Engaging Science, Technology, and Society

ORIGINAL RESEARCH ARTICLE

How Moderation Makes Science: Precautionary Valuation and Boundary-Making in the Early Circulation of Research

NATASCHA CHTENA
SIMON FRASER UNIVERSITY
CANADA

IRENE V. PASQUETTO
UNIVERSITY OF MARYLAND
USA

ALICE FLEERACKERS
UNIVERSITY OF AMSTERDAM
THE NETHERLANDS

STEPHEN PINFIELD
UNIVERSITY OF SHEFFIELD
UNITED KINGDOM &
RESEARCH ON RESEARCH INSTITUTE

MELANIE BENSON MARSHALL
UNIVERSITY OF SHEFFIELD
UNITED KINGDOM &
RESEARCH ON RESEARCH INSTITUTE

JUAN PABLO ALPERIN
SIMON FRASER UNIVERSITY
CANADA

Abstract

As preprint servers proliferate across scientific fields, moderation has emerged as a lightly structured form of gatekeeping that precedes—or bypasses—peer review. This article examines moderation as a site of precautionary valuation and boundary-making in the early circulation of research. Drawing on interviews with 14 individuals responsible for screening and oversight across 13 preprint servers, we analyze both the evaluative logics that organize moderation and the ways moderators themselves understand and frame their work. We find that credibility is enacted through precaution-driven threshold judgments—centering harm, trust, and community expectations—which determine whether a submission can responsibly circulate as scholarship. While moderators often reject the label of "gatekeeper," our findings show that their work performs a subtle but consequential filtering that shapes what and who enters the scholarly record—even as servers work to distance these decisions from the domain of peer review. By tracing how precautionary judgments and boundary-making structure decisions about what enters circulation at the preprint stage, we contribute to STS debates on valuation, peer review, and the governance of scholarly communication in open science.

To cite this article Chtena, Natascha, Irene Pasquetto, Alice Fleerackers, Stephen Pinfield, Melanie Benson Marshall, and Juan Pablo Alperin. 2026. "How Moderation Makes Science: Precautionary Valuation and Boundary-Making in the Early Circulation of Research." *Engaging Science, Technology, and Society* 12: e3187. URL

To email contact Natascha Chtena: Natascha.chtena@sfu.ca.

Keywords
preprints; moderation; open science; knowledge infrastructures; boundary work; valuation

Introduction
Preprints—or versions of research papers that have not yet undergone formal peer review—have become an increasingly prominent part of scholarly communication. Once confined to fields like physics and mathematics, preprints have gained traction in biology, medicine, and social sciences (Berg et al. 2016; Chiarelli et al. 2019). During the COVID-19 pandemic, they enabled rapid dissemination and public access to emerging scientific knowledge (Puebla, Polka, and Rieger 2021; Fleerackers et al. 2024). Today, preprints are integral to the publication pipeline in many disciplines and are increasingly incorporated into infrastructures for research evaluation, policy-making, and scholarly communication (Chiarelli et al. 2019; UNESCO 2021).

Despite their openness and speed, most preprint servers do not publish submissions without first subjecting them to some form of screening (Malički et al. 2020)—commonly referred to as moderation. This process is typically described as light-touch and administrative: a check for spam, plagiarism, overt pseudoscience, or whether the submission falls within the server's stated scope (Ginsparg 2021). Yet in practice, moderation can also involve more subjective judgments about what counts as credible science, what might cause harm, or what falls outside the bounds of a given research community. These decisions often go undocumented, and little is known about the epistemic, social, or institutional logics behind them.

Moderation also plays a key role in what preprints are expected to do. While publication in a peer-reviewed journal can take over a year in some disciplines (Christie et al. 2021; Powell 2016), preprint servers typically post articles within 24–48 hours. Bridging this time gap is often described as a core function of preprints (Ginsparg 2011; Wang, Chen, and Glänzel 2020). However, not all preprints are posted with the intention of journal publication; authors also use them to share letters, null result studies, or brief papers that may not fit within conventional journals (Maggio and Fleerackers 2023; Funk et al. 2020). Others view preprints as tools for reforming scholarly communication (Vianello 2021), offering opportunities for public commenting and alternative peer review models that might counteract biases in traditional review systems (Oliveira Henriques et al. 2023; Curry 2015). These varied functions make moderation a key—but under-examined—component of how credibility and visibility are negotiated in open science.

This article investigates how moderation is understood and practiced on preprint servers, focusing on the values, boundaries, and forms of judgment it entails. While often framed as a neutral or technical task (e.g., Ginsparg 2021), moderation operates as a form of infrastructural gatekeeping—performing boundary work in the absence of formal peer review. Drawing on interviews with 14 individuals involved in moderation across 13 preprint servers with varying disciplinary scopes, governance structures, and geographic locations, we examine how credibility is enacted through everyday decisions, criteria, and infrastructural design. We extend prior STS scholarship that has illuminated how moderation on arXiv helps maintain disciplinary boundaries (Delfanti 2015, 2016, 2017; Gunnarsdóttir 2005; Reyes-Galindo 2016) by offering a comparative account of how scientific legitimacy is negotiated across a broader range of preprint platforms. We situate this work within STS scholarship on boundary work and expert evaluation (Lamont 2009; Shapin and Schaffer 1985; Gieryn 1983), knowledge infrastructures (Bowker and Star 1999; Jasanoff 2004), and the sociology of valuation (Helgesson and Muniesa 2013; Lamont 2012). In doing so, we ask: *How do preprint*

servers define, structure, and enact moderation, and how do these practices contribute to broader processes of valuation and legitimacy-making in scholarly communication?

Situating Moderation within Scholarly Communication

Moderation as a Site of Epistemic Governance

Several large-scale studies have shown that preprint screening practices vary widely across servers. For example, Kirkham et al. (2020) examined 44 biomedical and medical preprint servers and found that most performed some screening prior to posting, but only a minority involved researchers with subject matter expertise. Similarly, analyses by Malički et al. (2020) and Hamade et al. (2022) found that while screening is widespread across preprint servers, guidance on research integrity and transparency is limited, and checks for scientific misconduct rarely extend beyond plagiarism. These studies, however, focus on written policy, which is often vague and inconsistently implemented; they tell us little about how moderation is organized, interpreted, and practiced on the ground.

STS scholars have addressed this gap between official policies and lived practices. Digital governance and technoscience researcher Kristrún Gunnarsdóttir (2005) shows how arXiv's semiautomated moderation process serves as a form of distributed gatekeeping that protects scientific relevance by filtering out non-certified actors. Sociologist Luis Reyes-Galindo (2016) further demonstrates how arXiv's classification tools and endorsement protocols translate tacit community norms into computational heuristics. Related work examines how preprint infrastructures mediate authorship, visibility, and scientific temporality. Digital labor scholar Alessandro Delfanti (2016) shows how physicists use arXiv to manage temporal competition, coordinate priority, and perform reputational labor. Additionally, his study of viXra (Delfanti 2017) illustrates how actors who feel excluded by arXiv mobilize alternative infrastructures to contest disciplinary hierarchies. Together, this work highlights how the design and governance of preprint platforms redistribute visibility and shape who is recognized as a legitimate contributor.

Our study expands this body of work by analyzing moderation practices across thirteen preprint servers spanning diverse disciplines and organizational models. While earlier research tends to focus on written policies or a single disciplinary case, our interview-based approach captures how moderation is conceptualized and enacted behind the scenes—often informally, and under conditions of considerable ambiguity. We show how moderation decisions unfold across platforms, geographies, and epistemic cultures.

Understanding how moderation is enacted—how decisions are made about what counts as acceptable science, what poses harm, and what serves a given community—offers insight into the evolving politics of openness in contemporary scholarly communication and into how science is positioned in relation to the public, particularly in information environments where scientific knowledge can be misrepresented (Swire-Thompson and Lazer 2020; Pereira and de Oliveira 2024).

Valuation and Epistemic Discretion in Scholarly Judgment

Our analysis is also informed by the sociology of valuation, which examines how worth is ascribed, justified, and enacted through social practices and classificatory systems (Lamont 2012; Helgesson and Muniesa 2013). Valuation scholars have shown that acts of evaluation are rarely standardized; instead, they are shaped by moral and symbolic frameworks, institutional constraints, and the situated judgments of actors.

Peer review, in particular, has been analyzed as a site where such logics of worth are negotiated. Cultural anthropologist Michèle Lamont's (2009) comparative study of multidisciplinary review panels illustrates how reviewers draw on tacit disciplinary sensibilities—rhetorical style, methodological conventions, perceived seriousness of purpose—alongside explicit criteria like rigor or clarity. Her work also illustrates how notions like "excellence" and "significance" are differentially understood across epistemic cultures, and how evaluative practices are shaped by discipline-specific repertoires, emotional investments, and fairness norms. Similarly, sociologists David Pontille and Didier Torny (2014) show how peer review operates as a moral economy in which evaluations of legitimacy intertwine with judgments about decorum, genre, and authorial positioning. Later work (Pontille and Torny 2017) shows how evaluative routines continue even after publication, suggesting that the status of knowledge is not fixed at the point of acceptance, but continually reshaped.

