

VU Research Portal

How Online Dispute Resolution can be used to negotiate social media bans

Lodder, Arno R.; Zeleznikow, J.

published in

ODR conference PORTO
2021

document version

Peer reviewed version

document license

Unspecified

[Link to publication in VU Research Portal](#)

citation for published version (APA)

Lodder, A. R., & Zeleznikow, J. (2021). How Online Dispute Resolution can be used to negotiate social media bans. In *ODR conference PORTO* Online Dispute Resolution (ODR).

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal ?

Take down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

E-mail address:

vuresearchportal.ub@vu.nl

How Online Dispute Resolution can be used to negotiate social media bans

Arno R. Lodder, Amsterdam Law and Technology Institute, alti.amsterdam, Vrije Universiteit Amsterdam, The Netherlands

John Zeleznikow, La Trobe University Law School, Melbourne, Australia

Abstract

For many people and companies, being banned from a social media platform can have a major impact, in terms of reputation, relationships and financially. So, is it possible to use Online Dispute Resolution to help parties negotiate disputes relating to social media bans? By investigating issues of free speech and a model for intelligent online dispute resolution, we propose a platform for helping resolve social media bans.

1 Introduction

For many people and companies, social media is an important communication channel. Being banned from a social media platform can have a major impact, in terms of reputation, relationships and even financially.

So why are people occasionally banned from using social media platforms? And can we use Online Dispute Resolution to challenge a ban from a platform. Or can we allow users who might be concerned about being banned from Social Media, to negotiate and regulate how they use social media?

In the United States, the First Amendment to the Constitution protects citizens from government censorship of speech. However, social media platforms are not under the direct control of governments. They are businesses that are run by companies. Their terms of reference do not oblige them to service everybody. They can decide where to set the boundaries on online speech¹.

Valerie Brannon, a legislative attorney at the US Congressional Research Service has said:

“As the Supreme Court has recognized, social media sites like Facebook and Twitter have become important venues for users to exercise free speech rights protected under the First Amendment. Commentators and legislators, however, have questioned whether these social media platforms are living up to their reputation as digital public forums. Some have expressed concern that these sites are not doing enough to counter violent or false speech. At the same time, many argue that the platforms are unfairly banning and restricting access to potentially valuable speech. Currently, federal law does not offer much recourse for social media users who seek to challenge a social media provider’s decision about whether and how to present a user’s content. Lawsuits predicated on these sites’

¹ <https://news.bloomberglaw.com/us-law-week/does-free-speech-have-limits-on-social-media-podcast?context=article-related> Bloomberg Law last viewed 22 September 2021

decisions to host or remove content have been largely unsuccessful, facing at least two significant barriers under existing federal law. Individuals have sometimes alleged that these companies violated their free speech rights by discriminating against users' content, courts have held that the First Amendment, which provides protection against state action, is not implicated by the actions of these private companies.”²

One example on how speech has been limited on social media, was the banning of then US president, Donald Trump, from Twitter and Facebook following the insurrection at the US Capitol Building on January 6 2021. The statement from the Twitter management on 8 January 2021, claimed that Trump was permanently banned. No statement was made about how long the ban would be in force, how Trump could expiate this decision and how he could appeal the decision.

Whilst the authors do not support Trump's criticism of the conduct of the 2020 US presidential election, nor do we disagree with his ban from Twitter and Facebook, we do believe that reasons need to be given for such action. And, there need to be avenues to appeal results and bans should be for a fixed time.

At a previous meeting in Portugal, the International conference e-Society 2013 in Lisbon, Irmgard Wetzstein and Peter Leitner discussed the “combination of social media and online dispute resolution”, and that this should be considered an “innovative way of dealing with conflicts and disputes”.³ In this paper we discuss issues of how social media should be regulated with regards to free speech and suggest an Online Dispute Resolution framework to help negotiate restriction upon online speech in social media. We believe that Online Dispute Resolution (ODR) can play a role in resolving such conflicts.

Katsh & Rabinovich-Einy start their chapter on ‘The Challenge of Social and Anti-Social Media’⁴ with a fictive example of a user being banned for using her everyday name on Facebook that is different from her official name. However, their chapter does focus on disputes between users. The main focus in our paper is on dispute resolution in the case of social media bans, so a dispute between the social media platform and the user.

