

friendly (technical) guide to COUNTER

COUNTER

This guide is a non-intimidating manual for publishers and vendors implementing COUNTER-compliant usage statistics for the first time, or for more experienced hands, implementing new reports or preparing for COUNTER audit.

Author: **Tasha Mellins-Cohen**

semantic 

AAP } ASSOCIATION
OF AMERICAN
PUBLISHERS

contents

Introduction	2
What is COUNTER?	2
Who uses COUNTER reports?	2
How is COUNTER funded and organized?	2
How do I become COUNTER compliant?	2
How will my customers know I am COUNTER compliant?	3
How do I find a COUNTER auditor?	3
Auditing	4
The audit process	4
Categories of audit result	4
Tracking usage	5
Page tags	5
Logfiles	5
Cookies	6
Abnormal spikes in usage	6
Processing your data	7
Return codes	7
Double clicks	7
Federated and bot searches	7
Data changes	8
Delivering COUNTER reports	9
SUSHI	9
Delimited files	10
Customer categories	10

What is COUNTER?

COUNTER stands for Counting Online Usage of NeTworked Electronic Resources. Our website is at <http://www.projectcounter.org/>

COUNTER was one of the first, if not the first, standards organization established for the modern information environment. It has succeeded in bringing together a collaboration of publishers and librarians to develop and maintain the standard for counting the use of electronic resources. It has also ensured that most major publishers and vendors are compliant by providing their library customers around the world with COUNTER usage statistics.

COUNTER publishes the Code of Practice, which is the standard for counting the use of electronic resources. It also maintains and publishes the register of COUNTER-compliant vendors and publishers.

The COUNTER reports

The COUNTER reports in release 4 are broken down by content types: there are reports for journals, books, databases and multimedia, as well as title reports which combine book and journal information. Details of the report types can be found in the accompanying Friendly Guide to COUNTER.

Who uses COUNTER reports?

The COUNTER standard was originally developed to provide a service to librarians and others who purchase subscriptions to publishers' content. The intention was to enable librarians to easily compare their usage across different publishers' content, and allow them to use that information to calculate a cost-per-download for their subscriptions. COUNTER reports were not originally intended to be used by publishers as a way of measuring usage across their client base, but are increasingly being used for that purpose.

All academic libraries across the world use COUNTER usage reports to:

- Inform renewal decisions
- Inform faculty about the value of the library and its resources
- Understand user behaviour and improve the user experience

Most major vendors and publishers also use COUNTER reports to:

- Provide reliable and consistent usage data to their customers
- Upsell using COUNTER data about access denied as the result of a content item not licensed or because concurrent/simultaneous user licence limits were exceeded.
- Inform editors and authors about the usage of their publications

How is COUNTER funded and organized?

COUNTER is a not-for-profit membership organization, funded by membership fees and sponsorship.

The membership – publishers, vendors and librarians – lead COUNTER. A Board of Directors has an overview of financial matters and appoints the Executive Committee to oversee COUNTER's operation. A Project Director, reporting to the Executive Committee, is responsible for the day-to-day management. The publisher, intermediary and librarian communities are all represented on the Board and on the Executive Committee, as well as on the Technical Advisory Board.

How do I become COUNTER compliant?

Select the usage reports that are relevant to your products

Some of the usage reports are obligatory for COUNTER compliance; these are the ones listed as 'standard' in this guide. However, only the 'standard' reports which are relevant to the categories of content that you publish are required for COUNTER compliance. For example, if you publish full-text journals but no books or databases, then you need to supply the 'standard' set of journal reports, but no book or

database reports. If, on the other hand, you publish multimedia databases, then you will need to supply database and multimedia reports but no journal or book reports.

The 'optional' usage reports are genuinely optional: they are not required for COUNTER compliance.

Prepare your COUNTER-compliant reports

The information in this guide, and in the full specification, will help you.

Enable SUSHI

There is a brief introduction to SUSHI in this guide, and more information at <http://www.niso.org/workrooms/sushi/tools/>.

Send your reports to COUNTER

COUNTER will arrange for a library test site to check them. Alternatively, one of the COUNTER-approved auditors will check them for a modest fee.

