Comparing Measures of Internet Censorship: Analyzing the Tradeoffs between Expert Analysis and Remote Measurement

Terry Fletcher, *Millennium Challenge Corporation (MCC), USA, FletcherTA@mcc.gov*; **Andria Hayes-Birchler,** *Independent Consultant for the Millennium Challenge Corporation (MCC), USA*

Abstract

As the internet increasingly becomes a leading tool for exchanging information, governments around the world sometimes seek methods to restrict citizens' access. Two of the most common methods for restricting the internet are shutting down internet access entirely and filtering specific content. We compare the tradeoffs between measuring these phenomena using expert analysis (as measured by Freedom House and V-Dem) and remote measurement with manual oversight (as measured by Access Now and ONI). We find that remote measurement with manual oversight is less likely to include false positives, while expert analysis is less likely to include false negatives.

Keywords – Remote Measurement; Internet Shutdown; Expert Analysis; Internet Censorship; Internet Filtering;

1 What is Internet Censorship?

Since the 1990s, the internet has spread around the world, reaching 3.8 billion people in three decades and fundamentally changing the way information is produced, disseminated, and consumed (Shahbaz & Funk, 2020; Cohen-Almagor, 2013). Policymakers, civil society, and academics have praised the internet as a tool for encouraging freedom of speech and information globally (USAID, 2020; Corduneanu-Huci & Hamilton, 2018; Howard et al., 2011; *Reno v. ACLU*, 1997). The Arab Spring in the early 2010s is often cited as an example of how the internet can help facilitate information sharing across civil society and hasten transitions to democracy (Farrell, 2012; Howard et al. 2011; Roberts et al. 2011; Stepanova, 2011). However, just as quickly as information has spread across the digital world, governments have found ways to restrict access through various forms of internet censorship (Gopaldas, 2019; Zeleke, 2019; Lakshmana, 2018; Clark et al. 2017b).

We define internet censorship as any method used to intentionally prevent information or services from reaching users over the internet. We focus on government censorship, as censorship by internet service providers (ISPs) is rare and often directed by the government (Taye, 2020; Clark et al. 2017b). As opposed to traditional censorship, which often involves arresting or attacking members of the media to stop content production (Karatnycky et al. 2003; McColm et al., 1991), internet censorship poses new challenges for repressive governments, as they often cannot stop the global production of information,[1] prevent it from entering their country, or stop their citizens from engaging with it (Clark et al., 2017a Clark et al., 2017b). In place of these traditional methods, governments often censor the internet through internet filtering and internet shutdowns.[2]

1.1 Internet Filtering

Internet filtering is used to restrict users' access to specific websites, domains, or IP addresses through technical blocks, including but not limited to DNS poisoning, HTTP filtering through middleboxes, and IP filtering (Yadav & Chakravarty, 2018; Clark et al. 2017b). Governments may deploy internet filtering software themselves, or they may compel ISPs to block or filter certain content within their country (Clark et al. 2017b; Puyosa & Chaguaceda, 2017).

Governments block content for a variety of reasons. Some governments want to restrict the flow of information by blocking e-mail, social media, or video calling services (Clark et al., 2017b; Carsten, 2014). Other governments block online content that expresses certain political views, such as content from opposition parties, civil society, human rights advocates, or specific minority groups (Shahbaz & Funk, 2020; Clark et al., 2017b). Some restrict

[1] Websites with content which is illegal throughout the world (such as child pornography, copyright infringement, and scams) are often taken down or seized by authorities (usually in the developed world) instead of being blocked (Farivar & Blankstein, 2019; Immigration and Customs Enforcement, 2018; Sisario, 2010). Taking down websites that are illegal around the world is not considered censorship in these datasets and is not the focus of this paper. This is also generally not an option for authoritarian governments as domains must be registered in the country (Kravets, 2012).

[2] There are additional methods of censorship, including raising the price of internet above market-value to restrict access (through government monopolies or excessive fees); passing restrictive laws on online activities; utilizing internet surveillance, harassment, arrests, legal action or attacks to intimidate, punish or induce self-censorship in content producers or consumers; and pressuring ISPs to engage in censorship. Due to this paper's focus on remote measurement, however, it examines only censorship methods that can be measured through remote, machine-based methods.

content for social, cultural, or religious reasons, such as content related to sexuality, gambling, drugs, alcohol, or other content that is perceived to be offensive (Clark et al., 2017b). Governments may block content continuously, or only at specific times, such as around an election (Anthonio, 2020; Taye, 2020; Clark et al., 2017b). They may be transparent – noting that access to certain sites is not permitted – or they may try to disguise the filtering so that the lack of access appears to be a technical problem – such as displaying "file not found" on a restricted website (Taye, 2020; Clark et al., 2017b; Dalek et al, 2015).

In recent years, improvements in security protocols and circumvention tools, have made filtering challenging. Encryption makes it difficult for censors to see which portions of a website a user is attempting to access (Rahimi & Gupta, 2020; Clark et al., 2017a). HTTPS in particular has made it challenging for governments to restrict certain pages without censoring the entire website (Rahimi & Gupta, 2020; Clark et al., 2017a; Clark et al., 2017b). This leads governments to restrict either the entirety of a website or none of it (e.g. all of Wikipedia or none of it, instead of just select pages). Circumvention tools like VPNs can also get around this selective filtering but are ineffective against full internet shutdowns (Al-Saqaf, 2015).

