Evaluating the legitimacy of platform governance: a review of research and a shared research agenda

Nicolas Suzor, Queensland University of Technology (QUT) School of Law, n.suzor@qut.edu.au.

Tess Van Geelen, Queensland University of Technology (QUT) School of Law, tessvangeelen@gmail.com.

Sarah Myers West, University of Southern California Annenberg School for Communication and Journalism, sarahmye@usc.edu

This paper provides an overview of the key values that we argue should underpin an index of the legitimacy of the governance of online intermediaries. The aim is ultimately to allow scholars to rank the policies and practices of intermediaries against core human rights values and principles of legitimate governance in a way that enables comparison across different intermediaries and over time. This work builds on the efforts of a broad range of researchers already working to systematically investigate the governance of social media platforms and telecommunications intermediaries. In this paper, we present our review and analysis of the work that has been carried out to date, using the digital constitutionalism literature to identify opportunities for further research and collaboration.

The theme of this special issue is the deeply contested political question of how human rights should be protected in a digital environment. ‘Digital constitutionalism’ encompasses “a constellation of initiatives that have sought to articulate a set of political rights, governance norms, and limitations on the exercise of power on the Internet” (Gill et al., 2015: 2). One of the key challenges of the digital constitutionalism project is to articulate appropriate limits on the private exercise of power by online intermediaries, how those limits may be imposed, and by whom.

This is a complex problem, and requires a new understanding of constitutionalism in an era where regulation is not only, or even not primarily, done by the state (Black, 2001). We take a broad view of ‘governance’ that encompasses all ‘organized efforts to manage the course of events in a social system’ (Burris et al., 2008: 3). This broad definition focuses attention particularly on the policies and practices of online intermediaries as key actors in the governance of online behaviour.

Intermediaries play a critical role in governing the internet by developing and managing its infrastructure. By ‘intermediaries’, we mean the broad range of entities that “bring together or facilitate transactions between third parties on the Internet” (OECD, 2010: 9). Intermediaries of all types --- the owners of physical pipes, the providers of core routing services, the search engines that make content visible, the content hosts, and the social media platforms --- all shape how people communicate in important but different ways. They are the focal points of control, where pressure can be most effectively deployed to influence user behavior (Goldsmith and Wu, 2006), and the decisions that they make have a real impact on public culture and the social and political lives of their users (DeNardis, 2014). In order to progress the political project of digital constitutionalism, more needs to be known about how intermediaries govern their networks and how their decisions impact the basic rights of individuals in different contexts. In this paper, we present an outline of the values we suggest ought to be measured, a guide to the work that has already been completed, and a sketch for future projects to build upon in their own research design.

Our approach: Evaluating governance and measuring change

Global interest in how digital intermediaries are governed is growing rapidly, and both governance practices and the intermediaries themselves are continuously evolving (Burgess, 2015). We suggest that an index is a promising approach to structure thinking about future research. Indices can help to identify trends over time and across different intermediaries in a way that furthers discussions about the role of private firms in promoting human rights, and when states have a responsibility to intervene (Hyden and Samuel, 2011; Langford and Fukuda-Parr, 2012). Well-designed indices can also help encourage firms to develop better governance practices (MacKinnon et al., 2016; Maréchal, 2015), as well as empower users to make more informed choices (at least in markets with effective competition). Ultimately, this evidence base can help lawmakers and advocates consider specific opportunities to improve laws, policies and practices (Cooley and Snyder, 2015). We are particularly hopeful that an index may help to enable collaboration and comparison between future independently designed research projects that address different aspects of governance and different sites of study.

In this paper, we seek to identify the human rights values that might inform an index of legitimacy of the governance of online intermediaries. To generate this list of values, we rely on the very useful analysis of Gill et al (2015), which extracted a list of values from thirty declarations of internet rights and similar documents. We supplement this list with human rights values drawn from the major international human rights instruments. We also include any relevant procedural governance values from major international rankings of state governance and from existing projects that have begun the work of ranking intermediaries on these values.

From this list of values, we focus on those that will be relevant to assessing the legitimacy of governance by online intermediaries. We understand legitimacy in the normative sense here: the contested political question of how the exercise of power can be morally justified. In this sense, legitimacy is a ‘meta-concept’ that enables comparisons between competing conceptions of morality and “opens up space for imagining and adopting alternative visions of legitimacy” (Thomas, 2014: 738). While human rights are universal, the extent to which firms owe duties to protect or promote these rights is contested and will differ according to context and jurisdiction (Green, 2001). Any theory of the responsibility of online intermediaries must accommodate the substantial differences between different providers, each with different degrees of visibility and control over the communications that pass over their networks. By including a broad range of values, we seek to help ground disparate political discussions about what societies expect from online intermediaries --- which rights we expect them to actively protect, in what contexts, and how.

