

# Guidelines for Test Lists Researchers

Test lists are machine-readable files made up of multiple websites (URLs) that are tested for accessibility by network measurement tools (e.g. OONI Probe, Centinel) to uncover censorship on the ground. The entries in them are organized into 31 categories that cover most spheres of social life. There are three types of test lists:

## Local test list

A local test list includes websites that contain websites specifically relevant to given country and that are predominantly in local languages. The local list is comprised principally of websites that are reported to be blocked and of websites that are vulnerable to blocking because of the nature of the content. Additional sites that are less likely to be blocked are also included to provide a baseline for future comparisons and to avoid selection bias.

## Global test list

The global list is a standard universal list that we test in all countries. A majority of the global list is comprised of English-language sites. The global list covers many of the most prominent tools and websites, including social media platforms, major global media organizations, free email, anonymizers, hacking, free web hosting, search engines, and translation sites.

## I. Updating test lists with fresh URLs

### 1. Organization of test lists

Test lists have peculiar structure. They include a broad range of local sites which content ranges from neutral (i.e. culture) to objectionable (i.e. porn). Updating of test lists requires knowledge of local social and political factors. They do not include “Alexa top 1K” sites for given country. Instead. Test lists include 3 categories of URLs: blocked, likely to be censored and not blocked. It is important to remember that test lists are not block lists! Most URLs in them are actually not blocked locally. Therefore, test lists help confirm both the *blocking* & *accessibility* of websites. Further information on the test lists organizing principles can be found [here](#).

The URLs in the lists are organized into 31 different categories included at the end of this document for reference purposes. You can see their description [here](#).

A country researcher should update the list with the most relevant websites from the following categories:

- Opposition political parties;
- Militant or extremist movements that are in conflict with the state;
- Media outlets known to be independent or critical of the government;
- Websites that cover political opponents of the current government, such as exiled or imprisoned dissidents;
- Websites of NGOs that work on freedom of expression, human rights, political rights, technology, and censorship issues in the country;
- If the country has a state religion, a list of websites of other religious denominations that are followed by some significant share of the population, including atheist websites;

## 2. How to identify fresh URLs?

- Start by researching the main political, social and economic issues pertaining to the country of interest (e.g. identify its political system and party organization, probe for presence of ethnic and/or religious minorities, etc.) Wikipedia and the [World Factbook](#) provide a good starting point.

- Identify topics that the country's government has previously indicated as sensitive or prohibited, then search to find websites corresponding to these topics. [OpenNet Initiative](#) reports, [ONI Access](#) and [Freedom House](#) are useful starting resources.

- Familiarize yourself with the [30 categories](#) in which the URLs in the lists are organized – this will give you indication what topics to research. Try to add URLs from each category.

- Review existing *research* on Internet censorship topics (i.e. books, papers, conference presentations) produced by investigative journalists and research institutions (i.e. [Citizen Lab](#), [Internet Monitor](#)) to explore what resources have been blocked in a given country;

- Review *reports* on Internet-related topics by international (e.g. [Reporters Without Borders](#), [Article 19](#)) and local (e.g. [Sinar Project](#), [RosKomSvoboda](#), [Venezuela Inteligente](#)) digital rights organizations;

- Follow articles on Internet censorship topics by *local & global media* (i.e. [Global Voices](#)). You may want to set up Google alerts with key phrases to automate the search process;

- Add the platforms websites of local independent media, political opposition, dissidents, minority communities, human rights, etc.