This body of work shows that while peer review is organized through formal procedures, evaluative decisions remain interpretive and context-dependent. Acts of valuation are rooted in disciplinary and institutional cultures, sustained by moral and symbolic repertoires, and enacted through situated practices of assessing, ranking, and justifying worth. We draw on this framework to examine moderation as a potentially analogous site of valuation. While peer review has been widely studied as a mechanism of academic judgment, less attention has been paid to how value is assigned in spaces that operate without formal reviewers or standardized protocols. Valuation theory offers a lens for understanding how such informal, low-visibility assessments contribute to the broader project of epistemic and institutional legitimacy.

Although valuation theory helps us understand how worth is enacted through situated judgments, it does not fully capture how these judgments are also boundary-producing. To explore how credibility assessments shape what counts as science, we turn to STS scholarship on boundary work.

Boundary Work and the Politics of Scientific Legitimacy

Philosopher and sociologist of science Thomas F. Gieryn (1983) first described boundary work as the discursive strategies scientists use to separate legitimate science from pseudoscience or "non-science," often to defend their professional authority. This seminal work also described how boundaries and the strategies used to draw them are fluid, with the qualities ascribed to "science" changing depending on what institution or knowledge system it is contrasted with. For instance, science has been portrayed as "practically useful" to distinguish it from religion, but as theoretical when contrasted to mechanics: a pursuit of knowledge even in the absence of technical (i.e., practical) applications. Later work extended the concept beyond discursive strategies to include the material and procedural tools through which boundaries are produced during everyday scientific work (Gieryn 1999; Jasanoff 2004; Star and Griesemer 1989).

Building on boundary-work theory, sociologist of science Graham Howard (2012) examines peer review as a socially and historically contingent practice through which epistemic authority and value are produced. Rather than treating peer review as a neutral or standardized procedure, he shows that its meaning and function vary across disciplines, geographies, and institutional contexts. These variations reveal that peer review operates less as an objective indicator of scholarly quality than as a symbolic resource for asserting legitimacy and maintaining hierarchies of credibility within particular epistemic cultures.

Reyes-Galindo (2016) extends boundary-work theory into the digital infrastructures of scholarly communication, examining how arXiv performs epistemic demarcation through algorithmic and human moderation. He shows that arXiv's automated filters and endorsement systems put tacit community norms into practice, capturing who "writes like a physicist," uses the "right" language, and is seen as belonging. These classifiers translate collective judgments about credibility and style into computational heuristics that flag submissions that deviate from disciplinary expectations for human review. Reyes-Galindo shows that automation and human discretion are co-constitutive; algorithms codify shared standards of legitimacy, while moderators draw on these outputs to calibrate their assessments. The result is an aesthetic and linguistic boundary that reproduces disciplinary hierarchies without explicit exclusion.

Taken together, the sociology of valuation and STS research on boundary work offer a useful lens for examining moderation as a situated form of epistemic and institutional judgment. Instead of asking whether moderation is "effective," we treat it as a site where norms from journal-based publishing are selectively inherited, reconfigured, or quietly adapted to fit the temporal, institutional, and infrastructural conditions of preprint dissemination. We approach these dynamics through the following research questions:

1. How do servers articulate the role and scope of moderation and how do these framings position preprints within the broader ecosystem of scholarly communication?
2. How is moderation enacted in the assessment of submissions, and what forms of evaluative and classificatory work does this entail?
3. How is moderation organized as an institutional practice, and how are expertise, discretion, and authority configured across different moderation arrangements?

Methods

Sample

We used a purposive sampling approach, guided by principles of maximum variation. While the sample does not encompass all existing servers—estimated at somewhere between 60 and 100 (Kirkham et al. 2020; COAR and CCSD n.d.)—it was assembled to capture the heterogeneity of preprint infrastructures and institutional imaginaries. In line with interpretive traditions, sampling was conceived as an analytic rather than a procedural process to make visible the organizational, epistemic, and infrastructural arrangements through which moderation is practiced and justified.

We selected servers that varied along several intersecting dimensions: governance model (community-led, publisher-led, cooperative-/consortium-led, independent, or government-operated), disciplinary scope (single-discipline, multidisciplinary), geographic location (Americas, Africa, and Asia), infrastructural configuration (open-source versus proprietary; differing workflows and degrees of formalization in moderation), and institutional age (from long-established servers to recent entrants). We also aimed to balance scale and visibility, including highly influential platforms that shape the broader discourse on preprints (e.g., bioRxiv, arXiv) alongside smaller, community-led initiatives with fewer resources (e.g., MediArXiv, RINarxiv). This allowed us to consider how moderation practices differ across institutional forms and varying levels of infrastructural maturity and public visibility.

We excluded funder-led open research platforms hosting preprints (e.g., Wellcome Open Research, Gates Open Research), as they function primarily as post-publication review venues, rather than preprint dissemination servers.

The final configuration was also shaped by prior studies (e.g., Chaleplioglou and Koulouris 2023; Chiarelli et al. 2019; Balaji and Dhanamjaya 2019) and pragmatic considerations, such as availability and willingness of staff. Table 1 summarizes the key characteristics of the preprint servers included in the study (see also Chtena et al. 2025b).

Table 1. Characteristics of preprint servers studied. Source: author's own.

	Server	Year Established	Country	Discipline	Host/ Platform	Governance Model	Ownership
1.	AfricArXiv	2018	Pan-African	Multi-disciplinary	UbuntuNet Alliance (AfricArXiv.ubuntunet.net) + legacy multi-platform (OSF, Zenodo, etc.)	Community-led with institutional technical hosting (UbuntuNet)	Community-owned, non-profit
2.	arXiv	1991	United States of America (USA)	Mathematics, information science, quantitative biology, quantitative finance, electrical engineering and systems, science, physics, and economics	Cornell University Library; proprietary platform	Institutionally governed (research institution) with community participation	Institutionally owned (non-profit service under Cornell's administrative umbrella)
3.	bioRxiv	2013	USA	Biology and life science/Clinical research	OpenRxiv; proprietary platform	Independent non-profit	Independent non-profit
4.	ChemRxiv	2017	USA	Chemistry	Cambridge Open Engage (Cambridge University Press)	Society consortium-governed	Co-owned and collaboratively managed by five chemical societies; non-profit
5.	EarthArXiv	2017	USA	Earth science	California Digital Library, eScholarship platform	Community-led (governance by volunteer council)	Community-owned; institutionally hosted
6.	Jxiv	2022	Japan	Multi-disciplinary	Japan Science and Technology Agency; Open Preprint Systems	Institutionally governed (government agency)	Institutionally owned; government-funded; non-profit
7.	MediArXiv	2019	USA	Media, film, and communication studies	Open Science Framework operated by the Center for Open Science	Community-led and volunteer-run (governance by volunteer Steering Committee)	Community-owned; non-profit
8.	medRxiv	2019	USA	Biology and life science/ Clinical research	OpenRxiv; proprietary platform	Independent non-profit	Independent non-profit

9.	Research Square	2013	USA	Multi-disciplinary	Research Square Company (part of Springer Nature Group); proprietary platform	Corporately governed (for-profit subsidiary governance)	Commercially owned; publisher-affiliated
10.	RINarxiv	2017	Indonesia	Multi-disciplinary	BRIN (National Research and Innovation Agency of Indonesia), using Open Preprint Systems	Community-led and government-hosted	Community-owned (Indonesian Open Science Team); non-profit
11.	SciELO Preprints	2020	Brazil	Multi-disciplinary	Open Preprint Systems, operated by the SciELO Program	Governed via a publishing cooperative model within the SciELO Program	Institutionally hosted (FAPESP/BIREME) under a cooperative publishing program; non-profit
12.	SocArXiv	2016	USA	Social and behavioral sciences	University of Maryland Libraries (on OSF, Center for Open Science)	Community-led and volunteer-run (governance by a volunteer Steering Committee)	Institutionally owned; non-profit
13.	SSRN	1994	USA	Multi-disciplinary	Elsevier; proprietary platform	Corporately governed (publisher-operated)	Commercially owned; publisher-affiliated

Preprint Server Typologies

To describe variation across our sample, we employ organizational typologies that capture how authority, decision-making, and infrastructural responsibility are distributed across preprint servers. We use the term *community-led* to describe servers operated and managed directly by volunteer collectives (e.g., MediArXiv, SocArXiv, RINarxiv), as well as arXiv, which represents a distinctive case—institutionally hosted and operated with paid staff, but shaped by strong community participation in governance and policy development. *Publisher-led* servers (e.g., SSRN, Research Square) are owned and directed by commercial publishers. *Publisher-adjacent* refers to servers governed by scholarly publishing organizations or cooperative publishing programs (e.g., ChemRxiv, SciELO Preprints), reflecting their close ties to the publishing ecosystem without being directly commercial. *Independent* servers are publisher-independent and operated by stand-alone non-profit entities (e.g., bioRxiv and medRxiv). Lastly, *government-operated* servers, such as Jxiv, represent a rare model in which the platform is fully managed by a state agency.