We exclude measures that can only effectively be implemented at a social level. Specifically, we exclude two categories: Access and Education (including, for example, the ‘right to access the internet’ and ‘media and digital literacy’), and Openness and Stability of networks (including network neutrality, open standards, and interoperability). We similarly exclude human rights obligations that have only tangential application to online intermediaries – for example, rights related to human trafficking and literacy, and broader categories of economic rights that focus primarily on state obligations. While telecommunications intermediaries have some role to play in promoting each of these values, we see them as primarily an arena for state legislation and policy, at least in the immediate future.

Our analysis approximates a comprehensive identification of the components of an index of the legitimacy of online intermediaries. For each of the values we identified from human rights instruments and digital constitutionalism declarations, we identify existing methodologies and datasets from academic studies and civil society projects that could be used to inform the content of an index. Specifically, we cross-reference the methodology of each project to our set

of procedural and substantive governance values and human rights, in order to identify with some particularity what research exists and where further work is required.

It is important to note two limitations at this stage. First, the list of values we have generated is contestable. We have attempted to err on the side of inclusivity, on the basis that only some of the values we identify will be applicable in any given context. We include a number of rights that have not traditionally been applied to intermediaries, like personal security and dignity, arguing that these rights can have important online applications. We hope that this analysis provides an opportunity to ground a collaborative discussion about which of these (and which additional) rights and values should be included in an index. Second, the identification of values and potential data sources is a precursor to future work to develop a set of quantitative measures that can be deployed in practice to evaluate governance policies and practices. While several of the projects we outline have completed important work in this regard, we leave the task of synthesizing existing indicators and developing the missing measures for a future project.

Value	Human Rights Instruments and supporting documents	Existing measures from scholarly and civil society indices and evaluations
<i>Fundamental Rights and Freedoms</i>		
Freedom of Expression	UDHR, arts 18, 19; ICCPR, art 19.	Extensively covered in RDR & ToSHR.
Freedom of Religion	UDHR, art 18; ICCPR, art 18.	No clear measure.
Freedom of Association	UDHR, art 20; ICCPR, art 22.	No clear measure.
Personal Security and Dignity	UDHR, arts 3, 12; ICCPR, art 17.	No clear measure.
Racial Discrimination	CERD; UDHR, art 7; ICCPR arts 4, 20, 26; ICESCR art 2; Special Rapporteurs, 5, 6.	No clear measure.
Discrimination against Women	CEDAW; UDHR, art 7; ICCPR arts 4, 20, 26; ICESCR art 2; Special Rapporteurs, 5, 6.	No clear measure.
Discrimination against Persons with Disabilities	CRPD; UDHR, art 7; ICCPR arts 4, 20, 26; ICESCR art 2; Special Rapporteurs, 5, 6	No clear measure.
Protection of Children	CRC; UDHR art 25; ICCPR arts 14, 24; ICESCR arts 10, 12; Special Rapporteurs, preamble, 3a, 6e.	No clear measure.
Cultural and Linguistic Diversity	UDHR, art 27; ICCPR arts 1, 27; ICESCR arts 1, 15.	No clear measure.
Privacy	UDHR, art 12; ICCPR art 17.	Extensively covered in RDR & ToSHR.
<i>Procedural Values</i>		
Rule of Law	UDHR, arts 7, 10, 11; ICCPR, arts 2, 14, 26; Guiding Principles, II.B.16, II.B.18, II.B.20, II.B.21, II.B.23.	RDR (F1-4, 6; P1, 2, 11); ToSHR (Freedom of Expression 1; Due Process 1, 3). No clear measures for consent; stability of rules; equality of application; sanctions for misconduct of employees.
Due Process	UDHR, arts 10, 11; ICCPR, arts 2, 14, 26; Guiding Principles, II.A.11, II.A.14, II.A.15, II.B.16, II.B.17, II.B.19, II.B.20.	RDR (F5); ToSHR (Freedom of Expression 3, 4; Due Process 1-9).
Right to Legal Remedies	UDHR, art 8; ICCPR, art 2.	RDR (C6)
Transparency and Openness	Special Rapporteurs, preamble, 3, 6.	RDR (P11, F7, F8, F9, C3.2)
Multistakeholder and Participatory Governance	UDHR, art 21; ICCPR, art 25; Guiding Principles, II.B.18.	RDR (C5)

Table 1: Summary extract of proposed list of fundamental rights and procedural governance values. Complete table available at <appendix URL>.

Fundamental rights and freedoms

The first set of values to include in our index are the universal rights embodied in international human rights instruments. International human rights are generally binding against nation states, not private actors. In recent years, however, there has been increasing recognition that private actors have a key role to play in protecting human rights. The United Nations Guiding Principles on Business and Human Rights sets out a responsibility that private firms “should avoid infringing on the human rights of others and should address adverse human rights impacts with which they are involved” (2011: 13). This obligation comes from a ‘tripartite’ model (Ruggie, 2008) developed to understand the responsibility of non-state actors to respect human rights.

