

# Guidelines for Test List Researchers

## What are Test Lists?

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. The URLs in them are organized into 30 categories. There are 3 types of lists:

### *Country*

They are made up of country specific URLs - predominantly in the local language(s) and are tested only in the states they have been designed for.

### *Global*

It is populated with globally relevant platforms (mostly in English language) such as social media (Facebook, Twitter, YouTube, etc.), global media (BBC, CNN, DW, etc.), email (Gmail, Hotmail, etc.), anonymizers, web hosting, search engines, translation sites, etc. This list is tested automatically in any country where measurements are collected along with that country's local list.

### *Custom*

Custom lists are designed for specific projects and are not relevant to this assignment.

## I. Updating test lists with fresh URLs

Test lists are not block-lists. They are populated by 3 types of URLs: a) blocked, b) likely to be blocked, and c) not blocked. In fact, most URLs in them are not blocked, which allows researchers to confirm both *blocking & accessibility* of websites. Test lists are not Alexa top 1,000 sites for individual countries either. The URLs in them are organized into 30 different [categories](#), encompassing all spheres of social life. They include a broad range of local sites, which content ranges from neutral (e.g. culture) to objectionable (e.g. porn). Further information on lists' organizing principles can be found [here](#).

### 1. How to identify fresh URLs?

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

- Familiarize yourself with the [30 categories](#) in which the URLs are organized. This will give you indication what topics to research. Add URLs from each category.
- Add the websites of as many news media platforms as possible. These can be pro-government, neutral or opposition sites. We want to probe them all! You can find a list of country-specific media here: [Top 30 News Sites](#) and [ABYZ Links](#). Please open each URL manually in a browser (rather than just copy-pasting published links) as many URLs have changed.
- Add the websites of all local political parties. These websites need not be blocked and could include pro-government, neutral as well as opposition parties. Please note that sometimes the addresses of these sites may be defunct or have changed, so it is important that you manually check and verify each website in your browser.
- Using similar techniques, identify and add the websites of local minority groups, human rights organizations, political dissidents, Internet freedom and other civil society groups.
- The “Government” category should ideally include key government and military websites of rival states (when applicable) as they might be blocked locally. For example, the Iranian test list includes Israeli government sites as both countries are in conflict. It is ok to include few local government websites too, but that shouldn't be no more than 5 except if there is a good reason to include more.
- Identify topics that the country’s government has previously indicated as sensitive or prohibited. Then, search to find websites corresponding to those topics. [OpenNet Initiative](#), [ONI Access](#) and [Freedom House](#) reports are great points of departure.
- Review existing research on Internet censorship topics (i.e. books, papers, conference presentations) produced by research organizations (i.e. [Citizen Lab](#), [Internet Monitor](#)) to explore what resources have been blocked in your country. Consult reports 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 published in local & global media (i.e. [Global Voices](#)).
- Review forums & chats where citizens discuss Internet censorship topics.
- Follow blogs & social media channels of local digital rights activists.
- Outreach to blocked media and NGOs to request information about blocked or about to be blocked platforms.
- Perform analysis of national block lists (if publicly available) or community platforms listing blocked websites.

- Review press releases and social media channels of the local communications authorities who sometimes report what is or going to be blocked.

## **2. Adding URLs to test lists**

- Add URLs exactly as they appear in your browser. For example, if you access “tacticaltech.org” you get “https://tacticaltech.org/” in your browser and that is you should add to the list.

- Add backslash “/” at the very end of each URL.

- If given website has HTTP and HTTPS addresses, enter only the HTTPS address

- Include the URL of the main (home) page of given website. If the platform has country-specific pages (e.g. BBC service for Kazakhstan are in English, Russian and Kazakh languages), add the URLs of different services as well, as in some cases censors block only country-specific URLs and leave the main page accessible.

- Every time you enter a new URL, you should add the following information relevant to it:

\* Category code (column B)

\* Description (column C) – consult table on p.10

\* Date in “YYYY-MM-DD” format

\* Contributor – Netalitica

\* Notes – brief description of intriguing URLs & other relevant information that can be of assistance to researchers (such as that given site is blocked, editors jailed, etc.)