1.2 Internet Shutdowns

In part due to the increasing difficulty of filtering select content, governments are more often turning to blunt censorship tools, such as dramatically slowing the speed of the internet (also known as throttling) or shutting down the entire internet (Taye, 2020; Al-Saqaf, 2015). Internet shutdowns were rare in the early 2010s (Subramanian, 2012; Roberts et al. 2011) but have become increasingly common (Taye, 2020; CIPESA, 2019), often occurring around specific events such as an election or large protest (Anthonio, 2020; Taye, 2020; Clark et al., 2017b). Governments often cite concerns about violent protest or instability as a reason for shutting down the internet (Taye, 2020; Clark et al. 2017b), although studies have demonstrated that shutting down the internet tends to increase the likelihood of violence, rather than decrease it (Rydzak, 2019; Rydzak, 2018). Like filtering, shutdowns may be done in a way that makes it difficult to differentiate between intentional shutdowns and technical issues. Internet shutdowns may be country-wide or targeted, so that only certain regions are shutdown, and they may last only a few hours or months (Taye, 2020). Internet shutdowns are often cited as more harmful than internet filtering since they impact the entire internet economy (NetBlocks, 2020; Woodhams & Migliano, 2020; Raveendran & Leberknight, 2018; West, 2016).

Today, both internet filtering and internet shutdowns are widespread practices, with some sources estimating that some form of internet censorship currently exists in more than half of countries (Bischoff, 2020; Mechkova, 2020). Internet filtering has been widespread for many years, but the number of internet shutdowns has increased dramatically each year since the mid-2010s (Selva, 2019; Clark, 2017b). Estimating the exact number of governments that utilize internet filtering or internet shutdowns is challenging, since many governments attempt to hide or disguise their internet censorship, and technical failures can be mistaken for censorship (VanderSloot et al., 2018; Pearce et al., 2018; Gueorguiev et al., 2017; Crandall et al, 2015). This paper explores two main methods of measuring internet censorship – expert analysis and remote measurement - and examines the pros and cons of each. We compares the findings from four of the most accessible datasets on internet censorship, and discusses the tradeoffs faced by policy-makers, civil society, and academics that use these data.

2 Measuring Internet Censorship

Government censorship of the internet is inherently focused on the removal and obfuscation of information. Governments often work to hide both the content of the internet from their citizens, and the methods they are using to hide that content (VanderSloot et al., 2018; Gueorguiev et al., 2017). This means that measurement of internet censorship can be both challenging and dangerous (Pearce et al., 2018; VanderSloot et al., 2018; Weinberg 2018., Narayanan & Zevenbergen 2015; Crandall et al. 2015). However, having accurate measures of internet censorship is important for a range of stakeholders, including users attempting to subvert it, academics attempting to better understand it, and donors or advocates attempting to address it or incentivize policies that limit it.

Despite a need for accurate measures of internet censorship, we find that almost no work has been done comparing the consistency of available data on internet censorship. We conduct a systematic review of the literature on internet censorship using Google Scholar. We search the full text of all articles containing the terms "internet censorship," "internet filtering", or "internet shutdowns" and choose four datasets that are often cited.[3] We then repeat the search with the same terms and the name of each dataset: namely Freedom House's Freedom of the Net; Varieties of

[3] We focus on datasets that may be useful to donor and advocacy organizations which prioritize public and accessible data with broad country coverage as described in section 2.1. Other datasets identified in the literature include data from the Open Observatory of Network Interference (OONI, 2020), Censored Planet (2020), Howard et al. (2011), and ICLab (Niaki et al. 2020).

Democracy's Digital Society Project (V-Dem); OpenNet Initiative (ONI); and Access Now's #KeepItOn data.

We find that at least one of these datasets is featured in a quarter of articles including the words "internet shutdowns", and an eighth of articles including the words "internet censorship" or "internet filtering." However, we find that only one article compares the results of two of these datasets (Frantz et al. 2020 compares ONI and V-Dem), and no articles compare any three or all four.

2.1 Consumers of Internet Censorship Data

In our review, we find there are many reasons consumers seek data on which governments censor the internet. Academics have an interest in understanding trends in internet censorship and its relationship to other phenomena (e.g. Sutterlin 2020; Sagir & Varlioglu 2020; Freyburg & Garbe 2018; Howard et al. 2011). Some consumers are technical experts and internet users working to circumvent censorship practices (e.g. Al-Saqaf, 2015; Leberknight et al. 2012; Roberts et al. 2011). Other consumers are advocacy or donor organizations that use the data to pressure governments to stop internet censorship (e.g. Sayadi & Taye 2020; Parks & Thompson, 2020; SK 2020; Millennium Challenge Corporation, 2019).

While all consumers of these data value an accurate reflection of the world, they may place more or less value on other characteristics of a dataset. An academic researcher may value a dataset that includes many years of historical data for the purpose of running regressions. A user of circumvention tools in an authoritarian country might value data that is constantly updated. Some donors are interested in a dataset with broad country coverage, publicly available and accessible data, and measurements explicitly linked to governance (Millennium Challenge Corporation, 2019; USAID 2019). We focus on datasets and criteria of interest to global donor and advocacy organizations.

2.2 Producers of Internet Censorship Data

We find that there are two broad methods of measuring internet censorship referenced in the literature: expert analysis and remote measurement. We define expert analysis as a process where one or more experts answer specific questions, which are used to create quantitative scores about internet censorship in a country. Remote measurement uses software and user reports to sense and catalog specific censorship events, often with human oversight. The datasets we use from Freedom House and V-Dem are expert analyses (Pemstein et al. 2020; Freedom House, 2019). The datasets we use from ONI and Access Now are remotely measured (Access Now, 2017; Faris, R. & Villeneuve, N. 2008). While some work has been done to compare individual datasets or new tools with existing tools for validity (Frantz et al. 2020; Raman et al., 2020), we find no work comparing methodologies as we do here.

