



Information Research – Vol. 31 No. iConf (2026)

Intermediary or gatekeeper? Defining content responsibility boundaries of open research data platforms

Zhengzheng Tian

DOI: <https://doi.org/10.47989/ir31iConf64168>

Abstract

Introduction. Open research data platforms often position themselves as ‘intermediaries’ yet exercise ‘*gatekeeper*’ power through design and policy choices. Despite extensive content governance scholarship on social media platforms, research data platforms remain underexplored.

Method. I employ reflexive thematic analysis of publicly available documents from three platforms: Zenodo (publicly funded generalist), ICPSR (membership-based disciplinary), and Mendeley data (commercial publisher-integrated). This comparative approach examines how distinct platform types navigate content responsibility.

Results. Platforms transcend binary intermediary-gatekeeper distinctions through multi-dimensional responsibility allocation. I construct two themes: curatorial boundary-drawing between form and substance, where platforms moderate formal compliance while adopting different approaches to scientific curation; and selective responsibility allocation, whereby platforms delegate immediate legal liability to users while internalizing long-term archival stewardship.

Conclusion. I contribute a framework distinguishing ‘*epistemic responsibility*’ from ‘*archival responsibility*’ and propose a typology of infrastructural intermediary, disciplinary gatekeeper, and commercial hybrid models. I further argue for ‘*directional flexibility*,’ where gatekeepers can scale by adopting intermediary functions, whereas intermediary platforms face economic barriers to adopting substantive gatekeeping.

Introduction

Open research data platforms face a fundamental tension in scholarly communication. They often position themselves as neutral ‘*intermediaries*’ that facilitate data sharing, yet simultaneously exercise substantial ‘*gatekeeper*’ power through technical design, policy frameworks, and quality control mechanisms.

This duality creates distinct governance challenges. Platforms must ensure scientific integrity while maintaining openness, balance immediate accessibility with long-term preservation, and navigate complex stakeholder relationships involving researchers, institutions, funders, and publishers. In this context, traditional intermediary-gatekeeper distinctions become insufficient. When platforms claim intermediary status while practicing gatekeeping, curatorial decisions determining which research contributions gain visibility, validation, and preservation become obscured.

Despite extensive critical scholarship on content governance in social media platforms (Gorwa et al., 2020), research data platforms remain largely unexamined. As research data sharing increasingly relies on platforms, understanding how platforms define content responsibility becomes essential.

This study addresses this gap through a comparative analysis of Zenodo, ICPSR, and Mendeley data. By examining how these distinct platforms manage the intermediary-gatekeeper tension, this research extends content governance theory to scholarly infrastructure contexts, contributing to both platform studies and scholarly communication literature. Specifically, this study addresses three research questions:

- How do open research data platforms define their content responsibility across the intermediary-gatekeeper spectrum?
- What strategies do platforms employ to manage the tension between neutrality claims and curatorial practices?
- How do these content governance strategies differ across platform models?

Literature review

Scholarship recognizes that digital platforms are not neutral conduits but active mediators shaping social practices through technical design and governance architectures (Frapporti, 2024; Hesmondhalgh et al., 2023). A key concept for understanding this mediating role is content governance, which Gorwa (2019a; 2019b) distinguishes from content moderation. While moderation often focuses on identifying and addressing illegal content, information incompatible with providers’ terms and conditions, and other forms of harmful or undesirable material (Flew et al., 2019; Busch, 2022), governance encompasses comprehensive strategic frameworks that shape information environments through policy design, curation, and stakeholder orchestration (Celeste et al., 2023; Gorwa, 2019c). Central to content governance is content responsibility, the normative boundaries a platform draws around its accountability for hosted content, addressing the fundamental question of ‘*who is responsible for what?*’ (Gorwa, 2019b). This question is critical because platforms routinely perform gatekeeping functions through technical architectures while claiming intermediary legal protections, rendering pure intermediary status practically unsustainable (Frapporti, 2024; Gillespie, 2018). Consequently, each seemingly technical or neutral governance choice embodies specific assumptions about responsibility allocation, rendering content governance practices operational expressions of defined content responsibility.

