

# TEST LISTS FOR MEASURING INTERNET CENSORSHIP

---

**APPLICABILITY, PROBLEMS & SOLUTIONS**



# IMPROVING LISTS OF CENSORED ONLINE CONTENT



- 
- Update test lists for 48 states in Latin America, Africa, MENA, Asia and CIS
  - Develop methodology (<http://netalitica.com/>)
  - Establish collaboration among tool developers, researchers and community organizations
  - Advance knowledge and research on Internet censorship

# WORKSHOP STRUCTURE

---

- What are Citizen Lab test lists?
- Types and organization
- How to clean lists?
- How to update lists?
- Potential risks and safety tips

# WHAT ARE THE TEST LISTS?

- Machine-readable files
- Made up of URLs tested for blocking by network measurement tools (e.g. OONI Probe, Centinel)
- Include sample of websites
- History

url	category_code	category_description	date_added	source	notes
http://www.alahrambeverages.cc	ALDR	Alcohol & Drugs	2019-03-31	netalita	
https://zenvpn.net/	ANON	Anonymization and circumvention tools	2017-06-23	AFTE	
https://te.eg/wps/portal/te/Persc	COMT	Communication Tools	2019-05-16	Netalita	
https://thegeekdaily.com/	CTRL	Control content	2017-12-06	AFTE	
http://www.culturewheel.com/ar	CULTR	Culture	2019-06-02	Netalita	
http://www.aljisir-news.com/	COMM	E-commerce	2017-07-01	AFTE	in past this website v
http://www.eaee-eg.com/	ENV	Environment	2018-02-09	OONI	not likely to be block
https://torrents.me/	FILE	File-sharing	2018-08-02	AFTE	
https://www.casinoarbi.com/	GMB	Gambling	2019-03-31	netalita	
http://www.mod.gov.il/	GOVT	Government	2017-08-10	Egypt List	
https://www.ampproject.org/	HOST	Hosting and Blogging Platforms	2018-02-07	AFTE	
https://www.eipr.org/	HUMR	Human Rights Issues	2014-04-15	citizenlab	
https://www.ochaopt.org/	IGO	Intergovernmental Organizations	2019-03-31	netalita	
https://www.globalgayz.com/	LGBT	LGBT	2014-04-15	citizenlab	
https://www.zeedrama.net/	MMED	Media sharing	2019-06-02	Netalita	Reportedly blocked
https://arabist.net/	NEWS	News Media	2014-04-15	citizenlab	
https://www.zawaj.com/	DATE	Online Dating	2019-05-16	Netalita	
http://www.gamal-mubarak.com,	POLR	Political Criticism	2014-04-15	citizenlab	
https://6-ar.com/	PORN	Pornography	2019-03-31	netalita	
https://www.victoriassecret.ae/e	PROV	Provocative Attire	2019-05-16	Netalita	
https://www.altibbi.com/	PUBH	Public Health	2019-05-16	Netalita	
https://islamonline.net/	REL	Religion	2014-04-15	citizenlab	
http://www.misrlinks.com/	SRCH	Search Engines	2019-05-16	Netalita	
https://www.zanzu.de/ar/	XED	Sex Education	2019-05-16	Netalita	
https://twitter.com/ikhwanweb	GRP	Social Networking	2014-04-15	citizenlab	
https://dawaalhaq.com/	MILX	Terrorism and Militants	2017-08-21	AFTE	



# TYPES OF TEST LISTS

---


## COUNTRY

- Tested in a single country
- URLs relevant to that country
- In local language(s)

## GLOBAL

- Tested in all countries
- Globally relevant URLs
- Predominantly in English

## CUSTOM

- Made to answer specific RQ
  - Customizable ([OONI Run](#))
  - Limited N of URLs
- 

# STRUCTURE - COUNTRY LISTS

---

- Broad range of local sites
- Not “Alexa top 1,000 sites”
- **Not block lists:** include blocked + likely to be blocked + not blocked URLs
- Help confirm *blocking* as well as *accessibility* of websites
- Organized into 30 categories

url	category_code	category_description	date_added	source
http://www.alahrambeverages.cc	ALDR	Alcohol & Drugs	2019-03-31	netaltica
https://zenvpn.net/	ANON	Anonymization and circumvention tools	2017-06-23	AFTE
https://te.eg/wps/portal/te/Persc	COMT	Communication Tools	2019-05-16	Netaltica
https://thegeekdaily.com/	CTRL	Control content	2017-12-06	AFTE
http://www.culturewheel.com/ar	CULTR	Culture	2019-06-02	Netaltica
http://www.aljissr-news.com/	COMM	E-commerce	2017-07-01	AFTE
http://www.eaee-eg.com/	ENV	Environment	2018-02-09	OONI
https://torrents.me/	FILE	File-sharing	2018-08-02	AFTE
https://www.casinoarbi.com/	GMB	Gambling	2019-03-31	netaltica
http://www.mod.gov.il/	GOVT	Government	2017-08-10	Egypt List
https://www.ampproject.org/	HOST	Hosting and Blogging Platforms	2018-02-07	AFTE
https://www.eipr.org/	HUMR	Human Rights Issues	2014-04-15	citizenlab
https://www.ochaopt.org/	IGO	Intergovernmental Organizations	2019-03-31	netaltica
https://www.globalgayz.com/	LGBT	LGBT	2014-04-15	citizenlab
https://www.zeedrama.net/	MMED	Media sharing	2019-06-02	Netaltica
https://arabist.net/	NEWS	News Media	2014-04-15	citizenlab
https://www.zawaj.com/	DATE	Online Dating	2019-05-16	Netaltica
http://www.gamal-mubarak.com/	POLR	Political Criticism	2014-04-15	citizenlab
https://6-ar.com/	PORN	Pornography	2019-03-31	netaltica
https://www.victoriassecret.ae/e	PROV	Provocative Attire	2019-05-16	Netaltica
https://www.altibbi.com/	PUBH	Public Health	2019-05-16	Netaltica
https://islamonline.net/	REL	Religion	2014-04-15	citizenlab
http://www.misrlinks.com/	SRCH	Search Engines	2019-05-16	Netaltica
https://www.zanzu.de/ar/	XED	Sex Education	2019-05-16	Netaltica
https://twitter.com/ikhwanweb	GRP	Social Networking	2014-04-15	citizenlab
https://dawaalhaq.com/	MILX	Terrorism and Militants	2017-08-21	AFTE

