Fediverse Governance Opportunities for Funders & Developers

Erin Kissane and Darius Kazemi with the support of the Digital Infrastructure Insights Fund
August 20, 2024

A 100-plus-page analysis of ethnographic research is a tall order for anyone to read, so although we've sought to provide a modular reading experience in our full report, we understand that not everyone is going to pick through it! This document is directed at two groups of people who are well placed to improve the infrastructure around governance on the Fediverse: independent software developers and philanthropic funders.

While there is no shortage of software developers pursuing projects on the Fediverse, these developers understandably often focus on problems they encounter as day-to-day users of Fediverse, or are unsure how best to contribute to the human/governance side of the network without—or in addition to—spinning up their own servers. Their work tends to be about enabling publishing and communication between individual users, which makes sense since most developers come to the work having participated in traditional social media ecosystems; thinking about online community governance does not necessarily come naturally if a person has not

participated in governance processes. We hope this document provides a good summary for developers who are looking for ways to help build a Fediverse that is more welcoming, more inclusive, and easier to adapt to thoughtful modes of governance.

Philanthropic organizations, on the other hand, are often on the outside of the Fediverse looking in, frequently with concerns about aspects of centralized platform governance that aren't viable on the Fediverse, but with little corresponding visibility into the Fediverse's existing strengths and ways of leveraging those strengths to build a more resilient network. This document highlights gaps in governance infrastructure we identified, in the hope that funders can use it to guide future grant-making and sponsorships.

We recommend reviewing **Section One: Overall Observations, Risks and Mitigations** in our report to get a broader sense of the characteristics of the landscape we encountered in our research, along with an array of governance-related risks—and recommendations for addressing those risks.

Institutions for Collective Governance

Many of the gaps that persist in governance tooling, policy, and practice on the Fediverse persist because they're difficult or impossible to address at the level of a single server or small coalition of servers. We believe that the network is entering a potentially fruitful period for the development and maintenance of multiple, thoroughly Fediverse-integrated institutions designed to support governance in a genuinely decentralized way, and that funding a multiplicity of new and existing institutions will be a way to provide an outsized boost to the network's development.

Institutions for identifying and addressing illegal content and collective threats

By design, the Fediverse has no central authority to adjudicate, report, or remove illegal content (like CSAM and, in many jurisdictions, graphic terrorist/extremist material) across the whole network. It also lacks the centralized telemetry that would allow for the detection and mitigation of network-wide adversarial behavior (like spam, scams, and coordinated inauthentic account networks) that Yoel Roth and Samatha Lai term "collective threats" (Roth and Lai 2024).

Research from the Atlantic Council's Task Force for a Trustworthy Future Web, the Stanford Internet Observatory, and the Journal of Online Trust & Safety highlights known vulnerabilities of federated architectures and current and probable future threats to Fediverse server teams and their attempts at good governance.

Fediverse server operators focused on good governance have dealt with present-day levels of adversarial and illegal content and behavior by moderating their membership, exercising the option to defederate from problem servers that—willingly or by neglect—host adversarial accounts, and

sharing relatively simple tools like lists of problem servers or more sophisticated databases of commonly reported servers and material. For the server teams we interviewed, these methods work well nearly all of the time. Nevertheless, many server operators believe that current methods won't work forever, or even much longer—especially because even small, tightly moderated servers will be negatively affected when high volumes of unwanted or illegal material overwhelm the resources of previously trustworthy servers.

The data-analysis methods and systems that work within and between central platforms to address threats like these are unlikely to be technically feasible or culturally acceptable in the Fediverse, which is—again, by design—resistant to centralized surveillance and centralized governance. Our research focused only peripherally on these threats, and prior work in Roth and Lai (2023), Roth and Lai (2024), and Thiel and DiResta (2023) offers detailed recommendations on both technical and institutional remedies for current gaps in Fediverse systems. Rather than reiterate or suggest revisions to their perspectives, we offer **a short series of cultural recommendations for building institutional and technical solutions that are more likely to be accepted within the Fediverse's many fiercely independent, highly localized, and strenuously anti-surveillance cultures**.

Based on our discussions with server teams, we believe that many (though not necessarily a majority of) server operators would be willing to work with at least one institution focused on data analysis, threat assessment/investigation, and reporting that is limited, transparent, and ethical. However, because even semi-centralized threat detection raises hot-button questions for the Fediverse—and emerges from a technical and cultural landscape associated with the big technology companies many people came to the Fediverse to avoid—there are real barriers to the success of such institutions.

