The Vulnerabilities of Trusted Notifier-Models in Russia: The Case of Netoscope

Sivetc, Liudmila; Wijermars, Mariëlle

Veröffentlichungsversion / Published Version
Zeitschriftenartikel / journal article

Empfohlene Zitierung / Suggested Citation:
Sivetc, L., & Wijermars, M. (2021). The Vulnerabilities of Trusted Notifier-Models in Russia: The Case of Netoscope. *Media and Communication*, *9*(4), 27-38. https://doi.org/10.17645/mac.v9i4.4237

Media and Communication (ISSN: 2183–2439)
2021, Volume 9, Issue 4, Pages 27–38
https://doi.org/10.17645/mac.v9i4.4237

Article

The Vulnerabilities of Trusted Notifier-Models in Russia: The Case of Netoscope

Liudmila Sivetc [1,*] and Mariëlle Wijermars [2]

[1] Faculty of Law, University of Turku, Finland; E-Mail: liusiv@utu.fi
[2] Faculty of Arts and Social Sciences, Maastricht University, The Netherlands; E-Mail: m.wijermars@maastrichtuniversity.nl

* Corresponding author

Submitted: 28 February 2021 | Accepted: 25 June 2021 | Published: 21 October 2021

Abstract
Current digital ecosystems are shaped by platformisation, algorithmic recommender systems, and news personalisation. These (algorithmic) infrastructures influence online news dissemination and therefore necessitate a reconceptualisation of how online media control is or may be exercised in states with restricted media freedom. Indeed, the degree of media plurality and journalistic independence becomes irrelevant when reporting is available but difficult to access; for example, if the websites of media outlets are not indexed or recommended by the search engines, news aggregators, or social media platforms that function as algorithmic gatekeepers. Research approaches to media control need to be broadened because authoritarian governments are increasingly adopting policies that govern the internet *through* its infrastructure; the power they leverage against private infrastructure owners yields more effective—and less easily perceptible—control over online content dissemination. Zooming in on the use of trusted notifier-models to counter online harms in Russia, we examine the Netoscope project (a database of Russian domain names suspected of malware, botnet, or phishing activities) in which federal censor Roskomnadzor cooperates with, e.g., Yandex (that downranks listed domains in search results), Kaspersky, and foreign partners. Based on publicly available reports, media coverage, and semi-structured interviews, the article analyses the degree of influence, control, and oversight of Netoscope's participating partners over the database and its applications. We argue that, in the absence of effective legal safeguards and transparency requirements, the politicised nature of internet infrastructure makes the trusted notifier-model vulnerable to abuse in authoritarian states.

Keywords
authoritarian states; internet governance; internet sovereignty; news personalisation; Netoscope project; platformisation; Roskomnadzor; Russia; trusted notifier-model

Issue
This article is part of the issue "Media Control Revisited: Challenges, Bottom-Up Resistance and Agency in the Digital Age" edited by Olga Dovbysh (University of Helsinki, Finland) and Esther Somfalvy (Research Centre for East European Studies at the University of Bremen, Germany).

1. Introduction

Current digital ecosystems are shaped by platformisation, algorithmic recommender systems, and—increasingly—news personalisation (van Dijck, 2020). These (algorithmic) infrastructures influence the online dissemination of news and therefore necessitate a reconceptualisation of how online media control is or can be exercised in states with restricted media freedom. Indeed, the degree of media plurality and journalistic independence becomes irrelevant when reporting is available but difficult to access; for example, if the websites of media outlets are not indexed or recommended by the search engines, news aggregators, or social media platforms that function as algorithmic gatekeepers (Napoli, 2015). This is all the more important since authoritarian governments increasingly adopt policies that govern the internet *through* its infrastructure

(Sivetc, 2021) and use the power of private infrastructure owners to achieve effective, but less easily perceptible, control over online content dissemination. For example, research indicates that, in Russia, Yandex's search engine and news aggregator demonstrate a bias and "referred users to significantly fewer websites that contained information" about protests (Kravets & Toepfl, 2021, p. 1). Zooming in on a concrete case where a trusted notifier-model (Schwemer, 2019) has been employed to counter online harm in Russia, we argue that, in absence of effective legal safeguards, accountability mechanisms and transparency requirements, the politicised nature of internet infrastructures makes this model vulnerable to abuse in authoritarian states.

Governments increasingly seek to control their "national" digital spaces by introducing online content regulations and expanding their influence over critical internet resources, such as Domain Name Systems (DNS; Mueller, 2010). When critical internet resources belong to private companies or (non-profit) organisations, governments, therefore, seek to cooperate with or co-opt them to decide inter alia on the accessibility of online content (Balkin, 2014). For instance, establishing control over the national DNS infrastructure enables one to control connectivity among internet users: The DNS system, similar to a telephone book, connects names (URLs) with corresponding numbers (IP address where the resource is hosted) and therefore serves as "a necessary prelude to communication" (Klein, 2002, p. 195). Since country-code top-level domains (ccTLDs) such as .ru, are governed by relevant authorities at the national level, studying the relations between national governments, private parties, and not-for-profit organisations in this sphere is important as they determine the availability of online content (Schwemer, 2018). Moreover, national domain name registries, government bodies, and various private or public partners can be involved in online content control by creating "trusted notifier-models" for flagging suspicious domain names (Schwemer, 2019, p. 3).

Russia, a country in which media freedom is significantly restricted, actively seeks to expand its control over internet infrastructure and thereby strengthen its capacity to censor online content (Sivetc, 2021; Wijermars, 2021). Under the 2019 Russian Internet Sovereignty Act, Russia became one of the stakeholders of its national registry (the Coordination Center for top-level domains .ru and .рф) in June 2020 (Coordination Center, 2020). Responsible for, among other tasks, the allocation and deallocation of domain names, the Coordination Center occupies a powerful position which may be a valuable asset in its cooperation with other stakeholders, including the Russian state. This article analyses to what extent a governance model which relies on trusted notifiers, and in which the Russian internet regulator Roskomnadzor (Federal Service for Supervision of Communications, Information Technology, and Mass Media) cooperates with the Coordination Center and other key internet infrastructure owners, could be used for alternate ends (by the regulator or other actors). Our argument builds upon an examination of the Netoscope project (a database of .ru domain names suspected of malware, botnet, or phishing activities). The project was launched by the Coordination Center in 2012 and, by 2021, involved 17 partners, including Roskomnadzor, who contribute to Netoscope's database of harmful domain names, thereby affecting their reputation and, potentially, their algorithmic ranking.