- Web-ethnography - follow forums & chats where netizens discuss Internet censorship;

- Follow the *blogs & social media* channels of local digital rights activists;
- Outreach to targeted media and NGOs to request information about recently blocked or about to be blocked platforms;
- Perform analysis of *national block lists* (if publicly available) or community platforms listing blocked websites (i.e. [Rublacklist](#) - website where a Russian Internet freedom organization publishes URLs of blocked resources);
- Follow the press releases on the websites and social media channels of local communications authorities who in some cases report what they are planning to block. For example, the Russian communications watchdog [Roskomnadzor](#) does so on their website;

### 3. Adding URLs to the test list

- The main thing to remember is **adding URLs exactly as they appear** in your browser. For example, if you access tacticaltech.org in a browser and you get <https://tacticaltech.org/>, that is what should be added to the test list.
- If given website has HTTP and HTTPS addresses, enter only the HTTPS address.
- Websites with many pages - include the URL of the home (main page) of the website. However, when a site has a country-specific page (e.g. BBC News service for China in Mandarin), add the URL of that page to the list. This is because in some cases the censors will block only the country-specific URL and leave the main page accessible.
- Every time you enter a new URL to the list, you should add the following information:
  - \* category code (column B) and description (column C)
  - \* date in the conventional format: YYYY-MM-DD
  - \* Contributor – Netalitica
  - \* Notes – enter brief description of the URLs added and any other useful information that can help researchers (for example that the site belongs to political party and is blocked).

## II. Cleaning test lists from problematic URLs

### The problem

Network measurements collected with the help of OONI Probes in the CIS region helped reveal different problems that existing test lists suffer from. Similar problems were found later to affect Citizen Lab test lists in general. Problematic URLs in them were found to range from 62% to 94%! Using data collected with such lists may result in making distorted conclusions and wastes resources (e.g. time, traffic expenses) of testers on the ground. Faulty URLs may well populate the lists of your country as well.

### Solution

Identify and flag problematic URLs. To do so, researchers are going to be working with the .CSV country lists available on [GitHub](#). Upon downloading the lists, please expand existing columns as suggested in the attached sample spreadsheet. Please enter your updates in a copy of the original file. After you finish with all updates, please submit your spreadsheet for review to Netalitica. We will then send it to OONI and Citizen Lab for public inclusion on GitHub.

### Steps

- 1) Download the most recent *country test list* from [GitHub](#). To do so, identify the country code of your target state (i.e. “EG” for Egypt), click on the relevant test list to open it, right-click on “Raw” button and select “Save link as...” to save the list in a .CSV file on your drive. Firefox browser works great for this purpose (Chrome sometimes may save the file in .txt format).
- 2) Conduct content analysis of each URL to determine if the entries are properly categorized and problem-free.
- 3) Provide useful and brief information about reviewed URLs in the “Notes” column (F). Such information may include the following: what the site is about (e.g. platform of opposition political party), why it should be tested (e.g. URL was blocked in the past), or other information that makes the site relevant for the test list.
- 4) In the process of review, if you come across a problematic URL:
  - a) flag it
  - b) categorize the problem using “problem category” column (G). Suggested problem codes include “focus changeover”, “outdated”, “domain for sale”, “dead website”, “purged

blog”, “irrelevant content”, “miscategorization”, “URL without HTTPS”, “Global list”. You may add other problem categories.

c) Use “problem notes” column (J) to provide brief details about the issue(s) that identified problematic URLs suffer from. This section is very important as it will give the final reviewer information to decide what to do with the problematic URL.

e) Use “Recommendations” column (L) to describe changes you have made to reviewed URLs (“domain updated”, “categorization updated”, “updated to https”) or suggest recommendations for final reviewers to make (e.g. “delete”, “monitor”, “move to Global”)

## URL problems and fixes

This list offers categorization of problematic URLs and provides instructions how to address them. In the process of reviewing existing URLs, when you come across any of the mentioned below problems, please briefly describe them in “Problem notes” column (J). You should enter **manually** each URL in your browser to check for the following issues:

- **Focus changeover** – URLs that no longer offer the content for which they were included in the lists. Their domains were probably acquired by new owners who populated them with irrelevant content. To identify this transformation, please perform content analysis of each suspicious URL. To determine the original focus of the site, check its URL in [Internet Archive](#). Examples:

- [Free Belarus Initiative](#) – former pro-democracy [initiative](#), now directs to irrelevant content
- Former [website](#) of a Belarussian dissident, now serves irrelevant articles in German;
- Former news platform now operates as online [store](#)

*Solutions:* 1) Determine if the original site has moved to another domain and if it has, update its URL in column “A”. If you think the site continues to be interesting for testing (e.g. site used to belong to a prominent organization or individual), keep it in the list and recommend “monitoring” (L). If the site is not significant, recommend deletion.