## **II. Cleaning test lists from problematic URLs**

### **The problem**

Content analysis of websites that URLs in Citizen Lab lists direct to helped reveal a number of problematic URLs which range between 62% to 94% per list depending on the country! Using such lists to conduct research will result in distorted conclusions and waste of resources of local testers (e.g. time, traffic). Faulty URLs populate the lists of your country too.

## Solution

Review and flag problematic URLs in your country list. The list of your country is available for download in .CSV format on [GitHub](#). You are going to enter your updates in a copy of the original file. After you finish with the updates, please submit your spreadsheet for review to Netalitica and we will prepare it for publishing.

## Steps

- 1) Download the most recent version of your *country test list* from [GitHub](#). To do so, identify the country code of your target state (for example “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 format on your computer. Firefox browser works great for this purpose (Chrome sometimes saves the file improperly).
- 2) Conduct content analysis of each URL in your list to determine if the included entries suffer from any problems (discussed below).
- 3) Provide brief information about reviewed URLs in the “Notes” column (F). Such information may include the following: what is the site about, why it should be tested (e.g. URL was blocked in the past), other useful information. See the attached sample spreadsheet for examples.
- 4) When you come across a problematic URL:
  - a) flag it.
  - b) categorize the problem in column G using the following category codes: “focus changeover”, “outdated”, “domain for sale”, “dead website”, “purged blog”, “irrelevant content”, “miscategorized”, “URL without HTTPS”, “Global list”. You may add other problem categories.
  - c) Use “problem notes” column (H) to provide brief details about the issue(s) that faulty URLs suffer from. This section is important as it will give the final reviewer information to decide what to do with the URL in question.
  - d) Enter the date the review was conducted (column I)
  - e) Use “Recommendations” column (J) to describe changes you have made to problem URLs (e.g. “domain updated”, “categorization updated”, “updated to https”, “monitor”) or recommendations for changes that need to be made (e.g. “delete”, “move to Global list”). Please do NOT delete bad URLs from the list as the deletion will be performed by the final reviewer.

## Types of URL problems and solutions

You should review manually each URL from your test list by entering it into your browser. Please do not use scripts or other automated methods in the review process.

The following list offers categorization of problematic URLs and provides instructions how to handle them. When coming across problematic entries, please briefly summarize your findings in the “problem notes” column (H).

- **Focus changeover** – URLs that direct to irrelevant content and no longer offer the original content for which they were included in the list. Most likely these domains changed owners and were populated with irrelevant content by the new owners. To determine the original focus of the site, enter its URL in [Internet Archive](#) to see a snapshots of the original website. Examples:

- [Free Belarus Initiative](#) – formerly a pro-democratic [initiative](#), now serves irrelevant content
- The [website](#) of a Belarussian dissident, [now](#) serves irrelevant content
- Former platform of youth [organization](#) now operates as an online [store](#)

*Solutions:* 1) Determine if the original site has moved to a new domain and if this is the case, update its URL in column “A”. 2) If the site continues to be interesting for testing, keep it in the list and recommend “monitoring”. Otherwise, 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:* Keep blocked sites in the lists even if they appear dead. 1) First, check if the dead URL you discovered has been blocked. TO do so, go to [OONI Explorer](#) → click “search” (top right on the page) → select your state from “Country” list → select time-period (e.g. 1 year) → select “Web connectivity” from “Test Name” → enter the domain name of the site (remove the “http://” from the front and “/” from the back of the URL: <http://adb.org/> → adb.org ) → hit “Filter results”. 2) If OONI Explorer shows instances of “confirmed” blocking or “anomalies”, keep the URL in the list, insert “blocked” in the Notes (column F) and recommend monitoring (column J). 3) If the site is not blocked, recommend deletion.

*Tip:* If the test list has too many dead URLs that are still blocked, we should keep monitoring just the most important of them like the websites belonging to important organizations (human rights) or individuals (political dissidents, activists).

**Outdated website** - platforms that have been abandoned or not updated for years.

- Moroccan environment [site](#) not updated since 2006
- Belarussian 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 such URLs as “outdated” in column “G”. 2) Check if the site has moved to a new domain and if so, update its URL. 3) If the URL is blocked and interesting for testing (belongs to politically significant organization or individual), keep the URL and recommend “monitoring”. 3) Check the site’s ranking on Alexa – keep websites that still generate noticeable traffic and delete unpopular ones. If you recommend “deletion” of given URL, include brief justification to do so in “Problem notes” (column H). 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.