Data Collection

We interviewed 14 individuals involved in content moderation or oversight across 13 preprint servers between February and April 2023. Interviews were conducted by the first two authors, following research ethics approval (#30001471) by the SFU Research Ethics Board and participant consent.

Participants were contacted via email using contact information found through public online sources (e.g., server websites, blogs, reports, slide decks, etc.). Each participant chose from three options for how their identity would be handled: (1) be named as a participant, and have their name linked to individual quotations; (2) be named, but not have individual quotations linked to them by name; and (3) remain

anonymous outside the research team. Ten participants chose full attribution, three chose partial attribution, and none opted for full anonymity. We respect these preferences in our presentation of the findings and refrain from identifying participants when doing so might expose them to professional risk. For the four participants who opted not to have individual quotations linked to them by name, we use pseudonymous codes (i.e., P01, P02).

Interviews explored how moderation is organized and justified—what steps are involved, who participates, and how decisions are made. We stopped conducting interviews once we felt we had gathered enough insight to answer our research questions and were no longer encountering significantly new perspectives (Vasileiou et al. 2018). Interview data were triangulated with data from secondary sources, such as preprint servers' websites, GitHub pages, and publicly available moderation checklists. For a full list of interviewees see table 2. Our interview guide is available via Open Science Framework (OSF) (Chtena et al. 2025c).

Table 2. List of preprint servers and server representatives[1] included in the study. Source: author's own.

	Server	Country	Representative Name	Representative Role
1.	AfricArXiv	Pan-African	Nicholas Outa	Head of Submissions Moderation
2.	arXiv	United States of America (USA)	Jim Entwood	Head of Content and User Support
3.	bioRxiv/ medRxiv	USA	John Inglis	Co-founder and Principal Investigator
4.	bioRxiv/ medRxiv	USA	Samantha Hindle	Content Manager
5.	ChemRxiv	USA	Ben Mudrak	Senior Product Manager
6.	EarthArXiv	USA	Bruce Caron	Co-founder
7.	Jxiv	Japan	Soichi Kubota	Manager
8.	MediArxiv	USA	Jeff Pooley	Co-founder and Co-lead
9.	Research Square	USA	Amye Kenall	Vice President, Product and Publishing
10.	Research Square	USA	Roma Konecky	Associate Product Manager
11.	RINarxiv	Indonesia	Dasapta Erwin Irawan	Manager
12	SciELO Preprints	Brazil	Alex Mendonça	Online Submission and Preprints Coordinator

[1] Roles reflect participants' positions at the time of data collection. Some individuals have since moved to different roles or institutions.

13.	SocArXiv	USA	Philip N. Cohen	Founding Director
14.	SSRN	USA	Gregg Gordon	Managing Director

Data Interpretation

All interviews were conducted over Zoom (lasting 60-90 minutes) and recorded for later transcription and analysis. The interviews were iteratively analyzed by the lead author using a hybrid inductive and deductive approach (Fereday and Muir-Cochrane 2006; Swain 2018). Coding was done in NVivo 12 and in a separate coding matrix, where relevant statements from the interviews were charted into a matrix with coding labels (Rosen et al. 2023) and supplemented with data extracted from secondary sources. Throughout the process, the team met regularly to discuss interpretations, refine themes, and reflect on our own assumptions. Our final codebook is available via OSF (Chtena et al. 2025d).

Findings

The ensuing sections unfold as follows. The first section examines how servers articulate the role and purpose of moderation, showing how ideas of harm avoidance, credibility, and trust shape how it is understood and communicated. The second section turns to how these purposes are enacted in practice—how moderators interpret, classify, and sometimes exclude submissions using checklists, heuristics, and tacit cues of scientific legitimacy. The third section analyzes the organizational dimensions of moderation, showing how authority and expertise are distributed across different governance models and infrastructural arrangements, and how complex cases are deliberated. Finally, the fourth section analyzes how servers position moderation in relation to journal peer review, highlighting the tension between rhetorically disavowing evaluation and the practical blurring that occurs when moderation requires discretionary judgment.

Articulating Purpose: Why Servers Moderate

During interviews, participants described moderation using terms such as *filtering*, *screening*, or *safeguarding*. Rather than evaluating the quality of a paper (for example, by assessing the appropriateness of a sampling strategy or statistical test), moderation was framed as determining whether a submission fits the genre of "science," aligns with the mission of a server, and can be circulated responsibly.

Specifically, participants described moderation as performing a selective sorting that excludes content that is irrelevant, misleading, or potentially harmful. These exclusions were not framed as censorship or judgment, but as necessary to maintain the server as a legitimate space for scholarly communication. In other words, moderation helps define what is seen as relevant, useful, and acceptable—and by doing so, helps determine what counts as science/scholarship.

Some participants also described moderation as a way to guard misuse of preprint servers, especially in contexts where metrics and institutional expectations shape how researchers engage with these platforms. For example, RINarxiv's Manager, Dasapta Erwin Irawan, shared a case where student assignments were submitted as faculty-authored preprints. This highlights how moderation sometimes requires responding to local research environments:

> It is related to the citation mass we have in the region, and also the productivity notion in the regulation that we have to publish something each semester if we are lecturers. [Researchers] try to make it simple by the existence of our server. We don't apply peer reviewing in the preprint server, so it can be very easy to post something. During our lifetime managing the server we get a rush of submissions that are really student assignments. They send the document to our server, including their lecturer's name as author.

Even though servers like RINarxiv do not strictly forbid student work, most expect submissions to contribute—or at least claim to contribute—something original to scientific knowledge. In such cases, moderation works like a kind of institutional triage—deciding what fits the server's mission, and what might weaken its credibility.

For medical and health-focused servers like bioRxiv and medRxiv, moderation was strongly linked to the "do no harm" principle of clinical research ethics and medical publishing (NCPHSBBR 1979). Here, the line between science and non-science involved questions about risk—especially how journalists or the public might interpret unreviewed findings. As co-founder of bioRxiv and medRxiv John Inglis explained:

> We are not making any judgments about the conclusions that the manuscript comes to, or the quality, the expertise and the knowledge that have gone into the work. But we can make a judgment about the consequences of posting it as a preprint without the intervention of properly qualified experts. And there are times when we think, "No, we're not going to. . . . This is better distributed after peer review."

Similar concerns were echoed by other servers handling content with public health and safety implications, including ChemRxiv, Research Square, and SSRN. Here, valuation hinges on anticipated consequences—what matters is not whether findings are correct, but whether premature circulation could cause harm, mislead publics, or erode institutional trust.

Representatives from the two publisher-owned servers in our sample (Research Square and SSRN) also emphasized moderation as a means of liability protection—against copyright infringement or the circulation of offensive, hateful, harassing, or otherwise objectionable material. Much like the do no harm principle, liability protection reflects a risk management logic that mitigates plausible harms through safeguards and responsibility boundaries.

Moderation was also described as a means of establishing credibility or protecting the reputation of servers themselves. This was especially true for newer servers seeking to establish trust and legitimacy with their audiences. As Alex Mendonça from SciELO Preprints explained:

> We are very careful because we're just starting, and we want to start with a good reputation. . . . We also want people to see that having your paper in SciELO Preprints has some credibility to it.

Relatedly, Mendonça noted that moderation also serves to avoid reinforcing the perception that preprints are low-quality or inferior to journal articles. As the following excerpt illustrates, concerns about the broader reputation of preprints intertwine with efforts to protect the SciELO brand:

> Some people . . . think that preprints are not good research, or they're rubbish, and we don't want to contribute to that. . . . We don't want to have [SciELO] associated with a server that doesn't move science forward. . . . We are very careful.

This concern about reputation was echoed by Inglis, who explained that building trust was central to the design of both servers—one reason why their moderation process includes screening by independent experts. Even basic screening, according to Inglis, is meant to create confidence:

> What we were trying to do, always, was give a sense of trust, give a sense of responsibility that we were doing things in a way that was not going to undermine the scientific process [and that] was not going to cause harm to human beings.

Across these accounts, participants drew on vocabularies of care, responsibility, and precaution that position moderation as a practice of maintaining the credibility of preprints and the integrity of the infrastructures through which they circulate. Moderation in this sense can be understood as a form of infrastructural care—an ongoing, behind-the-scenes labor that sustains the material and social conditions that make preprints a viable and legitimate mode of scholarly communication. Seen through a valuation lens, this care work involves assessing when the release of unreviewed research is responsible, appropriate, and institutionally safe—decisions that quietly shape which forms of knowledge servers are willing to stand behind.

