Three Eras of Digital Governance

Three Eras of Digital Governance

Jonathan Zittrain[1]

To understand where digital governance is going, we must take stock of where it's been, because the timbre of mainstream thinking around digital governance today is dramatically different than it was when study of "Internet governance" coalesced in the late 1990s.

Perhaps the most obvious change has been from emphasizing networked technologies' positive effects and promise – couched around concepts like connectivity, innovation, and, by this author, "generativity" – to pointing out their harms and threats. It's not that threats weren't previously recognized, but rather that they were more often seen in external clamps on technological development and upon the corresponding new freedoms for users, whether government intervention to block VOIP services like Skype to protect incumbent telco revenues, or in the shaping of technology to effect undue surveillance, whether for government or corporate purposes.

The shift in emphasis from positive to negative corresponds to a change in the overarching frameworks for talking about regulating information technology. We have moved from a discourse around *rights* – particularly those of end-users, and the ways in which abstention by intermediaries is important to facilitate citizen flourishing – to one of *public health*, which naturally asks for a weighing of the systemic benefits or harms of a technology, and to think about what systemic interventions might curtail its apparent excesses.

Each framework captures important values around the use of technology that can both empower and limit individual freedom of action, including to engage in harmful conduct. Our goal today should be to identify where competing values frameworks themselves preclude understanding of others' positions about regulation, and to see if we can map a path forward that, if not reconciling the frameworks, allows for satisfying, if ever-evolving, resolutions to immediate questions of public and private governance.

The Rights Era

The original consideration of threats as external to the otherwise-mostly-beneficial uses of tech made for a ready framing of Internet governance issues around rights, and in particular a classic libertarian ethos of the preservation of rapidly-growing individual affordances in speech – "now anyone can speak without a gatekeeper!" – against encroachment by government censorship[2] or corporate pushback motivated by the disruption of established business models.

---

[1] George Bemis Professor of International Law, Harvard Law School and Harvard Kennedy School of Government; Professor of Computer Science, Harvard John A. Paulsen School of Engineering and Applied Sciences. I thank John Bowers for top-notch research assistance.

[2] Two noteworthy entries in this genre are Timothy May's "The Crypto Anarchist Manifesto" (1988), and John Perry Barlow's "A Declaration of the Independence of Cyberspace" (1997). Both May and Barlow inveigh against governments bent on applying conventional rules and governance standards to the new terrain of cyberspace, a practice which they portray as being both intellectually bankrupt and doomed to fail. Both argue that cyberspace represents a fundamentally new and different sort of social and political construct, with a rights paradigm that is entirely its own. Per Barlow, "Governments derive their just powers from the consent of the governed. You have neither solicited nor received ours. We did not invite you. You do not know us, nor do you know our world. Cyberspace does not lie within your borders. Do not think that you can build it, as though it were a public construction project. You cannot. It is an act of nature and it grows itself through our collective actions." Both – in these writings and elsewhere – are fascinated by questions of intellectual property, fertile ground for the disruption of traditional structures of right and privilege. Gleefully anticipating the reconfiguring effects of cryptography, May

A good example in the first category are the debates around the U.S. Communications Decency Act of 1995, which sought to keep indecent material away from minors by penalizing those who indiscriminately made it available online. The Supreme Court struck down the core provisions of the CDA in 1997 on First Amendment grounds, holding that too much protected speech would be chilled by the law, and successor laws met a similar fate.[3] Another example can be found in the early and then not-officially-acknowledged efforts by the Chinese government to block citizens' access to web sites critical of the state, something viewed among those studying Internet governance as an unalloyed wrong, not least because of the lack of due process, including notification, in effecting any blocks.

When the Internet's affordances for near-instant file transfer led to objections by publishers and other copyright holders over copyright infringement, those against stepped-up enforcement or new requirements for intermediaries relied on a rights-centric account.[4] Copyright itself establishes legally protected interests – rights – but the sorts of interventions required to continue to secure those rights in practice were described early and often as overly burdening individual rights, whether through content takedown schemes to be effectuated by intermediaries, or individual lawsuits filed against those engaged in the sharing of copyright material.

