Information Law and Policy Lab

Taming the “Free”: Content moderation in the Fediverse and the Role of the DSA

A practical guide for server administrators in the Fediverse.

Information Law and Policy Lab

Amsterdam, April 24, 2024

This report is written by Emese van Rijnswou and Charlotte Verboom under the supervision of Dr. João Pedro Quintais & Ot van Daalen of the Glushko & Samuelson Information Law and Policy Lab (ILP Lab) of the Institute for Information Law (IViR) of the University of Amsterdam. The ILP Lab is a student-run, IViR-led institution which develops and promotes research-based policy solutions that protect fundamental rights and freedoms in the field of European information law.

The report has been written in partnership with the DSA Observatory and European Digital Rights Initiative (EDRi). It reflects the recommendations and conclusions of the authors of the ILP Lab. As part of this research, the authors conducted interviews with Ilaria Buri, Dr. Ronan Fahy, Dr. Paddy Leerssen and Jan Penfrat. The authors would like to thank them for their time and valuable feedback.

Executive summary

The current legal framework for content moderation in the Digital Services Act (DSA) is focused on centralized digital services. This makes it challenging for decentralized services, such as instances in the Fediverse, to know how to comply with the DSA. To address this issue, this report offers a practical guide for server administrators in the Fediverse to meet the DSA's content moderation obligations.

In this practical guide, you will find:

- categorization of Fediverse instances under the DSA;
- content moderation obligations for *all* intermediary services;
- content moderations obligations for *hosting* services in particular; as well as
- for *online platforms* in general.

In this report instances in the Fediverse are classified as hosting services, to be precise, as online platforms. As an online platform, instances will have to comply with the general content moderation obligations for all intermediary services, as well as the additional obligations for hosting services and online platforms. At the same time, micro or small enterprises are exempt from the additional obligations that online platforms have, which means that instances meeting this exemption are not subject to the additional obligations.

To simplify the steps that a server administrator can take to comply with the DSA, this report provides checkboxes to help server administrators determine if they fall under the DSA and, if so, what their obligations are. Additionally, platforms are encouraged to take on further responsibilities by, for example, adopting voluntary codes of conduct. We also conclude that even if an instance does not meet the exact requirements of a micro or small enterprise, full compliance with the additional obligations for online platforms may be less of a focus point for enforcement if you are a relatively small service.

Finally, and most importantly, we advise server administrators of small instances to provide transparency in their content moderation practices.

Table of Contents

1. Introduction....4
2. Background on the DSA....5
3. The Fediverse....7
4. Categorizing the Fediverse under the DSA....9
4.1. Classification of service....9
4.2. The exemption under Article 19 DSA....11
5. Content moderation....13
5.1. Obligations for all intermediary services....13
5.2. Additional obligations for online platforms....17
5.3. Additional voluntary actions....20
5.5. Checklist content moderation obligations....22
6. Conclusion and advice....23

1. Introduction

The Fediverse, abbreviated from 'Federated Universe', provides a groundbreaking alternative to the conventional paradigm of social media platforms. Unlike traditional networks, the Fediverse operates on the principles of federation and decentralization. Federated, because the Fediverse embodies a multitude of independent servers, or instances, blending together to facilitate seamless communication and data-sharing. Decentralized, because there is no central authority in charge of the Fediverse. Since each instance within the Fediverse functions autonomously, it is governed by its own set of rules, policies, and communities, under the stewardship of various server administrators.

This decentralized nature of the Fediverse, however, has a major impact on how user content is effectively moderated across different platforms and it raises questions about what server administrators are supposed to do. Server administrators hold primary responsibility for setting and enforcing content moderation policies specific to their instances, but not every server administrator has the tools or the know-how to do so. Bigger instances with fewer resources struggle, for example, more with content moderation, leading to the potential spread of illegal and harmful content via their services.

A response to concerns regarding the spread of illegal or harmful content, in general, is provided by the European Union (EU) in the Digital Services Act (DSA), a regulation that entered into force on 16 November 2022 and applies to all platforms since 17 February 2024.[1] Amongst its objectives, the DSA aims at regulating content moderation practices. It imposes obligations on intermediary services, including online platforms, to enhance transparency, oversight, and accountability. This means that server administrators in the Fediverse are also expected to align their content moderation practices with the obligations as set out in the DSA. This report aims to clarify how to do so.

This report includes practical guidelines for administrators in the Fediverse, offering insights into how DSA rules apply to their platforms, with a view to facilitate compliance. The report starts with background information on the Fediverse, for those who would like to understand the workings of instances and servers in a decentralized and federated nature. This introduction is followed by an analysis of how instances in the Fediverse may fall under the definitions and scope of application of the DSA. Roadmaps (in the form of check-boxes) are provided to help you assess your status on a step-by-step basis. Once you have classified your service, the following part supplies an overview of the content moderation obligation per type of service applicable in the Fediverse. There is a checklist at the end that can help you confirm if you have followed through on all the necessary obligations. The report concludes with general advice.

