

MATTHIAS C. KETTEMANN, JOSEFA FRANCKE, CHRISTINA DINAR AND LENA HINRICHS (EDS.)

Platform://Democracy

Research Report Americas



PLATFORM://DEMOCRACY

Platform://Democracy

Perspectives on Platform Power, Public Values and the Potential of Social Media Councils: Research Report Americas

edited by Matthias C. Kettemann, Heidi Tworek and Josefa Francke

LEIBNIZ INSTITUTE FOR MEDIA RESEARCH | HANS-BREDOW-INSTITUT, HAMBURG, GERMANY
HUMBOLDT INSTITUTE FOR INTERNET AND SOCIETY, BERLIN, GERMANY

Cite as: Matthias C. Kettemann, Heidi Tworek and Josefa Francke (eds.), Platform://Democracy – Perspectives on Platform Power, Public Values and the Potential of Social Media Councils: Research Report Americas (Hamburg: Verlag Hans-Bredow-Institut, 2023)

CC BY SA 4.0

This publication is part of the project *Platform://Democracy: Platform Councils as Tools to Democratize Hybrid Online Orders*. The project was carried out by the Leibniz Institute for Media | Hans-Bredow-Institut, Hamburg, the Alexander von Humboldt Institute for Internet and Society, Berlin, and the Department of Theory and Future of Law of the University of Innsbruck und funded by Stiftung Mercator.

Publisher: Leibniz Institut für Medienforschung | Hans-Bredow-Institut (HBI)
Rothenbaumchaussee 36, 20148 Hamburg
Tel. (+49 40) 45 02 17-0, info@leibniz-hbi.de, www.leibniz-hbi.de

Contributors

Name(s)	Affiliation
Heidi Tworek	University of British Columbia, Canada
Luca Belli	Fundação Getulio Vargas Law School, Brasil
Katie Harbarth	Anchor Change, USA
Kate Klonick	St. Johns University Law School, USA
Emma Llansó	Centre for Democracy and Technology, USA
David Morar	Open Technology Institute, USA
Aviv Ovadya	Berkman Klein Center, USA
Peter Routhier	Internet Archive, USA
Fabro Steibel	Institute for Technology and Society, Brasil
Alicia Wanless	Carnegie Endowment for International Peace, USA

Table of Contents

Contributors.....	3
Table of Contents.....	4
Public Values and Private Orders in Social Media Councils – Perspectives from the Americas	6
Platform Councils: Solving or Creating Regulatory Vulnerabilities? A Brazilian Perspective	9
Platform councils and their diverse rationales	9
Contextualising Brazilian evolutions	10
Platform councils: Regulatory trick or treat?	10
Conclusion	12
How Platform Councils Can Bridge Civil Society and Tech Companies	13
Unique Role of Platform Councils	13
The Evolution of Social Media Councils.....	16
Introduction	16
History of Formal and Informal Speech Platform Relationships with Social Media Councils	17
Current Web 2.0 Multistakeholder Intervention	19
Current Moment and the Lessons from the Past	19
Technical Difficulties: Incorporating independent technical expertise into platform council decision-making	21
Finding Technical Expertise	21
Tensions and Tradeoffs	22
Models for Incorporating Technical Expertise	23
Conclusion	25
Enforcement as a necessity in platform councils.....	27
Critiques of similar institutions without strong enforcement	27
Blue/Greenwashing	27
Designed/heavily influenced by businesses	28
Lack of potential to grow stronger	28
Meta Oversight Board	29
Enforcement critique	29
Conclusions	30
Interoperable Platform Democracy: How deliberative democratic processes commissioned by corporations can interact with nation-state, multilateral, and multistakeholder decision-making.....	31
What is platform democracy?	31
Interoperability with existing institutions	34
Impacts of platform democracy outputs	34

Conflict with platform democracy outputs	36
Inputs to platform democracy	37
Why we might want platform democracy	37
Public Values and the Private Internet	39
Introduction	39
Social Media Councils and the Broader Internet	39
An Internet with Public Interest Values	40
Conclusion	41
Soft Power and Platform Democracy: How social media councils could shape government and corporate strategies and preferences.....	42
What is soft power?	42
The elements of soft power in platform democracies	43
What next on framing platform democracy's soft power?	45
A Council for Consilience: How could a council foster a field researching the information environment?	46
What's the problem?	46
It takes a village...	47
...or a village council	48
Conclusion	52

Public Values and Private Orders in Social Media Councils – Perspectives from the Americas

Heidi Tworek

UNIVERSITY OF BRISITH COLUMBIA, VANCOUVER, CANADA

It is a platitude now to emphasize the need to insert civil society voices and public values into social media platforms' decision-making. What is not a platitude is thinking through how to make this happen concretely. The nine papers in this collection from the Americas regional clinic provide concrete and specific insights into how institutions like social media councils might inject public values into the private order of platforms.

Social media councils are not a new idea. The council part of social media councils goes back a century, if not more, depending upon whether we want to count councils as similar to commissions. The social media aspect of councils emerged in what Kate Klonick sees as the second phase of multi-stakeholder involvement in platforms. This occurred from around 2016 with multi-stakeholder intervention rather than the model of multi-stakeholder influence that held sway from 2006 to 2016. Twitter, for instance, announced its Trust & Safety Council in February 2016.

From around 2018, researchers and civil society organizations suggested that councils could not just be based within platforms, but could embed public values and broader concerns around freedom of expression. Pierre François Donquir, then at Article 19, suggested that social media councils could entrench human rights into content moderation processes. At a similar time, I started to write about the potential for social media councils in Canada (what Chris Tenove, Fenwick McKelvey, and I called a "Moderation Standards Council") or as one institution that could work on both sides of the Atlantic to resolve disputes around content moderation in a fair, accountable, independent, transparent, and effective way.

Much of this discussion initially intertwined with debates around Germany's Network Enforcement Law (NetzDG), which came into force in 2018. Some considered whether there were other methods beyond government requirements for social media companies to take down speech flagged by users as potentially illegal within 24 hours. Others highlighted the many speech issues that takedowns could not solve, such as the intransparency of recommender systems.

As more regulation has emerged, including with the Digital Services Act in the European Union, the discussion around social media councils has continued. Highly prominent self-regulated councils, principally the Facebook Oversight Board, also came into being. This sparked new debates about the limits of both self-regulation and global content moderation councils. It also highlighted the ultimately private nature of the Board, which contrasted with the spirit of suggestions for social media councils to incorporate broader public values.

The nine papers of this collection provide nuanced and practical views on how to incorporate those public values into suggestions for social media councils, now that we live in a world of increasing regulation and platform self-regulation. As the papers show, there is much work to be done.

Fabro Steibel's paper points out that policy framing around social media councils could shape their values and the institutions themselves. Values such as human rights and digital constitutionalism have discursive power; how we talk about the values behind social media councils can shape their institutionalization.

If we discuss social media councils through public values, then, we might also remember the contested nature of the idea of a “public” or to use the classic phrase “public sphere.” Jürgen Habermas’ ideas around the evolution of a “public sphere” in the 18th century has long been criticized for focusing on white, middle-class men. The critiques of Habermas over the last thirty years are one example of how definitions of publics and public values evolve over time. Feminist and class critiques in the early 1990s, a few years after Habermas’ work was translated into English, noted that women and working-class milieus formed publics and held their own values too, even if they didn’t fit into Habermas’ normative framework. More recently, Wendy Willems has pointed out that Habermas’ concept of a bourgeois public sphere elided slavery and colonialism, reproducing racialized hierarchies without even mentioning them. The silences in the canonical text showcase the importance of not assuming that any one voice defines “the public” or public values.

Peter Routhier’s and Aviv Ovadya’s papers take up the challenge of inserting the public in social media councils in two different ways. Ovadya’s paper takes a pragmatic approach to inserting democracy into platforms through processes of deliberative democracy such as sortition or temporary advisory bodies. Drawing from a broader literature on democratic processes like citizens’ assemblies, Ovadya suggests how platforms themselves could democratize their processes, even before public institutions step in to regulate.

Meanwhile, Routhier’s paper suggests that we continue to think more broadly about public interest values as a constitutive part of our online lives. He warns that in the worst case, social media councils could legitimize and entrench current social media platforms. Rather than rely solely on councils, Routhier provides broader ideas of how to embed public interests into the internet, for example by empowering libraries and non-profit sites like Wikipedia.

Although social media councils could offer the chance to bring publics into private orders, Emma Llanso, Katie Harbath, and Alicia Wanless provide important insights into areas where expertise may be necessary. Llanso’s paper considers whether and how to incorporate technical expertise into any such councils. Most people, including researchers and policy-makers, may not understand the technical infrastructure of platforms. Llanso offers practical ideas on how to include technical expertise into a council’s design and operations. Meanwhile, Katie Harbath’s paper sees councils as a space of bridge-building, potentially between platforms and civil society. This could be a crucial role for a council, given that only employees and former employees have access to certain forms of knowledge.

Wanless offers a complementary perspective on the challenges of building councils or council-like bodies in a participatory way. She reflects on the (ongoing) process of creating a CERN for the information environment. Such an international body would foster interdisciplinary research, so vital given the vast number of open questions around platforms’ influence on democracies. But it also presents challenges of how to build an entity that is bottom-up, while still having some coherent direction.

The final three papers consider the implementation of any council model. David Morar points out the limitations of any models that do not embed sufficient capacity for enforcement. Using examples from other industries, Morar lays out the pitfalls of weak enforcement, including greenwashing or bluewashing (“functionally toothless labelling schemes” on human rights and labor standards). Morar demonstrates the necessity of enforcement mechanisms, while suggesting that any council will find it hard to enforce its decisions if it does not incorporate platforms in some way.

For Luca Belli, such considerations help to explain why regulation will be required to insert public values into platforms. To solve problems, Belli argues, legislators would have to design social media councils to be “meaningfully accountable” instead of self-regulatory. Finally, though with very different

caveats, Kate Klonick warns that more voluntary social media councils may swiftly fall victim to cost-cutting during economic downturns. Her paper thus suggests mandates for any public-private versions of social media councils to avoid sudden dissolutions of voluntary bodies.

Overall, the nine papers from the Americas regional clinic tackle the practicalities of the oft-proclaimed idea that publics deserve to be involved in platforms' decision-making. Some papers offer pragmatic solutions for the present, while others point to future modes of institution-building. Still others remind us of pitfalls to avoid. They do not pretend that there is a panacea to the privatized order of the present. And they do not come to a consensus on one route ahead. But they show that there are many more paths forward than the pessimists might postulate.

Platform Councils: Solving or Creating Regulatory Vulnerabilities? A Brazilian Perspective

Luca Belli

FUNDAÇÃO GETULIO VARGAS LAW SCHOOL, RIO DE JANEIRO, BRASIL

An old Soviet adage recommends that “If you do not want to solve a problem, then create a commission.” The hypothesis of this brief paper is that platform councils may not necessarily solve digital platforms’ deficit of democratic values accountability, and they might contribute to creating further regulatory vulnerabilities, unless legislators design them to be meaningfully accountable.

To this end, this essay provides a brief and non-exhaustive introduction to the types of platform councils developed so far. Subsequently, it explores the existing Brazilian platform regulation framework, and utilises the three main democratic values included in the title of the Brazilian Bill for Platform Regulation i.e., “freedom, responsibility and transparency”, to analyse how far platform councils might be useful to foster democratic values. Lastly, the conclusion highlights some relevant caveats, regarding the effectiveness – and, ultimately, advisability – of relying on self-regulation versus public regulation to establish platform councils.

Platform councils and their diverse rationales

As highlighted by the Platforms and Democracy position papers, several types of digital platforms, besides social media, have been experimenting with platform councils over the past five years. The most renowned example is Facebook’s Oversight Board, which has established such a body to – supposedly – regulate the content moderation practice of the platform in a more participatory and accountable manner.

That is by no means the only existing example. Platform scholars are aware of gaming platforms experimenting with player councils, such as EVE Online’s Council of Stellar Management, a group consisting of ten EVE Online players democratically elected by community of game players, to advise and assist the continuous development of the platform, provide analysis, share suggestions, and give feedback.

A less-known instance is the one created by the Chinese data protection law, the Protection of Personal Information Law (PIPL), whose article 58 establishes an obligation for large platform providers – defined as those having “a huge number of users and complex business models” –to establish an “independent supervision board” composed by external members, responsible for monitoring the correct implementation of the law.

This normative provision represents a remarkably original feature among data protection frameworks that, together with the large platforms’ obligation to periodically publish reports on their data processing activities, also foreseen by PIPL article 58 aims at increasing accountability through what is described as “society’s supervision.”

The increase of platform accountability, in truly meaningful ways, ideally embedding democratic values and oversight in these private entities, has indeed been a recurrent preoccupation of scholars and policymakers alike for the past decade. Such preoccupation has been repeatedly expressed, for instance, by Brazilian policymakers that, since 2020, have been discussing policy efforts aimed at regulating digital platforms, although without reaching a consensus on how to do so, at the time of this writing.

Contextualising Brazilian evolutions

An important element of context regarding Brazilian digital platforms regulation is that social media platforms are already regulated by the Brazilian Civil Rights Framework for the Internet, Law n. 12,965/2014, a.k.a. “[Marco Civil da Internet](#)” (MCI), which the Brazilian Congress wishes to supplement with [Draft Bill n. 2,630/2020, on Freedom, Responsibility and Transparency on the Internet](#), a.k.a. the “Fake News Bill.”

The MCI is Brazil’s primary law regulating the Internet and the first and only general law for Internet governance and Internet rights adopted in Latin America. MCI article 19 establishes a general regime of a judicial notice-and-takedown system where “application providers,” i.e. platforms, are deemed liable for user-generated content only if failing to comply with court orders for the removal of specified content within 24 hours, granted they have the technical capacity to do so.

