts ("Experts") practicing in six countries. The round table conversations were structured to illustrate the empirical issues that researchers face, and also for the Practitioners to benefit from guidance on legal and ethical challenges. Upon the completion of the round tables, the LLTDM-X project team created a robust and hypothetical case study that (i) reflects the observed cross-border LLTDM issues and (ii) contains analysis to facilitate the development of future instructional materials.

As more fully described below in the Takeaways & Recommendations section of this white paper, LLTDM-X surfaced seven key themes:

1. Uncertainty about cross-border LLTDM issues indeed hinders U.S. TDM researchers, confirming the need for education about cross-border legal issues;
2. The expansion of education regarding U.S. LLTDM literacies remains essential, and should continue in parallel to cross-border education;
3. Disparities in national copyright, contracts, and privacy laws may incentivize TDM researcher "forum shopping" and exacerbate research bias;
4. License agreements (and the concept of "contractual override") often dominate the overall analysis of cross-border TDM permissibility;
5. Emerging lawsuits about and regulatory assessment of artificial intelligence may impact future understanding of fair use and other research exceptions;

6. Research is needed into issues of foreign jurisdiction, likelihood of lawsuits in foreign countries, and likelihood of enforcement of foreign judgments in the U.S. However, the overall "risk" of proceeding with cross-border TDM research may remain difficult to quantify; and
7. Institutional review boards (IRBs) have an opportunity to explore a new role or build partnerships to support researchers engaged in cross-border TDM.

Through blog posts, social media engagement, and presentations, we will broadly share this white paper and our case study to begin helping U.S.-based TDM researchers navigate cross-border LLTDM hurdles. And, we will continue to encourage the integration of LLTDM literacies into U.S. disciplinary curricula and library and archive professional development, to facilitate both domestic and cross-border DH TDM research.

Project Origins and Goals

Growth of TDM in Digital Humanities

Digital Humanities practitioners increasingly rely on automated techniques and algorithms to extract revelatory information from large sets of unstructured or thinly-structured digital content—a process known as text and data mining (TDM).[1] TDM allows researchers to identify and analyze patterns, trends, and relationships across volumes of data that would otherwise be impossible to sift through, enabling exploration of issues like: racial disparity evidenced through police body camera footage;[2] changes in gender significance in fiction;[3] and public discussions of social justice issues like violence against women.[4] TDM methodologies and tools continue to expand, enabling advancements across education, literature, society, politics, and beyond.[5]

Training for U.S. Law and Policy Hurdles

While TDM methodologies offer great potential for advancing research, they also present research practitioners with nettlesome law and policy challenges. Consider the example of a researcher mining and analyzing harassing speech within social media posts,[6] and then seeking

---

[1] Hearst, Marti A. (2003, October 17). What is text mining? http://people.ischool.berkeley.edu/~hearst/text-mining.html

[2] Voigt, R., Camp, N. P., Prabhakaran, V., Hamilton, W. L., Hetey, R. C., Griffiths, C. M., Jurgens, D., Jurafsky, D., and Eberhardt, J. L. (2017). Language from police body camera footage shows racial disparities in officer respect. *Proceedings of the National Academy of Sciences*, 114(25), 6521. https://doi.org/10.1073/pnas.1702413114

[3] Underwood, T., Bamman, D., & Lee, S. (2018). The transformation of gender in English-language fiction. *Cultural Analytics*. https://doi.org/10.22148/16.019

[4] Xue, J., Macropol, K., Jia, Y., Zhu, T., and Gelles, R. J. (2019). Harnessing big data for social justice: An exploration of violence against women-related conversations on Twitter. *Human Behavior and Emerging Technologies*, 1(3), 269–279. https://doi.org/10.1002/hbe2.160

[5] Hassani, H., Beneki, C., Unger, S., Mazinani, M. T., Yeganegi, M. R. (2020). Text Mining in Big Data Analytics. *Big Data and Cognitive Computing*, 4(1). https://doi.org/10.3390/bdcc4010001

[6] Suomela, T., Chee, F., Berendt, B., & Rockwell, G. (2019). Applying an ethics of cre to internet research: Gamergate and digital humanities. *Digital Studies/le Champ Numérique*, 9(1), 1-28. http://doi.org/10.16995/dscn.302

to share these datasets to encourage research reproducibility. This scholar would need to address matters of: (i) *copyright* (e.g. Are the posts protected by copyright? Does an exception like fair use enable TDM regardless?); (ii) *contracts* (e.g. Do social media websites impose terms of use? Do such website agreements override copyright exceptions?); (iii) *privacy* (e.g. Do the posts reveal information that infringes upon federal and state privacy rights of the persons described in the posts? Is republishing data a further privacy violation?), and (iv) *ethics* (e.g. Could downloading and recirculating the content exacerbate harm to the subjects of the posts?). The copyright, contracts, privacy, and ethical issues that TDM practitioners must navigate can be considered "legal literacies for text data mining," or "LLTDM."

For years, there was a need in digital humanities curricula and professional development programming for guidance and strategies on navigating LLTDM in DH research.[7] In 2019, UC Berkeley Library received NEH funding through the Institutes for Advanced Topics in the Digital Humanities program for Building LLTDM,[8] an institute designed to address this knowledge gap. We hosted Building LLTDM virtually from June 23-26, 2020, and subsequently published a white paper[9] and an open educational resource[10] to extend the institute's reach and impact. As reflected in the powerful and positive reviews of the institute, Building LLTDM demonstrated the effectiveness of training U.S. digital humanities researchers in navigating law, policy, ethics, and risk within their TDM projects.

Similar Need for Cross-Border Guidance

In both the institute's instructional sessions and post-institute evaluations, participants identified cross-border DH research collaborations as an ongoing LLTDM problem, noting that foreign law and ethics issues pervaded their research. As Building LLTDM focused on U.S. law, and only lightly touched on cross-border issues within the context of a single legal literacy (copyright), it became apparent that the U.S. DH TDM research and practitioner community lacks guidance on how to navigate these cross-border concerns—thus paving the way for LLTDM-X.

Indeed, U.S.-based DH scholars do not conduct TDM research only in or about the United States. Digital humanities research is marked by collaborativeness across institutions and

---

[7] Certainly, the LLTDM issues arise in disciplines beyond the penumbra of digital humanities, but for purposes of this grant, we focused our research and guidance on supporting DH practitioners given the pressing need evidenced both in the literature and our professional experiences.

