Julian Fietkau (2025): ORCID and the Fediverse: What Can We Do with Public Information? In: 23rd EUSSET Conf. on Computer-Supported Cooperative Work – Posters and Demos, DOI: 10.48340/ecscw2025_pd04

ORCID and the Fediverse: What Can We Do with Public Information?

Julian Fietkau
University of the Bundeswehr Munich, Germany
julian.fietkau@unibw.de

Abstract. ORCID is an identification scheme and bibliographic database for academics that aims to make information about researchers' works widely and easily accessible. The Fediverse is a collection of interoperable social media platforms where people can follow each other across platform boundaries to read and share text posts or other media. Between these two environments, we observe contrasting social norms around "public data" and conflicting expectations on how personal information may be stored and republished. During the early design phase of a tool to bridge ORCID data into the Fediverse and make individual ORCID records followable on open social platforms, we face a need to connect and resolve these differences to prevent avoidable conflicts. This article documents these norms and expectations as well as our approach to connect and bridge them.

1 Introduction

ORCID[1] ("Open Researcher and Contributor ID") maintains an opt-in database of academics and their published works. ORCID's goal is to enable clear identification of individual researchers, as well as to provide a hub for bibliographic data on each researcher's publications (through accessible data structures and APIs) in order to reduce duplicated data entry efforts. At this point, ORCID is very well established with about 15 million people having an ORCID record.

[1] https://orcid.org

The Fediverse is a loosely defined constellation of social media platforms running software such as Mastodon[2], Akkoma, or Pixelfed to host interoperable social communities. People can follow and interact with each other across server and software boundaries using the W3C standardized ActivityPub protocol (Lemmer-Webber et al., 2018). Most of the prominent Fediverse platforms resemble various commercial platforms such as Twitter or Facebook, although some experiment with different kinds of social design. Automated accounts ("bots") are hosted and used to make external data sources, such as weather information, accessible as followable social media accounts (Marinelli, 2025).

In 2024, inspired by Cope (2024), we began work on a software project to make public ORCID record data accessible and followable via ActivityPub. Any existing ORCID record would be mapped on-demand to a labeled ActivityPub bot account compatible with existing platforms. Whenever the researcher in question would add a new publication to their ORCID record, a summary thereof would appear in the social feeds of their ActivityPub followers. At the time of submission of this article, our project concept has just been made available to the general Fediverse user base and is open for public feedback. The initial responses have been highly supportive, with many academics expressing excitement about the possibility to follow ORCID records through social media.

From our personal experiences using both ORCID and ActivityPub services, we noticed some stark differences in how personal information that is ostensibly "public" is treated. (In this context, we will use "personal information" to refer to information that is attached to a natural person, without necessarily being private in nature.) For ORCID records, it is part of the platform's mission statement and pursuant to its purpose (which is to make bibliographic metadata as easy to share and acquire as possible) for personal bibliographic information to be publicly accessible as well as easily machine-readable and redistributable. We have been unable to find any cases of people objecting to external use of their ORCID record data. In contrast, people building technical systems that handle public ActivityPub data have on several occasions been met with hostility, which appears to be rooted in cultural norms around consent management for personal information access and republication permissions (Payne, 2024; Pincus, 2024). These norms are scarcely documented and have, to our knowledge, only been tangentially discussed in the research literature (Theophilos, 2024).

This work summarizes our insights into the differences between the two spheres, contrasts the technical norms codified in the ActivityPub specification with the cultural norms encountered on today's Fediverse, and showcases our attempt at mitigating anticipated conflicts around our integration of ORCID with the Fediverse by designing our system with transparency around data retention and data minimization. For the sake of clarity of scope, we point out that this text is not intended to present our software project in full. It focuses on insights regarding the

[2] https://joinmastodon.org/

expectations around personal information handling. For details on other aspects, we recommend consulting the project website[3].

2 Existing norms and practices

ORCID's self-declared understanding of what it means for record data to be public and how it may be republished and reused is detailed in their terms of use[4], their privacy policy[5] and their privacy settings documentation[6]. Notably, record holders are free to mark portions of their personal information as non-public. ORCID shares the public parts of its record data through live web views for humans, APIs for automated real-time access, and static files for long-term storage. They disclaim any copyright on this data and permit reuse without restrictions.

On public addressing, the official ActivityPub specification states that objects marked as public "shall be accessible to all users, without authentication" (Lemmer-Webber et al., 2018, section 5.6). Furthermore, it permits servers to send information to any other server participating in the network if it is addressed to the public (Lemmer-Webber et al., 2018, section 7.1.3), even if there is no specific recipient (e.g. "follower") there. Popular implementations regularly diverge from this in practice, for example by requiring authenticated requests even for public information (Barrett and nightpool, 2024, section 3.1). Blocking specific remote servers from accessing public information is a common administrative action (Theophilos, 2024). There is at least one notable project, the Website League[7], that promotes an allowlist-based ActivityPub network.