Hence, the Brazilian legislator has framed an approach to regulate digital platforms, although this initial approach has shown limits. The constitutionality of MCI article 19 is currently [challenged at the Brazilian Supreme Court level](#). Plus both the Brazilian Legislative and Executive powers are actively pursuing efforts to supplement the MCI provision with specific normative provisions aimed at improving “freedom, responsibility and transparency”.

The Brazilian policymaking initiatives do not mention councils, but for our discussion the Brazilian experience is particularly relevant, as it allows us to understand the democratic values that the Brazilian nation, through its democratically elected Congress, considers as the most relevant and in need of protection, when it comes to regulating platforms.

Democratic values, which could be supposedly baked into platform governance architectures through platform councils, are a very large and heterogeneous set of values. The Brazilian Legislator deems “freedom, responsibility and transparency” as so relevant for platform regulation – and clearly missing from platforms private orderings – to include them in the very name of the Bill that aimed at regulating them. This inclusion leads us to assume that such values are the most relevant ones from the Brazilian perspective.

It is not clear, however, how and to what extent the establishment of platform councils could strengthen the aforementioned values. To provide an initial answer to such query, the following section will briefly analyse how platform councils could be used and designed to positively contribute to improving Brazilian democratic values.

Platform councils: Regulatory trick or treat?

For the sake of clarity and conciseness, this essay will only examine three specific elements of the entire spectrum of democratic values – i.e., freedom, responsibility, and transparency – based on the importance that they seem to play in the Brazilian context. This section provides some observations and recommendations on how each of these democratic dimensions could fit into platform governance through the action of a platform council.

Freedom

This first item evokes the full spectrum of fundamental freedoms granted to each and every individual at the domestic level by national constitutions and at the international level by binding international law frameworks.

From a pragmatic standpoint, it seems highly unlikely that a sound protection of fundamental freedoms can be granted by solely relying on global international law frameworks, such as the ICCPR or ICSECR, as this type of international law framework inevitably relies on national systems to specify and implement normative provisions and, critically, their exceptions. Importantly, the subject of international law obligations are states (i.e., public bodies which must guarantee the full enjoyment of rights to physical persons). Thus, international jurisprudence, while offering important guidance, might be of limited use when trying to establish how corporations (i.e., legal persons) must respect individual rights.

Regional fundamental rights frameworks exist, and in some regions, might be more active than others, even regulating the behaviours of corporate actors, but they are typically renowned for their lethargic processes. In this perspective, the establishment of regional platform councils might be an interesting option to translate the existing regional approach from fundamental rights to guidelines for responsible platform behaviour, as I will discuss in the next point.

Lastly, national constitutional law and domestic jurisprudence are usually well suited to address the specificities of local culture, particularly as regards local juridical sensitivities and traditions. Therefore, if any national platform council must be established, I suggest they should be created at a national level, as convincingly argued also by other authors. Such national bodies could coordinate at the regional level if needed, considering that most international disputes may typically occur at the regional level.

Responsibility

As suggested in the previous paragraph, one of the core elements of a responsible behaviour from private entities in general and digital platforms in particular is the respect of fundamental rights, but also the provision of effective remedies, strongly recommended by the UN Principles for Business and Human Rights.

In this respect, platform councils might serve as a useful additional mechanism for users to seek redress when any of their rights is unduly violated. Considering the existence and notable advancement in terms of sophistication of Online Dispute Resolution (ODR) mechanisms, it seems that national platform councils could be designed to be appellate bodies of the existing ODR mechanism, so that ideally users are granted the most effective and just resolution of their potential controversies.

The ancillary benefit of establishing a system of a national platform council acting as an appeal body would be the decongestion of national juridical systems as regard platform disputes. This could simultaneously improve the full enjoyment of platform users' rights, increase the (corporate social) responsibility of platforms, and positively contribute to alleviating notoriously overburdened national judicial systems.

Transparency

This point deserves special attention as transparency is frequently touted as a sort of silver bullet, able to solve – or at least contribute to solving – a wide range of issues. But de facto transparency may not be as effective as we might think, being usually very poorly defined and even more poorly implemented.

A telling example in this regard are platform terms of service, which should supposedly be a tool of transparency – and indeed are considered so by data protection frameworks around the world. Yet, these end up as instruments cleverly engineered to mislead and confuse the user with lengthy and highly technical terms, without providing any meaningful information that could increase accountability for the provider and oversight for the user – and society.

In this respect also the establishment of platform councils may be helpful, should such organs be mandated to publish regularly reports on specific issues – e.g. content moderation, product defects, game features, etc – that enable an increased understanding, monitoring and, ultimately accountability of these private actors. Importantly, to achieve a situation of meaningful transparency, such reporting should also adopt shared – ideally standardised – formatting requirements so that data can be more easily compared and studied by regulators, researchers and users, and even be machine readable.

Conclusion

While a good dose of scepticism regarding the potential benefits of platform councils must be of order, this brief analysis has demonstrated that platform councils might prove to be useful, largely depending on how they are designed. A system composed of national and regional platform councils might prove to be an interesting choice as they could usefully assist in guaranteeing users' rights and provision of effective remedies, as well as improving platform transparency and, consequently, accountability.

However, a very strong caveat is that such system – and whatever other platform council system – would have a cost. Such a cost might be easily borne by large platforms, as explained by the Chinese legislator's choice to only target large platform with its obligation to establish "independent supervision boards." Small players and, particularly, new entrants in the platform business would be very unlikely to have the resources to establish such a complex and costly system and, therefore, stakeholders' expectations as to the impact of platform councils should be lowered to a minimum, adopting the most pragmatic approach possible.

Lastly, stakeholders should be aware that should the design and implementation of platform council be delegated to self-regulation, with no specific indications from legislators on how they should be established, their roles, responsibilities, and accountability, platforms might use such bodies as a strategy to avoid or postpone important regulatory action.

Indeed, as long as national law does not mandate the establishment and regulation of such bodies, any large platform executive would have a fiduciary obligation towards shareholders – which exist in every country for every publicly traded company, such as most large platforms – to prioritise the maximisation of their profits rather than the full enjoyment of user rights. In this perspective, it seems reasonable to posit that national platform councils regulated by domestic law could play a useful role, improving platform governance, while it seems at best naïve to argue that global platform councils, purely based on self-regulation will be able to offer any meaningful solution.

How Platform Councils Can Bridge Civil Society and Tech Companies

Katie Harbarth

ANCHOR CHANGE, WASHINGTON D.C., USA

Meta's Oversight Board is a first-of-its-kind experiment where a platform council can overturn the company's decisions on content. However, its powers are limited to individual pieces of content, though the Board can make policy recommendations to Meta and serves an important oversight role to hold Meta accountable for its promises.

While the Oversight Board is first in many ways, in others, it is not. For over a decade, platforms have pulled together groups of individuals with various backgrounds to provide guidance and expert advice on content policies, product development, human rights standards, regional and cultural differences, and many more. While these groups can counsel companies on their work, they have no power to force them to do anything. This lack of checks and balances on the companies and their decisions lies at the crux of many of the arguments debated today about holding tech companies accountable.

Moreover, global civil society groups have documented ways platforms can harm society for many years. They regularly lament that the platforms do not listen to them, nor do they even know how to contact someone who works at each company.

Meanwhile, the platforms struggle because if they are small to medium size, they likely don't have even one person dedicated to being a liaison with the civil society community. Even large platforms with multiple employees will struggle to manage requests from the thousands of groups around the world who want to talk to them.

This is where platform councils can play an important bridging role between the expertise of these two communities. These councils can act as a connector, help to translate civil society concerns into things that tech companies understand how to adopt, help civil society understand what's possible technically, help mediate competing priorities, create frameworks for engagement and provide valuable oversight on both communities.

This paper will explore all the roles platform councils could play and what they would need to be effective.

Unique Role of Platform Councils

Building a platform that protects freedom of expression but protects people from harm is a simple thing to say but an incredibly complicated balance to put into practice.

It goes beyond debating what a platform's policies should be on the types of content and behavior they do or do not allow. It also encompasses the following:

- What should the penalties be for violating the rules?
- Should there ever be exceptions, such as for newsworthiness?
- What the local law requires.

- How companies should respond to pressure from the government to remove content – especially when employees' safety might be at risk if they do not comply.
- What to do about borderline content that doesn't violate but is still problematic?
- The platform's ability to enforce said policies – especially at scale.
- The design of the product itself – including how machine learning algorithms are built.
- Transparency into the decision-making processes and an ability for researchers to study how the platforms work and their impact on users.
- And many more aspects.

Debates have ensued for years about how various platforms should answer these questions. Platform councils can potentially play a role in settling these debates by playing the role of:

- **Bridge.** Often, civil society organizations do not know how to engage tech companies – nor have the resources to talk to all of them, and vice versa many tech companies don't know how to reach out to civil society organizations nor have the resources to handle incoming from everyone who might want to engage them. Platform councils can help bridge this gap by organizing civil society concerns in a way that those at the platforms can easily digest. They can also help the tech platforms explain more about how they operate to the civil society organizations.
- **Oversight.** It's not enough for any organization to just promise that they will do something. They need to be held accountable on how they follow through. Platform councils can stay on top of tech companies and civil society groups to understand if their actions match their promises. Regular transparency reports can also help people understand how different companies compare in their efforts and if they are making regular progress.
- **Frameworks.** One of the biggest challenges for both civil society and technology companies is that they don't know how to speak to each other in a way that each will understand. Civil society often doesn't understand how the platforms work, and tech employees don't necessarily understand the human rights world. Platform councils can help to develop frameworks to facilitate conversations and collaboration. An example of how this works could be how platforms prioritize which elections they will work on. A platform council could work with tech companies to understand the various considerations and resources they have to work with and how they prioritize this work today. A platform council could make recommendations on various data sources the companies should use – such as [Variety of Democracies](#) – as well as other considerations based on input from civil society.
- **Reconciling priorities.** While better collaboration is needed, that doesn't necessarily mean that everyone will agree on the prioritization in which certain issues are worked on at the companies. Not everything can be done at the same time and reasonable people will disagree on where to draw the lines. Platform councils can play a role in reconciling those differences to help provide guidance on which things are more pressing than others.

To be effective in playing this role described above, the structure and scope of the platform councils must also be considered. Some crucial aspects would be needed:

- **Across platforms.** These problems are not siloed to any platform. To be effective and to ensure consistency where needed a platform council would need oversight and cooperation across various tech platforms and companies.
- **Product and policy oversight.** One major flaw in the Meta Oversight Board structure is that they only have the ability to tell the company to leave up or take down a particular piece of

content. They can make policy recommendations, but the company does not have to follow them. Nor can the Board make any recommendations on the product itself, such as how it is designed or works. To be effective platform councils would need to be able to participate in these areas too.

- **Diversity of backgrounds and skills.** The final important thing would be the make-up of the council itself. The people must be diverse not just in race and gender but also in geography, ideology, and skill sets. You need human rights experts, former platform employees, former government officials, journalists and others to adequately balance decision making.
- **Financing.** Both the platform council and civil society group would need to be compensated for this work. Ideally, funding would come not just from the companies themselves, but also from other philanthropic money. Government funding from USAID or similar agencies could also be considered. These funds would be put into a trust not controlled by the companies that could then fund the work of the platform council but also be used as stipends for the civil society groups who participate.

While platform councils have been around for a while as advisory groups, only recently, with the [Meta Oversight Board](#) do they have any power to overturn the decision of the platforms. Even that is very limited so councils do need to be reimagined and given real powers to be effective. By doing this a platform council can provide many meaningful roles, including being a bridge between the expertise that the tech companies have and civil society. By serving in this role they can fill a gap that continues to persist to this day.

The Evolution of Social Media Councils

Kate Klonick

ST JOHNS UNIVERSITY LAW SCHOOL, NEW YORK, USA

Introduction

Jack Balkin's Free Speech Triangle, provides a useful heuristic for understanding the old and new powers in tension to control free expression. Before the internet, the old model of free expression in a democratic society, he argues, was not a triangle at all, but a dyadic model between the State, which threatened to censor using its police power, and its citizens, who used voting, voice, exit, and protest as a means of pushing back against that power. The internet, Balkin argues, ushered in a new corner – and hence the triangle – radically altering the power structure between State and citizens. In this “new school” world, online private speech platforms diminished the power of the state to censor, allowing users to route around such controls via these new communications technologies. But this shift of routing around law was not only in citizens’ favor; it also enabled States to circumvent limits on their ability to surveil citizens by using private platforms.

The Free Speech Triangle is useful because it also provides a framework for mapping the current limitations of each corner’s power, and how the dangers of certain remedies can exacerbate power inequities particularly for the group we want to empower the most in a democracy: users/citizens.

For instance, users/citizens have little legal recourse against private online platforms, though they do ostensibly have the power of protest, boycott, and reputational shaming. An easy solution might seem to be to empower the law against online platforms on behalf of users/citizens. The irony is that doing so often requires state regulation of speech – that is, using the state to police the online platforms. But state control of speech brings us straight back to the dyadic model, pre-triangle. Especially when coupled with the private surveillance power that platforms enable, state control of platforms is an incredibly dangerous proposition. Indeed, to avoid regulation which might damage their business model, platforms might be more willing to cooperate with governments, despite potential harm to users. Though calling for regulation of a powerful industry might seem a sensible and obvious move, in fact, driving these two nodes of power together might be the most dangerous outcome for users/citizens.

What the triangle reveals is that the problem we’re trying to solve for – empowering users/citizens voices and their say in the private structures that control and govern those voices – is maybe not best answered by traditional solutions of regulation through governments. Instead, a new kind of solution might be necessary to strengthen users/citizens’ existing powers against private platform governance. This could take the forms of markets, norm enforcement, or, as this essay will argue, the creation of institutions to stand in for the direct representative capacity of individual users/citizens. One such mechanism to do this is social media councils, multi-stakeholder constituencies that themselves reflect the new pluralistic environment within which speech rights now exist.

This paper will reflect briefly on the history of social media councils and their varied successes in fulfilling this role as a point of leverage for users/citizens on speech platforms. It will conclude with a few key lessons from these experiments.