First, it is important that social media platforms specify the reasons for such bans and indicate under what category a particular banned user falls. Second, the user should be allowed to provide an explanation for the alleged cause for the ban and have the opportunity to argue why the ban is unjustified. The input of the users who object to the ban can be automatically analysed and grouped into various categories. Third, the platform should make a decision about the proposed ban and leave an option for appeal against its decision.

The process should take place online, and within a reasonable period of time. Given the importance of social media accounts, the disputes should be dealt with not within in weeks

² Brannon, Valerie C. "Free speech and the regulation of social media content." *Congressional Research Service* 27 (2019).

³ Irmgard Wetzstein and Peter Leitner, Conflict and dispute resolution in the age of social media, P. Kommers and Pedro Isaías (eds.), *PROCEEDINGS OF THE INTERNATIONAL CONFERENCE e-Society*, IADIS Press, p. 494-498.

⁴ Ethan Katsh and Orna Rabinovich-Einy, *Digital Justice: Technology and the Internet of Disputes*, Oxford University Press 2017.

but within days and probably even hours. In our paper we provide a preliminary description of how such an ODR process can be developed.

2 Social Media and Free Speech

One hundred years ago Herbert Goodrich wrote that many thoughtful men and women in the USA, witnessing the suppression of speech, by means both judicial and extra-judicial, in the period following the First World War,⁵ have reluctantly concluded that their hard won right of freedom of speech has been lost, swept away in the flood tide of war enthusiasm.

They point to the example of the candidate for the presidency in 1920, Eugene Debs, who at the time was confined in a federal prison for words he uttered during the First World War. They call attention to the fact that the fate of Mr. Debs was no worse than that of scores of other persons, members of his and other minority groups, who had gone to jail since April 1917, for giving utterance to unpopular opinions. Subsequently there was a widespread wave of "anti-disturbance" legislation among USA state legislatures during and immediately after the First World War.⁶

The First Amendment to the United States Constitution protects the right to freedom of religion and freedom of expression from government interference. It prohibits any laws that establish a national religion, impede the free exercise of religion, abridge the freedom of speech, infringe upon the freedom of the press, interfere with the right to peaceably assemble, or prohibit citizens from petitioning for a governmental redress of grievances. It was adopted into the Bill of Rights in 1791. The US Supreme Court interprets the extent of the protection afforded to these rights.

Genevieve Lakier argues that the US First amendment only protects speech from government censorship. This includes not only lawmakers and other elected officials, but also public schools, universities, courts, and police officers. However, it does not include private citizens and organisations or businesses.⁷ The US First Amendment has been interpreted by the US Supreme Court as applying to the entire federal government even though it is only expressly applicable to the US Congress. Furthermore, the Court has interpreted the Due Process Clause of the Fourteenth Amendment as protecting the rights in the First Amendment from interference by state governments⁸.

Wu argues that it is no longer speech itself that is scarce, but the attention of listeners. Emerging threats to public discourse take advantage of this change.⁹ Tufekci claims "censorship during the internet era does not operate under the same logic [as] it did during the heyday of print or even broadcast television."¹⁰ Instead of targeting speakers directly, it targets listeners or undermines speakers indirectly. More precisely, the emergent techniques of speech control depend on new punishments, like the unleashing of "troll armies" to abuse critics, the fabrication of news, and "flooding" tactics that distort or drown out other speech

⁵ The First World War and its immediate aftermath.

⁶ Goodrich, Herbert F. "Does the Constitution Protect Free Speech?." *Michigan Law Review* 19.5 (1921): 487-501

⁷ Lakier, Genevieve. "The Non-First Amendment Law of Freedom of Speech." *Harv. L. Rev.* 134 (2020): 2299.

⁸ https://www.law.cornell.edu/wex/first_amendment last viewed September 14 2021

⁹ Tim Wu, *Is the First Amendment Obsolete?*, 117 MICH. L. REV. 547 (2018)

¹⁰ Zynep Tufekci, *Twitter and Tear Gas: The Power and Fragility of Networked Protest* 226 (2017)

through the payment of fake commentators or the deployment of propaganda robots. Powerful actors, both public and private, have adopted speech itself as a weapon for controlling speech, yielding challenges for which the First Amendment is unprepared.