Content governance research has been driven primarily by social media analyses, documenting shifts from community self-regulation to corporate-controlled moderation systems (Celeste et al., 2023; Gorwa et al., 2020). However, research data platforms operate under different conditions

that make social media frameworks inadequate. Unlike user-generated social media content, research data constitutes the formal scholarly record, requiring long-term preservation, version control, and provenance tracking (National Academy of Sciences et al., 2009, pp. 26-27; Schöpfel et al., 2017; Schreier et al., 2006).

This stewardship context intensifies the intermediary-gatekeeper tension. Research data platforms are often expected to exercise substantive quality assessment befitting scholarly gatekeepers, yet they frequently operate under resource constraints that limit such oversight. Moreover, their positioning within the intermediary-gatekeeper spectrum carries direct implications for how responsibility is allocated among platforms, depositors, and the broader scholarly community. However, how platforms navigate this tension, balancing stewardship expectations against operational realities, remains undertheorized.

Method

This study adopts a comparative case study design, utilizing reflexive thematic analysis (RTA) to explore how open research data platforms define and practice their content responsibility.

Case selection and data collection

I purposefully sampled three platforms representing distinct governance models across the intermediary-gatekeeper spectrum:

- Zenodo: A publicly funded generalist repository, chosen for its permissive deposit standards, broad disciplinary scope, and commitment to minimal content intervention.
- ICPSR: A membership-based disciplinary archive, selected for offering distinct ‘*professional curation*’ (rigorous review) and ‘*self-publishing*’ (immediate user-managed release) options.
- Mendeley data: A commercial publisher-integrated platform, chosen for its integration with manuscript submission workflows and connection to Elsevier’s broader academic ecosystem.

For each case, I collected documents between June and August 2025, including platform policies (e.g., terms of service), operational guidance (e.g., user guides), and platform presentation texts (e.g., official blogs).

Data analysis

I employed RTA (Braun & Clarke, 2019, 2024), which demands reflexive examination of the researcher’s interpretive lens. As a researcher in scholarly communication with involvement in open science initiatives and internship experience in academic publishing, I recognized that my background might predispose me to naturalize certain platform practices, such as viewing preservation as self-evident rather than a governance choice, to hold normative assumptions about data openness, or to be particularly attuned to commercial integration strategies. I remained alert to how these orientations might privilege certain interpretive pathways.

As the sole researcher, I adopted a latent, primarily inductive approach, focusing on implicit assumptions and strategic positioning underlying platform discourse rather than surface-level content descriptions. I generated codes through iterative, reflexive engagement, focusing on platforms’ construction of responsibility, curation practices, and self-positioning. I moved recursively between data and codes, continuously refining interpretations. For instance, initial codes distinguishing active versus passive moderation were progressively refined to capture the nuanced boundary-work between formal and substantive curation. Through interpretive analysis, I constructed candidate themes by identifying patterns and questioning what platform discursive moves accomplish.

Throughout this process, I documented analytic decisions and reflexive thoughts to ensure interpretive rigour. After developing themes for each platform, I conducted comparative analysis examining how responsibility attribution, content control, and self-positioning strategies varied across cases. I also utilized temporal distance as a reflexive strategy, revisiting the analysis after initial submission to refine interpretations.

Findings

I constructed two themes that capture how platforms navigate content responsibility, each comprising two subthemes (see Appendix, Figure A1).

Theme 1 curatorial boundary: form versus substance

All three platforms exercise curatorial control but draw different boundaries regarding the focus of their intervention. While all platforms actively regulate formal compliance, they position themselves differently regarding scientific curation. This boundary-drawing reflects how platforms navigate content responsibility through selective intervention, locating them differently along the intermediary-gatekeeper spectrum.

Subtheme 1.1 active moderation of formal compliance

Zenodo and Mendeley data implement moderation to maintain platform order. Mendeley data mandates a moderation period to verify submissions against formal requirements, ensuring they contain no executable files or sensitive personal information. Zenodo similarly employs automated and manual measures to mitigate spam and misuse, maintaining takedown procedures for violations like copyright infringement.

ICPSR presents a more complex picture through its dual-track model. Its self-publishing option via openICPSR aligns with the minimal intervention approach of Zenodo and Mendeley data, allowing depositors to make data immediately available as submitted. However, its professional curation track implements more stringent formal gatekeeping, extending to data formats and documentation requirements. It encourages submission in specific statistical formats such as SAS and SPSS, and requires accompanying documentation such as codebooks, user guides, and README files.