History of Formal and Informal Speech Platform Relationships with Social Media Councils

Multi-stakeholder influence has existed at platforms since the very beginning of social media. For the sake of simplicity, I'll roughly group the nature of this influence into two eras: the 2006-2016 era, which I'll call early "Early Web 2.0" and then 2016 to present, which I'll call "Current Web 2.0."

Early Web 2.0 Multi-stakeholder Influence 2006-2016

For the large and now dominant user-generated content and social media platforms – Facebook, YouTube, Twitter – 2006 is a crucial year. In 2006, Facebook first became publicly available to a global audience and YouTube was created. While MySpace, LiveJournal, Orkut, Friendster, and many other platforms already existed, YouTube and Facebook are the most relevant examples to draw from because:

- They are still dominant user-generated content platforms making them highly relevant to both government and private entities;
- They are still dominant social media platforms making them highly visible in media and society;
- They operate globally;
- For better or worse, their content moderation practices have become the most followed models for U.S. social media companies.

As early trust and safety employees at Facebook and YouTube encountered issues around content moderation for non-intellectual property content, they developed standards based on American norms around freedom of expression. These later developed into more formal rules (and exceptions to rules) formed recursively in a common-law like pattern in response to new fact patterns of content moderation that presented themselves over time.

Though the early internet is often characterized in tech utopian terms by many early scholars and technologists for the freedom of speech and access to information that it provided, research and reflection has shown that the "Wild Wild West" of the early internet also enabled a massive amount of harmful behavior. Unfortunately such behavior – hate speech, harassment, stalking, doxxing, racism, defamation, misogyny, antisemitism – was underreported and underacknowledged by many high-profile academics and policy-makers at the time, in part, unsurprisingly, because such bad acts targeted historically vulnerable populations.

But two groups of individuals had front row seats to these ugly sides of the internet: the individuals at platforms working to remove such content or prevent its posting, and civil society groups that already specialized in helping individuals exposed to such harms in their offline capacity.

It took time, but eventually these groups found each other and began to become informed by each other. In particular, well-funded and respected organizations like the Anti-Defamation League, Electronic Frontier Foundation, and ACLU had early impact in bringing together individuals working on the internal content moderation policy at platforms and helping them collaborate to form best practices. Those in charge of content moderation and trust and safety at platforms were eager for external advice on how to set community standards for their sites and manage trust and safety flows for a number of practical reasons:

- Generally speaking these employees' roles were relatively low profile in the company, and at the time operated independently. Their reasons for joining tech platforms were social planner motivated and not tied to or in answer to revenue generation directly.

- These employees lacked both practical, legal, and philosophical expertise in these areas of harmful speech, or how to make tradeoffs between safety and principles of freedom of expression and democracy. They welcomed input from representatives of people affected by their policies and rules.
- As the influence of content moderation grew along with the size and scale of the platforms, recognizing and accepting input from expert groups seemed not only a practically helpful but more legitimate means of setting global speech policy.
- At scale it was easier and more efficient to speak to civil society, academic, and government stakeholders who served as gate-keepers, than to individuals that were affected by such policies.

Though early collaborations between such groups and platforms were at first informal, platforms formalized these relationships both by creating group councils and by formally empowering individual stakeholders.

Group Councils

Formalization of multi-stakeholder groups came in the form most recognizable today at Social Media Councils. The role and public face of such councils varied. Some, like the Twitter Trust & Safety Council, were private groups of expert individuals convened by the platforms to be on call to answer questions on particularly problematic pieces of content or to review forthcoming changes in community standards or internal speech policies. Others, like the Facebook (now Meta) Oversight Board, was more public-facing and included both a private and platforms process to request review and input on speech decisions and policies. The work demanded by these councils was often both reactive (providing input on particular pieces of content that had already been flagged as problematic which needed expertise in making a decision) and proactive (providing input on proposals to change the speech rules of the platforms in the future).

Individual Trusted Flagger and Partners Programs

Trusted flagger or trusted partner programs were also developed at YouTube and Facebook to do the same type of work as Group Councils but not in a committee format. Rather than group collaboration, such programs were lists of individuals who could be queried – or proactively inform the platforms – in a one-on-one capacity of issues arising on the site. Identification of such trusted flaggers and partners was both meritorious and legacy driven. Trusted flaggers ranged from

- Users who had established themselves as reliable signallers of problematic content or content erroneously flagged for removal by platforms over time;
- Former trust and safety employees of the platform;
- Government officials;
- Expert individuals from academia and civil society whose groups were flagged as “trusted” or who had existing relationships with platforms in other capacities;
- Members of the formal Group Councils in their individual capacity.

There are benefits and drawbacks to these more informal individual relationships. On the one hand, they are easy, cheap, fast, and reliable signal mechanisms for platforms. But such informal networks lack transparency, deliberation, and are prone to cronyism and nepotism.

Though both group councils and trusted partner programs still exist, the growing public attention and awareness around content moderation have led to both popular and government pressure for more transparency on such programs and increasing reliance on more formal councils.

There has also been more active engagement and development of civil society efforts to engage platforms through both these formal and informal mechanisms. This was due in part to the media attention and journalism around content moderation, but also a number of high-profile academic works that highlighted the realities of these pluralistic systems at work, making them more of a target for impact and intervention.

Current Web 2.0 Multistakeholder Intervention

A turning point in both journalism and academic work in this field occurred around 2016 when a number of events happened simultaneously to raise awareness:

- Publication of a number of foundational academic research texts describing the field of content moderation and the pluralistic forces that impacted it;
- High profile examples of both “erroneous” removal of content by platforms and platforms failure to remove other kinds of problematic content;
- The rise in this public, government, academic, and media awareness during a key U.S. presidential race, which had a controversial outcome that many blamed on social media.

The sudden scrutiny of content moderation policies and practices at platforms threatened to hurt the brand reputations of companies and accordingly, their stock valuations. It also threatened to disrupt the user engagement model that funded the advertising revenue on which such platforms were reliant. And of course, content moderation controversies created threats of litigation (however unsuccessful they might be due to Section 230 immunity) and more worryingly for platforms: regulations.

In order to stem some of the reputational damage and user dissatisfaction – as well as pre-empt potential regulation – many social media platforms decided to invest in more self-regulation and to do so more publicly. That took the form of more transparency around speech rules, increased cooperation with outside researchers, academics, and journalists, revamping of content policies and process – and most relevantly for this paper, increased investment in governance structures and institutions, in particular formal social media councils. Facebook’s creation and investment in the Oversight Board to review appeals from users on speech removal and non-removal, is perhaps the most high profile example.

Current Moment and the Lessons from the Past

Despite the call for and prevalence of formal social media councils in the last seven years, it is unclear what the long term future holds. A number of factors have changed the environment which – at least temporarily – made platforms voluntarily engage in social media councils. They include:

- A decline in market valuation of tech companies, resulting in widespread layoffs, hiring freezes and other cost-cutting measures, which include the dissolution or elimination of many social planner governance and research cooperation programs;
- The sense of relative lack of impact that platforms governance investments – or “governance washing” – did to repair harm to brand from content moderation scandals and election controversy;
- The failure of self-regulatory efforts like social media councils to stave off regulation like the European Union’s Digital Services Act and Digital Markets Act.

For those optimistic about the potential for new public-private institutional development like social media councils to protect users/citizens rights and represent their interests, these developments seem particularly disheartening. But the history of how they developed to this point and the environmental changes that led to their relative rise and fall can illuminate some valuable new tools and lessons to either preserve existing councils, encourage wider adoption, or empower similar pluralistic mechanisms on behalf of users/citizens. These include:

- State regulation of social media and speech platforms is particularly fraught where such platforms enable users/citizens rights around freedom of expression, but also perpetuate certain harms. Empowering the state to regulate platforms must be careful to preserve and protect the rights of users/citizens, not consolidate or strengthen, the power of government and the platforms against users/citizens.
- To this end, the best mechanisms for state power on behalf of users/citizens rights are those that promote process, users/citizens participation in platform governance, and transparency – rather than any substantive regulations around speech.
- Multi-stakeholder input to platforms on their speech policies and decisions developed organically because the relationship was mutually beneficial to both the interests of civil society groups – particularly in the United States and Global North – and the interests of platforms.
- Despite this mutual benefit, platforms have very little economic incentive to formalize or publicize such relationships with multi-stakeholders in any capacity, but this is particularly true for social media councils.
- If investment in social planner initiatives like social media councils has little to no economic benefit for the company and are only cost centers, they will quickly eliminate “self-regulatory” programs in economic downturns.
- Given these realities, mandated creation of social media councils for platforms is likely the only and best route to their continued existence and the best hope for continued development of new public-private institution building that empowers users/citizens.

Technical Difficulties: Incorporating independent technical expertise into platform council decision-making

Emma Llansó

CENTRE FOR DEMOCRACY AND TECHNOLOGY, WASHINGTON D.C., USA

Why was a cartoon critiquing police violence in Colombia suddenly removed from Facebook 16 months after it was first posted? What sorts of safeguards can Meta put in place to ensure that clearly labeled breast-cancer awareness images are not removed as violations of its anti-nudity policy? How should a service use automated filtering tools to swiftly enforce rules in a crisis situation, and what kinds of after-the-fact remedies should users have access to? These are just a few of the questions raised by actual cases that have been taken up by the Meta Oversight Board, established in 2020. They help to illustrate the significant technical component of many of the questions that the Board, or any variation on a platform council, will address about how an online service provider's systems are affecting and reflecting public values.

Reaching useful answers to these questions—investigating the provider's systems, evaluating their impacts on human rights, developing effective and implementable recommendations—will require the evaluators on such a council to have access to a wide range of expertise, including everything from interpretations of international human rights law to the lived experience of the people most directly affected by the online service provider's decision making.

For the purposes of this essay, I focus on the challenge of incorporating technical expertise in the design and operation of a service into the deliberations of a platform council. I identify a tension between two key features of this technical expertise: that it should be pertinent or specific to a given online service, and that it should be provided independent of that service. I also discuss the challenges of confidentiality when working with technical information pertaining to a specific service, and the role of transparency in ensuring both the accuracy and the legitimacy of council decisions. I then explore several options for making technical expertise available to the members of a platform council.

Finding Technical Expertise

Running an online service that provides a platform for user-generated content involves multiple technical systems: systems for carrying out the basic functions of the service, including hosting content, authenticating logins, and enabling accounts to interact with one another; systems for sorting, organizing, finding, and recommending content; and systems for detecting, evaluating, and enforcing moderation decisions against content that violates the service's policies, to name a few. Each of these systems, operating individually and in conjunction with one another, has the potential to affect the human rights of the service's users and may come under the scrutiny of a platform council.

Thus, a platform council will need access to many kinds of technical expertise. It requires different technical training to adequately evaluate, for example, whether a service's approach to moderating hate speech disproportionately affects users from a certain background, or whether the service's content recommendation algorithm routinely suppresses certain topics or perspectives. But technical expertise in the operation of online services is not seen as a primary, or even necessary, qualification for a member of a platform council. Among the Meta Oversight Board members, for example, the most common areas

of expertise include human rights, law, freedom of expression, and policy; Board members generally do not have a background in computer science or in working in trust & safety at a social media service.

Expertise in human rights law is certainly vital to evaluating online services' impacts on users' rights, but it illuminates only part of the picture. Moreover, while it can be helpful to have a general knowledge of how technical features such as machine learning classifiers or recommender systems work, most of the questions that will go before a platform council will be specific to the details of how a particular service operates. There is no single off-the-shelf technical approach to content moderation or recommendation that is employed across platforms. While some of the tools and techniques may be used in common across various services, each service modifies and implements them in unique ways, typically by incorporating them into a bespoke system. The questions of how Meta's systems affect its users' rights are specific to Meta's systems, just as they will be for any individual platform.

So not only do platform councils need access to subject-matter specific technical expertise, but it likely needs to be platform-specific as well. Much of the most relevant expertise will lie within the online services themselves, among the staff who have worked to build and run the very systems being evaluated. The field of online trust & safety is relatively young, and unlike in other technical industries, there do not yet exist massive consultancies of experts with decades of experience in the field who are available to weigh in as independent advisors. (There are several recent initiatives that seek to bring trust and safety professionals together to foster the development of the field (e.g. the Trust & Safety Professional Association) and to contribute their expertise to policymakers and others (e.g. the Integrity Institute). Given the pace at which a service's systems are updated and modified, the longer an expert has been away from working within a given platform, the more out-of-date their deep knowledge of its systems may be. And given the commercial sensitivity of the information at issue, both current and former staff are likely to be limited by non-disclosure agreements in what they can discuss about the technical inner workings of an online service.

Tensions and Tradeoffs

This, then, highlights a key tension in the quest to ensure that the requisite technical expertise is available to platform councils: independence versus pertinence. Platform councils need to trust that the expert technical advice they receive is accurate and not unduly influenced by financial, reputational, regulatory, or other considerations of the platform. However, the individuals with the most specific, detailed, and current understanding of how the platform's systems operate are highly likely to be people currently employed by the platform. And technical evaluations do not necessarily separate cleanly from other operational considerations: the feasibility of a particular recommended change or intervention in a system is likely to depend on considerations beyond just whether the change is technically possible.

A second tension is familiar from other discussions of platform accountability: transparency versus confidentiality. The work of a platform council will derive its legitimacy in part from the transparency of its operation and decisions. Transparent decision-making enables the affected parties and the outside world to understand the facts and rationale the council relied on to reach its decision. It also fosters additional learning and norm-development within the broader field of platform accountability. This includes transparency in the technical details of a case and information about how the underlying systems work.

Public details about the technical systems that underlie major online services remain relatively scarce and providing more information to the public is one of the many goals of the platform-council concept. It has already proven a useful output of the Meta Oversight Board, whose decisions have brought to

light new information about Facebook’s “cross-check” program, “media matching bank”, and machine learning classifiers for detecting nudity. Transparency in technical details also enables independent consideration of the tradeoffs involved in a given question, which can be useful for other online services grappling with similar issues as well as for researchers, regulators, and others trying to understand the online information environment as a whole.