[8] Samberg, R. G. (2019, August 14). Team Awarded Grant to Help Digital Humanities Scholars Navigate Legal Issues of Text Data Mining. *Berkeley Library Update*. https://update.lib.berkeley.edu/2019/08/14/team-awarded-grant-to-help-digital-humanities-scholars-navigate-legal-issues-of-text-data-mining/

[9] Samberg, R. G., Althaus, S., Bamman, D., Butler, B., Cate, B., Courtney, K. K., Flynn, S., Gould, M, Hennesy, C., Koehl, E.D., Padilla, T., Reardon, S., Sag, M. Schofield, B. L., Senseney, M., Vollmer, T, & Worthey, G. (2021). Building Legal Literacies for Text Data Mining: Institute White Paper. https://docs.google.com/document/d/107Qmu595-7aOc2DPWc4vVDbz7PTULzn_20NOeNfBbWM/edit?usp=sharing

[10] Althaus, S., Bamman, D., Butler, B., Cate, B., Courtney, K. K., Flynn, S., Gould, M, Hennesy, C., Koehl, E.D., Padilla, T., Reardon, S., Sag, M. Samberg, R. G., Schofield, B. L., Senseney, M., Vollmer, T, & Worthey, G. (2021). Building legal literacies for text data mining. https://berkeley.pressbooks.pub/buildinglltdm/

geographical boundaries.[11,12,13] U.S. DH practitioners encounter expanding and increasingly complex cross-border problems. For example, as Fernández-Molina et al. describe,[14] U.S. contract law may supersede rights under copyright, such that a U.S. database license agreement may prohibit TDM and other fair uses, whereas UK licenses cannot. U.S. DH practitioners collaborating with UK-based colleagues face impactful choices about which agreements (and underlying corpus content) to rely upon, as this may determine whether TDM is permitted. Likewise, in the U.S., "breaking" technological protection measures to conduct TDM is now authorized within certain parameters.[15] Other jurisdictions prohibit such work or apply different conditions.[16],[17] U.S. DH TDM researchers must accordingly consider how they work with internationally-held or -licensed materials or collaborators.

We observed at least three such "cross-border" TDM scenarios that digital humanities practitioners must parse, including: (i) if the materials they want to mine are housed in a foreign jurisdiction, or are otherwise subject to foreign database licensing or laws; (ii) if the human subjects they are studying or who created the underlying content reside in another country; or, (iii) if the colleagues with whom they are collaborating reside abroad, yielding uncertainty about which country's laws, agreements, and policies apply. These may collectively be considered the "cross-border" DH TDM scenarios.

U.S.-based DH practitioners are uncertain about how to navigate each of these scenarios. As evidenced in an informal survey that we conducted, 70% of respondents reported cross-border copyright questions, 72% reported uncertainty about cross-border licensing terms, 52% noted privacy issues, and 48% identified ethical concerns. This confusion impacted their DH TDM research. Some scholars "slowed down the project because [they] didn't know what problems it might lead to," or tried "not to ask too many questions" because they were concerned that the law would not allow them to proceed. Twenty-eight percent (28%) of respondents confirmed that these cross-border copyright, licensing, privacy, or ethical issues impeded or prevented their project entirely. Of equal concern is that 40% of responding practitioners reported hesitation to share their workflows, methodology, or sources because of possible cross-border LLTDM

---

[11] Su, F. (2020). Cross-national digital humanities research collaborations: structure, patterns and themes. *The Journal of Documentation*. 76(6): 1295-1312. https://doi.org/10.1108/JD-08-2019-0159
[12] Nyhan, J., & Duke-Williams, O. (2014). Joint and multi-authored publication patterns in digital humanities. *Literary and Linguistic Computing*, 29(3), 387-399. https://academic.oup.com/dsh/article/29/3/387/986317
[13] Kemman, M. (2019, August). Boundary practices of digital humanities collaborations. *Digital Humanities Benelux Journal*. 1-24. https://journal.dhbenelux.org/journal/issues/001/Article-Kemman/kemman-main.tex.html
[14] Fernandez-Molina, J. C., Eschenfelder, K. R., Rubel, A. P. (2021). Comparing use terms in Spanish and US research university e-journal licenses: Recent trends. *College & Research Libraries*, 82(2), 158-181. https://crl.acrl.org/index.php/crl/article/view/24830/32667
[15] United States Copyright Office (2021). Final rule: Exemption to Prohibition on Circumvention of Copyright Protection Systems for Access Control Technologies. https://www.govinfo.gov/content/pkg/FR-2021-10-28/pdf/2021-23311.pdf
[16] Simoes, 2019. How we fixed DRM in Portugal, and so can you. https://fsfe.org/news/2019/news-20191113-01.en.html
[17] Flynn, S, Palmedo, M., Izquierdo, A. (2021). Research exceptions in comparative copyright law. PIJIP/TLS Research Paper Series no. 72. https://digitalcommons.wcl.american.edu/research/72

issues. Indeed, in an even more recent and formal survey of international TDM practitioners, the lack of understanding of cross-border issues was cited as a challenge and resulted in researchers: dropping their foreign research partners, ignoring complex legal questions (and hoping their research team would not be disciplined later), and abandoning particular research questions.[18]

Without methodological transparency, findings are deemed unreliable and scholarship may be rejected for publication. And without researcher and practitioner confidence in traversing cross-border LLTDM, knowledge and cultural advancement are stymied. These problems will only mount given the increasing collaborativeness of DH research and the substantial amount of cross-border DH research occurring.[19, 20]

We understood that DH TDM practitioners would benefit from guidance on navigating cross-border LLTDM. But, before law and ethics experts are able to create practicable educational materials, it is first important to assess and document the scope of cross-border issues that practitioners face. We thus designed LLTDM-X to elicit those issues, and yield at least some preliminary guidance while also identifying topics that would benefit from additional research. To that end, we held a series of three virtually-hosted round tables that facilitated the exchange of practitioner narratives and expert feedback. We anonymized and extrapolated information from the round tables, practitioner narratives, and expert analyses to create a case study that identifies key LLTDM issues that U.S. cross-border researchers face. We intend for this case study, along with its preliminary guidance, to inform and facilitate the development of educational resources and to help set a future research agenda.

