

Measuring Data Use:

An Analysis of Data Portal Web Traffic

Open Data Watch

December 2018



Executive Summary

Overview

The Internet and World Wide Web have become the principal gateway to official statistics distributed by national statistical offices (NSOs) and other government agencies. Almost every NSO maintains a website or data portal—some more than one—that offers access to official statistics. Websites are an efficient means of dissemination because they can distribute large volumes of data at very low cost. However, their effectiveness depends whether data users are aware of the websites or their contents and, if so, whether those users are able to find the information they are seeking.

Web analytics are a means of measuring, collecting, and analyzing website data and the behavior of users on a website. They are widely used by commercial and non-profit organizations to measure the performance of websites and to improve the experience of users online. There are a variety of free and proprietary software systems that can be used to generate and monitor web traffic data. Google Analytics is a widely adopted analytics tool because it is free, easy to implement, and it effectively reports actionable data that allows organizations to optimize their web content.

This report was undertaken to help NSOs and their partners better understand the benefits of web analytics tools and assist them in implementing web analytics on their websites or data portals. These tools are an important part of the data revolution and can contribute to increasing the availability and accessibility of data. NSO web managers and partners should understand that collecting and analyzing web traffic data with an analytics tool is a critical first step in establishing a complete dissemination strategy to better serve a website's users.

Open Data Watch, in partnership with PARIS21, invited seven NSOs in low- or middle-income countries to participate in a study analyzing web traffic on their principal websites or data portals using Google Analytics. They were selected to

Five Lessons for NSO Web Managers

- Every NSO should install Google Analytics or a similar tool on their website. Detailed guidance on how to correctly implement website analytics can be found in the annexes to this report.
- NSOs that have installed Google Analytics should check their settings to fix some common errors identified in Annex 1 of this report.
- A well-structured website should allow the NSO to easily identify and group all pages related to data use and access, which requires changes to the organization of the website and web page naming conventions. This study identifies possible improvements, and we encourage web managers to adopt these best practices.
- Users of NSO websites and data portals were primarily interested in economic data, but poverty and population data were of interest as well. NSOs should attempt to make available the full range of data their users seek.
- We encourage web managers to adopt and use web analytics to improve the accessibility of data on their websites. Implementation of any or all of the recommendations will help achieve that goal.

include at least one NSO from major geographic regions: South America (1), Sub-Saharan Africa (3), Eastern Europe (1), Southeast Asia (2). The selected countries differ in size and in their level of economic development and are not necessarily representative of their regions or of low- and middle-income countries in general. However, the results reveal patterns of website use that provide a useful reference for other NSOs and their website managers. They also serve as an example for analyzing web traffic data that can be applied to other websites.

According to a survey conducted jointly by AidData and Open Data Watch,¹ NSO managers identified web analytics as the most important way of monitoring their data dissemination programs. Importantly, this study revealed that website users are seeking and do find data through the websites and data portals we analyzed, confirming that NSO websites and data portals are not graveyards but serve a vital role in making data available to the public and other stakeholders. However, the study also found that most participating websites are not configured to

collect or make optimal use of the available web analytics data. A set of specific recommendations to improve the structure and configuration of these websites, along with detailed instructions for their implementation, are included with this report. A customized dashboard providing a selection of the most-relevant web traffic data has also been created to serve as a resource for NSOs to monitor the use of their websites and data portals. These are available as annexes to this report.

Key Findings

- This study provides clear evidence that NSO websites and data portals are not data graveyards; they are widely used to access data and serve a vital function in the effort to disseminate data to the public.
- While NSO websites receive a higher volume of web traffic than dedicated data portals, they offer a wider variety of information and services that attract visitors, not all of whom are seeking data.
- Although dedicated data portals receive much lower levels of web traffic than NSO websites, visitors to those sites make more in-depth use of them.
- Search engines drive the most traffic to NSO sites while direct referrals drive most traffic to data portals. We encourage NSOs to implement a search engine optimization plan and should consider strategies to elevate awareness of data portals where necessary.
- Web traffic data can fluctuate because of a variety of external factors. These factors should be considered by NSO web managers when reviewing analytics and updating content to make data more accessible
- Many countries showed decreases in web traffic during the months of December and January, while other variations were unique to each country. These increases likely correspond to the release of data or other key moments on the NSO's calendar.
- Most users arrive at NSO websites through search engines, while most arrive at data portals through direct channels. However, a high amount of traffic to data portals that are closely tied to NSO websites comes from referrals, as users arrive on the data portal through links available throughout the NSO website.
- Domestic users are the source of the greatest number of site visits, which explains why websites from countries with larger populations have higher traffic. Some sites receive significant numbers of foreign visitors, usually from countries with social and economic ties to the country.
- Because of limitations of the structure and naming conventions of the websites and data portals included in this study, it was not possible to estimate the proportion of site visits that sought data or resulted in a data download.

¹ *Counting on Statistics*. <https://aiddata.org/publications/counting-on-statistics>

Acknowledgements

This report was made possible through generous financial support from the William and Flora Hewlett Foundation and is a collaboration between Open Data Watch, AidData at William & Mary, and PARIS21.

Part of a broader project to increase the impact of official statistics, this report focuses on measuring the use of data by analyzing data portals to gather use metrics. It is a companion to *Counting on Statistics: What can data producers and donors do differently to increase use?* which provides evidence of the perspectives of national statistical offices and government ministries in low- and middle-income countries on the use of official statistics.

The principal authors of this report were Eric Swanson and Amelia Pittman of Open Data Watch with advice and assistance from Shaïda Badiëe, Tawheeda Wahabzada, Deirdre Appel, Caleb Rudow, Jamison Crowell. William Hensley assisted in the final editing. Francois Fonteneau and Rajiv Ranjan of PARIS21 collaborated in the design of the project and provided invaluable feedback throughout. Geoffrey Greenwell, formerly at PARIS21, provided useful advice in the early stages. Ben Crawford, Jessica Cooper, Autumn Rose of Forum One provided technical support on the analysis of web analytics and the design of the dashboard template.

Table of Contents

Executive Summary	i
Acknowledgements	iii
Introduction	1
Study Participation and Data Collection	2
Key Performance Indicators	4
Audience Analysis	4
Overview	4
TABLE 1: Number of users, sessions, page views, and the bounce rate	5
Duration and depth	5
TABLE 2: Number of sessions by duration	5
FIGURE 1: Distribution of sessions by duration	6
TABLE 3: Number of sessions by page depth	7
FIGURE 2: Distribution of sessions by duration	7
TABLE 4: Session duration by page depth	8
Variation over time	8
FIGURE 3: Number of sessions per month	8
Audience engagement channels	9
TABLE 5: Number of sessions by channel grouping	10
FIGURE 4: Percent of sessions by channel grouping	10
Geographic origins	11
TABLE 6: Number of domestic users and sessions	11
FIGURE 5: Percent of users that are domestic users	11
Content analysis	12
Overview	12
TABLE 7: Share of page views among top 30 pages	12
Data content	13
TABLE 8: Share of data pages by contents (%)	13
Data search activity	14
TABLE 9: Website search activity	14
TABLE 10: Data-related search terms	15
Data collection challenges	15
Conclusions	16
 Annex 1: Recommendations for Improving Web Traffic Data	20
Annex 2: Web Traffic Dashboard	24
Annex 3: Customize Dashboard Content	34
Annex 4: Instruction Package	40

Introduction

Twenty years ago, most developing countries relied on CD-ROMs, DVDs, or paper publications to make their data available to users. Even large statistical offices in high-income countries were likely to use disks or even magnetic tape as their medium for large datasets. Today, dissemination of official statistics takes place on the website of a statistical agency or a third-party website that repackages data obtained through an application program interface (API). Recognizing this important advance in the data revolution, the 2017 Cape Town Global Action Plan for Sustainable Development² calls for the development of technological infrastructure for better data dissemination and to expand the use of online methods for the dissemination of SDG statistics.

In the most recent round of assessments for the Open Data Inventory (ODIN),³ NSO offices in 178 out of 180 countries, representing 99 percent of the world's population, had a functioning website to distribute official statistics. A survey conducted jointly by Open Data Watch and AidData⁴ showed that a majority of staff in government ministries (who are important users of the statistics produced by NSOs) said they preferred to obtain statistical information from online sources. In the same survey, 72 percent of NSO staff and leaders said they are currently using web analytics to monitor use of their websites. Among those not currently monitoring data use, 85 percent identified web analytics as the tool they would prefer to use.

In 2016, PARIS21 published a report⁵ that highlighted the proliferation of data portals for disseminating official statistics. On average, they found more than two such sites per country, and in African countries, where donors and international organizations have encouraged

their development, there were an average of 3.4 websites. The report raised concerns about the duplication of effort, the confusion among users caused by multiple sites, and “the high costs for demonstrably low usage of these portals.” The study was able to obtain evidence on the use of one country's data portal. The results were not reassuring. They found that 89 percent of 54,000 monthly visitors to an African NSO website stayed on the home page. Of those that navigated more deeply into the website, 6.5 percent went to a page with public job offers; 4.4 percent downloaded data publications in PDF format; and fewer than .005 percent – approximately 2.5 people each month – went on to a dedicated data portal to download data.

This study was designed to extend the work of PARIS21 by capturing more information about the volumes and patterns of web traffic on NSO websites and data portals. In addition, it provided an opportunity to demonstrate the tools available to NSOs for monitoring their web traffic. Notwithstanding interest expressed by NSOs in web analytics, few of the websites included in this study were set up with highly optimized versions of Google Analytics. As a contribution of the study, we provided recommendations to participating NSOs for adjustments to their implementation of Google Analytics to improve their ability to monitor and analyze web traffic. The recommendations are summarized in Annex 1 of this report. Additionally, we have provided participating NSOs with a custom dashboard that displays key indicators from Google Analytics. For other NSOs that might wish to improve their monitoring efforts, the Diagnostic Toolkit included with this report provides access to a dashboard template that can be copied and connected to web traffic from any Google Analytics account.

² Cape Town Global Action Plan for Sustainable Development Data. <https://unstats.un.org/sdgs/hlg/cape-town-global-action-plan/>

³ Open Data Inventory (ODIN). <http://odin.opendatawatch.com/>

⁴ Counting on Statistics. <https://aiddata.org/publications/counting-on-statistics>

⁵ Greenwell, Geoffrey, et. al. *Making Data Portals Work for SDGs: A view on deployment, design, and technology*. Discussion Paper No. 8. PARIS21. 2016. http://paris21.org/sites/default/files/Paper_on_Data_Portals%20wcover_WEB.pdf

Study Participation and Data Collection

Web analytics refers to the collection of data about users and their behavior on a website. Data may be collected in log files maintained by the server that hosts the website, but more commonly, the record of visits to the site is created by “tags,” script files that are encoded in the site and activated when a user visits a web page. The data generated can include the location of the visitor; whether the same visitor (using the same computer and web browser) has visited the site before; how long the visitor remained on a specific web page and where they went next; what search terms they employed while on the site; and other details of the way they interacted with the website content. The quality of the data produced depends on how well designed the website is, the placement of the tags, and how the analytics software, in this case Google Analytics, has been set up. While there are many vendors that provide web analytic services, Google, in addition to being a free service, is the most well-known and widely adopted analytics software available.

The quality of the data produced depends on how well designed the website is, the placement of the tags, and how the analytics software has been set up.

The countries invited to participate in this study were selected from among low- and middle-income countries. The data collected from Google Analytics for the study cover the period October

2017 through September 2018. Because of the need to obtain historical data, only NSOs that had previously installed Google Analytics on their statistical websites were included. Countries in Africa were given preference, but countries from Latin America, Europe, and Asia were also included. Eight countries were invited to participate in this study, selected based on past relationships each one had with PARSi21 and Open Data Watch. Seven agreed to participate in the study and to grant access to the Google Analytics account installed on their websites so that Open Data Watch could directly access web traffic data. In addition to the seven NSO websites, three countries provided access to data portals.⁶ All participating countries agreed to allow their anonymized data to appear in this summary report, which provides cross-country analysis.

The data available on the seven NSO websites and on two of the data portals are primarily statistical indicators compiled from censuses, surveys, and administrative data. One data portal is a specialized archive of microdata containing documentation of surveys and census records. Although the sample of countries and websites is not representative of all countries or even of countries in their regions, some of the observed patterns of use are likely to be commonly found.

Forum One, which served as advisor on web analytics, conducted further analysis prior to the beginning of the study to identify potential issues of website design that might impede analysis. The review of the websites showed that most were not set up to filter traffic generated by bots — automated programs that search for and download information from the internet — and most had not enabled site search tracking, which reports search words used by visitors. To improve data collection, participants were advised on how to filter bots and set up site search tracking.

⁶ There is no practical difference between a website and a data portal, except that the latter may be expected to focus more specifically on providing access to data, while the general-purpose websites maintained by statistical offices often include other services and news reports.

This outreach resulted in a total of seven of the websites (from five countries) with enabled search tracking and bot filtering. However, three websites experienced difficulties that will require further efforts to implement changes. Guidance for resolving these issues is provided in the instruction package included as an annex of this report. Unfortunately, none of the websites had implemented tracking of data downloads. One NSO required extensive assistance in resolving issues related to granting access to Google Analytics. To not overburden NSO staff, no further changes in their website were suggested.

In some cases, the data appearing in this report differs from raw Google Analytics reports because extraordinary traffic from the Reuters network has been excluded. The initial analysis of the geographic origins of users showed that the United Kingdom was one of the largest sources of web traffic on some websites, even surpassing

traffic originating in the country itself. However, the bounce rate for this traffic was high, in some cases over 90 percent, suggesting that whoever accessed the site from the Reuters network left without interacting with the page or navigating to another page. The unusual activity took place from July to November 2017 and originated from London. Further analysis revealed that it came from the Reuters network. Although it is possible to identify the national origin of most visitors, IP addresses are not reported by Google Analytics, so individual visitors cannot be identified. However, when traffic originates from a network that is clearly labeled, as in the case of Reuters, it is possible to create filters that trap the traffic from that source. Why Reuters generated so many web visits is not known, and the pattern was not familiar to the web analytics experts at Forum One. To keep the data analysis from being skewed, Reuters web traffic was excluded.

How can the Measuring Data Use report improve access to data?

We conducted this study to help NSOs and their partners understand the benefits of web analytics and their ability to improve data accessibility and dissemination online. This report provides specific and actionable recommendations that can be implemented by web managers to improve the accessibility of online data for users.