# CATEGORIES OF URLS

---

- Alcohol & drugs
- Anonymization and circumvention tools
- Communication tools
- Control content
- Culture
- E-commerce
- Economics (economic development)
- Environment
- File-sharing
- Gambling
- Gaming
- Government
- Hacking tools
- Hate speech

# CATEGORIES OF URLS

---

- Hosting and blogging platforms
- Human rights issues
- IGOs
- LGBT
- Media Sharing
- News Media
- Online dating
- Political Criticism
- Pornography
- Provocative attire
- Public health
- Religion
- Search engines
- Sex education
- Social networking
- Terrorism and Militants

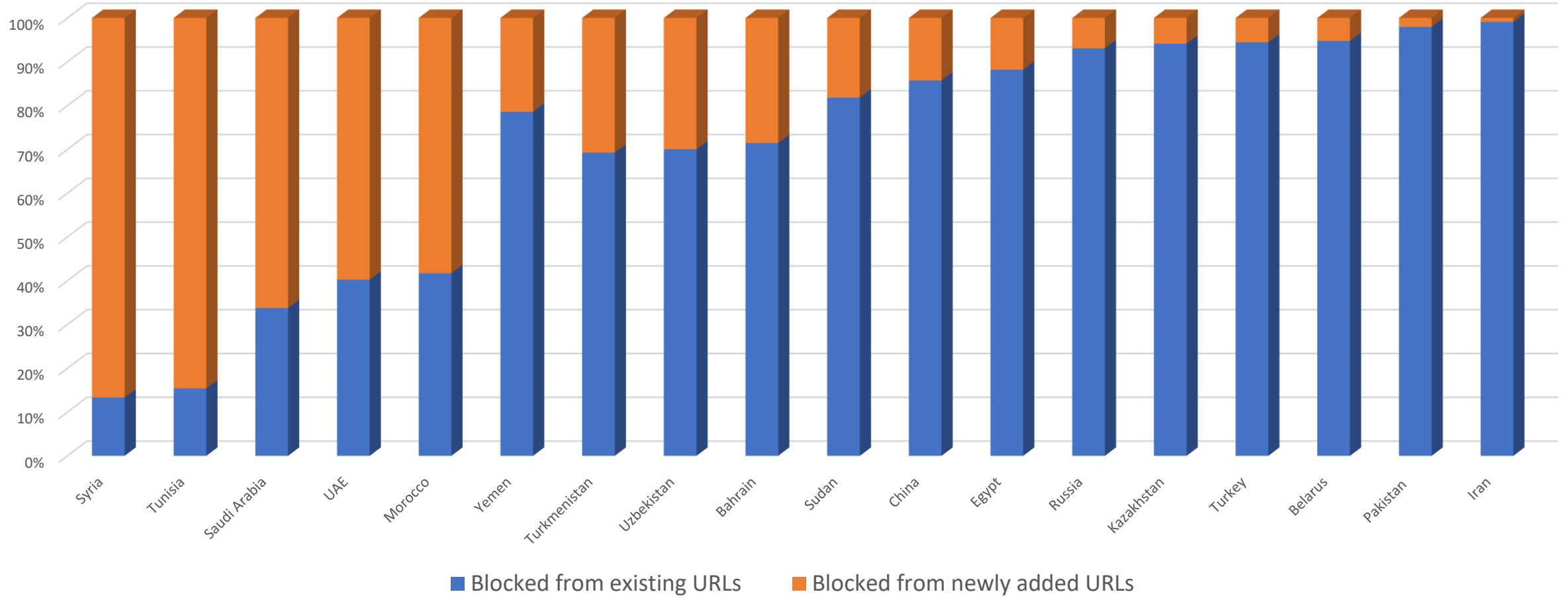


# WHY BOTHER?

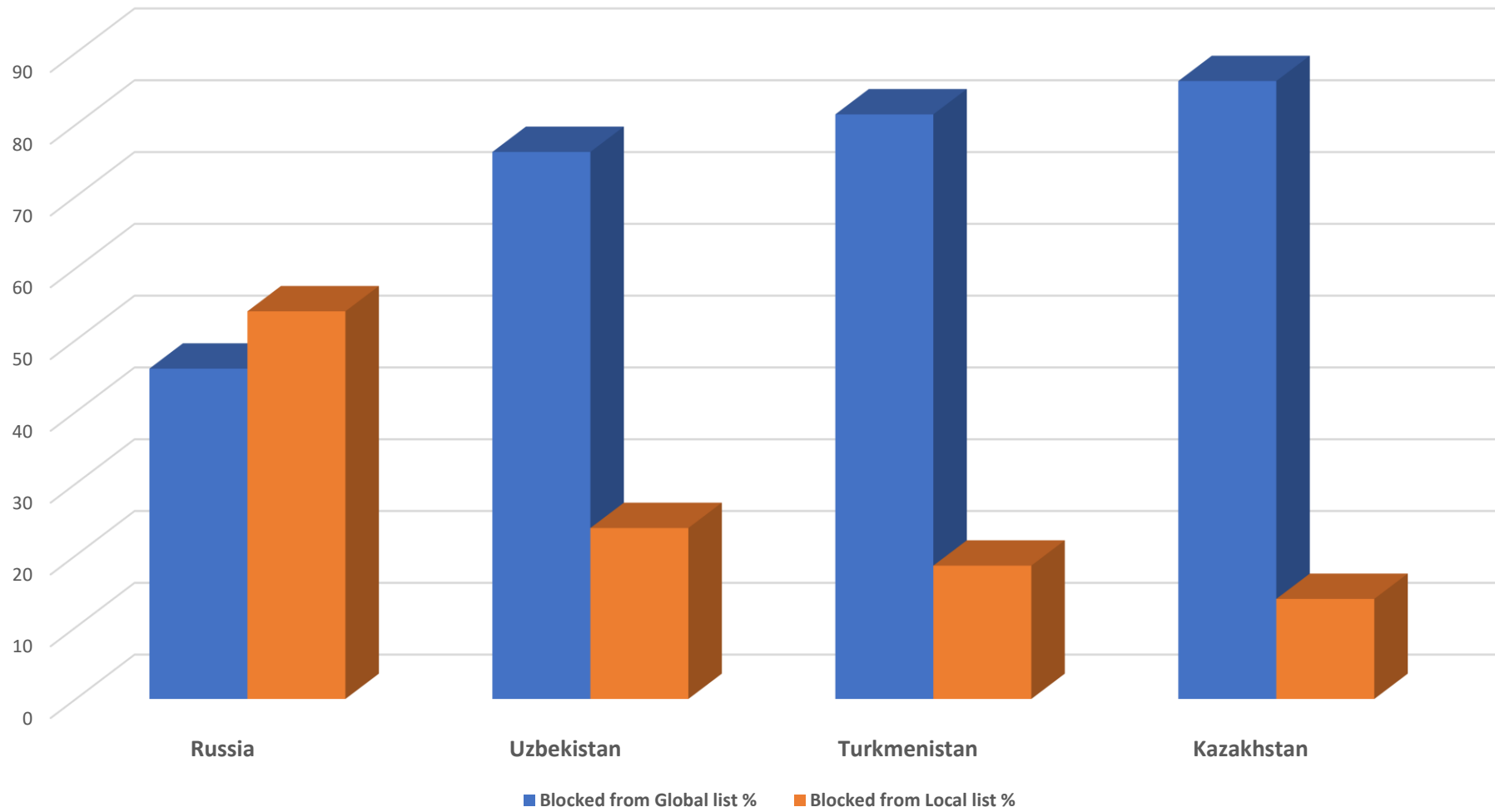
---

- “Censorship findings are only as interesting as the sites you test” (OONI)
- Raise awareness
- Keep authorities accountable
- Help local testers
- Advance Internet censorship research