Project Contributors and Activities

Project Contributors

Contributors to LLTDM-X included the team that proposed and advanced this grant project ("Project Team"), U.S.-based cross-border TDM Practitioners, and law and/or ethics Experts in cross-border TDM issues.

Project Team

The Project Team was responsible for contributing to project development and delivery, including: (1) identifying and securing participation from Experts and Practitioners; (2) designing, hosting, and moderating the round tables; and (3) drafting and curating written products,

[18] Aufderheide, P., & Butler, B. (n.d.). *The Chilling Effects of Obstacles to Accessing, Using, and Sharing In-copyright Data for Quantitative Research*. Forthcoming scholarly article on-file with the authors.
[19] Kemman, M. (2019, August). Boundary practices of digital humanities collaborations. *Digital Humanities Benelux Journal*. 1-24. https://journal.dhbenelux.org/journal/issues/001/Article-Kemman/kemman-main.tex.html
[20] Poole, A. H., & Garwood, D.A. (2018). Interdisciplinary scholarly collaboration in data-intensive, public-funded, international digital humanities project work. *Library & Information Science Research*, 40 (3-4), 184-193. https://doi.org/10.1016/j.lisr.2018.08.003

including: the Practitioners' two-page written accounts of their research project(s), Experts' written analyses, and the development of a case study and white paper.

LLTDM-X's Project Team included:

- Thomas Padilla, Internet Archive (Director)
- Rachael Samberg, UC Berkeley Library (Co-Director)
- Stacy Reardon, UC Berkeley Library[21] (Project Team Member)
- Timothy Vollmer, UC Berkeley Library (Project Manager)
- Catherine Falls, Internet Archive (Honoraria processing)

Practitioners

Practitioners were self-identified humanities TDM researchers whose projects included or were impacted by one or more cross-border issues as explained in the project description. From the outset of the project, we agreed to maintain confidentiality of the participating Practitioners. The reason for this was to make Practitioners feel comfortable in openly sharing their TDM challenges, both in the project write ups and round table conversations. Practitioners included researchers from the following U.S. institutions:

- Bowdoin College
- Massachusetts Institute of Technology
- Michigan State University
- North Carolina State University
- Stanford University
- Temple University
- University of Arizona
- University of California, Berkeley
- University of California, Los Angeles
- University of Michigan
- University of Minnesota Twin Cities
- University of Pennsylvania

Experts

Recruited by the Project Team, Experts contributed knowledge and experience in one or more cross-border LLTDM literacies (i.e. copyright, licensing, privacy, and ethics). While many Experts are professionally recognized as specialists in multiple LLTDM literacies, for the purposes of ensuring sufficient expertise across research project subjects, we assigned Experts to the particular domains noted below:

- Andrew Charlesworth, University of Bristol (privacy)
- Juan Carlos Fernández-Molina, Universidad de Granada (licensing)
- Sean Fiil-Flynn, American University Washington College of Law (copyright)

[21] Stacy Reardon has since become Instruction Librarian and Director, St. Michael's College Writing and Research Help Centre, University of Toronto.

- Lucie Guibault Dalhousie University (copyright, licensing)
- Heidi McKee, Miami University of Ohio (ethics)
- Argyri Panezi, IE Law School & Stanford University (privacy)
- James Porter, Miami University of Ohio (ethics)
- Matthew Sag, Emory University School of Law (copyright)
- Ben White, Bournemouth University (copyright)
- Fernando Esteban de la Rosa, Universidad de Granada (licensing)
- João Quintais, University of Amsterdam (copyright)
- Ryan Calo, University of Washington (privacy)

Financial support

LLTDM-X's core project activities included: (i) round table dialogues, preceded by Practitioner written statements to inform discussion, and (ii) preparation of Experts' written analysis responsive to Practitioners' statements and round table discussions; (iii) the Project Team's development of a case study, and (iv) this white paper.

All activities were hosted and undertaken in a fully virtual format (no in-person component), and the Project Team collaborated remotely for all of the written products. Project funds were dedicated to providing participation stipends to Experts and Practitioners, as follows:

- Experts received $1,500/person (reflecting a commitment estimated at 15 hours per person, and $100 per hour) as a financial incentive for participating in the project. Experts were expected to review and analyze the practitioners' two-page written accounts of their research project, conduct limited independent research as needed, attend and contribute to at least two round tables, and prepare Expert written analyses for distribution to the Practitioners. Project Team Members also served as Experts, and were compensated only for their contributions in this capacity.

- Practitioners (U.S.-based cross-border TDM researchers) received $800/person (reflecting an estimated 8 hours per person at $100 per hour) as a financial incentive for participating in the project. Practitioners prepared a two-page written account of their research project, and attended and contributed to one round table.

Activities

Identifying Practitioners and Experts

The Project Team identified 13 practitioners who have conducted TDM research involving one of the three "cross-border" scenarios explained above. Practitioners were recruited in part through the responses to our exploratory survey which we distributed through various DH email listservs. In order to solicit additional Practitioners, we published a blog post and created a short video to communicate project goals, eligibility, the application process, and remuneration.

We invited a group of 12 Experts with backgrounds supporting legal analysis of cross-border or international copyright, licensing, privacy, or ethics within TDM research. We identified the Experts through a combination of a literature review, their prior participation as faculty for the Building LLTDM Institute, and our professional networks.

Pre-Round Table Preparation & Statements

The Project Team coordinated round table preparation with the Practitioners and the Experts. We asked each Practitioner to write a two-page description of their TDM research, methodology, and any questions or challenges they faced related to cross-border LLTDM. We circulated the Practitioners' written statements to the Experts in advance of the first round table so that Experts could familiarize themselves with Practitioners' projects and self-identified challenges, and so Experts could prepare probing questions to ask during the first round table.