Key Performance Indicators

Audience Analysis

Overview: *Measuring web analytics sounds complicated – but useful and actionable data can be easily understood by looking at a few basic metrics.*

Basic indicators of website traffic, including the number of users visiting the site, the number of sessions that the users have initiated, the number of page views occurring through those sessions, and the bounce rate provide useful overview of site traffic. This section analyzes these indicators for each national statistical office (NSO) website and data portal included in the study and can be seen in Table 1.

The number of users shows how many uniquely identified individuals have visited the site during the studied period, although users accessing the site with multiple devices may be counted more than once. Considerable variation exists between the sites. Websites that belong to countries with larger populations tend to receive a larger number of visitors. The seven NSO sites received an average of 1.1 million users with a range of 4.2 million at the highest and 91,000 at the lowest. NSO sites receive much higher web traffic activity than data portals. The three data portals received an average of 39,265 users with a range of 59,000 at the highest and 3,900 at the lowest. An NSO site serves many functions in addition to providing data, while the dedicated data portals serve the sole function of providing access to data, which limits the type of users that would seek to access the site.

The number of sessions shows how many times users loaded the website, and the number of page views shows how many times the pages throughout the site were accessed. The seven NSO sites received an average of 2.0 million sessions with a range of 6.5 million at the highest and 177,000 at the lowest. They received an average of 5.7 million page views with a range of 15.9 million at the highest and 544,000 at the lowest. The three data portals received an average of 71,000 sessions with a range of 122,000 at the highest and 7,000 at the lowest. They received an average of 527,000 page views with a range of 1.0 million at the highest and

30,000 at the lowest. While the number of users provides an understanding of how many people accessed the site, the number of sessions and page views represents the frequency and depth at which the website was accessed.

Key Terms

Users: A new or returning individual accessing the website.

Session: A group of user interactions with a website that take place within a given time frame.

Page views: The number of times a page was loaded on a website during a session.

Bounce rate: The proportion of users that arrive on a website page and immediately leave without navigating further into the website.

The bounce rate shows the proportion of users that arrive on a website page and immediately leave without going farther into the site. The bounce rate may capture several different scenarios: users may arrive on the website by accident; they may be checking for an update; or they may be interested only in the information on that page. The seven NSO sites experienced an average bounce rate of 47.4 with a range of 56.0 at the highest and 39.7 at the lowest. The three data portals experienced an average bounce rate of 34.1 users with a range of 45.5 at the highest and 25.2 at the lowest.

TABLE 1: Number of users, sessions, page views, and the bounce rate

Website	Users	Sessions	Page views	Bounce Rate
Country 1 – NSO Site	4,182,770	6,464,157	15,915,032	49.0
Country 1 – Data Portal	58,507	84,131	531,738	31.6
Country 2 – NSO Site	925,329	1,564,049	4,882,915	48.0
Country 3 – NSO Site	231,897	545,007	1,820,504	39.7
Country 3 – Data Portal	55,387	121,837	1,017,674	25.2
Country 4 – NSO Site	320,576	524,435	1,355,585	47.1
Country 5 – NSO Site	2,031,370	4,183,264	14,809,298	46.8
Country 6 – NSO Site	104,450	177,102	613,766	47.6
Country 6 – Data Portal	3,901	6,989	30,379	45.5
Country 7 – NSO Site	91,439	184,933	543,570	56.0
Average – NSO Sites	1,126,833	1,948,992	5,705,810	47.7
Average – Data Portals	39,265	70,986	526,597	34.1

Duration and depth: *Our study discovered that users explore data portals in more depth than NSO websites. The deeper they go, the longer users remain in a session, but they take slightly less time to navigate through data portals.*

Table 2 shows the number of sessions by their duration. Some users access the website only briefly (0-10 seconds), indicating use without exploration. Others explore the website for a few minutes (11-180 seconds), indicating moderate exploration. And some access the website for extended periods of time (180 seconds or more), indicating in-depth use. However, some of these users may have opened the website and forgotten to close it, resulting in inflated duration times.

Key Terms

Session Duration: Amount of time a user spends on the website during a session.

Session Depth: Number of pages a user clicks during a session.

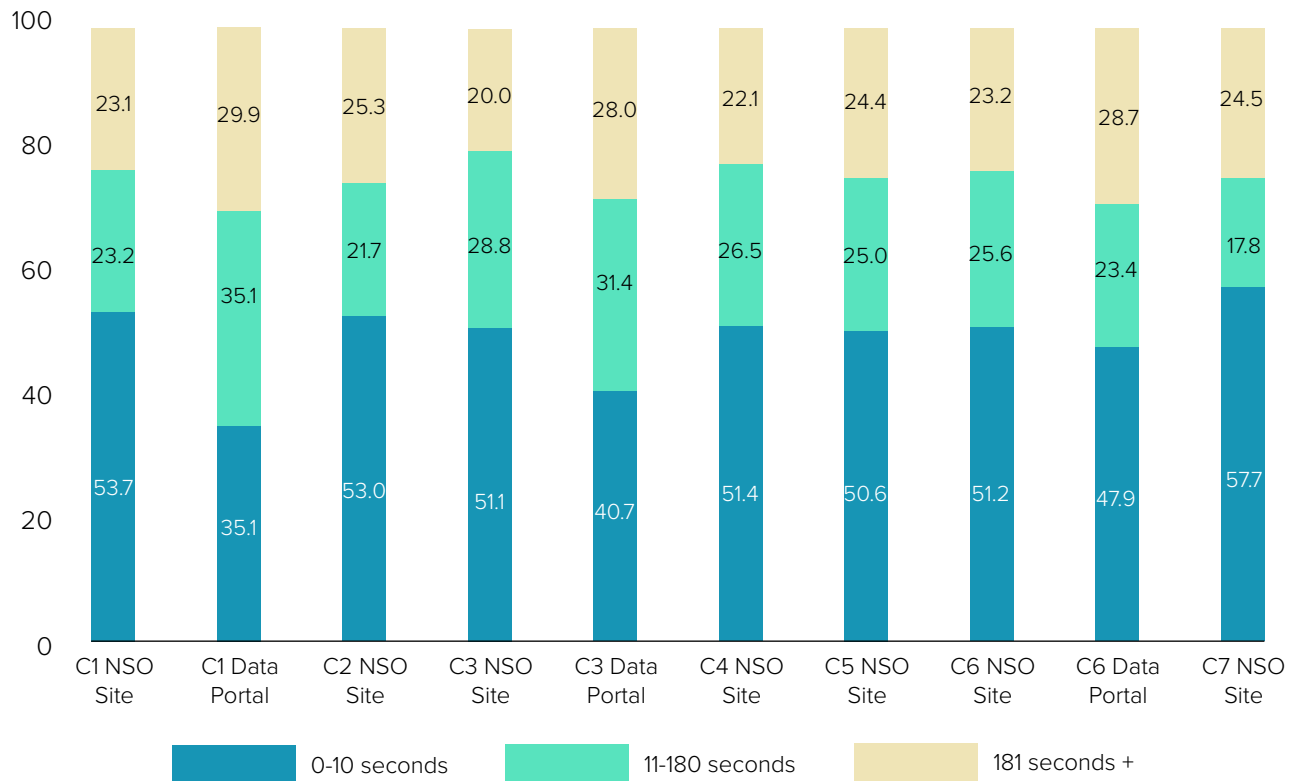
TABLE 2: Number of sessions by duration

Website	0-10 sec	(%)	11-180 sec	(%)	181+ sec	(%)
Country 1 – NSO Site	3,471,177	53.7	1,501,052	23.2	1,491,928	23.1
Country 1 – Data Portal	29,498	35.1	29,504	35.1	25,129	29.9
Country 2 – NSO Site	828,726	53.0	338,935	21.7	396,388	25.3
Country 3 – NSO Site	278,585	51.1	157,231	28.8	109,192	20.0
Country 3 – Data Portal	49,528	40.7	38,240	31.4	34,069	28.0
Country 4 – NSO Site	269,305	51.4	139,074	26.5	116,056	22.1
Country 5 – NSO Site	1,329,683	50.6	655,784	25.0	641,927	24.4
Country 6 – NSO Site	90,623	51.2	45,320	25.6	41,159	23.2
Country 6 – Data Portal	3,348	47.9	1,636	23.4	2,005	28.7
Country 7 – NSO Site	106,719	57.7	32,875	17.8	45,339	24.5
Average – NSO Sites	910,688	52.7	410,039	24.1	405,998	23.3
Average – Data Portals	27,458	41.2	23,127	30.0	20,401	28.8

For the seven NSO sites, the majority of sessions lasted 10 seconds or less. On average 52.7 percent of sessions lasted for 0-10 seconds; 24.1 percent of sessions lasted for 11-180 seconds; and 23.3 percent of sessions lasted for more than 181 seconds. For the three data portal sites, 41.2 percent of sessions lasted for 0-10 seconds; 30.0

percent of sessions lasted for 11-180 seconds; and 28.8 percent of sessions lasted for more than 181 seconds. This shows that on average, users explored for a longer time on the data portals sites than on the NSO sites, suggesting greater depth of usage. Figure 1 visualizes the distribution of sessions on each site.

FIGURE 1: Distribution of sessions by duration



The number of pages viewed during a session provides a complementary measure of the depth of exploration. Exploration of one page indicates sessions during which a user clicked to only one other page before exiting the site. Sessions involving two to four pages indicate moderate exploration of the site. Sessions that involved five or more pages reflect in-depth exploration of a site content. As Table 3 shows, for the seven NSO sites, an average of 49.5 percent of sessions involved exploration of one page, while

an average of 32.8 percent of sessions involved two to four pages, and an average of 17.7 percent of sessions involved five or more pages. For the three data portal sites, an average of 35.7 percent of sessions involved exploration of one page, while an average of 27.3 percent of sessions involved two to four pages, and an average of 37.1 percent of sessions involved five or more pages. Figure 2 shows the variation in the proportion of in-depth website usage across the various websites.

TABLE 3: Number of sessions by page depth

Website	One or fewer	(%)	Two to four	(%)	Five or more	(%)
Country 1 – NSO Site	3,725,855	57.6	1,961,199	30.3	777,104	12.0
Country 1 – Data Portal	30,539	36.3	21,235	25.2	32,357	38.5
Country 2 – NSO Site	750,478	48.0	530,901	33.9	282,670	18.1
Country 3 – NSO Site	216,163	39.7	215,026	39.5	113,819	20.9
Country 3 – Data Portal	30,687	25.2	35,222	28.9	55,928	45.9
Country 4 – NSO Site	249,230	47.5	202,954	38.7	72,252	13.8
Country 5 – NSO Site	1,958,892	46.8	1,362,422	32.6	861,950	20.6
Country 6 – NSO Site	89,632	50.6	48,442	27.4	39,028	22.0
Country 6 – Data Portal	3,178	45.5	1,931	27.6	1,880	26.9
Country 7 – NSO Site	103,593	56.0	50,675	27.4	30,665	16.6
Average – NSO Sites	1,013,406	49.5	624,517	32.8	311,069.7	17.7
Average – Data Portals	21,468	35.7	19,463	27.3	30,055.0	37.1

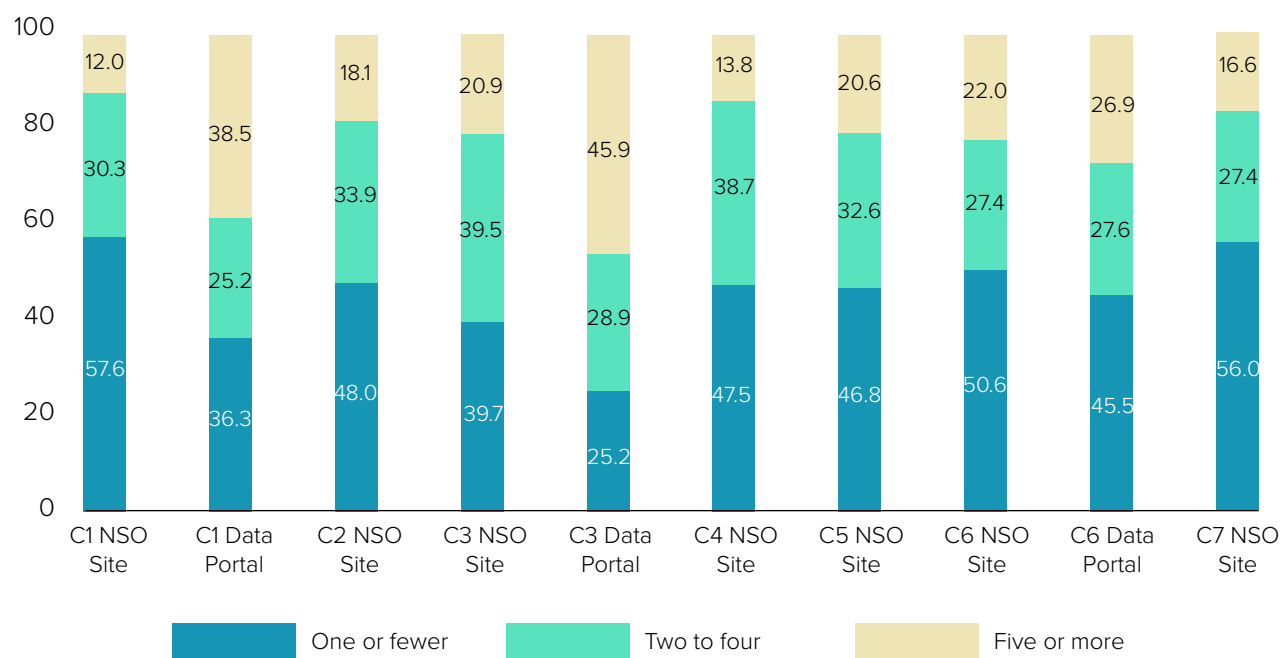
FIGURE 2: Distribution of sessions by page depth

Table 4 shows the average of duration of sessions by page depth. Across NSO sites, one or fewer pages are viewed for an average of 3 seconds. Sessions that involve an exploration of two to four pages last for an average of 4 minutes and 9 seconds. And sessions that involve five or more pages last for an average of 14 minutes

and 17 seconds. Across data portals sites, one or fewer pages are viewed for an average of 0 minutes and 5 seconds. Sessions that involve an exploration of two to four pages last for an average of 2 minutes and 58 seconds. And sessions that involve five or more pages last for an average of 11 minutes and 57 seconds.

