
Submitted August 31, 2023

2023 EU Systemic Risk Assessment

- *Cover note* -

Executive Summary

Wikipedia’s inaugural 2023 EU Systemic Risk Assessment—in spreadsheet form—is appended to this cover note. It sets out the Wikimedia Foundation’s assessment of “systemic” risks, in the EU, linked to the use of Wikipedia. This is a legally-required document under the new European Union Digital Services Act (DSA).

Out of 11 EU systemic risks that were assessed, the 2023 risk assessment designates three immediate priorities:

1. Disinformation around conflicts, and/or civic and electoral processes
2. Disinformation around historical/geographical narratives
3. Harassment of Wikipedia’s volunteer users

The Foundation is now documenting a range of existing and planned mitigations for all 11 risks. Once complete, the mitigations plan will also be submitted to regulators—as a follow up to this Risk Assessment—without delay. This will be repeated at least once a year.

Globally, the Foundation deals with a much wider range of risks—such as Wikimedia projects being blocked in some non-EU countries. Despite those global risks’ importance, we understand them to be out of scope for this DSA-specific exercise.

[The Wikimedia Foundation’s 2023-24 Annual Plan](#) provides a more holistic overview of the issues our volunteer editors and audience of readers are facing every day, and what is being done to protect and defend the rights of our audiences and editors to access and share free knowledge across borders. Furthermore, the Foundation’s [Human Rights Policy](#) describes the organization’s commitment to identify and mitigate human rights risks globally.

About the Wikimedia Foundation

The Wikimedia Foundation is the nonprofit organization that hosts [Wikipedia and other free knowledge projects](#). The vision of our free knowledge communities is a world in which every single human being can freely share in the sum of all knowledge. To this end, we support a vibrant community of more than 300,000 volunteers around the world, who contribute to the

Wikimedia projects by adding, editing, and verifying content in over 55 million articles across more than 300 languages, all for free and without ads.

Contents

Executive Summary	1
About the Wikimedia Foundation	1
Background	2
<i>About this document</i>	2
<i>About the Wikimedia model: advantages and challenges when it comes to EU DSA SRAM</i>	4
<i>How are relevant risks selected?</i>	6
<i>What risks are excluded by this approach?</i>	7
<i>How are the selected risks assessed?</i>	8
What are the key risks highlighted in this 2023 exercise?	9
Conclusion	10

Background

About this document

Wikipedia's inaugural 2023 EU Systemic Risk Assessment sets out the Wikimedia Foundation's assessment of systemic risks linked to the use of Wikipedia in the EU. This is a legally-required document under the new DSA. It has been completed by the Foundation's Legal department in consultation with a range of internal subject matter experts and stakeholders across the Foundation.

The Risk Assessment is an intermediate step in a wider exercise. The DSA's overall Systemic Risk Assessment and Mitigation (SRAM) process requires the Foundation (as the hosting provider of Wikipedia) to not only assess risks, but also determine what risk mitigations are appropriate.

This process of assessment and mitigation is repeated at least annually. Interim updates may be appropriate, in response to significant changes in the risk landscape.