Some examples of cross-border LLTDM concerns put forward by the Practitioners included:

- Whether practitioners can assemble and mine a TDM corpus composed of materials published in or licensed from foreign countries, and whether foreign countries' copyright rules apply (or take precedence) when doing so;
- Whether practitioners can create (i.e. download or reproduce) and share (i.e. distribute) a TDM corpus with researchers located at institutions outside of the U.S., and particularly when such corpora contain materials licensed to a specific institution;
- Whether practitioners' cross-border colleagues can conduct TDM on a corpus and share that corpus and/or the results with the U.S. colleagues;
- Whether it is permissible in the U.S. to circumvent technological protection measures and/or transcoding on DVDs originally released in foreign countries, and whether foreign-based colleagues can export DVDs to U.S. colleagues to decrypt;
- Whether TDM researchers in the U.S. must comply with privacy laws in other countries when those privacy laws govern the people who are the subjects of the TDM research;
- How to work with scholars and research subjects on TDM projects in countries with authoritarian regimes, and the potential implications of sharing "big data" information from social media platforms if such data could endanger local populations due to government surveillance;
- Whether it is ethical to scrape social media if that content was posted by authors outside of the U.S., and to what extent should institutional review boards (or their equivalent) be involved from foreign countries; and
- How to address privacy and ethical concerns when doing TDM on materials like diaries or personal letters when the authors of those diaries and letters live(d) abroad and did not create the materials with the intention that they be used in TDM research.

We created the Zoom links and detailed agendas for all of the round tables and circulated them to the Practitioners and Experts. The Project Team also created an abbreviated summary of each Practitioner's project so that Experts would have easy-to-consult notes to guide them during the round table discussion.

Round Table 1

To facilitate participation from a diverse pool of both U.S.-based Practitioners and Experts (with some Experts participating from Europe), all round tables were conducted over Zoom. To encourage a comfortable environment conducive to sharing, Practitioners and Experts were also encouraged to introduce themselves asynchronously prior to the first round table.

The Project Team divided Round Table 1 into two segments, with the first segment focusing on Practitioners whose primary research challenges related to copyright or licensing, and the second segment focusing on those who predominantly faced privacy or ethical challenges.

In each segment, Practitioners began by sharing a 3-minute story in which they discussed their TDM cross-border challenges in response to pre-provided prompts. Through this storytelling exercise, we asked Practitioners to convey some or all of the following:

- How do or did cross-border legal or ethical problems affect your research?
- What are you worried about other researchers or the experts finding out about you or your processes?
- What specific questions do you want answered?
- What advice would you give to other TDM researchers involved in cross-border work?

Following their 3-minute mini-presentations, we invited the Practitioners to comment on and discuss each others' challenges, to identify areas of commonality, suggest guidance if possible, and highlight issues for which further guidance would be useful. After brief discussion amongst the Practitioners, we opened the floor for the Experts to comment on and discuss the Practitioners' projects and challenges, and ask questions.

At the close of Round Table 1, we engaged in a plenary group reflection enabling Practitioners to highlight learnings from the session.

Round Tables 2 and 3

In Round Tables 2 and 3, we relied on Experts to: (1) identify and describe the specific legal & ethical challenges they observed in the Practitioners' cross-border TDM research, and (2) reflect on what kind of guidance or education researchers will need to navigate those challenges. For Round Table 2, we convened those Experts whose focus is on copyright or licensing issues, grouping these experts together because these matters can be conceptually linked through jurisdictional variations on the permissibility of "contractual override" (i.e. when licenses circumscribe rights granted by copyright law). In Round Table 3, we brought together the privacy and ethics Experts because, conceptually, national variations in what is protected by privacy laws can be addressed relative to what is considered "private" in each country from an ethical perspective.

Project Outcomes

Expert feedback for Practitioners

Following the round tables, we charged each of the Experts with providing written feedback (consisting of a few paragraphs) to at least two Practitioners. The purpose of these brief Expert analyses was to provide responsive and tailored analysis to the Practitioners about how they might address specific issues relevant to and reflected in each Practioner's stated research project. Naturally, we aligned each Expert with a Practitioner whose project raised issues within that Expert's domain. So, for example, an Expert whose primary expertise was copyright law was asked to craft feedback for a Practitioner whose project evoked cross-border copyright challenges.

Case study & white paper

Extrapolating from the issues discussed in the round tables, the Practitioners' statements, and the Experts' written analyses, the Project Team developed a hypothetical case study reflective of "typical" cross-border LLTDM issues that U.S.-based practitioners encounter. The case study provides basic guidance to support U.S. researchers in navigating cross-border TDM issues, while also highlighting questions that would benefit from further research.

We then prepared this white paper to reflect upon the issues and guidance in the case study, and to make preliminary recommendations for future development of LLTDM-X training modules.

Project documentation

We have made the following LLTDM-X materials publicly available:

- Case Study
- Writing prompts for LLTDM-X Researchers/Practitioners
- Website and blog posts
- Roundtable 1 slide deck with transcript notes

In order to promote candid discussion over the course of the project, we have withheld the following materials from public disclosure:

- Practitioners' two-page project write-ups detailing their cross-border LLTDM challenges
- Experts' written analyses tailored to particular Practitioners' projects
- Round table meeting recordings
- Round table notes

Takeaways & Recommendations

Project Takeaways

1. Uncertainty about cross-border LLTDM issues hinders U.S. TDM researchers, confirming the need for further research and education.

LLTDM-X emerged from scholars' stated needs during the Building LLTDM Institute that cross-border law and ethics issues pervaded their research. The Project Team knew from our informal survey analysis[22] (which has since been supported through more formal research conducted by others[23]) that DH TDM scholars' uncertainty about cross-border LLTDM issues deter and even preclude their research. The LLTDM-X project has confirmed that this uncertainty about cross-border legal and ethical issues indeed hinders U.S. TDM researchers both from taking on cross-border research questions and from partnering with scholars abroad, and that further educational guidance and advocacy is needed.

In Practitioners' written statements and round table discussions, the majority of LLTDM-X Practitioners noted they could not see any way forward with their DH TDM research due to concerns regarding cross-border legal and ethical issues—and, in particular, concerns about copyright. Regarding copyright, Practitioners expressed surprise to learn that typically these perceived copyright hurdles were not insurmountable because of (i) the availability of U.S. fair use exception, and (ii) the opportunity for researchers to disseminate analysis or derived data outputs rather than the underlying corpus. Conversely, Practitioners expressed equal surprise to learn that often the more decisive hurdle in their research would be negotiating *contractual rights* to share corpus content with other researchers. As such, educational modules that provide step-by-step but also reassuring guidance about copyright matters, along with specific recommendations to address or negotiate around contractual limitations, might give scholars a more actionable approach for pursuing their TDM research projects.

That said, certain parameters of law and resulting risk associated with cross-border TDM projects still remain largely unknown. For this reason, the case study captures all questions that Practitioners identified as impediments, not only to guide the development of instructional content but also to flag those issues that need further research to address more conclusively.