Internet and telecommunications intermediaries are deeply involved in a range of issues with relevance to substantive human rights (Kaye, 2016; UNESCO, 2014). The requirements that the Guiding Principles imply for intermediaries are set out in a report by the European Commission (2013). This framework has six basic components: companies must make a firm commitment to respect human rights and embed that commitment into their culture; continuously assess their impact; prevent or mitigate harms in which they are involved; track how well harms are addressed; communicate how harms are addressed; and remediate harm that they have caused or contributed to. These principles have been adopted by groups of state and non-state actors in relation to freedom of speech, rights of individual privacy, rights to be free from harassment and abuse, and obligations to resist pressure to unduly limit freedom of speech (e.g. IPRC, 2014; Global Network Initiative, 2012; UNESCO, 2014; Manila Principles on Intermediary Liability, 2015).

Both Ranking Digital Rights (2015) and the Terms of Service and Human Rights (Venturini et al., 2016) projects provide extensive evaluations of how freedom of expression and privacy rights are dealt with by intermediaries in formal documents. Together, these projects examine in great detail when and for what purpose intermediaries collect, retain and/or share user data, how much control users have over that process and the data itself, how the company responds to third party requests for information, and whether information is encrypted. They also examine whether rules about acceptable content are clear, whether moderation processes are disclosed, and whether there are adequate systems for resolving disputes. Importantly, however, neither project addresses the more contested substantive scope of restrictions on freedom of expression. Article 19 of the ICCPR (United Nations, 1966), for example, not only requires that limits on speech are clear and accessible, but also that they are designed to serve a legitimate purpose, and are both necessary and proportionate to achieving that goal. While restrictions on speech imposed by online intermediaries will often be acceptable where similar restrictions imposed by states would not be, this may not always be the case. Users of particular services often consent to more restrictive rules about acceptable speech than would be legitimate if imposed by a nation-state, so any measurement of this value should always be viewed in context (Suzor and Woodford, 2013). Nevertheless, the breadth of speech permitted by intermediaries is fundamentally important to the user experience. We suggest that this gap should be filled by evaluating the relative scope of restrictions on speech and the adequacy of reasons set out by intermediaries in justification.

There remain several substantive human rights based in international law and digital constitutionalism documents that are not considered in rankings undertaken to-date. These represent the most immediate opportunities to extend the work of existing indices. The right to personal security and dignity is a key component of a number of digital constitutionalism documents but not specifically used to evaluate intermediaries in existing studies. This right, particularly when conceptualized as freedom from harassment and abuse, is increasingly relevant in controversies about how online networks are governed. Recent work has highlighted the pervasive and massive extents of harmful abuse and hatred online (Citron, 2014; Woodlock, 2016). Intermediaries are increasingly being challenged on their role in facilitating the spread of abuse and their perceived failures to limit it (Citron, 2014; Langlois and Slane, 2017; Massanari, 2015; Shepherd et al., 2015). To drive positive change on this issue we need to begin to systematically measure their performance in this regard.

The set of human rights against discrimination also deserve special inclusion in evaluations of intermediary governance. The design of technology may itself be discriminatory, and any index of human rights should certainly rank intermediaries on the extent to which their systems are accessible, particularly to people with disabilities. Intermediaries should also be evaluated for their role in enabling or supporting discrimination perpetuated through their systems, whether as a direct consequence of design or a failure to adequately address harmful patterns of user behavior. Now that digital media platforms play such a large role in organizing labor and markets for goods and services, it will become increasingly important to examine how they might enable or prevent discrimination (Edelman and Luca, 2014; Witt et al., 2015; Zuckerman et al., 2015).

The impact of online intermediaries on freedom of association is also important to measure now that social media in particular plays such an important role in coordinating collective action (Castells, 2012; Tufekci and Wilson, 2012). The recent experience of the Arab Spring, for example, highlighted how states may seek to interfere with telecommunications services to disrupt protests, and it may be possible to rank intermediaries on their willingness to resist such pressure unless validly compelled by law (Brown and Korff, 2012; DeNardis, 2012).

There are a number of other rights that are not well covered and should ideally be incorporated into future research. Digital constitutionalism documents commonly reference the right to cultural and linguistic diversity, and this has been a significant, ongoing challenge for internet infrastructure. Recent work has sought to ensure that the technical standards that underpin the internet are more inclusive (Milan and Ten Oever, 2016), and social media platforms too are often critiqued on their design, policies and practices (Panigrahi, 2015). A number of major digital constitutionalism documents also specifically reference the protection of children (UNCRC, 1989), and intermediaries have long faced a great deal of pressure to protect and empower children using their networks. Other rights may be lower priorities for evaluation, although this may change in future. Freedom of religious belief, for example, tends to fall under the umbrella of freedom of expression in the digital context, and may not warrant specific inclusion at this time.