Metadata requirements further reflect the governance of form, as qualified metadata is central to data sharing (Hansson & Dahlgren, 2021; Kim, 2021). All platforms impose metadata requirements, though their logic differs. Zenodo encourages rich metadata primarily for technical reasons, such as helping users avoid automated spam filters, while leaving responsibility for metadata quality with uploaders. Mendeley data mandates specific fields such as a title and at least one contributor, and although it reviews metadata for completeness, it strongly recommends that users verify accuracy themselves. Its requirements appear designed to ensure interoperability within Elsevier's commercial ecosystem, turning metadata into a bridge linking datasets with published articles, enhancing the asset value of both.

ICPSR's approach to metadata governance further exemplifies its dual-track strategy. Its Metadata Documentation Portal provides guidance for all depositors (ICPSR, 2025), but the two tracks diverge in implementation. Self-published depositors retain responsibility for metadata creation, whereas the professional curation track involves active review and enhancement by ICPSR staff to ensure quality and long-term usability. This active enhancement, rather than passive acceptance, positions ICPSR's professional curation service as a more interventionist gatekeeper even at the formal level.

Subtheme 1.2 strategic positioning on scientific substance

Zenodo and Mendeley data disclaim responsibility for scientific quality, positioning themselves as neutral intermediaries regarding data substance. Mendeley data states: *'We do not currently*

validate or curate the contents of valid research datasets,' relying instead on user reporting to flag non-compliant datasets. Zenodo echoes this, noting that content is provided *'as-is'* and that moderators *'are not domain experts, and thus cannot evaluate the scientific quality of content'*. It outsources this responsibility through its *'community'* feature, enabling domain experts to create curated areas. By delegating the power to judge data quality to researchers and the academic community, these platforms handle massive data volumes while avoiding the burden of substantive curation.

ICPSR, conversely, employs its dual-track model to offer intermediary and gatekeeper pathways. Through its self-publishing option, the platform mirrors Zenodo and Mendeley data, making data immediately available without curatorial review. However, ICPSR's professional curation track represents the most interventionist model. It defines curation as the process it uses to add value to data, maximize access, and ensure long-term preservation (ICPSR, 2022). This comprehensive intervention into data substance constitutes a clear expression of its gatekeeper role, extending beyond data review to active enhancement. ICPSR's curators actively enhance data to make them optimally usable, for instance, by recoding data to reduce disclosure risk and enforcing strict requirements on data content, such as mandating that for quantitative data, *'each variable in the data collection should have a set of exhaustive, mutually-exclusive codes.'* These curatorial interventions shape data substance by determining how research findings can be interpreted and reused, thereby demonstrating a substantial commitment to improving and vouching for the scientific usability of hosted data.

Theme 2 responsibility allocation: delegated liability versus internalized mission

This theme explores who is ultimately responsible, revealing how platforms navigate epistemic and archival responsibility. The analysis constructs a pattern of *'selective responsibility allocation'* whereby platforms delegate immediate legal liabilities to users while internalizing long-term archival stewardship.

Subtheme 2.1 delegation of legal liability to users

All platforms position themselves as legal intermediaries by contractually delegating responsibility for content legality to uploaders. Zenodo's Terms of Use are the most explicit: *'the uploader is exclusively responsible for the content that they upload... and shall indemnify and hold CERN free and harmless in connection with their use of the service'*. Mendeley data affirms that the user *'retain[s] complete control and copyright over the data'* and is responsible for ensuring they are not uploading copyrighted content they do not own. Even within ICPSR's curated track, the platform shifts initial ethical and legal burden to depositors. Acceptance is conditioned on the depositor's assurance that data are *'consistent with the study participants'* consent and relevant Institutional Review Board approval'. Furthermore, depositors are *'asked to review their data and documentation for information that could identify respondents'* prior to submission. This reinforces that the primary duty of confidentiality remains with the data producer.

This delegation extends beyond legal agreements to community-led governance. Zenodo distributes evaluative responsibility through its *'community'* feature, while Mendeley data empowers users to report a dataset that *'does not meet requirements or is in conflict with publication ethics.'* By assigning legal liability to uploaders and providing tools for community oversight, all platforms externalize content-related risks. This legal delegation simultaneously constitutes epistemic delegation: platforms disclaim not only liability but also the authority to judge content validity, shifting evaluative responsibility to depositors and the broader research community.