## Identified cases of blocking



## Blocking of websites in the CIS - global vs. local content



---

## UPDATING LISTS AS A 3-STAGE PROCESS

Review & removing “bad” URLs

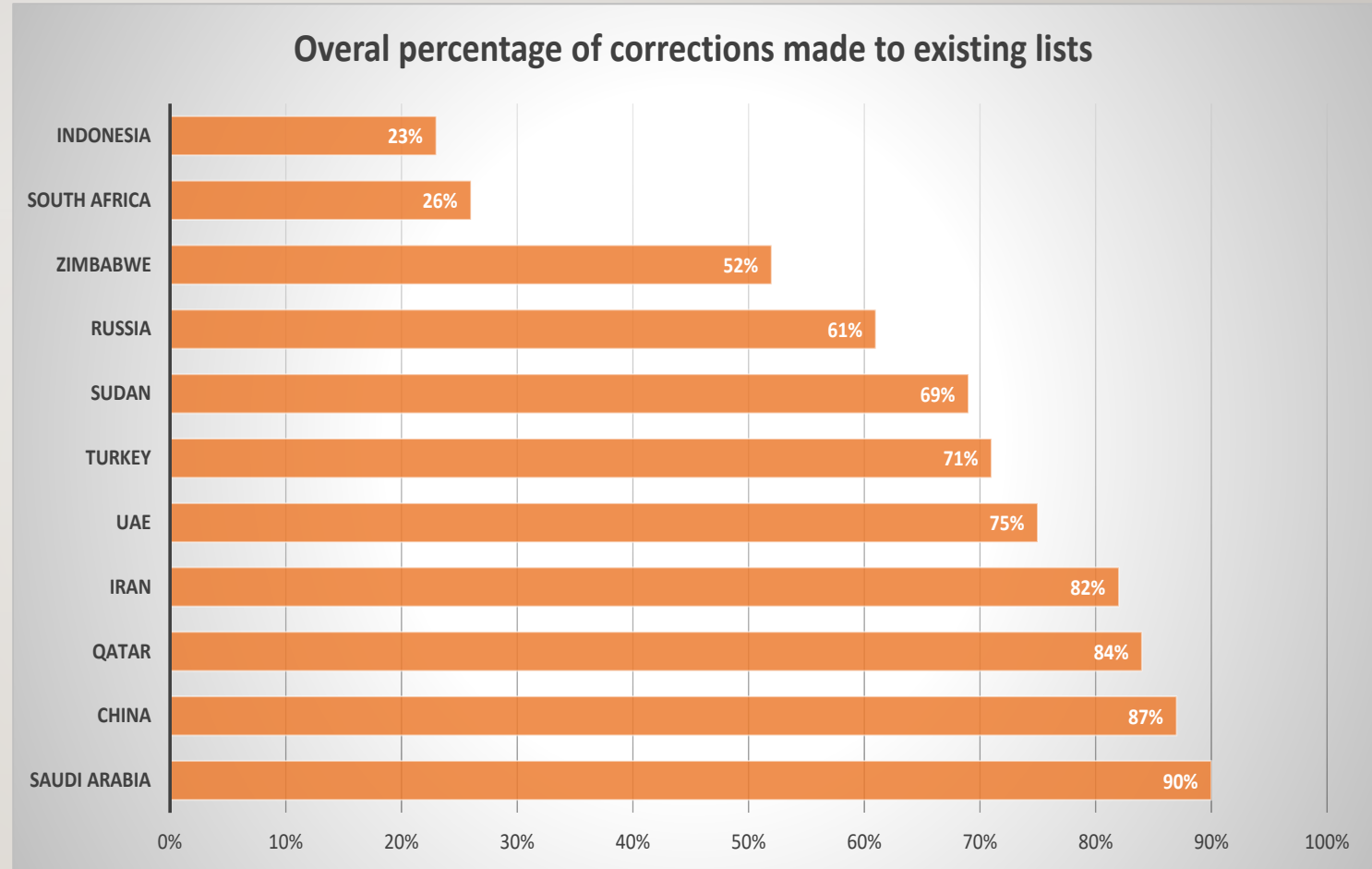
Adding fresh websites

Balancing



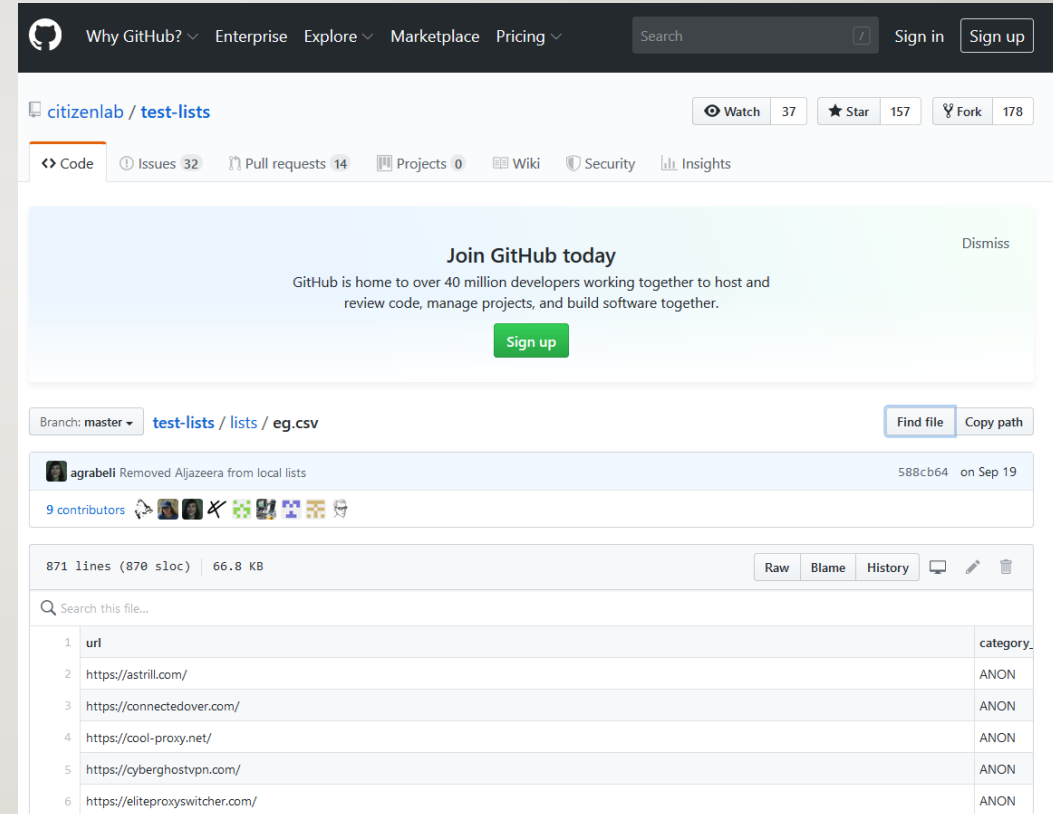
---

# STEP I: CLEANING LISTS FROM “BAD” URLS



# HOW TO CLEAN LIST?

- Consult our [Guideline \(netalitica.com\)](https://netalitica.com/guideline)
- Download country list from [\\_GitHub](https://github.com)
- Conduct analysis of each website in-browser
- Record findings
- Get in touch with Netalitica ([netalitica@gmail.com](mailto:netalitica@gmail.com))



