Analyzing IETF Data: Changing affiliations[1]

Nick Doty
Center for Democracy & Technology
29 September 2021

Technical standard-setting of protocols for the Internet and the Web can be laborious and difficult to understand, but it's also an intensely documented process. We should take advantage of that rich corpus to illuminate how and why Internet standard-setting works and how it can be more effective and more responsive to user needs and human rights. A potentially insightful subset of this data is the organizational affiliation of participants in technical standard-setting processes. While individuals may be the fundamental unit of the IETF's work,[2] the organizations and employers of participants influence how resources are allocated and investments made in different standards.

Which companies are seeing increased participation in the IETF? In what areas are participation and authorship dominated by particular large companies? How much of IETF standards and discussions come from corporate-affiliated individuals vs. academics or non-profit participants, and how has that changed over the IETF's history? (While some findings may sound obvious to deeply-engaged participants, even then there may still be valuable quantifications of those implicitly understood trends.)

Of key concern are the rare or missing voices of organizational types including consumer advocates, civil rights organizations, academics, or small companies. Having good data on where civil society has or has not engaged will let us test how public interest technologists can influence conversations on human rights values. And we can identify gaps (in individual working groups, larger areas, or whole SDOs) where further support of user advocates is needed.

Datatracker, mailing list and document data can also give us insight into *changes* in affiliation. Standard-setting fora in particular can be boundary organizations[3] where individuals from

---

[1] Yes, the title also reflects a personal, personnel update: I have recently changed my affiliation from UC Berkeley School of Information, where I completed my dissertation on technical standard-setting processes, to the Center for Democracy & Technology, where I'll contribute to civil society participation in supporting privacy and interoperability in Internet and Web standards.

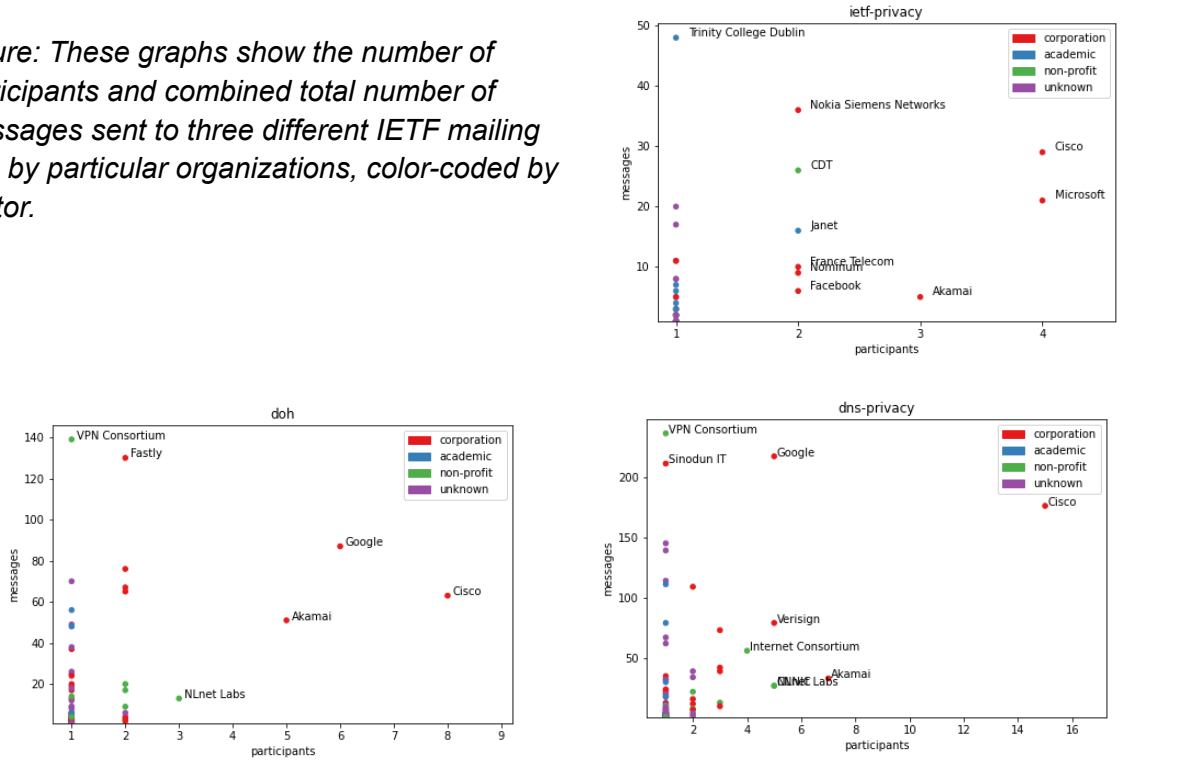
[2] https://datatracker.ietf.org/doc/html/rfc3935