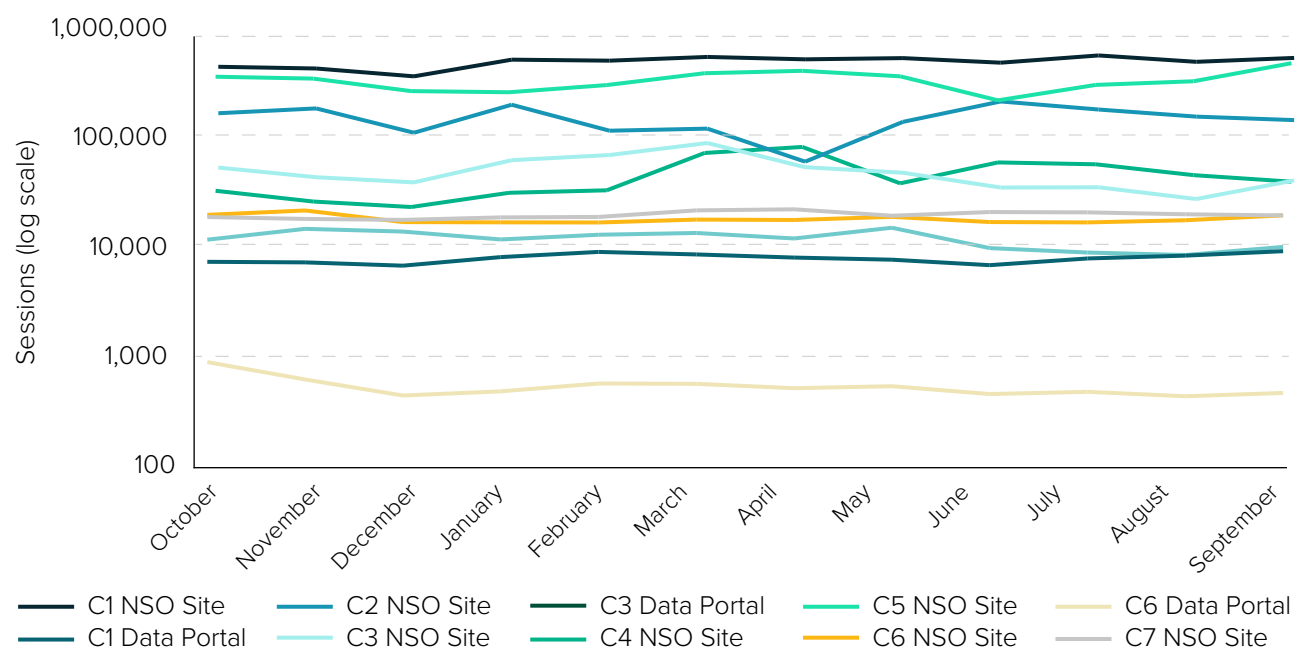
TABLE 4: Session duration by page depth

Website	One or fewer	Two to four	Five or more
Country 1 – NSO Site	0:00:13	0:05:13	0:16:59
Country 1 – Data Portal	0:00:14	0:03:33	0:10:02
Country 2 – NSO Site	0:00:01	0:04:09	0:15:26
Country 3 – NSO Site	0:00:01	0:02:31	0:10:56
Country 3 – Data Portal	0:00:00	0:01:28	0:10:27
Country 4 – NSO Site	0:00:01	0:03:44	0:12:45
Country 5 – NSO Site	0:00:01	0:04:11	0:15:23
Country 6 – NSO Site	0:00:05	0:04:14	0:12:58
Country 6 – Data Portal	0:00:00	0:03:53	0:15:21
Country 7 – NSO Site	0:00:01	0:05:04	0:15:32
Average – NSO Sites	0:00:03	0:04:09	0:14:17
Average – Data Portals	0:00:05	0:02:58	0:11:57

Variation over time: Most websites experience seasonal variation with lower traffic during holiday periods and higher traffic with new data releases and publications.

Understanding the patterns of variation that a website experiences will help provide context for web traffic data viewed in a specific month. Figure 3 shows that most websites experience a reduction in the number of sessions during the months of December and January. This may reflect reduction in working hours during holidays or schools being out of session. Other drops may reflect issues such as website down time.

Viewing how the number of sessions rise and fall across the months is also an effective way to view the success of dissemination efforts. After releasing key statistical publication there should be an increase in the number of sessions, reflecting the activity of users seeking to access the new publication. Some of the months during which the overall web traffic is higher are the result of sharp spikes in user activity during a few days that may coincide with data releases or other events.

FIGURE 3: Number of sessions per month

Audience engagement channels: NSO websites were most often reached by organic search; data portals were accessed directly or by referral from another website.

Engagement channels record how users find their way to a given website: through a search engine, a bookmarked link, a link shared through social media, or other routes. Table 5 provides a breakdown of the channels through which sessions begin on each website. Figure 4 further illustrates the variation in engagement channels.

Across the seven NSO sites, an average of 66.4 percent of sessions began through organic searches. This indicates sessions initiated by users who found the NSO website using keywords through a search engine such as Google or Bing. There are many users who know the website name, but not the address and therefore access it through a search engine. For the three data portals, an average of 14.2 percent of sessions began through organic searches.

Many sessions also began through direct channels when a user typed the URL into the browser or clicked on a bookmarked link. For the seven NSO sites, an average of 26.1 percent of sessions began through direct channels, and 54.0 percent of sessions for the three data portals. Accessing a site through direct channels suggests more purposeful usage, because the user must already know how to access a given site, whereas accessing a site through organic search suggests users less familiar with how to access a site.

Sessions that began through referral involve users that clicked on a link from another website to reach the NSO website. This does not include links through major search engines, which are counted as organic searches. On average, data portal sites have a much higher percentage of referrals than NSO sites. For the seven NSO sites, an average of 6.4 percent of sessions began through referrals, while 30.7 percent of sessions on the three data portals began through referrals. This most likely a result of data portal users being referred through the country's NSO site.

Few sessions began through social, email, or other channels. Of those that did, most came from social media, such as Facebook or Twitter. The “other” channel refers to sources that Google Analytics is not able to classify. The fewest number of sessions began through email, which may be the result of limited email outreach from NSOs or because of the correct Google Analytics tags that would report data were not included in emails.

Audience Engagement Channels

Direct: Typing URL into browser or clicking bookmarked link.

Email: Links clicked in emails using special Google Analytics tags.

Organic Search: Using search engines like Google, Bing, or Yahoo.

Other: Sources Google Analytics cannot classify.

Referral: Links clicked on another website.

Social: Links clicked through social media, like Facebook or Twitter.

TABLE 5: Number of sessions by channel grouping

Website	Direct	Email	Organic Search	Other	Referral	Social
Country 1 – NSO Site	14.3	0.0	79.5	0.0	5.0	1.3
Country 1 – Data Portal	65.9	0.0	31.2	0.0	0.7	2.2
Country 2 – NSO Site	21.7	0.0	73.3	0.0	3.9	1.1
Country 3 – NSO Site	29.3	0.0	61.3	0.0	7.8	1.7
Country 3 – Data Portal	11.0	0.0	0.5	0.0	87.5	1.0
Country 4 – NSO Site	29.0	0.0	64.4	0.5	4.4	1.7
Country 5 – NSO Site	19.2	0.0	63.8	0.0	16.7	0.4
Country 6 – NSO Site	31.8	0.0	64.5	0.0	3.0	0.7
Country 6 – Data Portal	85.1	0.0	11.0	0.0	3.7	0.1
Country 7 – NSO Site	37.5	0.0	57.7	0.0	4.1	0.7
Average – NSO Sites	26.1	0.0	66.4	0.1	6.4	1.1
Average – Data Portals	54.0	0.0	14.2	0.0	30.7	1.1

FIGURE 4: Percent of sessions by channel grouping



Geographic origins: Across all sites, users most often come from within countries hosting the websites.

As Table 6 shows, for the seven NSO sites, the average percentage of users coming from within the country is 73.4, while for the data portals, the average is 65.4. The only instance in which less than half of web traffic came from domestic sources was the NSO site for country 6. This high rate of international web traffic suggests

that the country may have lower domestic capacity to access data and or receives support from an especially high number of international sources, resulting in their need to access data. Figure 5 shows the variation in the percentage of domestic users across the NSO websites and data portals.

TABLE 6: Number of domestic users and sessions

Website	Domestic Users	Proportion of all users (%)	Domestic Sessions	Proportion of all sessions (%)
Country 1 – NSO Site	3,489,081	83.0	5,477,639	84.7
Country 1 – Data Portal	49,328	83.8	70,734	84.1
Country 2 – NSO Site	773,515	83.9	1,377,374	88.1
Country 3 – NSO Site	146,039	62.5	431,475	79.2
Country 3 – Data Portal	34,569	62.3	81,532	66.9
Country 4 – NSO Site	265,977	82.4	434,597	82.9
Country 5 – NSO Site	1,772,764	87.5	3,799,943	0.8
Country 6 – NSO Site	50,443	49.2	104,252	58.9
Country 6 – Data Portal	1,988	50.0	3,636	52.0
Country 7 – NSO Site	60,211	65.1	134,979	73.0
Average – NSO Sites	936,861	73.4	1,680,037	79.7
Average – Data Portals	28,628	65.4	51,967	67.7

FIGURE 5: Number of domestic users



Further analyzing the geographic origins of users may yield additional insights regarding user characteristics. It is even possible to view the data by city rather than by country. Unfortunately, providing a full list of the top international sources of web traffic to each site would compromise the anonymity of the countries participating in this study because most countries received a great deal of web traffic from neighboring countries, regardless of the income level of these countries.

The analysis of the geographic origins of users revealed the irregularity in the data stemming from the Reuters network. The raw data from Google Analytics for most websites showed the United Kingdom as providing the second highest number of users, with a bounce rate of more than 90 percent in most cases. Further analysis revealed that this was due to an unusual increase in activity coming from the Reuters network from July to November 2017, almost all of which immediately left. It is not clear what this spike in user activity from Reuters means. To provide data that show the behavior of actual users, the web traffic data from the Reuters network has been filtered out of the analysis.

Content Analysis

Overview: Web traffic analysis provides evidence that websites and data portals are being used to access a wide variety of statistical data.

Are website visitors looking for data and do they find what they are looking for? Without knowing the content of every page, it is impossible to draw a definitive conclusion, but web traffic data provides useful insights.

To estimate the share of traffic going to data pages on NSO websites, the study examined the 30 pages receiving the most traffic, not including the home page. The number of page views these top 30 pages received reflected 24 to 84 percent of all page views on the site. This shows that even in cases of the smallest range of percentages, the 30 pages still reflect a significant portion of web traffic. See table 7. Among the top 30 pages on

each site, the share of page views on data-related pages ranged from 18 to 97 percent. NSO sites with a high percentage of data pages viewed had a more specialized focus on providing access to statistics, functioning much like a data portal. But other sites provided other services that attracted a large share of total traffic, such as access to official records, information on employment opportunities, and forms for uploading data. (Although these are certainly data-related, in this analysis we only considered pages used to disseminate data.) These percentages based on the 30 most heavily visited pages are not necessarily representative of all traffic on the sites, but they confirm that along with the less heavily used data portals, the NSO sites are important vehicles for data dissemination.

TABLE 7: Share of page views among top 30 pages

Website	Total page views excluding home page	Page views of top 30 Pages	Top 30 share of all page view (%)	Top 30 page views for data pages	Share of top 30 page traffic going to data (%)
Country 1	14,296,742	3,879,733	27.1	969,715	25.0
Country 2	-	2,054,124	-	1,615,857	78.7
Country 3	1,283,971	1,076,738	83.9	195,066	18.1
Country 4	1,072,922	591,991	55.2	311,094	52.6
Country 5	12,795,843	3,058,421	23.9	2,981,735	97.5
Country 6	533,729	204,809	38.4	159,665	78.0
Country 7	388,987	154,954	39.8	134,980	87.1

It was a challenge to identify the volume of web traffic directed solely to data-related pages because not all sites used a clear directory structure in their URLs. A clear directory structure on a website organizes all data under /statistics/ or a similar key word in the URL, then organizes them under categories and themes using descriptors in the URLs that describe their contents, such as, /statistics/education/ or /statistics/health/. On many

of the sites, the page URLs are differentiated only by a serial number, necessitating a closer examination to determine their contents. Another challenge to the analysis of data use is that none of the websites participating in this study were set up to track data downloads. Moreover, some websites have been designed using iframes to provide access to data, which make it impossible to ever track data downloads.

Data content: Pages with economic statistics were most frequently visited

The contents of the top 30 pages were identified by examining each page and categorized to show trends in data access by page contents. Table 8 shows the results. Many pages provide access to publications such as statistical yearbooks that contain a variety of statistics. These were classified as “general statistics.” “Economic statistics” include national accounts, price indexes, employment statistics, and foreign trade data. “Social statistics” include demographic data, poverty statistic, and health and education data. The pages classified as “environment statistics” included agricultural

statistics, geographic information or maps. Pages in the top 30 that did not provide data, were classified as “other.” With the exception of Country 2, pages with economic statistics were the most often visited among those pages that could be classified. No pages were found that included environmental indicators such as land use data, pollution indicators, or indicators of the built environment. This may reflect the interest – or lack of interest – of website visitors but may also reflect a lack of data available on the websites.

TABLE 8: Share of data pages by contents (%)

Website	Economic statistics	Environment statistics	Social statistics	General statistics	Other
Country 1	11.3	0	10.3	3.3	75
Country 2	18.2	2.1	41.3	17.2	21.3
Country 3	10.5	0	8.2	9.7	71.6
Country 4	29.5	0	1.2	21.9	47.4
Country 5	34	0	16.2	45.8	4
Country 6	27.4	3.7	4	42.8	22
Country 7	52.3	7	7	20.8	12.9
Average	21.1	0.6	17.2	20.8	40.4

Besides the number of page views, other measures may be useful for gauging user interest in these pages such as the exit rate, bounce rate, and time spent on page. The interpretation of these measures and their usefulness in assessing how users respond to the site differs depending on the nature of the page. A low exit rate is good for pages that guide users to access data on pages located elsewhere on the site, because it means that users most likely continued on to access those

pages. A higher exit rate on a page that provides direct access to data may mean that users have discovered the data they need and left the site with their needs satisfied. A high bounce rate on a page providing data could indicate that users have a that page bookmarked and regularly check it for updates. More time spent on a page could be positive if it includes a great deal of information, but it could also indicate a page that users find difficult to navigate.