Previous studies of Russian internet governance concerning media control have focused on federal legislation, media ownership structures, censorship, and surveillance (Litvinenko & Toepfl, 2019; Lokot, 2018; Sherstoboeva, 2020; Vendil Pallin, 2017). The "infrastructural turn" in internet governance scholarship (Musiani et al., 2016) has only recently started to be addressed with regard to Russia (Daucé & Musiani, 2021). Control mechanisms that function through infrastructures and the governance models involved have yet to be substantially investigated. This is of particular relevance as Russia seeks to create a "sovereign" internet whose successful realisation relies on state control over the Russian internet's infrastructure, including the creation of a national DNS (Stadnik, 2021).

To address this gap and demonstrate the need to complement existing approaches to studying media freedom with research into the governance of the algorithmic and physical infrastructures that shape online news dissemination, we examine the relations between the Coordination Center and the various partners it collaborates with within Netoscope. Based on publicly available reports, media coverage, and semi-structured interviews with representatives from the Coordination Center, Kaspersky (national partner), and SURFnet (international partner), we seek to understand the nature and dynamics of the trusted notifier-model which underlies the partnership and to explore the extent to which various Netoscope partners can influence, control, and have insight into the database and its applications. We interpret the implications of the governance structures we uncover and argue that, as a result of limited transparency, this governance model may be vulnerable to manipulation or abuse towards media control or other restrictive objectives.

2. The Place of Trusted Notifier Systems in Internet Governance

The introduction of state regulation of online content is by now a common trend across political systems, illustrating a gradual shift in the balance away from the multistakeholder approach, long held to be inherent to internet governance, towards more state-centred tendencies. The multistakeholder approach emphasises the global nature and complex interdependence of the internet; its governance therefore should involve not only states but also businesses, civil society, and communities of technical experts (Dutton, 2015). In contrast, the

state-centred approach to internet governance, referred to as internet balkanisation (Hill, 2012), fragmentation (Drake et al., 2016), or sovereignisation (Möllers, 2021) focuses on state regulation or "self-determination" with regard to local internet arenas. Notwithstanding the push towards sovereignisation, self-regulatory models continue to be prevalent, for example in efforts to limit the dissemination of illegal and harmful content on social media platforms.

In this context, trusted notifier-models have emerged as a way to disable access to illegal online content on the basis of notices sent by "trusted flaggers" or "trusted notifiers" (Schwemer, 2019, p. 2). This expertise can come from individuals, private organisations, civil society organisations, semi-public bodies, and public authorities (Schwemer, 2019, p. 3). For example, the trusted notifier-model is supported by the European Commission as it encourages platforms to collaborate with public authorities and trusted notifiers to take down illegal content (European Commission, 2017). Although trusted notifiers can act in different contexts (from flagging terrorist speech to identifying copyright infringements) several general features of such governance models can be identified (Schwemer, 2019): (1) Trusted-notifier models emerge as voluntary arrangements; (2) trusted notifiers act as privileged parties with a direct channel to the intermediary that has the capacity to affect the accessibility of flagged content; (3) there is no requirement of preliminary judiciary assessment of content flagged by trusted notifiers; and (4) as a form of privatised enforcement, the model suffers from a democratic deficit and can be challenged from the perspective of the rule of law, legal certainty, accountability, right to due process, as well as freedom of expression. In the context of initiatives aimed at countering disinformation, for example, outsourcing decisions on politically contentious issues to trusted notifiers may result in overcensoring with limited or no opportunity for redress.

Various international examples exist of the creation of public-private partnerships with the specific aim of countering malware and botnets, similar to the case under examination in this article (Dupont, 2017). Examining such anti-botnet initiatives launched between 2005 and 2010 in Australia, Japan, South Korea, Germany, and the Netherlands, Dupont (2017) explains that they centre around the engagement of internet service providers (ISPs) and anti-virus companies, typically encompassing private entities who are each other's direct competitors and are "often implemented by public Internet regulatory agencies attached to economic development and telecommunications ministries" (Dupont, 2017, p. 109). At their core is the establishment of information-sharing systems between telecommunications regulatory agencies and ISPs to aggregate data on botnets and identify infected devices. In South Korea, the Netherlands, and the United States, ISPs are known to place infected machines, whose users are "unable or unwilling to rectify the situation" in a "digital quarantine" by disrupting their internet access until the infection has been addressed (Dupont, 2017, p. 109); as a form of private enforcement, such practices give rise to legal and ethical concerns.

3. Russian Internet Governance and Media Control

Up until 2012, the Russian state demonstrated a relatively hands-off approach regarding internet regulation. Rather than employing filtering, restricting internet access, or blocking online content, the online domain was governed through more subtle means as Russia sought to shape online discourses "through effective counterinformation campaigns that overwhelm, discredit, or demoralize opponents" (Deibert & Rohozinski, 2010, p. 27). Therefore, the internet was able to function as a counterweight to the increasingly restricted traditional media (federal television, newspapers) and flourish as a platform for independent journalism and political activism (Wijermars & Lehtisaari, 2020). Russia had already taken several "preparatory steps" by enhancing state ownership of internet companies, attaching the status of mass media (and thereby the restrictions applicable to them) to their online counterparts, and floating the first proposals to establish a "national firewall" (Lonkila et al., 2020).

Since 2012, Russia intensified internet control, for example, by introducing website blocking legislation. Roskomnadzor was established in 2008 to regulate mass media and telecommunications and issue licences, and has since played a central role in website blocking procedures (Sivetc, 2020). In June 2020, the European Court of Human Rights criticised this practice when it ruled in two separate cases (*Kharitonov v. Russia*, 2020; *OOO Flavus and others v. Russia*, 2020) that Russia's website blocking legislation violates Article 10 of the European Convention on Human Rights. The Court found that the legal framework for website blocking jeopardises freedom of expression. It grants Roskomnadzor the ability to, without preliminary court oversight, block access not only to the allegedly unlawful content but also the entire website on which any such content is published (in these cases, e.g., grani.ru, an oppositional online media outlet). Moreover, implementation procedures affect innocent websites hosted on the same server as the targeted website (on overblocking and the ban of messenger Telegram, see Ermoshina & Musiani, 2021). Roskomnadzor's prerogatives in restricting access to online content without preliminary court oversight are expanding. The federal agency also partakes in extralegal internet governance practices, as is the case in the example we examine.