2. Broader education regarding U.S.-centric LLTDM literacies should also continue.

In addition to demonstrating the need for education on cross-border literacies, LLTDM-X revealed the need for ongoing education regarding U.S.-centric LLTDM literacies.

---

[22] We distributed a survey through local and national DH email listservs and the UC Berkeley Library's scholarly communications Twitter account. Twenty-nine researchers responded. Survey results are on file with the authors.

[23] Aufderheide, P., & Butler, B. (n.d.). *The Chilling Effects of Obstacles to Accessing, Using, and Sharing In-copyright Data for Quantitative Research*. Forthcoming scholarly article on-file with the authors.

In the previous Building LLTDM Institute, we demonstrated the efficacy of design thinking as a way to teach legal literacies for TDM, as this approach yielded increased researcher confidence.[24] That Institute, offered in 2020, supported a small cohort of only 32 scholars and, although it was followed by a comprehensive open educational resource (OER)[25] that expanded its reach and impact, the vast majority of DH TDM researchers continue to lack formal education about the legal and ethical nuances of text and data mining in the U.S.[26]

LLTDM-X revealed that the need to expand U.S.-centric LLTDM training in DH curricula is paramount. This was evidenced by the fact that many Practitioners in LLTDM-X described having certain "cross-border" LLTDM problems that were readily addressable with further guidance about and foundations in U.S. law alone. For instance, multiple Practitioners described fears and hesitancy about proceeding with mining copyright-protected materials that were published in foreign countries. They felt that foreign copyright laws would prohibit them from conducting TDM. Yet, particularly given that the U.S. participates in multilateral treaties like the Berne Convention, the law of the country in which the TDM activities are *performed* governs the infringement analysis; the law of the country in which the works were published is not controlling. As such, U.S. TDM researchers can rely entirely on U.S. copyright law and the parameters of its fair use exception. Researchers expressed relief at learning this.

While Building LLTDM yielded a useful instructional model and materials by which to educate U.S.-based researchers on LLTDM, U.S. research institutions and universities would benefit from additional funding to incentivize and facilitate the integration of these materials into their curricula. Particularly in light of forthcoming findings from other scholars studying perceptions of legal impediments to TDM research,[27] we view ongoing U.S.-related LLTDM training as an essential counterpart to the cross-border instructional modules that must be created.

---

[24] Samberg, R. G., Althaus, S., Bamman, D., Butler, B., Cate, B., Courtney, K. K., Flynn, S., Gould, M, Hennesy, C., Koehl, E.D., Padilla, T., Reardon, S., Sag, M. Schofield, B. L., Senseney, M., Vollmer, T, & Worthey, G. (2021). Building Legal Literacies for Text Data Mining: Institute White Paper. https://docs.google.com/document/d/107Qmu595-7aOc2DPWc4vVDbz7PTULzn_20NOeNfBbWM/edit?usp=sharing

[25] Althaus, S., Bamman, D., Butler, B., Cate, B., Courtney, K. K., Flynn, S., Gould, M, Hennesy, C., Koehl, E.D., Padilla, T., Reardon, S., Sag, M. Samberg, R. G., Schofield, B. L., Senseney, M., Vollmer, T, & Worthey, G. (2021). Building legal literacies for text data mining. https://berkeley.pressbooks.pub/buildinglltdm/

[26] Indeed, scholars remain challenged to understand their legal rights to develop and run TDM projects in the U.S. today, particularly in a fast-changing environment of artificial intelligence. See "Taking Down Prosecraft.io" at https://blog.shaxpir.com/taking-down-prosecraft-io-37e189797121

[27] Aufderheide and Butler conducted an international survey to assess the challenges faced by text data mining researchers in using in-copyright data for quantitative study. One-third of the respondents engage in TDM in the United States. The forthcoming article notes: "Researchers experience challenges in accessing, using, sharing, and storage of in-copyright data. The sources of the problems are high prices for proprietary data, terms of use that inhibit research, and legal policies including copyright, privacy, and anti-hacking. Consequences of facing this range of obstacles include changing research design, delaying research, abandoning research, and failure to collaborate across institutional or jurisdictional borders." Aufderheide, P., & Butler, B. (n.d.). *The Chilling Effects of Obstacles to Accessing, Using, and Sharing In-copyright Data for Quantitative Research*. Forthcoming scholarly article on-file with the authors.

3. Disparities in national laws may incentivize TDM researcher "forum shopping" and exacerbate scholarly bias.

In Round Tables 2 and 3, LLTDM-X Experts observed that national differences in copyright, contracts, and privacy laws across jurisdictions appear to have outsized impact on, or perhaps unintentionally incentivize, researchers' selection of particular corpora or their research focus on certain jurisdictions.

National variations in copyright laws can be used to demonstrate this point. All countries have implemented copyright exceptions to support activities like scientific or scholarly research. Some of these exceptions—like fair use in the United States—may also via statute or through judicial interpretation authorize TDM research. However, approximately only one fifth of countries' research exceptions are broad enough to permit the full range of TDM research, which requires the ability to copy, share, and analyze whole works in collaboration with others.[28] As explained by Flynn et al. (2022),[29] "some countries have research exceptions that permit uses only of excerpts of a work (e.g., Argentina), do not apply to uses of books or other kinds of works (e.g., most post-Soviet countries), or require membership in a specific research institute (e.g., Sweden)."

The resulting impact of these variations is underscored by the following hypothetical: Imagine a U.S.-based DH TDM researcher who desires to partner with a scholar in Spain on a TDM research project, with some corpus content to be downloaded or reproduced by each researcher in their respective countries and then distributed across borders. The trajectory of legal analysis is as follows:

- These acts are first governed by the Directive on Copyright in the Digital Single Market (DCDSM), which generally supports the research and TDM uses being described here, provided that among other things there is no subsequent dissemination of the underlying corpus publicly. But applying the DCDSM is not the end of the inquiry. The DCDSM "is a legislative act that sets out a goal that all EU countries must achieve. However, it is up to the individual countries to devise their own laws on how to reach these goals."[30] And, although the DCDSM imposes minimum requirements, "national laws still have a margin of discretion on how they implement the different elements of the legal regime, especially at this early stage of implementation, before the Court of Justice of the EU steps in."[31]

- As such, the next step of the inquiry is to apply the national law of the country. In this case, the copyright law of Spain limits copyright exceptions to a "personal" or "private"