Another way to identify pages that are important to data users is to look at entry and exit data for the overall site. Data-related pages on which higher numbers of sessions begin suggest that users arrived at the site searching specifically for those

data. Exit rates may indicate where users gave up on trying to find what they needed, but they also indicate where users were able to find what they were looking for, which helps NSOs identify types of data that are in the highest demand.

Data search activity: Site search analytics provides further evidence that users are most interested in economic data.

Although only a small segment of users accesses the search functionality of a website, analyzing web traffic related to search activity can provide valuable insight into the data-related needs of users. Keywords that users have typed in order to find data are a direct message from users asking for those data. They also suggest which types of data may be difficult to find, leading them to the use of the website's search functionality in the first place.

Only four of the ten websites included in this study had search data to analyze. Although the availability of website search activity is limited, it is possible to carry out some initial analysis. The data in Table 7 show that few sessions involved the use of the

site search functionality. The average percent of sessions that involved site search on the sites below is 1.9. Visitors viewed a search result page 1.5 times after performing a search. Following a search, the average number of pages viewed was 4.6 with 2.2 being the lowest number of pages and 7.7 being the highest. The highest number came from the data portal for which search data were available. The average duration of a session following a search was 6 minutes and 6 seconds, which suggests that sessions that involve site search correspond to more in-depth exploration of the website. The exit rate of 20.9 percent may reflect users who upon viewing the results did not find what they were looking for and gave up searching.

TABLE 9: Website search activity

Website	Sessions with search	Share of all page views (%)	Page views/search	Average search depth	Time after search	Exit rate (%)
Country 1 – NSO Site	3,615	0.1	1.3	3.7	0:06:43	18.4
Country 3 – NSO Site	6,214	4.2	1.1	2.2	0:03:55	18.6
Country 3 – Data Portal	1,236	3.4	2.4	7.7	0:05:57	19.8
Country 5 – NSO Site	6,426	0.2	1.1	4.6	0:07:48	26.6
Average – All	4,373	1.9	1.5	4.6	0:06:06	20.9

Note: For Country 1 and Country 5, search data are available October 1, 2017 to September 30, 2018, and for Country 3, data are available May 17, 2018 to September 30, 2018.

The data-related search terms from the four websites with a full year's data reflected similar findings from the analysis of the key data pages. Table 10 shows the top five data-related searches for each website. The exact terms have been adjusted to remove any words that might identify the website or country to preserve anonymity. On the Country 1's NSO site, "calendar," which leads to release-dates for data, was the most popular search term, accounting for 15.4 percent of all searches.

For the Country 3's NSO Site and Data Portal, the most popular search term was GDP, accounting for 1.9 percent and for 0.9 percent of searches respectively. For Country 5's NSO Site, the most popular search term was GDP, which accounted for 4.1 percent of searches. Most data-related search terms refer to economic statistics such as GDP and foreign trade. There is additional interest in data-release dates and data on poverty and population.

TABLE 10: Data-related search terms

Website	Search Terms	Number of Searches	Percent of Searches
Country 1 – NSO Site	Calendar	701	15.4
	CRVS Data	26	0.6
	Poverty	13	0.3
	Population by Age	13	0.3
	Annual Survey	13	0.3
Country 3 – NSO Site	GDP	87	1.9
	Exports	36	0.8
	Calendar	34	0.7
	Tourism	34	0.7
	Inflation	24	0.5
Country 3 – Data Portal	GDP	43	0.9
	Salaries	13	0.3
	2017 Data	11	0.2
	Inflation	7	0.2
	Region	6	0.1
Country 5 – NSO Site	GDP	186	4.1
	Inflation	143	3.1
	Economic Growth	76	1.7
	Life Expectancy	59	1.3
	Exports	42	0.9

Data Collection Challenges: *This study identified common mistakes in the implementation of Google Analytics that, in some cases, made it difficult or impossible to draw definitive conclusions.*

As noted earlier, NSO websites offer more diverse content than data portals. Because of that, some users of an NSO website may ignore statistical offerings and look for news announcements, job openings, or other services provided by the NSO online. Tracking patterns of activity on NSO websites that lead to data pages can be difficult because most sites provide multiple paths for accessing data, with some hosting data in different formats at separate locations.

Tracking data use requires special implementation of an analytics program, and in some cases, may require some adjustments to the website itself.

The task is made more difficult when the website is not structured around easily identified landing pages or when the page URLs are made up of unrecognizable codes. A clear directory structure on a website organizes all data under a recognizable keyword, such as /statistics/ or a similar keyword in the URL, then organizes them under categories and themes using descriptors in the URLs that describe their contents, such as, /statistics/education/ or /statistics/health/. On many of the sites, the page URLs are differentiated only by a serial number, necessitating a closer examination to determine their contents. Most analytical features on Google Analytics come automatically enabled but tracking data downloads and the search activity of users must be separately enabled. None of the countries in the study were able to implement download

tracking. Moreover, some websites have been designed using iframes to provide access to data, which make it impossible to ever track data downloads.

Some participating countries already had site search tracking enabled and could provide more than a year of data for analysis while other countries were encouraged to enable this functionality. In most cases, enabling search functionality simply requires clicking a button under the settings of Google Analytics. However, there are a few instances where default search parameters are not included within the URL structure. As a result, some adjustment is needed in Google Analytics settings for tracking site searches. In some cases, custom coding may be needed to successfully enable this feature.

Conclusions

If the first rule of communication is “know your audience,” then web analytics should form the foundation for NSOs’ digital dissemination strategies. The analytic tracks left by website users provide evidence of what users are looking for and what they found. To better anticipate what users need, NSOs should consider using feedback forms, user surveys, focus groups, and advisory councils complemented by web traffic data. Changes made in response to user demands should always be reflected in the data. An effective web analytics and optimization program creates a robust feedback loop that connects data providers, content creators, website managers, and IT staff back to their users. This study reaffirms our belief that NSOs should view web analytics as a vital resource available to them to help disseminate data effectively.

The study also confirmed that NSO websites and data portals themselves are important vehicles for data dissemination. In our analysis of the 30 pages receiving the most traffic, not including the home page, 18 to 97 percent of those pages offered data-related content, while these pages represented 24 to 84 percent of all page views on the sites. This clearly demonstrates that NSO websites and data portals are effectively disseminating data to users who seek it.

Volumes of traffic ranged from 3 million on the most heavily used NSO website to 3 thousand on the least heavily used microdata portal. The number of visitors to NSO websites tends to be proportionate to the population of the country, which is consistent with the observation that a large share of the traffic to the websites originates in the country (exact proportions are not given because they could breach the confidentiality agreement with participating countries.) Across all sites, users accessing the websites most often come from within the country hosting the website. International visitors come predominantly from neighboring countries or from countries that have traditional ties of trade or aid flows with the host country. The

only website and data portal for which domestic users made up less than two-thirds of website visitors were located in a small country receiving extensive donor support.

The study confirmed that NSO websites and data portals themselves are important vehicles for data dissemination.

NSO sites often serve many functions in addition to providing data, which leads to users with a variety of needs from and expectations of the sites. Conversely, data portal sites serve a singular purpose and have a more limited type of user engaging with the sites. The study discovered that data portals had lower bounce rates than the NSO websites, suggesting more deliberate access of their content. On average, users explored data portal sites for longer and with greater page depth than NSO sites, which suggests users have greater interest in exploring their contents, or that it may be more difficult for users to find the content they are looking for. The data comparing the number of sessions by page depth also confirm that, on average, users explored not only for a longer time on the data portals sites, but also with greater depth.

The variation in the number of sessions over time reveals traffic changes due to seasonal patterns, important content updates, and other external factors. The decrease during the months of December and January may reflect reduction in working hours or school attendance, as many individuals observe holidays at that time of year. Decreases may also reflect periods of time when the website was down. Increases may reflect releases of key statistical publications. Website

managers should consider external factors such as these when assessing data from web analytics to gauge the impact of changes in their websites and their contents.

The study had difficulty determining the proportion of visitors sought and found data on both the NSO and data portal sites. Most sites provide multiple paths to accessing data, with some hosting data in different formats at separate locations throughout the site. It was a challenge to separate the web traffic to these pages because most sites did not provide a clear directory structure in their URLs. To help find the data-related sections of NSO sites, the Open Data Inventory (ODIN) was used to identify specific pages containing key data-related content. These included landing pages for publications or sections dedicated to providing statistics. Although this may not include all possible data-related pages, analyzing pages included in the ODIN assessments shows web traffic for pages that provide access to highly sought data.

Economic indicators were the most in demand content among website users. Pages containing employment statistics were among the most frequently accessed on the NSO sites. Foreign trade indicators, including export and import data, were also frequently accessed. Other frequently accessed economic indicators included GDP data and consumer price indices. Among social indicators, population data was the most frequently viewed, in some cases more frequently than economic data. Other social indicators concerned poverty, education, and health. The search terms from the websites with search tracking led to similar results. Among the most popular data-related search terms were “GDP,” “inflation,” “exports,” “population,” and “poverty.”

Improvements to the structure of websites and implementation of website analytics will increase the value of the data captured by web analytics. NSO managers and their website teams should use this information, along with other avenues for feedback, to respond to user demands for more and better data and to update and improve their websites and data portals accordingly.

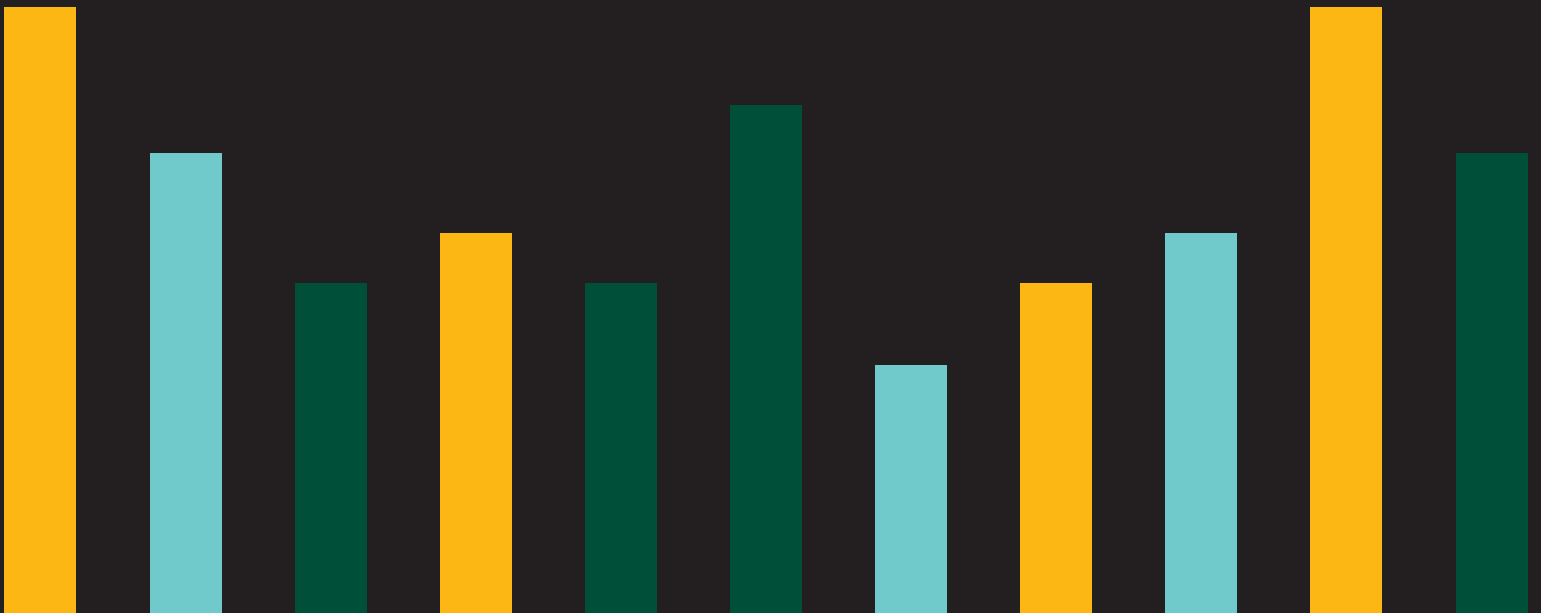
Annexes:

Annex 1: Recommendations for
Improving Web Traffic
Data

Annex 2: Web Traffic Dashboard

Annex 3: Customize Dashboard
Content

Annex 4: Instruction Package



Annex 1: Recommendations for Improving Web Traffic Data

The process of analyzing web traffic data for ten NSO websites and data portals revealed a number of opportunities to further improve the quality of data available through Google Analytics. These changes involve adjustments in the configuration of Google Analytics on the website pages as well as changes to the NSO websites and data portals themselves. Not all the recommendations will be relevant in every context, but those included here tend to be common issues across websites. In addition to tracking which recommendations may apply to which country participating in the study, Forum One provided an estimate of how frequently they have encountered these opportunities for improvement when working with other websites. Although these recommendations were developed through the analysis of the ten websites and data portals included in this study,

they are widely applicable and can be used by any NSO seeking to improve their site.

The recommendations below have been organized according to the effort expected to accomplish them and further sorted by the impact the changes will have on the quality of data obtained. Those that require the least effort can be implemented immediately, while those that require high effort should probably be saved for consideration until a site redesign is undertaken. More information on how to implement many of these recommendations is available in the Annex: Instruction Package.

Table A summarizes and organizes the recommendations according to the effort that each will require.

TABLE A: Summary of recommendations

One or fewer	Two to four	Five or more
Consistently install Google Analytics on all pages	Track data file downloads	Avoid the use of iframes
Filter bot traffic	Capture web traffic key to data sites	Ensure consistent use of language in URL structure
Maintain an up-to-date list of properties and views	Implement canonical URL structures	Create human-readable URL structures
Implement site search tracking	Create a single “roll-up” Google Analytics view	Create a clear directory structure
Filtering out internal traffic	Enable cross-domain tracking	
	Label each website page with a unique and clear title	
	Remove spaces in filenames or URL structures	

Effort required: Low

Consistently install Google Analytics on all pages

Analytics tracking codes must be installed across every page within a website. This is not a common issue, but when tracking codes are missing, it has a high impact on the quality of the data that Google Analytics collects. If the tracking code is missing across a single page or across a section of content, page views per session, entrance pages, bounce rate, and referral sources cannot be accurately recorded. One of the NSO websites participating in this study did not have Google Analytics installed on the home page or on a number of other key landing pages throughout the site. This is an extremely urgent issue that needs to be resolved.