It is in intermediary liability that the most significant regulatory battles have unfolded, and that is likely to remain so. The shaping of end-user behavior through rule and sanction was, and is, difficult. But intermediaries can be persuaded or required to shape users' technological experiences to channel them away from objectionable or illegal behavior, whether through hardware or operating system design of smart phones, or the shaping of software and services used by billions, such as by the most prominent social media platforms. The rights framework generally finds that such shaping should be limited, and in the late 1990s that was reflected in American law. For example, section 230 of the Communications Decency Act -- a part of the Act that remained after the Supreme Court struck down the rest -- provided for immunity by platforms against many forms of potential liability occasioned by those platforms hosting and amplifying the speech of others, including end-users. And the notice-and-takedown safe harbors of the Digital Millennium Copyright Act offered a low-impact, routinized way for platforms to respond on a case-by-case basis to copyright complaints for others' material. Still, some scholars advocating for a rights framework thought these provisions went too far.[5]

---

writes that "just as a seemingly minor invention like barbed wire made possible the fencing-off of vast ranches and farms, thus altering forever the concepts of land and property rights in the frontier West, so too will the seemingly minor discovery [of cryptography] come to be the wire clippers which dismantle the barbed wire around intellectual property."

[3] For an account of the protracted 10-year judicial struggle over the constitutionality of the Child Online Protection Act, itself developed in the wake of the CDA strikedown, refer to this blog post from Lauren Gelman at Stanford Law School.

[4] Here too Barlow's writings comprise a well-known exemplar. In his 1992 essay "Selling Wine Without Bottles: The Economy of Mind on the Global Net," Barlow once again takes aim at the lawyers and corporations vigorously defending what he sees to be entirely obsolete copyright doctrine: "Intellectual property law cannot be patched, retrofitted, or expanded to contain the gasses of digitized expression… Most of the people who actually create soft property – the programmers, hackers, and Net surfers – already know this. Unfortunately, neither the companies they work for nor the lawyers these companies hire have enough direct experience with immaterial goods to understand why they are so problematic. They are proceeding as though the old laws can somehow be made to work, either by grotesque expansion or by force. They are wrong." Barlow's words contra governments and corporations alike presage the copyright wars of the late 1990s and early 2000s, in which old and new theories of rights competed – and ultimately compromised – to become doctrine.

[5] Many of these concerns relate to the DMCA notice-and-takedown system's potential utility as a private censorship tool. In a *New Republic* article from 2000 entitled "Call it the Digital Millennium Censorship Act: Unfair Use", intellectual property scholar Julie Cohen argued that "DMCA's notice and takedown provisions – which don't

It was also in this rights-centric era that ICANN came about, chartered to bring consistency and "stakeholder" representation to policy-inflected decisions around global Internet naming and numbering, such as the number and nature of top-level domains (TLDs) like .com and .uk, including who would be charged with giving out or selling second-level names under those domains, and under what conditions. Apart from the simple desire to establish and regularize who would be earning money from the sale of domain names, the main concern aired as ICANN came into its own was about whether ICANN would itself become a censor of Internet content.[6] ICANN could, the theory went, use its certification of TLD registries to, through a cascade of contracts, make for the suspension or transfer of domain names comprising or pointing to "bad stuff." Describing material in more precise terms of outright illegality has been difficult, since it would require a choice of which jurisdiction's definition of illegality to apply.[7]

As it has happened, concerns about ICANN becoming the Internet police – infringing on individual rights – has so far seen ICANN's catalyzation of a suspension power to be only in the area of domain names whose very nature indicate a bad faith registration amounting to a form of trademark infringement.[8]

---

require prior court review of takedown demands – threaten to substitute private censorship for judicial process." These concerns persisted as the DMCA matured. In 2004, Siva Vaidhyanathan wrote that "DMCA...has emerged as the law of choice for censoring criticism and commentary in the electronic environment." In 2019, Eugene Volokh wrote about the continuing use of fraudulent DMCA takedown requests to suppress online content – including, bizarrely, his own writings on fraudulent DMCA takedowns.