[1] Regulation (EU) 2022/2065 of the European Parliament and of the Council of 19 October 2022 on a Single Market For Digital Services and amending Directive 2000/31/EC (Digital Services Act), hereafter "DSA".

2. Background on the DSA

As mentioned above, the DSA applies to all platforms from 17 February 2024 onwards. It is a new EU regulation that updates the legal framework for digital services in the EU. It builds upon the previous E-Commerce Directive (ECD)[2] from 2000, which primarily applies to intermediary service providers that offer services such as hosting, caching, and mere conduit. The ECD is known for providing a legal framework containing liability exemptions or 'safe-harbors' for certain services of intermediaries.

However, it is important to recognize that the digital landscape has significantly evolved since then. The emergence of big tech companies and the proliferation of disinformation, hate speech, and harmful content online made it necessary to revise the rules, eventually leading to the proposal of the DSA in 2020.

The DSA does not fully replace the ECD, but complements it. The DSA includes provisions related to platform liability, content moderation, advertising transparency, data access, and user rights, amongst others. The main objective of this regulation is to establish a digital environment that is fair, safe, and transparent for society as a whole. It expands on the principles and provisions of the ECD by covering a wider range of digital services, including online platforms and search engines. The legislation imposes new obligations and requirements on these services, such as measures to combat illegal content with notice and action mechanisms, enhance transparency, and protect user rights.

A distinctive feature of the DSA is that it follows a layered approach, which means different rules and obligations apply based on the type of digital service provider, depending on their role, size and impact on the online ecosystem (see figure 1). Larger online services have a bigger significant impact on the online ecosystem and society, which necessitates them to have more stringent obligations to consider.

[2] Directive 2000/31/EC of the European Parliament and of the Council of 8 June 2000 on certain legal aspects of information society services, in particular electronic commerce, in the Internal Market ('Directive on electronic commerce').

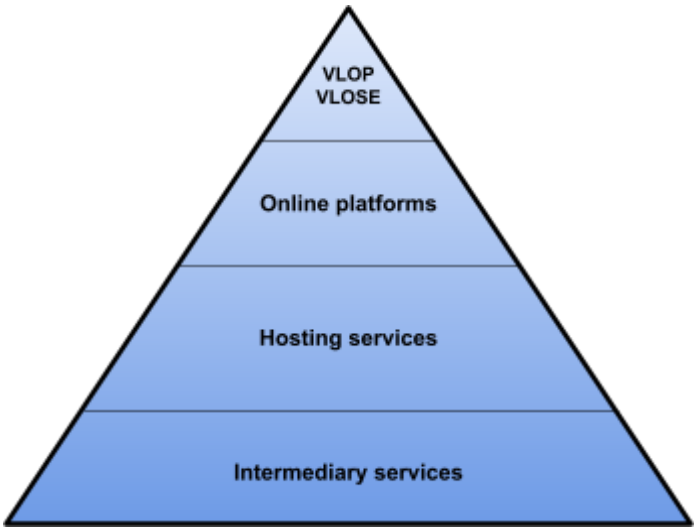

Figure 1: The layered pyramid structure of the DSA.

All digital services located or provided in the EU must comply with basic obligations. Additional obligations are introduced as you move up the pyramid; the provisions in Section 1 of the DSA apply to all intermediary services, Section 2 to hosting services, and so on. The most comprehensive set of rules and oversight apply to the ones in the top of the pyramid, which are the Very Large Online Platforms (VLOPs) and Very Large Online Search Engines (VLOSEs). See paragraph 4.1. of this report for more detail on the definitions of services.

3. The Fediverse

In the vast landscape of the internet, the Fediverse has a unique status. Its roots can be traced back to the early 21st century, where this model emerged amidst a growing frustration with the monopolistic control of traditional social media giants, currently exemplified by Facebook, Instagram, TikTok, and YouTube. These online platforms have ascended to unprecedented power over the past two decades, reshaping not only the digital landscape but also the social and economic structures in the world. What began as platforms promising connectivity and community, quickly evolved into digital empires, wielding immense influence over information dissemination, market dynamics, and our own individual liberties, such as privacy and freedom of expression. The ideology behind the Fediverse is to empower users to reclaim control over their digital interactions and content, and provide them the freedom to curate their own online experiences.

To understand the position of platforms within the Fediverse, it is important to know how it operates, technically and practically. The Fediverse is a federated universe. This means that it is based on a federative model with multiple independent servers or instances that run compatible software that adheres to a decentralized communication protocol. Such a protocol is the ActivityPub protocol ('ActivityPub'), which defines a set of standards for interaction between different social media platforms and servers in the Fediverse. ActivityPub enables the federation of internet platforms and interoperability between instances, which allows users from one instance to interact and share content with users from another instance.

So each instance or server operates independently. Servers can host a specific type of service, such as microblogging, blogging, or video-sharing, and often have a unique domain name. Popular Fediverse services include Mastodon, Lemmy, PeerTube, and Pixelfed. Because of its independent nature, each instance has its own rules, policies, and communities, which their server administrators set and manage. This type of architecture, which includes a variety of specific moderation rules per instance, is in stark contrast with that of centralized services, such as Facebook, whose content moderation policies are also centralized. In other words, every Facebook user, account, or page has to follow the same top-down formulated rules, which are enforced through a number of measures, such as content removal or account suspension. The decentralized nature of the Fediverse allows its users to choose their community based on a preference for a set of rules and policies amongst the choice of multiple instances.