Procedural values

The second component of our proposed index focuses on procedural limits to the exercise of governance power. Generally, these fall into two broad categories within the digital constitutionalism literature: first, concerns about the rule of law, due process, and transparency; and second, governance and participation concerns about multi-stakeholderism and participation in decision making. Rule of law values, as applied to online intermediaries, require evaluating the legitimacy of private governance against the extent to which it is consensual, transparent, equally applied, relatively stable, and fairly enforced (Suzor, 2011, 2012). Due process includes the right to access to justice, the right to be heard, and a meaningful

ability to challenge decisions (Citron and Pasquale, 2014). Transparency incorporates the right of users to know when, why and by whom decisions that affect them are being made. While many of these values are articulated in the major international human rights instruments, many of these documents were written before the digital age. This is where the digital constitutionalism documents are most relevant, in recognising the need for legitimate governance of intermediaries in an era where so much of the public sphere has moved online.

All of these values are relatively well covered by the combination of the RDR index and the ToSHR study. Between them, these studies examine whether there are clear and accessible rules which are applied fairly, whether users are notified of changes, whether there are internal procedures to challenge decisions (including a presumption of innocence), and whether users can access independent adjudication by a court or tribunal. The most immediate limitation of existing studies is that these procedural concerns are at times examined primarily in relation to rules about content (freedom of expression) or privacy, and not always the full range of substantive issues detailed above. There are a number of additional values within the broad definition of the rule of law and due process that could also be included in future work, including measures of stability and consent to the rules, and equality in the application of rules.

Questions of participatory governance are somewhat less well incorporated into work that has been carried out to date. Many digital constitutionalism documents consider multi-stakeholder and participatory governance to be important, but there remains a significant lack of clarity about what a right to participation might look like in the context of private online governance (Stein, 2013; Suzor and Woodford, 2013). RDR has one indicator that examines whether intermediaries are involved in multi-stakeholder initiatives, but does not rank firms further on the extent to which they incorporate input from relevant stakeholders on decisions that impact users. The efforts of intermediaries towards participatory governance practices is worth exploring further and tracking in future work.

Potential Approaches for Studying Governance-in-practice

In this section, we turn to the next major opportunity for future research: studying and evaluating governance-in-practice. We suggest that more research is required to understand how the design, policies, and processes of intermediaries impact human rights in practice, and how users experience these processes. Our review of existing work to date shows that this is a relatively under-explored component of the digital constitutionalism project. So far, the large systematic examinations of governance across multiple intermediaries have focused on legal rules, contractual terms, and policy documents. Meanwhile, many scholars have contributed excellent detailed studies of the practices of particular intermediaries. We suggest that there is an opportunity to begin to systematize this work. This is where an index can be a particularly useful in providing the structure needed to enable comparisons across intermediaries and evaluate changes over time. In this section, we review some of the more prominent efforts to systematically understand the governance of online intermediaries in practice and provide a review of methodological challenges and opportunities to expand this work.

Transparency reporting

The starting point of these efforts has been the growth of transparency reporting initiatives from intermediaries themselves, a welcome practice used by intermediaries to cultivate legitimacy with users and civil society organizations. Transparency reports typically include quantitative information on legal requests received by intermediaries, breaking down this data on a country by country basis and often providing information on what proportion of requests have been complied with. This information sheds light on various issues regarding third party interference, such as government requests for data and copyright takedown requests. Increasingly, standardization of reporting procedures enables comparison among

intermediaries, though there is still a long way to go (Woolery et al., 2016). Access Now has also undertaken important work synthesizing transparency reports and legal complaints across companies (Access Now, 2016). Lumen, a project at the Berkman-Klein Center for Internet and Society, acts as a clearinghouse for legal complaints and requests for removal of online materials, and is particularly robust for information such as DMCA takedown requests received by a small number of participating intermediaries. Researchers have been able to make good use of this data to analyze how the notice and takedown system works in practice, and how it is changing over time (Urban et al., 2016; Urban and Quilter, 2006).

Given the large volumes of information that are now being made available by major telecommunications providers, there is ample room for more academic analysis of governance practices that make use of transparency data. Unfortunately, transparency efforts offer only a partial picture of the legitimacy of companies' practices. In particular, transparency reports often fail to capture the context within which decisions are made. Moreover, legal requests make up only part of the broader landscape of decisions that implicate legitimacy: policies set by intermediaries and design-level choices also have an impact on the legitimacy of intermediary practices, but are not reflected in transparency reports. The next steps in expanding this research are to (a) increase the number and range of intermediaries issuing transparency reports, and (b) encourage intermediaries to provide more transparency about their own terms of service enforcement and extra-legal requests from third parties.