On the other hand, Everett argues that social media are public fora regulated by quasigovernmental actors seeking to filter certain speech.¹¹ He argues for advocates for congressional action as the preferred avenue of conferring free speech rights on end-users of social media because of the difficulty of judicial review in state action and due process claims.

In 1951, the then Liberal Party government of Australia introduced a constitutional amendment to ban the Communist Party of Australia.¹² The referendum was? unsuccessful. Seventy years later, members of the House of Representatives from that same Liberal party are arguing free speech should allow them to promote scientifically proven facts about the dangers of Astra Zeneca and Pfizer vaccines against the COVID vaccine and the efficacy of Ivermectin.¹³ In early twentieth century USA, appeals to the right of free speech primarily came from those on the left wing of politics. Today it comes from the right wing: whether it be former US president Donald Trump's spurious claims about vote fraud in the 2020 US presidential elections¹⁴, or the anti-vaccination claims of COVID19 deniers.

Michael Luca illustrates that former US president Donald J Trump's accounts have been banned on Twitter, Facebook, and a host of other platforms¹⁵. Every last one of @realdonaldtrump's 47,000 tweets vanished from the site in an instant, from the birther lies¹⁶ and election conspiracy theories to the 2016 taco bowl tweet¹⁷. In an explanatory blog post, the company cited the attack on the US Capitol on January 6 2021 and "the risk of further incitement of violence" that might occur by permitting further Trump tweets. Whilst this action has been widely applauded, it has also raised concerns that the companies had gone too far in exercising their power to shape what users see.

Katherine Gelber argues that there was no free speech defence to the speech that Donald Trump gave in Washington, DC on 6 January¹⁸. This was because of the key issue of the incitement of violence. No coherent free speech argument has ever suggested that its protections extend to the incitement of violence. There is no free speech protection for speech that is imminently and causally connected to the incitement of a mob to violent lawlessness. This is because such speech is evidently, immediately, and virulently harmful.

She justifies her claim by indicating that five lives were lost at Capitol Hill that day and there have been subsequent reports that improvised explosive devices were left in the building, and

¹¹ Colby M. Everett, Free Speech on Privately-Owned Fora: A Discussion on Speech Freedoms and Policy for Social Media, 28 KAN. J.L. & PUB. POL'y 113 (2018).

¹² Macintyre, S 1998, *The Reds*, Allen & Unwin, St Leonards, Sydney

¹³ <https://www.abc.net.au/news/2021-09-14/george-christensen-ivermectin-feeds-vaccine-hesitancy-says-gp/100457530>

¹⁴ Note that there have been no claims of fraud for concurrent US elections for the Senate and the House of Representatives. These elections were conducted using the same ballot papers.

¹⁵ [Social Media Bans Are Really, Actually, Shockingly Common | WIRED](#) last viewed 13 September 2021

¹⁶ Arguing former US president Barack Obama was not born in USA

¹⁷ Francia, Peter L. "Free media and Twitter in the 2016 presidential election: The unconventional campaign of Donald Trump." *Social Science Computer Review* 36.4 (2018): 440-455.

¹⁸ <https://www.abc.net.au/religion/katharine-gelber-social-media-and-free-speech/13093868> last viewed 22 September 2021

suggestions that some of the insurrectionists had planned (including by bringing along equipment) to take hostages and to execute them.

Currently there has been little academic research, or indeed litigation about social media and free speech.

3 Online Dispute Resolution: six facilities to be provided

Lodder and Zeleznikow indicate, ‘whilst there is no generally accepted definition of Online Dispute Resolution (ODR), we can think of it as using the Internet to perform Alternative Dispute Resolution (ADR).’ The one factor common to all descriptions of Online Dispute Resolution is the existence of a fourth party – namely the technology¹⁹.

Brown claims that human-centered design is a methodology that is primarily concerned with the users’ experience. It uses these experiences to develop new solutions that are both experimental and iterative²⁰. To build user-centric ODR systems, Zeleznikow²¹ built upon the Thiessen-Zeleznikow²² and Lodder-Zeleznikow²³ classification schemes for designing Intelligent ODR Systems.