The technical obstacles Roskomnadzor encountered in putting in place effective website blocking (Ermoshina & Musiani, 2021; Stadnik, 2021) have led to the restructuring of Russian internet governance through the Russian Internet Sovereignty Act (2019). This law transferred the implementation of website blocking from

ISPs to the state. Through the obligatory placement of devices equipped with deep packet inspection technologies, Roskomnadzor was empowered to directly and more accurately filter and block websites, which should limit overblocking. However, lessening the dependence on ISPs and using state-controlled deep packet inspection filters may turn the website blocking mechanisms into a black box that is non-transparent to public and providers' scrutiny (Stadnik, 2021).

In addition to controlling online speech through legislative measures, the Russian government has co-opted internet gatekeepers to use their private rules to affect online content (Daucé & Loveluck, 2021). Their efforts to control which news items and sources are recommended by news aggregators, resulting in the law "On News Aggregators" (Wijermars, 2021), clearly indicate that the authorities are aware of the centrality of platforms and algorithmic infrastructures in online news dissemination. Empirical research suggests that Yandex's search engine and news aggregator indeed "forwar[d] users to fewer websites that regularly featured criticism of Russia's authoritarian leadership" (Kravets & Toepfl, 2021, p. 1). Yet, within scholarship on media freedom in Russia, the role of these intermediaries and governmental efforts to control them has received limited scrutiny. While for many Russian technological companies, their degree of independence vis-a-vis the Russian state has been (rightfully) questioned, the emergence of trusted notifier-models (as exemplified by Netoscope) within Russian internet governance and its possible implications necessitates further scrutiny as both part of and separate from the general sovereignisation trend.

4. Methodology

To gain insight into Netoscope and its governance structure we triangulated multiple sources. First, we analysed Coordination Center reports (2013–2020) that contain a section dedicated to counteracting illegal activities that use domains .ru/.рф, providing concise, general information about Netoscope and its main achievements. Second, we examined media coverage using the INTEGRUM Profi database, which provided additional information on the development of Netoscope, its partners, and the applications of the database. We queried the database with the Russian project name (*Нетоскоп*) for the period 1 January 2011–30 September 2020. Upon manually assessing relevance and removing duplicates, this resulted in 48 unique results. Media coverage was most frequent in 2013 (11 unique results) when the project's first results were published, and 2018 (10 results) in connection to the project's collaboration with FIFA. A substantial number (16) concerned publications by IT websites and magazines. Overall, media coverage can be characterised as being limited in frequency and largely guided by press releases.

Third, we conducted semi-structured interviews with Netoscope partners; all partners were invited, yet only three accepted the invitation. We interviewed a representative from the Coordination Center, who requested anonymity; Andrey Yarnykh, the Director of the Strategic Development Project of Kaspersky in Russia; and Roland van Rijswijk-Deij, who was employed as a researcher at SURFnet at the time when their agreement with Netoscope was signed and involved in the coordination of the collaboration. Each interviewee was asked to answer the same set of pre-prepared questions. All interviews were conducted online, in January and August 2020. This interview guide included several groups of questions: general questions regarding Netoscope; questions related to the motivations for joining and the role of the interviewee's organisation or company in Netoscope; questions about the relations among the project partners; questions about the Netoscope database (e.g., whether the interviewee's organisation contributes to the Netoscope database, uses it, has access to and control over it, ability to see which partner has flagged a certain domain name, whether it has any verification or safeguard mechanisms to prevent or remedy mistakes); and finally, a question concerning Roskomnadzor's participation in Netoscope. At the end of the interview, interviewees were invited to add anything else they would like to share regarding Netoscope. Since SURFnet's involvement in Netoscope is limited, this interview generated much less information and correspondingly features less prominently in our analysis. All translations were carried out by the authors.

The fact that many project partners declined our interview request presents a clear limitation to our study; yet, this is a condition that is commonly shared by research in this area focusing on Russia. For example, both Yandex, Russia's leading technology company, and Roskomnadzor are notoriously closed to information requests from researchers. Since the conducted interviews present three distinct perspectives (the Coordination Center, a Russian partner, and an international partner) we are nonetheless able to present a sufficiently comprehensive picture of how Netoscope functions. In interpreting these interviews, one also has to consider that, given the politicisation of internet infrastructure in Russia, interviewees may present an incomplete/one-sided view of the situation. Therefore, we compared and complemented findings with data from Coordination Center reports and media coverage whenever possible (again, taking into consideration the limitations in the availability and reliability of the latter sources). In the next section, we first present the insights gathered from these sources to tell a coherent story about the development and functioning of Netoscope. Since the way in which interviewees narrativise their positions is an important source for understanding the project, these statements are presented comprehensively. A critical discussion of the picture emerging from our sources then follows in Section 5.3.

5. Netoscope

5.1. History and Functionality of Netoscope

Netoscope was launched in 2012 by the Coordination Center. As stated on its official website, the project "aims at making the Russian domain space safer for users" (Coordination Center, 2012). In our interview, the representative of the Coordination Center, who is directly involved in the functioning of Netoscope, explained that the project was not intended for the regulation of the Russian internet; rather, it was deemed necessary for improving the reputation of the Russian top-level domains, since they did not rank among the safest domains in 2009–2011. Although this low ranking, according to the representative, lacked a proper justification, they admitted the validity of some of the security concerns; the .ru domain was indeed used for malicious activities, such as malware and the creation and operation of botnets. In the representative's view, these malicious activities may be explained by the low prices of domain name registration and the (according to them, incorrect) impression that the Coordination Center was indifferent to activities in the Russian ccTLDs. On the contrary, the representative emphasised that the Coordination Center was very much interested in making the domain safe for internet users but the issue was that the Coordination Center lacked the necessary competencies to identify domain names involved in malicious activities. Therefore, it proposed Netoscope as a platform for cooperating with cybersecurity experts.