Qualitative analyses of governance

There is now a wealth of qualitative research that examines governance practices of internet intermediaries. These studies provide insight into context and practice that transparency data cannot convey. There are excellent in-depth studies that examine the institutions involved in internet governance, particularly the major multi-stakeholder groups responsible for technical infrastructure (e.g. Mueller, 2002) and the role of nation states in influencing telecommunications technology and processes (e.g. DeNardis, 2009; Mueller, 2010). Researchers have also recently been able to develop rich understandings of governance from the perspectives of the digital laborers tasked with enforcing policies (e.g. Roberts, 2014). Investigative journalists have also played a critical role in examining particular controversies, such as discriminatory content moderation policies, abuse of platforms by bad actors, or the need for algorithmic transparency and accountability (e.g. Buni and Chemal, 2016; Dewey, 2016; Rosen, 2013). This growing body of work is too large to adequately represent here; it suffices to note that this burgeoning field promises to greatly improve our understanding of governance, particularly as it pivots to more comprehensively examine the role of infrastructure and of private actors in internet governance (DeNardis, 2013; DeNardis and Hackl, 2015; Hofmann et al., 2016; Musiani et al., 2016).

In general, however, there are still gaps for rigorous studies of governance-in-practice, particularly studies that take into account governance outcomes of private intermediaries. For the majority of larger private intermediaries, gaining access to examine governance processes is exceedingly difficult. This lack of transparency about outcomes makes it challenging to assess the impact of policies at scale. Research in this area often proceeds in a way that is somewhat ad hoc: at present, we lack general methodologies for evaluating how well intermediaries are performing on human rights criteria in a manner that is comparable over time and across intermediaries.

User-oriented studies can help assess the outcomes of intermediary policies. For example, OnlineCensorship.org utilizes a survey approach to crowdsource reports from users who have had their content removed or accounts shut down on social media platforms in order to gain a fuller view of how content moderation impacts users. This method may also prove useful for evaluating the effectiveness of intermediary due process by documenting and comparing the

experiences of users who have sought redress through the systems intermediaries make available to them.

We suggest that a possibly fruitful avenue for future work would be to crowdsource the collective work of different groups of scholars and civil society advocates. Given the growing number of studies of different controversies, it seems that it may be useful to combine the efforts of many in order to obtain a more systematic view of the legitimacy of governance practices of online intermediaries. There is a significant methodological challenge here that presents an opportunity for much more extensive analysis: how might a metareview of existing qualitative analyses, all with different cases and methods, seek to extract a set of evaluations that can be used to compare governance practices across controversies, across sites, and over time? We hope that our preliminary work in developing a shared analytical framework can help as a first step to leveraging these existing studies. At any rate, a shared framework can at least help to form the basis of future collaboration.

Quantitative approaches and network measurement

An additional body of research has begun to use network measurement to assess levels of censorship, prodding and probing intermediaries in order to measure outcomes. The major challenge in studying governance by private intermediaries and the technical and social systems they create is that so often, decisions are opaque, and gaining access to observe internal processes is extremely difficult. Some of the most promising avenues for understanding internet governance in practice require the development of new methods to assess the inputs and outputs of these ‘black box’ systems (Pasquale, 2015; Rosenblat et al., 2014; Sandvig et al., 2014).

There are interesting initial examples of this type of research in recent years. Scholars examining Sina Weibo in China, for example, have collected large numbers of microblog posts shortly after they are posted, and tracked them over time in order to see how certain forms of political speech are removed due to government pressure (Zeng et al., 2016). Diakopoulos (2014, 2015) provides a good overview of how computational journalists are using reverse engineering techniques to make algorithms more accountable, focusing on many of the values outlined above. Perel and Elkin-Koren (2016) have outlined a methodology for systematically experimenting with the content moderation processes of online intermediaries by uploading prohibited and permitted content and observing the results. Much of this work is still new, and often requires interdisciplinary collaborations, but it is part of a range of exciting opportunities to collect the digital traces of governance decisions in order to systematically measure the performance of these systems against human rights principles.

Conclusion

We have presented a high level sketch of the landscape of studies that examine the legitimacy of governance by online intermediaries. We situate these studies within an explicit framework of digital constitutionalism, in order to identify the potential to extend existing work and opportunities to address gaps in the research. Our analysis suggests several important avenues for future research and collaboration to develop the evidence base that will be required to progress the project of digital constitutionalism. We note first that formal analyses of governance documents of online intermediaries have well developed methodologies that can be extended to examine the policies of a larger number of internet and telecommunications firms. We suggest that these studies can be enhanced by turning to more explicitly address human rights other than freedom of expression and privacy, including personal security and dignity, freedom from discrimination, and freedom of association.