Filter bot traffic

To gain the most accurate data within Google Analytics, known bot traffic must be filtered. This allows Google Analytics to collect the real traffic that comes from human users rather than tracking the activity of bots. This is also a very simple issue to fix. Under the management user interface, it is possible to simply check a box labeled "Exclude traffic from known bots and spiders." If this option is selected, then Google Analytics will filter out the activity of all known bots and spiders.

Maintain an up-to-date list of properties and views

Google Analytics properties and views must be clearly labeled and those no longer in use should be deleted. If it is not possible to delete them,

then titles should be updated to reflect the fact that they have been retired. This is a common issue that makes it challenging to identify the correct property and view to use when drawing conclusions about user behavior. Unless the properties and views have clear titles, outdated or incorrect ones may be included in the analysis.

Implement site search tracking

Most websites allow users to search for data using an internal search engine, and each website should track the web traffic related to these searches. Not only is it possible to view user activity leading up to and following the use of the search engine, site search tracking also allows analysis of the search terms used. This may reveal content that users are not readily able to locate or identify gaps within a site's content. In most cases, resolving this issue is a simple matter of enabling site search tracking. However, in certain circumstances, where search parameters are not included within the URL structure, the level of effort to implement site search can be high, as it may involve custom coding.

Filtering out internal traffic

An IP address filter should be used to keep from tracking internal traffic. This will prevent the Google Analytics data from being skewed by the activity of internal users updating and managing the site. Websites that receive lower web traffic will see a greater change to their analytic reporting than others, but it will ensure the data being reported is relevant and actionable.

Effort required: Moderate

Track data file downloads

It is necessary to track file downloads to monitor users when they access and download data files. Although this may be a more complex recommendation to implement, it is crucial for monitoring data use through NSO websites and data portals. Otherwise it is only possible to record what pages were visited on a site, but there is no way to tell whether a user accessed the data a page offered before leaving the site. It is possible to use Google Analytics' outbound file download tracking events to capture when a file download has occurred.

Capture web traffic to key data sites

It is important to capture how often users are traveling from one site to the other when data or key assets are hosted within other sites. To obtain a more complete understanding of usage, it is possible to use Google Analytics' outbound link tracking events to capture when a file download has occurred. Unless this is implemented, it will not be possible to record when a user leaves an NSO site to access data contained on another data portal.

Implement canonical URL structures

A website that contains a single page that is accessible through multiple URLs or different pages with similar content must have one of the URLs designated as canonical. In addition to impacting the analytics data for a site, having multiple URLs for the same page can dilute search rankings and prevent users from locating relevant content. A canonical tag lets search engines know which page should be indexed and which page can be ignored.

Create a single “roll up” Google Analytics view

If content is provided across multiple domains, at least one property and view should be created within Google Analytics where all domains feed into one set of data. This will provide a holistic view of how users are accessing content across the various data portals. This is particularly important for NSO websites that provide access to data through dedicated data portals. When creating this "roll-up" view across multiple domains, it is important to implement cross domain tracking to ensure data is accurately collected.

Enable cross-domain tracking

In cases where content is provided across multiple domains, cross-domain tracking should be enabled to gain a better understanding on where users came from, their path to the data, and how long they spent on the site. Without cross-domain tracking, user session duration, pages per session, landing page and entrance

source will not be accurately measured and reported through Google Analytics.

Label each website page with a unique and clear title

Each page of a website should have a title that is different from any other page and that clearly describes the content contained within it. Many websites use a default title on multiple pages across the site. Other websites may use simple titles such as “Table” or “Data” that don’t describe the content of the page. Unique and clear page titles serve two important purposes; first, it improves search engine optimization by allowing search engines to correctly index pages and second, it allows Google Analytics to correctly display data by page title. When pages have the same title, Google Analytics combines data for those pages. Without clear page titles, it is only possible to analyze data by URLs which are not always clear. Although relabeling website pages does not pose a high technical challenge, it may be a time-consuming process depending on the number of pages involved.

Remove spaces in filenames or URL structures

The filenames and URL structures should be a continuous string of characters and symbols. This is necessary to ensure that search engines can index URLs and to prevent them from inserting additional characters into the spaces. Dashes or hyphens should be used to separate words instead of spaces. Although correcting this issue is not technically difficult, it can be time consuming.

Effort required: High

Avoid the use of iframes

Where possible, iFrames should be avoided in favor of posting content directly within a page or using a downloadable document. Web designers use iFrames to change content without needing the user to reload the surrounding page. However, iFrames cannot be indexed by search engines and cannot be easily captured within analytics. They can also create accessibility issues. It is rare for a website to provide access to data through iFrames, but when this happens, it is impossible to track data file downloads.

Ensure consistent use of language in URL structure

When providing content within two or more languages, it is a best practice to append a variable to the page URL to quickly identify the user's language preference and to calculate the number of pageview content received across multiple languages. The level of effort it will take to address the issue will be affected by the method used to build and manage the site.

Create human-readable URL structures

Instead of a complex series of ID characters and symbols, URLs should include recognizable and descriptive keywords. This will help ensure that users can find the content and data they are seeking. Human-readable URL structures facilitate ranking within search engine results. To ensure users can find the content and data they are seeking, ensure each URL includes recognizable and descriptive keywords. Dashes or hyphens should be used to separate words. The level of effort will depend on the method used to build and manage the site.

Create a clear directory structure

Specific issues or sets of data should be grouped and organized under a clear directory structure. This applies to human-readable URLs that include words descriptive of the content of a given page. Creating this structure will facilitate the analysis of web traffic. A directory structure can group statistics by category and type. Thus, a dataset under /statistics/ could be further grouped under subheadings like /health/ or /education/. Example: <https://www.example.net/statistics/health/2017/dataset-name> Having such a directory structure will make it possible to analyze web traffic related to specific categories and subsets of data.

Annex 2: Web Traffic Dashboard

Each country participating in this study has received access to a dashboard presenting a selection of web traffic data for their data portal available through their Google Analytics accounts. A template for this dashboard is available for anyone to access and use to create their own dashboard. Like the country dashboards, the template was built using Google Data Studio.⁷ The instructions given below for setting up the dashboard template may also be of help to website managers who wish to further customize an existing dashboard.

Because an enormous amount of data is available through Google Analytics, accessing web traffic data directly through Google Analytics can be a time consuming process, particularly given the need for selection and configuration every time the data need to be accessed. The dashboard template presents a selection of indicators that are relevant to monitoring user activity on a

website or data portal. Using this dashboard is as simple as opening a document on Google Drive. (Any dashboard created by Google Data Studio will appear within Google Drive.) It is possible to further customize each dashboard to reflect the monitoring needs for a particular website.

Although the dashboard displays Google Analytics data, the data displayed through Data Studio may differ from data returned by Google Analytics due to sampling.⁸ Sampling may happen more frequently and result in larger gaps when a site experiences hosts more than 100,000 sessions within a reporting period. Rather than provide pin-point accuracy, the goal of the dashboard is to uncover trends and provide tangible, actionable data on how users interact with the website that can be used to inform decisions about how to optimize web properties to provide easier access to the data users are seeking.

Accessing Dashboard Template

To use the template, it is necessary to have a Google-enabled email address. This is also necessary for accessing Google Analytics itself, so it may be convenient to use the email address associated with the Google Analytics account to access the dashboard template.

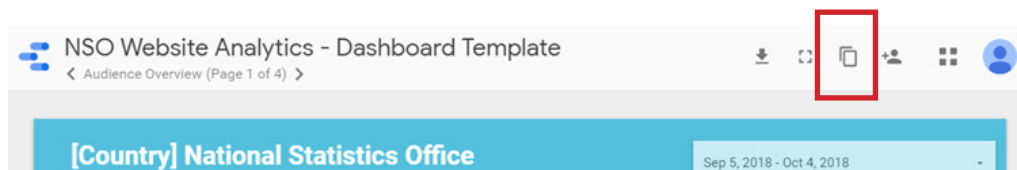
The dashboard template presents sample Google Analytics Data as a placeholder. The contents of the dashboard will adjust according to the data source. A number of sections in the template are blank, which is due to the need for initial configuration of dashboard settings.

Step 1: Open the dashboard template

The link for accessing the dashboard template is: https://datastudio.google.com/open/1a4pVANVYrQVwiaPM_PTSHSjbJwUHkwln

Step 2: Make a copy and connect to new Google Analytics data

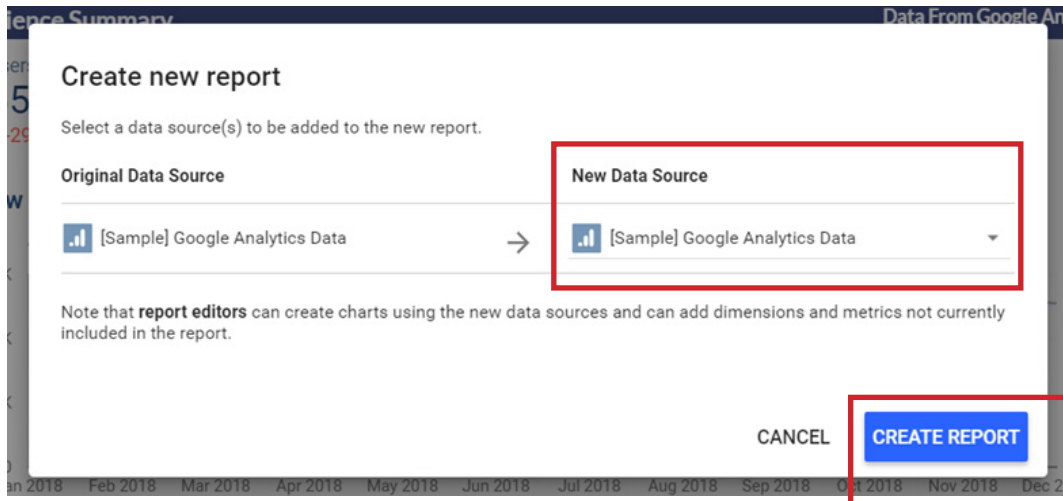
The dashboard template is set as public so that anyone can view it. Initially it displays dummy data from Google. To connect it with data from a new Google Analytics account, make a copy by clicking the copy icon in the upper right-hand corner as highlighted in the screenshot below.



⁷ A general introduction video to Google Data Studio is available at: <https://www.youtube.com/embed/6FTUpceqWnc>

⁸ More information on sampling is available at: <https://support.google.com/analytics/answer/2637192?hl=en>

Clicking this icon opens a dialogue window that invites the user to connect the dashboard template to a new source of data. If the template is being accessed from an email account associated with a Google Analytics account, it will be possible to select the relevant account and view from the drop-down menu. Once the new data source is selected, select Create Report.



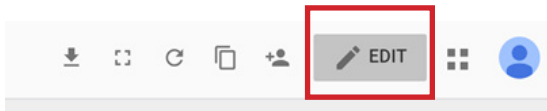
Step 3: Configure dashboard content for new data

The content of the dashboard will update automatically to display web traffic data from the new Google Analytics account. Please note that the Search Usage page (page 4) may be blank if search tracking is not enabled for the website or properly configured on Google Analytics. In certain circumstances where search parameters are not included within the URL structure, additional configuration of the search settings may be required. More information is available on this in the Instruction Package. Once search tracking has been enabled and properly configured, the Search Usage will automatically update.

Additional configuration of the dashboard is necessary for page three, Key Data Page Analysis. This is because it presents web traffic from selected website pages that provide access to official statistics, which will differ from site to site. Some countries may have a single landing page for accessing official statistics, while others may have a several that they wish to monitor. The first table on page 3, “Traffic for Data Landing Pages,” provides summary data for the key pages, while the following tables show the next table that web traffic moves to after viewing a particular landing page. The number of these “Next Page” tables should match the number of landing pages provided in the first table above.

Step 3.1: Updating Traffic for Data Landing Pages

Select EDIT in the header menu.



Select the table. This will change the content in the side panel to the right. Scroll down the Data column to the bottom to find the Filter menu. Select ADD A FILTER. A new menu of filters will appear. Select CREATE A FILTER at the bottom of the side panel.

The screenshot displays the Google Analytics interface for the [Country] National Statistics Office website. The main header shows the date range "Sep 5, 2018 - Oct 4, 2018". Below the header, the "Key Data Page Analysis" section is visible, followed by the "Traffic for Data Landing Pages" table.

Page Title	Pageviews	Avg. Time on Page	Bounce Rate	% Exit
Home	54,951	00:00:46	22%	20%
Google Online Store	26,018	00:01:22	59%	58%
Grand total	359,862	00:00:45	40%	21%

Below the traffic table, there are two sections labeled "Next Page After Viewing: [Data Landing Page Title]", each containing a table of page navigation data.

Page	Pageviews	Avg. Time on Page
/	72,623	00:00:57
/basket.html	18,468	00:00:45
/google+redesign/bags	12,168	00:00:40

The right sidebar shows the "Filter" section with a red box highlighting the "ADD A FILTER" button.

In the dialog box that pops up, create a name for the new filter, possibly [Country] Data Pages. The following steps will allow key data pages to be selected. On the Select a field menu, type “Page” into the Search fields. Scroll down to find and select the option that only says “Page”. On the Select a condition menu, choose Equal to (=). In the field that says, “example: value field,” type the URL information for a given data page. Note if the full address of a data landing page is www.example.org/statistics-page/, it is only necessary to include /statistics-page/ in this field. The result should appear as in the screenshot below. To add additional pages, select AND to repeat the process for other pages.

Create Filter ✕ CLOSE

Name

[Country] Data Pages

Data source

[Sample] Google Analytics

Inclusion

ABC Page

Equal to ...

/statistics-page/

OR

AND

Step 3.2: Updating Next Page After Viewing tables

In the dialogue box that pops up, create a name for the new filter that will make it easy for future users to identify it. On the Select a field menu, type “Previous Page Path” into the Search field and select this option from the list below. On the Select a condition menu, choose Contains. In the field that says, “example: value field,” type the URL information for the data page in question. As before, it is only necessary to include the sections of the URL following the main website address. The final result should appear as in the screenshot below.

Repeat this process for the other tables. To delete a table, select the table and press the backspace or delete keys on the keyboard. To create a table, make a copy by selecting the table and right clicking to access copy and paste options.