2.2.1 Expert Analysis

The methodology for expert analyses involves periodically surveying one or more experts and aggregating that information into a quantitative measure. These analyses are published regularly, usually on an annual basis. Sometimes they include disaggregated data on certain responses or narratives explaining the rationale for score changes. These data are used to provide general context (e.g. Maréchal, 2017), variables in regressions (e.g. Sagir & Varlioglu, 2020), and to determine funding and incentivize reform (Millennium Challenge Corporation, 2019). While the reports produced by these organizations can provide helpful context for censorship, there are some drawbacks to expert analyses, given that they do not document specific events nor provide information as to exactly how the internet was censored in particular instances (Roberts et al., 2011).

Examples of expert analyses with questions on internet censorship include Freedom House's Freedom on the Net report, Freedom House's Key Internet Controls, Reporters Without Borders' Press Freedom Index, and V-Dem's Digital Society Project. In this report we focus on V-Dem and Freedom House's Key Internet Controls, as these have disaggregated data which examines the same questions on internet censorship. These two datasets use different methods of expert analysis: Freedom House trains a single expert or organization in their methodology and how to create a narrative report (Freedom House, 2019), whereas V-Dem surveys multiple experts and then aggregates their responses into a single score (Pemstein et al. 2020).

2.2.2 Remote Measurement

Remote measurement of internet censorship involves sensing and cataloging specific instances of censorship (such as certain pages that were blocked or moments when the internet was shut down in a particular place). We divide remote measurement into three categories: no oversight, manual oversight, and automated oversight. No oversight methods generally involve a program testing for a particular type of censorship in a given country, and the raw data being made available for use by other researchers. Examples include OONI, which uses software installed on computers of volunteers around the world to sense censorship instances, or ICLab (OONI, 2020; Niaki, 2020). However, without some degree of oversight, these methods are prone to false positives, false negatives, and other technical challenges (Yadav & Chakravarty, 2018; Weinberg, 2018; Pearce et al. 2018; Weaver et al. 2009).

To mitigate these challenges, many datasets turn to some type of oversight. Manual oversight methods are those which involve some level of human testing or aggregation of instances of censorship. This may involve a machine identifying a possible instance of censorship and a human checking to see if it can be confirmed, or a human reviewing a series of automated tests and aggregating them into a single score. The two remotely measured datasets reviewed here both use manual oversight. ONI has volunteers download software on their computers that tests a list of potentially censored pages. These automated results are then reviewed by humans and aggregated into a score for each of four policy categories (Faris, R. & Villeneuve, N. 2008). Access Now uses both volunteer reports and machine sensing methods to detect potential shutdowns and then uses local volunteers, internet companies, and the media to manually confirm shutdowns (Access Now, 2017).

A comparatively new method for detecting internet censorship includes both automated sensing and oversight, where various methods are used to alleviate the challenges of automated remote sensing without requiring human oversight or in-country volunteers (Raman et al. 2020; Hoang et al. 2019; Weinberg, 2019; VanderSloot et al. 2018; Pearce et al. 2018; Sfakianakis et al. 2011). These methods are lauded as being more efficient and ethical, as they do not endanger in-country volunteers (Pearce et al., 2018; VanderSloot et al., 2018; Crandall et al. 2015). However, despite promise for academics and users of circumvention tools, the current forms of these data are too inaccessible and disaggregated to be useful to donors or advocates, and therefore we do not include any in our analysis.

3 Comparing Censorship Data

Given the importance of these data for researchers, donors, policymakers, and civil society, it is vital they be as accurate as possible. Without omniscience we cannot know whether any of these data are perfectly accurate (in that they capture all and only instances of internet censorship). However, it is possible to assess the likelihood that datasets include false positives (they capture censorship that did not actually occur) or false negatives (they do not capture censorship when it occurs)[4] by examining their methodology, as well as comparing how often and where they agree or disagree with one another. While we cannot determine with certainty whether false positive or false negatives occur in any given dataset, the findings from our empirical analyses, combined with each dataset's methodology, and our broader literature review, all suggest that remotely measured data with manual oversight are less likely to contain false positives, but may be more vulnerable to false negatives. Conversely, some expert analyses appear more likely to include false positives but may be less vulnerable to false negatives. Recognizing this may help consumers of these data identify tradeoffs when selecting which datasets to utilize.

3.1 Methods

In order to compare datasets, we focus on three concepts covered by multiple datasets: 1) did a country's government filter political content on the internet in a given year?, 2) did a country's government block social media in a given year?, and 3) did a country's government shutdown the internet in a given year? The exact questions asked by each source, as well as the scales used to score them, are described in Annex A. We use Stata to compare answers from each dataset for the same countries and years.