The screenshot shows the GitHub interface for the repository 'citizenlab / test-lists'. The repository has 37 watchers, 157 stars, and 178 forks. The current branch is 'master'. A commit by 'agrabeli' is shown, which removed 'Aljazeera' from local lists. The file 'eg.csv' is selected, showing 871 lines (870 sloc) and 66.8 KB. The file content is a CSV with two columns: 'url' and 'category'. The visible rows are:

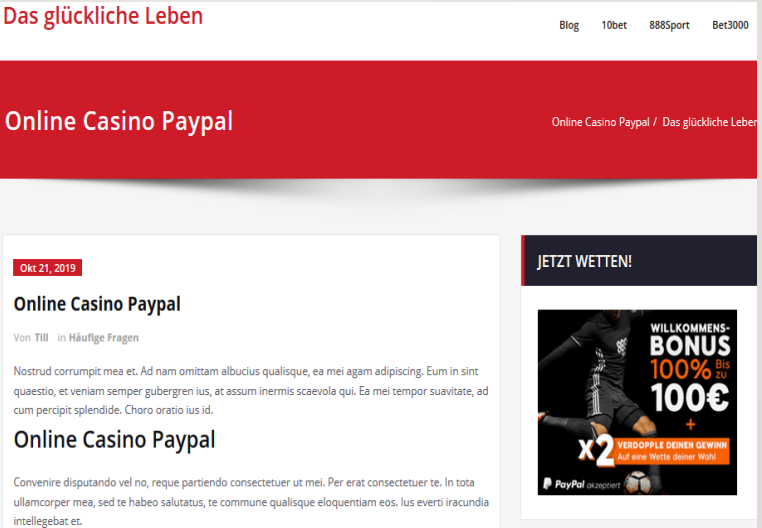
url	category
https://astrill.com/	ANON
https://connectedover.com/	ANON
https://cool-proxy.net/	ANON
https://cyberghostvpn.com/	ANON
https://eliteproxyswitcher.com/	ANON

# FOCUS CHANGEOVER

## BEFORE



## NOW



## SOLUTION

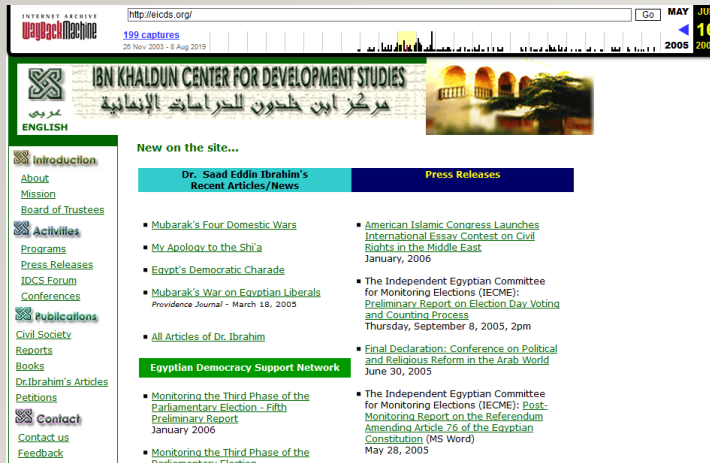
- News site (Egypt)
- Is the site blocked ([OOONI Explorer](#))?
- Keep blocked URLs
- Delete the rest