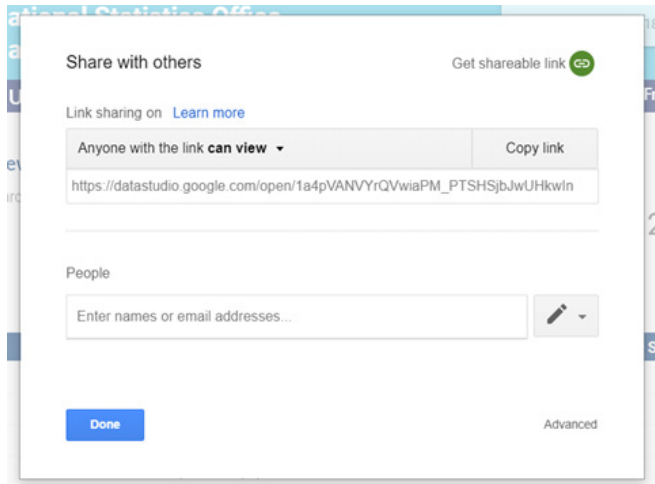
If there are too many tables and they become too large to fit on the dashboard page, it is possible to reduce the number of rows of data displayed on each table. Select a table and in the side panel to the right, access the Rows per Page menu. The default number of rows per page is 10, but it is possible to select fewer.

It is also possible to adjust the size of the table by selecting it and dragging the edges to be larger or smaller.

Step 4: (Optional) Share dashboard with others

The dashboard content may be shared with other users. To grant access to new users, open the dashboard and click the top right-hand icon of a person as seen in the screenshot below.

In the dialogue box that appears, enter the email address of the individual with whom the dashboard should be shared and select Done. The newly added person will receive instructions via email on how to access the dashboard.



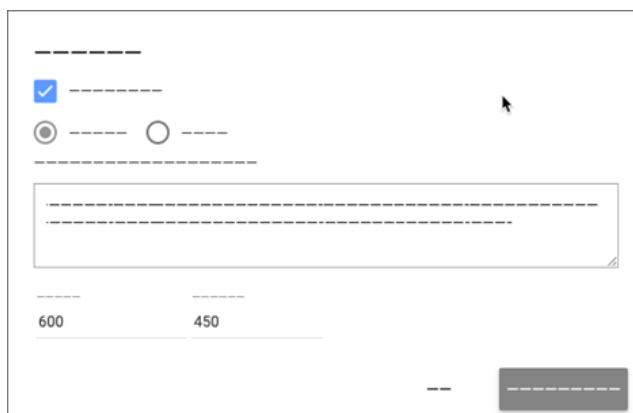
Please note, if the default privacy setting is “Anyone with the link can view” as appears in the screenshot above, there is no limitation on how widely the dashboard can be shared. To increase the privacy and security of the data, select “can view” and choose “OFF – only specific people can access.”

Embedding the Dashboard

It is also possible to make the dashboard fully public through embedding it on the NSO site itself. This would allow others to understand how the websites are being used and their official statistics are being accessed. It is possible to share the dashboard as broadly as possible through iframes, as these provide a window to the dashboard hosted on Google Studio. However, as discussed in the recommendations, iframes will not make it possible to track how users interact with the data.

Instructions:

1. Sign in to Data Studio.
2. Edit the report to be embedded.
3. Share the report:
 - For a public report, use the link sharing options "Public on the web" or "Anyone with the link can view"
 - For a private report, enter one or more users or Google Groups
 - G Suite users: To limit the report to users in your domain, turn link sharing on and select the "Anyone at [domain] with the link can view"
4. Select File > Embed report, or in the upper right, click Embed.
5. Click Enable embedding. A dialog box appears: Enable embedding dialog



Note: To embed a specific page of the report, navigate to that page before generating the embed link.

6. Ensure that the Embed Code option is selected. The iframe code appears in the text area below.
7. (Optional) Adjust the size of the iframe using the width and height controls
8. Click COPY TO CLIPBOARD. Note that the link generated by this dialog must be used. It isn't possible to embed a report using the standard shareable link.
9. Edit the HTML source of the page on which the report will be embedded. Paste the iframe code where the report should appear.

More information is available at: <https://support.google.com/datastudio/answer/7450249?hl=en>

Reviewing Dashboard Content

The dashboard contains four pages: Audience Overview, Content Analysis, Key Data Page Analysis, and Search Usage. Below is a brief review of the content of each of these pages.

Audience Overview

The Audience Overview page focuses on how many people are accessing the portal, where they are coming from, the method of finding the site, and how long they spend engaging with the site. All comparisons, unless otherwise noted, are made year over year.

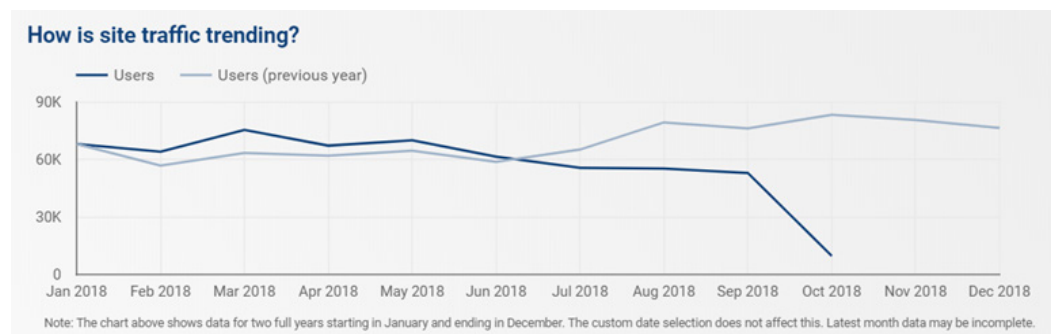
Audience Summary

The audience summary provides the three primary methods of measuring the reach of the site. “Users” tracks individuals. “Sessions” tracks how many times those individuals came back to the site. “Page views” indicates the overall number of pages those individuals accessed during their session. It also provides the bounce rate (how often visit only one page during their session) which is an overall indicator of site health.

Audience Summary			Data From Google Analytics
Users	Sessions	Pageviews	Bounce Rate
55,617	75,340	359,862	39.7%
↓ -29.6%	↓ -26.9%	↓ -13.4%	↓ -24.8%

User Trend Chart

The user trend chart provides a full year summary of how many users the site received each month compared to the previous year. This chart can uncover cyclical trends and can provide insights into how quickly the site is gaining in popularity over time. Because data for the current month are incomplete, the chart may show a sharp drop in the final month that will be corrected when complete data are available.

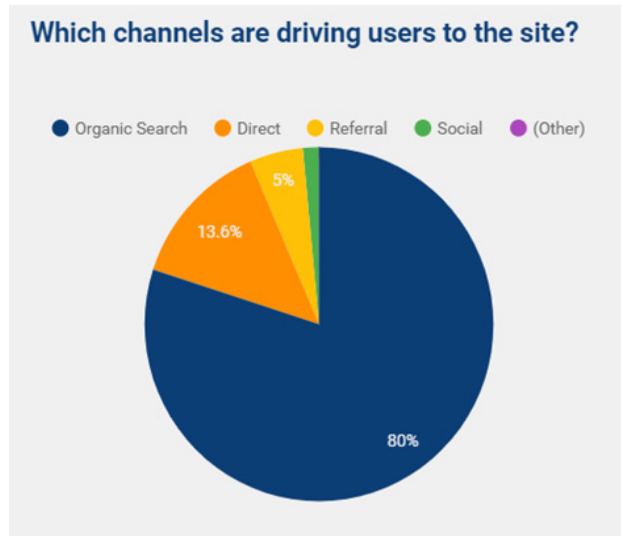


Channel Pie Chart

The channel pie chart shows how users are finding the site. The majority of users will be using search engines, which are reflected in the label “organic search”.

Channel Pie Chart

The channel pie chart shows how users are finding the site. The majority of users will be using search engines, which are reflected in the label “organic search”.



User Engagement Table

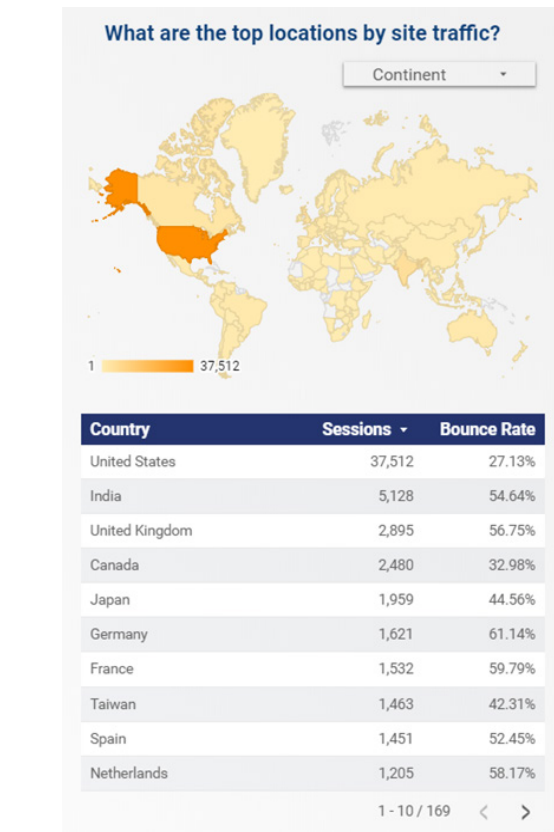
User engagement measured by page depth and time on site is an excellent measure of site health. While it is not possible to tell if users found what they are looking for, it is possible to see if there is a proportionate average session duration related to how many pages a user visited. If there are many pages visited in a short amount of time, this can indicate users are lost and quickly bouncing from page to page. An average of 1 min 30 seconds or longer per page would be a quality ratio.

What is the distribution of user engagement?

Visit a Single Page	Avg. Session Duration
346,035	00:00:12
Visit 2 Pages	Avg. Session Duration
102,787	00:03:47
Visit 3 to 4 Pages	Avg. Session Duration
82,502	00:06:25
Visit 5 or More Pages	Avg. Session Duration
82,522	00:15:42

User Location Map

The user location map can be used to view the site’s popularity by selecting a continent and viewing the number of sessions within each country.



Content Analysis

The Content Analysis page focuses on page-level interactions including how many pages individuals are accessing while exploring the portal, the pages they enter the site through, the most popular pages visited, and the last pages they view before leaving.

Audience Summary

The audience summary summarizes page-level interactions, including the total number of pageviews, the average time on page, the average number of pages accessed by a user during a single session, and the overall bounce rate. It's important to note that all users who bounced have viewed only one page before leaving the site, so their numbers skew the average number of pageviews per session downward.

Audience Summary		Data From Google Analytics	
Users	Sessions	Pageviews	Bounce Rate
55,617	75,340	359,862	39.7%
↓ -29.6%	↓ -26.9%	↓ -13.4%	↓ -24.8%

Top Entrance Pages

The top entrance pages table indicate how users are primarily finding and entering the site and how long they stay within the site before leaving. Review the top pages to ensure they serve well as a starting point for users entering the site. This includes having clear navigation and links to related content and data sets.

What are the top entrance pages to the site?		
Landing Page	Sessions ▾	Avg. Session Duration
/	43,350	00:03:00
/google+redesign/shop+by+brand/youtube	5,732	00:01:39
/google+redesign/apparel/mens/mens+t-shirts	2,764	00:02:56
/store.html?tpt=4532&sortci=orderscount+desc	2,644	00:02:28
/signin.html	1,906	00:02:16

Most Popular Pages

The most popular pages table lists the pages most frequently accessed by users and the average length of time spent within the page. Monitor this list to identify whether there are any “popular” pages that receive less than an average of one minute on page, because this may indicate that users are not finding what they hoped for and are leaving quickly.

What content is most popular with your audience?

Page Title	Pageviews ▾	Avg. Time on Page
Home	54,951	00:00:46
Google Online Store	26,018	00:01:22
The Google Merchandise Store - Log In	19,356	00:00:26
Shopping Cart	19,009	00:00:45
Bags Google Merchandise Store	18,571	00:00:32
Men's Apparel Google Merchandise Store	14,674	00:00:45

Top Exit Pages

The exit pages table displays the pages users most frequently visit just before leaving the site. A high exit rate for a page is not necessarily bad: it may indicate that users found the information or resources they came to the site to retrieve.

Where are people leaving the site?		
Exit Page	Sessions ▾	% Exit
/	25,643	47.86%
/google+redesign/shop+by+brand/youtube	4,352	46.34%
/google+redesign/apparel/mens	2,274	24.85%
/basket.html	2,261	9.7%
/google+redesign/apparel/mens/menst-shirts	2,143	29%
/google+redesign/bags	2,132	22.52%

Key Data Page Analysis

The Key Data Content Analysis page focuses on the site's key data landing pages. Some countries may have a single landing page for accessing official statistics, while others may have a several that they wish to monitor. This page provides summary data on all landing pages as well as web traffic to "next pages" viewed after a given landing page. These tables are affected by the structure and organization of a given site. If the data landing pages are not included within a specific directory, or section, of the site, they cannot be easily pulled out for examination.

Traffic for Data Landing Pages

The traffic for the data landing pages table provides summary data for all key pages. It shows the number of page views each landing page has received, the average time users spent on the page, the bounce rate, and the percent of sessions that ended on that page.

Traffic for Data Landing Pages				
Page Title	Pageviews ▾	Avg. Time on Page	Bounce Rate	% Exit
Home	54,951	00:00:46	22%	20%
Google Online Store	26,018	00:01:22	59%	58%
The Google+ Redesign Store - Home	19,955	00:00:45	40%	21%
Grand total	359,862	00:00:45	40%	21%

1 - 20 / 413 < >

Next Page After Viewing

The next page after viewing tables show web traffic to the other pages after viewing a landing page. It shows the number of page views each page receives as well as the average amount of time users spent on each page.

Next Page After Viewing: [Data Landing Page Title]		
Page	Pageviews ▾	Avg. Time on Page
/	72,623	00:00:57
/basket.html	18,468	00:00:45
/google+redesign/bags	12,168	00:00:40
/yourinfo.html	11,807	00:00:34
/google+redesign/apparel/mens	11,114	00:00:51
/signin.html?loginway=header	10,213	00:00:24
/google+redesign/shop+by+brand/youtube	9,634	00:01:04

Search Usage

The search usage page uncovers the top data-related searches conducted within the site and the pages users then went to. Review these results to see if users are landing on the best results that would correspond with their search. If it appears that there is other content that could be a better fit, review the page content to ensure it is properly labeled and contains keywords. It also presents the top data related searches that generated no results.