[6] Indeed, ICANN's potential capabilities in the censorship domain struck some as being particularly troubling exactly because ICANN was formed as a non-governmental entity, and therefore functioned at some distance from conventional modes of democratic recourse. In a widely-cited article titled "ICANN and the Problem of Legitimacy" – published in 2000, soon after ICANN's 1998 founding – Jonathan Weinberg warned that "ICANN's role is one generally played in our society by public entities. It is setting rules for an international communications medium of surpassing importance. That task had historically been performed by a U.S. government contractor in an explicitly public-regarding manner. ICANN is addressing important public policy issues. Further, it is implementing some of its choices via means that look uncannily like command-and-control regulation. If ICANN is to establish its legitimacy, it must be able to answer the charge that its exercise of authority is inconsistent with our ordinary understandings about public power and public policymaking." Milton Mueller's 1999 paper "ICANN and Internet Governance: Sorting Through the Debris of 'Self-Regulation'" offers analysis of a range of other potential issues with a "self-regulating" ICANN: "ICANN looks and acts more like an incipient inter-governmental agency than a private sector corporation. The process of forming ICANN has been mired in so much factionalism and political controversy that references to 'consensus based' self-regulation are laughable." Skepticism towards ICANN even had a homepage: icannwatch.org (2002 archive) was launched as a sort of watchdog to monitor these concerns and others.

ICANN's charter established that substantive decisions regarding the rights and privileges of individuals seeking to take part in a transformative communications technology would be delegated to a non-public entity. As this essay will go on to argue further, we've seen similar – and perhaps less technocratic, more visible – tensions play out in controversies around the content governance practices of contemporary internet platform companies.

[7] What's more, the impracticality of the notion of an "extraterritorial" internet unbound by the laws of any given country – favored by many rights mavens of the early internet era – factors into the analysis here. These narratives were complicated by the fact that, in the early days of the internet and now, governments, and particularly those with authoritarian characteristics, have a tendency to pressure internet companies to police content and expression in accordance with localized laws and norms. Jack Goldsmith and Tim Wu offer a thoughtful reflection on this tempering of the dreams of the techno-utopians in their 2006 book *Who Controls the Internet: Illusions of a Borderless World*. So even if ICANN were to attempt to establish itself as an interventionary global governance body, its actions would nonetheless remain subject to those of governments themselves.

[8] ICANN's process for adjudicating copyright disputes over domain names is the Uniform Domain-Name Dispute-Resolution Policy, launched in December of 1999. The UDRP holds that, in order to wrest control of a domain name from a registrant, a complainant must prove three elements:

Domain names that are not so infringing, but that are used as mnemonics for destinations containing harmful or illegal content, have generally not been touched by ICANN's policies.[9]

The Public Health Era

I was among those who celebrated the benefits of a rapidly-expanding Internet, both in scope and capability, thanks to the generative contributions of millions of users in code and content. For example, Internet protocols made possible the growth of the World Wide Web as an Internet application without

---

> "(i) your domain name is identical or confusingly similar to a trademark or service mark in which the complainant has rights; and
> (ii) you have no rights or legitimate interests in respect of the domain name; and
> (iii) your domain name has been registered and is being used in bad faith."

The apparent ambiguity of the UDRP has long been a topic of consternation amongst lawyers and policymakers. A Berkman Klein Center analysis from soon after its implementation points to dozens of precedential proceedings, with each one offering refinements to an interpretation of concepts like "bad faith" and "legitimate interests" under the UDPR.