To simplify, in the Fediverse, you make an account on, for example, one instance within the service Mastodon. You choose your instance based on your personal preferences. Maybe you like science and you wish to connect to other people with an interest or background in science so you choose to make an account on the instance 'sciences.social', also knowing that they enforce deliberately minimal rules but do not allow racism, sexism, homophobia,

transphobia or casteism and precludes abusive behavior and doxxing. Now you have a Mastodon account up and running on one instance.

With this account, you can see and interact with your mother's post on another Mastodon instance, for example, 'historians.social' (unless two instances have not disconnected from each other). Your instance may have other rules and policies than your mother's choice of instance, yet, communication and visibility remain possible so far as each instance allows it. With your account, you can also move to other services to interact with users on other Fediverse services. For example, with your Mastodon-related account, you can engage with your colleague's video on any instance on PeerTube or your partner's meme on Lemmy. You only have to make one account on one instance to communicate with or post on any other instances. Concurrently, your followers should be able to view and engage with all your posts through the instances that they prefer. A true federation of instances.

As we know centralized platforms, such as Facebook and Instagram, are owned by Meta and Meta decides what content is moderated and how. Because there is no central authority in the Fediverse, neither the non-profit 'Mastodon gGmbH' (Mastodon's developer) nor 'Framasoft' (PeerTube's developer) or any entity is in charge of defining what is allowed and what is not. Each instance, no matter the size of it, decides upon its own rules. Server administrators hold primary responsibility for setting and enforcing content moderation policies specific to their instances. Some choose to be strict on limiting the spread of possible false information (historians.social does not allow posting content that says that 5G causes coronavirus and penalizes this with suspension[3]), whereas other sites may be more lenient towards a certain kind of content. The decentralized nature of the federation makes it unlikely to moderate content consistently across the entirety of a platform. This could make instances more prone to the spread of illegal or harmful content. In fact, practical cases reveal varying degrees of community supervision, with open-rule communities at risk of abuse. The Gab-case is a well-known example of this.[4]

However, despite its decentralized and federated nature, Server administrators within the Fediverse also have to follow a standard of rules and regulations. For instances or servers established in the EU, those who have users in the EU, or those who offer services in the EU, the DSA applies. The DSA regulates the obligations of digital services and is directly applicable across the EU, replacing any overlapping national laws and establishing uniform rules for all EU member states to ensure equal rights and protections for citizens online. In the next section, the categorization of the Fediverse under the DSA is explained, followed by the specific content moderation obligations per service.

[3] see: https://historians.social/about.

[4] As a Mastodon instance, Gab was a platform for people who felt marginalized or censored on other platforms. Gab was criticized for hosting extremist content, and due to that, it got isolated from other instances due to their unwillingness to engage with its content. This isolation manifested itself in a practice known as 'defederating', in which Gab voluntarily severed ties with the communities that did not want to interact with the platform. This intentional disconnection highlighted the complexities of content moderation within the decentralized social media landscape, demonstrating instances' ability to isolate and regulate themselves in response to divergent content policies.

4. Categorizing the Fediverse under the DSA

Federated decentralized social networks without effective content moderation run the risk of facilitating the spread of harmful and illegal content, potentially leading to an unregulated and unsafe online environment. As mentioned before, the DSA sets out specific obligations for intermediary services. If your instance or server is established in the EU, or if you have users in the EU, or you offer services in the EU, then the DSA applies to you. However, before understanding your obligations, it is crucial to understand your position as the server administrator of an instance. The DSA does not mention 'instances' or 'servers', nor does it cover the term 'server administrator'. Instead, the DSA regulates the obligations of online 'intermediary services' and speaks of 'providers of a service'. Server administrators in the Fediverse are considered providers of services, and therefore, as providers of online services, they have obligations. The obligations applicable to your service depend on the nature of the service you offer. The following section will explain the type of intermediary service an instance in the Fediverse is.

4.1. Classification of service

There are three defined types of intermediary services: 'mere conduit', 'caching', and 'hosting' services. The DSA also refers to 'online platforms' and 'very large online platforms' (VLOPs) and 'very large search engines' (VLOSEs), which are differentiated by the number of users and the reach of their services.[5]

Mere conduit and caching services are key types of online intermediaries, however, do not apply to the Fediverse. Mere conduit services involve simply transmitting information over a communication network or providing access to such a network, meaning they just pass along data provided by users without altering it. Caching services also transmit information but include automatic, temporary storage of data. This storage is only to make future data transmission more efficient for users. Both types of services are important for ensuring data moves smoothly and quickly online.

In contrast, a hosting service provides users with a platform or infrastructure where they can store information or data and access it whenever needed. In other words, a hosting service is an intermediary service that offers server space, bandwidth, and resources to those service providers who want to manage a website or application. Well-known web-hosting services or cloud-computing services are, for example, GoDaddy or IONOS.