# DEAD WEBSITES

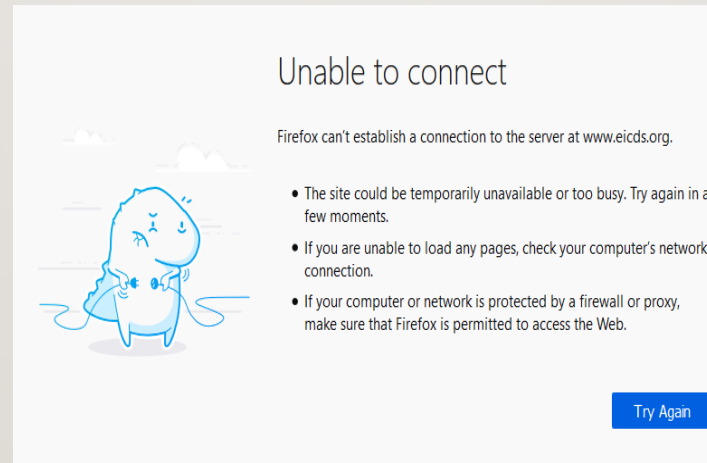
BEFORE

NOW

SOLUTION



The screenshot shows the website's homepage with a navigation menu on the left and a main content area. The navigation menu includes links for Introduction, About, Mission, Board of Trustees, Activities, Programs, Press Releases, IDCS Forum, Conferences, Publications, Civil Society, Reports, Books, Dr. Ibrahim's Articles, Petitions, Contact, Contact us, and Feedback. The main content area features a header with the center's name in Arabic and English, and a section titled "New on the site..." with two columns of recent articles and press releases. The articles include "Mubarak's Four Domestic Wars", "My Apology to the Shi'a", "Egypt's Democratic Charade", "Mubarak's War on Egyptian Liberals", "All Articles of Dr. Ibrahim", "Final Declaration: Conference on Political and Religious Reform in the Arab World", "Monitoring the Third Phase of the Parliamentary Election - Fifth Preliminary Report", and "Monitoring the Third Phase of the Parliamentary Election - Fifth Preliminary Report".



The screenshot shows a Firefox error message with a cartoon character that looks sad. The text reads: "Unable to connect. Firefox can't establish a connection to the server at www.eicds.org." Below the text are three bullet points: "The site could be temporarily unavailable or too busy. Try again in a few moments.", "If you are unable to load any pages, check your computer's network connection.", and "If your computer or network is protected by a firewall or proxy, make sure that Firefox is permitted to access the Web." A "Try Again" button is located at the bottom right of the error message.

- Human rights platform (Egypt)
- Keep blocked URLs
- Delete the rest



# OUTDATED WEBSITES

Political criticism [site](#) (Sudan - 2005)

SOLUTION



الهدف

نداء الي القاري  
إذا رجيت في مساندة المرحلة البعيدة للتغلب على احدى الطغيات الرئيسية امامها يدعم مالي بمكثك الاتصال بالهوان الموجود في البريد الالكتروني .. مع الشكر المقدم.

محمد علي جادين الامين العام

02.06.2005 العدد الاول يصدرها حزب البعث السوداني

مفكرين وأراء  
عبد العزيز الصاوي  
محمد علي جادين  
د. محمد محمود  
منصور خالد  
حيدر ابراهيم علي  
فرنسيس دبتق  
عبد السلام نور الدين

الأخبار

مندوبو مباحثات دارفور يصلون الى نيجيريا لمفاوضات دارفور تعاود اليوم في أبوجا

أوصل مندوبون من كل الأطراف في مباحثات السلام بشأن منطقة دارفور المضطربة في السودان الى العاصمة النيجيرية ابوجا امس الخميس لحضور جولة جديدة من المفاوضات من المقرر ان تبدأ يوم الجمعة بعد توقف دام ستة أشهر

سالم أحمد سالم

[تكملة](#)

- Keep URLs that are blocked & generate traffic
- Delete the rest

# DOMAIN FOR SALE

## BEFORE



## NOW

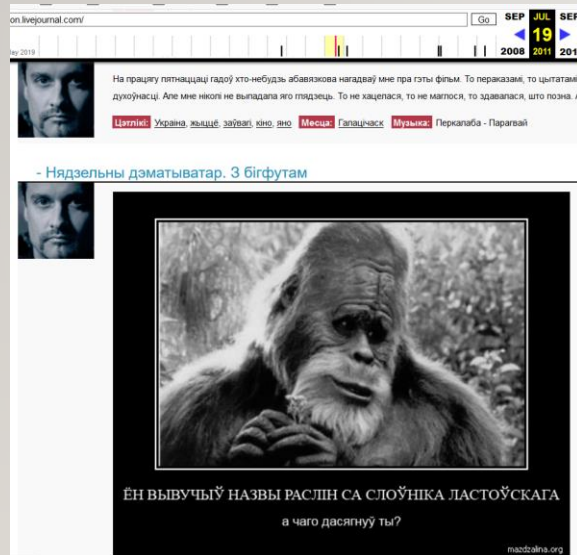


## SOLUTION

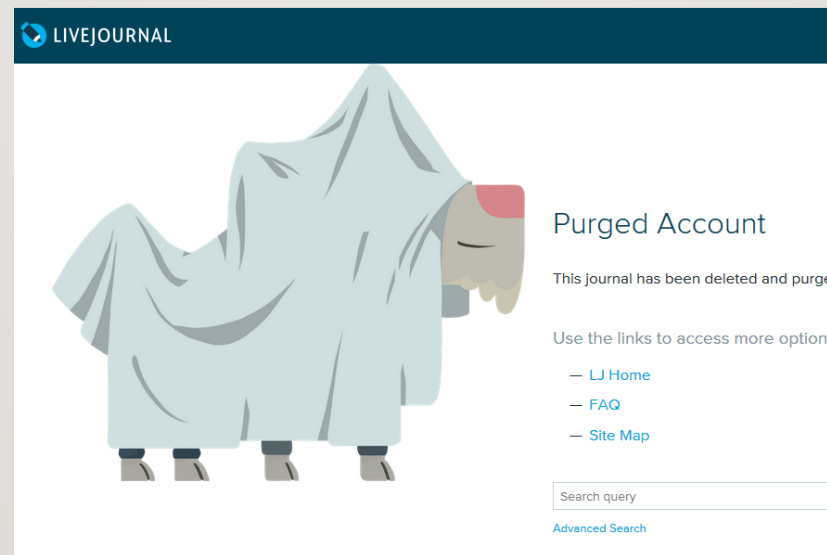
- Blogging platform (Tunisia)
- Update URL if service migrated to new domain
- **Delete**

# URL PROBLEMS – PURGED SOCIAL MEDIA

BEFORE



NOW



SOLUTION

- Update URL if page moved to new platform
- **Delete**



# DOMAIN REDIRECTS

## BEFORE

infoforhealth.org About INFO Project Publications Tools & Services Best Practices Knowledge Management Networks & Communities Topics

## END VIOLENCE *against Women*: INFORMATION AND RESOURCES

Home Resources FAQs Search About E-Newsletter Contribute Partners Contacts

This site was developed by the INFO Project to collect and share in one central location information on the latest research, struggle to end violence against women. We aim to especially cover the intersection of violence against women and subse researchers, health communication specialists, policy makers, and others.