We think institutions most likely to succeed in this work will be:

- **Multiple and culturally specific, rather than monolithic and global.** The Fediverse operates by connecting local systems running under local norms, and we think institutional efforts to combat collective threats will be more accepted and ultimately more effective if they work at the level of regional, topical, or cultural coalitions.
- **Transparent about their relationships with central platforms and state actors.** Clarity about what kind of information an institution will share with law enforcement bodies, with governments, and with other online platforms will allow Fediverse server teams to make decisions about their participation that are suitable to the needs of their members, some of whom live under authoritarian governments and many of whom are deeply resistant to data-sharing of any kind with dominant technology companies.
- **Respectful of and curious about the needs of marginalized and racialized communities and their members.** Organizations that take seriously the concerns of those most likely to be unfairly punished or excluded by traditional governance systems are likely to be

understood as better citizens of the Fediverse than those which write off increased harm to these groups as a necessary cost of achieving governance goals.

- **At least partially integrated into the Fediverse, rather than approaching from a position of distant or external authority.** Institutions that count active Fediverse server operators and members among their core staff will have an easier time navigating the network's cultural and technical quirks and complexities than teams parachuting in from "outside." IFTAS (Independent Federated Trust & Safety) is a great model of this kind of organizational development.

We encourage potential funders (and developers) interested in building out these much-needed capacities for decentralized networks to focus on projects that demonstrate a commitment to these principles as a way of increasing uptake and acceptance.

Institutions for pragmatic sustainability

Our research report focuses on the unique capabilities and possibilities of human-scale, medium-sized social media servers on the Fediverse. Many of these medium-sized servers rely on small-scale fundraising for vital financial support and struggle to connect disparate, isolated systems for financial, administrative, and communication work. Legal responsibilities are often unclear, and many small and medium-sized servers in our sample have required paid or pro-bono legal advice.

Particularly in the wake of the Open Collective Foundation's abrupt dissolution, there are few options for Fediverse server teams who want to formalize their organizational, legal, and financial structures without taking on the financial and administrative burden of incorporating as a business or forming a recognized non-profit entity. Because decentralized networks are relatively new and often ill-understood, it's difficult for many US-based servers to obtain fiscal sponsorship, an organizational structure that is accessible to even small teams.

We encourage funders concerned with governance on the Fediverse to consider supporting the development of organizations that can provide fiscal sponsorship and potentially other forms of financial, administrative, and legal advisory services. In the interim, we also recommend funding research into and comprehensive and transparent documentation of subjectively successful financial structures and sustainability campaigns for medium-sized Fediverse servers.

Institutions for participatory and democratic governance

Many server teams in our research sample indicated that they're interested in building out more participatory or democratic forms of server governance, but with the exception of the two formal cooperatives we spoke with, most teams find it difficult to understand what it takes—in legal and financial but especially social/cultural terms—to get from point A (interest in participatory governance) to point B (a flourishing self-governed server). One Fediverse cooperative,

Social.coop, provides starter documentation in this direction, but server teams are hungry for detailed advice, templates, and potentially even mentoring to help them move toward participatory governance.

We think there's a fascinating opportunity to fund the provision of additional documentation, focused training on setup and maintenance of participatory systems, and coaching for the participatory governance of Fediverse servers.

Bringing external institutions into the the Fediverse

We think the potential benefits of participation by (subjectively benevolent) institutions in the Fediverse are benefits to the commons: if more institutions offer financially sustainable, appropriately staffed servers and services, Fediverse users gain access to broader sources of information, more connection with people and entities they value, and potentially to servers that provide stable, long-term community hubs less likely to be subject to arbitrary shutdowns or mass defederations than are many hobby servers.

For these reasons, we recommend that funding organizations consider sponsoring long-term, committed participation in the Fediverse by stable institutions including civic and governing bodies, cultural and media organizations, technology and philanthropic organizations, and potentially labs or projects within higher learning and research institutions. With equal emphasis, we recommend supporting only those institutional Fediverse projects that demonstrate their ability to run for at least two to four years past launch, and which launch with a plan for orderly shutdown and account migration after this initial period if it becomes necessary. Short-term experiments in the Fediverse may be valuable to those running the experiments, but we're unclear about their value to the broader network and its members.

Software and Systems

Software is not the end all be all of socio-technical systems, but there remain problems that can be solved by funding socio-technical projects. This section collects a few of these.

Helping users pick a server

We believe funders could support infrastructure that would allow users to indicate what sort of governance (moderation, leadership, diplomacy) they are interested in. One could imagine a "wizard" type interface that asks a small number of questions and then gives the user a short list of servers that meet their criteria. Or anything else! This is a problem space that almost by definition can't be solved by a single Fediverse project. It will require coalition building across projects, servers, and perhaps even on the protocol level. (We write more on the rationale and need for this in Helping existing and potential Fediverse members find well-governed servers.)