[5] VLOPs are for example Facebook, TikTok and Youtube, as well as Amazon Store and Booking.com. VLOSEs are Bing and Google Search. In order to be defined as a VLOP and VLOSE, certification by the EC is required. Because of their enormous reach, and therefore influence, VLOPs and VLOSEs have more stringent obligations under the DSA. However, it is not necessary to further consider VLOPs and VLOSEs in light of the scope of this report since none of the services in the Fediverse reach the threshold of VLOPs, which is 45 million active users per month (Article 33 DSA), nor does any service fall under the definition of a search engine (Article 3(j) DSA).

Online platforms and VLOPs are types of hosting services, but they are more than that. An online platform is a hosting service that stores and shares information or data when asked, but its main difference is that it serves to actively disseminate information to the public. Online platforms enable individuals and businesses to publish, share, and access content on the internet. Famous online platforms, which are also VLOPs, are Instagram, YouTube, and TikTok, as well as business-to-consumer (B2C) marketplaces such as Amazon and Alibaba.[6] Online platforms that are not VLOPs are, for example, Strava and Etsy.

Instances within the Fediverse provide users with the infrastructure to store and access data, create content, and interact with others, thus embodying the essence of hosting services. Moreover, they display the key features of online platforms as defined under the DSA by allowing users to share posts, images, and other content widely, thereby actively distributing information across the network and making it accessible to the broader public. In other words, Fediverse instances qualify as online platforms under the DSA. This means that each individual instance within a particular service platform (whether that is Mastodon or PeerTube instance) may in theory qualify as an online platform since each individual instance runs on its own server, storing and allowing access to information related to their instance and other instances they interact with. Their users can upload their content to the instance's server, which is then made accessible to other users via the internet. Even when instances store data at third-party servers, the server administrators of these instances are still responsible for the data of their users, and therefore, each instance may still qualify as an online platform under the DSA.

Roadmap I: Application & Classification

1. *Are you established in the EU, do you have users in the EU, or do you offer services in the EU?*
   ➔ Yes: DSA applies, proceed to *step 2*.
   ➔ No: DSA does not apply.

2. *Is the storage of information and making content available upon request by a recipient of the service a core part of your service?*
   ➔ Yes: Your service probably qualifies as hosting[7] (e.g. web-hosting services), proceed to *step 3*.

[6] See the list of all the designated VLOPs: <https://digital-strategy.ec.europa.eu/en/policies/list-designated-vlops-and-vloses>.
[7] Hosting services consist of the storage of information provided by, and at the request of, a recipient of the service, Article 3(g)(iii) DSA.

➔ No: Your service probably either qualifies as caching[8] or mere conduit[9]. This report may not be of help for you.

3. *Does your hosting service actively disseminate information to the public?*
   ➔ Yes: Your service probably qualifies as an online platform[10] (e.g. social networks; content-sharing platforms).
   ➔ No: Your service probably qualifies as hosting only.

If your instance qualifies as a hosting service or an online platform, you, as the server administrator (service provider), will have to comply with the content moderation obligations under the DSA as mentioned in paragraphs 5.1 and 5.2 of this report. These obligations include reporting obligations and putting an internal complaint-handling system in place, amongst others, *unless* you fall under the exemption as mentioned in Article 19 DSA. The next paragraph will explain the exemption under Article 19 and how to determine if your instance falls under this exception.

4.2. The exemption under Article 19 DSA

Article 19 of the DSA exempts online platform providers that qualify as micro or small enterprises from the extra obligations established by the DSA for online platforms.[11] This exemption is in place because micro and small enterprises typically have limited resources and capacity compared to larger entities. The additional rules for online platforms are designed to ensure greater accountability and transparency, which can be resource-intensive to implement and maintain. Consequently, to prevent undue burden on these smaller entities, the DSA does not apply these additional rules to micro or small enterprises.

The European Commission (EC) has defined an 'enterprise' as: "any entity engaged in an economic activity, irrespective of its legal form".[12] This includes "self-employed persons and family businesses engaged in craft or other activities, and partnerships or associations regularly engaged in an economic activity".[13] It is worth noting that in the Fediverse, non-profit instances may not be engaging in 'economic activity' in the strict sense, but the broad definition of economic activity means that even non-profit organizations are likely to be considered as engaging in an economic activity. In other words, non-profit entities should qualify for this exemption.

[8] Caching services consist of the transmission in a communication network of information provided by a recipient of the service, involving the automatic, intermediate and temporary storage of that information, performed for the sole purpose of making more efficient the information's onward transmission to other recipients upon their request, Article 3(g)(ii) DSA.
[9] Mere conduit services consist of the transmission in a communication network of information provided by a recipient of the service, or the provision of access to a communication network, Article 3(g)(i) DSA.
[10] An 'online platform' is a hosting service that stores and shares information with the public when asked, Article 3(i) DSA.
[11] Article 19(1) DSA.
[12] Article 1, Recommendation 2003/361/EC.
[13] *Ibid*.