[28] Flynn, S., Schirru, L., Palmedo, M., & Izquierdo, A. (2022). Research Exceptions in Comparative Copyright. *Joint PIJIP/TLS Research Paper Series*. https://digitalcommons.wcl.american.edu/research/75
[29] Fiil-Flynn, S. M., Butler, B., Carroll, M., Cohen-Sasson, O., Craig, C., Guibault, L., Jaszi, P., Jütte, B. J., Katz, A., Quintais, J. P., Margoni, T., de Souza, A. R., Sag, M., Samberg, R., Schirru, L., Senftleben, M., Tur-Sinai, O., & Contreras, J. L. (2022). Legal reform to enhance global text and data mining research. *Science*, *378*(6623), 951–953. https://doi.org/10.1126/science.add6124
[30] https://european-union.europa.eu/institutions-law-budget/law/types-legislation_en
[31] Written project feedback from João Pedro Quintais, Institute for Information Law, University of Amsterdam. On file with authors.

right or use, and has been interpreted to mean that the Spanish researcher would be restricted from reproducing and distributing the copyright-protected corpus to other researchers.[32] This result might discourage a U.S.-based researcher from partnering with a Spanish colleague for TDM due to the distribution restriction, and incentivize the U.S. researcher to partner instead with a scholar from England or Germany, which have open research exceptions that would allow the desired corpus sharing within the research group.[33]

- Further, copyright law in England prohibits license agreements from taking away (or overriding) rights granted under copyright law.[34] In the United States, by contrast, licensing agreements can circumscribe the rights afforded by the Copyright Act. The prohibition on derogation of statutory rights in England might incentivize a U.S.-based TDM researcher to partner with an England-based researcher who is more likely to have favorable fair dealing or research-sharing provisions in their institutional license agreements.

Overall, the implications of these variations in national laws—whether as to copyright, licensing, or privacy matters—may exacerbate bias in the nature of research questions being studied (e.g. perhaps leaving research questions affecting countries like Spain underexplored relative to those impacting more copyright "permissive" countries like England) or the types of materials being used to study them (e.g. perhaps favoring use of public domain works not protected by any copyright laws).[35]

The World Intellectual Property Organization is considering this fragmented landscape at least with respect to copyright law, reviewing whether (or how) to harmonize research exceptions to facilitate cross-border TDM research.[36] But while harmonization of copyright exceptions faces a possible (though uncertain) route forward, licensing and privacy laws are unlikely candidates for synthesis altogether. Overall, TDM researchers will need substantial instructional guidance on understanding these shifting cross-border and extraterritorial implications of copyright, contracts, and privacy laws—underscoring the importance of future educational materials.

---

[32] "The most common of these exceptions extend to research uses as a category of "private" or "personal" use. By virtue of the use of the term "private" or "personal," we assume that none of these exceptions authorizes sharing with other researchers…" Flynn, S., Schirru, L., Palmedo, M., & Izquierdo, A. (2022). Research Exceptions in Comparative Copyright. *Joint PIJIP/TLS Research Paper Series*. p. 26, Table 4. https://digitalcommons.wcl.american.edu/research/75

[33] Ibid, at p. 17, Table 1.

[34] "To the extent that a term of a contract purports to prevent or restrict the making of a copy which, by virtue of this section, would not infringe copyright, that term is unenforceable." Copyright, Designs and Patents Act 1988. Statute Law Database. Retrieved August 23, 2023, from https://www.legislation.gov.uk/ukpga/1988/48/part/I/chapter/III

[35] Levendowski, A. (2018). How Copyright Law Can Fix Artificial Intelligence's Implicit Bias Problem. *Washington Law Review*, *93*(2), 579. https://digitalcommons.law.uw.edu/wlr/vol93/iss2/2/

[36] Fiil-Flynn, S. M., Butler, B., Carroll, M., Cohen-Sasson, O., Craig, C., Guibault, L., Jaszi, P., Jütte, B. J., Katz, A., Quintais, J. P., Margoni, T., de Souza, A. R., Sag, M., Samberg, R., Schirru, L., Senftleben, M., Tur-Sinai, O., & Contreras, J. L. (2022). Legal reform to enhance global text and data mining research. *Science*, *378*(6623), 951–953. https://doi.org/10.1126/science.add6124

4. License agreements often dominate analysis of cross-border TDM permissibility.

For the majority of the research scenarios presented by the LLTDM-X Practitioners, copyright exceptions like fair use or fair dealing would enable the Practitioners' intended TDM research methodologies, except to the extent the researchers also desired to share or distribute corpora (rather than extracted or derived research outputs or analysis). More often, the controlling impediment to Practitioners' research plans were those limitations arising through institutional license agreements. Although various countries' national laws prohibit contracts from overriding copyright exceptions like TDM,[37] the United States does not. The effect of this is that TDM researchers at U.S. institutions who are licensing content from or partnering with researchers in other "override" countries may be contractually prohibited from scraping, reproducing, or sharing corpora with other researchers, or even conducting TDM all-together—while their international collaborators may not be similarly bound, or may be subject to contractual override only for certain copyright-protected acts (e.g. distribution of a corpus) but not others (e.g. the actual text mining of the corpus).

It typically remains up to the U.S. research institution to negotiate to preserve these rights for their researchers, and many universities are not always successful in such endeavors. This is because publishers and vendors are sometimes unwilling to license such rights at any costs (for fear that they will not be able to "control" the dissemination of their content), or alternatively the publishers seek to charge institutions additional (and increasingly out-of-reach) sums to authorize TDM, reproduction, or other distribution rights for database content. When either these costs or license restrictions are unworkable for the institution as a whole, the publisher or vendor may then offer similar contractual terms directly to research teams, who may feel obliged to agree in order to get access to the content they need.

Overall, this landscape of trying to negotiate around contractual override in the U.S. can result in cross-border TDM problems like: (i) incentivization of research using "low friction" materials (e.g. public domain works online) that are not otherwise subject to license agreements, potentially leaving important cross-border research questions unanswered; and (ii) siloization of information, if institutions cannot reach suitable cross-border TDM terms, and project teams are left to independently try to license database content at cost. The latter scenario would also impede research by teams who lack grant or other funds to cover these database fees directly—penalizing research in or about underfunded disciplines or geographical regions, and potentially resulting in bias as to the topics and regions studied.