[9] The idea that ICANN might for some reason begin to police forms of abuse – illegal or otherwise – unrelated to trademark protections has long concerned advocates for individual online rights. In 2013, ICANN revised its agreement with registrars to include what Electronic Frontier Foundation Jeremy Malcolm called, in a series of blog posts titled "EFF to ICANN: Don't Pick Up the Censor's Pen," a "provision requiring registrars to 'receive reports of abuse involving Registered Names' and to 'take reasonable and prompt steps to investigate and respond appropriately.'" Much was made of the ambiguity around this phrasing – what might a "report of abuse" entail beyond the domain of copyright? – prompting ICANN to release a lengthy 2015 blog post decisively titled "ICANN Is Not the Internet Content Police." In the post, ICANN Chief Contract Compliance Officer Allen R. Grogan argues that

> "Though the appropriate interpretation of 2013 RAA is the subject of debate, there are clear-cut boundaries between ICANN enforcing its contracts and the enforcement of laws and regulations by the institutions mentioned earlier. A blanket rule requiring suspension of any domain name alleged to be involved in illegal activity goes beyond ICANN's remit and would inevitably put ICANN in the position of interpreting and enforcing laws regulating website content. At worst, it would put ICANN squarely in the position of censoring, or requiring others to censor, Internet content."

In 2017, however, EFF and other allied organizations were raised a cry – detailed in the same EFF blog post – over ICANN's appointment of a former law enforcement official to the post of Consumer Safeguards Director. Per Malcolm,

> "a draft report [PDF] of ICANN's Competition, Consumer Trust and Consumer Choice Review Team recommends that strict new enforcement and reporting obligations should be made compulsory for any new top-level domains that ICANN adopts in the future. ICANN's Non-Commercial Stakeholder Group (NCSG) has explained [PDF] why many of these recommendations would be unnecessary and harmful.
>
> A subteam of this same Competition, Consumer Trust and Consumer Choice Review Team has also recently released a draft proposal [PDF] for the creation of a new DNS Abuse Dispute Resolution Procedure (DADRP) that would allow enforcement action to be taken by ICANN against an entire registry if that registry's top-level domain has too many "abusive" domain names in it... If this proposed DADRP goes ahead, registries could come under pressure to go on a purge of domains if they wish to avoid being sanctioned by ICANN."

any approvals sought or needed; the Web facilitated the rise of online wikis, and those wikis made possible the phenomenon of Wikipedia, which in turn invited contributions of content from people who themselves were not interested in coding software. Even amidst this celebration, in my case circa 2007, lay a new round of problems, which I described as part of the Generative Pattern:

1. An idea originates in a backwater.
2. It is ambitious but incomplete. It is partially implemented and released anyway, embracing the ethos of the procrastination principle.
3. Contribution is welcomed from all corners, resulting in an influx of usage.
4. Success is achieved beyond any expectation, and a higher profile draws even more usage.
5. Success is cut short: "There goes the neighborhood" as newer users are not conversant with the idea of experimentation and contribution, and other users are prepared to exploit the openness of the system to undesirable ends.
6. There is movement toward enclosure to prevent the problems that arise from the system's very popularity.

Indeed, the cutting short of success by those who subvert the system and take advantage of its now-many users – a problem arising from the very openness of the system itself – began in earnest by 2010.[10] Cybersecurity had been my central worry; it was clear those problems were no longer wholesale, business-to-business issues, but something touching all of users' online activities. Without urgent attention given to developing a collective, generative defense, I worried about the Generative Pattern's conclusion: top-down enclosure to protect everyone by curtailing everyone's freedoms, demanded by the users themselves.[11]

These kinds of concerns and how to meet them don't much benefit from a rights discourse, especially as they involve the mutual (if surely not symmetric) violation of rights by users against users, at least from a technical network point of view. Rather, they have much in common with how we talk about public health.[12] They emphasize the interlinkages among us, the way that problems can all too easily spread from

[10] In many cases – before and after 2010 – the abuse of the internet's affordances by users was met with applause or laughter. The practice of "trolling,' intentionally seeking to shock, annoy, or enrage other internet users, became both a hobby and a sort of spectator sport, with content consumers watching, often gleefully, the sowing of chaos. Whitney Phillips argues in a 2019 paper titled "It Wasn't Just the Trolls: Early Internet Culture, 'Fun,' and the Fires of Exclusionary Laughter" that the widespread acceptance (even embrace) of an internet culture comfortable with many forms of insensitivity and abuse laid much of the groundwork for the toxic online dynamics of today. Her account encourages us to look starkly at the internet libertarians of the rights era, many of whom were innoculated against all but the worst of these tendencies by their status as white men.