The table below illustrates the differences between the categories of micro and small enterprises. In any way, instances will fall under this exemption if they have less than 50 employees and do not have an annual turnover and/or an annual balance sheet exceeding 10 million euros.[14] This also includes donations that some instances get from their users. The vast majority of Fediverse instances will fall under the exemption and therefore do not have to comply with the transparency reporting obligation (see paragraph 5.1) and the additional obligations as set out in section III in the DSA, which are covered in paragraph 5.2.

Size	**Employees**	**Annual turnover/annual balance sheet**
Micro	< 10 persons	< 2 million EUR
Small	< 50 persons	< 10 million EUR

In summary, it becomes evident that each individual instance within the Fediverse can potentially fall under the exemption outlined in Article 19 of the DSA. By qualifying as micro or small enterprises, these instances are spared from the additional obligations imposed on larger online platforms.

Roadmap II: Assess the application of Article 19 exception to your service

1. *Are you an online platform (see Roadmap I)?*
   - ➔ Yes: Proceed to *step 2*.
   - ➔ No: The exception of Article 19 DSA does not apply to you.

2. *Do you have less than 10 employees and an annual turnover of less than 2 million euros?*
   - ➔ Yes: You qualify as a micro-enterprise. The exception of Article 19 applies, meaning you do not have to apply the extra obligations for online platforms.
   - ➔ No, more: Proceed to *step 3*.

3. *Do you have more than 10 but less than 50 employees and an annual turnover of less than 10 million euros?*
   - ➔ Yes: You qualify as a small enterprise. The Article 19 exception applies, meaning you do not have to apply the extra obligations for online platforms.
   - ➔ No, more: You do not qualify as a micro or small enterprise and Article 19 does not apply to you. This means that you do have to comply with the extra obligations for online platforms.

[14] Recommendation 2003/361/EC.

5. Content moderation

Content moderation is referred to in the DSA as the actions taken by intermediary service providers to find and deal with illegal or inappropriate content to ensure compliance with community guidelines, terms of service, and/or legal requirements. It involves the review, editing, removal, or approval of content uploaded by users to maintain a safe, respectful, and appropriate online environment. These actions can be done manually or through automated tools and may involve detecting, identifying, and addressing this content. Some platforms involve community engagement in their content moderation practice through user reporting systems and community guidelines while others employ professional moderators. Content moderation measures can include making the content less visible, removing it, or even suspending the account of the person who shared it.

The DSA contains several obligations that *all* intermediary services must comply with, as well as additional obligations that service providers must comply with depending on, and proportionate to the size and risk of the service. This report highlights below the specific obligations covering content moderation practices for *all* intermediary services, including the obligations for hosting services, followed by the additional obligations for online platforms. The obligations for VLOPs and VLOSEs are not mentioned since Fediverse services so far have not reached the threshold of an average of 45 million active monthly users.

5.1. Obligations for *all* intermediary services

All intermediary services under the DSA, including hosting services and online platforms, have the general obligations to designate points of contact (**obligation A**), to establish 'terms and conditions' for their users (**obligation B**), and to publish transparency reports[15] (**obligation C**). Additionally, as a *hosting service*, you will also have to implement 'notice-and-takedown' mechanisms (**obligation D**) and act upon illegal or incompatible content (**obligation E**).

Below, we have set out the specifics of these obligations. Note that these are the minimum obligations for any (hosting) service provider in the Fediverse. If you qualify as an *online platform* that does not fall under the exclusion of Article 19 (see **Roadmap I** and **II**), you will have to also comply with the *additional* obligations established for online platforms (see paragraph 5.2. on **obligations F-H**).

Further note that these obligations are independent from the question of liability of services.

[15] *Note* that the reporting obligations do not apply to online platforms that qualify as micro or small enterprises (see Roadmap II).

All intermediary services, which include mere conduit, caching, hosting services, and online platforms, have the following content moderation obligations:

A. Establish points of contact.

Server administrators will have to designate single points of contact to enable them to have contact with the dedicated authorities, as well as the recipients of their service:

First of all, you will have to enable direct contact with your States' authorities[16], the European Commission and the European Board for Digital Services through electronic means.[17] It is necessary to make the contact information easily accessible, public and to keep it up-to-date. Make sure you present this information in the official language of your State and a language that is broadly understood by the largest possible number of EU citizens.

Additionally, you will have to enable your users to communicate with you directly and rapidly, through electronic means and in a user-friendly manner.[18] This includes allowing the users to choose the means of communication. It is not regarded as user-friendly to rely solely on automated tools. It is necessary to make the contact information easily accessible, public and to keep it up-to-date.

If you do not have an establishment in an EU country, but you do offer services in the EU, you will have to appoint and mandate a legal or natural person to act as your legal representative in the Member State where you provide your service.[19] The name, postal address, email address and telephone number of the legal representative will have to be communicated with the Digital Services Coordinator in the Member State where that legal representative resides or is established.