There is no easy solution to these problems, as jurisdictions like the United States endorse the private right to contract, and confine oversight of agreements to circumstances in which fundamental freedoms are at stake or where contractual provisions contravene public policy. Some scholars have proposed an "efficient breach" theory that could (if endorsed or adopted) enable TDM researchers in the U.S. to breach license agreements that override fair use and

[37] Band, J. (2023). Protecting User Rights Against Contract Override. *Joint PIJIP/TLS Research Paper Series*. https://digitalcommons.wcl.american.edu/research/97

other fundamental copyright exceptions.[38] In all events, contractual override often remains the controlling limitation on cross-border TDM research.

5. Emerging lawsuits about generative artificial intelligence may impact understanding of fair use and other research exceptions in cross-border TDM.

Cross-border TDM researchers, and those institutional professionals providing guidance to them, may be affected by the outcomes of pending litigation involving so-called generative artificial intelligence ("generative AI")—particularly for TDM projects that create new content through an algorithmic engine rather than merely analyze existing content.

As background, it is helpful to understand that there has been a surge in the development of generative AI tools and platforms over the last few years, including for instance the chatbot ChatGPT, and of text-to-image generators like DALL-E, Midjourney, and Stable Diffusion.[39] These tools rely on the creation and use of large language models, which in turn must be trained on a wide swath of content, including copyrighted works. Rightsholders have expressed concerns over (1) AI tools being trained on copyright-protected works without their consent, and (2) the new or generative output that the AI tools can be used to create. Accordingly, they have filed several lawsuits alleging copyright infringement by such artificial intelligence platforms.[40]

One question that will need to be resolved in these lawsuits is whether the ingestion and training of the large language models that power generative AI platforms is an infringement of copyright, or whether this training conduct is considered a fair use under U.S. law (and under what circumstances, such as research contexts versus commercial ones). Courts will naturally look to legal precedents on TDM. Up until now, court decisions in the U.S.—including *Authors Guild v. Google*[41] and *Authors Guild v HathiTrust*[42]—have ruled that in certain contexts, compiling a corpus of copyrighted works and conducting TDM on those works is a transformative fair use under copyright law. Specifically, these previous cases confirmed that making copies of in-copyright books for the purposes of creating a full-text searchable database to glean new

---

[38] Samuelson, P., & Carroll, M. (2023, May 19). *Fair Breach*. American University University College of Law User Rights Network Symposium. Presentation materials on-file with the authors.
[39] The U.S. Copyright Office has discussed this further in its recent Notice of Inquiry and Request for Comments. See United States Copyright Office. (2023, August 30). *Notice of inquiry and request for comments: Artificial Intelligence and Copyright*. Federal Register. https://www.federalregister.gov/documents/2023/08/30/2023-18624/artificial-intelligence-and-copyright
[40] For example, see Vincent, J. (2023, February 6). *Getty Images sues AI art generator Stable Diffusion in the US for copyright infringement*. The Verge. https://www.theverge.com/2023/2/6/23587393/ai-art-copyright-lawsuit-getty-images-stable-diffusion and Setty, R. (2023, January 17); and *AI Art Generators Hit With Copyright Suit Over Artists' Images*. Bloomberg Law. https://news.bloomberglaw.com/ip-law/ai-art-generators-hit-with-copyright-suit-over-artists-images.
[41] Authors Guild, Inc. v. Google, Inc., 804 F.3d 202 (2d Cir. 2015). Available at https://casetext.com/case/guild-v-google-inc-1
[42] Authors Guild, Inc. v. HathiTrust, 755 F.3d 87 (2d Cir. 2014). Available at https://casetext.com/case/authors-guild-inc-v-hathitrust-1

insights and understandings by using the works in a non-expressive, non-consumptive fashion is considered a fair use, thus not infringing copyright in the original works.

But some observers question whether the legal framework of computational text analysis that underlies these past suits will remain relevant to the sometimes more complicated implementations of generative artificial intelligence platforms.[43] A key difference in the new litigation is that, while the more "traditional" TDM activities (like those at issue in the Google Books and HathiTrust cases) allow researchers to run algorithms in order to extract new understandings across an existing corpus, those TDM activities do not actually potentially create derivative and/or de facto infringing works from the original content. As Prof. Matthew Sag notes: "these [generative AI] systems produce much more than information *about* expression; they are now the engines of *new content creation*."[44] And one question at hand is whether the outputs of AI tools might infringe on the copyrights of the training data used as inputs.

At the time of the call for participation in LLTDM-X, issues surrounding generative AI had not yet surfaced in Practitioners' cross-border TDM projects. Therefore, generative AI legal and ethical topics were not part of the round table discussions; as such, we are not able to offer significant guidance yet pending the outcome of litigation or other regulatory changes. But with the spate of recent lawsuits, and the Copyright Office's Notice of Inquiry[45] inviting comments about copyright and AI, we anticipate that any future cross-border educational materials must confront generative AI issues directly.

6. Research is needed into issues of foreign jurisdiction, likelihood of lawsuits in foreign countries, and likelihood of enforcement of foreign judgments in the U.S. However, overall "risk" of proceeding with cross-border TDM research may remain difficult to quantify.

While the impacts of some national variations in copyright, contract, and privacy laws could reasonably be addressed by LLTDM-X experts, it was clear from the round tables that the key concern needing further research before educational materials can be prepared is that of overall "risk." In particular, research is needed into the risk-related issues of: (i) the propriety of foreign courts' jurisdiction over U.S. TDM researchers, (ii) the likelihood of lawsuits arising in foreign countries, and (iii) the likelihood of a U.S. court agreeing to enforce a foreign judgment against a U.S. researcher.[46] Even if these research questions are explored through case law, however, the

[43] In a forthcoming article, Prof. Matthew Sag writes, "sweeping claims that generative AI is predicated on massive copyright infringement are misplaced; but it also acknowledges that in specific—but perhaps rare—contexts, the process of creating generative AI may cross the line from fair use to infringement because large language models sometimes "memorize" the training data rather than simply "learning" from it." Sag, M. (2023). *Copyright Safety for Generative AI* (SSRN Scholarly Paper 4438593). https://doi.org/10.2139/ssrn.4438593, p. 6. See also United States Copyright Office. (2023, August 30). *Notice of inquiry and request for comments: Artificial Intelligence and Copyright*. Federal Register. https://www.federalregister.gov/documents/2023/08/30/2023-18624/artificial-intelligence-and-copyright
[44] Ibid, at 11.
[45] *Notice of inquiry and request for comments: Artificial Intelligence and Copyright*. Federal Register. https://www.federalregister.gov/documents/2023/08/30/2023-18624/artificial-intelligence-and-copyright
[46] Complicating the matter is that U.S. state laws differ in their approach to enforcement of foreign judgments. For instance, in California, the "California Recognition Act" allows a court to decline to

overall "risk" of proceeding with cross-border TDM research may remain difficult to quantify. That is because there are a number of different types or perceptions of risks that violating laws and policies may impose.[47]

As a preliminary matter, lawsuits may impose either injunctions (i.e. orders to stop behavior), or damages (i.e. monetary sanctions), or both. Researchers may perceive the threat of injunctions to be less "risky" than the potential for damages. Unfortunately, laws relating to damages and injunctions, and their availability and scope, vary by state and country—making universal guidance difficult.