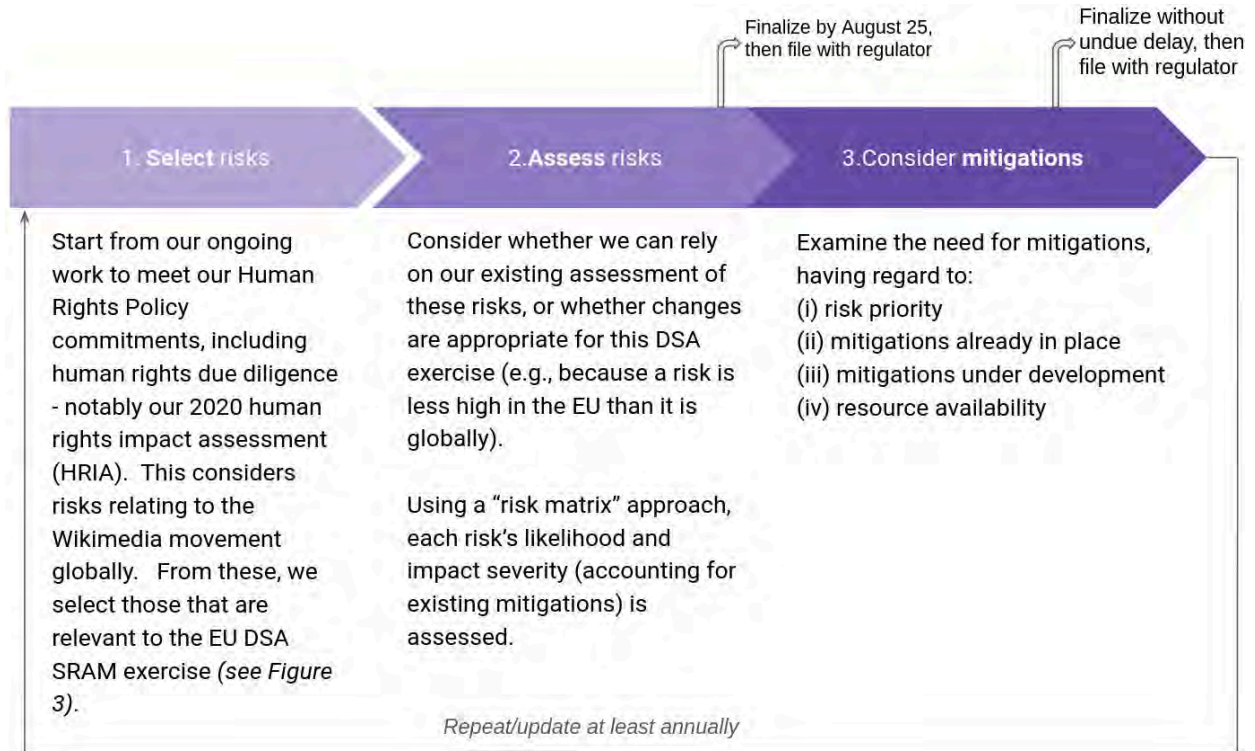


Figure 1: Diagram showing the **stages of the EU DSA SRAM process**.

The Foundation is opting for a single, living SRAM document which can be updated on an ongoing basis. This is also sometimes referred to as the Foundation’s DSA Risk Register. Each year, an annual snapshot will be taken of that document, and will be filed with regulators.

The DSA sets out risk assessment and mitigation as a two-step process (requiring two successive filings: first the risk assessment, and then the mitigation plan—see Figure 1, above). Despite this, our DSA SRAM Register has been structured so that it can be used in a single-step strategy in future years if appropriate (see Figure 2, below).

For the first year, the Register will first be submitted to the European Commission without the Mitigations content, while we assess and plan our mitigations in response to the assessed (submitted) risks. Once the mitigation planning is complete, a consolidated document (i.e. including the final column, for Mitigations) will be submitted.

Category	Risk	Risk Description	Priority (Derived from Likelihood x impact)	Risk Status	Quarter Added	Quarter Updated	Mitigations
	Disinformation regarding civic and electoral processes, and conflicts	Actors interested in a particular political/electoral/civic outcome could launch coordinated campaigns to insert misleading content into Wikipedia, reducing the broader reliability of content, misleading readers, and spreading disinformation. Such risks interfere with and diminish users' freedoms of expression and thought, right to participate in civic and political life, and right to good administration (UDHR 18, 19, 21 / C/REU 10, 11, 41).	Immediate Priority	Mitigation Ongoing	Q2 2023	Q2 2023	<ul style="list-style-type: none"> - Volunteer moderators maintain editorial policies, and remove L - Use and policies (Status: In place) - Development of an institutional strategy to counter disinforma - Establishment of temporary task forces, composed of both sta - Coordinated disinformation campaigns during significant events - Implemented as needed) - Development of tools to support moderation and content-cur - Deployment of an e-learning module to assist volunteer editors

Figure 2: Example extract from the Foundation’s EU DSA SRAM Register. Its components will be updated and filed at least once a year with EU regulators. The DSA foresees the Mitigations column (blue) being completed and filed separately, as a second step following Risk Assessment. In practice, our DSA SRAM Register—once complete, i.e. from 2024 onwards—will record both our assessment of risks and their mitigations; this example, which shows all columns, illustrates our planned end-state for the Register.

About the Wikimedia model: advantages and challenges when it comes to EU DSA SRAM

Wikipedia and other Wikimedia projects provide free access to neutral, well-sourced information about science, culture, history, and other encyclopedic subjects. The information in question is added, organized, and edited by a decentralized community of volunteers who engage in open debate to reach consensus around content decisions and policies.