Cybersecurity experts, in turn, needed the cooperation with the Coordination Center because only they are able to terminate the domain name delegation of resources involved in the "epidemic" dissemination of, for example, malware, as Andrey Yarnykh, the Director of Strategic Development Project of Kaspersky in Russia, indicated in the interview. The termination of the domain delegation does not cancel the registration of a domain name; it terminates the connectivity between the domain name and the corresponding address, which makes the respective website inaccessible until the delegation is restored. Experts employed by Kaspersky, he indicated, can detect malware being spread by such resources and identify which domain names are used for coordinating command points. To prevent such epidemics from developing, the resources behind them should be disabled directly by terminating the delegation of the domain names involved. Therefore, Kaspersky had an interest in being able to inform the Coordination Center on domain names engaged in malicious activities and request the termination of their delegation. According to Yarnykh, Netoscope provided the necessary mechanisms for that purpose and Kaspersky sends information on malicious domain names to the project to enable the Coordination Center to expeditiously react to cyberthreats. Here, his account differs from that provided by the Coordination Center representative, who pointed out that Netoscope is only the basis for technical and scientific collaboration. The termination of domain name delegation, which indeed lies within the mandate of the Coordination Center, is realised through a separate trusted-notifier mechanism that is more formalised and transparent in its procedure and in which Kaspersky and other Russian Netoscope partners are authorised to request undelegation.

According to Yarnykh, Netoscope effectively combats the viral spread of malware, botnets, and phishing by disabling coordinating command points, which decreases the levels of malicious activities in the Russian ccTLDs as well as globally. The Coordination Center representative also indicated that the cooperation within Netoscope has led to a decrease in the number of malicious activities in the Russian ccTLDs and thereby improved their reputation. If, in the beginning, Netoscope flagged a hundred thousand malicious domains per year, by 2020, the numbers had decreased significantly and the domain became "cleaner" (measuring the impact of such partnerships is, however, difficult; Dupont, 2017).

Yarnykh highlighted that Kaspersky does not gain commercial benefit from participating in Netoscope but acts as a "donor." The company's interest, he said, consists solely in contributing to stable internet development. To this aim, the company cooperates with Netoscope partners to make the Russian ccTLDs "cleaner and more protected." Kaspersky cooperates with partners involved in Netoscope outside of the project as well, but these processes are conducted "in different formats" than those within Netoscope.

The Coordination Center representative explained that Netoscope was created upon several meetings with experts. Some of them had shown interest in cooperating, others were specially invited by the Coordination Center. Initially, the representative indicated, Netoscope involved such partners as RU-CERT, Kaspersky, Group IB, and the Technical Center "Internet"; i.e., Russian cybersecurity companies. The Coordination Center's 2012 report indicates that Yandex, which can be considered the Russian counterpart and competitor of Google, providing a broad array of digital services, including internet browser, search engine and news aggregation, joined in 2012 (Coordination Center, 2013, p. 11). Gradually, additional companies and organisations also joined the project, including three foreign partners: IThreat Cyber Group (United States), SURFnet (the Netherlands), and FIFA.

Table 1 presents an overview of the 17 project partners listed on the website and indicates their main areas of activity. It shows that Netoscope differs from the anti-botnet public-private partnerships described by Dupont (2017) in several respects. First, while cybersecurity companies make up a substantive proportion of partners, the central role of ISPs Dupont identified is lacking. The only partner involved in providing internet services is Rostelecom. However, its membership may be explained by its involvement in the creation of the

Table 1. Netoscope partners.

Partner	Organisation type	Role within Netoscope
Coordination Center	Russian domain name registry organisation	Coordinator
Technical Center "Internet" (TCI)	Russian organisation maintaining the main registry for ccTLDs .ru, .рф and .su	Participant
Roskomnadzor	Russian federal executive authority for media and internet regulation	Participant
National Computer Incident Response and Coordination Center	Russian Computer Emergency Response Team responsible for the protection of governmental networks of the Russian Federation	Participant
RU-CERT	Russian autonomous non-profit organisation. Computer Emergency Response Team	Participant
Group IB	Russian private cybersecurity company	Participant
Kaspersky	Russian private cybersecurity company	Participant
SkyDNS	Russian private cybersecurity company	Participant
Dr. Web	Russian private cybersecurity company	Participant
BI-ZONE	Russian private cybersecurity company. Daughter company of Sber (previously, Sberbank)	Participant
MasterCard Members' Association	Russian non-profit organisation	Participant
Rostelecom	Russian private telecommunications company. Market leader in provision of (mobile) internet services	Participant
Yandex	Russian multinational corporation offering a wide array of digital services. Owner of Yandex browser and search engine	Participant
Mail.ru Group	Russian corporation active in email services, e-commerce, B2B, media, instant messaging. Owner of VKontakte and Odnoklassniki	Participant
IThreat	American private cybersecurity company	Participant
SURFnet	Cooperative association of Dutch educational and research institutions aimed at the development and procurement of information and communication technology facilities and knowledge sharing	Participant
FIFA	French non-profit organisation. Organiser of the FIFA World Cup	Participant

Russian browser and search engine Sputnik, launched in 2014, which filtered various harmful materials from its search results through its collaboration with Kaspersky, Netoscope, and Roskomnadzor ("'Sputnik' iskliuchaet iz poiskovoi," 2014). Because of its limited success, the Sputnik search engine was discontinued in 2020, yet the company continues to provide search solutions to corporate and government clients ("Poiskovik 'Sputnik' prekratil," 2020). Second, it includes two key players of the Russian internet: Yandex, previously introduced, and Mail.ru Group, which (among many other activities) is the owner of the popular social media platforms Vkontakte and Odnoklassniki and a (much less popular) search engine and news aggregator. Rambler Media Group, another prominent digital media company owned by Sber (a state-owned bank), is not included. Finally, there are three non-Russian partners, whose partnership appears to be motivated differently, as will be discussed below.

According to the Coordination Center's 2016 report, Roskomnadzor joined Netoscope on 19 April 2016 (Coordination Center, 2017, p. 12). Roskomnadzor and Netoscope concluded an agreement on cooperation aimed inter alia at "the joint investigation of content, types, and features of unlawful online information and the development of means of precluding it from dissemination on the Internet" (Coordination Center, 2016, p. 2), a formulation which suggests a scope that extends beyond botnets and malware. Despite becoming an official partner in 2016, Roskomnadzor, as the Coordination Center representative clarified, was involved in Netoscope from the very beginning and was an active participant both before and after concluding the agreement. Their cooperation practices were not affected by the changed status, the representative stated: "[T]here have not been any cardinal changes. Instead, there has been active, annual, everyday, on-time performance." Andrey Yarnykh also indicated that the practices

of cooperation within Netoscope did not change when Roskomnadzor joined; at least, Kaspersky did not notice any changes. The company continues to send information to the database in accordance with its own expertise: phishing, spam, and malware. Yarnykh assumes that Roskomnadzor, just as other partners, contributes to the project within the agency's expertise, in a way that benefits Netoscope's overall objective.