Sorting Science: How Moderators Decide What Counts

While the previous section examined how servers describe the moral, epistemic, and institutional purposes of moderation, this section turns to how those purposes are enacted in practice. Moderation here operates as a form of *precautionary valuation* through which ideas of credibility, scientificity, and relevance are interpreted and applied in context, as well as a form of *boundary work*, through which the contours of credible science are negotiated and maintained.

These forms of judgment are made concrete through moderation workflows and the tools that support them—notably moderation checklists that codify expectations of scope, relevance, and scientific form.

Moderation Checklists as Valuation Devices

Drawing on interviews and public documentation, we identified a recurring set of criteria that servers draw upon when deciding whether to post a preprint. While no single checklist is shared, most servers screen submissions with reference to some combination of the following considerations:

1. **Scope/relevance**: The article is written in a language accepted by the server (e.g., English, French), fits within its stated subject/geographic coverage, and is of anticipated interest to its target readership.
2. **Scholarly/scientific nature**: The article constitutes "scholarly content," i.e., is presented in a scholarly format and part of the scholarly discourse in its subject area.
3. **Reasonableness**: Claims made appear sensible; those that challenge "common sense" or prevailing understandings are supported by evidence (e.g., openly available data).
4. **(Minimally) informative**: The article meets basic standards of expression and presentation and provides at least some informative value.

5. **Technical compliance and completeness**: The article complies with the server's formatting and file requirements, submission metadata is complete, and required supplemental files have been included.
6. **Reporting standards**: The article includes information on ethical oversight, conflicts of interest, and clinical trial registration; no patient-identifying information is included.
7. **Risk and safety**: The article does not risk public safety or the health of individual patients.
8. **Author credentials or endorsement**: At least one of the authors must be affiliated with a recognized research institution and have a publication record in the field or be endorsed by a member of the relevant scholarly community.

In addition, most servers also perform plagiarism and copyright checks, either manually or using automated tools. While these moderation criteria are found across servers, their weight and interpretation vary substantially. These differences reflect disciplinary risk profiles (e.g., public health vs. media studies), available expertise and time, and organizational philosophies about openness and harm-avoidance. For example, publisher-led servers and those working in high-stakes health domains (e.g., medRxiv, Research Square) apply stricter checks related to ethics, clinical trial registration, and risk of public misunderstanding, while community-led servers are selective in context-specific ways. AfricArXiv, for instance, uses light-touch checks focused on geographic inclusion—whether the work is produced in, about, or relevant to Africa—and welcomes a broad range of formats and early-stage contributions, reflecting a mission of inclusivity and capacity-building rather than adherence to strict scientific "form."

However, even in cases where servers draw on seemingly identical criteria, the threshold for what satisfies that category is interpretive and context-specific. As Lamont (2009) notes, "evaluation is not based on stable comparables, and ... various competing criteria with multiple meanings are used to assess academic work" (18). Thus, while moderation checklists project an impression of procedural neutrality, they do not mechanize decision-making. Checklists structure evaluation by naming shared categories, but what counts as "scientific," "reasonable," or "in scope" hinges on tacit disciplinary sensibilities rather than fixed definitions.

Checklists, then, do not settle whether submissions satisfy the listed criteria; they designate the sites where judgment must occur. In this sense, checklists operate as devices through which precautionary valuation is enacted. They codify institutional commitments such as credibility and responsibility, but their application depends on moderators' situated, interpretive work.

Demarcating Belonging: Scientific Content and Credible Contributors

While all moderation criteria require situated judgment and interpretive flexibility, determining whether a submission is "scientific" or "scholarly" is where discretion becomes most visible—and where boundary work most clearly takes shape. In deciding whether a manuscript can count as science, moderators perform both valuation (assessing credibility and epistemic fit) and boundary work (marking the limits of legitimate scholarship), often through tacit criteria. Because these judgments determine whether a piece of writing enters the scholarly record, their unstated nature becomes especially consequential. Participants often described this process as "fuzzy" or "intuitive," shaped more by feel than fixed rules: "It's got a little bit of a feeling to it, a little bit of thinking to it, but you can kind of tell, right? I mean, does it look like a scientific paper? Does it feel like a scientific paper?" (P01). This admission of intuition stands in tension with how

servers publicly present moderation—as a neutral, checklist-driven procedure—highlighting the gap between the discursive ideal of procedural screening and the interpretive labor that actually sustains it.

In practice, this interpretive labor often hinged on stylistic and rhetorical cues, wherein format, tone, and writing style served as proxies for legitimacy—suggesting that what "looks" or "sounds" like science often passes as science. Several moderators, for example, emphasized considering scholarly presentation (e.g., structured methods) to determine what counts as research. As one participant noted, "basically, if it has structured methods, a motivated, reasonable research question, and presents the results, then we will consider that 'research' and we will post it." (Roma Konecky, Research Square).

Disciplinary fluency also played a role in determining the bounds of legitimate scholarship, as familiarity with a field's vocabularies, tools, and norms were central to whether a submission could be considered scholarly:

> Mostly [moderators are] not looking to see, did they do their experiments right? Did they have the right controls? We can't police [that] or evaluate with that sort of rigor. But they can give everything an actual read, and a lot of it is based on [moderators'] extensive experience in the field and saying, "Yep, may not be right." (P02).

In some servers, questions of credibility and belonging extended to *who* was submitting rather than *what* was submitted. Much like manuscript-level assessments, author screenings function as a form of boundary work (Reyes-Galindo 2016), where affiliation, publication history, and even personal correspondence become proxies for epistemic credibility. These checks reinforce institutionalized boundaries of participation—filtering not just ideas, but the actors deemed legitimate to circulate them.

arXiv, for example, uses an endorsement system to verify that a submitting author "belongs to the scientific community" (ArXiv n.d.). Authors from known institutions typically receive automatic endorsement, while those without institutional email addresses must be endorsed by a peer active on arXiv. Endorsers must have recently posted a certain number of preprints in the relevant subject area and must either know the author or have read the submission, though detailed reading or verifying correctness is not required. This system, while scalable, enacts a form of credential-gating that privileges known actors and recognized institutions:

> We don't have the capacity to open up for everyone in the world to hit a submit button without passing through any kind of gatekeeping, for lack of a better word there. The endorsement system basically determines who can submit to arXiv. (Jim Entwood, arXiv).

SciELO Preprints, on the other hand, implements a more bureaucratic model, using authors' past publication record as a verification tool that expedites manuscript processing. To bypass moderation by a SciELO expert, authors submitting original research must have at least two articles indexed in Scopus or SciELO, or five articles with Digital Object Identifiers (DOIs) indexed in Google Scholar. Interestingly, authors submitting other manuscript types (e.g., reviews, essays, opinion articles) have to meet a higher bar: at least five articles with DOIs indexed in SciELO or Scopus, or 10 articles with DOIs indexed in Google Scholar. SciELO Preprints also offers an endorsement option for authors who do not meet these criteria, though endorsement is not required for acceptance. Endorsements are manually checked by staff to verify the endorser's expertise, reflecting an additional layer of credential-based oversight: "we check where the endorsement is coming

from because it might come from someone who is not even an expert in the area" (Alex Mendonça, SciELO Preprints).

In contrast, RINarxiv takes a highly relational approach. As Dasapta Erwin Irawan explained, moderation involves "direct communication for each submission just to get enough story so we don't publish . . . non-eligible documents." If the author fails to respond to the email inquiry, a submission is deemed untrustworthy and automatically deleted, making responsiveness itself a proxy for legitimacy.

RINarxiv's approach reflects the local Indonesian research context, where awareness of preprints remains low (Irawan, Zahroh, and Puebla 2022). The server's age likely also plays a role; launched in May 2020, it averages fewer than 20 submissions per month as of this writing (compared to arXiv's 20,000). It also depends on the moderator's ability to connect with authors and assess their "story," a task that requires local knowledge and cultural sensitivity. As such, the approach is neither scalable nor easily transferable.

Lastly, some servers, like SocArXiv, apply author screening in a more selective manner—making use of institutional affiliation or prior publications only in ambiguous cases:

> I think one thing we would consider is the status of the researcher, not as a rule, but as one factor among many. Undergraduate student versus a professor at a university would be a difference. . . . We do use some reputational things to add into the mix if we're already on the fence about something. (Philip Cohen, SocArXiv).

These accounts expose moderation as a layered system of epistemic filtering that reaches beyond the manuscript alone, where *who* is permitted to contribute becomes a proxy for *what* is considered credible knowledge, a dynamic well-documented in STS literature on expertise and legitimacy (Collins and Evans 2007; Shapin 1995).