In tension with this transparency, however, are the pressures towards keeping core technical information about an online service confidential. These pressures are many, including the legitimate need to maintain trade secret protection over certain intellectual property, the risk of manipulation of this information by bad actors seeking to abuse the service, and financial and regulatory concerns about the disclosure of derogatory information. As noted above, online services routinely require non-disclosure agreements from current and departing employees, which can significantly limit the ability of these individuals to share their specific knowledge of the operation of these systems. Platform council members may also have certain duties of confidentiality towards the services they evaluate (as well as the people whose claims they assess) that can further limit their ability to disclose information in published opinions.

The next section lays out a variety of options for bringing technical expertise into the deliberations of platform councils and discusses the tradeoffs between these various tensions.

Models for Incorporating Technical Expertise

There are multiple options for incorporating technical expertise into the deliberations of a platform council. A council will need technical advice and input at a variety of stages of its work:

- In reviewing and selecting the cases it hears, to understand the kinds of questions that are being put to it and to identify trends that may point to broader, systemic issues within the online service;
- In reviewing and evaluating the evidence and information provided by parties, especially that provided by the online service;
- In formulating questions and requests for further information from the online service; and
- In making a final decision and developing recommendations for how the online service should change its policies, practices, and design and use of different technologies.

In each of these phases (and likely others), council members will need to understand the technologies at issue, be able to ask thoughtful questions that identify how the operation of that technology may impinge on human rights, understand the information being provided by the online service and evaluate its legitimacy, and work towards technically feasible resolutions.

Appointing council members with relevant technical expertise

One way to ensure that relevant expertise is on-hand during the deliberations of a platform council is to appoint council members with a background in the kinds of technologies and processes the council will review. Depending on the size, composition, and remit of the council, this could include selecting council members with particular technical sub-specialties (much as one might appoint council members with expertise in the human rights law of different regions) or selecting members who have a general technical background and who can generally assess the information provided and ask relevant questions.

Positive aspects of this approach include ensuring that someone with technical expertise is present at every step of the process, and giving that person similar standing in the proceeding to other members of the council. Potential drawbacks include the risk of overreliance on one technical expert’s perspective or understanding, particularly among non-experts. There is also a need to ensure a balance of expertise

across members of a council, to avoid either over-burdening a handful of technical experts or crowding out other types of expertise.

The pertinence versus independence trade-off may be particularly salient for the technical expert-as-council member; an individual with direct experience working on the systems of a given service will have the most pertinent expertise but may also face the strongest questions about the independence of their perspective from their former employer. These questions could, in turn, shape public perception of the independence and legitimacy of the entire council. Recent former employees may also face the most stringent confidentiality obligations, which could limit their ability to apply their particular expertise. It may be most useful to focus on appointing council members with more general expertise, e.g. from working on trust & safety issues across multiple online services, rather than designating specialists in each relevant service.

Receiving technical briefings from current platform employees

Current platform employees can provide council members with technical briefings on the general operation of the online service and on specific questions that arise in a given place. This approach is used by the Facebook Oversight Board, which describes a process in its charter for Board members to receive information from and ask questions of the staff of Facebook and Instagram about the cases on its docket.

These briefings can help with the council members' overall familiarity with the service's systems and processes and can provide opportunities for a more efficient exchange of questions and answers. They can also be helpful in building a relationship between the service provider and the council and fostering an environment of good-faith interaction. The downside is that such presentations will necessarily come from the point of view of the online service, which may not be self-critical or wholly forthcoming. Council members may not know the most incisive questions to ask, or be equipped to catch inconsistencies or over-generalizations in the explanations they receive from the online service.

In other words, this approach maximizes pertinence of the expertise at the expense of independence; the information will likely be exactly on point to the question under consideration, but it will lack independent verification. Online services may find it easier to reach confidentiality agreements to allow current employees to conduct these briefings, given their existing employment relationship with the staff members. But, especially given the lack of independence of this information, it would be essential for the council to be as publicly transparent as possible about the technical information that forms the basis of their decisions, in order to enable independent evaluation and critique.

Seek technical input from outside parties

Platform councils could embrace a role that exists in many court systems around the world: that of an *amicus curia*, or "friend of the court". Also called "intervenors" in many legal traditions, these amici are independent of the parties involved in a case and serve to highlight interests, interpretations of law, and factual information that may be useful to the court in reaching its decision. The Facebook Oversight Board employs a version of this function in the open call for public comment that accompanies each case announcement.

A council's openness to input from third parties can help socialize its work to a broader audience, potentially reinforcing the norm-development work of the council, and can yield unexpected information and perspectives to inform the council's work. It may be difficult to attract high-quality comments from third parties, however; in the case of the Facebook Oversight Board, a high-profile case may draw hundreds or thousands of comments (the case of Donald Trump's suspension received 9,666 comments)

while many cases see only a handful of contributions. There is also no guarantee that commenters will answer the questions the council members pose.

Third-party commenters may typically fall on the other end of the spectrum between independence and pertinence: they can be wholly independent of the online service being evaluated, but the information they have to provide may be general, out-dated, or off-topic for what the council really seeks to explore. One way of improving the utility of third-party commenter technical information is for the council to be as clear and specific in its call for comment about what it knows and understands already about the systems involved in the case, and to identify precise questions that amici may ask. Some third-party commenters may be limited in their comments by pre-existing confidentiality agreements (e.g. if they previously worked at an online service), but they will have considerable leeway in deciding how to share what information they can. Comments from third parties should generally be made publicly available.

Appoint technical amici to work with council members

Independent technical experts could also serve as special advisors to council members, as technical amici who engage in a sustained way in the deliberations around a case. Technical amici with appropriate skills and expertise could be identified at the beginning of a case, or even beforehand if the council intends to explore a particular set of questions about a service's systems. These experts could be on-hand to answer questions and provide general explanations about their take on the technical factors in a case and could be present at any briefings provided by the online service. The technical amici could also provide perspectives on the proposed solutions and recommendations of the council.

The availability of a technical amicus would allow for immediate clarification and discussion of technical information provided to the council throughout the deliberation and could enable deeper inquiry from the council and a swifter resolution of technical questions. It would ensure that multiple perspectives, not only the online service's, inform the assessment of the provider's systems. This structure could be vulnerable to over-reliance on a single expert's perspective, and the online service may not be as forthcoming with an independent third party involved in the discussions.

Appointing case-specific technical amici could provide councils with the flexibility to ensure that the individual's particular expertise is pertinent to each case. (As with every option for bringing technical expertise into council discussions, councils may find the pool of technical experts with service-specific insights limited by pre-existing non-disclosure agreements.) Concerns about the independence of a technical amicus with recent ties to the online service (whose advice would be highly pertinent) could be balanced by employing multiple technical amici, or by using an additional method for sourcing technical expertise. Technical amici would likely need to agree to some level of non-disclosure agreement to participate in case deliberations, in order to ensure robust disclosure from the online service provider. Once again, this necessary confidentiality can be balanced by transparency in the ultimate decision, with the council including information about the advice received from the technical amicus as well as the information provided by the online service.

Conclusion

Each of these models for incorporating technical expertise into platform council decision making involves tradeoffs between getting the most pertinent information, presumably from someone with ties to the online service being evaluated, and ensuring the independence of the perspective received from the technical expert. The need for access to specific information about the operation of a given platform's systems also raises issues of confidentiality, as online services may be reluctant to share certain information about their systems outside of the confines of non-disclosure agreements.

Ultimately, platform councils will likely need to use multiple sources of technical expertise, to account for the benefits and drawbacks of each approach. Technical briefings from the online service itself can be complemented by input from independent sources, including through public comments or an independent technical amici role, and at least some council members could be selected for their specific technical expertise and training. Whatever the source of the technical input, it will be vital for councils to explain their understanding of the systems and processes at issue, and how technical considerations inform their decisions and recommendations, as part of their published opinions. This will enable independent evaluation of the role of technical systems in various platform accountability questions and will help to increase the overall transparency of our information environment.

Further investigation into this issue should involve a deeper exploration of other oversight and decision making processes and their approaches to involving different types of expertise in their work. Media councils for print and broadcast journalism, for example, have been in operation in different parts of the world for many decades, and have served as the inspiration for today's debates about social media or platform councils. Deliberative democratic and collective dialogue processes seek to bring the people directly affected by governance decisions into the process of making those decisions; such processes must also grapple with the varying kinds of expertise that participants bring to deliberations, especially when considering technical questions. And regulators in many other industries, from aviation to pharmacology, routinely deal with highly technical questions.

Finally, it should be noted that platform councils are far from the only institutions who will be grappling with this need for enhanced technical understanding of the systems that make up online platforms. As legislators worldwide craft new intermediary liability laws and platform regulatory frameworks, judges and regulators will need to interpret and apply them, and will encounter many of these same questions. Platform councils, then, represent a useful opportunity to work through various options for incorporating technical expertise into third-party oversight of online services, and to balance these competing tensions of pertinence, independence, confidentiality, and transparency.

Enforcement as a necessity in platform councils

David Morar

OPEN TECHNOLOGY INSTITUTE, WASHINGTON D.C., USA

It is a worthy goal to democratize online communication spaces, but it will not be easy to achieve. By definition, any external institution built to infuse public values into the functioning of online platforms threatens the current make-up of the companies running those platforms, as they would have to cede the power they currently hold unilaterally. If platform councils are to perform this crucial task, one way or another, they would require enforcement capacity, or rather a way to ensure accountability.

Public values, rooted deeply in democracy, derive a significant amount of their legitimacy or worth from, among others, functions of accountability, as well as a right to rule, via consent and acceptance. For the purposes of this paper, these functions and rights are bundled together under the umbrella of enforcement.

If an institution inhabits a role of “governing” or “regulating” another or a specific group (or industry), that implies the existence of accountability, or enforcement, at a minimum through reporting, auditing, monitoring, and verification procedures. In the space of private governance, accountability measures range from being baked into the purpose of the institution, such as public assessments of compliance with a certain set of best practices, or certification schemes, all the way to exclusion from the institution. While necessary, these mechanisms are not sufficient to legitimize any institution, and even with stellar enforcement private governance structures may end up being ineffective due to a whole slew of other concerns.

Against this backdrop, this paper will first surface examples of critiques of similar institutions from other industries and their enforcement structures. It then looks at the most well-known example of a platform council, the Facebook Oversight Board, and explores its enforcement mechanisms. Finally, it problematizes the idea of the need for enforcement, by assessing industry perspectives on relinquishing its power and ways to alleviate concerns.

Critiques of similar institutions without strong enforcement

Multi-stakeholder governance institutions are institutions where power is distributed among actors from different stakeholder groups. These institutions have a very wide range of structures and varying degrees of enforcement/accountability mechanisms. Historically, those without strong, effective enforcement have been decried as marginally better if not similar to self-governance schemes, which is to say ineffective at allowing other stakeholder groups access to influence the actions of industry (whether individually or at industry level). Three critiques in particular can be very useful for envisioning enforcement within the platform council perspectives.

Blue/Greenwashing

The most damning response to multi-stakeholder initiatives is the accusation of greenwashing: to deflect blame for a misstep, a company creates or adopts a broad, substance-less set of principles which have no actual effect on the original issue. A similar version is that of bluewashing, where corporations join institutions or structures in functionally toothless labeling schemes where they benefit from participating alongside the UN, without making any meaningful changes to their social or environmental performance.

In cases when industry is not in any way beholden to making significant changes, a direct result of lack of enforcement capacity, multi-stakeholder initiatives are usually described by critics as a form of greenwashing. Businesses consent and accept only inasmuch as they want to be seen as doing so externally, but without any accountability, the institutions are considered toothless. The most famous example of bluewashing is the [Global Compact](#), which is predicated on companies adhering to nine principles related to environmental protection, human rights and labor standards. However, it has very few ways of actually enforcing this adherence.

Designed/heavily influenced by businesses

Another criticism of multi-stakeholder institutions is that industry's role in creating, implementing, or running the institution overpowers any kind of significant input by others, usually civil society. Beyond having the power to actually influence or make the rules, industry would also have the power to dilute the accountability aspect of the institution. While providing its consent and acceptance, the company stakeholder group becomes the one in charge of its own accountability, either directly by wielding power within the institution, or indirectly, by building the structure themselves.

For instance, industry-led institutions that claim to be multi-stakeholder and multi-stakeholder institutions with a strong if not overpowering industry role are truly a difference without a distinction. This critique is usually leveraged against multi-stakeholder institutions regardless of their actual make-up, so identifying real examples is crucial. [The Sustainable Forestry Initiative](#) is an industry-led standards body that claims to have equal representation of stakeholders, but its output is far less stringent than the [also flawed](#) but multi-stakeholder [Forestry Stewardship Council](#), which makes its true nature apparent.

Lack of potential to grow stronger

A more nuanced critique of multi-stakeholderism is that its powers rarely expand beyond its original scope. It's important to remember that multi-stakeholder institutions usually exist because companies have ceded some of their own power. As previous critiques argue, what is presented as a legitimate power of the institution is, in fact, not real, or hobbled by industry (alone or with others), so the hope would be that with time the institution would amass real influence. However, an institution built with limited enforcement can't, by itself, decide to increase its own enforcement capability unless the entity being governed further agrees to it. What potentially can happen is the decrease of the enforcement capacity and thus authority of the institution.

Internet governance has famously been experimenting with multistakeholderism. One particular institution of the internet governance ecosystem, the multistakeholder Internet Governance Forum (IGF) was a peculiar outgrowth of power struggles between the previous multilateral UN-based regime and the upstart multistakeholder community taking part in the ICANN (Internet Corporation for Assigned Names and Numbers) which de facto governed the infrastructure layer of the internet. The outcome was bizarre: after a failed attempt at wresting power from ICANN, in 2005 the UN ended up with a compromise that its newly created Internet Governance Forum would not actually have any outputs, thus effectively neutering the organization, and subsequently removing any kind of enforcement. The IGF remains an institution whose main goal is to bring stakeholders together, without any kind of policy output.