Due to a lack of overlap in the years and concepts covered by these datasets, it is not possible to compare all variables across all datasets. V-Dem is the only dataset that overlaps temporally with ONI, but ONI does not contain any measure of internet shutdowns. Therefore V-Dem and ONI are compared on the two internet filtering questions (political content and social media) from 2007-2012. Freedom House, V-Dem, and Access Now overlap temporally from 2016-2019, but Access Now does not contain information on the filtering of political content. Therefore, Freedom House, V-Dem, and Access Now are compared on the concepts of social media blockages and internet shutdowns from 2016-2019. For all comparisons, only countries and years covered by all datasets are included.[5]

Since these datasets are on different scales, we first convert all of the scores into binary yes/no responses for the three questions, except for Freedom House's dataset, which is already binary.[6] In the Access Now dataset, any country listed as having a "full-shutdown" in a particular year is counted as shutting down the internet. If a country is listed as having a "service-based" shutdown it is counted as having blocked social media (Access Now, 2017). If a country is listed as having "full and service-based shutdowns" this indicates the government both shutdown the internet for some location or period of time and also blocked social

[4] False negatives occur when a dataset covers a country but does not identify censorship that occurs. This is distinct from not covering a country.

[5] This includes 325 observations compared between Freedom House, V-Dem and Access Now and 74 observations compared between V-Dem and ONI. We test the impact of this small sample size by also comparing just V-Dem and Access Now, using all 716 observations shared by both datasets and find that V-Dem continues to find many more instances of censorship.

[6] The time period for Freedom House data is slightly offset from the calendar year. To test the impact of this, we use Access Now's monthly data to compare the exact time period for one year. We find that the dataset agreement was identical for shutdowns, and almost identical for filtering (off by only 6 percentage points). Access Now continues to identify censorship much less frequently. This indicates that the differences in these datasets are not wholly due to the differences in time periods covered.

media at another location or period of time; as such, it is counted in both categories. In order to convert V-Dem data to binary values we consider any response other than "Never or almost never" as censorship occurring in the country. Similarly, for ONI, any score other than "No evidence of filtering" is counted as censorship occurring in the country.

We then compare these binary scores across each relevant dataset to determine whether responses for each variable are the same across datasets. In other words, if V-Dem states that a given country filtered political content, blocked social media, and shutdown the internet in a given year, do the other datasets agree with this assessment? Where datasets disagree, we then categorize how often each dataset uniquely identified instances of censorship (indicating potential false positives) and how often each dataset uniquely identified instances of non-censorship (indicating potential false negatives).

The literature suggests that remote measurement without oversight is likely to result in false positives and false negatives (Yadav & Chakravarty, 2018; Weinberg, 2018; Pearce et al. 2018; Weaver et al. 2009). However, we anticipate that the remotely measured datasets we examine will result in fewer false positives due to the manual oversight and emphasis on verifiability in their methods. In an attempt to guard against the false positives common to the automated elements of remote measurement, these datasets establish for themselves a burden of proof to verify specific instance of filtering or shutdowns. The same burden of proof does not apply to the expert analysis methodology. We anticipate that this burden of proof may result in more false negatives in these remote measured datasets, as they may believe a country is censoring its internet but cannot verify it and therefore do not count it.

3.2 Findings

Our findings support the hypothesis that remotely measured datasets are less likely to contain false positives than expert analyses. Table 1 depicts the findings of the analysis of Freedom House (FH), V-Dem (VD), and Access Now (AN) on the concepts of social media blockage and internet shutdown. For each concept, each cell is mutually exclusive and together they are jointly exhaustive.

While all three datasets agreed as to whether a country shut down the internet in a given year in the majority of cases, they disagree in at least a third of cases. The disagreement is more pronounced for social media blockages, where the three sources agree only slightly more than half the time. For both concepts, Access Now has the fewest instances (1.15% and 0.8%) of uniquely identifying censorship. The fact that in 99% of cases at least one of the expert analyses agrees with Access Now – combined with the verifiable evidence Access Now publishes for each occurrence – indicates fewer false positives in the Access Now dataset.

Conversely, V-Dem is the sole dataset to identify censorship in 19.6% of instances for internet shutdowns and 14.6% of instances for social media blockages. This combined with the lack of verifiable evidence for its scores may indicate a higher rate of false positives. It may alternatively suggest that V-Dem finds instances of censorship missed by other datasets, and the other datasets include some false negatives.

Table 1. Freedom House, V-Dem, & Access Now Comparison

		FH	VD	AN
Internet Shutdown	Agree	65.4%		
	Unique Censor.	2.3%	19.6%	1.15%
	Unique Non-Censor.	1.5%	1.9%	8.1%
Social Media Blockage	Agree	56.5%		
	Unique Censor.	7.3%	14.6%	0.8%
	Unique Non-Censor.	1.15%	1.5%	18.1%

In order to investigate the issue of false negatives, we examine the cases where a dataset was the only one *not* to list a country as censoring the internet. As hypothesized, Access Now is the most likely to omit a country from its list of censors when both other datasets find that censorship occurred (8.1% of cases for internet shutdowns and 18.1% for social media blockages). V-Dem and Freedom House each have low shares of cases where they uniquely identified a country as not censoring the internet (1.9% and 1.5% for V-Dem, and 1.5% and 1.15% for Freedom House), indicating potentially fewer false negatives.

Table 2. V-Dem & ONI Comparison

		VD	ONI
Social Media Blockage	Agree	72.6%	
	Unique Censorship	17.8%	9.6%
Political Filtering	Agree	68.5%	
	Unique Censorship	28.8%	2.7%

Table 2 presents the findings of the analysis of V-Dem and ONI on the two issues they both covered: social media blockages and political filtering. The two data sources agree in over two-thirds of cases, but once again V-Dem is more likely to uniquely identify censorship (in 17.8% of cases for social media blockages and 28.8% of cases for political filtering). Although there are also cases where ONI uniquely identified censorship, they were far less frequent (9.6% of cases for social media blockages and 2.7% of cases for political filtering.) A comparable analysis completed by Frantz et al., which did not convert the scores into binary measures, similarly found that ONI and V-Dem's measures of censorship have very little correlation (2020.)