Volunteers address most everyday content issues on the Wikimedia projects, such as intentional vandalism or edits that do not meet Wikipedia’s reliability and neutrality standards, without interference from the Wikimedia Foundation.

This volunteer-led, self-governing model provides for several advantages that allow our movement to advance knowledge equity and to combat disinformation through collective contributions and open debate about content and content moderation. A global community of committed volunteers with specific expertise and local knowledge and language skills have built an unprecedented repository of encyclopedic information that is available and culturally relevant to more than 300 language communities around the globe. These same qualities allow these communities to monitor Wikipedia pages for disinformation, and to rapidly remove content that may range from common vandalism to artifacts of coordinated disinformation campaigns. All of this work adding, editing, and verifying content takes place in the open and is governed by policies developed and implemented democratically by these volunteer communities themselves.

This unique model has allowed the Wikimedia projects to achieve the prominence and high quality they have now, and the volunteer-led, distributed decision-making that occurs on the Wikimedia projects is a manifestation of the ideals of online participation and freedom of expression.

Despite these advantages, some challenges naturally arise.

One challenge relevant to this work emanates from our community governance model. Beyond the Foundation’s Terms of Use and other high-level policies, most policies and decision making related to platform governance and content moderation are developed and implemented by these volunteer communities. Thus, any significant changes to platform governance and content moderation policies or practices—such as possible recommendations stemming from human rights impact assessments—require consultation with, buy-in from, and leadership by these communities. This model of governance, therefore, can be slower to effect change than that which is feasible for platforms with top-down models.

Another challenge is that the Wikimedia Foundation operates as a non-profit organization. The Foundation’s largest share of revenues comes from individual donations; it does not generate ad revenue or sell user data to generate income. As a result of this model, the Foundation has fewer financial and human resources available to tackle significant technical changes or upgrades than do other VLOPs. Any reallocation of resources to make significant changes to Wikipedia for risk mitigation requires pulling resources from other priority areas, such as requested feature development, fixing bugs or broken tools, enhancing the overall stability and reliability of the websites we host, or—more fundamentally—continuing to create an environment in which volunteers want to engage and improve the projects.

This presents a challenge with respect to the EU DSA SRAM process: the DSA expects greater intervention from platform operators (“accountability” / “responsibilization” of platforms). This is a sensible objective for more conventional, for-profit, social media platforms, whose content policies and moderation are imposed on a top-down basis. However, when it comes to the Wikimedia projects, regulatory obligations should not be interpreted in a way that reduces the Wikimedia communities’ autonomy, enthusiasm, and control.

Accordingly, in stark contrast to other commercial platforms, the Foundation intends to generally refrain from dictating changes to content policy, or displacing effective community mechanisms (e.g. for complaint handling, or efforts to tackle certain categories of problematic content). Our focus is instead on creating the right conditions for success. This is achieved by continuing the Foundation’s catalytic, framework role, for instance by creating useful tools, supporting community structures (such as key committees), offering training, fostering

discussion, offering guidance and support, and engaging in the co-creation of new policies and systems, such as the Universal Code of Conduct and its enforcement processes.

The focus is therefore on ensuring that the Wikipedia communities have the means to tackle these important problems—rather than usurping their leading role.

How are relevant risks selected?

Our starting point is the Foundation’s global, non-DSA-specific view of the risk landscape for Wikimedia projects. In particular, our DSA exercise builds on the earlier results of a detailed, third-party-led, enterprise-wide [Human Rights Impact Assessment \(HRIA\)](#), last conducted in 2020, which laid out a detailed analysis of risks relating to the Wikimedia projects generally, and how they could be better mitigated. This is now being complemented by follow-up work, such as a Child Rights Impact Assessment, assessments related to the use and deployment of specific technologies such as machine learning, and the integration of routine due diligence processes into work streams across the organization.