According to the Coordination Center representative, experts contribute to Netoscope by submitting information on domain names involved in phishing, malware, and botnet activities to a database that accumulates the information and stores all suspicious domain names. This means that once a domain name is included in the Netoscope database, it will never be excluded from it, even when the flagged domain name no longer hosts the malicious content. If the domain name ceases to exist (if its registration in one of the Russian ccTLDs is discontinued) this also does not affect the information stored in the database. These structural characteristics leave the issue of how to interpret the information about a domain's entry into the database up to the user of the database. The principle of permanent storage, the Coordination Center representative explained, is based on the assumption that a domain name that has been used for malicious activities in the past is likely to be used again and therefore retains its dangerous potential. Yet, it means there is no possibility for domains that have been falsely flagged or flagged as a result of manipulation (e.g., a malicious actor simulating an attack and connecting it to the domain of an opposition-related website) to rid themselves of the reputational damage and its possible consequences. The available information also suggests domain name owners are not necessarily informed if they are added to the database.

The Netoscope database serves as the basis for the Domain Checker available on the Netoscope website. Any internet user can use it to find out whether a domain name registered in the .ru, .su, or .рф domains has been flagged by Netoscope. For example, (oppositional) online media outlet grani.ru was blocked in March 2014 on the allegation of publishing calls to participate in unauthorised mass protests. The Domain Checker (Netoscope, 2021) indicates the following result for the domain: "On the domain name grani.ru project partners recorded the following malicious activities: Formerly Malware."

In December 2020, the Netoscope database contained approximately 4,7 million domain names (Netoscope, 2020). The Coordination Center representative explained that this figure should not be understood as an indicator of a high level of malicious activity. Only a small number of these, around five thousand, represent domain names flagged as "currently malicious." In the case of grani.ru, its inclusion in the database indicates that the domain name is outside the scope of currently malicious websites, yet possessed this status at some point in the past. This status should signal to users that the website is safe to access. However, the fact that it was previously flagged by Netoscope may also give rise to questions regarding the website's safety. For example, according to the representative, companies involved in the domain name business adjust their decision to purchase a certain domain name if it has been flagged by Netoscope.

Netoscope has another direct and intended effect: The Coordination Center's 2014 report states that Yandex, the provider of Russia's most popular search engine, has been using the Netoscope database since 2014 to exclude links to malicious websites from its search results (Coordination Center, 2015, p. 11; see also Kudriavtseva, 2020). The Coordination Center representative confirmed that Yandex can use the Netoscope database to adjust how its algorithms decide on which websites are prioritised in search results, yet stressed it is but one of many resources Yandex uses as an input source for its algorithms. Yandex also contributes to the database: According to the representative, the Yandex Safe Browsing database has been used by Netoscope to enrich and refine data about domain names included in the Netoscope database.

5.2. Netoscope as a Trusted Notifier-Model

The Coordination Center representative highlighted an important feature of Netoscope: The project facilitates collaboration among competitors. Most of the partners involved in Netoscope are commercial entities active in adjacent fields; therefore, they prefer not to share information with other (cybersecurity, technology) companies. Yet, as partners in Netoscope, they are willing to share information with the Coordination Center and contribute to the database. According to the representative, the partners cooperate because they share the common goal of making the Russian ccTLDs safer. Moreover, by cooperating, they develop "mechanisms" for identifying malicious activities, which enhances their competencies and thereby their competitiveness in the market. However, each Netoscope partner is unaware of what information the other partners share with Netoscope. As the Coordination Center representative explained, Group IB, for instance, does not know which domain names have been flagged as malicious by Kaspersky: The partners have agreed on this practice because, as competitors, they "do not support the idea that some of them donate the information, while the others only use it without contributing."

Kaspersky's Andrey Yarnykh also mentioned market competition among Netoscope partners as the reason for the fact that there is only unilateral communication between Netoscope and the company. He said that, because Kaspersky's databases with information on malware, phishing, and botnets are used in conducting its projects, this information should be kept secret from competitors. Although Kaspersky sends the information to Netoscope, this information is available only to the project but not to its partners. According to Yarnykh,

"it would be incorrect and wrong if Netoscope presented a resource that shares the information we provided." Rather, "Netoscope was initially designed as a resource to which the partners contribute information but do not take from it"; he also indicates that Netoscope accumulates information but does not disseminate it.

Another aspect affecting information-sharing practices within Netoscope is the different competencies respective Netoscope partners have, which, according to the Coordination Center representative, is noted in the agreement on cooperation. They explained the actual cooperation occurs as follows: The Netoscope database is located at the Coordination Center. Each partner submits information on those domain names that it identifies as being involved in malicious activities to the database. The representative stressed that partners decide whether to flag a domain name, in accordance with their particular expertise. Yarnykh indicated that Netoscope aggregates information sent by the partners and issues reports on the levels of malicious activity. These reports are purposely designed not to reveal the size and content of each partner's contribution to the project. As Yarnykh said, reports provide "statistics rather than analytics." Netoscope does not enable Kaspersky to see which partner flagged a certain domain name.

Importantly, as the Coordination Center representative indicated, Netoscope relies on partners' expertise and does not verify inputs into the database. They explained that such verification falls outside of the Coordination Center's remit and they do not employ experts to perform such verification checks. If a Netoscope partner "says that this domain name is connected with phishing at this moment, it means that the partner answers for [the accuracy of] its words."

According to the Coordination Center representative, Netoscope also relies on the partners' expertise in deciding on notifications about malicious activities received from internet users. Users can inform Netoscope by pressing the button "report malware" on the Netoscope website. Netoscope then sorts out notifications about botnets, phishing, and malware and forwards this information to the relevant partner specialising in identifying the respective malicious activity. Netoscope has received many complaints on malicious activities from users, the representative mentioned, without specifying whether any NGOs or organised groups of internet users are known to submit such notifications (online vigilante groups have in the past played a significant role in flagging online content, thereby initiating website blocking procedures; Daucé et al., 2019).