While such general critiques are important in the abstract, far more useful is to understand how one of the few platform councils in existence fares on these issues.

Meta Oversight Board

The Meta Oversight Board's structure shows that it was built with a focus on deliberation, rather than governance. The enforcement mechanism and thus the legitimacy of the Board is limited in both scope and dimensions. Designed by and for a corporate entity, with the ostensible and amorphous goal of oversight, the Board has become a beacon of primarily externalizing responsibility for difficult disparate choices instead of a democratizing check on industry power.

Providing its deliberative action with a contractual mandatory accountability role signifies that the company has to implement, within the boundaries of predetermined rules, a decision by the Board's panel. While no penalties exist for non-compliance, it would certainly lead the entire edifice to crumble. The legitimacy of the Board has as one of its many necessary but not sufficient attributes the concept of the enforceability of its judicial decisions; Facebook consents and accepts these decisions *a priori*.

Its role was always first and foremost to provide a limited and pointed check on specific thorny cases of content issues. Thus, it was invested with contractual power to provide enforceable decisions in those cases. Given a perfunctory and wholly performative power to comment on matters of policy, the Board fails to inhabit a role of meaningfully inserting public values in its relationship with Facebook. Facebook does not automatically consent or accept the comments from the Board that are outside of the judicial cases. Even more, there are no explicit or implicit accountability mechanisms, as evidenced in the 2021 X-Check [case](#), where the Board learned from media reports that Facebook lied about one of its programs, and subsequently issued an [advisory opinion](#) in December 2022 with no way of reprimanding the company.

Confined to its limited role, FBOB was properly designed as an external adjudication body. Extrapolated to the discourse surrounding it as a governance or rather "oversight" structure, the FBOB was faulty, to say the least. It can be argued that a platform council with a limited remit and with even more limited mechanisms for enforcement, starting with a right to rule and ending with accountability, would have a hard time actually infusing public values into the corporate space. Certainly its existence by itself can denote a shift towards democratic values, as a small portion of the company's power is diminished and re-routed this way. However, the important decisions related to the functioning of the platform are still taken by the business. The FBOB's founding documents, built by the company, which was relinquishing power, did not provide for a way to actively embody and wield legitimate power over the overall functioning of the company.

However, adding such mechanisms to a platform council can potentially lead to the exact opposite scenario, where companies feel they do not have a way to respond.

Enforcement critique

A crucial aspect of decision making at any level is the enforceability of its outputs. A decision that can be ignored without consequence undermines the structure. At the level of the state, Max Weber famously posited that a monopoly of legitimate use of force would be necessary to enforce order. However, the state, and especially a democratic state, has throughout its mechanisms important checks and balances. Be it, like in the US government, where the three branches serve as each other's enforcement mechanisms, or between citizens and the government, where those who have consented and accepted to be governed have the opportunity to use their own power to hold those in power accountable.

Whereas a democratic nation-state's duty is to secure basic rights for its citizens, a corporation's incentive is to uphold its duty to its shareholders. Thus, within democratic nation-states the enforcement

mechanism is, at least ostensibly, wielded by entities with similar incentives to those being overseen, ultimately securing said basic rights. Not only are they not similar, but platform councils would be the product of the company relinquishing its power to a group of people who are not aligned in their incentives. Even more, why would a company give up their power to an institution which it has no way of reprimanding, or at least calling out potential concerns? Arguing for a strong enforcement mechanism would make it even less palatable to corporations to relinquish their influence over their own products and services.

The counter to such an argument should be in the details of the organization of the councils. Building effective enforcement would also require ways for those being governed to object, and do so in a way that allows for legitimate complaints, but does not burden or grind to a halt the structure. Including as equal input the desires of the companies, or allowing them to have a voice in the process would be a way to do so. For instance, the Internet Corporation for Assigned Names and Numbers, ICANN, has a robust commercial stakeholder group with a nominally equal voting bloc in matters of policy.

Conclusions

When attempting to understand an experiment like platform councils, it is useful to learn lessons not just from the limited examples in the field but also from similar cases from other industries.

This paper has three major takeaways. First, built-in mechanisms for consent, acceptance and accountability, from clear contractual terms (in the case of ICANN) all the way to dissociation (in cases like the FSC) are important and they can be one of the deciding factors in whether the infusion of public values is effective. Second, as certain actors (most likely industry) can take advantage of a weak structure or of their own power, institutions built on decoupling the powerful (in this case platforms) from their power require enforcement mechanisms to ensure actual governance or oversight and to avoid pitfalls of perceived or real legitimacy gaps. Third, platform councils that do not include in some legitimate way either the voice of the platform in its deliberations or mechanisms for complaints may find it difficult to implement enforcement mechanisms.

Interoperable Platform Democracy: How deliberative democratic processes commissioned by corporations can interact with nation-state, multilateral, and multistakeholder decision-making

Aviv Ovadya

BERKMAN KLEIN CENTER, CAMBRIDGE, USA

Is there a world where corporations not only run democratic processes for their decision-making—but where that process is actually a *good* thing? A world where important and controversial choices facing corporate platforms and AI organizations are decided not by leadership fiat but by a truly representative deliberation (largely outside of government)—and where this is not just ‘democracy washing’?

This piece explores what that world might look like, and how such democratic processes—potentially commissioned by corporations—might beneficially interoperate into our existing institutions of national, transnational, and global governance (hereafter referred to simply as institutions).

Such questions are particularly salient given both Meta’s concrete actions to use such processes (initially to develop greenfield policies in the Metaverse), and AI leaders’ exhortations to “align their interests to that of humanity”—where such processes might be particularly applicable.

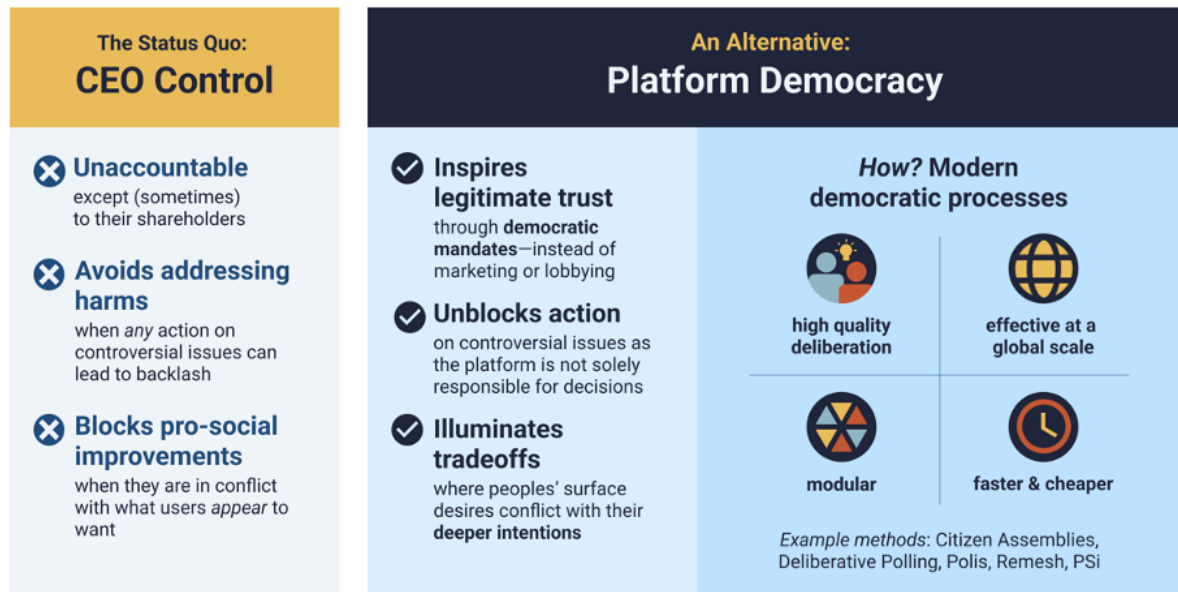
There were several key guiding questions which led to the approach outlined below:

- Who *should* actually be in charge?
- Is it possible to govern tech in a way that moves power to the people being impacted—and away from *both* corporate leadership and oppressive governments?
- What are **pragmatic** approaches we can try **today** to rapidly improve the governance of transnational technologies—in a world where such international coordination seems increasingly difficult?

What is platform democracy?

Platform democracy: Governance of the people, by the people, for the people—except within the context of an internet platform (e.g. Facebook, YouTube, or TikTok) instead of a physical nation.

More formally, platform democracy refers to the use of democratic processes to include the populations impacted by a platform, in the governance of that platform in a representative fashion.



In particular, two approaches to such platform democracy are considered here: intensive deliberative democratic platform assembly processes for complex decisions and lighter-weight collective dialogue processes for decisions that need less context. In both cases, an organization (such as a platform) needs to make a decision that would benefit from democratic legitimacy from the decision.

Such questions might include:

- What if anything should be done about content that is not strictly false, but which is meant to be misleading?
- Under what conditions, if any, should audio or video be recorded in online spaces in order to identify potential harassment, and if so, who should have access to such recordings?
- What kinds of content, if any, should not be shown as ‘trending’?
- What kinds of outputs are acceptable from generative AI systems?

None of these are theoretical. Meta has already directly explored a version of the first two of these questions through such processes; Twitter would likely have asked the third question had there not been an acquisition, and OpenAI’s CEO has described the fourth as a question for which he would like global democratic input.

How are these questions then answered such that the processes are “democratic”?

A microcosm of the impacted population is convened and facilitated by a neutral 3rd party, such that everyone being impacted by the decision (might be e.g. the user base, or the countries the organization operates in) has roughly the same opportunity to be selected (through sortition: stratified random sampling). The selected people make the ultimate recommendation to the decision-maker—and unlike a poll, they are given the opportunity to learn from each other's perspectives (and for decisions that involve significant tradeoffs or context, they also learn from stakeholders and experts). These ‘deliberators’ are paid for their time, and ideally child care, elder care, travel, etc., to reduce the self-selection bias.

Platform Democracy via Platform Assemblies¹

Platforms (like Facebook and YouTube) can use the same “citizen assembly” processes validated by countries around the world to **incorporate democracy into decision-making**. A kind of on-demand platform **legislature**.

platformdemocracy.com
Aviv Ovadya | aviv.me | aviv@aviv.me | [@metaviv](https://twitter.com/metaviv)

How does a ‘Platform Assembly’ work to address a controversial issue?

Example issue: *What should the platform’s rules be for political ads?*



PHASE 1: Representation

A democratic lottery is used to select members of a platform assembly such that they match the makeup of those impacted by the platform.²

\$ Paid 30–800 People



PHASE 2: Deliberation

The assembly learns from experts, stakeholders, and each others’ perspectives. They produce a set of recommendations for the platform with rough consensus.

4+ days Neutral Facilitation Expert testimony

The platform implements the recommendations or provides specific reasons they could not.³

¹ There are other potential approaches to platform democracy, e.g. deliberative polls and digital governance tools, but they involve less agency and context.
² Democratic lotteries use sortition—stratified sampling of a population—to create an assembly which matches the population being governed along designated criteria.
³ One ideal involves the platforms binding themselves to the recommendations and even having assemblies provide oversight of implementation.

Why are these processes legitimate?

The potential democratic legitimacy of such processes comes from their representative nature—instead of every single person having an opportunity to vote, but spending fairly small amounts of time per person, a much smaller number of people vote, but they each are supported with the time and resources to make the best possible decision (without the often perverse incentives of electoral politics or corporate profit). Such processes are also not just some techno-optimistic idealistic dream but are being used by existing governments around the world. Moreover, as any individual has only a small chance of being selected, it is far more feasible to imagine such processes working across many platforms, even globally, than an electoral representation system.

Can deliberative processes work across many languages and cultures?

Such deliberative democratic processes have now been run with many languages a number of times, including several across the EU. Admittedly, interlingual and intercultural deliberation is still imperfect, but there are both process approaches and tools that can help mitigate the risks, and ongoing experimentation to develop best practices around the most challenging aspects (e.g., subtle differences in word connotations across languages).

What happens after the process is complete?

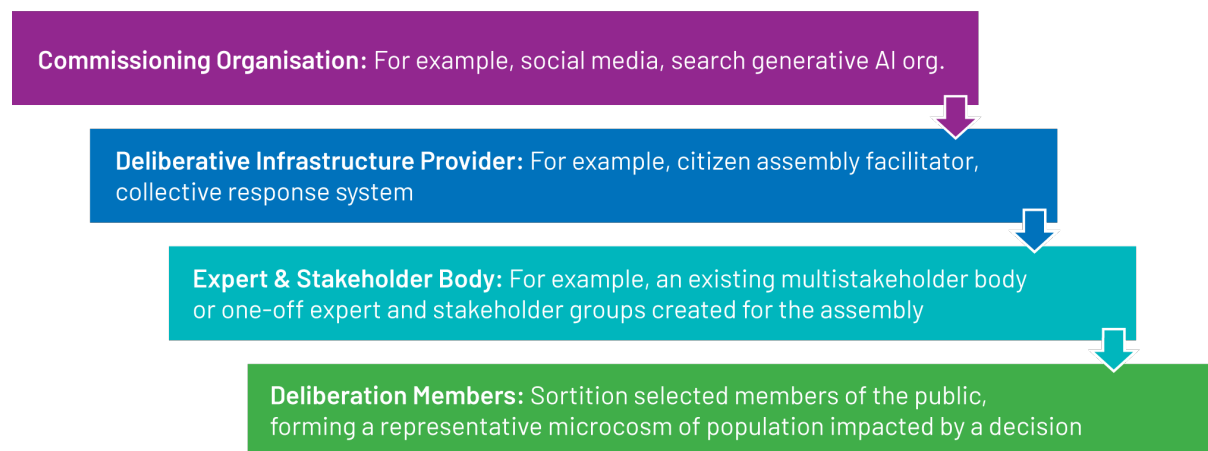
As a slight generalization, when governments run such deliberative processes, they usually serve as recommendations, which must be either implemented, or receive a response from the government about why the recommendation is not being followed. The same can apply when the commissioning organizations are companies like Meta or OpenAI; though it is also likely possible to make the results binding.