- **Dead Website** - domains that cannot be reached for technical reasons not related to filtering such as problems locating their servers’ IP addresses, bad gateways, etc. Examples:
  - [Site](#) about culture
  - News [site](#)

*Solutions:* 1) Check if the site is blocked: go to [OONI Explorer](#) -> click “search” (top right on the page) -> select your state from “Country” list -> select time-period (e.g. one year) -> click “Test Name” and select “Web connectivity” -> enter the domain name of the site -> hit “Filter results” -> check the measurements. 2) If OONI Explorer shows instances of “confirmed” blocking or “anomalies”, keep the URL in the list, insert “blocked” in Notes column (F) and recommend monitoring (column L). 3) If the site is not blocked, recommend deletion.

*Tip:* Generally, when given URL is blocked, we want to keep it in the list and recommend "monitoring". However, if the test list has many outdated URLs that are still blocked, then we should keep monitoring just the most important of them - like the websites belonging to important organizations (political parties) or individuals (political dissidents, activists). Please pick only the important sites for monitoring and recommend "delete" for the rest.

- **Outdated website** - platforms that have been abandoned or not updated for years.
  - Moroccan environment [site](#) not updated since 2006
  - Belarusian news [website](#) not updated since 2011
  - Egyptian [platform](#) not updated since 2011
  - Moroccan news [site](#) – last updated 2017 (keep monitoring – imprisoned journalists)

*Solutions:* 1) Categorize URL as “outdated” in “Problem category” (G). 2) Check if the site has moved to a new domain and update its address. 3) Check if the site is blocked. If the URL is blocked and interesting for testing (e.g. belongs to politically significant organization or individual), keep the URL, recommend “monitoring (L) and put “blocked” in “Notes” column (F). 3) Check site’s ranking on Alexa – keep still popular sites and delete unpopular ones. If recommending deletion, include brief justification in “Problem notes” column (J).

*Tip.* Alexa has a handy [extension](#) for [Firefox](#) browser that allows to check in the browser the rankings of sites as well as access Internet Archive. Use your judgment to determine if given outdated site continues to be interesting from political, social or human rights perspective to keep it in the list.

- **Domains for sale** - URLs that may appear accessible in the tests, but do not offer access to any content. Their domain names are up for sale. Examples:
  - Former [news platform](#)
  - Former [site](#) about culture

*Solution:* Update the URL if site has migrated to new domain. Otherwise, recommend deletion.

- **Dead blogs & social media channels** – accounts that have been suspended or purged for years. Examples:

- purged [account](#) in Live Journal
- purged media [account](#) in Live Journal

*Solution:* Check if the account was moved to a new platform and update its URL in the list. Otherwise, recommend deletion.

- **Domain redirects** – may appear accessible in the tests but in reality, the reader is redirected often to irrelevant content

- Former human rights [platform](#), redirects to irrelevant [content](#)
- Former human rights [site](#), now directs to Chinese [site](#)

*Solution:* Check if the platform has migrated to a new domain and update the latter in the list. Otherwise, recommend deletion.

- **Wrong categorization** – some URLs are put in the wrong category

*Solution:* Put “miscategorized” in “Problem Category” column (G), indicate proper category code and description in columns “H” and “I”, **update** category code and description in columns “B” and “C”, insert “updated categorization” in “Recommendations” column (L).

*Important:* avoid using “MISC” (Miscellaneous) category. Try to find the closest match from the existing categories (listed at the end of this document).