Legal compliance tools

There are opportunities to provide technical tools that aid server operators who wish to remain compliant with relevant laws. For example, in the US, operators are required to report apparent child sexual abuse material (CSAM) to NCMEC's CyberTipline. While there are online forms and even an API for filing these reports, there are also legal requirements in the US around retaining CSAM content for 90 days to aid law enforcement. Many Fediverse admins understandably do not feel comfortable retaining this information. An encrypted lockbox-style system with a robust audit trail indicating whether and by whom the data was accessed and decrypted could be developed to assist Fediverse admins in their legal compliance. This is just one example; local laws vary and while corporations with legal teams can reasonably be counted on to develop their own compliance tools, an open source ecosystem requires outside funding. For a start, one could take a look at this IFTAS guide to Fediverse server operators' responsibilities under the EU Digital Services Act, make a list of every place where Fediverse infrastructure and software falls short, and fund projects to fill those gaps.

An example of one such legal/technical support system that exists today is the Time Zone Database (TZDB). Time zones are legal constructions that can change at any time when a country or region decides to tweak them. Corporations and other entities can't be expected to keep track of time zone related legislation in all countries and municipalities of the world. So ICANN manages/funds TZDB, a project that keeps track of all this legislation and coordinates with governments around the world to release updated time zone information as soon as it is relevant.

Fiscal tools

Small social media communities that want to run themselves in a nonprofit manner currently lack the financial infrastructure to do so. In a US context, there is a lack of fiscal sponsorship for server operators, especially after the shuttering of OpenCollective Foundation. There is clearly room for one or more 501(c)3 organizations to step up and provide fiscal sponsorship for Fediverse servers that don't want to be hosted out of personal bank accounts but also cannot afford the steep costs of incorporating as a nonprofit. There may even be a need for a nonprofit dedicated to exactly this kind of fiscal sponsorship. Without it, noncommercial federated social media that aspires to move beyond the limits of "hobby servers" may die out.

Moderation tools

The federated social media ecosystem needs generic open source content moderation tools and workflows. These projects have been mostly built out as proof of concept and swiftly abandoned. Long term funding of an open source content moderation system that can ingest content from multiple sources (including ActivityPub servers) and put the content into automated and manual

moderation queues would be invaluable to the Fediverse and the social internet in general. Our main report identifies the following risks that a project like this could mitigate: lack of moderator training, unsophisticated spam campaigns, unsophisticated trolling, time-consuming moderation tools, insufficient appeals tooling, lack of ability to communicate easily or well with other server teams, lack of readily-available tools to detect and report illegal content.

Some key capabilities missing from the current Mastodon tooling that we have identified in the main report include:

- **Support for collaborative moderation.** Moderation tooling on Mastodon is mostly designed as a single-user experience. (This is not unique to Mastodon; other Fediverse projects assume moderation happens in a vacuum as well.) Building out open source moderation tools with *teams of moderators* in mind would provide a huge benefit to especially medium-sized and large servers, by allowing moderation teams to act in concert. Collaborative moderation could even happen across servers, given flexible enough tooling.

- **Better communication channels between moderators and members.** "Transparency" on large social media platforms often looks like quarterly reports, documentation of policy, or large data sets provided to governments or third party watchdog organizations. But for our study participants, *transparency means a human touch, an ability to ask "why was this specific decision made", and a sense of connection to and even collaboration with the moderation team*. Put another way, moderators on medium-size Fediverse servers may have as much in common with community managers as they do with trust and safety workers. Designers of moderation tools need to keep in mind that moderation work on these servers is not just about having a queue of content to efficiently approve or reject, but is also about investigative work and open communication with users.

- **Shared blocklists and/or federated moderation.** While server operators remain skeptical of defederation lists or shared blocklists, many operators we spoke to indicated that if they could enable *social* sharing of blocks they would do so. Sophisticated projects in this realm would provide many different layers of trust and mechanisms for input that moderators could tweak. For example, while the basic solution is to allow servers to trust some authority and subscribe to their blocks, there are plenty of other options including but not limited to trusting the decision of coalitions rather than single entities, or providing suggestions of other servers that make similar moderation decisions to one's own. There are many exciting possibilities and proposals in this area and there is room to *fund all of them* and see what works.