[11] The "walled gardens" of today's platforms are, in some sense, a manifestation of this natural conclusion. But these ostensibly tightly-controlled spaces have been the site of some of the most sustained claimed abuses and most immediately-apparent harms of the public health era, from allegedly social media-fueled genocides in Myanmar to the Cambridge Analytica scandal. Concentration creates new levers for governance over what might otherwise be messy generative networks, but it also offers up target-rich environments to those seeking to do harm.

[12] One clear marker of the shift from a discourse of rights to a discourse of public health has been careful reevaluation of the stipulations of CDA 230, with critics arguing that it often unduly insulates culpable internet platforms from responsibility for the harms arising from their actions. In their 2010 book *The Offensive Internet*, for example, Martha Nussbaum and Saul Levmore describe how the internet has generated unprecedented opportunities for reputational harm to individuals. This harm, they argue, has been enabled in large part by CDA 230: "A withdrawal of [CDA 230] immunity could, without constitutional difficulty, restore the symmetry between website operators and publishers of newspapers, which can of course be sued for damages if they publish defamatory material." Even many scholars uncomfortable with an aggressive rollback of CDA 230 have placed its provisions under a microscope. In her paper "The New Governers," Kate Klonick details the often unwieldy mixture of constitutional analogies, one-off decisions, and economic and political incentives which drive platforms' content governance paradigms under CDA 230.

one person or node to another, and the need for systemic intervention and shaping to prevent harm from accruing, regardless of who might be to blame for first injecting harm into the system. Worries around viral malware hopping from one server to another have grown to be worries about mis- and disinformation hopping from one credulous person to another, abetted by social network intermediaries who amplify controversial or outright false content if it increases user engagement with the platforms. Indeed, there is a literal public health dimension to misinformation today, as screeds and videos against even basic public vaccination, long proven to be beneficial, circulate and previously-near-defeated illnesses like measles make a startling comeback.[13]

A public health framework is much more geared around risks and benefits than around individual rights. Pointing out harmful speech in a rights discourse might typically result in what amounts to a shrug and a declaration that such excesses are the "price of freedom," a sign that our commitment to rights requires sacrifice precisely where people would otherwise find the exercise of rights objectionable. In the public health frame, we instead are asked to gather empirical data about benefits and harms, and to brainstorm ways that the latter might be decreased without unduly trimming the former.

The Process, or Legitimacy, Era

Reconciling rights and public health frameworks is not easy, not only between two people whose normative commitments fall into the respective camps, but also often within a single person: each framework can speak powerfully to us, favoring both individual liberty – including a skepticism over the responsible exercise of state power – while also sensitive to the fact that we live in a tightly-coupled, interlinked society, all the more so with the rise of networked technologies, and there are times when collective security calls for organized and perhaps even mindful architectural intervention. Moreover, the rise of intermediaries that not only facilitate communication with people we already know we want to reach – think email, or instant messaging – but also discovery of new ideas and people, means that there's a less-agreed-upon conception of neutrality or non-intervention. When Facebook or Twitter has millions of candidate items with which to salt a feed, any decision about what to show or recommend to you next is going to be freighted in a way that speeding delivery of a note between two discrete people is not.[14]

The recent case of Herrick v. Grindr has furnished an illustration of how CDA 230 protections can insulate companies complicit in facilitating real-world harms from liability. The plaintiff's ex-boyfriend used the gay dating app to manipulate upwards of 1,000 men into threatening and harassing the plaintiff, often in real life, over the course of almost a year. The U.S. Court of Appeals for the Second Circuit confirmed a dismissal of the complaint on the grounds of Grindr's CDA 230 immunity in March of 2019.

[13] Many criticisms of platform behavior relating to vaccine controversies center on the sorting and ordering of content feeds, whether in the context of a search engine or social media site. It's worth noting, however, that public health concerns relating to content ordering are nothing new. In 2004, a Google search for 'jew' would return the anti-semitic website jewwatch.com. Google refused to alter its results, stating that "We find this result offensive, but the objectivity of our ranking function prevents us from making any changes." In the case of vaccine misinformation, however, pressure from lawmakers and the public has driven commitments to action on the part of Facebook and Twitter, among others.