The major research gap that we have identified in this work is the lack of systematic analysis of governance-in-practice in a form that enables comparison across contexts and over time. While there is a great deal of interesting work that has started on this challenge, there are many more opportunities to develop new projects and methods to better understand how the decisions of private intermediaries impact the human rights of their users (or third parties). We call on intermediaries themselves to follow best practices in releasing transparency data to enable researchers to carry on this work. Since comprehensive transparency data is still difficult to find, however, we suggest that new experimental methods and reverse engineering techniques are likely needed to better illuminate the governance practices of intermediaries. Because the challenge facing scholars here is so great, and the social issues so pressing, we also call for more collaboration on these research questions. In developing the outline of an index of intermediary governance that addresses the core human rights challenges of digital constitutionalism, we hope to help guide efforts to develop interoperable research methodologies in the near future. Finally, we encourage the many diverse scholars working on these issues to adopt the open data practices of groups like Ranking Digital Rights, to enable, as far as possible, the development of new analyses that seek to understand governance in rapidly changing environments.

Funding acknowledgements

Suzor is the recipient of an Australian Research Council DECRA Fellowship (project number DE160101542).

References

- Access Now (2016) *Transparency Reporting Index*. Available from: <https://www.accessnow.org/transparency-reporting-index/> (accessed 28 February 2017).
- Black J (2001) Decentring Regulation: Understanding the Role of Regulation and Self-Regulation in a 'Post-Regulatory' World. *Current Legal Problems* 54(1): 103–146.
- Brown I and Korff D (2012) Digital Freedoms in International Law. *Global Network Initiative*.
- Buni C and Chemal S (2016) The secret rules of the internet: The murky history of moderation, and how it's shaping the future of free speech. *The Verge*. Available from: <http://www.theverge.com/2016/4/13/11387934/internet-moderator-history-youtube-facebook-reddit-censorship-free-speech> (accessed 11 May 2016).
- Burgess J (2015) From 'broadcast yourself' to 'follow your interests': making over social media. *International Journal of Cultural Studies* 18(3): 281–285.
- Burris S, Kempa M and Shearing C (2008) Changes in Governance: A Cross- Disciplinary Review of Current Scholarship. *Akron Law Review* 41: 1–66.
- Castells M (2012) *Networks of outrage and hope: social movements in the internet age*. Cambridge: Polity Press.
- Citron DK (2014) *Hate Crimes in Cyberspace*. Cambridge: Harvard University Press.
- Citron DK and Pasquale FA (2014) The scored society: due process for automated predictions. *Washington Law Review* 89. Available from: http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2376209 (accessed 30 June 2015).
-

-
- Cooley A and Snyder J (eds) (2015) *Ranking the World*. Cambridge Core. Available from: /core/books/ranking-the-world/CC428C9015C2D6656218795F649D9BBC (accessed 19 February 2017).
- Council of Europe (2011) Declaration on Principles for Internet Policy-Making.
- DeNardis L (2009) *Protocol Politics: The Globalization of Internet Governance*. MIT Press.
- DeNardis L (2012) Hidden levers of internet control. *Information, Communication & Society* 15(5).
- DeNardis L (2013) The Emerging Field of Internet Governance. In: Dutton WH (ed.), *The Oxford Handbook of Internet Studies*.
- DeNardis L (2014) *The global war for internet governance*. Yale University Press.
- DeNardis L and Hackl A (2015) Internet governance by social media platforms. *Telecommunications Policy*, Special Issue on the Governance of Social Media 39(9): 761–770.
- Dewey C (2016) Facebook is tracking ‘trends,’ so we’re tracking Facebook. *The Washington Post*, 31st August. Available from: <https://www.washingtonpost.com/news/the-intersect/wp/2016/08/31/facebook-is-tracking-trends-so-were-tracking-facebook/> (accessed 17 February 2017).
- Diakopoulos N (2014) Algorithmic Accountability: On the Investigation of Black Boxes. *Tow Centre for Digital Journalism*. Available from: <http://towcenter.org/research/algorithmic-accountability-on-the-investigation-of-black-boxes-2/> (accessed 20 April 2016).
- Diakopoulos N (2015) Algorithmic Accountability. Journalistic investigation of computational power structures. *Digital Journalism* 3(3): 398–415.
- Edelman BG and Luca M (2014) Digital Discrimination: The Case of Airbnb.com. *Harvard Business School, Harvard Business School Working Papers: 14-054, 2014, 21 pp.*: 21.
- European Commission (2013) *ICT Sector Guide on Implementing the UN Guiding Principles on Business and Human Rights*. Available from: https://ec.europa.eu/anti-trafficking/publications/european-commission-sector-guides-implementing-un-guiding-principles-business-and-hum-0_en (accessed 14 February 2017).
- Gill L, Redeker D and Gasser U (2015) Towards Digital Constitutionalism? Mapping Attempts to Craft an Internet Bill of Rights. *Berkman Centre for Law and Society Research Publication No. 2015-15*. Available from: <http://papers.ssrn.com/abstract=2687120> (accessed 1 February 2016).
- Global Network Initiative (2012) Principles on Freedom of Expression and Privacy. Available from: <https://www.globalnetworkinitiative.org/principles/index.php> (accessed 2 June 2015).
- Goldsmith J and Wu T (2006) *Who controls the Internet?: illusions of a borderless world*. Oxford University Press.
- Green M (2001) What we mean when we talk about indicators: Current approaches to human rights measurement. *Human Rights Quarterly* 23(4).
-