Complete the paperwork

Two pieces of paperwork are needed for COUNTER compliance: you need to become a COUNTER member and to complete the Declaration of COUNTER Compliance. You will find both the Membership Application and the Declaration on the COUNTER website.

Undergo an independent audit

You will need to pass an audit within six months of signing the Declaration of COUNTER Compliance, and annually thereafter. If you are a very small publisher and will find an annual audit particularly difficult, please get in touch with COUNTER to discuss permission to be audited every other year.

How will my customers know I am COUNTER compliant?
COUNTER will list you on its website and you can use the COUNTER logo on your website.

How do I find a COUNTER auditor?

There are two approved COUNTER auditors:

- ABC: <http://www.abc.org.uk>
- BPA Worldwide: <http://www.bpawww.com>

COUNTER will also accept an audit by any Chartered Accountant (UK), CPA (USA) or their equivalent elsewhere.

An important feature of the COUNTER Code of Practice is that compliant vendors must be independently audited on a regular basis in order to maintain their COUNTER-compliant status. We have tried to ensure that the audit meets the need of libraries for credible usage statistics without making the process too onerous. For this reason, audits are conducted online using the detailed test scripts included in the auditing standards and procedures.

The audit process

COUNTER-compliant vendors are notified in writing by COUNTER that an audit is required. We send this notification at least three months before the audit is due.

You have one month to respond to the notice, telling us:

- Your planned timetable for the audit
- The name of the organization that will carry out the audit
- Any queries you have about the audit process

Regardless of the auditor selected, the audit must adhere to the requirements and use the tests specified in Appendix E of the COUNTER Code of Practice. The audit is carried out in three stages:

- 1 The format and structure of the usage reports
- 2 The integrity of the reported usage statistics
- 3 The delivery of the usage reports

On completion of a successful audit, the auditor must send a signed copy of the audit report to the COUNTER office (lorraine.estelle@counterusage.org).

If the audit is not successful, the auditor will send an interim report to the COUNTER office, outlining the reasons for failure. The auditors will work with you to correct the areas of failure within a time agreed with COUNTER.

Categories of audit result

Pass

No further action is required as a result of the audit. In some cases the auditor may add observations to the audit report. These are designed to help improve your COUNTER usage reports, but they are outside the scope of the audit itself.

Qualified Pass

The audit has been passed, but with a minor issue which needs to be addressed in order to maintain COUNTER-compliant status. A minor issue does not affect the reported figures; for example, it may be related to the presentation of the report. Minor issues need to be resolved within three months of the audit to maintain COUNTER-compliant status.

Fail

There is an issue that must be rectified for you to maintain COUNTER-compliant status. You will be given a grace period of one month to rectify the reasons for the failure from the date of notification [Date of Report] and achieve a pass.

Usage data can be generated in a number of ways and COUNTER does not prescribe which approach should be taken. The two most common approaches are page tagging and logfile analysis. Both have advantages and disadvantages, summarized below. It is important to remember that data collected for COUNTER reports only records intended usage: because every platform records usage slightly differently, it is not possible for us to describe all of the available filters for cleaning up the data. This guide therefore only outlines the requirements.

Page tags

Page tags are small pieces of code embedded in each page of your website. They are usually written in JavaScript, though other languages such as Java are used at the discretion of the site developers. Data is gathered via these code tags and passed to a database. Scripts written in languages such as JQuery and AJAX can then be used in conjunction with a server-side scripting language such as PHP to manipulate and store the data, allowing complete control over how the data is represented. The data storage and manipulation script may have access to additional information about the web client or the user, for example by reading information from your access management system. Access management systems authenticate users as being affiliated with a specific organization, and then authorize their access to resources licensed by that organization.

Page tagging is fairly standard in web analytics; for example, Google Analytics makes use of it. One key difference between logfile analysis and page tagging is that with page tags a usage count is activated by opening the page, not by requesting it from the server. This means that you are likely to see a more accurate reflection of usage through using page tags, because cached pages are counted in the same way as server calls.

Page tags are particularly useful for companies that do not have access to their own web servers; with the increasing use of cloud storage, page tagging is becoming a

preferred way to obtain analytics information. Page tagging and tag analysis can be done in house, but are also widely available as third-party services.