*Tip.* Alexa has a handy [extension](#) for [Firefox](#) browser which allows you to check the country popularity of sites as well as access “Internet Archive” through the browser.

- **Domains for sale:**

- Former [news platform](#)

*Solution:* If the site has migrated to a new domain, update its URL. Otherwise, recommend deletion.

- **Dead blogs & social media channels** – accounts that have been suspended:

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

*Solution:* Check if the account has moved to another platform and update its URL in column “A”. Otherwise, recommend deletion.

- **Domain redirects** to irrelevant content:

- [End Violence Against Women URL](#) redirects 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** – often URLs are put in the wrong category

*Solution:* Put “miscategorized” in “Problem category” (column G), update category code and description in columns “B” and “C”, insert “categorization updated” in column “J”.

*Important:* do not use “MISC” (Miscellaneous) category – it is being phased out. 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 URLs have global relevance, recommend them to be moved to the Global list. Otherwise, recommend deletion:
  - Irrelevant [blog post](#) in the Egyptian list.
- **Duplicates** with entries in the country list and with URLs from the Global list should be recommended for deletion. “Duplicates” is also used for websites that have both HTTP and HTTPS addresses. If this is the case, keep only the HTTPS address and recommend the HTTP one to be deleted.
- **URL without HTTPS** – used for websites that have migrated from HTTP to HTTPS protocol. You can identify such instances by observing if the URL you enter in your browser is automatically changed to HTTPS. Example:
  - <http://www.alalamtv.net/> → automatically changes to <https://www.alalamtv.net/>  
*Solution:* 1) Put “URL without HTTPS” in “problem category” column (G); 2) update the address to the latest HTTPS version; 3) indicate “HTTPS updated” in column (J).
- **Old domains** – refers to websites that altered their domain names. Update their new URLs in column “A”, categorize as “old domain” (column G) and put “domain updated” in “J”.
- **Global test list** – refers to websites that are included in country lists but have global relevance and are better match for the “Global” list. When you come across such URLs, recommend “move to global list” (column J) and check the [Global](#) list to make sure they are not already present in it (in which case recommend their deletion from the country list as “duplicates”).
  - [Wikipedia simple](#) was included in the Belarusian test list. It should be moved to Global and deleted from the local list.
- **Facebook, Twitter & YouTube pages** - since all these platforms nowadays support 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. The latter, however, are already included in the “Global” test list and will be tested automatically in your country. Thus, you should recommend deletion of individual social media pages included in the list of your country. For control purposes, keep in total up to 10 Facebook, Twitter and YouTube pages of important organizations/individuals.

In the process of research, you may run into new types of faulty URLs. If so, try to categorize them and discuss your findings with Netalitica.

### **III. Balance test lists**

Test lists need to be internally balanced between and within 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 test lists [here](#)). Test lists should include URLs that are blocked, might be blocked and those that are freely accessible. Try to add all three types of URLs to each category if possible.

2. *Balance between categories.* Lists should be balanced to include representative number of URLs from each category. Ideally, there should be 20+ websites per category. It is expected that some categories (e.g. mass media, human rights, political criticism) will be populated by more URLs than others (e.g. provocative attire, porn). Politically sensitive websites and those known to be blocked are of greater interests to this project and shall receive greater representation in the list.

**IV. Outreach** to local digital rights organizations and researchers to ask them to review the URLs you added to the list and suggest additional websites. Please seek to establish lasting relations with such organizations in order to continue relying on their expertise and voluntary contributions in the future.

### **V. Generate report**

At the end of your research, please fill out the “Report” tab on the spreadsheet we provided you with.

### **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 list in your browser. 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 the 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](#), for [Mac](#).
- Use [Tor browser](#) and a reliable VPN (e.g. [Psiphon](#)) to open individual URLs safely
- Consider using an alias
- Here are some popular online resources offering instructions how to boost the digital security of individuals and organizations. We recommend reviewing them. [Security Planner](#), [Security in a Box](#), [Digital Hygiene](#).
- If you determine at anytime that conducting this research is not safe, please stop immediately and inform Netalitica!

## VII. 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](#)