- **Content filtering**. Server operators need access to automated content filtering. This could be an integration with off-the-shelf spam identification tools, matching of known-illegal content to hashed databases, or any number of other solutions. While many solutions already exist in the enterprise software ecosystem, these types of tools need to become accessible to small communities that do not have millions (or thousands) of dollars to spend on trust and safety.

- **Accessible limited/allow-list federation.** We've argued in our main findings report that the Fediverse is best conceptualized not as a platform, but as a social component of the open web. This model confers benefits, but also drawbacks—most obviously that the worst things on the open web will also be present on the Fediverse. Our research participants have found ways to protect their members by defederating—often proactively and aggressively—from the Fediverse's worst actors, but especially for new and small teams and/or servers hosting communities frequently targeted for abuse and harassment, this approach isn't always sufficient. Many proposals and possibilities for building Fediverse tools that make it easy to run microblogging services that federate only with approved servers; we'd like to see Mastodon itself integrate these options, and also to see funding for and development of other approaches along these lines.

Cross-project governance tools

We'd like to see greater recognition of governance needs and trade-offs from core-software projects like Mastodon, but there is also room for third-party, ActivityPub-compliant but software-agnostic projects to be funded that provide enhanced governance mechanisms to Fediverse servers. Think of something like Loomio, but designed with Fediverse servers in mind.

Commodity hosting support

There is a conspicuous gap in the ecosystem around federated social media software: namely, there is not a robust selection of commodity hosts for Fediverse servers. This is due to a combination of factors: Mastodon is the most popular open source service and is expensive and resource-intensive to host; most Fediverse server software is not designed from a UX perspective with commodity hosting in mind (for example, a lot of configuration happens in text files a managed hosting customer would not have access to); there is no clear "second place" open source software winner in the Fediverse server space (there is demand for non-Mastodon software hosting but it's hard to point to one or even five software projects where most of that demand is concentrated).

It is necessary for developers to create user-friendly administrative configuration support for commodity hosted Fediverse software. Highly configurable GUI software like cPanel helped enable the web hosting boom, and Fediverse software should be similarly configurable to admins who are not Unix wizards with root access to their host servers.

Ways to Help Fediverse Members Find Well-Governed Servers

Many of the server teams we spoke with expressed concern that the Fediverse is at risk of losing many potential new members who would benefit from and enjoy the network's benefits, but who end up on centralized platforms instead. Our participants cited many reasons for this, including that it's easier to get started and rebuild a social graph on other networks, that other networks have built stronger user-bases within various social and professional communities, and that the Fediverse is experienced as hostile by some users and communities. All of these factors would benefit from more study and community work; we'll focus here on the lowest-hanging fruit, which is helping Fedi-curious people find their way onto well governed servers.

The first problem a potential new Fediverse member encounters is not learning how to use new software, but confronting a daunting step zero: "pick a server." Unfortunately, many potential members have no idea how to evaluate even the curated list of servers listed on JoinMastodon.org, a site run by Mastodon gGmbH: they can filter by region, topic, or registration approval method, but if they aren't going to choose at random, they'll have to click through to many potential servers and try to evaluate them based only on their brief public server rules. These rules rarely offer context crucial to understanding how a given server actually operates, including what the server information about the server's leadership team, whether it's governed from the top-down or in more democratic ways, what its uptime/performance stats and data retention policies are, and how it approaches moderation and the kinds of (de)federation decisions that significantly shape each member's experience of the Fediverse.

As importantly, novice members are given no guidance on what to look for in a server or how their choice of server will affect their experience of the Fediverse. Instead, they're assured that they'll be able to follow other accounts from any server (which isn't true if their server doesn't federate with the server that a friend is on) and that if they're unhappy with their choice, they can move later without losing followers—which is true, but fails to mention that they'll lose all their old posts, direct messages, and other settings and content if they switch servers.

Mastodon gGmbH tries to alleviate some of this complexity by steering new members to join its own server, Mastodon.social, which is an understandable choice. Joining that server, which is by far the network's largest, will prevent them from picking a server that immediately shuts down, leaving them stranded without their data or relationships, or from accidentally joining a server that will negligently expose them to other servers that host abusive accounts. But by shuttling new members into Mastodon.social, JoinMastodon.org puts them on a relatively lightly moderated server that works more like a centralized platform and offers fewer of the high-touch, high-context benefits offered by many well governed small- and medium-sized Fediverse servers.

We think the Fediverse would benefit from multiple new approaches to guiding potential members (and unsatisfied current ones) to servers that suit them and will help provide a positive experience of the Fediverse, and that it would be of immediate value to the network to fund community and infrastructure projects focused on this problem.