[14] Some platforms have struggled to develop workable frameworks for navigating the (algorithmically mediated) spectrum between driving the virality of content and taking it down. When it comes to public health considerations, platforms now have a tendency to lean on the language of demotion rather than that of removal. Whether this tactical shift represents a move towards or away from censorship is very much up for debate.

In a November 2018 blog post entitled "A Blueprint for Content Governance and Enforcement," Mark Zuckerberg asserted – with visuals! – that "[internal Facebook] research suggests that no matter where we draw the lines for what is allowed, as a piece of content gets close to that line, people will engage with it more on average – even when they tell us afterwards they don't like the content." He proposes that this problem might be solved by demoting content as it approaches the line, inverting this engagement pattern and penalizing borderline content. But there may

We also happen to be in a time of very little trust in many if not most civic and private institutions, especially national and transnational ones. A simple vote in a legislature, or split decision from a court, seems not to well settle the complex and deeply debated issues that spring around digital governance.

This may be why we've lately seen some of today's most powerful private intermediaries, such as Facebook, Google, and Cloudflare, expressing uncertainty or contradiction about their own policies for intervention, a.k.a. intermeddling, vs. abstention, a.k.a. abdication.[15] The rise of mainstream AI means that even detailed policies can be applied – or misapplied – in real time to the activities of billions of people so voluminous to otherwise be beyond moderation.

These companies have made some attempts to take decisions about content or user behavior out of their terms-of-service, customer support channels, and into some new institutional configuration meant to match the gravity of the questions around abuse, harassment, and the promotion or stifling of political speech.[16] Facebook has proposed an independent review board, whose decisions would be binding upon the company. Others have sought internal boards to reflect upon ethically-freighted decisions before making them. And regulators, loathe to try to make the decisions themselves at scale, have sought to require private intermediaries to impose particular standards without offering much by way of detail, such as in the current implementation of the European right to be forgotten.

What the field of digital governance, and indeed the world at large, needs, are ideas for new institutions and institutional relationships that can come to closure, however temporary, on some of these questions, and, like the project of law and political processes themselves, understand that all views will not and cannot be reconciled. But ideally even those who feel they have lost in a particular dispute or debate will not feel that they have been taken advantage of, or that the project to which they are contributing and are subject to – some digital expression of ideas and power – is not morally bankrupt.

The key to the next era of digital governance lies not in some abstract evaluation of whether our affordances are structured in ways that are correct or incorrect on one person's view, but rather if they are legitimate because of the inclusive and deliberative, and where possible, federated, way in which they were settled.

---

be good reason to believe that provocative content that plays close to Facebook's boundaries without violating them serves an important discursive function. Controversial forms of speech may well verge into toxicity much of the time, but such speech can also communicate strong emotions, drive changes in norms, and generally constitute free and productive expression.

[15] In an August 2019 post describing Cloudflare's decision to halt service to 8chan, a discussion board associated with hate groups and the perpetrators of a number of mass shootings, Cloudflare CEO Matthew Prince appealed for guidance from public decisionmakers: "Cloudflare is not a government. While we've been successful as a company, that does not give us the political legitimacy to make determinations on what content is good and bad. Nor should it. Questions around content are real societal issues that need politically legitimate solutions. We will continue to engage with lawmakers around the world as they set the boundaries of what is acceptable in their countries through due process of law. And we will comply with those boundaries when and where they are set."

[16] A number of scholars including Thomas Kadri and Kate Klonick have argued that the specificity and impact of these decisionmaking processes call for a form of constitution-building within the platforms. In "Facebook v. Sullivan: Public Figures and Newsworthiness in Online Speech," the two argue for the articulation of clear, oversight-friendly processes for the establishment and review of content governance standards. Striking a balance between the representation of user interests and the complex operational realities of administering a platform will be one of the greatest challenges to face internet platforms to date.