Interoperability with existing institutions

Platform democracy does not exist in isolation—it should be structured to support existing institutions instead of fighting them. There are several places where this can happen.

To contextualize the options for interoperability with existing institutions, it's useful to understand the different organizations potentially involved in such a process.

- First, there must be a **commissioning organization**, e.g. Meta, Twitter, Google, or OpenAI. This could also be a combination of organizations, or even organizations and governments together.
- They commission the deliberation with “**deliberative infrastructure providers**”—organizations that run these sorts of processes as neutral third parties for governments (and now companies) around the world. The deliberative infrastructure providers also select the members of the deliberation body using sortition and facilitate the assembly itself.
- These deliberation infrastructure providers may work with **existing expert and stakeholder bodies** to provide context for the deliberators or create a **new temporary advisory or stakeholder body** to help support the deliberation.
- Finally, the **deliberation members** learn from those stakeholders, experts, and each other in order to make the final recommendation.



Impacts of platform democracy outputs

The most obvious touchpoint where such a process interacts with the broader world beyond the commissioning organization relates to the impacts of the process *outputs*.

Media pressure

Let's not beat around the bush here—platform democracy, in its most limited form, can be considered a form of self-regulation. However, it is different from most forms of self-regulation in that the power of creating the mandate is not directly in the hands of the platform. It is instead put to people chosen at random, without any incentive to accede to the platform's wishes, and facilitated by a 3rd party deliberation organization. (A rigorous process aiming for strong legitimacy would also use as impartial a method as possible of choosing experts and stakeholders.)

Moreover, even if the recommendation is not binding, the legitimacy of the mandate created by such a representative democratic process makes this kind of self-regulation rather awkward for a company to ignore. In plainer words, it looks very bad to the media (and the governments that follow it) if an organization convened a process to have the people tell it what they want in a democratic fashion—and the organization ignored the outputs.

Perhaps even more exciting, a sufficiently transparent and high-profile process not only educates the members of the deliberation themselves, but allows the broader public, media, and even regulators to follow along through broadcast and social media, enabling learning from experts and stakeholders alongside the members. This can potentially help elevate the overall level of conversation on issues with complex trade-offs more effectively than hearings used to score political points, and also help the public see itself through its reflection in the deliberative microcosm.

Raising the responsibility baseline

In fact, recommendations that come out of a process that is seen as broadly legitimate are likely to not only affect the organization that is convening it, but also any other organization facing similar questions and advocacy and interest groups that relate to the question (assuming it is not too specific to the convening organization). If the question is for example around potential responsibility actions, this can help create a corresponding *responsibility baseline*—a minimal level of action that is seen to be broadly acceptable, which may be higher than the current industry default, raising pressure to implement responsibility practices across the board.

Even if the responsibility baseline is lowered, that is potentially indicative that the impacted population does not actually believe that that level of ‘responsibility’ is warranted (for example, you can imagine a deliberative process that determines that there should actually be less content moderation around a particular issue—which would be a good thing to know).

Creating a responsibility ‘north star’

Some processes may not change the baseline, but may instead create a north star— responsibility practices that might be too difficult to fully execute on, but can be aspired to and approximated. Such north stars may also exert pressure on the entire industry of the commissioning organization.

Identifying global ‘moral high ground’

For some issues, the challenge is not around the ideal north star, or the minimal baseline for responsibility. Instead, there might be deeply competing notions of what responsibility even is. For example, some organizations developing powerful AI systems say that the responsible thing to do is to share as much as possible—maximizing openness. Others are extremely cautious and barely release any information about their research. Both sides say that they are acting for the good of humanity—in other words that they have the ‘moral high ground’. Both argue their perspective to the public and regulators with the intent of shaping perception and law. Similar differences in approach occur in many domains, including in tradeoffs between privacy and security.

A rigorous global deliberative process can create something closer to an idealized public sphere to actually identify what ‘humanity’ believes is the moral high ground (that such companies should be aiming for). There are thus potentially strong incentives for organizations that believe that they are closer to the ‘true moral high ground of humanity’ to convene such processes, in order to have their approach validated (assuming that they are correct).

Regulatory and institutional suggestions

Such responsibility baselines, north stars, and moral high grounds may then directly impact the actions of legislators, regulators, standards bodies, multilateral bodies, multi-stakeholder bodies, trade associations, etc., in ways that may be binding. In other words, the commissioning organization is essentially fronting the cost for deep research and input gathering that can then directly feed into these existing processes, some of which may have more binding force. Concretely, this might look like, for example, the UK government, the EU, UNESCO, or the Partnership on AI developing recommendations (or, for governments, even laws) directly based on and referencing those deliberative outputs. This could be true even if the deliberative process that was originally convened by Meta or OpenAI—assuming that the process was seen as rigorously impartial and democratic.

Bindingness

There is the option that the convening organization can pre-commit to making an output binding (when otherwise legal), using the legal infrastructure of the jurisdiction(s) they are operating in. There are likely a number of legal instruments that can be used to do this depending on the relevant jurisdiction (e.g. [a golden share arrangement](#)).

Conflict with platform democracy outputs

There are some common questions about how this might play out in practice:

What happens when there is conflict with existing law or regulation?

In situations where there are conflicts between the outputs of deliberations and existing law or regulations, the situation is roughly analogous to when a company's strong ideological stance conflicts with that of a government. In some cases, this may be seen as good, e.g. when a company avoids sharing location information about democracy activists, thus violating the laws of an authoritarian country. In other cases this may be seen as problematic, e.g. when a ride sharing company ignores local safety regulations. Either way, if organizations do not follow the laws of the nations they are based in, they face the consequences. The main difference is that if the legitimacy of the process used to create the deliberate outputs is higher than that used by the government (for example in an authoritarian or extremely partisan context), then there may be significant pressure, both externally and internally pushing for the more democratic outcome.

Could the governments or regulators themselves actually be involved in the process?

Definitely, though of course this can become more challenging with more global processes (and thus more governments). It's also worth noting that one of the benefits of the platform itself running a process is that the process can be specific to features that only that platform has, and it may not be worth the time for government officials to be involved with every platform in such a manner. That said, especially for processes which involve multiple platforms or industry consortia, governments may want to act as co-convenors, and platforms may want that also in order to increase legitimacy of the outcomes.

Could there be permanent deliberative bodies?

There are many potential models beyond the simple temporary platform assembly or collective dialogue, including institutionalized permanent models built on approaches such as multibody sortition, the

Ostbelgien model, the Paris model, and which could directly interact with existing institutions in much more sophisticated ways. It feels somewhat presumptuous to explore this in the context of platforms and companies without more understanding and exploration with the basic model, but it is important to know that it may be possible to have key decisive power over an entire company through such processes, as they are refined and combined. One could even imagine augmenting or replacing a traditional corporate board structure with carefully designed deliberative bodies in order to truly enable democratic governance, with no higher executive or board level power (though feasibility might depend on the jurisdiction).

What happens if there are multiple representative deliberations with conflicting outcomes, perhaps even some run by the governments themselves?

There is no clear answer to this as this entire regime is too nascent. It is perhaps roughly analogous to having multiple treaties or non-binding agreements which are in conflict in a multilateral context. The ideal is likely that the process which is most rigorous and thus most legitimately democratic wins out—but there are many potential interpretations of rigorous, legitimate, and democratic, and no clear arbiter. This suggests that it is particularly important to create international standards for such processes in order to ensure consistent evaluation.

More generally, any time there are multiple competing decision-makers, potentially of varying quality, and no official hierarchy, there is bound to be tension, (ideally productive tension) and there is value in creating institutions to navigate those tensions.

Inputs to platform democracy

Beyond simply interoperating with other organizations through the outputs, democratic processes *inputs* can also interact with existing institutions and organizations at other stages of the process.

These include:

- The *commissioning organization* could actually be a joint body involving a partnership of a platform (or platforms) with a government or even governments, multilateral institutions etc. The commissioning organization could itself be an existing multi-constituency body such as the Digital Trust and Safety Partnership.
- The *expert and stakeholder body* could also be an existing multi-stakeholder body such as the Partnership on AI.
- Governments could help support the actual *process of sortition selection* if they already have ‘sortition infrastructure’ (as e.g. Mongolia has, as illustrated by its incredible turnout for their [deliberative democratic process on constitutional amendments](#)).

Why we might want platform democracy

I might prefer a world where purely public institutions fully govern our technological developments and have kept up with the rate of technological change—change that respects no borders. But we have not evolved our existing governance institutions to take on the challenge of legislating at the speed of technology, and that is unlikely to change very quickly.

The realist question we thus face is:

“How can we practically govern an onslaught of technological disruption—and what are the consequences if we fail to do so?”

This is not theoretical—platforms like Facebook, YouTube, and TikTok have shaped society through their policies, but even more impactfully, they have shaped the incentives of society through their ranking systems. These ranking systems determine what kinds of politicians, journalists, or entertainers succeed and shape the kinds of content they produce. Our existing governance institutions, over a decade after this became clear, have done very little to improve the impact of such systems on society outside of the narrow scopes of personalization and privacy.

We can and must do better, both to tackle belated issues and the emerging governance challenges around new technologies. This is especially salient for advances in AI, such as foundation models like GPT-4 and the products like ChatGPT built on top of them, which are likely to rapidly transform our lives. Perhaps deliberative democracy can help us find a way forward. Given a steady rhythm of convened processes, decisions might be made democratically, even at global scale, within months instead of years or decades.

Platform democracy alone cannot solve our problems, but it perhaps provides a useful new governance option between our status quo of platform autocracy and platform chaos.

Public Values and the Private Internet

Peter Routhier

INTERNET ARCHIVE, SAN FRANCISCO, USA

Introduction

The [Platform://Democracy](#) project asks how we might “ensure public values in private and hybrid communication orders,” and what role social media councils could play in achieving this result. This paper takes as its starting point that the relevant “communication order” is the internet itself—not just the “walled gardens” of today’s social media companies. From this standpoint, we can see a few ways social media councils could inadvertently hinder the development of public values in the broader online ecosystem. And we can suggest one way to address this: by enabling our existing public interest institutions, like libraries, to perform their traditional functions in the online environment. In this way, we can build public values into the broader internet.

Social Media Councils and the Broader Internet

For many years now, [Tim Berners-Lee](#) and other early internet pioneers have identified certain aspects of social media as presenting “threats” to the World Wide Web—and the internet beyond the Web itself. Others have been more blunt in their assessment: “The internet has a Facebook problem, but the internet is not Facebook.” While often critical, these kinds of voices also typically express optimism about our ability to make the Web whatever we want it to be: “We create the Web . . . We choose what properties we want it to have and not have.”

From this standpoint, we should consider what impact social media councils could have not only on today’s dominant social media platforms, but on the broader development of the internet and the World Wide Web. This paper suggests three possibilities: they could unduly legitimize the already-existing social media platforms; they could entrench these platforms into their dominant positions; and they could distract from the potential to build public interest values into a broader and more diverse online ecosystem.

First, social media councils might unduly legitimize today’s social media corporations. These companies are largely duty-bound to maximize their profits, and there is reason to doubt that they are willing or able to make this duty subservient to the advice of a council or otherwise. If one accepts that at least some of the problems with social media companies that councils are seeking to solve stem from the profit-maximizing behavior of those companies vis-à-vis the operation of their platforms, then—under Delaware law, at least—the influence of a social media council will be muted if it is not able to supersede or at least somehow coexist with the profit-seeking directives at the heart of their corporate governance. In these circumstances, social media councils might lend to platforms the imprimatur of their members and their organizers, all while being structurally incapable of imbuing them with their values.

Second, social media councils can work to entrench today’s platforms into their dominant positions. If these corporations are unwilling or unable to voluntarily alter their problematic behavior, as the first point holds, then the natural conclusion of the regulator is to make mandatory and superseding rules. But in a variety of ways—most notably, the cost of compliance—such rules can entrench these companies in a position of dominance online. That is, imposing new compliance obligations on platforms of all stripes can have counterproductive structural consequences. It is possible, for instance, that they would increase the cost of participation in the relevant online environment to such a degree

that only corporations with a certain level of financial resources can bear the costs. This could increase, rather than reduce, the profit-maximizing behaviors of the relevant firms which led to the initial problem to be solved. And it could remove the possibility that alternative organizational models could arise or persist in the new environment.

Finally, social media councils can distract from the potential for different kinds of reform. The best example here may be Facebook's creation of the limited liability corporation that it calls the Oversight Board. The process leading to the development of the Oversight Board corporation was no doubt fascinating and extraordinary: the company has spent hundreds of millions of dollars on its creation and development, capturing the attention and interest of a variety of actors from around the world. Meanwhile, the Oversight Board corporation itself has written only a handful of guidance documents with which it requests Facebook's voluntary compliance. Has this amounted to much more than a distraction? Was it intended to be anything but?

None of this is to say that social media councils cannot do real good—it would seem to be the least we could ask of some of today's dominant platforms—but rather to suggest that we shouldn't let them legitimize our existing platforms, entrench them, or distract us from other possibilities for imbuing public interest values into our online information ecosystem. One such possibility would be to empower existing public interest institutions, like libraries, to play a larger role—even simply their traditional roles—in the digital environment.

An Internet with Public Interest Values

When considering how we might imbue the broader internet with public interest values, a natural starting point would be our traditional public interest institutions, like libraries. To be sure, social media companies have displaced an older model of information production, publication, and curation that was populated by a variety of different actors and constructed in a fundamentally different way; we cannot “put the genie back in the bottle.” But it is worth considering whether and to what extent we could empower public institutions like libraries so that they can, at least, play their traditional role in the digital environment.