Subtheme 2.2 internalization of the archival mission

In contrast, these platforms internalize responsibility for being durable, trustworthy archives. This *'archival mission'* represents a different yet equally powerful form of gatekeeping. It is not a

gatekeeping of admission, which controls what is published, but a gatekeeping of permanence, which controls what remains part of the long-term scientific record.

Archival gatekeeping operates through several mechanisms. First, these platforms act as gatekeepers by assigning persistent DOIs, admitting datasets into citable academic outputs. Second, they guard record integrity against subsequent alteration, only allowing changes through a formal versioning process. For instance, ICPSR's metadata schema mandates a '*version*' number for all deposits, binding the dataset to a specific temporal state (ICPSR, 2025). Similarly, Zenodo's practice of leaving a '*tombstone page*' for withdrawn records further protects citation network integrity.

Yet archival responsibility reveals internal differentiation. Both Zenodo and Mendeley data distinguish between file preservation and functional preservation. While both commit to file preservation, they disclaim responsibility for functional preservation. Zenodo makes no promises regarding the usability and understandability of deposited objects over time, and its policy of never modifying original content reinforces this bounded responsibility. Mendeley data similarly positions archiving through Data Archiving and Network Services (DANS) as focused on storage rather than ongoing interpretability.

ICPSR's professional curation track, conversely, internalizes functional preservation through active data transformation. By converting data into multiple statistical formats including SAS, SPSS, Stata, and R, and creating standardized codebooks and comprehensive documentation, ICPSR guarantees not merely physical preservation but ongoing usability and interpretability. Its self-publishing track, however, mirrors Zenodo and Mendeley data's file-preservation-only approach.

Beyond these differences in preservation scope, all three platforms commit to long-term preservation: Zenodo promises retention for the '*lifetime of the repository*', Mendeley data relies on DANS, and ICPSR demonstrates this through CoreTrustSeal certification and encrypted off-site backups. These commitments position platforms as '*archival gatekeepers*' preserving the stability and integrity of scientific knowledge.

Discussion

Research data platforms navigate content responsibility through sophisticated strategies that challenge traditional intermediary-gatekeeper distinctions. Rather than occupying fixed positions on this spectrum, platforms strategically allocate different types of responsibility to manage the tension.

In defining their content responsibility, platforms operationalize a distinction between formal compliance and scientific quality. While platforms universally govern form to maintain operational order, they adopt divergent strategies regarding substance. Zenodo and Mendeley data strategically decouple these dimensions to avoid substantive evaluative burdens. ICPSR, conversely, incorporates both logics within its dual-track model: integrating substantive curation as a core mission in its professional track, while mirroring the decoupled, intermediary logic in its self-publishing service. This variable engagement with substance suggests that the gatekeeper role is not a fixed identity, but a scalable service layer.

Building on this definition, platforms employ '*selective responsibility allocation*' as a core strategy to manage the tension between neutrality claims and curatorial practices. While the form-substance boundary defines what they govern, this allocation strategy defines how they handle the resulting liabilities. Specifically, platforms delegate high-risk legal liabilities to users while internalizing the archival mission. This temporal division of labour allows platforms to claim intermediary status for immediate content risks while asserting gatekeeper authority over the scholarly record's long-term integrity.

As illustrated in Figure 1, this analysis reveals three distinct content governance models. Zenodo represents infrastructural intermediation, operating as a public good with low barriers to maximize access. ICPSR embodies disciplinary gatekeeping, yet its labour-intensive curation necessitates a club good model, introducing ‘*economic gatekeeping*’ where sustainability depends on membership fees. Mendeley data occupies a commercial hybrid position, mandating pre-publication review to ensure ecosystem interoperability. This positions it closer to the gatekeeper end regarding form while maintaining intermediary rhetoric on scientific substance. Finally, ICPSR’s dual-track model demonstrates directional flexibility: established gatekeepers can absorb intermediary functions to scale, whereas intermediary platforms face economic barriers to adopting substantive gatekeeping.