In practice, ActivityPub users are sometimes surprised to see their public personal information replicated in unfamiliar places, especially if those places are not well-established participants in the network, and doubly so if they forward information to network nodes outside of the ActivityPub space. To our knowledge, the disparity between the technical norms set by the protocol specification and the social expectations by people using ActivityPub platforms has not yet been researched systematically, but moments where conflicts around this topic have bubbled over have been reported in the enthusiast press (Tilley, 2024a,b, 2025) and further contextualized e.g. by Pincus (2024) and Payne (2024). Dash (2023) uses the consent-based data ownership stance as a springboard to outline a design for a privacy-respecting social media search engine. All these perspectives offer valuable insight about why people may feel protective about their social media profile and how and where it gets shared, even if the information in it is technically public. Stated reasons can be roughly summarized as a desire for posts to be publicly viewable, but not necessarily republishable; being driven by motivation to use the Fediverse as a means of escape from "big tech" surveillance and value

[3] https://encyclia.pub/

[4] https://info.orcid.org/terms-of-use/

[5] https://info.orcid.org/privacy-policy/

[6] https://support.orcid.org/hc/en-us/articles/25538671642519

[7] https://websiteleague.org/

extraction; and attempts to keep online social interactions accessible but localized, contained in specific social contexts for the sake of community safety.

It is worth noting that, while the current user base of the Fediverse leans privacy-conscious and skeptical of large corporate social platforms, the described stances on public profile information are far from unanimous. For every documented backlash against a software project with a perceived lack of respect for personal information autonomy, there has also always been vocal support for data discovery and reuse from different groups of people using these platforms. There are no reliable estimations for how prevalent each opinion is among Fediverse users and we are hesitant to guess without evidence, so all we can confidently say is that the privacy-focused position is common enough that neglecting it outright would not be pragmatically feasible, even if one were hypothetically willing to ignore any and all ethical concerns.

3 Our approach

Our software project started with the goal of bringing public ORCID data into the Fediverse. It is an open question how much, if at all, the cultural norms around data ownership summarized above could be extended to ORCID-style publication metadata, given that making this data easily publicly accessible and reusable is much more overtly "the point" of an ORCID record than of a social media profile. Nonetheless, while preparing the public launch of our project, we were motivated to reduce the likelihood of a possible escalation to the point of personal threats or harassment (Tilley, 2024a) as much as possible.

The central question posed by our bridge project was whether it should be opt-out or opt-in – that is, whether ORCID records should be accessible unless the owner indicates otherwise, or only made viewable after the owner has given informed consent. We performed a theoretical impact analysis[8] of the two possible choices and estimated the consequences for personal safety and for data ownership, as well as the potential benefits for ORCID record holders, and presented it to the community for public commentary and discussion in order to incorporate feedback before the activation of the bridge system itself. This process is ongoing at the time of submission.

We determined that the nature of the information contained in an ORCID record is substantially too different from a typical social media account for the personal risk assessment to be comparable. Social media posts may divulge information about interpersonal relationships, about personal life events, about geographical locations, organizational commitments, and so on. In contrast, the contents of an ORCID record offer only a much reduced and typically not real-time based perspective on a researcher's professional output. Hence, our conclusion was that making our ORCID bridge opt-out would not cause a

[8] https://encyclia.pub/optin-optout-analysis

discernible increase in risk for ORCID record holders, but that there were certain safety caveats that we intended to honor:

1. Our project assumes a self-imposed responsibility to ensure that the public ORCID record data we republish is as up-to-date as possible and that if someone removes a piece of information from their public ORCID record, that change is promptly reflected in our system.

2. We need to offer options to each ORCID record holder to perform "defederation"-type moderation actions for their own account. This enables them to quickly and accurately decide which ActivityPub servers are allowed to access their bridged ORCID record. Since ORCID offers a technical interface for ID verification, these settings can be offered on our project's website while keeping the barrier to entry low for ORCID record holders.

3. Furthermore, we restrict our platform from receiving free-form content (social media replies or mentions) from external ActivityPub servers. Countable discrete signals like favorites and shares are still aggregated, but free-form text from external actors (which may be abusive in nature) is not stored or forwarded to ORCID record holders. In this, we match ORCID's stance of not attempting to be a communication platform.