While we cannot determine conclusively which dataset is most accurate, our findings suggest that remote measurement with human oversight results in fewer false positives than expert analysis, although it may be more vulnerable to false negatives. Conversely, our analysis also suggests that expert analyses, V-Dem in particular, include more false positives than remote measurement, though they may provide a more complete picture of internet censorship.

One explanation of the false negatives found in remote measurement data could be that these methods are constrained by the number of in-country volunteers or journalists who can manually confirm each instance of censorship. In countries with significant limitations on civil liberties or press freedoms journalists and civil society organizations may lack the capacity to confirm censorship. This is borne out in the data, as the vast majority of censorship identified by Freedom House and V-Dem but not Access Now come from North Africa, the Middle East, and Central Asia. These regions have some of the strictest limitations on the media in the world (Reporters Without Borders, 2020).

There are several possible explanations for the many instances of censorship that are identified by V-Dem, but no other datasets. One is that V-Dem is taking an average of multiple experts' survey responses, and therefore it is less likely to result in a score at either end of the spectrum.[7] Looking at news reports from the countries where V-Dem uniquely identifies internet censorship, it may also be the case that some experts are conflating social media blockages with full internet shutdowns in places like Venezuela (Gold, 2019), Saudi Arabia (Dahan, 2019), and Cuba (Amnesty International, 2017) or civil liberties in general with internet censorship in the Philippines (Engagemedia & Sinar Project 2018). This could explain the high degree of internal correlation in V-Dem noted by Frantz et al. (2020), and would explain why V-Dem's data shows that shutdowns were almost as prevalent in 2000 as they are today, in spite of the literature suggesting they were very rare before 2011 (CIPESA, 2019; Subramanian, 2012; Roberts et al. 2011) and very prevalent today (Taye, 2020).

However, there do appear to be instances where V-Dem is accurately picking up on censorship that is not captured in the other two datasets, such as in Lebanon (Hall 2019)) and Rwanda (McDevitt, 2017), which suggests that instances of V-Dem uniquely identifying censorship are a mix of false positives *and* V-Dem picking up on censorship that the other datasets miss.

4 Conclusion

There are a range of considerations that go into determining which dataset is the best measure of internet censorship for a given purpose. While perfect accuracy is desirable, given the high levels of disagreement it is unlikely that any one dataset is completely accurate. Consumers of these data should therefore consider whether they prefer a higher likelihood of false positives or false negatives. Consumers may prefer that their data be fully verifiable: that there should be a high burden of proof to show a country has censored their internet, even if that means missing some instances. Our analysis indicates that such consumers would be best served by a remotely measured dataset with manual oversight, such as Access Now or ONI, where each instance of censorship has been verified, and where an average of 96.5% of cases are supported by two or more datasets. Other consumers may prefer that as many instances of censorship are captured as possible, even if that means including some countries which may not have engaged in censorship. Given that many governments try to hide their censorship – and that many are quite sophisticated in doing so – these consumers may worry that the burden of proof of remote measurement methods with human oversight leaves them vulnerable to missing too many cases of censorship. These consumers would be advised to use an expert analyses dataset instead, such as Freedom on the Net or V-Dem.

There are of course other, logistical considerations that users consider when choosing a particular dataset. Users that want universal country coverage may avoid Freedom House's data, which only covers 65 countries. Users looking to create a time series may prefer V-Dem's data, as it starts in 2000. Some users may prefer data at the incident level as opposed to the country level, which makes Access Now's dataset more appealing. Others may prefer that the data include a clear index and ranking of countries to better "name and shame" to leverage policy change.

Progress has been made to create remotely measured datasets with automated oversight, which may be more accurate than either of the methods reviewed here (Raman et al. 2020; Hoang et al. 2019; Weinberg, 2019; VanderSloot et al. 2018; Pearce et al. 2018). However, the current versions of these datasets fail to meet many of the logistical considerations above. They are often too technical or disaggregated to be useful to donors and advocacy organizations. Therefore, there is an opportunity for future work in the aggregation of these, potentially more accurate, datasets into annual ranked indices of censorship that are more accessible to donor and advocacy organizations.

[7] When we ran the analyses again including both the V-Dem categories of "rarely" and "never or almost never" as instances of non-censorship, V-Dem uniquely identifies many fewer instances (fewer than Freedom House, in fact, although still as many or more than Access Now or ONI).

Acknowledgements

Thanks to Daniel Barnes, Jennifer Sturdy, Cindy Sobieski, Maïté Hostetter, and Alexandra Berry for reviewing. The views expressed herein are those of the authors and should not be construed as an express or implied endorsement of those views by the MCC nor the U.S. Government.

References

Access Now (2017). Shutdown Tracker Optimization Project. https://www.accessnow.org/cms/assets/uploads/2017/09/How-to-view-the-Access-Now-Internet-Shutdown-Tracker-2017.pdf

Al-Saqaf, W. (2015). Internet Censorship Circumvention Tools: Escaping the Control of the Syrian Regime. *Media and Communication.* 4(1), 39-50.