While the use of credential-based heuristics is efficient, it raises questions about consistency, bias, and equity—similar to those raised in critiques of peer review itself (Huber et al. 2022). These concerns were not acknowledged by participants, perhaps because preprint servers are designed with lower barriers to entry than journals and thus seen as less vulnerable to the gatekeeping excesses of formal peer review. Yet the reliance on stylistic and reputational cues shows how moderation performs boundary work at the intersection of epistemic and social judgment. What "counts" as scientific is not only a matter of method or argument but also of presentation, tone, and perceived authorial credibility.

In this sense, moderation reproduces hierarchies of legitimacy even as it aims to lower barriers to participation. This dynamic resonates with Delfanti's (2016) analysis of open science infrastructures, which shows how participatory systems reconfigure, rather than eliminate, hierarchies of expertise. Preprint moderation embodies this paradox of openness: it extends access and visibility while reinstating evaluative mechanisms that distinguish legitimate contributors from those deemed peripheral.

A detailed description of the heuristics utilized to demarcate "real" science from non-science or pseudoscience can be found in table B1, appendix B (Chtena et al. 2025e).

Organizing Moderation: Authority, Expertise, and Institutional Form

Moderation does not operate in a vacuum. It depends on specific forms of expertise, organizational norms, and the infrastructures available to support them. Across our interviews, participants described who is

allowed—or expected—to perform moderation, offering insight into how discretion and epistemic authority are distributed across different server types and models.

Community-led servers such as EarthArXiv and MediArXiv emphasized the value of having moderators embedded in relevant scholarly communities. Across these initiatives—whether institutionally hosted or volunteer-run—moderation was framed as a form of epistemic labor grounded in disciplinary belonging. These moderators were seen as more attuned to disciplinary norms and better equipped to recognize what belongs. As Jeff Pooley (MediArXiv) put it, "the moderating task in some ways is best done by specialists in the field." Moderators are expected to identify relevance, uphold informal standards, and help protect the boundary of what "counts" as appropriate content. Nicholas Outa of AfricArXiv further underscored the importance of moderators having experience and knowledge of the scholarly publishing landscape:

> We mostly encourage people who have published, who are also very . . . aware of, you know, the ethics of publishing, those who understand how scientific communication is done. . . . Because somebody could be a professional, but when it comes to scientific publications, it needs somebody who has really published.

Outa did not explicitly mention early-career researchers or those without a publishing record, but his comments suggest they may not be seen as ideal moderators. In this framing, moderation relies on accumulated scholarly capital—knowing how to interpret conventions, and how to judge from within a given research culture. This aligns with sociological accounts of expertise as a situated, relational status rather than a purely technical skill (Collins and Evans 2007), and with STS scholarship that highlights how epistemic authority is linked to embeddedness within disciplinary communities (Jasanoff 2004; Lamont 2009). Moderators are not just evaluators—they are insiders whose judgments are informed by tacit knowledge and symbolic positioning within the field.

Compared to the publisher-led and publisher-adjacent servers in our sample, community-led servers were less likely to offer structured training to moderators. While participants were not asked explicitly why they do not train moderators, responses suggest lack of time and resources likely plays a role. Community-led servers almost exclusively employ volunteers, although some (e.g., AfricArXiv) provide small stipends to incentivize moderators and ensure manuscripts are processed in a timely manner. Most community servers assign a single moderator to each manuscript, unless a submission is escalated (e.g., due to racial undertones or potential for public harm).

Publisher-led (e.g., SSRN), publisher-adjacent (e.g., ChemRxiv), government-operated (e.g., Jxiv), and independent (e.g., bioRxiv) preprint servers, by contrast, follow more formalized workflows. These servers typically use a two-pronged moderation process, where in-house screening staff first performs administrative checks and PhD-level experts perform a second check more narrowly focused on a preprint's content. These moderators were expected to understand the scientific field and bring knowledge of publishing ethics, intellectual property, and clinical research reporting guidelines. As one participant described, training is thus seen as essential:

> There are some preprint servers . . . they have academics who have full-time jobs looking at content. It's not like they're going through any kind of training program, any kind of onboarding. They're being given maybe a document to look at to say, like, "This is what you [need to check]." That's not enough. That's

> crazy to me. *Of course* you are going to be releasing content you shouldn't be releasing. (Amye Kenall, Research Square).

For similar reasons, bioRxiv and medRxiv's screeners are paid and carefully trained over several months by a Content Team that is responsible for both servers:

> We start them off very slow . . . go[ing] through the motions of looking through the manuscripts . . . not actually making any decisions. That will happen for several months until we feel satisfied they're making the right . . . decisions. And we give them a lot of feedback.

Meanwhile, the Volunteer Affiliates that work with bioRxiv and medRxiv—and who are doctoral-level experts—are onboarded on a call with members of the Content Team where they can ask questions about the process.

Legal risk was not a major theme in our interviews, but the strong emphasis on consistency and training suggests that liability concerns are quietly shaping how moderation is designed—especially in commercial or publisher-affiliated contexts.

The ownership and governance model of a server shapes not just who performs moderation but how authority is exercised when judgment becomes uncertain. How complex or borderline cases are handled—through peer discussion, escalation, or expert consultation—often reflects the server's broader governance structure.

In community-led servers with horizontal structures, where moderators are often active researchers, uncertainty was negotiated collaboratively, typically through informal peer discussion:

> If cases are complicated or difficult, we have a listserv. . . . People send emails and say, "Here's a link to a paper. I'm not sure what to do with it." . . . Then we might have a little email chit chat about the paper. (Philip Cohen, SocArXiv).

By contrast, community-led servers with more vertical structures tend to centralize discretion, with ambiguous cases escalated to founders or senior figures:

> Sometimes we get papers that we don't know how to place. . . . We have to sit down and with [our founder's] advice, we sit down and say, "Yeah, this paper is one of those tricky papers." (Nicholas Outa, AfricArXiv).

Publisher-led servers rely on more formalized workflows, distributing responsibility across defined staff roles. At Research Square, difficult cases are escalated to senior editorial staff who convene broader discussions:

> I help with consulting on more difficult cases that the screeners encounter. So, [the screeners] will come to me first, I'll look them over, and sort of open that up to a broader discussion with our editorial staff to make decisions on trickier preprints that come through our table. (Roma Konecky, Research Square).

On the other hand, publisher-adjacent platforms are more likely to involve subject matter experts in screening ambiguous submissions. These experts are typically situated at the infrastructural periphery of

the server—not part of the core operational team, yet mobilized when domain-specific judgment is required:

> We get some preprints that are very delicate in terms of what they are communicating. It can be a sensitive subject. . . . So, we are not very confident, and we prefer to send it to an expert. For that, we use SciELO's own editors from the [SciELO] collection [of journals]. (Alex Mendonça, SciELO Preprints).

These arrangements reveal how the work of moderation is enacted through different configurations of authority, where decisions about what "counts" as credible and safe-to-circulate science are inseparable from decisions about who holds the authority to decide within the infrastructure.

A detailed overview of how the moderation process is structured across our server sample can be found at Chtena et al. 2025f.

Drawing Institutional Boundaries: Moderation vs. Peer Review

As described above, participants primarily framed moderation as boundary work around what and whom belongs within legitimate scholarly communication. However, another important boundary emerged in the interviews concerning the distinction between moderation and peer review. This distinction was articulated clearly by Roma Konecky of Research Square, who underscored the non-evaluative function of moderation: "We're completely agnostic to the quality of the research. . . . That's not for us to decide. . . . We're not validating the method."

The differentiation here is not merely semantic; it serves to define the institutional place of preprint servers within the broader publishing system. By presenting moderation as a light-touch, non-evaluative process, servers sustain the claim that preprints remain "unreviewed," and therefore do not constitute prior publication—an assurance they viewed as necessary to ensure their compatibility with journal workflows. At the same time, the chosen language ("not for us to decide") signals deference to the symbolic authority of peer review as the dominant marker of scientific legitimacy. As SocArXiv's Philip Cohen explained:

> If we do anything that counts as peer review, then some journals would not accept our papers anymore, and authors might not want to post their papers on SocArXiv. It's important for us to not really evaluate the quality of the work, because that would be peer review, and that would be problematic. At the same time, we do, but at a very low bar.

Yet once participants described how screening actually unfolds, the boundary they drew between moderation and peer review appeared less clear in practice. This blurring was particularly evident in the involvement of disciplinary experts, as discussed above. Servers such as bioRxiv and ChemRxiv incorporate expert checks systematically, whereas platforms such as Jxiv and SciELO Preprints rely on specialists only when initial checks raise concerns about plausibility or disciplinary appropriateness.

Across these different models, participants described a wide range of expert involvement. Judgments could take the form of brief plausibility assessments or, in other cases, resemble near-editorial appraisals. This variability reveals how moderation slides along a continuum between infrastructural maintenance—routine checks that sustain the operability and credibility of the platform—and epistemic evaluation.