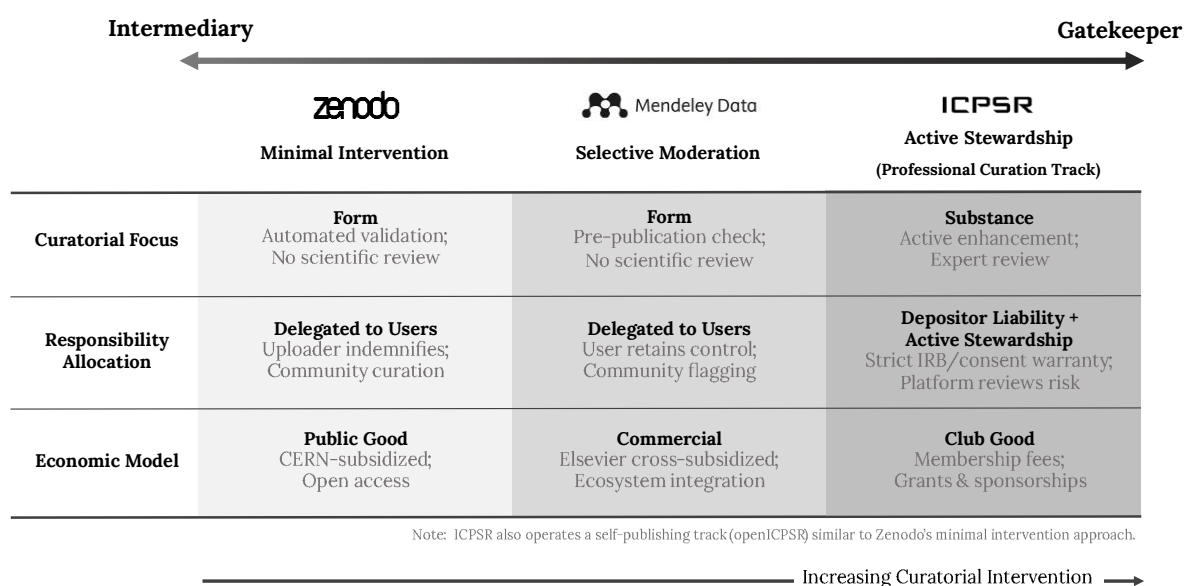


Figure 1. Defining content responsibility across the intermediary-gatekeeper spectrum.

These findings make several theoretical contributions. First, distinguishing epistemic from archival responsibility provides analytical precision absent from binary intermediary-gatekeeper models, revealing how platforms navigate governance tensions through multi-dimensional responsibility allocation. Second, identifying ‘*gatekeeping of permanence*’ extends the theoretical scope of gatekeeping beyond the traditional spatial metaphor of admission to include a temporal mechanism of preservation, capturing how platforms exert power over the scholarly record without substantive content review.

Additionally, for researchers, platform-mediated data sharing creates new dependencies, as governance mechanisms directly shape data discoverability and workflow integration. By exercising archival responsibility, platforms concentrate significant power over research output permanence and citability. Consequently, for open science policy, understanding content responsibility and its underlying economic models is crucial for assessing ecosystem sustainability.

Conclusion

This study advances content governance theory by demonstrating how scholarly infrastructure platforms transcend binary intermediary-gatekeeper distinctions through strategic multi-dimensional responsibility allocation. Research data platforms navigate content responsibility through boundary-drawing between form and substance, and between immediate liability and long-term stewardship. The analysis contributes a new theoretical framework distinguishing ‘epistemic responsibility’ from ‘archival responsibility.’

The study also proposes a typology of infrastructural intermediary, commercial hybrid, and disciplinary gatekeeper models that provides an analytical lens for understanding platform governance strategies across different operational contexts. Notably, the findings suggest a logic of '*directional flexibility*,' whereby gatekeeper platforms can adopt intermediary functions more readily than intermediary platforms can develop substantive gatekeeping capacities. These findings have practical implications for platform design and open science policy. As research data sharing shifts toward platform-mediated systems, understanding content responsibility becomes crucial for ensuring scholarly infrastructure integrity.

Future research should examine actual moderation practices beyond policy documents and develop contextual frameworks that help platforms calibrate governance strategies based on disciplinary needs, data sensitivity, and community expectations. Such work would support more responsive and accountable research data infrastructure.

Acknowledgements

The author thanks the anonymous reviewers for their constructive feedback.

