Monitoring of digital rights and freedoms

Guide for recording and monitoring violations of digital rights and freedoms

by SHARE Foundation

Author: Bojan Perkov, Policy Researcher, SHARE Foundation

SHARE Foundation Monitoring coordinator

Introduction

SHARE Foundation is a nonprofit organization established in 2012 in Serbia to advance human rights and freedoms online and promote positive values of an open and decentralized Web, as well as free access to information, knowledge, and technology.

Since 2014, SHARE Foundation has been monitoring the state of digital rights and freedoms in Serbia and so far we have collected more than 400 cases of digital rights violations in our Monitoring database. During this time, we have developed a practice of publishing periodical monitoring reports, which highlight the state of digital rights and freedoms in a given time period (quarterly and annually). We started with this practice in the spring of 2014, when great floods in Serbia caused loss of lives and substantial material damage. During those days, there were widespread cases of suppressing citizens' freedom of expression and silence sources of information which tried to provide their reports of what was happening in the flooded areas of Serbia. The pressures went as far as arresting citizens because of their posts on social media, which could allegedly induce “panic and riots”.

Why monitor cases of violations?

Through active monitoring and reporting on cases of violations of digital rights and freedoms you can follow trends, try to connect the dots and figure out why certain types of violations are happening more often than others. Which websites in your country are mostly being targeted with technical attacks, such as Distributed Denial-of-Service (DDoS)? What are the most common types of pressures bloggers and activists face in the online environment - are they being threatened or is someone, for instance, making fake accounts on social media to misrepresent them? In order to do an analysis and draw conclusions, you need to collect data on as many cases of digital rights violations as you can find, according to your capacities.

The methodology and the database

The SHARE Foundation team has developed a methodology for categorising cases of violations according to the type of violation, means used and which consequences they can produce. Below is the list of categories (and subcategories) of breaches, means of technical attacks and legal consequences.

Categories of violations

A. Information security breaches

1. Making content unavailable through technical means
2. Destruction and theft of data and programs
3. Unauthorized access - unauthorized alterations and insertions of content
4. Computer fraud
5. Disabling control over an online account or content

This category refers to violations which are in essence technical or require technical knowledge (coding, infrastructure, device and data manipulation) and practically based on criminal acts of cybercrime. This may include Denial-of-Service attacks, malware injections, breaching an online account and similar violations.

B. Information privacy and personal data breaches

1. Publishing information about private life
2. Illegal interception of electronic communications
3. Citizens' personal data breaches
4. Illegal personal data processing
5. Breaches of information privacy in the workplace
6. Other information privacy breaches

Processing, collecting and analysing citizens' personal data and private information is done by many actors, public and private, and on some occasions they fail to protect the data in accordance with the law. This category refers to breaches of information privacy, in the sense of private life, correspondence and people's personal data.

C. Pressures because of expression and activities on the internet

1. Publishing falsehoods and unverified information with the intention to damage reputation
2. Insults and unfounded accusations
3. Threatening content and endangering of security
4. Hate speech and discrimination
5. Freedom of expression on the internet and the workplace
6. Pressures because of publishing information

This category is used for categorising threats, insults, calls to violence and hate speech, which we often see on social media platforms. In many cases, this type of content is directed at journalists, bloggers, human rights activists, and sometimes other public figures such as artists, sportsmen/women or politicians.

D. Manipulations and propaganda in the digital environment

1. Creating fake accounts and paid promotion of false content
2. Organized reporting of accounts and content on social media
3. Changes or removal of content in the public interest
4. Placement of commercial content as news

We felt the need to include this category since the digital environment, as we have seen in the case of Serbia, can easily be turned into an "information battlefield". Violations that can be filed under this category include promoting fake Facebook pages to show someone in a bad light, mass reporting of legitimate content on social media to induce a takedown, removal of online articles or making significant changes so it loses its primary meaning, or publishing an article which promotes a product or a service as an informative piece.

E. Holding intermediaries liable

1. Pressures because of user-generated content

This category is related to pressures from public or private on ISPs and online platforms to suppress or remove content of their users, usually with threats of litigation, monetary fines or service blocking/suspension.

F. Blocking and filtering of content

1. Blocking/filtering on the network level
2. Algorithmic blocking or suspension of content

This category is reserved for cases for where certain content is blocked on the network level, i.e. either nationwide or in a certain organisation, or in cases where algorithms on social media blocked or suspended legitimate content (e.g. a parody video).

G. Other violations

This is a category used to sort violations that do not fall under any of the existing categories, usually reserved for emerging threats that might require a new category.