Amnesty International (2017). Cuba's Internet paradox: How controlled and censored internet risks Cuba's achievements in education. *Amnesty International*. https://www.amnesty.org/en/latest/news/2017/08/cubas-internet-paradox-how-controlled-and-censored-internet-risks-cubas-achievements-in-education/

Anthonio, F. (2020). A Shutdown taints Togo's 2020 presidential election: what happened and what's next. *Access Now*. https://www.accessnow.org/a-shutdown-taints-togos-2020-presidential-elections-what-happened-and-whats-next/

Bischoff, P. (2020). Internet Censorship 2020: A Global Map of Internet Restrictions. *Comparitech.* https://www.comparitech.com/blog/vpn-privacy/internet-censorship-map/

Carsten, P. (2014). Google's Gmail blocked in China. *Reuters*. https://www.reuters.com/article/us-google-china/googles-gmail-blocked-in-china-idUSKBN0K70BD20141229

Censored Planet (2020). Censored Planet Raw Data. https://censoredplanet.org/data/raw

CIPESA (2019). State of Internet Freedom in Africa 2019. *Collaboration on International ICT Policy for East and Southern Africa.* https://cipesa.org/?wpfb_dl=307

Clark, J., Faris, R. Jones, R.H., (2017a). Analyzing Accessibility of Wikipedia Projects Around the World. *Berkman Klein Center for Internet & Society Research Publication*.

Clark, J., Faris, R., Morrison-Westphal, R., Noman H., Tilton, C., & Zittrain, J. (2017b). The Shifting Landscape of Global Internet Censorship. *Berkman Klein Center for Internet & Society Research Publication.*

Cohen-Almagor, R. (2013). Internet History. In Luppicini, R. (Eds.), *Moral, Ethical, and Social Dilemmas in the Age of Technology: Theories and Practice* (pp. 19-39). IGI Global. http://doi:10.4018/978-1-4666-2931-8.ch002

Coppedge, M., Gerring, J., Knutsen, C. H., Lindberg S. I., Teorell, J., Altman, D., Bernhard, M., Fish, M. S., Glynn, A., Hicken, A., Luhrmann, A., Marquardt, K. L., McMann, K., Paxton, P., Pemstein, D., Seim, B., Sigman, R., Skaaning, S., Staton, J., Wilson, S., Cornell, A., Alizada, N., Gastaldi, L., Gjerløw, H., Hindle, G., Ilchenko, N., Maxwell, L., Mechkova, V., Medzihorsky, J., von Römer, J. Sundström, A., Tzelgov, E., Wang, Y., Wig, T., & Ziblatt, D. (2020). V-Dem Dataset v10. *Varieties of Democracy (V-Dem) Project*. https://doi.org/10.23696/vdemds20.

Corduneanu-Huci, C., & Hamilton, A. (2018). Selective Control: The Political Economy of Censorship. *World Bank Group.*

Crandall, J. R., Crete-Nishihata, & M., Knockel, J. (2015). Forgive Us our SYNs: Technical and Ethical Considerations for Measuring Internet Filtering. NS Ethics '15: Proceedings of the 2015 ACM SIGCOMM Workshop on Ethics in Networked Systems Research.

Dahan, N. (2019). Internet, interrupted: How network cuts are used to quell dissent in the Middle East. *Middle East Eye.* https://www.middleeasteye.net/news/internet-interrupted-how-middle-east-countries-use-network-restrictions-clamp-down-dissent

Dalek, J., Deibert, R., McKune, S., Gill, P., Senft, A., & Noor, N. (2015). Information Controls during Military Operations The case of Yemen during the 2015 political and armed conflict. *University of Toronto.* https://citizenlab.ca/2015/10/information-controls-military-operations-yemen/

Engagemedia & Sinar Project (2018). Internet Censorship Monitoring: Duterte's Drug War. https://sinarproject.org/digital-rights/updates/internet-censorship-monitoring-dutertes-drug-war

Faris, R. & Villeneuve, N. (2008). Measuring Global Internet Filtering in R. Deibert, J. Palfrey, R. Rohozinski, & J. Zittrain, (Eds.), Access Denied: The Practice and Policy of Global Internet Filtering, Cambridge: MIT Press. http://opennet.net/sites/opennet.net/files/Deibert_02_Ch01_005-028.pdf

Farivar, C., & Blankstein, A. (2019). Feds take down the 'world's largest dark web child porn marketplace.' *NBC News*. https://www.nbcnews.com/news/crime-courts/feds-take-down-world-s-largest-dark-web-child-porn-n1066511

Farrell, H. (2012). The Consequences of the Internet for Politics. *Annual Review of Political Science*, 15, 35-52.

Frantz, E., Kendall-Taylor, A., & Wright, J. (2020). Digital Repression in Autocracies. *The Varieties of Democracy Institute*.

Freedom House (2019). Freedom on the Net Research Methodology. https://freedomhouse.org/reports/freedom-net/freedom-net-research-methodology

Freyburg, T. & Garbe, L. (2018). Blocking the Bottleneck: Internet Shutdowns and Ownership at Election Times in Sub-Saharan Africa. *International Journal of Communication*, 12(2018), 3896-3916.

Gold, H. (2019). Information war escalates as Venezuela tries to contain uprising. *CNN Business*.

https://www.cnn.com/2019/05/01/media/media-venezuela-information-war/index.html

Gopaldas, R. (2019). Digital Dictatorship versus Digital Democracy in Africa. *SAIIA Policy Insights* 75.

Gueorguiev, D., Shao, L., & Crabtree, C. (2017). Blurring the Lines: Rethinking Censorship Under Autocracy. 10.13140/RG.2.2.29037.08160.