Search Overview

The search overview metrics provide information on user behavior in relation to the searches performed on the website. It provides the total sessions with search, the average number of times visitors viewed a search result page after performing a search, the average number of pages visitors viewed after performing a search, the average time spent on the website after search, and the percent of sessions that ended following a result from an internal search

Search Overview				
Sessions with Search	Results Pageviews / Search	Average Search Depth	Time after Search	% Exit
4K	1.1	3.4	02:53	20.8%

Search Terms

The search terms table shows the terms that users entered into the search engine in order to find content on the website that they were looking for. This table also shows the top page that each term led users to and the number of times users entered in each key term.

Search Terms		
Search Term	Search Destination Page	Total Unique Searches ▾
nest	(exit)	294
nest	/	166
nest	/signin.html?loginway=header	128
nest	/\${escatttitle}/quickview	119
gopher	(exit)	64

Annex 3: Web Traffic Dashboard

The dashboard template is provided as a starting place for countries to analyze the web traffic data most relevant to monitoring the access and use of an NSO website or data portal. It is possible to completely change the design and content of the dashboard according to the unique monitoring needs of each website.⁹

The sections below provide instructions from Google Data Studio support pages.

Change layout of the report

It is possible to control the way the dashboard appears on the viewer's screen using the options in the Layout tab of the Layout and Theme properties panel on the right-hand side of the screen. This is the default properties panel; it appears when no other component is selected.

View mode options

Header visibility

The header visibility options control the behavior of the report header in view mode. The header contains things like the report title, the share button, and edit button.

- **Always show**—the report header is always visible.
- **Auto hide**—the report header appears briefly, then is hidden. Hover your mouse over the header area to redisplay it.
- **Initially hidden**—the report header is hidden until the mouse hovers over the header area.

Navigation position

The navigation position options control how the pagination controls appear in your report.

- **Left** displays the report page navigation in a collapsible panel on the left.
- **Top** displays the report page navigation above the report.

Display mode

The display mode options govern how your report fits in the browser window.

- **Fit to width** scales the components in the report in proportion to the width of the window.
- **Actual size** displays the report components at the same size they were created.
- **Has margin** shows or hides the margin area surrounding the report. Hiding the margin makes the report appear to occupy the entire window.

Canvas Size

The canvas size options let a user select from preset sizes or enter custom display dimensions.

Report-level component position

This option determines how report-level components interact with other components on the page.

- Top positions report-level components in front of all other components (similar to using the Arrange > Order > Bring to front menu).
- Bottom positions report-level components behind all other components (similar to using the Arrange > Order > Send to back menu).

More information on how to use the report editor is available at: https://support.google.com/datastudio/answer/6371822?visit_id=636743650895496899-4004008079&rd=2

⁹ Source: https://support.google.com/datastudio/answer/7355651?hl=en&ref_topic=6369007

Add a new chart or graph

(Source: <https://support.google.com/datastudio/answer/6291062>)

The existing dashboards are just a starting place, add your own table, pie chart, trend chart, bar chart, bubble charts, and pivot tables to your dashboards. It is a simple process to create visually attractive charts in Data Studio. To do this, click the chart type to be added on the toolbar and draw it on the page. Customize the chart's data in the DATA properties panel. Use the STYLE panel options to adjust design elements.

Add a chart to a report

Use the toolbar at the top of the editor to add charts, as well as text boxes, controls, shapes, and images to your report. It is also possible to select the desired component from the **Insert** menu.

After selecting the chart on the toolbar, **click and draw** on the editor page where the chart should appear.

Resize the chart by clicking and dragging a corner or edge. For some charts, such as the scorecard and pie chart, changing the chart size increases or decreases the font size used for metrics and labels.

It is possible to **move the chart** using a mouse or keyboard arrows. When using the keyboard, fine-tune the placement by holding down the Shift key while pressing an arrow key (this moves the selected component in single pixel increments).

When two or more components are selected, use the **Arrange** menu to align them horizontally or vertically. It is possible to also group selected components, change their visual order, and make them appear on every page.

More information on how to arrange and group components on a page is available at: <https://support.google.com/datastudio/answer/6370253>

Configure chart data

Charts get their data from a data source. The fields in a data source come in 2 forms: dimensions and metrics. Data Studio automatically picks dimensions and metrics when a chart is first drawn, but it is possible to easily change these:

To change a chart's data:

1. Select the chart
2. Select the DATA tab on the right
3. Use the properties panel to select new dimensions (in green) or metrics (in blue).

Some charts allow multiple dimensions and metrics to be added. Click +Add a dimension or +Add a metric (in gray) under the existing ones.

More information on dimensions and metrics is available at: <https://support.google.com/datastudio/answer/6402048>

Switch chart type

It is possible to switch between chart types to see how the data look using different visualization techniques. Data Studio adjusts the default dimensions and metrics to best suit the selected chart type. Any customizations you've made to the chart, including adding or changing dimensions and metrics, style options, filters, segments, etc., are preserved so it is possible to switch back to the originally selected chart.

To switch chart types:

1. Select the chart
2. Select the DATA tab on the right
3. Click the desired chart in the Chart type section at the top of the properties panel.

Select a different data source for a chart

Every chart added will be based on the current default data source. It is possible to select a different data source using the Data Source options in the DATA tab. Charts can only visualize data from a single data source, but it is possible to have as many different data sources in a report as the creator has access to.

To change the data source used by a chart:

1. Select the chart
2. Select the DATA tab on the right
3. Click the current data source
4. Select an existing data source or click CREATE NEW DATA SOURCE at the bottom of the list.
5. If prompted, add the data source to the report.

Changing the data source for one chart has no effect on the other charts in your report. If the new data source has a different structure (schema) from the original one, the chart may display errors. It is possible to fix those by selecting dimensions and metrics from the new data source.

More information about connecting to data is available at: <https://support.google.com/datastudio/answer/6390659>

Change chart date properties

The chart date options are located below the dimensions and metrics in the DATA tab. These options let the time frame of the chart be controlled. It is also possible to tell the chart whether or not to display previous period performance.

More information on working with dates and time is available at <https://support.google.com/datastudio/answer/6401549>

Style a chart

The STYLE tab options provide a great deal of control over the appearance of charts. Chart style options include shared properties, such as font styles, background and border colors. Individual charts will also have properties specific that chart, such as whether to show a time series as a line or bars, or how wide to make the donut option of a pie chart.

It is possible to style multiple charts at once by selecting them, then applying any available styles to the selection. Only shared style properties can be set this way.

It is possible to copy styling from one component to another component:

1. Copy the component that has the styling that should be applied.
2. Right click on the component to which that style should be applied.
3. Select Paste Special > Paste style only.

Add a filter control to data

View just a subset of results within a dashboard and easily compare mobile vs. desktop users or view only content that includes a specific keyword within the URL by adding a new filter control to the dashboard. Filter controls allow report users to apply and remove filters on-demand.

It is possible to add filters to an entire report, page, chart, filter control, or to a group of charts and controls. To add a filter, the following steps can be followed:

1. Add a new filter to a chart or filter control
2. Edit your report.
3. Select the chart or control to filter.
4. Select the DATA tab in the Properties panel.
5. Click +Add a filter.
6. Click CREATE NEW FILTER. (It is also possible to reuse an existing filter)
7. Name the filter.
8. Select Include or Exclude. Include filters include only data that matches the expression provided. Exclude filters remove all data that matches the expression.
9. Select the dimension or metric by which to filter.
10. Select the comparison operator (e.g. Equals, Contains, Greater than, etc.).
11. Provide a comparison value for the expression.
12. (Optional) To add an OR clause, click OR on the right of the condition.
13. (Optional) To add an AND clause, click AND below the clause.
14. Click SAVE.


More information on filter conditions is available at: <https://support.google.com/datastudio/answer/7326859?hl=en>

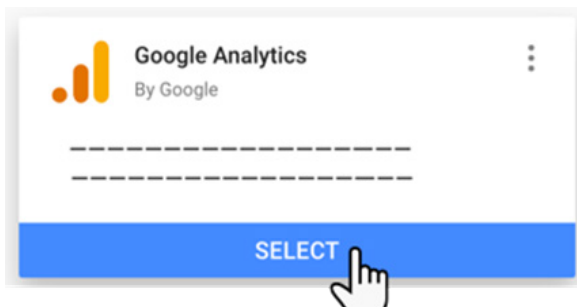
Add new data source

It is possible to consolidate results from multiple accounts associated with a website by adding another data source to the dashboard. Data from a range of other tools, including spreadsheets, can also be added. Below is a tutorial from Google Data Studio on how to create a data source that connects to a Google Analytics view.

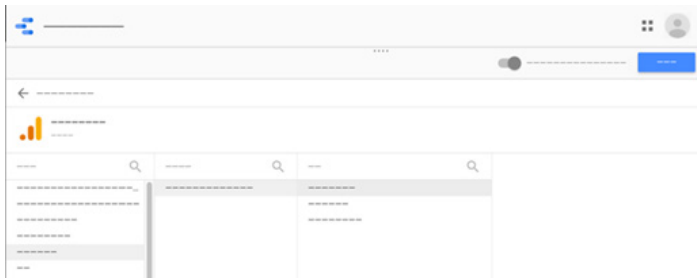
Step 1: Connect to your Google Analytics account

The first step in creating a data source is to connect to your data set.

1. Sign in to Data Studio. Access Google Data Studio at this link: <https://datastudio.google.com/>
2. Navigate to the DATA SOURCES tab. Available here: <https://datastudio.google.com/#/navigation/datasources>
3. Click  .
4. In the connectors panel, select Google Analytics.



5. Select an **Account**, **Property**, and **View**.



6. Click **CONNECT** in the upper right. The *fields panel* appears.

Step 2: Configure the data source

The fields panel contains all the dimensions (the green chips) and metrics (the blue chips) from the data set. The data source can be fine-tuned by renaming or disabling fields, adding calculated fields, and changing the aggregation and data type. Note: To create a basic report using Google Analytics data, it usually isn't necessary to make any changes to the data source.

Data source credentials

Credentials determine who can see the data provided by this data source. The default setting is Owner's Credentials, which allows anyone using this data source to see the data as if they were the owner of the data source. This means viewers of a report that uses this data source can see the data without needing their own access to the data set.


It is possible to instead require each user of the data source and any reports built on it to provide their own credentials to the data set. To do this:

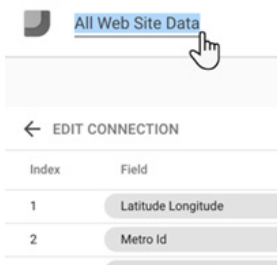
1. In the upper right of the fields panel, click Using Owner's Credentials
2. In the Data source access dialog, click Viewer's Credentials
3. Click SAVE.

More information on working with data source fields is available at: <https://support.google.com/datastudio/topic/7441655>

Step 3: Name the data source

Data Studio automatically names your data source the same as your data set name. To change that, click the name in the upper left and enter a new name.

It is also possible to rename the data source later from the DATA SOURCES home page by clicking the context menu  and selecting Rename.



Step 4: Share the data source

It is possible to let other editors add this data source to reports by sharing it with them.

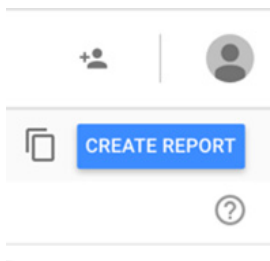
To share a data source

1. In the upper right, click Share icon
2. Enter the email addresses or Google Groups to be shared with
3. For each person or group, select the access (permissions). Permissions determine what other people can do with the data source. Your choices are:
 - Can view. People with this permission can view but not edit the data source.
 - Can edit. People with this permission can both view and edit the data source.

Step 5: Use the data source in reports

It is now possible to create reports that get their data from the Google Analytics view:

1. Click CREATE REPORT in the upper right



2. Click ADD TO REPORT in the *You are about to add a data source to this report* dialog.

Source: <https://support.google.com/datastudio/answer/6295036?hl=en>

Annex 4: Instruction Package

This is a collection of guidance notes for implementing the recommendations made in the Measuring Data Use report. Each section contains instructions from Google Analytics help files. It is provided to help NSO managers, IT staff, and consultants working on the maintenance and improvements of data portals to implement any recommendations that have been identified as helpful to a specific context.

Installing Google Analytics

To start collecting basic data from a website:

1. **Create or sign in to an Analytics account.**
 - a. Go to <https://google.com/analytics>
 - b. Do one of the following:
 - i. To create an account, click Start for free.
 - ii. To sign in to your account, Click Sign in to Analytics.
2. **Set up a property in your Analytics account.** A property represents a website or app, and is the collection point in Analytics for the data from a site or app.
 - a. Sign in to the Analytics account at <https://analytics.google.com>
 - b. Click Admin. The Admin page provides access to the Analytics administrative features.
 - c. In the ACCOUNT column, use the menu to select the account to which the property will be added.
 - d. In the PROPERTY column, select Create new property from the menu.
 - i. Note: Without edit permission on the account, the Create new property option won't be visible. Check that the correct account in the ACCOUNT column has been selected.
 - e. Select Website.
 - f. Enter the Website name.
 - g. Enter the Website URL.
 - i. Note: It isn't possible to create a property if the URL isn't formatted correctly.
 - h. Select an industry category
 - i. Select the reporting time zone. This will be used as the day boundary for reports, regardless of where the data originates.
 - j. Click Get Tracking ID. The property is created after this button is clicked, but the tracking code must be set up to collect data. The tracking ID will be used in the gtag.js tracking code snippet.
3. **Set up a reporting view in your property.** Views allows the creation of filtered perspectives of your data; for example, all data except from your company's internal IP addresses, or all data associated with a specific sales region.
 - a. Sign in to the Analytics account at <https://analytics.google.com>
 - b. Under Admin, navigate to the account and property to which a view will be added.
 - c. In the VIEW column, click the menu, then click Create new view.
 - d. Select Web Site
 - e. Enter a Name.
 - i. Note: Use a specific and descriptive name, to make it easy to tell what data is in this view when the name is seen in a list.
 - f. Select the Reporting Time Zone.
 - i. Note: If an Analytics account is linked to a Google Ads account, the time zone is automatically set to the Google Ads preference, and this option won't be visible.
 - g. (For User ID enabled Properties only) Click the toggle ON to create a User ID view. For a reporting view, leave the toggle set to OFF.
 - h. Click Create View.