-
- Hofmann J, Katzenbach C and Gollatz K (2016) Between coordination and regulation: Finding the governance in Internet governance. *New Media & Society*: 1461444816639975.
- Hyden G and Samuel J (eds) (2011) *Making the state responsive: Experience with democratic governance assessments*. New York: United Nations Development Program. Available from: http://www.undp.org/content/dam/undp/documents/partners/civil_society/additional_documents/Africa%20Forum%20on%20Civil%20Society%20and%20Governance%20Assessments/Making%20the%20state%20responsive.pdf (accessed 23 February 2017).
- IPRC (2014) *Charter of Human Rights and Principles for the Internet*. Internet Rights and Principles Coalition. Available from: <http://internetrightsandprinciples.org/wpcharter/> (accessed 18 February 2015).
- Kaye D (2016) *Report of the Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression*. Geneva, Switzerland: United Nations, Human Rights Council. Available from: <http://www.ohchr.org/EN/Issues/FreedomOpinion/Pages/PrivateSectorintheDigitalAge.aspx> (accessed 16 February 2017).
- Langford M and Fukuda-Parr S (2012) The turn to metrics. *Nordic Journal of Human Rights* 30(3): 222–238.
- Langlois G and Slane A (2017) Economies of reputation: the case of revenge porn. *Communication and Critical/Cultural Studies*: 120–138.
- LaRue F, et al. (2011) Freedom of Expression Rapporteurs Issue Joint Declaration Concerning the Internet. *UN, OAS, IACHR, OSCE*. Available from: <http://www.oas.org/en/iachr/expression/showarticle.asp?artID=848&IID=1> (accessed 27 June 2017).
- MacKinnon R, Maréchal N and Kumar P (2016) *Corporate Accountability for a Free and Open Internet*. Global Commission on Internet Governance Paper Series. Available from: <https://www.cigionline.org/publications/corporate-accountability-free-and-open-internet> (accessed 14 February 2017).
- Manila Principles on Intermediary Liability (2015). Available from: <https://www.manilaprinciples.org/> (accessed 27 November 2015).
- Maréchal N (2015) Ranking Digital Rights: Human Rights, the Internet and the Fifth Estate. *International Journal of Communication* 9(10): 3440–3449.
- Massanari A (2015) #Gamergate and The Fappening: How Reddit’s algorithm, governance, and culture support toxic technocultures. *New Media & Society*: 1461444815608807.
- Milan S and Ten Oever N (2016) Coding and Encoding Rights in Internet Infrastructure. Sociotechnical Imaginaries and Grassroots Ordering in Internet Governance. Available from: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2875816 (accessed 28 February 2017).
- Mueller M (2002) Ruling the root. *Cambridge: MIT Press*: 162.
- Mueller M (2010) *Networks and states the global politics of Internet governance*. Information revolution and global politics, Cambridge: MIT Press.
-

-
- Musiani F, Cogburn DL, DeNardis L, et al. (eds) (2016) *The turn to infrastructure in Internet governance*. Information technology and global governance, New York: Palgrave Macmillan.
- OECD (2010) *The Economic and Social Role of Internet Intermediaries*. Organisation for Economic Co-operation and Development. Available from: <https://www.oecd.org/internet/ieconomy/44949023.pdf> (accessed 8 July 2016).
- Panigrahi S (2015) Rising Voices: Indigenous Language Digital Activism. In: Shah N, Sneha PP, and Chattapadhyay S (eds), *Digital activism in Asia reader*, Lüneburg: meson press, pp. 155–158. Available from: <http://nbn-resolving.de/http://dx.doi.org/10.14619/013> (accessed 27 February 2017).
- Pasquale F (2015) *The Black Box Society*. Cambridge, Mass.: Harvard University Press.
- Perel M and Elkin-Koren N (2016) Black Box Tinkering: Beyond Transparency in Algorithmic Enforcement. *Florida Law Review (Forthcoming)*. Available from: <http://papers.ssrn.com/abstract=2741513> (accessed 1 September 2016).
- Ranking Digital Rights (2015) *2015 Corporate Accountability Index*. Ranking Digital Rights and Sustainalytics. Available from: <https://rankingdigitalrights.org/> (accessed 11 May 2016).
- Roberts ST (2014) Behind the screen: the hidden digital labor of commercial content moderation. University of Illinois at Urbana-Champaign. Available from: <https://www.ideals.illinois.edu/handle/2142/50401> (accessed 27 February 2017).
- Rosen J (2013) The Delete Squad. *New Republic*. Available from: <http://www.newrepublic.com/article/113045/free-speech-internet-silicon-valley-making-rules> (accessed 9 September 2014).
- Rosenblat A, Kneese T and boyd danah (2014) Algorithmic Accountability. New York. Available from: http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2535540 (accessed 14 July 2016).
- Ruggie J (2008) Protect, Respect and Remedy: A Framework for Business and Human Rights. *Innovations: Technology, Governance, Globalization* 3(2): 189–212.
- Sandvig C, Hamilton K, Karahalios K, et al. (2014) An Algorithm Audit. In: Gangadharan SP, Eubanks V, and Barocas S (eds), *Data and Discrimination: Selected Essays*, Open Technology Institute and New America, pp. 6–10. Available from: <https://www.newamerica.org/oti/policy-papers/data-and-discrimination/>.
- Shepherd T, Harvey A, Jordan T, et al. (2015) Histories of Hating. *Social Media and Society* 1(2): 2056305115603997.
- Stein L (2013) Policy and Participation on Social Media: The Cases of YouTube, Facebook, and Wikipedia. *Communication, Culture & Critique* 6(3): 353–371.
- Suzor N (2011) The role of the rule of law in virtual communities. *Berkeley Technology Law Journal* 25: 1819–1886.
- Suzor N (2012) Order Supported by Law: The Enforcement of Norms in Virtual Communities. *Mercer Law Review* 63: 523.
-