### FEATURED RESOURCE

**Abandoning Female Genital Mutilation/Cutting: Information From Around the World**

This CD-ROM, in English and French, is a collection of data and research on female USAID, this resource is a direct response to the findings of a 2004 survey, undertake respondents said that serious information gaps on FGM/C exist. The CD-ROM fills s unreliable method of research, by including material on various approaches used by training efforts, and policy aspects. This resource is being disseminated to various I of charge, please contact PRB at [prborders@prb.org](mailto:prborders@prb.org); by phone at 202-483-1100; or

### NEW RESEARCH

## NOW

痩せる運動やダイエットの食事

話題のダイエットや運動・食事の効果など、キレイに痩せるための情報を満載

## ダイエット BEAUTY!

トップ /

### 産後ダイエットを成功させる2つの要素

芸能人など産後ダイエットに成功している人と、どうしても妊娠前の体重へ戻せない人との違いとは何なのでしょう。芸能人だからという理由で無条件で産後ダイエットに成功しているわけではありません。産後ダイエ...

[“産後ダイエットを成功させる2つの要素”の詳細は >](#)

### 産後ダイエットでリバウンドしないコツ

産後ダイエットをする度にリバウンドしてしまう、という悩みを持つ人も多いと思います。毎回の出産で少しずつ体重が増えていき、2人目、3人目・・・と、気が付けば子供を産むたびに体重が増えていった・・・今回...

[“産後ダイエットでリバウンドしないコツ”の詳細は >](#)

## SOLUTION

- Update URL if site migrated to new domain
- **Delete**



# IRRELEVANT CONTENT

Irish medium in Ukraine list

SOLUTION

The screenshot shows the BreakingNews.ie website homepage. At the top, there is a search bar and navigation links for 'Ireland', 'Business', 'World', 'Showbiz', '#Discover', 'Sport', 'Tech', 'Jobs', and 'Property'. Below the navigation, there are several news articles with images and headlines. The main article on the left is titled 'Jastine Valdez's killer 'looked as if he knew everything was over', inquest hears'. Other articles include 'US reports first case of Wuhan coronavirus', 'Family concerned as woman goes missing from her home in Carlow', 'Gardaí appeal for help to find teenage boy missing for almost a week', 'Court finds captain and his co-defendant, who was arrested naked, guilty of sailing while drunk on the Liffey', and 'Court hears Cork farmer's dangerous dancefloor moves ended in arrest'. On the right side, there are sections for 'FOLLOW BREAKINGNEWS.IE' with social media icons, 'IRELAND WEATHER' with a forecast, 'HOROSCOPES', and 'LOTTO RESULTS' for Saturday, January 18, 2020, showing numbers 4, 13, 16, 23, 33, 42, and 27.