Zeleznikow²⁴ and Schmitz and Zeleznikow²⁵ argue that from their research on observing how ODR systems should be used, remembering that the current COVID19 focus is upon videoconferencing, with some systems such as the British Columbia Civil Resolution²⁶ Tribunal and systems developed by MODRIA²⁷ also proving advisory tools, they believe that a truly helpful ODR system should provide the following facilities:

1. **Case management:** the ODR system should allow users to enter information, ask users for appropriate data and provide users with templates to initiate the dispute.
2. **Triaging:** the ODR system should indicate which cases require immediate action and which cases can wait. The system must also advise upon in which forum the case should be heard. Triaging systems are important for initiating and expediting action in high-risk cases, leading to reduced risk to the applicants and community.
3. **Advisory tools:** the ODR system should provide processes for reality testing. Relevant tools could include, articles, BATNA advisory systems (which would inform

¹⁹ Lodder, A.R. and Zeleznikow, J. 2010. *Enhanced Dispute Resolution through the use of Information Technology*. Cambridge University Press, Cambridge, United Kingdom.

²⁰ Brown, T., 2008. Design thinking. *Harvard business review*, 86(6), p.84.

²¹ Zeleznikow, J. 2021. Using Artificial Intelligence to provide Intelligent Dispute Resolution Support. *Group Decis Negot* (2021). <https://doi.org/10.1007/s10726-021-09734-1>

²² Thiessen, E. and Zeleznikow, J., 2004, July. Technical aspects of online dispute resolution challenges and opportunities. In *Proceedings of the Third Annual Forum on Online Dispute Resolution, Melbourne, Australia* (pp. 5-6).

²³ Lodder, A.R. and Zeleznikow, J., 2005. Developing an online dispute resolution environment: Dialogue tools and negotiation support systems in a three-step model. *Harv. Negot. L. Rev.*, 10, p.287.

²⁴ Ibid

²⁵ Schmitz, A. and Zeleznikow, J. 2022. Intelligent Legal Tech to Empower Self-Represented Litigants. To appear in *Columbia Science and Technology Law Review*

²⁶ Salter, S. and Thompson, D., 2016. Public-Centered Civil Justice Redesign: A Case Study of the British Columbia Civil Resolution Tribunal. *McGill J. Disp. Resol.*, 3, p.113

²⁷ See <https://www.tylertech.com/products/Modria> last viewed 17 March 2021

litigants of the likely outcome of the dispute if it were to be decided by a decision-maker such as an arbitrator or judge), books providing useful parenting advice, calculators (such as those to advise upon tax and child support obligations), copies of legislation, reports of cases, and videos of desirable and undesirable behaviour.

4. **Communication tools:** the ODR system must provide for communication tools to support conciliation, facilitation, mediation and negotiation. Where appropriate, the tools could provide shuttle mediation. For many ODR providers, the provision of communication tools is their major or indeed only goal.
5. **Decision Support Tools** – if after substantial communications, the disputants still cannot resolve their conflict, as Lodder and Zeleznikow²⁸ suggest, computer programs that utilise artificial intelligence and/or game theory can be used to facilitate trade-offs.
6. **Drafting software or Agreement Technologies:** once the parties to a dispute reach an in-principle settlement, it is important to provide computer software that assists in drafting acceptable agreements.

We believe that much of this six-step framework can be used to help handle disputes over social media bans. There is no need for triaging for disputes about bans, or for drafting software or agreement technologies. The former because in case of a ban immediate action is always required. The latter because the agreement is straightforward. The ban is lifted, or not. In case the ban is lifted this may be conditional, but this condition will be connected to the reason for the ban and pre-formatted fields can be used.

4 Case management

The social media platform advocating the ban should give reasons why it took such action and provide timelines plus opportunities for redress. The effected party can then file their defence. The case management tool can provide updates as appropriate.

A social media ban not necessarily leads to a dispute between the user and the platform. In some cases the user knows and understands why he is being banned. Still, even then it is important the platform communicates at least:

- a) the reason for the ban;
- b) the duration of the ban;
- c) how the ban can be challenged.

At this moment sometimes platforms ban without providing any information. That is not acceptable, and unnecessary. The decision is not taken in vain. The platform will always have a reason, and this reason should be communicated. We are aware of the fact that since many years platforms use in their terms of reference the clause that you can be banned “for any reason or no reason at all”. No matter whether such a clause is legally valid, it should be common practice for platforms always to provide a reason for a ban.