Means of technical attacks (for category A) and legal consequences (for category C)

We used additional items to describe categories A (Information security breaches) and C (Pressures because of expression and activities on the internet) as they are more specific in their nature and they can provide more depth into the case in question. We used Roman numerals to designate them so they don't get confused with subcategories of violations.

Means of technical attacks:

I - Malware attacks
II - Code injection
III - Reconnaissance attacks

IV - Interception attacks
V - Access attacks
VI - Flooding/Denial of Service

These are the most common types of technical means used for violations of information privacy we encountered. For example, a news organisation website has been down for several days and it has been established by their tech support that the server where they host their website had been flooded by access requests from 20.000 IP addresses. In this hypothetical case, violation would be categorised under A1 (Making content unavailable through technical means) and the means would be flooding/denial of service (VI).

Legal consequences:

VII - Private lawsuits
VIII - Criminal complaints
IX - Arrests and detentions
X - Important judgments
XI - Confiscation and searches
XII - Misdemeanor complaints

These are the most common types of legal consequences in cases of pressures because of expression and activities on the internet. For example, a blogger has published a post about high-ranking official and criminal charges were brought against him/her. In this hypothetical case, violation would be categorised under C6 (Pressures because of publishing information) and the legal consequence would be criminal complaints (VIII).

Why we chose these categories of violations

Monitoring of digital rights and freedoms is a process specific for every country or region. Violations that often occur in one country, for example internet shutdowns[1], are not so common in other countries. That of course means that SHARE Foundation's methodology for monitoring the state of digital rights and freedoms is not exclusive. We however believe it is a good starting point for starting a similar process in another country or region and that it allows for comparisons and following trends around the globe. As new challenges to digital rights are constantly emerging, researchers are encouraged to build upon the methodology provided by SHARE Foundation and add new categories of violations.

[1] According to Access Now, an international organisation that keeps track of internet shutdowns around the world, the regions most affected by internet shutdowns are Asia and Africa: https://www.accessnow.org/keepiton/

How to track and collect cases

There are several possible methods of keeping track of cases of violations. Our approach is to keep out for social media posts describing violations or news reports from trusted sources which provide details of some violation. Partner organisations and associates also provide valuable insights when it comes to collecting and sorting cases. It is also good practice to verify the reports of violations received from targets.

SHARE Foundation’s online Monitoring database[2] holds cases with the following information (if available):

- Case number
- Title: short explanation of what happened (similar to a news article title);
- Target/Actors: who is being targeted/involved in the case?
- Attacker: who is the party initiating the violation, if known;
- Means/Legal consequences
- Category
- Start date: when did the violation occur;
- End date: if it is a case of continuous violation, when did it end;
- Image: image files which can give more details if the case requires;
- Related link: link(s) that provide more information about the case;
- Description: longer description with case details.

It is also important to note that if there are some information missing, it is better to sort the case as new information might come into light later. Also, when performing analysis of cases, we recommend using uniform values for targets and attackers, e.g. blogger, journalist, activist, state institution, public official, public figure, etc, as it has proven to be an important part of the analysis.

Further reading

SHARE@WORK 2016: Monitoring digital rights and freedoms in Serbia (2017). Available at: https://resursi.sharefoundation.info/wp-content/uploads/2018/10/share_yearly_monitoring_report_2016_eng_final.pdf

[2] Available in Serbian and English: monitoring.labs.rs

Mapping and quantifying political information warfare: Part 1, Propaganda, domination & attacks on online media (2016). Available at:
https://labs.rs/en/mapping-and-quantifying-political-information-warfare/

Mapping and quantifying political information warfare: Part 2, Social media battlefield, arrests & detentions (2016). Available at:
https://labs.rs/en/mapping-and-quantifying-political-information-warfare-2/

Social media as editors of public sphere: YouTube vs. Ombudsman (2016). Available at:
https://resursi.sharefoundation.info/en/resource/social-media-as-editors-of-public-sphere-youtube-vs-ombudsman/

Walking on the Digital Edge. A guide on online media autonomy: Security risks and protection mechanisms (2015). Available at:
https://resursi.sharefoundation.info/wp-content/uploads/2018/10/vodic_walking_eng_web.pdf

Arresting a Facebook account in Aleksinac (2015). Available at:
https://resursi.sharefoundation.info/en/resource/arresting-a-facebook-account-in-aleksinac/

Analysis of Internet freedoms in Serbia: June & July 2014. Available at:
https://resursi.sharefoundation.info/en/resource/analysis-of-internet-freedoms-in-serbia/

Internet remembers everything: Analysis of Internet freedoms during the emergency situation (2014). Available at:
https://resursi.sharefoundation.info/en/resource/internet-remembers-everything/