B. Provide your users with a set of 'terms and conditions' in which you clearly state your content moderation policy, procedures, measures, tools, and processes.[20]

[16] With regards to the State's authorities, the supervisory and enforcement responsibilities of the DSA are delegated to the Digital Services Coordinators (DSCs). These authorities are appointed at the national level within each Member State, find your DSC here: <https://digital-strategy.ec.europa.eu/en/policies/dsa-dscs#:~:text=Each%20Member%20State%20has%20to,designate%20their%20Digital%20Services%20Coordinators>.

[17] Article 11 DSA; The European Board for Digital Services ("the Board") is referred to in Article 61 for the application of the DSA.

[18] Article 12 DSA.

[19] Article 13 DSA.

[20] Article 14 DSA.

Server administrators must clearly state any restrictions they impose on user-provided information in their terms and conditions. This includes detailing your content moderation policies, procedures, measures, and tools, such as algorithmic decision-making and human review processes. For example, you define that your service does not allow discrimination of such-and-such sort, that you use such-and-such automated tools to detect discriminatory content, and that if such content is found you apply such-and-such measures.
The terms and conditions must be easy to understand for your general user, written in clear language, made publicly accessible, and machine-readable. The more user-friendly, the better. If your service is used by minors, you should explain your rules in a way that minors can comprehend.

You must always inform your users if you make any significant changes to the terms and conditions.

You must apply and enforce your restrictions diligently, objectively, and proportionately, considering the rights and interests of all parties involved.

C. You will have to publish 'transparency reports' with information on content moderation measures taken.[21]

*Note: These reporting obligations do not apply to online platforms that qualify as micro or small enterprises (see **Roadmap II**). However, if you do qualify as an online platform, then you will also have to take additional transparency reporting obligations into account (see **obligation I**).*

Server administrators must share clear and understandable reports at least once a year about their content moderation activities. These reports should be easy to access and published in a format that is machine-readable.

Your yearly transparency report must include at least:

- The number of orders you got from authorities in different EU countries, what kind of illegal content was involved, and how long it took you to respond to these orders;
- for hosting services (incl. online platforms) additionally how many notices you received about potentially illegal content (see **obligation D**), who sent these notices, what actions you took in response (based on the law or your own rules), how many and which notices were handled automatically, and how long it took you to act on them;
- details about your own content moderation efforts, such as using automated tools or not, trained moderators, and what actions you took to restrict the

[21] Article 15 DSA.

content. You should also categorize this information by (1) the type of illegal content, (2) how it was detected, and (3) what restrictions were applied;
- the number of complaints you received through your complaint systems (see **obligation F**), what these complaints were about, what decisions you made, how long it took to make these decisions, and how many times you have changed your decisions, and;
- details about any automated tools used for content moderation, including how accurate they are, and if, and if so, what kind of safeguards are in place.

Further content moderation obligations for server administrators of ***hosting services*** in the Fediverse, including online platforms, are:

D. You have to implement or update your 'notice-and-action mechanisms'.[22]

Server administrators should put mechanisms into place that allow anyone to notify about the presence of possible illegal content.

Those mechanisms should be easy to access, user-friendly, and exclusively through electronic means.

The mechanism in place should help the person or entity making the notification to do so in (i) a sufficient way in which a substantiated explanation of reasoning is encouraged, provided with (ii) the exact electronic location of the illegal information, (iii) the name and email address of the individual or entity submitting the notice[23], and (iv) a statement confirming that the notice is made in good faith.

You will have to send a confirmation of receipt of the notice to that individual or entity, without reasonable delay.

Additionally, you will need to inform the person or entity about what decision you have made regarding the notification and offer options for the person or entity to challenge or appeal this decision. If you use automated tools or systems to process these notices or make decisions about content, you must disclose this information in the notification you send to the individual or entity in question.

You are expected to handle these notifications promptly and carefully and you are expected to make fair, timely, and thorough decisions based on clear criteria, ensuring transparency in the processes.

[22] Article 16 DSA.
[23] Except in the case of information considered to involve one of the offenses referred to in Articles 3 to 7 of Directive 2011/93/EU.

E. If content is found to be illegal or incompatible with your terms and conditions, you will need to act.[24]

The server administrator has to offer a clear and specific 'statement of reasons' to the affected party if the contact details of the recipient are known.

You can then decide to restrict the visibility of certain content, remove the content, disable access to the content, or demote the content. Besides, the hosting provider can suspend, terminate, or restrict monetary payments, suspend or terminate the provision of the service in whole or in part, and or suspend or terminate the recipient's account.

The statement of reasons needs to include the following information in a clear and comprehensible way:

- What decision did the hosting provider take (removing content, suspending an account, etc.);
- on what basis this decision is made (due to a notification or because of voluntarily investigation);
- if automated means are used;
- a reference to either the legal ground or the terms and conditions as the contractual ground; and
- user-friendly ways to dispute the decision.

In the situation where you as the provider of a hosting service become aware of any possible criminal offense involving a threat to the life or safety of a person or persons, you will have to promptly inform any law enforcement agency.