About the author

Zhengzheng Tian is a PhD candidate in the Department of Smart Publishing and Digital Culture, School of Information Management, Wuhan University, Wuhan, China. Her research interests include data publishing and data governance. She can be contacted at 2016302480061@whu.edu.cn.

References

- Braun, V., & Clarke, V. (2019). Reflecting on reflexive thematic analysis. *Qualitative Research in Sport, Exercise and Health*, 11(4), 589-597. <https://doi.org/10.1080/2159676X.2019.1628806>
- Braun, V., & Clarke, V. (2024). Supporting best practice in reflexive thematic analysis reporting in Palliative Medicine: A review of published research and introduction to the Reflexive Thematic Analysis Reporting Guidelines (RTARG). *Palliative Medicine*, 38(6), 608-616. <https://doi.org/10.1177/02692163241234800>
- Busch, C. (2022). Regulating the expanding content moderation universe: A European perspective on infrastructure moderation. *UCLA Journal of Law and Technology*, 27(2), 32-79. <https://uclajolt.com/regulating-the-expanding-content-moderation-universe-a-european-perspective-on-infrastructure-moderation/> (Archived by the Internet Archive at <https://web.archive.org/web/20251227111350/https://uclajolt.com/regulating-the-expanding-content-moderation-universe-a-european-perspective-on-infrastructure-moderation/>)
- Celeste, E., Palladino, N., Redeker, D., & Yilma, K. (2023). The content governance dilemma. In D. L. Cogburn (Ed.), *The content governance dilemma: Digital constitutionalism, social media, and the search for a global standard* (pp. 7-25). Palgrave Macmillan. https://doi.org/10.1007/978-3-031-32924-1_2
- Flew, T., Martin, F., & Suzor, N. (2019). Internet regulation as media policy: Rethinking the question of digital communication platform governance. *Journal of Digital Media & Policy*, 10(1), 33-50. https://doi.org/10.1386/jdmp.10.1.33_1

- Frapporti, M. (2024). The politics of platforms: Exploring platforms' infrastructural role and power. In S. Mezzadra, N. Cuppini, M. Frapporti, & M. Pirone (Eds.), *Capitalism in the platform age* (pp. 109-126). Springer. https://doi.org/10.1007/978-3-031-49147-4_6
- Gillespie, T. (2018). Platforms are not intermediaries. *Georgetown Law Technology Review*, 2(2), 198-216. <https://georgetownlawtechreview.org/platforms-are-not-intermediaries/GLTR-07-2018/> (Archived by the Internet Archive at <https://web.archive.org/web/20251227105529/https://georgetownlawtechreview.org/platforms-are-not-intermediaries/GLTR-07-2018/>)
- Gorwa, R. (2019a). The platform governance triangle: Conceptualising the informal regulation of online content. *Internet Policy Review*, 8(2), 1-22. <https://doi.org/10.14763/2019.2.1407>
- Gorwa, R. (2019b). What is platform governance? *Information, Communication & Society*, 22(6), 854-871. <https://doi.org/10.1080/1369118X.2019.1573914>
- Gorwa, R. (2019c, October 23). The shifting definition of platform governance. Centre for International Governance Innovation. <https://www.cigionline.org/articles/shifting-definition-platform-governance/> (Archived by the Internet Archive at <https://web.archive.org/web/20250717191208/https://www.cigionline.org/articles/shifting-definition-platform-governance/>)
- Gorwa, R., Binns, R., & Katzenbach, C. (2020). Algorithmic content moderation: Technical and political challenges in the automation of platform governance. *Big Data & Society*, 7(1). <https://doi.org/10.1177/2053951719897945>
- Hansson, K., & Dahlgren, A. (2021). Open research data repositories: Practices, norms, and metadata for sharing images. *Journal of the Association for Information Science and Technology*, 73(2), 303-316. <https://doi.org/10.1002/asi.24571>
- Hesmondhalgh, D., Campos Valverde, R., Kaye, D. B. V., & Li, Z. (2023). Digital platforms and infrastructure in the realm of culture. *Media and Communication*, 11(2), 296-306. <https://doi.org/10.17645/mac.v11i2.6416>
- ICPSR. (2022, December 2). Data management & curation. <https://www.icpsr.umich.edu/pages/datamanagement/> (Archived by the Internet Archive at <https://web.archive.org/web/20251227104640/https://www.icpsr.umich.edu/pages/datamanagement/>)
- ICPSR. (2025, November 11). ICPSR metadata schema. https://icpsr.github.io/metadata/icpsr_study_schema/ (Archived by the Internet Archive at https://web.archive.org/web/20251227105017/https://icpsr.github.io/metadata/icpsr_study_schema/)
- Kim, Y. (2021). A study of the roles of metadata standard and data repository in science, technology, engineering, and mathematics researchers' data reuse. *Online Information Review*, 45(7), 1306-1321. <https://doi.org/10.1108/OIR-09-2020-0431>
- National Academy of Sciences, National Academy of Engineering, & Institute of Medicine. (2009). *Ensuring the integrity, accessibility, and stewardship of research data in the digital age*. National Academies Press. <https://doi.org/10.17226/12615>