The spectrum is illustrated particularly well by the SciELO Preprints case, where moderation involves a network of volunteer editors drawn from the SciELO journals collection. As Alex Mendonça explained, the depth and style of review can vary significantly across individuals:

> It varies a lot, as you can imagine. . . . Each researcher [has] their own way of reviewing. Some are more elaborate than others. Even though it's not a peer review, some moderators are very attentive and detailed in their moderation. . . . But others are more objective and quicker.

The SciELO case highlights how moderation exists on a spectrum between administrative screening and substantive evaluation, depending on who performs it and how they interpret their role. Even when moderators are explicitly instructed *not* to perform peer review, disciplinary training and professional norms may pull them toward more detailed forms of scrutiny.

There is also a difference in how scrutiny is distributed across outcomes. Acceptance tends to be light-touch and is justified by the idea that the wider community will ultimately arbitrate a preprint's value, whereas rejection demands greater justification and care:

> For acceptance, we don't require much from our moderators in terms of justifying the reason for acceptance. But for rejection, it's more delicate. . . . It's important to give a reason to the authors. (Alex Mendonça, SciELO Preprints).

It is notable that when describing expert checks, participants sometimes adopted language associated with peer review—such as "reviewing" a manuscript or providing "a reason for rejection"—while simultaneously insisting that such assessments do not constitute formal peer review.

Taken together, these dynamics reveal a gap between how moderation is described and how it is performed. Servers must preserve a rhetorical separation from peer review to remain legible within existing journal workflows, yet the handling of ambiguous submissions requires forms of discretionary judgment that draw on the same sensibilities that underpin peer review. Moderation thus occupies a hybrid space—publicly procedural, practically discretionary—in which the credibility of early research is negotiated without formally assuming the authority of peer review. It is within this space that preprint servers carve out their institutional identity while shaping what enters the scholarly record.

Between Openness and Precaution: How Epistemic and Institutional Logics Shape Moderation

Our analysis shows that preprint moderation unfolds along two interrelated dimensions. One is epistemic—how moderators interpret, filter, and sometimes exclude submissions based on tacit cues of credibility, relevance, and scientific form. The other dimension is institutional—how moderation is organized, legitimated, and positioned within the wider publishing system, including who is authorized to moderate, how responsibility and liability are managed, and how servers define their relationship to journal peer review. Discourse—the language and narratives through which moderation's purpose, scope, and legitimacy are articulated—cuts across both dimensions.

These two dimensions (i.e., epistemic and institutional) do not represent discrete or exhaustive categories, but rather overlapping planes of analysis that capture how moderation operates across sites of

judgment, governance, and meaning-making.[2] Boundary work and valuation, meanwhile, function as conceptual lenses, illuminating how servers negotiate what can circulate as scholarship and under what epistemic and institutional conditions. Table B2, appendix B (Chtena et al. 2025e) summarizes how the epistemic and institutional dimensions surface across different stages of moderation and points to where boundary work and valuation help explain the types of judgment being performed. These patterns reveal what appears as procedural screening but is, in practice, a form of interpretive labor.

Epistemically, moderation is a site where distinctions are made about what counts as legitimate scholarly communication, based on tacit sensibilities about plausibility, disciplinary form, interpretability, and the potential implications of public circulation. Through these judgments, moderators draw boundaries between science and non-science, between credible knowledge and content perceived as "noise," and between early-stage work that can responsibly circulate and material that may cause harm or confusion. As Lamont (2009) shows in her study of academic review panels, evaluators rely on shared sensibilities and implicit understandings of acceptable scholarly work. We observe similar repertoires in preprint moderation, but embedded into "lighter," faster, and less formalized infrastructures that make moderation's evaluative role only partially legible to wider publics. Yet the substance of these judgments differs in important ways. Whereas Lamont's reviewers deliberate over excellence—originality, rigor, methodological strength, and contribution—moderators in our study are less concerned with ranking quality and more with establishing whether a submission is credible enough to circulate without distorting public understanding or undermining trust in preprints. These judgments rely less on comparative merit and more on plausibility, trustworthiness, and anticipated consequences of dissemination. By foregrounding this mode of precautionary, threshold-oriented assessment, our study joins recent valuation scholarship in examining how worth is assigned through infrastructures, routines, and practices of maintenance rather than only through ranking or excellence (e.g., Kornberger et al. 2017; Denis and Pontille 2015). We show that valuation can also take the form of infrastructural care, where the central question is not "Is this work excellent?" but "Is this work safe to post?" Our contribution extends this line of inquiry to early-stage knowledge circulation, showing how these quieter, anticipatory judgments shape which forms of research become socially and epistemically viable.

These forms of anticipatory judgment operate at an epistemic level, but they also blur into institutional concerns—about responsibility, risk, and trustworthiness—which is where a second, institutional dimension of moderation becomes visible. As we have shown, moderation is not oriented toward a single problem; different servers are trying to solve different problems, shaped by their own, unique disciplinary contexts, governance histories, and imaginaries of risk. For instance, while biomedical servers tend toward more cautious, multi-layer screening to prevent harm, others emphasize inclusion and

[2] Although our analysis emphasizes epistemic and institutional dynamics, moderation is also deeply infrastructural. Platform affordances—e.g., submission portals, screening dashboards, queue structures, and communication systems—shape how judgments are made and enacted. Rather than treating the infrastructural as a third analytical category, we understand it as a mediating layer that conditions both epistemic and institutional practices (Bowker and Star 1999; Edwards 2010).

accessibility to preserve low barriers to entry. These differences reflect locally situated commitments to openness, precaution, and community expectations, resulting in a patchwork of moderation cultures rather than a standardized evaluative regime. Efforts to standardize moderation approaches thus risk imposing the assumptions, priorities, and risk sensitivities of one domain onto others.

Another institutional dynamic concerns how servers position moderation in relation to peer review. Here a form of boundary work emerges, distinct from the epistemic work of sorting science from non-science, which downplays rather than asserts evaluative authority. Moderators perform lightweight valuation work that relies on genre-specific cues—plausibility, tone, citation practice, and disciplinary form—but are quick to emphasize that these activities do not constitute peer review. Yet contradictions persist. Servers describe moderation as value-neutral but also note that only those with publishing experience or subject expertise are equipped to decide what counts as credible. They frame the evaluation they do as "very low bar" but simultaneously emphasize that making such judgments effectively requires extensive training. These tensions echo what Gieryn (1983) observed in his work on science's boundaries; the more actors attempt to draw clear lines, the more the fuzziness of those boundaries becomes visible.

Importantly, this rhetorical distancing plays a protective role for servers; it preserves the idea that preprints do not count as prior publication and will not interfere with journal submission. It also helps servers co-exist with journals rather than compete with them—reaffirming peer review's status as the gold standard (Berenbaum 2023) and avoiding the perception that preprints might undermine it (Chtena et al. 2025a). In this sense, servers engage in a defensive form of boundary work, safely positioning moderation outside the evaluative jurisdiction of journals, even when the underlying practices blur. While prior work has shown that boundary work can include minimizing or disclaiming expertise to manage institutional responsibilities (Gieryn 1999; Wynne 1992; Epstein 1995), this dynamic has not been analyzed in the context of preprint moderation. We show how servers' distancing from peer review authority operates as a boundary-making strategy that stabilizes moderation as separate from formal review.

Taken together, our findings demonstrate that moderation is neither a neutral checkpoint nor a simplified version of peer review, but a site where epistemic and institutional judgments converge in shaping what becomes admissible as early-stage scientific communication. What appears outwardly as procedural screening relies on forms of precautionary, classificatory, and discretionary labor directed toward multiple, and often competing, goals. Moderators must continuously balance responsibility with openness, legitimacy with speed, and evaluation with a strategic disavowal of peer review authority. These frictions make moderation an increasingly consequential site of scholarly judgment that shapes the boundaries of credible science and the politics of openness as preprinting becomes a default mode of dissemination. Moderation thus remakes the governance of scholarly communication by introducing a new moment of public release that demands forms of oversight and justification not found in journal-based publishing.

Conclusion

This paper offers the first comparative, interview-based account of moderation practices across diverse preprint servers, providing new insights into a low-visibility but increasingly important part of scholarly communication. While preprint platforms are often described as alternatives to journal publishing, our findings show that they, too, are shaped by processes of sorting, boundary-making, and valuation—although these happen under different infrastructural, temporal, and normative conditions. By showing

how epistemic and institutional judgments converge in moderation, our study highlights the evaluative work that precedes public circulation and shapes what becomes thinkable as scholarship. This perspective draws attention to how infrastructures that promise openness nonetheless depend on subtle forms of boundary-making, and to how these dynamics may evolve as preprints become more tightly integrated into journal workflows, funder mandates, and research assessment practices.

Future research could trace how moderation policies evolve in response to emerging challenges such as AI-generated content or politically sensitive research, and examine how authors interpret moderation processes and how these perceptions influence their submission behavior. More work is needed to understand how moderation shapes inclusion, exclusion, and epistemic diversity.