Libraries are relevant to the discussion for many reasons; they have long played “a fundamental role in society” as “gateways to knowledge and culture.” They have done so by, among other things, preserving and lending books for the public's benefit. Importantly, unlike social media companies and a variety of other actors to come before them, libraries often have a public character rather than a private, profit-maximizing interest.

As a result, it is perhaps no surprise that some have called for librarians to play a more active role in the online information ecosystem. But as with other experiments with placing traditional knowledge workers into social media environments, placing libraries and librarians directly into the social media framework may offer no salvation. Indeed, there is reason to believe it may not be desired by many librarians themselves. In 2022, Internet Archive and the Movement for a Better Internet convened a group of leading librarians and others at Georgetown Law Center for a workshop on digital library issues. As the resulting whitepaper reports, those present expressed little interest in working as fact checkers or content moderators for the Facebooks of the world. What they hoped for, instead, was to be empowered—or, at the very least, simply allowed—to fully execute their traditional public interest functions in the digital environment. Unfortunately, as outlined in the paper, a variety of economic and legal factors have so far worked against their doing so. But as Ethan Zuckerman has noted, we should not assume that the current “economic and legal system govern[ing] our online spaces” will persist in perpetuity; such thinking “obscures possible solutions to the challenges arising around the socially corrosive effects of new media technologies.”

For an example of how empowering libraries in the digital environment could improve our online ecosystem, consider Internet Archive's project to weave books into the web—and turn all references on Wikipedia blue. In short, this project has made hundreds of thousands of authoritative sources—books, webpages, and more—available to readers and contributors of Wikipedia. Vast quantities of Wikipedia citations are now cite-checkable, so that readers and contributors alike can check their sources and work within a more complete and accurate information environment.

Conclusion

While many have proposed structural reforms which would weaken the power of dominant social media companies or otherwise open them up to competition and reform, we should also consider empowering public-interest-minded actors so that they can play a larger role in the online information ecosystem. Social media councils may help bring public values into private companies, but we should not stop there. Empowering public interest organizations like libraries, so they can execute on their traditional functions in the online information environment and seek to innovate towards new ones, can help bring public values to the broader internet as well.

Soft Power and Platform Democracy: How social media councils could shape government and corporate strategies and preferences

Fabro Steibel

INSTITUTE FOR TECHNOLOGY AND SOCIETY, RIO DE JANEIRO, BRASIL

In previous [work](#), we have addressed the question of content moderation and accountability as well as how to use policy framing to understand how institutions are designed to solve policy challenges. The previous research showed how fact-checkers and judiciary institutions have framed the problem of negative advertising (the terminology used to refer to fake news back in 2011) in largely different ways. The relevance is related to accountability: if we set different policy questions to solve, we have different methods to evaluate what solution is ideal to cope with content moderation.

In this article we emphasize using policy framing to explore if the concept of platform democracy is going in a similar direction. Are social media councils being framed in largely different ways? And if not, how do the policy framings used to describe the values, policies, and institutions related to social media councils help us to understand where the content moderation agenda is going?

[Platform democracies](#) is a concept that explores the role of social media councils as external governance structures. The councils are more commonly initiated by companies themselves, as a self-regulation practice based on external oversight. One of the main cases is the [Facebook Oversight Board](#), but many other cases are worth noticing, from Ireland's planned "[Social Media Council](#)" to TikTok's "western" formed council.

Platform democracies are far from a stable model of regulation, but they foster intense debates around the limits of co-regulatory models, including unusual models of "regulated self-regulation". [Brazil](#) is one of the countries with advanced legislative debate. According to the model, legislation mandates platforms to set up councils (as self-regulation bodies), prescribe them with content moderation and procedural rules, and supervise them with a state-controlled agent. The idea has been presented in other contexts, including the [German policy debate](#).

But how can we understand the innovations in policy framing driven by these "platform democracies" debates? This is the overall question we pose in this short article.

[Soft power](#) refers to "the ability to affect others through the co-optive means of framing the agenda, persuading, and eliciting positive attraction in order to obtain preferred outcomes." Soft power focuses not on changing a subject's strategy (what hard power does), but rather on curtailing a subject's agenda, its first preferences. Hence the [relationship](#) between soft power and behavioral change, or constructivist theories such as framing, agenda-setting, and priming.

What is soft power?

Soft power is a term coined by [Joseph Nye](#) that has been widely applied as a resource for public diplomacy and state actors. For instance, Naren Chitty [notes](#) that soft power "as a term, represents a body of thought that is associated with resources invested in attraction-power as well as with strategies for using such resources to further actors' interests." The underlying idea is that while, in general,

governments remain the most powerful actors on the global stage, the stage is increasingly more crowded.

Critics of the terminology argue that other concepts such as ideology or culture are best fit to frame a subject's power preferences. Supporters of the terminology are aware of such criticism and argue that the concept fills a void in global international policy that helps to explain how governments (and large corporations) act.

In practical terms, soft power can be used for example to persuade individuals to join large groups, to cooperate in resource distribution, to facilitate political organization or even to frame prescribed actions. In this way, soft power is less a matter of governing and rather a matter of cooperation, learning, and growth, especially in larger contexts.

The elements of soft power in platform democracies

To measure how soft power influences platform democracy, we must rely on categories that are mostly intangible, such as culture, values, political ideals, and institutions. In this particular case, we decided to focus on three specific variables, namely values (e.g. human rights and democracy), policies (e.g. obligations to remove certain contents), and personalities (e.g. actors or institutions involved in platform democracy). The source of data (in particular the emphasis on variables used to frame the policy debate) is based on overall debate, but it is possible to evaluate in quali-quantum terms the reference to terms used in policy discourse.

Values

Three key values are frequently used to justify the policy framing of platform democracy: individual rights and human rights, digital constitutionalism (which encompasses a number of other values, including the rule of law, separation of powers, among others), and tech exceptionalism.

The values related to individual and/or human rights are a direct consequence of the selection of sources used to explain platform democracy. Voices and context are mostly selected from the EU and US regions, rather than coming from China and Russia, for example. This is true even for social media councils from companies based outside western regions, such as TikTok.

With this in mind, it is easier to understand the emphasis on fair procedures and individual rights protection (a reference in particular to the EU region) or from corporate social responsibility principles, such as external oversight (a reference in particular to the US context). In the same direction, platforms and governments in the debate emphasize the valorization of Human Rights standards, in particular those grounded in the United Nations Guiding Principles for Business and Human Rights (UNGP), framed as a "soft law" that "defines a corporate responsibility to respect human rights."

A second value is the rule of law, a cornerstone of the digital constitutionalism narrative. Looking at it this way, we can understand the relevance of state-like principles in the private domain, as in the case of platforms assuming state-like obligations usually attributed to state authority. Hence the importance of values such as procedural fairness, transparency, decision-making accountability, access to information, participatory governance structures, and more.

Lastly, we consider the value of "tech exceptionalism," the value that some companies are not traditional companies, but rather a special type of company. On the one hand, this value highlights that excessive regulation could stifle innovation and limit economic growth, which can be avoided with a more flexible and adaptive regulatory approach. On the other hand, the value focuses on the particularities of content moderation activities by online intermediaries, leading to debates around algorithmic curation, content

removal and agent behavior. The problem here is one of scope: in the digital ecosystem, social media platforms do not work mainly with online content intermediation.

Tech exceptionalism ends up setting a division between digital companies, one that may not have ground in real markets. In this policy framing value, platforms are not digital companies per se (because they work with content-moderation while others don't), but platforms may also profit from other digital markers (such as the link between digital payments and social media).

Policies

We identify three key policies frequently used in the overall debate to justify the policy framing of platform democracy: terms of service, multistakeholder governance, and co-regulation.

Rules and guidelines regulating social media platforms are usually codified as “terms of service,” a set of legal agreements between the provider and its users that outlines the rules, responsibilities, and guidelines that users must adhere to when using the platform's services. Those policies highlight for example how content is taken down or stays up on online platforms and are increasingly used by governments to justify actions where state regulation is yet to be adopted, with a potential hybridization between state and non-state governance.

Multistakeholder governance refers to the policies used by platform democracy settings to institutionalize decision-making models that promote participation from non-state and non-private actors, notably those coming from academia and civil society organizations. The focus on multistakeholderism promotes governance models that move away from classic industry-led self-regulatory organizations, and away from state-driven hard regulation, favoring a ‘hybrid governance’ system.

Lastly, we identify the emphasis on co-regulation. The policy frame here favors that platform democracy is best fit to avoid the pitfalls of both hard regulation and self-regulation. The innovation comes in the form of variations of the co-regulatory model, that lead to formats that emulate platform democracy as advisory, quasi-legislative, and quasi-judicial institutions.

Institutions

When looking at the plethora of institutions involved in platform democracy, we identify an emphasis on voices coming from journalists, civil society, academia, and platforms.

Fact-checkers, and journalists in general, are often believed to be mandatory actors to better understand mis/disinformation content-moderation. This group is portrayed in the overall debate as individuals who should be supported (including with economic incentives), and who have a role in social media councils to contribute for example as suppliers of evidence or as community voices.

Civil society and academia are emphasized in the overall debate as highly valuable members of social media councils, occupying roles related not only to advice, but also on rulemaking and decision-making. There is also emphasis on diversity of members composing the boards, a selection criterion set from start when creating such forms of external oversight.

Lastly, we highlight the role of platforms as institutions. Social media councils are mostly defined as non-binding forums of decision-making for platforms, which highlights the role of companies themselves to participate in the debate, as high-level voices, and decision-makers.

What next on framing platform democracy's soft power?

It seems likely that a consensus will emerge around the pros and cons of social media councils as content moderation oversight bodies. The level of consistency between values, policies and institutions shows very little variation indeed.

Priming values explain why platform democracy values focus on individual rights, human rights, and digital constitutionalism. The association favors the expansion of state-like regulation to private companies activities. Some read this new frontier as problematic, making private-sector potentially arbitrary and contradictory decisions as validated state-like mechanisms. Others read the scenario in a positive light, considering such councils as an innovation to overcome the worst uses of private sector or state sector for speech moderation.

Another crucial value is explaining social media companies within a tech exceptionalism light. Some will point to the risks of such an approach in overlooking market concentration aspects; others will see this as the necessary justification to advance in new models of co-regulation even further.

It is also important how the participation of civil society and academia is defined as mandatory. Some will see in this case an opportunity for better oversight of platform self-regulation, while others will consider the participation of such actors as insufficient.

A Council for Consilience: How could a council foster a field researching the information environment?

Alicia Wanless

CARNEGIE ENDOWMENT FOR INTERNATIONAL PEACE, WASHINGTON D.C., USA

Councils have long played a role in human history, providing an organising function to bring some community to bear on a problem. Today, democracies face a pressing and complex problem in a polluted information environment. Yet the very act of governmental intervention to address challenges within the information environment raises questions about democratic legitimacy. Many desired interventions, such as banning bad actors and disinformation from social media platforms, resemble authoritarian approaches and are being implemented simultaneously as trust in public institutions plummets. Moreover, little is known about the impact of many of these interventions because research of this type is just emerging, and its practitioners lack consilience and are not supported to do such work at scale and over time. Thus, we have a situation where policymakers need to understand the information environment which is largely controlled by private businesses, and researchers currently study in silos from various disciplines without common terms and methods. In democracies, this means that any attempt to do something about the information environment must be inherently multistakeholder, even if that first step is simply to understand that system. Could a council approach enable a multistakeholder effort to better understand the information environment in the context of democracy?

What's the problem?

Much of the problem is that we don't know what we don't know. The information environment isn't being studied as a system, making it difficult to put research into context. It is tempting to try to assess the effects of specific pieces of technology, individual threats like disinformation, or campaigns run by threat actors, but if we don't understand the system in which they operate, it is impossible to understand the effect of one variable. The two major reasons for this ignorance about the information environment are a lack of consilience, or shared understanding across disciplines, and gaps in resources supporting research.

The consilience problem stems from the fact that researchers assess the information environment through the lens of their own disciplines, whether in media studies, ethnography, computer science, or psychology. This leads to a lack of shared terminology and methods for studying the information environment. It's difficult to study a system if we aren't seeing its entirety and building on the work of others in a systematised manner. For a scientific understanding of the information environment to emerge, a convergence of disciplines must occur, with researchers speaking a shared language and working consistently together on the topic.

The second issue of resource gaps is wide-ranging. Gaps include researchers' lack of access to data generated by social media companies. Researchers also lack access to platforms to conduct studies, such as measuring the impact of interventions. Even if researchers had access to social media data, many lack the engineering resources and infrastructure to make sense of that data at scale. This is to say nothing of

a lack of diversity among researchers who mostly come from North America and Europe and, even then, do not represent the diverse communities and experiences of all those living there.

In short, massive gaps impede an understanding of the information environment at the very time that democratic societies must quickly get a handle on the role of that system, lest the interventions introduced today have unintended consequences.

It takes a village...

Filling these gaps requires a multistakeholder community. Stakeholders from across sectors must be brought to bear to fill the many resource and research gaps. There are four key types of stakeholders: researchers from academia and civil society who lead in framing understanding of the information environment; governments who set policies and enforce regulations related to the governance of the information environment; companies that often control major aspects of the information environment; and citizens whose agency to make free and informed decisions within the information environment is necessary to democracy and therefore should be a part of understanding and governing it as well.

Fostering consilience entails connecting the work and approaches of different fields to allow researchers to build on each other's work and collectively develop an understanding of how the information environment works as a system. It requires systematising what is already known about the information environment, finding consensus on shared terminology, and building frameworks for studying the information environment that can transcend time, geographies, and the inevitable introduction of new technology that changes how we process and communicate information. While efforts are emerging to systematise existing research, for example, through an [Intergovernmental Panel on Information Environment](#), other questions like how the information environment should be framed and studied as a system and what other types of data can be used to do so also need answering.