-
- Suzor N and Woodford D (2013) Evaluating Consent and Legitimacy Amongst Shifting Community Norms: an EVE Online Case Study. *Journal For Virtual Worlds Research* 6(3).
- Thomas CA (2014) The Uses and Abuses of Legitimacy in International Law. *Oxford Journal of Legal Studies* 34(4): 729–758.
- Tufekci Z and Wilson C (2012) Social Media and the Decision to Participate in Political Protest: Observations From Tahrir Square. *Journal of Communication* 62(2): 363–379.
- UNESCO (2014) *Fostering Freedom Online: the Role of Internet Intermediaries*. UNESCO. Available from: <http://unesdoc.unesco.org/images/0023/002311/231162e.pdf>.
- UN Internet Governance Forum (2014) Charter (2.0) of Human Rights and Principles for the Internet.
- United Nations (1965) International Convention on the Elimination of All Forms of Racial Discrimination.
- United Nations (1966) International Covenant on Civil and Political Rights.
- United Nations (1976) International Covenant on Economic, Social, and Cultural Rights.
- United Nations (1979) Convention on the Elimination of All Forms of Discrimination Against Women.
- United Nations (1989) Convention on the Rights of the Child.
- United Nations (2011) Guiding Principles on Business and Human Rights. Available from: http://www.ohchr.org/documents/publications/GuidingprinciplesBusinesshr_en.pdf.
- United Nations (1948) Universal Declaration of Human Rights.
- United Nations (2006) Convention on the Rights of Persons with Disabilities.
- Urban JM and Quilter L (2006) Efficient Process or ‘Chilling Effects’? Takedown Notices Under Section 512 of the Digital Millennium Copyright Act. *Santa Clara Computer and High Technology Law Journal* 22: 621.
- Urban JM, Karaganis J and Schofield BL (2016) *Notice and Takedown in Everyday Practice*. Available from: <http://papers.ssrn.com/abstract=2755628> (accessed 1 April 2016).
- Venturini J, Louzada L, Maciel M, et al. (2016) *Terms of Service and Human Rights: an Analysis of Online Platform Contracts*. Rio de Janeiro: Editora Revan. Available from: http://internet-governance.fgv.br/sites/internet-governance.fgv.br/files/publicacoes/terms_of_services_06_12_2016.pdf.
- Witt A, Suzor N and Wikström P (2015) Regulating ride-sharing in the peer economy. *Communication Research and Practice* 1(2).
- Woodlock D (2016) The Abuse of Technology in Domestic Violence and Stalking. *Violence Against Women*: 19.
- Woolery L, Budish RH and Bankston K (2016) *The Transparency Reporting Toolkit: Best Practices for Reporting on US Government Requests for User Information*. New America
-

Foundation and The Berkman Center for Internet & Society. Available from:
<https://dash.harvard.edu/handle/1/28552578> (accessed 23 February 2017).

Zeng J, Chan C and Fu K (2016) What is social media platforms' role in constructing 'truth' around crisis events? A case study of Weibo's rumour management strategies after the 2015 Tianjin blasts. In: Oxford Internet Institute, University of Oxford. Available from:
http://ipp.oii.ox.ac.uk/sites/ipp/files/documents/TianjinBlastFullPaper_14Aug.pdf (accessed 16 December 2016).

Zukin S, Lindeman S and Hurson L (2015) The omnivore's neighborhood? Online restaurant reviews, race, and gentrification. *Journal of Consumer Culture*. Available from:
<http://joc.sagepub.com/cgi/doi/10.1177/1469540515611203> (accessed 9 December 2015).