In addition to that, the duration should be clear for the user. There may be cases so serious a permanent ban makes sense, but this should be used only in exceptional cases. For all other situations the number of days, or hours should be communicated. Although for most users a

²⁸ Lodder, A.R. and Zeleznikow, J., 2005. Developing an online dispute resolution environment: Dialogue tools and negotiation support systems in a three-step model. *Harv. Negot. L. Rev.*, 10, p.287.

ban for 24 hours is not worth the effort of objecting, for sure for some commercial accounts every hour, and maybe even minute being online counts.

Then, the platform should indicate how the ban can be challenged. An option that should be available is that the account is put back online again immediately after a challenge.

Obviously, not for all cases this should be an option, but it could depend on e.g. the seriousness of the reason for the ban.

5 Advisory tools

The provision of reality testing is a vital task for resolving the dispute, especially given that such disputes are relatively recent. The platform should provide information about previous bans. Data analytics can help in discovering patterns. This is not only valuable information for users, but the platform can also benefit if users have an idea whether fighting the ban is promising. And the information gathered after analytics can also be helpful in categorizing cases by the platform and how to deal with the different categories.

There are some court cases about bans but it is doubtful whether BATNA advisory systems are useful. First, it is not easy to communicate outcomes of cases from different jurisdictions in a way understandable to all users. For one, because court cases are often drafted in the national language that in most countries is not English. Second, there will not be a large number of court cases so it is difficult to define BATNAs. Third, because platforms are private actors and freedom of speech is still largely restricted to governmental actors, judges will be hesitant in deciding in favour of the banned user.

6 Communication tools

No doubt, easy communication is central to resolving disputes no matter what type of method is used, e.g. conciliation, facilitation, mediation and negotiation. The COVID19 pandemic left no alternative, even to the often old fashioned courts, than to rely on online communication. When being banned from a social media platform it does not make sense to fight the ban other than with online communication tools. First, because speed in communication is needed and that is what online tools provide. Second, the origin of the dispute is online, so to continue online the moment the platform decides on a ban is a natural continuation.

The first level of communication should be with a limited numbers of options to choose from. A list of different categories of challenges. This helps in streamlining the dispute. Second, in the explanatory part open text should be allowed and the uploading of supporting evidence.

7 Decision Support Tools

Computer programs that utilise artificial intelligence and/or game theory can be used to facilitate trade-offs. Such tools assist to engage in tradeoffs, e.g. to change intemperate

language, not mention certain themes or accept a ban for a certain period. Useful examples include Family Winner²⁹ and Smartsettle.³⁰

However, a ban from a social media platform is at its core a single-issue dispute. Someone is banned from the platform, and the dispute is about lifting the ban. There are not that many trade-offs. In the communication phase eventual errors must become clear. One example is an Instagram account being banned because posts were coming from a wide variety of IP addresses. Under these circumstances the platform assumes that the account is run by bots. The users of the Instagram account were an international company with people from different countries having access and posting content. It should not be that difficult to make this clear during the communication phase. Note that this is a real case, and that Instagram did not provide why they banned and did not give the opportunity for explanation.

In case during the communication phase there is agreement about the reason, or there is agreement about the reason but not about the duration of the ban, decision support tools could be of help in presenting a range of possible solutions. For instance, to give conditional access to the user, the condition being that particular behaviour is not repeated.

8 Conclusion

In this paper we have introduced an ODR paradigm to help Social Media users and the relative platform providers to negotiate disputes about the appropriate use of platforms. Based on a need to provide users of social media platforms with the ability to engage in free speech which respects civil rights and criminal law, the system utilises the six-step model of Zeleznikow³¹. The focus has been on the opportunity ODR can offer to users of social media after being banned from the platform. The ODR paradigm provides support for case management, advisory tools and decision support tools as well as traditional communication tools.

²⁹ Bellucci, Emilia and John Zeleznikow: 2006, Developing Negotiation Decision Support Systems that support mediators: a case study of the Family_Winner system, *Journal of Artificial Intelligence and Law*, 13(2): 233-271.

³⁰ Thiessen, Ernest, and Peter Holt. "Resolving Insurance Claims with Smartsettle One." *IJODR* 4 (2017): 45.

³¹ Zeleznikow, J. 2021. Using Artificial Intelligence to provide Intelligent Dispute Resolution Support. *Group Decis Negot* (2021). <https://doi.org/10.1007/s10726-021-09734-1>