5.2. Additional obligations for *online platforms*

If you qualify as an online platform (see **Roadmap I**), you will also have to take additional content moderation obligations into account, *unless* you qualify as a micro or small enterprise (see **Roadmap II**), with the exemption of Article 24(3) DSA, which mentions the requirement 'to publicly share online how many people use your service each month in the EU' (see second paragraph **obligation I**). The obligations for online platforms include an internal complaint-handling system (**obligation F**), the priority of trusted flaggers (**obligation G**), the suspension of frequent wrongdoers or misusers (**obligation H**), and additional transparency reporting obligations (**obligation I**). As they are additional, they are supplementary to the

[24] Article 17 and 18 DSA.

general obligations listed above (**obligations A-E**). In more detail, the following additional obligations for online platforms are:

F. You need to set up an internal complaint-handling system.[25]

With regards to the notification of illegal or incompatible content (see **obligations D** and **E**), server administrators of online platforms need to provide an effective internal complaint-handling system to allow their users to file complaints against the following decisions:

- Your decision whether or not to remove, disable access to, or restrict the visibility of the information.
- Your decision whether or not to suspend or terminate the provision of the service, in whole or in part, to your users.
- Your decision on whether or not to suspend or terminate your users' account.
- Your decisions whether or not to suspend, terminate or otherwise restrict the ability to monetize information provided by your users.

Access to the internal complaint-handling system should be given for at least six months after your given decision, which term starts on the day your user has been notified.

The internal complaint-handling systems should be electronically and free of charge. It should also be easily accessible, user-friendly, and should facilitate the submission of precise and substantiated complaints.

It is necessary that you handle these complaints in a timely, non-discriminatory, diligent, and non-arbitrary manner.

When the complaint leads you to conclude that your decision should be reversed, you shall do so without undue delay. You will also have to make sure that such a decision is not solely based on automated tools and should be under the responsibility of qualified employees.

You will also need to inform the complainant of their right to choose any certified out-of-court dispute resolution body for settling disputes that arise from the handling of a complaint.[26] Ensure that information about these dispute resolution options is clear, user-friendly, and easily accessible.

[25] Article 20 DSA.
[26] See further Article 21 DSA for this.

G. You need to give priority to 'trusted flaggers'.[27]

Server administrators of online platforms need to implement technical and organizational measures to prioritize and handle complaints from so-called 'Trusted Flaggers' for them to report illegal content to which platforms will have to react with priority.

Trusted Flaggers are independent entities that have demonstrated particular expertise and competence, which they objectively use to detect and report upon allegedly illegal content. Trusted flaggers will be appointed from 17 February 2024 on by Digital Services Coordinators, the national authorities in charge of supervising and enforcing the DSA in Member States.

If you notice that a trusted flagger has submitted a significant number of insufficiently precise, inaccurate, or inadequately substantiated notices through your notice-and-action mechanisms, then you should inform the Digital Services Coordinator applicable to you.

H. Suspension of frequent wrongdoers or misusers.[28]

Server administrators of online platforms shall suspend the provision of their services to recipients of the service that frequently provide manifestly illegal content.

Prior to the suspension, you should first issue a warning to the user in question and the suspensions should be given for a reasonable period of time.

Your assessment of the frequent wrongdoers or misusers should be done so on a case-to-case basis, where you take into account all relevant facts and circumstances available to you. These circumstances should at least include the following:

- The absolute number of items of illegal content, or misusages of notice or complaint systems submitted within a certain time frame;
- the relative proportion thereof compared to all the information or notices shared during a specific period;
- the gravity of the misuses, including the nature of illegal content, and of its consequences;
- if possible, the intention of the individual or entity in question.

[27] Article 22 DSA.
[28] Article 23 DSA.

Note that you will have to add to your 'terms and conditions' (see **obligation B**) your policy regarding handling such situations, define what circumstances will lead you to suspension, and describe in what manner your assessment of such actions will be done.

I. Additional transparency reporting obligations.[29]

On top of the transparency reporting obligations (see **obligation C**), server administrators will have to include the following information in their yearly transparency reports:

- the details on the disputes submitted to the out-of-court dispute settlement bodies (this procedure is mentioned Article 21 DSA);
- the number of suspensions imposed following **obligation H**, including the mentioning of if they were for clearly illegal content, false notices, or baseless complaints.

Moreover, since February 17 2024, and every six months after, you will have to publicly share online how many people use your service each month in the EU. If asked, you have to tell the Digital Services Coordinator and the Commission these numbers, updated at that moment, but without sharing personal data. *Note: this part of the transparency obligation is not exempted by Article 19; you will have to do this even if you are a micro or small enterprise.*

As last, you will also have to share your statement of reasons (**obligation E**) with the Commission.

5.3. Additional voluntary actions

In addition to the general and additional obligation as mentioned above, stakeholders can establish voluntary rules ("codes of conduct") or other self-regulatory measures.[30] Codes of conduct are voluntary frameworks established to guide behavior and practices within certain industries or sectors. They serve as a tool for self-regulation and outline standards and principles in order to promote compliance with EU laws and regulations.