The DSA is clear that the Foundation’s global, non-DSA-specific work should not be used as-is for this specific exercise. For one thing, the HRIA (and related work) are not EU-specific. For another, not all risks in the HRIA have the EU-wide scope and scale to be “systemic”, as the DSA requires. And not all of them are also Wikipedia-specific.¹ For example, the use of algorithms to recommend content on commercial platforms has resulted in significant concern among the general public around the privacy of users’ data and how that data feeds into those algorithms, the internal workings of which are rarely visible to the public. On Wikipedia, however, the few algorithms used to recommend encyclopedic articles do not rely heavily on individuals’ user data because Wikipedia only collects the minimal amount of user data feasible. Furthermore, the internal workings of such algorithms are documented transparently using model cards for the public to scrutinize.

These differences represent an important distinction between how commercial platforms and Wikipedia use these technologies – a difference that is often underappreciated by the general public. We are therefore being careful to assess the seriousness of each risk identified in our risk register based on two factors: 1) how that risk could manifest on Wikipedia given how the platform actually operates; and 2) the likelihood of that risk arising within the EU.

¹ Of all the Wikimedia projects, only Wikipedia is in-scope for the DSA SRAM exercise—it is the only project large enough to have been designated as a Very Large Online Platform (VLOP); the other Wikimedia projects will be subject to several other DSA obligations, but not SRAM.

Based on the wording and structure of the law, our first step for this SRAM exercise was therefore to select risks by taking the global risks that were highlighted in our HRIA (and related work), and then applying DSA-specific filters:



Figure 3: Diagram showing how the Wikimedia Foundation *selects risks for inclusion* in the DSA SRAM process.

Based on the four factors shown in Figure 3, the global risks and their assessments are—or are not—transposed (modified, if appropriate) into the DSA SRAM Register. For the most part, the Foundation’s broader assessments of global risks, e.g. the analysis contained in our HRIA, will remain somewhat relevant to the risks transposed into the DSA SRAM Register. Those documents therefore provide important further reading for those with an interest in the various risks logged in the DSA SRAM Register.

What risks are excluded by this approach?

DSA-specific filtering means that important global risks, which are the focus of extensive Foundation and wider community activity, *are not reflected in the DSA SRAM Register*.

For example, country-wide blocking of the Wikimedia projects is a major risk for affected populations and for the Wikimedia projects themselves, which then have reduced participation from affected citizens. We do not currently consider this an EU systemic risk, nor is it clearly linked to Wikipedia’s design/functioning/use, so it has not been included in this DSA-specific exercise, despite its wider importance.

In addition, it should be noted that Wikipedia was designated as a VLOP based on its number of monthly active visitors from the EU. However, certain of the risks included in our DSA SRAM exercise are likely to be experienced only by actively-engaged volunteers, such as those making and discussing changes to Wikipedia articles. This is a much smaller number of people—far below the 45 million user VLOP threshold. By way of example, English-language Wikipedia only sees (approximately) 59,000 active editors a month; of these, *at least* 34,000 come from non-EU countries.² For French Wikipedia, there are only approximately 5,700 active editors from France, and another 420 from Belgium. Accordingly, even though the Foundation is focused on tackling risks such as volunteer harassment, globally, it was unclear to us to what extent that should be selected as a DSA-specific EU systemic risk. Future iterations of the EU DSA SRAM exercise may omit it, even though it was included for 2023. Similar considerations apply for risks that might concern an even smaller population of users, such as active EU editors aged under 18.

How are the selected risks assessed?

Starting from our existing (wider) assessment of risks in exercises such as our HRIA, we then conduct a second pass assessment from a DSA-specific perspective.

Our primary consideration is (1) the likelihood of a risk manifesting, and (2) the impact a risk is likely to have, if it manifests. Among other factors, this takes into account the mitigations which are already in place.

The distribution of Child Sexual Abuse Material (CSAM), for instance, can be damaging, but its prevalence on Wikipedia (as an EU systemic risk) appears to be low. Unlike other websites, the openness of Wikipedia, i.e. that all contributions and edits are logged publicly, and volunteer editors can identify and remove such content, makes it a highly unattractive platform for distributing such material. That transparency similarly makes Wikipedia a relatively unattractive environment for grooming, or other types of conduct that thrive in obscurity.