The Domain Checker available on the Netoscope website warns users about any malicious activity the checked domain name is/was involved in based on Netoscope partners' assessments. In line with the restricted disclosure and anonymised aggregation discussed above, the results received from the Domain Checker do not show which partner flagged the domain name in question nor when this occurred. The Coordination Center representative explained, making information non-traceable was "the main condition at the start of the project." It means that, although the Coordination Center has access to these details, information about partners' involvement is not disclosed, and this lack of transparency extends to all partners in the project. As Yarnykh explained, Kaspersky sends information "like an email" and is not able to trace how it is subsequently processed.

For some of the international partners, the motivations behind joining the project and the content of their contributions appear to be somewhat different. The partnership with FIFA was established in 2018 in the context of the World Cup that Russia hosted. According to FIFA's advisor on brand protection, Aleksei Shvetsov, "FIFA [would] identify and transfer data to Netoscope about domain names used for phishing in the illegal sale of tickets for the World Cup" ("FIFA i 'Netoskop' budut," 2018). The received data would be analysed by "participants of Netoscope" and resources blocked if illegal activities were indeed identified. SURFnet, a cooperative association of Dutch educational and research institutions aimed at the development and procurement of information and communication technology facilities and knowledge sharing, concluded their agreement with Netoscope in 2017. This followed upon initial contact between SURFnet and Technical Center "Internet" at the Internet Engineering Task Force meeting in Berlin in 2016 (interview with Roland van Rijswijk-Deij). SURFnet had an interest in obtaining access to data on the Russian ccTLDs as part of a larger open intelligence project. Following a year-long negotiation process, an agreement to this effect was signed with Netoscope, on condition that SURFnet reports any relevant threats it finds on the basis of the shared data with its Russian partner. According to Roland van Rijswijk-Deij, SURFnet contributed to the Netoscope database on a single occasion (spam detection) and did not receive information on how their notification was handled.

5.3. Discussion: Implications and Possibilities for Misuse

Yarnykh positively assessed the results of Netoscope since the Coordination Center managed to consolidate collaboration among the leading Russian internet companies in the project. Therefore, he considered Netoscope as "a valuable example to also be emulated on an international level, provided the level of trust, responsibility, and coordination is sufficient to use such a cooperation for the sake of internet stability." Yet, from an internet governance perspective, the project also creates a fundamental vulnerability, especially given the current politicisation of internet infrastructure in Russia, that is of relevance beyond our case. Our study shows that the Coordination Center indeed trusts its partners' assessments and does not check whether information sent to the database is correct. On the other hand, Kaspersky (and presumably, other partners) trust the Coordination

Center and cannot trace how the information they provide is processed by Netoscope. The Netoscope database is non-transparent for all but the Coordination Center and it is precisely this condition of non-transparency that served as the basis for establishing and preserving trust within the project. However, the same condition of non-transparency gives rise to concern related to how the database is/may be used by various end-users and the lack of any (legal) redress for domain name owners. Combined with the lack of verification mechanisms (except for domains flagged by internet users) it risks the trust in it being violated by malicious flagging, i.e., an innocent domain name being accused of containing malware by (an employee of) one of the partners or a targeted website being accused of intentional involvement in a (simulated) attack in order for it to be included in the Netoscope database.

In addition to the fact that most partners are either de facto controlled by the state or have had their independence from the state questioned, a particular area of concern is the lack of information on how Roskomnadzor, as the federal agency involved in executing (restrictive) internet regulation, contributes to the project. While there is currently no evidence suggesting that Roskomnadzor uses the Netoscope database to flag unwanted speech as well as malware (which would negatively affect the reputation of the domain name, which could affect its indexation and recommendation) the governance structure of the project, in as far as we were able to confirm, does not have safeguards against such misuse. Within its current scope of competence, Roskomnadzor may then use Netoscope as an implementation tool, instead of, or alongside the other means of enforcement at its disposal (legal action, fines, preemptive website blocking); although, again, their willingness to do so may only be assumed since, as of yet, no proof of its misuse is available. In such a case, using the governance particularities of Netoscope and the competencies of the partners involved (representing leading search engines, news aggregators, and social media platforms) may prove quite effective in extending internet control mechanisms to the level of DNS infrastructure. Similar to other algorithm-driven forms of hidden censorship (Makhortykh & Bastian, 2020), detecting and exposing such misuse is difficult; the lack of transparency and accountability limits possibilities for exposing misuse while trust in the (abused) system is continually reinforced through its usage. Given that Roskomnadzor did not respond to our interview request, information on its role remains limited.

Applying Schwemer's trusted notifier-model to the information we gathered shows Netoscope possesses all of the model's four features: First, Netoscope is based on voluntary arrangements; second, Netoscope partners act as privileged parties with a direct channel, through the database, to the Coordination Center as the intermediary with the capacity to affect the accessibility of flagged content; third, there is no requirement of preliminary judiciary assessment of whether content flagged by the trusted notifiers is indeed illegal (in this case, there is also no safeguard mechanism within Netocope to verify partners' notifications); fourth, Netoscope's non-transparency and the fact that its functioning is not restricted by a clear legal framework challenges the project from the perspective of the rule of law, legal certainty, accountability, right to due process, and freedom of expression. Netoscope appears to function as a black box not only for the public and scholarly community but also for the partners themselves. Since publicly available information suggests that project partners use the Netoscope database as an input for their algorithmic ranking systems, the inclusion of independent news sources may affect their online visibility.

Netoscope's governance structure emerged from the need to create a condition of trust among competitors in order to share data and collectively work towards reducing malware and phishing within the Russian domains. It emerged from the Coordination Center, which, as a technically-oriented non-profit organisation, is influenced by international practices of multistakeholderism in internet governance. Operating through collaboration with security professionals within its partner organisations, their shared understandings of and trust in the reliability of technical expertise provide the basis for the database and its use. However, the introduction of the Russian Internet Sovereignty Act and the (planned) creation of a national DNS are only the most recent signs of a shift from multistakeholderism to a state-centric tendency in Russian internet governance. This politicisation and securitisation of internet infrastructure in Russia mean that the project's neutrality and "technocratic" nature can no longer be assumed. As DeNardis (2014, p. 18) argued some years ago: "Internet governance structures were originally based on familiarity, trust, and expertise and on 'rough consensus and running code.' Things have changed." The fact that the Russian state has become a stakeholder in the Coordination Center is but one indicator of this trend. The lack of transparency—crucial to its involvement of private partners—creates a lack of accountability. Contrary to the procedural requirements and reporting obligations that pertain to, for example, website blocking, a similar degree of transparency is not provided when it comes to the contents and applications of the Netoscope database, which makes it hard to detect whether Netoscope has been used as a tool for online content control. Moreover, those applications of the database that are particularly relevant for indirect media control (algorithmic downranking of flagged domains) are considered company secrets. Recently, transparency concerns have also been expressed regarding website blocking (Stadnik, 2021). As was mentioned above, the Russian Internet Sovereignty Act enables website blocking through state-controlled deep packet inspection filters which may turn it into a black box. In this respect, both cases signal a worrying trend towards rendering

online content governance in Russia less transparent and thereby less accountable.