Along with taking a general stance of personal data minimization, our software thus matches or exceeds the capabilities for personal data ownership offered by popular Fediverse platforms. Initial discussions with several Fediverse safety experts and privacy advocates have been highly encouraging towards our approach, and have not resulted in significant criticisms or big adjustments to our platform's design. We are therefore confident that this approach is suitable to bridge personal information from an environment with very permissive norms around data ownership into one with more restrictive norms.

4 Conclusion

In this article we have examined the different expectations around the handling of publicly posted personal information in the ORCID ecosystem and the ActivityPub-based Fediverse, as well as the contrast in assumptions around this topic between the written ActivityPub W3C standard and the current lived Fediverse culture.

For tackling the problem of connecting the data flows between two contexts with such different cultural norms, we have showcased our approach of matching our software bridge's tooling to the more privacy-conscious side in order to best serve the interests of the people passively affected by our software. This gives them all the options they might desire to take ownership of how it presents their information while also shielding them from potentially malicious interactions.

Direct feedback from several Fediverse safety and privacy experts suggests that our approach does not have any immediately obvious flaws, and that the belief that consent-based, privacy-conscious design always necessitates an opt-in system is not necessarily accurate without exceptions. In the time since the public announcement of the project and the submission of this article, no objections have been raised to us on the basis of personal information handling. While it is not possible to prove whether any conflicts have in fact been prevented, we interpret this as a sign that our approach is suitable to reduce the risk of privacy-related conflict escalation between platform users and implementers. This embodies our current contribution to the ongoing conversation around the viability of the Fediverse as a mass medium (Struett et al., 2023).

Based on these conclusions, we are taking a slow approach with our bridge project, starting with only documentation and informational material, then moving to an opt-in testing phase with a small group of volunteers, before transitioning into the public opt-out design later in the year (assuming our plans do not change based on community feedback). As this is an ongoing case study, we are hopeful to document these experiences in the future.

Acknowledgments

We thank Esther Payne, Jon Pincus, and Emelia Smith for discussions and feedback about privacy, safety expectations, and data ownership on the Fediverse. Additional thanks to Laura Stojko for feedback on a draft version of this article.

References

Barrett, R. and nightpool (2024): ‘ActivityPub and HTTP Signatures’. W3C Social Web Incubator Community Group Draft Report. https://swicg.github.io/activitypub-http-signature/

Cope, A. (2024): ‘Holding Hands with the “Fediverse” – ActivityPub at SFO Museum’. SFO Museum. https://millsfield.sfomuseum.org/blog/2024/03/12/activitypub/

Dash, A. (2023): ‘How you could build a search that the fediverse would welcome’. anildash.com. https://www.anildash.com/2023/01/16/a-fediverse-search/

Lemmer-Webber, C., J. Tallon, E. Shepherd, A. Guy, and E. Prodromou (2018): ‘ActivityPub’. W3C Recommendation. https://www.w3.org/TR/activitypub/

Marinelli, S. (2025): ‘FediMeteo: How a Tiny €4 FreeBSD VPS Became a Global Weather Service for Thousands’. IT Notes. https://it-notes.dragas.net/2025/02/26/fedimeteo-how-a-tiny-freebsd-vps-became-a-global-weather-service-for-thousands/

Payne, E. (2024): ‘Consent and the Fediverse part deux: The Opt-out two-shuffle’. https://www.onepict.com/consentpartdeux20240215.html

Pincus, J. (2024): ‘Eight tips about consent for fediverse developers’. The Nexus of Privacy. https://privacy.thenexus.today/consent-for-fediverse-developers/

Struett, T., A. Sinnreich, P. Aufderheide, and R. Gehl (2023): ‘Can This Platform Survive? Governance Challenges for the Fediverse’. *SSRN Electronic Journal*, vol. August 2024. https://doi.org/10.2139/ssrn.4598303

Theophilos, J. A. (2024): ‘Closing the Door to Remain Open: The Politics of Openness and the Practices of Strategic Closure in the Fediverse’. *Social Media + Society*, vol. 10, no. 4. https://doi.org/10.1177/20563051241308323

Tilley, S. (2024a): ‘Content Nation Backlash Highlights Mastodon’s Toxicity’. WeDistribute. https://wedistribute.org/2024/03/contentnation-mastodons-toxicity/

Tilley, S. (2024b): ‘Maven Imported 1.12 Million Fediverse Posts’. WeDistribute. https://wedistribute.org/2024/06/maven-mastodon-posts/

Tilley, S. (2025): ‘Public Firehose Project Shutters After Backlash’. WeDistribute. https://wedistribute.org/2025/02/fedionfire-shutdown/