Schöpfel, J., Prost, H., & Rebouillat, V. (2017). Research data in current research information systems. *Procedia Computer Science*, 106, 305-320.
<https://doi.org/10.1016/j.procs.2017.03.030>

Schreier, A. A., Wilson, K., & Resnik, D. (2006). Academic research record-keeping: Best practices for individuals, group leaders, and institutions. *Academic Medicine*, 81(1), 42-47.
<https://doi.org/10.1097/00001888-200601000-00010>

© [CC-BY-NC 4.0](#) The Author(s). For more information, see our [Open Access Policy](#).

Appendix

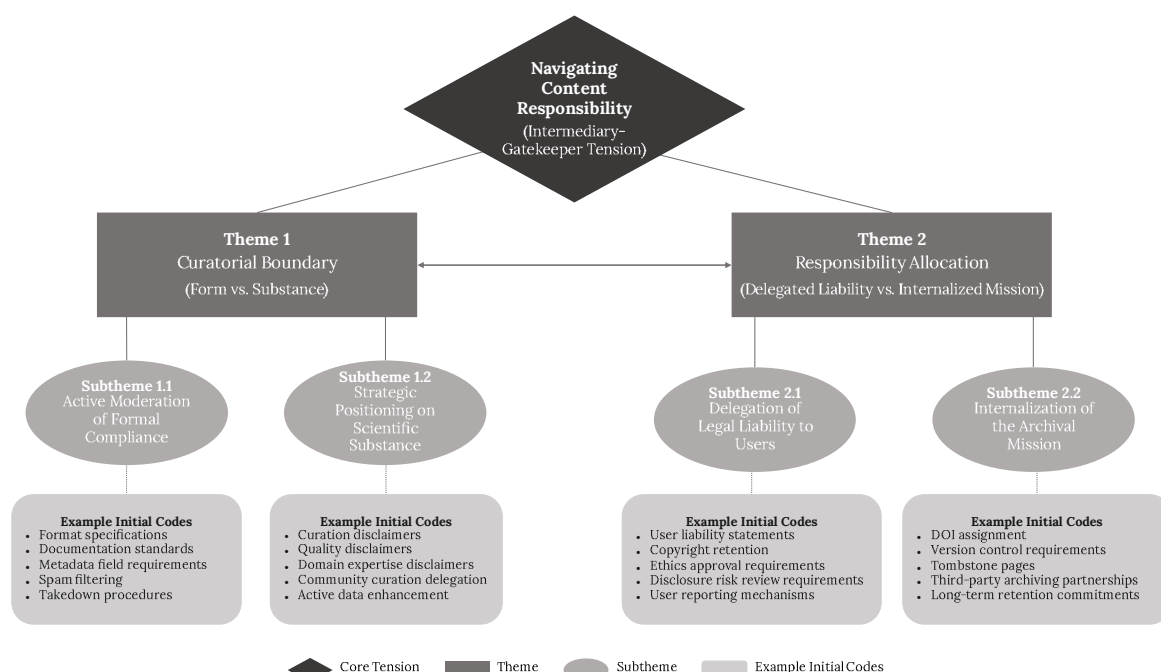


Figure A1. Thematic map of platform content responsibility navigation. It presents themes and subthemes as interpretive patterns of meaning, with example initial codes included illustratively to demonstrate analytic engagement rather than a linear or aggregative coding process.