6. Conclusion

Russia's push towards establishing a "sovereign" internet has garnered international attention in academic, policy, and rights advocacy circles alike. The possible impact of the policy on freedom of expression, among other rights, has been a key concern in these debates, resonating with the earlier concerns about overblocking such as those included in the *Kharitonov v. Russia* (2020) and *OOO Flavus and others v. Russia* (2020) decisions. Scholarship on media control in Russia, however, has yet to fully embrace the importance of internet governance as an enabling or prohibiting factor. Our aim has been to argue for a broadening of how authoritarian control of online media is studied by looking not just at legislation, media ownership, journalistic culture, or self-censorship, but also by critically examining how key technology and internet infrastructure players are involved in internet governance practices that may affect the online dissemination of news and other information. On the example of Netoscope, we argued that the use of a trusted notifier-model, which is currently gaining in popularity as a way to, for example, address online harm within social media, may be vulnerable to manipulation or abuse without effective legal/procedural safeguards and transparency requirements (although, as of yet, there is no evidence of misuse in this particular case). While further research is needed, our findings suggest there are grounds for questioning the general validity of using trust-based models in non-free media systems as they amplify their inherent weaknesses (e.g., limited accountability). To fully grasp the role and impact of such governance practices that exercise control via (physical, algorithmic) internet infrastructures, an analysis of further cases is required. For example, the recent initiative by Yandex to engage selected media and fact-checking organisations as trusted notifiers to counter "fake news" on its personalised content distribution platform (Yandex Zen) illustrates the urgent need to establish an understanding of media control that reflects the complexity of digital ecosystems today. Our analysis of Netoscope underscores the importance of transparency and accountability mechanisms to safeguard against (future) political instrumentalisation of ostensibly technical or specialist collaborations, systems, and governance structures.

Acknowledgments

The authors would like to thank Mykola Makhortykh and two anonymous reviewers for their insightful comments.

Conflict of Interests

The authors declare no conflict of interest.

References

Balkin, J. (2014). Old-school/new-school speech regulation. *Harvard Law Review*, *127*(8), 2296–2342.

Coordination Center. (2012). *About project*. Netoscope. https://netoscope.ru/en/about

Coordination Center. (2013). *Otchet Direktora ANO "Koordinatsionnyi tsentr national'nogo domena seti Internet" A.V. Kolesnikova* [Report by the Director of ANO "Coordination Center of the national domain of the internet" A.V. Kolesnikov]. https://cctld.ru/upload/files/dir_year_report_2012.pdf

Coordination Center. (2015). *Otchet Direktora ANO "Koordinatsionnyi tsentr national'nogo domena seti Internet" A.A. Vorob'eva* [Report by the Director of ANO "Coordination Center of the national domain of the internet" A.A. Vorob'ev]. https://cctld.ru/upload/files/dir_year_report_2014.pdf

Coordination Center. (2016). *Soglashenie o sotrudnichestve v sfere protivodeistviia rasprostraneniiu v seti Internet informatsii, priznannoi zapreshchennoi k rasprostraneniiu na territorii Rossiiskoi Federatsii, i rasprostraniaiushcheisia s narusheniem zakonodatel'stva Rossiiskoi Federatsii* [Agreement on scientific-technical cooperation in the sphere of counteracting the dissemination of information that is illegal to disseminate in the Russian Federation on the Internet]. https://cctld.ru/files/news/rkn_agreement.pdf

Coordination Center. (2017). *Otchet Direktora ANO "Koordinatsionnyi tsentr national'nogo domena seti Internet" A.A. Vorob'eva* [Report by the Director of ANO "Coordination Center of the national domain of the internet" A.A. Vorob'ev]. https://cctld.ru/upload/files/dir_year_report_2016.pdf

Coordination Center. (2020). *About the center*. https://cctld.ru/en/about

Daucé, F., & Loveluck, B. (2021). Codes of conduct for algorithmic news recommendation: The Yandex.News controversy in Russia. *First Monday*, *26*(5). https://doi.org/10.5210/fm.v26i5.11708

Daucé, F., Loveluck, B., Ostromooukhova, B., & Zaytseva, A. (2019). From citizen investigators to cyber patrols: Volunteer internet regulation in Russia. *Laboratorium: Russian Review of Social Research*, *11*(3), 46–70.

Daucé, F., & Musiani, F. (Eds.). (2021). Infrastructure-embedded control, circumvention, and sovereignty in the Russian internet. *First Monday*, *26*(5).

Deibert, R., & Rohozinski, R. (2010). Control and subversion in Russian cyberspace. In R. Deibert, J. Palfrey, R. Rohozinski, & J. Zittrain (Eds.), *Access controlled: The shaping of power, rights, and rule in cyberspace* (pp. 15–34). MIT Press.

DeNardis, L. (2014). *The global war for internet governance*. Yale University Press.

Drake, W., Cerf, V., & Kleinwächter, W. (Eds.). (2016). *Internet fragmentation: An overview*. World Economic Forum.

Dupont, B. (2017). Bots, cops, and corporations: On the limits of enforcement and the promise of polycentric regulation as a way to control large-scale cybercrime. *Crime Law Soc Change*, *67*, 97–116. https://doi.org/10.1007/s10611-016-9649-z

Dutton, W. (2015). *Multistakeholder internet governance?* World Bank. https://pubdocs.worldbank.org/en/591571452529901419/WDR16-BP-Multistakeholder-Dutton.pdf

Ermoshina, K., & Musiani, F. (2021). The Telegram ban: How censorship "made in Russia" faces a global internet. *First Monday*, *26*(5). https://doi.org/10.5210/fm.v26i5.11704

European Commission. (2017, September 28). *Security Union: Commission steps up efforts to tackle illegal content online* [Press Release]. https://ec.europa.eu/commission/presscorner/detail/en/IP_17_3493

FIFA i "Netoskop" budut borot'sia s moshennichestvom pri prodazhe biletov na ChM-2018 [FIFA and "Netoscope" will fight against fraud in ticket sales for the 2018 World Cup]. (2018, February 13). *RIA Novosti*.