Page tag examples

This is a small introductory selection of page tags from the Google Tag Manager collection, all of which have direct applications in COUNTER reporting.

Tag name	Google definition	Use in COUNTER reports
Page View	The most basic tag, Page View should fire on every page of your site	
Event	Used to track a specific action or event, such as a button click	Separating out PDF downloads, searches, recording clicks and the various other items required for full COUNTER reporting
Timing	Used to track loading speeds on your page	Distinguishing double clicks

If you are interested in using page tagging for generating COUNTER reports, <https://www.google.co.uk/analytics/> is a good place to start.

Logfiles

Logfiles are text files representing individual HTTP requests, including the client-side host name or IP address, the date and time of the request, the requested file name, the HTTP response status and size, the referring URL and the browser information.

Most web servers produce logfiles by default, in a pre-defined format which may differ by server type, so the raw data you need for your COUNTER implementation should already be available to you, without you having to make changes to your website.

Because logfile data is held on your own servers in a standard format, you can use

a variety of analytics programmes and receive consistent results over time. Logfiles are also independent of your users' web browsers, meaning you can reliably track all activity for the purposes of COUNTER reporting.

Be aware that cached pages are not counted in logfile analysis, because they are not requested from the server, although cached pages can account for a significant proportion of page views.

As different servers will deliver logfiles in different formats, most logfile analysis is performed by the people who support the server – usually the vendor's in-house team.

If you are interested in using logfile analysis for generating COUNTER reports, we recommend that you speak to your development team.

To learn more about logfiles generally, the AWS logfile documentation is excellent. <http://aws.amazon.com/documentation/>

Cookies

Cookies are small files that are stored on a user's computer. They are designed to hold a modest amount of data specific to a particular client and website, and can be accessed by either the web server or the client computer. Page tags can be used to manage the process of assigning cookies to visitors; with logfile analysis, the server has to be configured to do this. There are legal considerations around assigning cookies, so please check the requirements that apply to you before configuring your setup.

Abnormal spikes in usage

What is regarded as an abnormal spike in usage can vary from one institution to another; there are many occasions on which exceptionally high usage in a particular month is genuine. Because of this, we do not have a strict protocol for dealing with usage spikes. However the following approaches will provide an indication of possible abnormal usage or another unusual event and should therefore be used as a prompt for human intervention to take a closer look at the numbers.

Positive Spike in Usage: Reported usage may be too high (a Positive Spike) if, in a specific month, the reported usage at a particular customer for an individual product is at least 100 units of measurement greater than 300% (three hundred per cent) above the previous 12-month average.

Negative Spike in Usage: Reported usage may be too low (a Negative Spike), if, in a specific month, the reported usage at a particular customer for an individual product is less than 1% (one per cent) of the previous 12-month average usage. (Note: the average usage of that product in the previous 12 months should be at least 20 units of measurement).

Only genuine, user-driven usage should be reported.

Return codes

Only successful and valid page requests should be counted; pages which fail to load, or bad requests, must be excluded. Only relevant content items (e.g. book chapters and their supplementary files) should be counted. Other records generated by the server as part of the requested page (e.g. images or style sheets) should not be counted.

For web server logs, successful requests are those with specific NCSA return codes, namely 200 and 304. The standards for return codes are defined and maintained by NCSA. Where your server uses key events, their definition should match the NCSA standards. For more information, see www.ncsa.illinois.edu.

Double clicks

Double clicks on an http link should be counted as one request. For the purposes of COUNTER, the time window for a double click on an HTML page is set at a maximum of 10 seconds between the first and second mouse clicks. For example, a click at 10.01.00 and a second click at 10.01.09 would be considered a double click; a click at 10.01.00 and a second click at 10.01.15 would count as two separate single clicks.

Downloading and rendering a large file, such as a video clip, may take longer than rendering an HTML page. For PDFs, images, video clips and audio clips, therefore, the double-click window is 30 seconds. For example, a click at 10.01.00 and a second click at 10.01.15 would be considered a double click; a click at 10.01.00 and a second click at 10.01.35 would count as two separate single clicks.