4. **Add the tracking code** to a website to collect data in an Analytics property.
 - a. Sign in to the Analytics account at <https://analytics.google.com>
 - b. Find the tracking ID (generated during step 2).
 - i. Click Admin.
 - ii. Select an account from the menu in the ACCOUNT column.
 - iii. Select a property from the menu in the PROPERTY column.
 - iv. Under PROPERTY, click Tracking Info > Tracking Code. The tracking ID is displayed at the top of the page.
 - c. Paste the following snippet right after the <head> tag on each page of your site. Replace GA_TRACKING_ID with your own Google Analytics tracking ID:

```
<!-- Global Site Tag (gtag.js) - Google Analytics -->
<script async src="https://www.googletagmanager.com/gtag/
js?id=GA_TRACKING_ID"></script>
<script>
  window.dataLayer = window.dataLayer || [];
  function gtag(){dataLayer.push(arguments);}
  gtag('js', new Date());

  gtag('config', 'GA_TRACKING_ID');
</script>
```

- d. Verify that the tracking code is working.
 - i. Visit the website and check to see that this visit is being registered in the Real-Time reports.

Sources:

<https://support.google.com/analytics/answer/1008015>
<https://support.google.com/analytics/answer/1042508>
<https://support.google.com/analytics/answer/1009714>
<https://support.google.com/analytics/answer/1008080>

Filtering bot traffic

1. Sign in to the Analytics account at <https://analytics.google.com>
2. Click Admin.
3. Select View Settings on the view column
4. Under Bot Filtering, click the box to exclude all hits from known bots and spiders.

Updating Google Analytics properties and views

1. Sign in to the Analytics account at <https://analytics.google.com>
2. Click Admin.
3. Navigate to the property to be edited
4. In the PROPERTY column, click Property settings.
5. Edit the Property name.
6. Click Save.

Source:

<https://support.google.com/analytics/answer/3467852?hl=en>

Implementing site search tracking

1. Sign in to the Analytics account at <https://analytics.google.com>
2. Click Admin.
3. Select View Settings on the view column
4. Under Site Search Settings, set Site Search Tracking to ON.

In some cases, this may be sufficient to implement site search, but further configuration may be required. In the Query Parameter field, enter the word or words that designate internal query parameters, such as term, search, query, keywords. Sometimes query parameters are designated by just a letter, such as s or q. Enter up to five parameters, separated by commas. Do not enter any additional characters: for example, if the query parameter is designated by the letter q, enter only q (not q=). Read

How to identify search query parameters:

When users search a site, their queries are usually included in the URL. For example, if Google is used to search the phrase Mountain View, the URL contains q (Google's query parameter) followed by the query:

<http://www.google.com?hl=en&q=mountain+view...>

If the site uses categories, then the same principle applies. The webmaster may also be able to identify the query parameter for the site.

If the site uses a POST-based search engine, the search-results URL would look something like: http://www.example.com/search_results.php

There are two options to use Site Search for POST-based search engines:

- Option 1: Configure your web application to append the query keywords to the end of the URL (e.g., http://www.example.com/search_results.php?q=keyword) and then set up Site Search as described in the previous section.
- Option 2: Customize the tracking code on your results page to dynamically specify a virtual page path that includes the query keywords. The tracking code on the results page would look something like:

```
analytics.js: ga('send', 'pageview', '/search_results.php?q=keyword');
```

Source:

<https://support.google.com/analytics/answer/1012264?hl=en>

Filtering internal traffic

To prevent internal traffic from affecting the web traffic data, a filter can be used to filter out traffic by IP address.

1. Sign in to Google Analytics.
2. Click Admin
3. Navigate to the view in which the filter will be created.
4. In the VIEW column, click Filters.
5. Click + New Filter. (If this button is not visible, the necessary permission has not been granted.)
6. Select Create new Filter.
7. Enter a name for the filter.
8. Leave the Filter Type as Predefined.
9. Click the Select filter type drop-down menu and select Exclude.
10. Click the Select source or destination drop-down menu and select traffic from the IP addresses.

11. Click the Select expression drop-down menu and select the appropriate expression.
12. Enter the IP address or a regular expression. See the examples below.
13. From the Available views list, select the views to which the filter will be applied, then click Add.
14. Click Save.

Example of excluding a single IP address:

- Filter name: enter a name
- Filter type: Predefined
- Select filter type: Exclude
- Select source or destination: traffic from the IP addresses
- Select expression: that are equal to
- IP address: enter a single IP address

The screenshot shows the 'Filter Information' dialog box in Google Analytics. It has a 'Filter Name' field with the text 'Exclude internal IPs'. Below that, the 'Filter Type' section has two tabs: 'Predefined' (which is active) and 'Custom'. Under the 'Predefined' tab, there are three dropdown menus: the first is set to 'Exclude', the second to 'traffic from the IP addresses', and the third to 'that are equal to'. At the bottom, there is an 'IP address' field containing the text '192.168.0.1'.

Ensuring the filter works:

It is possible to use the [Google Tag Assistant](#) Chrome browser extension to record a flow on the website that sends hits to the property. The extension has an Analytics report that shows the effects of any filters set for a given view.

Sources:

<https://support.google.com/analytics/answer/1034840?hl=en>
<https://support.google.com/analytics/answer/1034823>

Tracking data file downloads

Clicks on links to downloadable files (e.g. PDF, DOCX, etc.) can be tracked as events though other methods are possible.

To track clicks as events:

1. Add a Google Analytics pageview tag if one does not already exist. This tag must fire on all pages.
2. Enable Tag Manager to capture clicked URL values:
 - a. Click Variables.
 - b. In the Built-In Variables section, click Configure.
- c. Select the Click URL built-in variable.
3. Create a single trigger to handle link clicks for each item in the table above:
 - a. Click Triggers and then New.
 - b. Click Trigger Configuration and then Click - Just Links
 - c. Configure the trigger with these settings:
 - i. Wait for tags: Enable
 - ii. Max wait time: 2000ms

- iii. Check Validation: Disable
 - iv. Enable this trigger when all of these conditions are true: Click URL and then matches RegEx and then contact\\.html\\buy\\.html\\about\\.html
 - v. Fire On: All Clicks
 - d. Name the trigger “Trigger - Nav Clicks” and save.
- 4. Create a single Tag to handle all clicks:
 - a. Click Tags and then New.
 - b. Click Tag Configuration and then Universal Analytics.
 - c. Set Track Type to Event.
 - d. Complete the Event Tracking Parameters:
 - i. Category: Nav
 - ii. Action: Select
 - iii. Label: {{Click URL}}
 - e. Click Triggering and select the trigger created in step 3, titled “Trigger - Nav Clicks”.
 - f. Save the tag.
- 5. Publish the container.

If Analytics tracking is added directly to the site, event tracking will need to be implemented with analytics.js. Developer instructions for this form of tracking are available at: <https://developers.google.com/analytics/devguides/collection/analyticsjs/events>

Sources:

<https://support.google.com/analytics/answer/1012044?hl=en>
https://support.google.com/tagmanager/answer/6106716?visit_id=636747063213409290-2725201082&rd=1

Capturing web traffic to key data sites

It is possible to use the Google Tag manager to manage Analytics tracking or to add tracking directly to a site.

If Google Tag manager is used, follow the instructions below to implement events:

1. Add a Google Analytics pageview tag if one does not already exist. This tag must fire on all pages.
2. Enable Tag Manager to capture clicked URL values:
 - a. Click Variables.
 - b. In the Built-In Variables section, click Configure.
 - c. Select the Click URL built-in variable.
3. Create a single trigger to handle link clicks for each item in the table above:
 - a. Click Triggers and then New.
 - b. Click Trigger Configuration and then Click - Just Links
 - c. Configure the trigger with these settings:
 - i. Wait for tags: Enable
 - ii. Max wait time: 2000ms
 - iii. Check Validation: Disable
 - iv. Enable this trigger when all of these conditions are true: Click URL and then matches RegEx and then contact\\.html\\buy\\.html\\about\\.html
 - v. Fire On: All Clicks
 - d. Name the trigger “Trigger - Nav Clicks” and save.
4. Create a single Tag to handle all clicks:
 - a. Click Tags and then New.

- b. Click Tag Configuration and then Universal Analytics.
 - c. Set Track Type to Event.
 - d. Complete the Event Tracking Parameters:
 - i. Category: Nav
 - ii. Action: Select
 - iii. Label: {{Click URL}}
 - e. Click Triggering and select the trigger created in step 3, titled "Trigger - Nav Clicks".
 - f. Save the tag.
5. Publish the container.

If analytics tracking code is added directly to the site, the script below can be copied and pasted into the pages to set up Event tracking for outbound links. This script should be added to the page header, but not within the basic Analytics tracking code snippet. Outbound clicks will appear in the Analytics Events reports with a Category of "outbound" and an Action of "click".

This example assumes that the website uses the analytics.js tracking code.

```
<script>
/**
 * Function that tracks a click on an outbound link in Analytics.
 * This function takes a valid URL string as an argument, and uses that URL
 * string
 * as the event label. Setting the transport method to 'beacon' lets the hit
 * be sent
 * using 'navigator.sendBeacon' in browser that support it.
 */
var trackOutboundLink = function(url) {
  ga('send', 'event', 'outbound', 'click', url, {
    'transport': 'beacon',
    'hitCallback': function(){document.location = url;}
  });
}
</script>
```

It will also be necessary to add (or modify) the `onclick` attribute to the links. Use this example as a model for links:

```
<a href="http://www.example.com" onclick="trackOutboundLink('http://www.
example.com'); return false;">Check out example.com</a>
```

Sources:

<https://support.google.com/analytics/answer/1136920?hl=en>

https://support.google.com/tagmanager/answer/6106716?visit_id=636747063213409290-2725201082&rd=1

Implementing canonical URL structures

For all canonicalization methods, follow these general guidelines.

- Don't use the robots.txt file for canonicalization purposes.
- Don't use the URL removal tool for canonicalization: it removes all versions of a URL from search.
- Don't specify different URLs as canonical for the same page using the same or different canonicalization techniques (for example, don't specify one URL in a sitemap but a different URL for that same page using `rel="canonical"`).
- Don't use `noindex` as a means to prevent selection of a canonical page. This directive is intended

- to exclude the page from the index, not to manage the choice of a canonical page.
- Do specify a canonical page when using hreflang tags. Specify a canonical page in same language, or the best possible substitute language if a canonical doesn't exist for the same language.

Prefer HTTPS over HTTP for canonical URLs

Google prefers HTTPS pages over equivalent HTTP pages as canonical, except when there are issues or conflicting signals such as the following:

- The HTTPS page has an invalid SSL certificate.
- The HTTPS page contains insecure dependencies (other than images).
- The HTTPS page redirects users to or through an HTTP page.
- The HTTPS page has a rel="canonical" link to the HTTP page.

Although Google Analytics systems prefer HTTPS pages over HTTP pages by default, this behavior can be ensured by taking any of the following actions:

- Add redirects from the HTTP page to the HTTPS page.
- Add a rel="canonical" link from the HTTP page to the HTTPS page.
- Implement HSTS.

To prevent Google from incorrectly making the HTTP page canonical, the following practices should be avoided:

- Bad SSL certificates and HTTPS-to-HTTP redirects cause us to prefer HTTP very strongly. Implementing HSTS cannot override this strong preference.
- Including the HTTP page in your sitemap or hreflang entries rather than the HTTPS version.
- Implementing your SSL/TLS certificate for the wrong host-variant: for example, example.com serving the certificate for www.example.com. The certificate must match your complete site URL, or be a wildcard certificate that can be used for multiple subdomains on a domain.

Tell Google to ignore dynamic parameters

Use Parameter Handling to tell Googlebot about any parameters that should be ignored when crawling. Ignoring certain parameters can reduce duplicate content in Google's index and make your site more crawlable. For example, the parameter sessionId should be ignored is selected, Googlebot will consider the following two URLs as duplicates:

- <https://www.example.com/dresses/green.php?sessionId=273749>
- <https://www.example.com/dresses/green.php>

Several specific methods exist for specifying a canonical URL for duplicate URLs or duplicate/similar pages:

- Set a preferred domain: Use Search Console to specify URLs on one domain as canonical over their counterparts on another domain. For example, example.com rather than www.example.com. Use this only when there are two similar sites that differ only by subdomain. Don't use this for http/https counterpart sites.
- Use rel="canonical" link tag: Add a <link> tag in the code for all duplicate pages, pointing to the canonical page.
- Use rel="canonical" HTTP header: Send a rel=canonical header in the page response.
- Use a sitemap: Specify canonical pages in a sitemap.
- Use 301 redirects for retired URLs: Use 301 redirects to tell Googlebot that a redirected URL is a better version than a given URL. Use this only when deprecating a duplicate page.

Source:

<https://support.google.com/webmasters/answer/139066?hl=en>

Enabling cross-domain tracking

If the flow of website user experience moves between more than one domain (e.g. example-products.com to example-checkout.com), Universal Analytics tags should be set up to track those visits as one. Without this functionality, Google Analytics will track the user as two separate visitors, which will inflate then number of users appearing in reports. This will occur even if both domains use the same Tag Manager container and Analytics account.

To configure Universal Analytics tags in Tag Manager to track across multiple domains:

1. In Tag Manager, open your Google Analytics Settings variable or Universal Analytics tag for editing.
 - a. To find your Google Analytics Settings variable, click Variables and then click the appropriate variable in the list.
 - b. To find your Universal Analytics tag, click Tags and then select the appropriate tag in the list.
2. Navigate to More Settings and then Cross Domain Tracking.
 - a. Note: The preferred method is to make changes to the Google Analytics Settings variable, so that changes are shared across multiple tags. In Universal Analytics tags, click Enable overriding settings in this tag to expose the More Settings option.
3. In the Auto Link Domains field, enter a comma-separated list of domains.
4. Save changes and publish the container.

Source:

https://support.google.com/tagmanager/answer/6164469?visit_id=636747063213409290-2725201082&rd=1



About Open Data Watch

Open Data Watch is an international, non-profit organization of data experts working to bring change to organizations that produce and manage official statistical data. We support the efforts of national statistical offices (NSOs), particularly those in low- and middle-income, to improve their data systems and harness the advancements of the data revolution. Through our policy advice, data support, and monitoring work, we influence and help both NSOs and other organizations meet the goals of their national plans and the sustainable development goals (SDGs). Our aim is to make lasting changes to improve the coverage and openness of official statistics.

Open Data Watch
info@opendatawatch.com