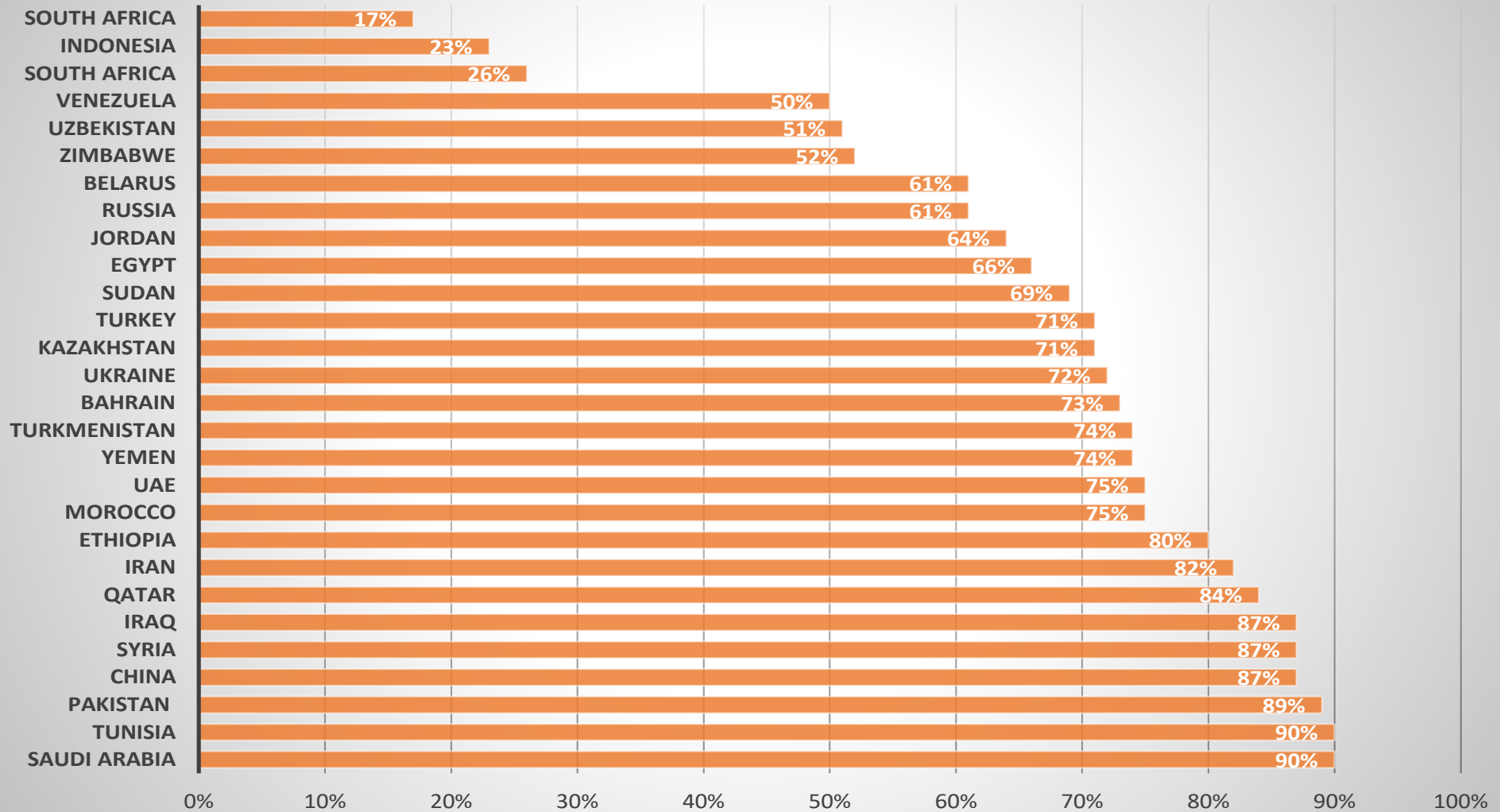
- Delete

# OTHER ISSUES

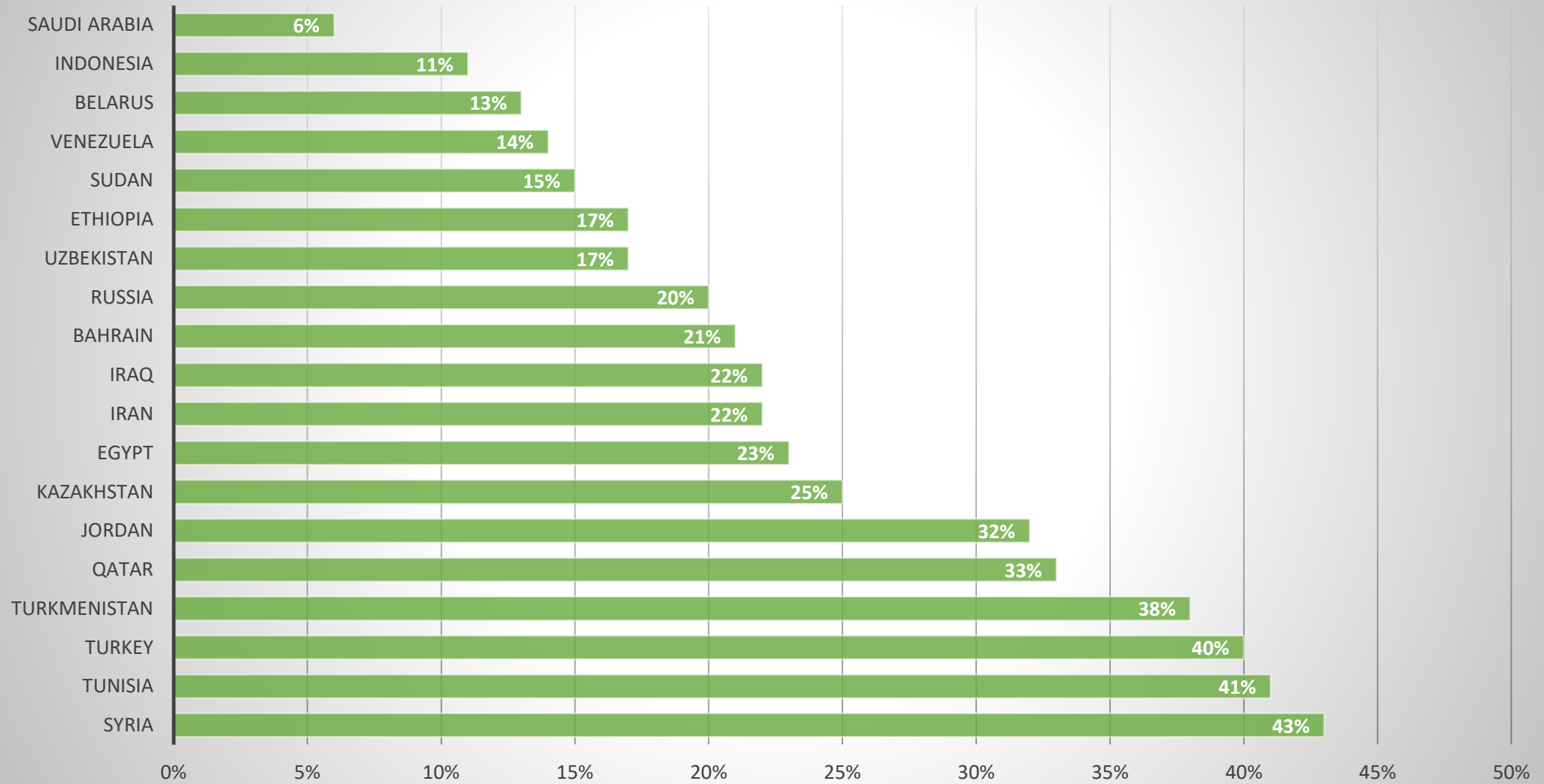
---

- Duplicates (**delete**)
- Wrong categorization (**correct**)
- Old domain names (**update**)
- URLs with global relevance (**move to Global list**)
- Facebook, Twitter & YouTube pages (**delete** - keep up to 5)

## Overall percentage of corrections made to existing lists



## Improved speed of web connectivity tests





## II. UPDATING LISTS WITH NEW WEBSITES

### How to Identify fresh URLs?

---

- Investigate socio-political & economic issues (Wikipedia, World Factbook)
  - Identify sensitive topics and corresponding sites
  - Research on Internet censorship topics (Citizen Lab, [OpenNet Initiative, ONI Access](#))
  - Reports by digital rights organizations
  - Articles in local & global media
  - Social media channels of activists and communications authorities
  - Outreach to targeted platforms
  - Analyze national block lists (if public)
- 

## II. UPDATING LISTS WITH NEW WEBSITES

### How to enter new URLs?

---

- Add URLs exactly as they appear in your browser
- If given website has both HTTP and HTTPS addresses, enter only HTTPS
- When entering a new URL, include:
  - Category Code (e.g. NEWS) and description (e.g. News Media)
  - date (YYYY-MM-DD)
  - Contributor – optional
  - Notes – useful information about added URL