A double click may also be triggered by pressing a refresh or back button. When two requests are made for the same article within the above time limits (10 seconds for HTML, 30 seconds for PDF), the first request should be removed and the second retained. Any additional requests for the same article within these time limits should be treated identically: always remove the first and retain the second.

There are different ways to track whether or not a user is double-clicking, depending on how the user is authenticated on your site. These options are listed in order of increasing reliability, with option 4 being the most reliable.

- 1** If the user is authenticated only through their IP address, that IP should be used as the field to trace double clicks. Where you have multiple users on a single IP address, this can occasionally lead to separate clicks from different users being logged as a double click from one user. This will only happen if the multiple users are clicking on exactly the same content within a few seconds of each other.
- 2** When a session cookie is implemented and logged, the session cookie should be used to trace double clicks.
- 3** When a user cookie is available and logged, the user cookie should be used to trace double clicks.
- 4** When an individual has logged in with their own profile, their username should be used to trace double clicks.

Federated and bot searches

The growing use of federated searches and the spread of web crawler robots have the potential to inflate usage statistics, so COUNTER requires you to show this type of usage separately in your reports.

Protocol for federated searches and automated search agents

Search activity generated by federated search engines and automated search agents should be categorized separately from regular searches. Any searches generated from such systems should be included in the separate 'Searches – federated and automated' counts within Database Report 1 and Platform Report 1, and must not be included in the 'Regular Searches' counts in these reports.

Activity generated by internet robots and crawlers must be excluded from all COUNTER usage reports.

The most common ways to recognize federated and automated search activity are as follows:

- A federated search engine may be using its own dedicated IP address, which can be identified and used to separate out the activity.
- If the standard HTML interface is being used (e.g. for screen scraping), the browser ID within the web logs can be used to identify the activity as coming from a federated search.
- For Z39.50 activity, authentication is usually through a username/password combination. Create a unique username/password that just the federated search engine will use.
- If an API or XML gateway is available, set up an instance of the gateway that is for the exclusive use of federated search tools; it is recommended that you also require the federated search to include an identifying parameter when making requests to the gateway.

COUNTER has lists of federated search tools and web robots in the appendices to the full implementation guide. COUNTER reviews and updates this list on a regular basis.

Protocol for tools that enable bulk downloading

Only genuine, user-driven usage should be reported. Usage of full-text articles that is initiated by automatic or semi-automatic bulk download tools, such as Quosa or Pubget, should only be recorded when the user has clicked on the downloaded full-text article in order to open it.

Data changes

Retrospective reporting of errors in usage data

If you identify errors in the usage statistics you have been providing in the COUNTER reports, you must correct the errors within three months of their discovery and inform your customers of the corrections.

Reporting of usage statistics when journal titles change

When the title of a journal is changed, but the DOI or ISSN stays the same, you should continue to provide a single COUNTER report for the journal. The report should be provided against the new title, with the original title being dropped from the list.

If a new DOI or ISSN is allocated to the journal when the title changes, you should provide two separate COUNTER reports, one under the old DOI or ISSN, and one under the new DOI or ISSN. You must not report usage for the same period under both the old DOI or ISSN and the new. Any report generated for the old DOI or ISSN should show zero usage from the month in which the new DOI or ISSN takes effect.

COUNTER reports are available in two formats: delimited files, which are readable using Excel and similar spreadsheet tools, and XML, which is delivered using SUSHI. It is important to remember that the XML file delivered by SUSHI is the primary COUNTER format: delimited files are required for libraries that have not yet enabled SUSHI.

There are some key factors to consider in delivering COUNTER reports:

- Reports should be readily available on a password-controlled website.
- There should be an option to receive an email alert when a new report is available.
- Reports must be provided monthly.
- Data must be updated within four weeks of the end of the reporting period.
- A minimum of the most recent 24 months of usage data must be available, unless the vendor is newly COUNTER compliant.
- It must be possible to request usage for a date range, in months, within the most recent 24-month period.
- Where no date range is specified, the default is to show the whole of the most recent 24-month period.
- Each report should reside in a separate file or page to avoid producing files of unwieldy size.
- For consortium usage reports the consortium administrator must be able to access both the consolidated consortium level usage statistics and the usage statistics for individual consortium member institutions, from a single login, using the same user ID and password (i.e. without having to log out and back in for each individual institution).
- Usage statistics reported in the COUNTER reports must not be browser-dependent. As a minimum, vendors must support current versions, compliant with World Wide Web Consortium (W3C) standards, of the following web browsers: Google Chrome, Internet Explorer and Mozilla Firefox.