Hall, R. (2019). Lebanon blocks Grindr in latest attack on LGBT+ community. *The Independent*. https://www.independent.co.uk/news/world/middle-east/grindr-lebanon-ban-lgbt-rights-dating-app-gay-a8933556.html

Howard, P., Agarwal, S.D., & Hussain, M. M. (2011). When Do States Disconnect Their Digital Networks? Regime Responses to the Political Uses of Social Media. *The Communication Review*, 14, 216-232.

Hoang, P. N., Doreen, S., Polychronakis, M., (2019). Measuring I2P Censorship at a Global Scale. In Proceedings of the 9th USENIX Workshop on Free and Open Communications on the Internet.

Immigration and Customs Enforcement (2018). Over a million websites seized in global operation. *ICE Newsroom.* https://www.ice.gov/news/releases/over-million-websites-seized-global-operation

Karatnycky, A., Piano, A., Puddington, A. (2003). Freedom in the World: The Annual Survey of Political Rights & Civil Liberties. *Freedom House.*

Kravets, D. (2012). Uncle Sam: If It Ends in .Com It's .Seizable. *Wired.com.* https://www.wired.com/2012/03/feds-seize-foreign-sites/

Lakshmana, K.V. (2018). A Coward's Political Weapon: Troll armies go on settling scores. *Common Cause*, 37(4), 27-29.

Leberknight, C.S., Chiang, M., Poor, H. V., Wong, F. (2012). A Taxonomy of Internet Censorship and Anti-Censorship.

Maréchal, N. (2017). Networked Authoritarianism and the Geopolitics of Information: Understanding Russian Internet Policy. *Media and Communication,* 5(1), 29-41.

McColm, R.B., Finn, J., Payne, D.W., Ryan, J.E., Sussman, L.R., Zarycky, G. (1991). Freedom in the World Political Rights & Civil Liberties. *Freedom House.*

McDevitt, D. (2017). Rwanda censors critical, independent media in targeted fashion. *Open Technology Fund*. https://www.opentech.fund/news/new-report-investigates-internet-censorship-during-rwandas-2017-presidential-election/

Mechkova, V., Daniel P., Brigitte S., & Steven W. (2020). Digital Society Project Dataset v2. *Varieties of Democracy (V-Dem) Project*

Millennium Challenge Corporation (2019). Guide to the MCC Indicators for Fiscal Year 2020. *Millennium Challenge Corporation.* https://www.mcc.gov/resources/doc/guide-to-the-indicators-fy-2020

Narayanan A, & Zevenbergen B. (2015). No Encore for Encore? Ethical questions for web-based censorship measurement. *Technology Science*. http://techscience.org/a/2015121501.

NetBlocks (2020). Internet cut in Ethiopia amid unrest following killing of singer. *NetBlocks Mapping Net Freedom.* https://netblocks.org/reports/internet-cut-in-ethiopia-amid-unrest-following-killing-of-singer-pA25Z28b

Niaki, A.A., Cho, S., Weinberg Z., Hoang, N.P., Razaghpanah, A., Christin, N., Gill, P. (2020). ICLab: A Global, Longitudinal Internet Censorship Measurement Platform. Proceedings of the 41st IEEE Symposium on Security and Privacy.

OONI (2020). OONI Data Explorer. https://explorer.ooni.org/

Parks, L. & Thompson R. (2020). The Slow Shutdown: Information and Internet Regulation in Tanzania From 2010 to 2018 and Impacts on Online Content Creators. *International Journal of Communication,* 14(2020), 1-21.

Pearce, P., Ensafi, R., Li, F., Feamster, N. & Paxson, V. (2018). "Toward Continual Measurement of Global Network-Level Censorship. *IEEE Security & Privacy*, 16(1), 24-33, 10.1109/MSP.2018.1331018.

Pemstein, D., Marquardt, K. L., Tzelgov, E., Wang, Y., Medzihorsky, J., Krusell, J., Miri, F., & von Römer, J. (2020). The V-Dem Measurement Model: Latent Variable Analysis for Cross-National and Cross-Temporal Expert-Coded Data. V-Dem Working Paper No. 21. 5th edition. University of Gothenburg: Varieties of Democracy Institute.

Puyosa, I. & Chaguaceda, A (2017). Cinco regímenes políticos en Latinoamérica, libertad de internet y mecanismos de control. *RETOS*, 7(14), 11-37.

Rahimi, N. & Gupta, B. (2020). A Study of the Landscape of Internet Censorship and Anti-Censorship in Middle East. EPiC Series in Computing: Proceeding of the 35th International Conference on Computers and Their Applications. 69, 60-68.

Raman, R. S., Stoll, A., Dalek, J., Sarabi, A., Ramesh, R. Scott, W., & Ensafi, R. (2020). Measuring the Deployment of Network Censorship Filters at Global Scale. Network and Distributed System Security (NDSS) Symposium 2020.

Raveendran, N., & Leberknight, C.S. (2018). Internet Censorship and Economic Impacts: A Case Study of Internet Outages in India. Proceedings of the Twenty-fourth Americas Conference on Information Systems.

Reno v. American Civil Liberties Union, 521, U.S. 811 (1997).

Reporters Without Borders (2020). World Press Freedom Index. https://rsf.org/en/ranking

Roberts, H., Zuckerman, E., & Palfrey, J. (2011). 2011 Circumvention Tool Evaluation. *The Berkman Center for Internet & Society Research Publication Series.*

Rydzak, J.A. (2018). A Total Eclipse of the Net: The Dynamics of Network Shutdowns and Collective Action Responses. *University of Arizona*.