Finally, our risk prioritization (including the ability to devote resources to a selected risk's mitigation, for DSA purposes) also takes into account the Foundation and wider editor and reader communities' global (non-DSA-specific) priorities. As noted above, the Wikimedia projects, globally, are exposed to important risks, such as mass censorship, or impediments to

²

[https://stats.wikimedia.org/#/en.wikipedia.org/contributing/active-editors-by-country/normal|table|last-month|\(activity-level\)~5..99-edits|monthly](https://stats.wikimedia.org/#/en.wikipedia.org/contributing/active-editors-by-country/normal|table|last-month|(activity-level)~5..99-edits|monthly)

global knowledge equity, that demand significant prioritization outside of this EU, DSA-specific exercise.

For all these reasons, it is important to also note that the results of the Foundation’s 2023 DSA SRAM exercise, including this inaugural Systemic Risk Assessment, cannot be compared on a like-for-like basis with those produced by other platforms and search engines that have been designated as VLOPs and VLOSEs. For instance, the number of EU-based users exposed to some of these risks on large social media platforms could be orders of magnitude greater than on Wikipedia.

What are the key risks highlighted in this 2023 exercise?

Our assessment for 2023 identifies three immediate priorities:

1. Disinformation regarding civic and electoral processes, and conflicts
2. Disinformation regarding historical/geographical narratives
3. Harassment among the volunteer community

Work on these is already underway across the whole communities of Wikipedia contributors and allies. More generally, Wikipedia is a broadly welcoming environment for contributors that results in reliable, neutrally-presented information on a vast range of topics. This is the result of longstanding efforts by the community, the Foundation, and others. But as is clear not just from our HRIA, but also community conversations and outside research, more can be done.

Although the full range of mitigations is yet to be fully documented and assessed for 2023, it is possible to preview the main ongoing mitigations for each of these risks:

1. **Anti-harassment efforts** are being supported by rollout of the [Universal Code of Conduct](#) and its related guidance and enforcement processes; and the Foundation is working on [a new, supplemental way for UCoC-violating incidents to be reported](#);
2. **Anti-disinformation efforts** are sufficiently diverse that the Foundation has developed an overarching Anti-Disinformation Strategy. Specific disinformation risks—for example ahead of certain elections—are responded to by supporting the formation of temporary task forces (by the volunteer community, or the Foundation, or both). This approach is backed by a program of work to improve technical tools to support moderation and content curation carried out by volunteer editors, and a [wider research program to support Knowledge Integrity](#) on Wikipedia and its sister projects. There are

also plans to develop an e-learning module to assist volunteer editors in identifying and combatting disinformation.

Other risks that were assessed and are being mitigated, as part of this exercise, include the following. More details on these risks and mitigations can be found in the Risk Register itself, and the underlying materials it is based on.

Medium-Term Priorities:

1. Disinformation regarding scientific information and conspiracy theories
2. Dangerous content: Exposure of vulnerable individuals to content that may be dangerous in that context
3. Child sexual abuse material (CSAM)
4. Terrorist or violent extremist content (TVEC)

Long-Term Priorities:

5. Over-reliance on moderation tech: Possible detrimental side-effects of the increased use of technological tools for content moderation on Wikipedia, in particular (i) algorithmic propagation of bias, and/or (ii) over-reliance on tools, that could become detrimental to volunteer skills and participation
6. Propagation of disinformation or bias through generative AI
7. Privacy risks for young contributors
8. Attacks on individuals profiled on Wikipedia

Conclusion

The Wikimedia Foundation appreciates the opportunity to discuss these risks with the European Commission, members of the volunteer community that make free knowledge projects like Wikipedia possible, and, indeed, the broader public. The Foundation looks forward to engaging in sustained dialogue with these stakeholders around these risks, how it is already working to mitigate them now and in the future, and how our work to identify and mitigate risks will continue in the coming years as the digital ecosystem evolves.

This process of systemic risk assessment has served as a timely incentive for the Foundation to take stock of the broader societal impacts of Wikipedia within the EU. We welcome the opportunity to describe to the Commission, and the EU citizens it represents, how the organization plans to mitigate risks that the DSA is intended to address. Indeed, we welcome efforts by governments to work with platforms on processes that foster greater transparency and accountability for online platforms.

As discussed above, many important actions to mitigate the risks identified in this submission will require buy-in and action from Wikimedia affiliates and volunteers throughout the EU. We hope our publication of material about the risks and mitigations will ultimately spark important discussions with the Wikimedia communities about their important and essential roles in mitigating risks to themselves and European society more broadly, thereby empowering Wikimedians across Europe to exercise leadership and initiative in our collective endeavor.

*