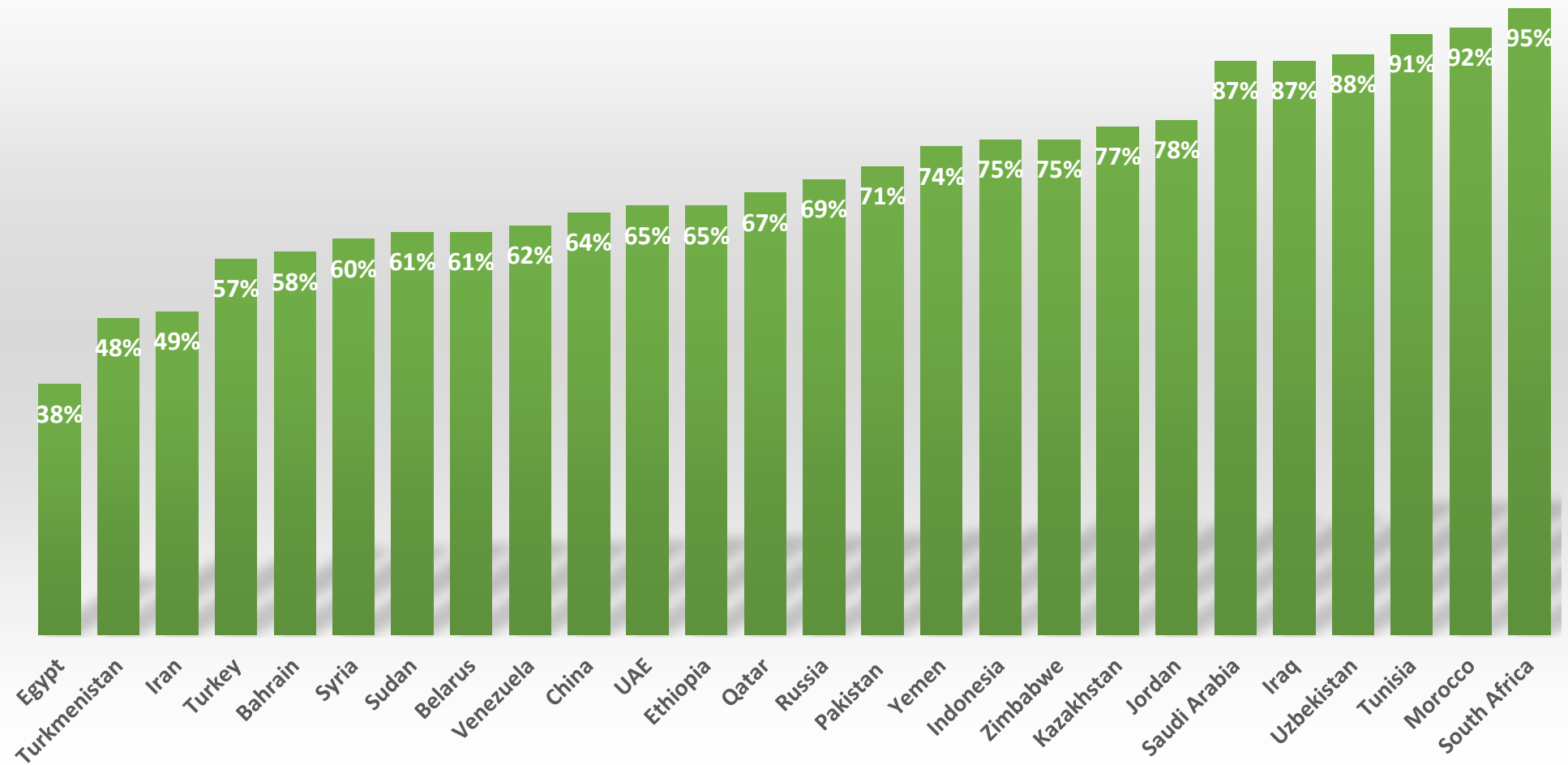
# III. UPDATING LISTS WITH NEW WEBSITES

## How to enter new URLs?

---

url	category_code	category_description	date_added	source	notes
<a href="https://roskomsvoboda.org/">https://roskomsvoboda.org/</a>	HUMR	Human Rights Issues	2014-04-15	citizenlab	influential digital rights organization
<a href="http://www.minjust.net/">http://www.minjust.net/</a>	NEWS	News Media	2014-04-15	citizenlab	dead site but still blocked
<a href="http://ipvnews.org/">http://ipvnews.org/</a>	NEWS	News Media	2014-04-15	citizenlab	blocked, critical of Putin
<a href="http://www.kasparov.ru/">http://www.kasparov.ru/</a>	NEWS	News Media	2014-04-15	citizenlab	political opponent of Putin
<a href="https://hromadske.ua/">https://hromadske.ua/</a>	NEWS	News Media	2017-09-28	igorv	Blocked in Crimea
<a href="https://www.bbc.com/russian">https://www.bbc.com/russian</a>	NEWS	News Media	2018-09-13	OONI	
<a href="https://cherurg.github.io/investing-coins/">https://cherurg.github.io/investing-coins/</a>	ECON	Economics	2019-08-06		Cryptocurrency article
<a href="https://fergana.agency/">https://fergana.agency/</a>	NEWS	News Media	2019-09-06	Netalitica	covers Central Asia states

## New URLs added as % of total





# III. BALANCING TEST LISTS

---

**Within categories** –  
each category should  
include blocked and not  
blocked URLs

**Between categories** –  
each category should  
include representative  
number of URLs

Some categories (e.g. news  
media) will be naturally  
more populated than  
others

---

## IV. SAFETY



Updating test lists may be risky in some countries



Use a reliable VPN (e.g. [Psiphon](#))



Use [Tor browser](#)



Encrypt your communication with Open PGP  
(instructions for [PC](#) and [Mac](#))



Check digital security resources - [Security Planner](#), [Security in a Box](#), [Digital Hygiene](#)

# HIRING RESEARCHERS – [NETALITICA@GMAIL.COM](mailto:NETALITICA@GMAIL.COM) NETALITICA.COM

---

## **Latin America**

- Bolivia
- Brazil
- Cuba
- Ecuador
- Nicaragua

## **Asia**

- India
- Hong Kong
- Myanmar
- Singapore
- Thailand

## **Africa**

- Burundi
- Cameroon
- Uganda