Hill, J. (2012). A Balkanized internet? The uncertain future of global internet standards. *Georgetown Journal of International Affairs*, *2012*, 49–58.

Klein, H. (2002). ICANN and internet governance: Leveraging technical coordination to realize global public policy. *The Information Society*, *18*(3), 193–207.

Kravets, D., & Toepfl, F. (2021). Gauging reference and source bias over time: How Russia's partially state-controlled search engine Yandex mediated an anti-regime protest event. *Information, Communication & Society*. Advance online publication. https://doi.org/10.1080/1369118X.2021.1933563

Kudriavtseva, V. (2020, February 26). Kak ne popast' v seti internet-moshennikov? [How not to get caught in the net of internet scammers]. *Telekanal Kul'tura*.

Litvinenko, A., & Toepfl, F. (2019). The "gardening" of an authoritarian public at large: How Russia's ruling elites transformed the country's media landscape after the 2011–2012 protests for fair elections. *Publizistik*, *64*(2), 225–240.

Lokot, T. (2018). Be safe or be seen? How Russian activists negotiate visibility and security in online resistance practices. *Surveillance & Society*, *16*(3), 332–346.

Lonkila, M., Shpakovskaya, L., & Torchinsky, P. (2020). The occupation of Runet? The tightening state regulation of the Russian-language section of the Internet. In M. Wijermars & K. Lehtisaari (Eds.), *Freedom of expression in Russia's new mediasphere* (pp. 17–38). Routledge.

Makhortykh, M., & Bastian, M. (2020). Personalizing the war: Perspectives for the adoption of news recommendation algorithms in the media coverage of the conflict in Eastern Ukraine. *Media, War & Conflict*. Advance online publication. https://doi.org/10.1177/1750635220906254

Möllers, N. (2021). Making digital territory: Cybersecurity, techno-nationalism, and the moral boundaries of the state. *Science, Technology, & Human Values*, *46*(1), 112–138.

Mueller, M. (2010). *Network and states: The global politics of internet governance*. MIT Press.

Musiani, F., Cogburn, D. L., DeNardis, L., & Levinson, N. S. (Eds.). (2016). *The turn to infrastructure in internet governance*. Palgrave Macmillan.

Napoli, P. (2015). Social media and the public interest: Governance of news platforms in the realm of individual and algorithmic gatekeepers. *Telecommunications Policy*, *39*(9), 751–760.

Netoscope. (2020). *Otchet proekta Netoskop za 4 kvartal 2020 goda* [Report of the Netoscope project on Q4 2020]. https://netoscope.ru/upload/stats/Netoscop_2020_4.pdf

Netoscope. (2021). *Domain checker*. https://netoscope.ru/en/check/?q=

OOO Flavus and others v. Russia, Nos. 12468/15, 23489/15, and 19074/16 (2020).

Poiskovik "Sputnik" prekratil rabotu [Search engine "Sputnik" ceased its operations]. (2020, September 8). *RIA Novosti*. https://ria.ru/20200908/sputnik-1576905844.html

Schwemer, S. (2018). On domain registries and unlawful website content: Shifts in intermediaries' role in light of unlawful content or just another brick in the wall? *International Journal of Law and Information Technology*, *26*(4), 273–293. https://doi.org/10.1093/ijlit/eay012

Schwemer, S. (2019). Trusted notifiers and the privatization of online enforcement. *Computer Law & Security Review*, *35*(6). https://doi.org/10.1016/j.clsr.2019.105339

Sherstoboeva, E. (2020). Regulation of online freedom of expression in Russia in the context of the Council of Europe standards. In S. Davydov (Ed.), *Internet in Russia: A study of the RuNet and its impact on social life* (pp. 83–100). Springer.

Sivetc, L. (2020). The blacklisting mechanism: New-school regulation of online expression and its technological challenges. In M. Wijermars & K. Lehtisaari (Eds.), *Freedom of expression in Russia's new mediashere* (pp. 39–56). Routledge.

Sivetc, L. (2021). Controlling free expression "by infrastructure" in the Russian internet: The consequences of RuNet sovereignization. *First Monday*, *26*(5). https://doi.org/10.5210/fm.v26i5.11698

"Sputnik" iskliuchaet iz poiskovoi vydachi saity, popavshie v "chernye spiski" Roskomnadzora ["Sputnik" excludes sites that are included in Roskomnadzor's "black lists" from its search results]. (2014, September 25). *SearchEngines.ru*.

Stadnik, I. (2021). Control by infrastructure: Political ambitions meet technical implementations in RuNet. *First Monday*, *26*(5). https://doi.org/10.5210/fm.v26i5.11693

van Dijck, J. (2020). Seeing the forest for the trees:

Visualizing platformization and its governance. *New Media & Society*, *23*(9), 2801–2819. https://doi.org/10.1177/1461444820940293

Vendil Pallin, C. (2017). Internet control through ownership: The case of Russia. *Post-Soviet Affairs*, *33*(1), 16–33.

Vladimir Kharitonov v. Russia, No. 10795/14 (2020).

Wijermars, M. (2021). Russia's law "On news aggregators": Control the news feed, control the news? *Journalism*. Advance online publication. https://doi.org/10.1177/1464884921990917

Wijermars, M., & Lehtisaari, K. (2020). Introduction: Freedom of expression in Russia's new mediasphere. In M. Wijermars & K. Lehtisaari (Eds.), *Freedom of expression in Russia's new mediashere* (pp. 1–14). Routledge.

About the Authors

Liudmila Sivetc is a lawyer and a doctoral candidate at the Faculty of Law, University of Turku. Her research interest is connected to internet regulation and freedom of expression. Her most recent work has appeared in *First Monday* as part of the Special Issue *Infrastructure-Embedded Control, Circumvention and Sovereignty in the Russian Internet* edited by Daucé and Musiani (2021).

Mariëlle Wijermars (PhD) is an assistant professor in cyber-security and politics at Maastricht University. She conducts research on algorithmic governance, media freedom, and the human rights implications of internet policy. She is co-editor of *The Palgrave Handbook of Digital Russia Studies* and *Freedom of Expression in Russia's New Mediasphere*.