- **Irrelevant** – used for websites that do not have relevance to the researched country or region. If such entries have global relevance, recommend them for the Global list. Otherwise, recommend their deletion. Example:

- Irrelevant [blog post](#) in the Egyptian test list.

- **Duplicates** – used for duplicates which should be deleted. Also used for websites that have both HTTP and HTTPS addresses included in the list - keep only the HTTPS address and delete the HTTP one.
- **URL without HTTPS** – used for websites that have migrated to HTTPS addresses but in the list is included only their HTTP address. You can check for this by inputting the original HTTP address in the browser and observing if it is automatically updated to HTTPS. Example:

- <http://www.alalamtv.net/> automatically changes in the browser to <https://www.alalamtv.net/>

*Solution:* 1) Put “URL without HTTPS” in “Problem Category” column (G); 2) update the URL to its HTTPS version in column “A”; 3) indicate “updated to HTTPS” in “Recommendations” column (L).

- **Old domains** – websites that migrated to new domains. Update their URLs in column “A”, categorize as “old domain” (column G) and put “domain updated” in “Recommendations” column (L)
- **Global test list** – refers to websites that are included in country lists but have global relevance and therefore are better suited in the “Global” test list. When you come across such URLs, recommend “move to global list” (column L) and check the Global list to make sure they are not already there.
  - [Wikipedia simple](#) was put in the Belarusian test list. It should be moved to the Global.
- **Facebook, Twitter & YouTube pages** - since all these platforms support now the secure HTTPS protocol, it is no longer possible for ISPs to block individual pages on them. In order to do so, censors need to block the entire *facebook.com*, *twitter.com*, *youtube.com* domains, which are already included in the “Global” test list. Therefore, you should recommend deletion of individual Facebook, Twitter and YouTube pages and check if separate websites exist for those pages (if so, please add them to the list). For control purposes, you should keep a total of 8-10 Facebook, Twitter and YouTube pages. Their testing will help determine if censors have learned to block individual pages on these platforms. For this control group, select only important pages like entries belonging to prominent liberal media, activists, dissidents, political opposition.

You may very well run into new categories of faulty URLs. Please try to categorize them and discuss how to systematically address them with Netalitica.

### III. Generate report

Please add a second tab to the spreadsheet titled “Report” to provide a summary of editions and changes made to the original spreadsheet. Check the sample spreadsheet for instructions how to organize this report.



#### **IV. Balance test lists**

Test lists need to be internally balanced between and within different categories.

1. *Balance within categories.* Test lists should *not* be limited only to listing blocked sites. They are not designed to be blocklists (see further information on the structure of the lists [here](#)). Ideally, each category should include URLs that are blocked, might be blocked and those that are freely accessible. Try to add URLs from all three types to different categories to achieve diversify and avoid “selection” bias.

2. *Balance between categories.* Lists should be balanced to include representative number of URLs from each category (if possible). Ideally there should be 20-40 websites per category, though it is expected that there won't be sufficient URLs for certain categories (e.g. provocative attire). Politically sensitive categories (e.g. Political Criticism, Human Rights Issues, News Media) and categories known to have been blocked locally are of greater interests and should receive more coverage.

**V. Outreach** to local digital rights organizations, researchers, activists and ask them to:

1. Review the new URLs you added to the list
2. Suggest additional websites
3. Please seek to establish lasting relations with community organizations in order to continue relying on their expertise and voluntary contributions in the future.

#### **VI. Potential risks and safety tips**

Safety first! Conducting research on Internet censorship may be risky in some countries. The risks involved in performing this research come from the need to open manually all URLs from given test list in your browser to determine their validity, categorization, etc. The test lists include a broad range of sites, including provocative or objectionable content (e.g. pornography, provocative attire), hate speech, political criticisms and other which might be illegal to access in some countries. Potential risks from accessing forbidden URLs are described [here](#). Please consider seriously the risks involved in conducting this research and if you decide to participate in this project, please take all necessary steps to conduct this research safely!