Rydzak, J. (2019). Of Blackouts and Bandhs: The Strategy and Structure of Disconnected Protest in India. https://ssrn.com/abstract=3330413

Sagir, M. & Varlioglu, S. (2020). Explaining the Relationship between Internet and Democracy in Partly Free Countries Using Machine Learning Models. IT Research Symposium 2020.

Sayadi, E. & Taye, B. (2020). #KeepItOn: As Yemen's war goes online, internet shutdowns and censorship are hurting Yemenis. https://www.accessnow.org/keepiton-as-yemens-war-goes-online-internet-shutdowns-and-censorship-are-hurting-yemenis/

Selva, M. (2019). Reaching for the off switch: Internet shutdowns are growing as nations seek to control public access to information. *Index on Censorship*, 48(3), 19-22.

Shahbaz, A. & Funk, A. (2020). Freedom on the Net 2019: The Crisis of Social Media. *Freedom House.*

Sisario, B (2010). U.S. Shuts Down Web Sites in Piracy Crackdown. *New York Times.*

SK, C. (2020). Those Unspoken Thoughts A study of censorship and median freedom in Manipur, India. *Open Observatory of Network Interference.*

Sfakianakis, A., Athanasopoulos, E., Ioannidis, S. (2011). CensMon: A Web Censorship Monitor. In Proceedings of USENIX FOCI 2011.

Subramanian, R. (2012). The Growth of Global Internet Censorship and Circumvention: A Survey. *Communication of the International Information Management Association.*

Sutterlin, E. (2020). Flipping the Kill-Switch: Why Governments Shut Down the Internet. *Undergraduate Honors Theses.* Paper 1493.

Stepanova, E. (2011). The Role of Information Communication Technologies the "Arab Spring." *PONARS Eurasia.*

Taye, B. (2020). Targeted, Cut Off, and Left in the Dark The #KeepItOn Report on internet shutdowns in 2019. *Access Now.* https://www.accessnow.org/keepiton/

USAID (2019). FY 2020 USAID Journey to Self-Reliance Country Roadmap Methodology Guide. https://selfreliance.usaid.gov/

USAID (2020). Digital Strategy 2020-2024. *United States Agency for International Development*. https://www.usaid.gov/usaid-digital-strategy

VanderSloot, B., McDonald, A., Scott, W., Halderman, J.A., & Ensafi, R. (2018). Quack: Scalable Remote Measurement of Application-Layer Censorship. In Proceedings of the 27th USENIX Security Symposium.

Wagner, B., Gollatz, K., Calderaro, A. (2013). Common Narrative – Divergent Agendas: The Internet and Human Rights in Foreign Policy. Proceedings of the 1st International Conference on Internet Science.

Weaver, N., Sommer, R., & Paxson, V. (2009) Detecting Forged TCP Reset Packets. *International Computer Science Institute.*

Weinberg, Z. (2018). Toward Automated Worldwide Monitoring of Network-level Censorship. *Carnegie Mellon University*.

West, D.M. (2016). Internet Shutdowns cost countries $2.4 billion last year. *Center for Technology Innovation at Brookings.*

Woodhams, S. & Migliano, S. (2020). The Global Cost of Internet Shutdowns in 2019. *Top10VPN.com.*

Yadav, T. K., & Chakravarty, S. (2018). Trends and patterns of internet censorship in India. *Indraprastha Institute of Information Technology Delhi.*

Zeleke, T. A. (2019). The Quandary of Cyber Governance in Ethiopia. *Journal of Public Policy and Administration*, 3(1), 1-7.

Annex A

Institution	Freedom House	V-Dem	ONI	Access Now
Dataset	Key Internet Controls	Digital Society Project	ONI Censorship Data	#KeepItOn
Collection Method	Expert Analysis	Expert Analysis	Remote Measurement	Remote Measurement
Years Covered	2015-2019	2000-2019 (pre-2018 data is retroactive)	2007-2013	2016-2019
Countries Covered	65 every year	173 every year (174 after 2011 with South Sudan). Includes all countries covered by all other datasets.	74 observations total. Between 3 and 37 observations per year.	All countries are covered every year. Only countries with censorship are listed.
Internet Filtering Question (Political Content)	(Freedom House's data contains a question on this, but it is not used in our analysis as it does not overlap temporally with a remotely measured dataset that asks this question)	How frequently does the government censor political information (text, audio, images, or video) on the Internet by filtering (blocking access to certain websites)?	How much does the government censor web sites that express views in opposition to those of the current government.	N/A
Internet Filtering Question (Social Media)	Are entire apps or key functions of social media, messaging, and calling platforms temporarily or permanently blocked to prevent communication and information sharing?	How often does the government shut down access to social media platforms?	How much does the government censor web sites that provide e-mail, Internet hosting, search, translation, Voice-over Internet Protocol (VoIP) telephone service, and circumvention methods	"Service-based shutdowns" where a government blocks social media or communication platforms
Internet Shutdown Question	Does the government intentionally disrupt the internet or cellphone networks in response to political or social events, whether temporary or long term, localized or nationwide?	How often does the government shut down domestic access to the Internet?	N/A	Full shutdowns of the internet or cellphone networks, whether temporary or long term, and localized or nationwide.
Scoring Scale	Yes / No	Ordinal Data used: 0 – 4 with 0 meaning "Extremely Often" and 4 meaning "Never or almost never."	0- 4 with 0 meaning "No evidence of internet filtering" and 4 meaning "pervasive internet filtering"	Lists specific instances of filtering or shutdowns