[3] A concept from Guston: Guston, David H. 2001. "Boundary Organizations in Environmental Policy and Science: An Introduction." *Science, Technology, & Human Values* 26 (4): 399–408. http://www.jstor.org/stable/690161.
which we applied to technical standard-setting fora in: Doty, Nick, and Deirdre K. Mulligan. 2013. "Internet Multistakeholder Processes and Techno-Policy Standards: Initial Reflections on Privacy at the World Wide Web Consortium." *Journal on Telecommunications and High Technology Law* 11. http://www.jthtl.org/content/articles/V11I1/JTHTLv11i1_MulliganDoty.PDF.

different organizations and different types of organizations interact, negotiate and share ideas. But because individuals may have significant latitude, expertise and personal perspectives that influence their work and the outcome of standard-setting discussions,[4] where an individual moves between organizations may indicate both how particular concepts or approaches are shifting between organizations (perhaps contributing to institutional isomorphism[5]) and how organizations may use hiring to influence standard-setting discussions.

Affiliation data is, like almost all data, messy and sometimes missing. At a recent IETF hackathon,[6] we used email domains, IETF meeting attendance records, GitHub profile data and IETF draft authorship data (via the Datatracker) to infer organizational affiliation for participants in a few different IETF group mailing lists. Entity resolution of organizations is necessary and non-trivial and sectoral (or industry) classification may have to be done manually but initial results suggest that this data analysis is feasible. There are interesting and non-obvious patterns in how many participants from a particular organization participate in IETF groups and how active they are on mailing lists.

Figure: These graphs show the number of participants and combined total number of messages sent to three different IETF mailing lists by particular organizations, color-coded by sector.



[4] Findings related to this difference between individuals and organizations are detailed here: https://npdoty.name/writing/enacting-privacy/themes/individuals-organizations.html, a part of: Nick Doty. *Enacting Privacy in Internet Standards*. Ph.D. dissertation. Advisor: Deirdre K. Mulligan. University of California, Berkeley. 2020. https://npdoty.name/enacting-privacy/.

[5] Famously, from: DiMaggio, P J, and W W Powell. 1983. “The Iron Cage Revisited: Institutional Isomorphism and Collective Rationality in Organizational Fields.” *American Sociological Review* 48 (2): 147–60.

[6] See BigBang presentation from IETF 111 Hackathon: https://github.com/IETF-Hackathon/ietf111-project-presentations/blob/main/ietf111-hackathon-bigbang.pdf

Ethical and privacy considerations of research

Organizational affiliation data has typically been made public during IETF participation: in publicly-archived mailing lists, in IETF attendance records, and in document authorship credits. That data is publicly accessible does not mean that people may not have privacy interests in that data. However, these IETF data sources are known by the participants to be publicly accessible (indeed, they are regularly accessed by the participants themselves in other contexts) and one intended purpose of publicly documenting standard-setting processes is to provide for external transparency and review.

As we collect more information on individuals, demographics, changing affiliations over time, and other characteristics, we may want to refrain from making those individual records public, in much the way that we might not want to mirror (even public) archives, as there may be a privacy interest in not having aggregated or inferred information about a particular individual more easily publicly accessible. Access controls and simple agreements for sharing data among researchers without publishing all such data may be a practical and respectful way forward. BigBang is starting to apply such a process in sharing collected mailing list archives; feedback on such approaches would be welcome.

Some next steps

BigBang provides tools for gathering, parsing and analyzing mailing list archives (and some other online community traces).[7] We can add libraries to efficiently use the Datatracker and other datasets to merge with mailing list activity and document publication data to measure and visualize organizational involvement and diversity.

As we add tools for inferring organizational affiliations over time, these techniques can be applied not only to IETF, but also W3C, 3GPP and other SDOs and organizations and some individuals will bridge those fora. For the sake of collaboration, it might be useful to have a simple common format for information on participants, including affiliations at points of time, so we can share findings from across different standard-setting or other Internet governance venues.

Finally, are there other data sets or existing tools we can use for understanding and analyzing the organizations that mediate participation in governance fora? Data on companies and their financial status, location, industry sector, etc. could all be useful supplements.

[7] https://github.com/datactive/bigbang