We recommend taking the following steps:

- If possible, conduct this research outside of the country whose test list you are updating - preferably from a democratic state. This will not only provide you with opportunity to access websites that might be blocked locally, but also will increase your online security.
- Encrypt your communication with Netalitica and other organizations by using Open PGP. Steps how to do this are described in these online tutorials – for [PC](#) and for [Mac](#).
- Use [Tor browser](#) and a reliable VPN (e.g. [Psiphon](#)) to conduct this research;
- Consider using an alias;
- Here are some popular online resources that provide tips and detailed instructions how to increase digital security of individuals and organizations. We recommend reviewing them. [Security Planner](#), [Security in a Box](#), [Digital Hygiene](#).

If you determine that conducting this research is not safe, please stop immediately and inform Netalitica!

## VIII. Categorization of websites and codes

Category Description	Category Code	Description
Alcohol & Drugs	ALDR	Sites devoted to the use, paraphernalia, and sale of drugs and alcohol irrespective of the local legality.
Anonymization and circumvention tools	ANON	Sites that provide tools used for anonymization, circumvention, proxy-services and encryption.
Communication Tools	COMT	Sites and tools for individual and group communications. Includes webmail, VoIP, instant messaging, chat and mobile messaging applications.
Control content	CTRL	Benign or innocuous content used as a control.
Culture	CULTR	Content relating to entertainment, history, literature, music, film, books, satire and humour
E-commerce	COMM	Websites of commercial services and products.
Economics	ECON	General economic development and poverty related topics, agencies and funding opportunities
Environment	ENV	Pollution, international environmental treaties, deforestation, environmental justice, disasters, etc.
File-sharing	FILE	Sites and tools used to share files, including cloud-based file storage, torrents and P2P file-sharing tools.
Gambling	GMB	Online gambling sites. Includes casino games, sports betting, etc.
Gaming	GAME	Online games and gaming platforms, excluding gambling sites.
Government	GOVT	Government-run websites, including military sites.
Hacking Tools	HACK	Sites dedicated to computer security, including news and tools. Includes malicious and non-malicious content.
Hate Speech	HATE	Content that disparages particular groups or persons based on race, sex, sexuality or other characteristics
Hosting and Blogging Platforms	HOST	Web hosting services, blogging and other online publishing platforms.
Human Rights Issues	HUMR	Sites dedicated to discussing human rights issues in various forms. Includes women's rights and rights of minority ethnic groups.
Intergovernmental Organizations	IGO	Websites of intergovernmental organizations such as the United Nations.
LGBT	LGBT	A range of gay-lesbian-bisexual-transgender queer issues. (Excluding pornography)
Media sharing	MMED	Video, audio or photo sharing platforms.
Miscellaneous content	MISC	Sites that don't fit in any category (XXX Things in here should be categorised)
News Media	NEWS	This category includes major news outlets (BBC, CNN, etc.) as well as regional news outlets and independent media.
Online Dating	DATE	Online dating services which can be used to meet people, post profiles, chat, etc
Political Criticism	POLR	Content that offers critical political viewpoints. Includes critical authors and bloggers, as well as oppositional political organizations. Includes pro-democracy content, anti-corruption content as well as content calling for changes in leadership, governance issues, legal reform. Etc.
Pornography	PORN	Hard-core and soft-core pornography.
Provocative Attire	PROV	Websites which show provocative attire and portray women in a sexual manner, wearing minimal clothing.
Public Health	PUBH	HIV, SARS, bird flu, centers for disease control, World Health Organization, etc
Religion	REL	Sites devoted to discussion of religious issues, both supportive and critical, as well as discussion of minority religious groups.
Search Engines	SRCH	Search engines and portals.
Sex Education	XED	Includes contraception, abstinence, STDs, healthy sexuality, teen pregnancy, rape prevention, abortion, sexual rights, and sexual health services.
Social Networking	GRP	Social networking tools and platforms.
Terrorism and Militants	MILX	Sites promoting terrorism, violent militant or separatist movements.

Check for category updates and description [here](#)