The EC endorses the creation of codes of conduct, mainly for protecting minors, tackling disinformation, and addressing gender-based violence.[31] These voluntary guidelines should consider best practices and available guidance, and function as an additional tool to help make sure the DSA is applied correctly. Codes of conduct should at least state the nature of

[29] Article 24 DSA.
[30] Crisis protocols, as well as awareness-raising action, are mentioned in the DSA in relation to VLOPs and VLOSEs, so therefore this report does not elaborate on these.
[31] Recital 71; 104, DSA.

their public goal, include independent evaluation mechanisms to check if these goals are met, and explain what authorities are involved and how.[32] In the DSA, Articles 45-47 provide the rules on codes of conduct for providers of online platforms and other relevant service providers.

Because codes of conduct are voluntary, there is no obligation on the server administrators of an intermediary service to engage in such self-regulatory efforts. However, because codes of conduct can shape the DSA's effectiveness in practice, they are highly appreciated. Therefore, server administrators are encouraged to form codes of conduct for content moderation practices in the Fediverse. Implementing codes of conduct into your service will also help you with your compliance and will most likely be considered regarding enforcement and supervision.[33]

[32] Recital 103, DSA.
[33] See Article 75(3) DSA.

5.5. Checklist content moderation obligations

Checklist: Content moderation obligations for hosting services and online platforms in the Fediverse

	You are a hosting-service	You are an online platform	
A. Points of contact	☐	☐	
B. Terms and conditions	☐	☐	
C. Transparency reports	☐	☐	Art. 19 exclusion may apply to you
D. Note-and-takedown mechanisms	☐	☐	
E. Act upon illegal or incompatible content	☐	☐	
F. Internal complaint-handling system		☐	Art. 19 exclusion may apply to you
G. Priority to trusted flaggers		☐	Art. 19 exclusion may apply to you
H. Suspension of frequent wrongdoers		☐	Art. 19 exclusion may apply to you
I. Additional transparency reporting obligations		☐	Art. 19 exclusion may apply to you[34]

[34] Article 19 clearly exempts Article 24(3), which is the requirement to publicly share online how many people use your service each month in the EU, and to, if asked, inform the Digital Services Coordinator and the Commission these numbers, updated at that moment, but without sharing personal data.

6. Conclusion and advice

The DSA aims to create a safer online environment, and its content moderation provisions play a crucial role in achieving this goal. Also in the Fediverse, a decentralized network of interconnected instances, server administrators will have to prevent the spread of illegal or harmful content through content moderation practices. This guide aims to assist server administrators in the Fediverse in understanding how the DSA obligations regarding content moderation apply to their instances.

As server providers, server administrators in the Fediverse have obligations for content moderation. Chapter 5 has covered the obligations that server administrators must comply with related to content moderation, which may vary depending on the classification of service. This report has classified instances in the Fediverse as hosting services, and in particular online platforms, see Chapter 4. This means that server administrators have to follow general intermediary rules (contact points, terms of service, transparency reports) and handle illegal content (notice-and-takedown). Additionally, online platforms require complaint systems, prioritizing trusted flaggers, and suspending repeat offenders, along with extra transparency reports.

If the instances qualify for the exemption as stated in Article 19 DSA, server administrators are not required to implement the additional obligations for online platforms. The European Commission defines 'enterprises' broadly as any entity involved in economic activity, including self-employed individuals, family businesses, partnerships, and nonprofit entities. However, even if they are exempt from the additional obligations, they still have to meet the additional transparency reporting part that requires service providers to publicly disclose the number of monthly users in the EU and provide this information to the Digital Services Coordinator and the Commission upon request. Despite Article 19, online platforms are not exempt from their *general* obligations as intermediary and hosting services.

The DSA can be challenging for small Fediverse instances to follow. Therefore, it may be beneficial to distinguish between full legal compliance and compliance in spirit to reduce the burden on small instances with limited resources. It is important to remember that greater compliance reduces the risk of liability issues.

The main goal of the DSA regarding content moderation is ensuring transparency. Therefore, this report recommends full compliance with the terms and conditions obligation. This includes detailing your content moderation policies, procedures, measures, and tools, such as algorithmic decision-making and human review processes. Additionally, transparency is meaningless without clear points of contact. Therefore, establishing designated points of contact is also crucial. Enforcing policies against illegal or harmful content, is another recommended obligation, particularly if the content violates your terms and conditions. It is

advisable to have notice and takedown mechanisms in place for this purpose, although the practical details of these mechanisms may vary. On the other hand, there should be less emphasis on transparency reports, as this falls under the Article 19 exception and small instances may not have the capacity to fulfill such reporting tasks.

Even if your instance does not meet the exact requirements of a micro and small enterprise for whatever reason, less weight could be given to compliance with the *additional* obligations for online platforms when you are a relatively small service. However, it is important to note that if your instance forms an important platform for the distribution of content and facilitating debate, that the impact or reach of your online platform may be taken into account by enforcement authorities.

Overall, compliance with the DSA requires a nuanced approach as a server administrator. In prioritizing transparency, it is important to strike a balance between legal obligations and practical implementation to contribute to a safer online environment for you and your users.