Researchers have also been working to address the resource gaps outlined above. The [European Digital Media Observatory Working Group on Platform-to-Researcher Data Access](#) is developing detailed guidelines for researcher data access. The University of Michigan is developing the [Social Media Archive](#), accepting data deposits, and making more data available for research purposes. [Social Science and Humanities Open Cloud for the Netherlands](#) will help researchers share data from images, surveys and social media. The [Observatory for Online Human and Platform Behavior](#) at Northeastern University “captures the online behavior of a large sample of volunteers” generating data for research. Princeton University has begun [developing shared engineering infrastructure](#). These initiatives should be commended, but if democratic societies want to understand the information environment at the speed necessary to address current challenges, this community must come together to foster consilience and champion shared resources at a scale that can support a larger community of researchers than in single countries or their own institutions.

Beyond researchers, governments are also an important stakeholder. Democratic governments have an obligation to understand the impact of the interventions they are making in the information environment, lest the cure is worse than the disease. Governments can affect significant change that leads to an understanding of the information environment faster through funding to fill resource gaps, and also by mandating that online services publish [operational reporting](#) to increase transparency on how they operate and share data with researchers.

Tech companies are also key. They bear a considerable responsibility to support research on the information environment, given their role within and the revenues derived from it. Regulation will also increasingly compel industry to find solutions to safeguard their users and ensure their operations are not degrading the information environment.

Another stakeholder that is frequently the focus of interventions but often forgotten in consultations is citizens. Finding ways to meaningfully engage them in research on the information environment is key for ensuring the longevity of democracy. Citizens should absolutely be informing the ethical principles that govern research of the information environment, not to mention the lifecycle of data generated by them through social media.

While much activity is underway across stakeholder types, these efforts are often disconnected from (and sometimes at odds with) each other. The paper now considers whether councils offer a method of efficiently convening multiple stakeholders without condemning efforts to elite capture, particularly by governments and industry.

...or a village council

Councils have been used by a variety of communities to solve problems. A council – the Council of Europe – was used to bring a fractured continent together after the ravages of war. But they have also shaped and controlled religions. Revolutionaries saw councils as a means to overthrow the existing order. Others pinned hopes for greater democracy on citizen engagement through councils. More recently, the concept has been applied to hold tech companies to account through social media councils. In short, councils are what people make them. So, what makes a council?

Councils represent a sort of community, coming together to collectively make decisions like solving a problem. Councils tend to be framed by rules governing their activity, persist over time, and comprise a limited number of selected people. For councils to emerge, they need a catalyst – they take work. These limitations can impede the level of diversity of council members, and of democracy, most councils can expect to employ. In other words, councils embody a paradox of top-down action to foster a bottom-up approach. What follows is an exploration, drawing on the author's current efforts to build a council in ways that leave room for a community to emerge in a ground-up fashion as members develop it together.

In this example, Jacob N. Shapiro and I are starting with an idea that other researchers have also expressed: creating the equivalent of a European Centre for Nuclear Research, a CERN for the information environment. This multinational research facility, the Institute for Research on the Information Environment (IRIE), would develop shared infrastructure to speed up research on the information environment to inform policymaking. A council guiding IRIE's development might form for a year or so to answer challenging questions that would inform the institute's development and operations.

Resist the urge to answer everything

I have never built a multinational, multistakeholder research centre before. Few probably have. I can tell you that our instinct, perhaps as North Americans, was to get a group together and build it. But as a small number of us puzzled through the next steps, the difficult questions needing answers piled up. Like, what ethical principles should govern this research or what funding models will ensure the

independence of IRIE? At the time of this writing, fifteen questions have emerged that directly relate to how something like IRIE should be structured. There are likely more. Many of these questions are challenging and given the relationship of the information environment to democracy, they shouldn't just be answered by a small group of researchers from elite institutions in one part of the world. But even if we wanted to answer all these questions ourselves, it would take years of research. My first lesson has thus been to resist the urge to try to have all the answers at the outset. This problem is immense. But if a wider research community comes together, as part of a council, to answer these questions, we can speed up research to support policymaking.

Start Small

Trying to build a council from the ground up can seem counterintuitive. The trick is constructing a tent that can remain open with enough space to welcome new and unforeseen community members as they emerge. With that in mind, maybe the council isn't designed to be completed with a fixed number of members and launched fully formed. Rather, it should develop over time based on the number of questions that arise needing to be answered before IRIE could ever operate. Council building could start smaller by identifying who might already be answering those questions identified or whose work might lead to answering them. This can be done through consultations with other stakeholders working on related issues. Through this process, a list can be generated to begin mapping the wider community. In some consultations, leaders might emerge who volunteer to take on questions. In other cases, the work of someone might be so advanced it makes sense to approach those leaders and gauge their interest in being part of a growing community working towards a common goal of building a field and providing it with shared infrastructure.

Keeping diversity in mind as the community forms is important, but we are all limited by our own networks that introduce a certain degree of exclusivity by virtue of whom we know. Moreover, my limited and privileged experience might preclude me from even conceiving of the barriers to entry others might face. Indeed, the way some of these questions are framed might already be introducing bias that makes it hard for researchers to engage from other countries and backgrounds than mine. To that end, it isn't sufficient to simply construct a council with a mix of backgrounds over geographies but to include questions as they arise from researchers with a diversity of perspectives. Likewise, it isn't enough to invite someone to a table under our own terms, but we must make space for those who come later to inform our collective efforts, quite possibly from a vantage point we haven't yet considered and help ensure that these insights are heard by others.

IRIE COUNCIL QUESTIONS

- What can science already tell us about the information environment?
- How should the information environment be framed and studied as a system?
- What kinds of data can be used to study the information environment?
- How will legal regimes governing aspects of the information environment affect IRIE?
- What is the mechanism by which data can be made available for research purposes?
- How can citizens be better engaged in research on the information environment?
- How can a greater diversity of research be fostered?
- What is the relationship of data storage to the research lifecycle?

- What are the most pressing research questions that need to be answered, and how can they be studied?
- What ethical principles should govern this research?
- What funding models will ensure the independence of IRIE?
- What kinds of large-scale engineering infrastructure can speed up that research?
- How can a feedback loop be best created such that research informs policy-making?
- How can capacity be developed to help researchers make use of IRIE and new opportunities to data access offered by the Digital Services Act (DSA)?
- What is the impact of this research on the physical environment?

This approach fosters a community to answer the above questions and work together towards a shared vision. While some community leaders might already be working to answer integral questions and be well-supported, others might need to be identified to start work and will need funding. The key to success is taking a collaborative approach to work together and matchmake interested donors with community members willing to undertake answering pressing questions while also being open to growing the council as new questions need to be answered. In answering their questions, researchers might be encouraged to take a mixed-method approach and engage as broad a community as possible to encourage diversity. Ultimately, it is up to that leader to decide how best to answer their question and whom to engage in so doing. Each leader would be asked to develop a short one-to-two-page proposal outlining their plans and resource requirements. Taken together, these proposals feed into the requirements for a wider council that can be fundraised together or in parts, with many donors engaged across geographies to support the emerging effort.

Wait for a Home

While most councils start with a home and are built top-down starting within, building from the ground up might mean waiting to find a place to house it. In this case, it involved putting an idea out to the community, bringing an initial group of committed leaders to bear, and stepping aside to allow that emerging community to choose its leader and a home for the future council in a third-party organisation that can carry the project forward.

For a multinational research facility like IRIE, it's important that such a home be non-partisan and not viewed by the wider research community as competition. If the aim is to speed up research on the information environment to generate evidence for policymaking, then it must be accessible to the widest community of researchers possible, facilitating their research instead of competing with them. It must foster collaboration. Philanthropies with a long track record of supporting or recognizing intellectual endeavours, like the Nobel Foundation and Kofi Annan Foundation, come to mind. Ultimately, the initial members should choose where this emerging council is housed.

Find Funding Carefully

It is up to leaders to choose funding sources that fit their comfort level. The least complicated sources usually are philanthropies. For the perceived integrity of the overall project, who funds the council itself and provides it with a home must be chosen very carefully and not be affiliated with politics. Donors who support leaders answering research questions can be more diverse, opening the door to include smaller philanthropies in this work. This approach also enables democratic governments to directly support researchers in their country, contributing to the overall effort.

Something at the scale of a multinational research facility can only be sustained over time following a similar approach to the CERN, whereby multiple democratic countries agree to contribute each year based on their GDP. To protect IRIE from perceived interference by donors, this commitment must be made over a long-term basis into decades and not up for political approval on a renewal basis. While few countries have the resources to fund such a centre alone, thirty countries sharing the burden of contributing \$1M yearly start to make a multinational research facility a reality. However, given the timelines most governments work on, it could take years before such commitments are made. Taking a ground-up approach to building and funding the council piecemeal helps the community collaborate now to have the answers needed when that day comes.

In the short term, drawing on industry money to answer IRIE questions will be discouraged. No matter how well-intentioned that support might be, the perception of industry support could taint the overall project. In the long term, one of the key questions that must be answered through this process is how IRIE can be sustainably funded and maintain its independence. It would be unreasonable and perhaps unjust not to expect social media companies and those who profited from them to support efforts to understand and improve their effects, but all parties must recognize the need to protect against the reality or appearance of research being shaped by donor interests.

Hinder Capture by Industry and Government

Unfortunately, many tech companies and governments are not without past transgressions when it comes to the information environment. Both stakeholder types also wield more power than those in civil society. For these reasons, researchers from across academia and think tanks bear the burden of leading on answering questions and engaging stakeholders from industry or government in a manner that can prevent capture by the latter. However, stakeholders from multilateral organisations representing democratic governments could potentially lead in answering questions, as their mandate to support member states often entails information gathering rather than shaping the agenda of an external effort to suit the needs of a specific country.

Build a Council

A council might begin to emerge after a certain percentage of the identified questions have leaders committed to answer them. Yet, to encourage diversity and enable a wider community to become engaged, this might be a point to stop and try other forms of outreach, such as hosting a meeting on the side of a related conference, such as the [International Conference on Computational Social Science](#) or the [Paris Peace Forum](#), depending on the community needing to be reached. Indeed, finding opportunities to bring leaders and the wider community concerned with the integrity of the information environment together creates a means for growth and fostering a field. It reinforces the purpose of existing conferences while also fostering community-building more efficiently. After all, the information environment is complex, and the problems are so numerous and challenging that there is more than enough work for everyone. We must work together. Moreover, through these engagements, more leaders will emerge willing to answer the remaining questions or new ones as they emerge.

The council is ultimately formed of those leaders answering questions in pursuit of a shared aim, such as building a multinational research facility to study the information environment. Together, the council chooses a chair and a home. The function of the council in the short term is to share information between those answering different questions to ensure each member's efforts help head in their shared direction.

In some cases, answering one question, such as how the information environment should be framed and studied, might feed into the answering of another, for example, what data might be used to study it and vice versa. Each leader must publish a final report as part of their effort.

While the initial purpose of the council is to share information between members to help the community work towards a shared aim, as a final step, it could also be tasked with drafting the plan to build a multinational research facility. Indeed, each of the final reports would feed into this process. How the institute would be governed and structured would ultimately be determined by this process, informing what role a future council might have, if any. And in this manner, a community can develop to tackle a complex issue more quickly.

Keep the Tent Open

Even with a council formed, it must not be viewed as a *fait accompli*. Questions that had not been considered will inevitably arise, and answers needed to inform plans for a multinational research facility. Council members should resist the urge to answer more questions themselves and engage others who might emerge later wanting to support the initiative. They can support other leaders but should lead on only one question. This will help build community and collaboration to address more outstanding questions faster. Similarly, others might emerge who want to contribute shared infrastructure or research help in finding a path to keep the tent open to newcomers so as to support as wide and diverse a community as possible across democracies. Indeed, several of the known questions relate to keeping the tent open, such as in finding ways to engage citizens in research on the information environment and fostering greater diversity of research.

Newcomers might fit two categories: those who take on answering key questions that inform the governance and structure of IRIE and those who contribute to practical operations. The former would consist of leaders tackling a new question that must be answered before a plan can be drafted for IRIE and should be incorporated into the council. The latter are core long-term partners for IRIE should it come to exist in whatever form it takes. This could include philanthropies interested in supporting the effort, policymakers with pressing problems that need research to address, as well as a broad range of actors from the research community, such as those making data available to study, those conducting measurements on aspects of the information environment, others building tools to support such work, and the wider community who would benefit from shared infrastructure. Essentially, the shared vision is the tent under which we build and foster a wider community committed to understanding the information environment in the context of democracies.

Conclusion

There are many ways to build a council, each with its pros and cons. As noted, the approach outlined in this paper is paradoxical; building a council from the bottom up still requires top-down action to start. It is an awkward chicken-and-egg situation. Not taking on the entire leadership to build a council can make the process unruly and not intuitive for others to follow. Likewise, funding in pieces to support the answering of individual questions might not be compelling for researchers to want to step forward and tackle problem sets, but on the flip side, it enables greater outreach to the community and the ability to engage a variety of donors and form networks to support future work.

Building a council from the ground up is inductive and, as such, must be open to continuous improvement. Indeed, existing council members should remain open to new findings. This doesn't necessarily mean changing course entirely, but being able to adapt as we all learn together. To that end, it should have an inherent feedback loop between council members enabling their findings to help inform the work of the whole and, ultimately IRIE. We all bring something to the table, and making sense of the information environment will require many perspectives and skills. Consilience will take compromise. We are working towards the same end. Instead of saying no, we could commit to finding solutions. Together these three ideas could form the principles framing the council.

If this approach works, a community working together can do more and faster. Likewise, working together as part of a wider community could, in turn, create a groundswell across democracies to champion IRIE and make it a reality – something none of us can do alone. Governments in democracies won't tackle something this big if researchers in their own backyards don't see a need for it. Moreover, while not intuitive, keeping the tent open could help identify a wide network of researchers, each studying aspects of the information environment from their own lens to help build the consilience and diversity in research that is so badly needed.

Only time will tell whether this approach will work or not. Although, at the time of writing this, of the fifteen known questions, researchers have stepped forward to answer nine. Even if the bigger goal of creating IRIE doesn't happen, the act of community building in pursuit of it could foster a field, and that alone would be huge.