Data Availability

Supplemental materials for this paper are available via OSF: https://doi.org/10.17605/OSF.IO/DRTJ6.

Acknowledgements

We thank the anonymous reviewers for their exceptionally careful, generous, and constructive engagement with the manuscript. Their comments significantly strengthened the clarity, framing, and contribution of this paper. One author (Stephen Pinfield) serves as an editor of MetaResearch Open Review (MetaROR). This role is not connected to the focus of this study. One author (Natascha Chtena) is a member of the editorial team at mediastudies.press, where one study participant (Jeff Pooley) also holds an editorial role. This affiliation began after the study was completed and did not influence the research. All other authors declare no conflicts of interest. Generative AI tools were used in the later stages of manuscript preparation to assist with editing for clarity, concision, and organization. AI tools were not used to generate empirical data, conduct analysis, or produce original interpretations; all substantive arguments, interpretations, and conclusions remain the authors' own.

Author Biographies

Juan Pablo Alperin is Associate Professor in the Publishing Program, Scientific Director of the Public Knowledge Project (PKP), and Co-Director of the Scholarly Communications Lab at Simon Fraser University. His research focuses on scholarly communication and the development of evidence-based approaches to open access and open science, with particular attention to equity, public engagement, and scholarship in Latin America and the Global South.

Melanie Benson Marshall is a Postdoctoral Research Associate at the University of Sheffield and a Research Fellow with the Research on Research Institute (RoRI). Her research in metascience focuses on the evolution of open science, scholarly communication, and the evaluation of research systems. She is currently conducting experimental research on distributed peer review models for grant funding.

Natascha Chtena is a Research Associate at the Scholarly Communications Lab and a Research Affiliate of the Public Knowledge Project (PKP) at Simon Fraser University. Her research spans open science, scholarly publishing, and knowledge infrastructures, examining how systems of publication and evaluation are organized, governed, and contested.

Alice Fleerackers is Assistant Professor of Journalism and Civic Engagement in the Department of Media Studies at the University of Amsterdam. Her research examines the intersections of journalism, health and science communication, and scholarly communication. She is also the Vice President of the Public Communication of Science and Technology Network (PCST) and a Research Associate of the Scholarly Communications Lab.

Irene V. Pasquetto is a scholar of information and communication studies who examines contemporary knowledge production and how digital technologies shape it. Her recent work focuses on open science and public participation in research. She is Assistant Professor at the University of Maryland.

Stephen Pinfield is Professor of Information Services Management at the University of Sheffield, UK, and Senior Research Fellow in the Research on Research Institute. He has a particular interest in scholarly communication, open science, research data management, science evaluation and research policy. He is the author of *Achieving Global Open Access: The Need for Scientific, Epistemic and Participatory Openness* (Routledge, 2024).

References

ArXiv. n.d. "About arXiv - arXiv Info." Accessed August 12, 2024. https://info.arxiv.org/about/index.html.

Balaji, B. Preedip, and M. Dhanamjaya. 2019. "Preprints in Scholarly Communication: Re-Imagining Metrics and Infrastructures." *Publications* 7(1): 1–23. https://doi.org/10.3390/publications7010006.

Berenbaum, May R. 2023. "On Peer Review—Then, Now, and Soon to Be?" *Proceedings of the National Academy of Sciences* 120(11): 1–3. https://doi.org/10.1073/pnas.2302593120.

Berg, Jeremy M., Needhi Bhalla, Philip E. Bourne, Martin Chalfie, et al. 2016. "Preprints for the Life Sciences: The Time Is Right for Biologists to Post their Research Findings onto Preprint Servers." *Science* 352(6288): 899–901. https://doi.org/10.1126/science.aaf9133.

Bowker, Geoffrey C., and Susan Leigh Star. 1999. *Sorting Things Out: Classification and Its Consequences*. Cambridge, MA: MIT Press.

Chaleplioglou, Artemis, and Alexandros Koulouris. 2023. "Preprint Paper Platforms in the Academic Scholarly Communication Environment." *Journal of Librarianship and Information Science* 55(1): 43–56. https://doi.org/10.1177/09610006211058908.

Chiarelli, Andrea, Rob Johnson, Stephen Pinfield, and Emma Richens. 2019. "Preprints and Scholarly Communication: An Exploratory Qualitative Study of Adoption, Practices, Drivers and Barriers." *F1000Research* 8: 1–77. https://doi.org/10.12688/f1000research.19619.2.

Christie, Alec P., Thomas B. White, Philip A. Martin, Silviu O. Petrovan, et al. 2021. "Reducing Publication Delay to Improve the Efficiency and Impact of Conservation Science." *PeerJ* 9: 1–24. https://doi.org/10.7717/peerj.12245.

Chtena, Natascha, Juan Pablo Alperin, Stephen Pinfield, Alice Fleerackers, et al. 2025a. "Preprint Servers and Journals: Rivals or Allies?" *Journal of Documentation* 81(4): 847–869. https://doi.org/10.1108/JD-09-2024-0215.

———. 2025b. *Appendix A*. Open Science Framework (OSF), last updated December 18, 2025, accessed February 22, 2026. https://osf.io/drtj6/files/uq5js.

———. 2025c. *Semi-Structured Interview Guide*. Open Science Framework (OSF), last updated December 18, 2025, accessed February 22, 2026. https://osf.io/drtj6/files/kg6y2.

———. 2025d. *Codebook*. Open Science Framework (OSF), last updated December 18, 2025, accessed February 22, 2026. https://osf.io/drtj6/files/8a45c.

———. 2025e. *Appendix B*. Open Science Framework (OSF), last updated December 18, 2025, accessed February 22, 2026. https://osf.io/drtj6/files/f4us9.

———. 2025f. *Moderation Workflow: Overview*. Open Science Framework (OSF), last updated December 18, 2025, accessed February 22, 2026. https://osf.io/drtj6/files/m37zj.

Collins, Harry, and Robert Evans. 2007. *Rethinking Expertise*. Chicago: The University of Chicago Press.

Confederation of Open Access Repositories (COAR) and Centre pour la Communication Scientifique Directe (CCSD) [Center for Direct Scientific Communication]. n.d. "Directory of Open Access Preprint Repositories." Last updated February 4, 2026, accessed February 24, 2026. https://doapr.coar-repositories.org.

Curry, Stephen. 2015. "Peer Review, Preprints and the Speed of Science." *The Guardian*, September 7, 2015, accessed February 20, 2026. https://www.theguardian.com/science/occams-corner/2015/sep/07/peer-review-preprints-speed-science-journals.

Delfanti, Alessandro. 2015. "Preprints and Peer Review: A Lesson from Physics." *Journal of Science Communication* 14(4): 1–4. https://doi.org/10.22323/2.14040305.

———. 2016. "Beams of Particles and Papers: How Digital Preprint Archives Shape Authorship and Credit." *Social Studies of Science* 46(4): 629–645. https://doi.org/10.1177/0306312716659373.

———. 2017. "Fake Archives: Doppelgängers and the Search for Openness in Scholarly Communication Platforms," preprint, Open Science Framework (OSF), 1–6. https://doi.org/10.31235/osf.io/5pfe4.

Denis, Jérôme, and David Pontille. 2015. "Material Ordering and the Care of Things." *Science, Technology, & Human Values* 40(3): 338–367. https://doi.org/10.1177/0162243914553129.

Edwards, Paul N. 2010. *A Vast Machine: Computer Models, Climate Data, and the Politics of Global Warming*. Cambridge, MA: MIT Press.

Epstein, Steven. 1995. "The Construction of Lay Expertise: AIDS Activism and the Forging of Credibility in the Reform of Clinical Trials." *Science, Technology, & Human Values* 20(4): 408–437. https://doi.org/10.1177/016224399502000402.

Fereday, Jennifer, and Eimear Muir-Cochrane. 2006. "Demonstrating Rigor Using Thematic Analysis: A Hybrid Approach of Inductive and Deductive Coding and Theme Development." *International Journal of Qualitative Methods* 5(1): 80–92. https://doi.org/10.1177/160940690600500107.

Fleerackers, Alice, Kenneth Shores, Natascha Chtena, and Juan Pablo Alperin. 2024. "Unreviewed Science in the News: The Evolution of Preprint Media Coverage from 2014–2021." *Quantitative Science Studies* 5(2): 297–316. https://doi.org/10.1162/qss_a_00282.

Funk, Kathryn, Alice Meadows, Alex Mendonça, Oya Rieger, et al. 2020. "Preprint Authors Optimistic About Benefits: Preliminary Results from the #bioPreprints2020 Survey," *ASAPbio*, July 27, 2020, accessed February 24, 2026. https://asapbio.org/biopreprints2020-survey-initial-results/.