In addition, there are other types of risks that may arise in cross-border TDM research:

- *Risks to researchers*: There could be reputational harms associated with violating agreements or knowingly infringing. Some publishers may also refuse publication or retract papers when violations come to light.
- *Risks to institutions*: Institutions could face litigation costs and loss of access to key resources (e.g. if access for the campus is terminated as a result of an individual's violation of a license agreement)
- *Risks to subjects / third partie*s: Rights holders, vulnerable or marginalized communities, and data subjects may face varying types and degrees of harm (e.g. danger, shame, ridicule) if their expectations of privacy or obscurity are breached or exceeded.

Any guidance developed for cross-border TDM researchers should account not only for known legal outcomes but also, to the extent possible, address perceptions of relative risk.

7. Institutional review boards (IRBs) have an opportunity to explore a new role or build partnerships to support researchers engaged in cross-border TDM.

Which campus entity or entities should support scholars with matters of cross-border LLTDM? What campus partnership models will best serve such researchers, and which staffing models are feasible? These are questions with no easy answer, but ones that must be addressed because researchers are currently being under-served by campus departments.

It is helpful to reflect first upon the fact that providing guidance on copyright and license agreements in research and publishing typically falls within the purview of scholarly communication offices, which are often situated within academic libraries. Sometimes these

---

recognize a foreign-country money judgment if the "judgment or the cause of action or claim for relief on which the judgment is based is repugnant to the public policy of [California] or of the United States." Cal. Civ. Proc. Code § 1716(c)(3). However, the bar for satisfying "repugnancy" (and thus declining to enforce the foreign judgment) is very high. *Ohno v. Yasuma*, 723 F.3d 984 (9th Cir. 2013). As explained most relevantly in *De Fontbrune v. Wofs*y, 39 F.4th 1214 (9th Cir. 2022), "The issue is not simply whether the 'foreign judgment or cause of action is contrary to our public policy,' Rather, the question is whether either is 'so offensive to our public policy as to be prejudicial to recognized standards of morality and to the general interests of the citizens.' [citations omitted]."

[47] Preliminary guidance prepared by Matt Sag, and on-file with authors.

scholarly communication offices also provide scholars instruction on privacy law and ethics within their research, but this varies by institution.

In all events, matters of copyright and licensing are typically well beyond what U.S. IRBs can or do support. IRBs instead focus on ensuring human subject consent and avoidance of harm under the so-called "common rule"[48]. With their regulatory mandates framed around consent, IRBs are seemingly well-positioned to address questions of privacy law, and possibly the intersection of privacy law and ethical concerns. And indeed, many Practitioners involved in LLTDM-X did try to seek support from their institutional IRBs—particularly if they were researching issues in or about free speech-restricted countries, given their fear that the TDM research could expose individuals to harm. (Highlighting even publicly-available materials such as social media posts could result in political reprisal for the posters or authors of such content in repressive jurisdictions.) But in not a single instance did a cross-border TDM practitioner report that their IRB had provided them with sufficient support or guidance on these challenges.

Were IRBs to begin providing or expand their existing support for privacy and ethical issues within cross-border TDM research, they would need sufficient expertise to be able to address variations in national privacy laws and the potential extraterritorial reach of such laws. In addition, IRBs would need to be able to address regional variations in *perceptions* of privacy (from an ethical perspective) that could result in harm to individuals even where national laws do not extend formal protections. Certainly, the recommendations from the Association of Internet Researchers (AoIR) may provide some independent guidance to researchers on these privacy or ethical considerations[49], but AoIR is not a form of oversight for institutions.

Determining which campus entities or units should be involved in supporting cross-border TDM researchers is challenging as higher education resources are increasingly depleted. Nevertheless, these are not problems that university research offices can afford to continue ignoring: Practitioners report being under-supported by their IRBs. If IRBs do not or cannot become involved in the full spectrum of cross-border LLTDM guidance, at a minimum campuses should have a plan for what other entities, or what alternative forms of support or guidance, are available.

Next steps & Recommendations

Through blog posts, social media engagement, and presentations, we will broadly share this white paper and the case study to begin helping U.S.-based TDM researchers navigate cross-border LLTDM hurdles. We will also speak publicly to educate researchers and the TDM community regarding project takeaways, and to advocate for legal and ethical experts to

[48] See discussion of the "common rule" discussed in https://berkeley.pressbooks.pub/buildinglltdm/chapter/ethics/; *see also* National Commission for the Protection of Human Subjects of Biomedical and Behavioral Research. (1979). Belmont Report: Ethical Principles and Guidelines for the Protection of Human Subjects of Research. https://www.hhs.gov/ohrp/regulations-and-policy/belmont-report/index.html.
[49] Association of Internet Researchers. (n.d.). *Ethics*. Retrieved September 13, 2023, from https://aoir.org/ethics/

undertake the essential research questions and begin developing much-needed educational materials. And, we will continue to encourage the integration of LLTDM literacies into DH curricula, to facilitate both domestic and cross-border TDM research.

A final note on the nature of educational materials to be created: In developing the case study we felt that, given the specialized nature of cross-border LLTDM issues, future guidance might benefit from instruction supported by visual aides. To experiment with this, we engaged in nascent efforts to graphically depict the cross-border legal parameters, but the complexity of the variables soon proved too unwieldy to readily chart in the context of this project. We believe it advisable for future cross-border LLTDM projects to specifically budget for the development of educational materials that include helpful (albeit perhaps challenging-to-develop) graphical representations of the cross-border LLTDM problems.