Gieryn, Thomas F. 1983. "Boundary-Work and the Demarcation of Science From Non-Science: Strains and Interests in Professional Ideologies of Scientists." *American Sociological Review* 48(6): 781–95. https://doi.org/10.2307/2095325.

———. 1999. *Cultural Boundaries of Science: Credibility on the Line*. Chicago: The University of Chicago Press.

Ginsparg, Paul. 2011. "ArXiv at 20." *Nature* 476(7359): 145–47. https://doi.org/10.1038/476145a.

———. 2021. "Lessons from arXiv's 30 Years of Information Sharing." *Nature Reviews Physics* 3(9): 602–3. https://doi.org/10.1038/s42254-021-00360-z.

Gunnarsdóttir, Kristrún. 2005. "Scientific Journal Publications: On the Role of Electronic Preprint Exchange in the Distribution of Scientific Literature." *Social Studies of Science* 35(4): 549–79. https://doi.org/10.1177/0306312705052358.

Hamade, Lama, Amr Ali, Ayat R. A. Hijazi, Dima Itani, et al. 2022. "Policies of Biomedical Preprint Servers on Conflicts of Interest, Authorship, and Research Integrity Lacked Important Details." *Journal of Clinical Epidemiology* 146: 47–59. https://doi.org/10.1016/j.jclinepi.2022.02.007.

Helgesson, Claes-Fredrik, and Fabian Muniesa. 2013. "For What It's Worth: An Introduction to Valuation Studies." *Valuation Studies* 1(1): 1–10. https://doi.org/10.3384/vs.2001-5992.13111.

Howard, Graham. 2012. "Peer Review as Boundary Work." *Journal of Scholarly Publishing* 43(3): 322–35. https://doi.org/10.3138/jsp.43.3.322.

Huber, Jürgen, Sabiou Inoua, Rudolf Kerschbamer, Christian König-Kersting, et al. 2022. "Nobel and Novice: Author Prominence Affects Peer Review." *Proceedings of the National Academy of Sciences* 119(41): 1–7. https://doi.org/10.1073/pnas.2205779119.

Irawan, Dasapta Erwin, Hilyatuz Zahroh, and Iratxe Puebla. 2022. "Preprints as a Driver of Open Science: Opportunities for Southeast Asia." *Frontiers in Research Metrics and Analytics* 7(September): 992942. https://doi.org/10.3389/frma.2022.992942.

Jasanoff, Sheila, ed. 2004. *States of Knowledge: The Co-Production of Science and Social Order*. London: Routledge.

Kirkham, Jamie J., Naomi C. Penfold, Fiona Murphy, Isabelle Boutron, et al. 2020. "Systematic Examination of Preprint Platforms for Use in the Medical and Biomedical Sciences Setting." *BMJ Open* 10(12): 1–8. https://doi.org/10.1136/bmjopen-2020-041849.

Kornberger, Martin, Dane Pflueger, and Jan Mouritsen. 2017. "Evaluative Infrastructures: Accounting for Platform Organization." *Accounting, Organizations and Society* 60: 79–95. https://doi.org/10.1016/j.aos.2017.05.002.

Lamont, Michèle. 2009. *How Professors Think: Inside the Curious World of Academic Judgment*. Cambridge, MA: Harvard University Press.

———. 2012. "Toward a Comparative Sociology of Valuation and Evaluation." *Annual Review of Sociology* 38: 201–21. https://doi.org/10.1146/annurev-soc-070308-120022.

Maggio, Lauren A., and Alice Fleerackers. 2023. "Preprints in Health Professions Education: Raising Awareness and Shifting Culture." *Academic Medicine* 98(1): 17–20. https://doi.org/10.1097/ACM.0000000000005001.

Malički, Mario, Ana Jerončić, Gerben ter Riet, Lex M. Bouter, et al. 2020. "Preprint Servers' Policies, Submission Requirements, and Transparency in Reporting and Research Integrity Recommendations." *JAMA* 324(18): 1901–3. https://doi.org/10.1001/jama.2020.17195.

Oliveira Henriques, Susana, Narmin Rzayeva, Stephen Pinfield, and Ludo Waltman. 2023. "Preprint Review Services: Disrupting the Scholarly Communication Landscape?" Open Science Framework (OSF), 1–29. https://doi.org/10.31235/osf.io/8c6xm.

Pereira, Fábio Henrique, and Raphael Sandes de Oliveira. 2024. "Journalists and Scientists Together: The Public Problem of Science Disinformation in Brazil." *Journal of Science Communication* 23(3): 1–19. https://doi.org/10.22323/2.23030204.

Pontille, David, and Didier Torny. 2014. "The Blind Shall See! The Question of Anonymity in Journal Peer Review." *Ada: A Journal of Gender, New Media, and Technology* (4): 1–8. https://papers.ssrn.com/abstract=2427461.

———. 2017. "Beyond Fact Checking: Reconsidering the Status of Truth of Published Articles." *EASST Review* 36(1).

https://easst.net/easst-review/36-1/beyond-fact-checking-reconsidering-the-status-of-truth-of-published-articles/.

Powell, Kendall. 2016. "Does It Take Too Long to Publish Research?" *Nature* 530(7589): 148–51. https://doi.org/10.1038/530148a.

Puebla, Iratxe, Jessica Polka, and Oya Y. Rieger. 2021. *Preprints: Their Evolving Role in Science Communication*. Charleston Briefings. Los Angeles, USA: Against the Grain (Media), LLC.

Reyes-Galindo, Luis. 2016. "Automating the Horae: Boundary-Work in the Age of Computers." *Social Studies of Science* 46(4): 586–606. https://doi.org/10.1177/0306312716642317.

Rosen, Rochelle K., Monique Gainey, Sabiha Nasrin, Stephanie C. Garbern, et al. 2023. "Use of Framework Matrix and Thematic Coding Methods in Qualitative Analysis for mHealth: The FluidCalc app." *International Journal of Qualitative Methods* 22: 1–10. https://doi.org/10.1177/16094069231184123.

Shapin, Steven. 1995. "Here and Everywhere: Sociology of Scientific Knowledge." *Annual Review of Sociology* 21: 289–321. https://doi.org/10.1146/annurev.so.21.080195.001445.

Shapin, Steven, and Simon Schaffer. 1985. *Leviathan and the Air-Pump: Hobbes, Boyle, and the Experimental Life*. Princeton: Princeton University Press.

Star, Susan Leigh, and James R. Griesemer. 1989. "Institutional Ecology, 'Translations' and Boundary Objects: Amateurs and Professionals in Berkeley's Museum of Vertebrate Zoology, 1907-39." *Social Studies of Science* 19(3): 387–420. https://doi.org/10.1177/030631289019003001.

Swain, Jon. 2018. "A Hybrid Approach to Thematic Analysis in Qualitative Research: Using a Practical Example." In *Sage Research Methods Cases Part 2*. London: SAGE Publications Ltd. https://doi.org/10.4135/9781526435477.

Swire-Thompson, Briony, and David Lazer. 2020. "Public Health and Online Misinformation: Challenges and Recommendations." *Annual Review of Public Health* 41: 433–51. https://doi.org/10.1146/annurev-publhealth-040119-094127.

The National Commission for the Protection of Human Subjects of Biomedical and Behavioral Research. 1979. *The Belmont Report: Ethical Principles and Guidelines for the Protection of Human Subjects of Research*. U.S. Office of the Secretary, Department of Health, Education, and Welfare, April 18, 1979. Accessed February 20, 2026. https://www.hhs.gov/ohrp/regulations-and-policy/belmont-report/read-the-belmont-report/index.html.

United Nations Educational, Scientific and Cultural Organization (UNESCO). 2021. *UNESCO Recommendation on Open Science*. Paris: UNESCO. https://doi.org/10.54677/MNMH8546.

Vasileiou, Konstantina, Julie Barnett, Susan Thorpe, and Terry Young. 2018. "Characterising and Justifying Sample Size Sufficiency in Interview-Based Studies: Systematic Analysis of Qualitative Health Research over a 15-Year Period." *BMC Medical Research Methodology* 18(1): 1–18. https://doi.org/10.1186/s12874-018-0594-7.

Vianello, Stefano Davide. 2021. "The 'Pre' in [my] 'Preprint' Is for Pre-Figurative." *Commonplace*, October 7, 2021. https://doi.org/10.21428/6ffd8432.5de25622.

Wang, Zhiqi, Yue Chen, and Wolfgang Glänzel. 2020. "Preprints as Accelerator of Scholarly Communication: An Empirical Analysis in Mathematics." *Journal of Informetrics* 14(4): 1–24. https://doi.org/10.1016/j.joi.2020.101097.

Wynne, Brian. 1992. "Public Understanding of Science Research: New Horizons or Hall of Mirrors?" *Public Understanding of Science* 1(1): 37–43. https://doi.org/10.1088/0963-6625/1/1/008.