SUSHI

‘The SUSHI protocol, allied with COUNTER compliance, has been a real godsend for the provision of standardized journal usage statistics’ (JUSP, 2012).

The Standardized Usage Statistics Harvesting Initiative (SUSHI) protocol is an ANSI/NISO standard designed to work with COUNTER usage statistics. It enables the fast and efficient collection of usage statistics. SUSHI allows libraries and providers of usage consolidation products to download quality-checked data from many publishers, removing the need to visit numerous websites.

The SUSHI protocol is designed to simplify the gathering of usage statistics by librarians, and it uses a series of XML schemas to do this. For the purposes of COUNTER release 4, we are interested in only two of these:

- the core SUSHI schema, which is generalized to retrieve any compatible XML-formatted usage statistics reports.
- the COUNTER-SUSHI schema, which references COUNTER to generate XML versions of COUNTER4 reports.

Please see <http://www.niso.org/workrooms/sushi/> for more information on SUSHI. The SUSHI schema for COUNTER reports can be found at <http://www.niso.org/schemas/sushi/#counter>

SUSHI server response times

A SUSHI server must respond to requests from a client within 120 seconds. If the publisher’s server is unable to consistently deliver a complete COUNTER report within 120 seconds, the server should instead be set up to respond to the initial request with a ‘Server Busy’ exception, and queue the request for background processing. Since most SUSHI clients will wait minutes or hours before retrying the request, the report will be ready to be delivered on the subsequent request.

The SUSHI Harvester for Consortia

Publishers delivering COUNTER reports to large consortia, particularly when the publisher has a great deal of content, may find that the standard Consortium report files become excessively large. In these instances, the SUSHI Harvester for Consortia should be used instead. The Harvester is a free Microsoft Access application from EBSCO that leverages the open-source SUSHI MISO client (developed by Serials Solutions) to batch download reports for the member organizations of the consortium, covering

- Journal Report 1
- Database Report 1
- Book Reports 1 and 2
- Multimedia Report 1
- Other COUNTER reports as needed

It is at the discretion of the publisher/vendor and their customers to decide which approach to take.

The SUSHI Harvester for Consortia, together with a detailed User Guide containing instructions on how to implement it, may be found on the NISO website at: http://www.niso.org/apps/group_public/download.php/4774/SUSHI-Harvester.zip

Delimited files

With the exception of consortia and title reports, the reports specified in COUNTER release 4 can all be delivered as delimited files:

- Comma separated, or .csv
- Tab separated, or .tsv

Delimited files can be opened and read in all spreadsheet tools, including Excel, OpenOffice Calc, Google Sheets and Numbers for Mac. Formatting, in the sense of typeface and colour, are irrelevant in delimited files, but it is important to adhere to the layout described in the COUNTER specification for each report.

There must be a column for every month that falls within the Reporting Period covered by the report:

- Where there was no usage recorded in a given month, '0' must be included in the relevant cells.
- Where usage has not yet been recorded for a given month, the relevant cells must be left blank.

Customer categories

Publishers must provide COUNTER reports on a per-customer ID basis. For example, if a business school has a separate customer ID from its parent university, the school and the university must be sent separate COUNTER reports, if it is feasible to do so. Most authentication is through IP address recognition. In the example above, if the business school does not have a unique IP range, it is not possible to distinguish usage from the school from that of the university, and therefore only the university should receive a COUNTER report.

This guide is published by COUNTER in 2016

About the author:

Tasha Mellins-Cohen is the Head of Platform at Semantico in Brighton, England.



Prior to joining Semantico Tasha worked for a variety of publishers, from small societies to large commercial units, with a focus on optimizing the operational aspects of scholarly publishing. She has extensive experience of the full publishing life cycle for both journals and books, from submission and peer review through to production and downstream data feeds. This has covered working with client publishers, offshore suppliers and end users. She is also a member and deputy Chair of the COUNTER Technical Advisory Group.

On the technology front, Tasha was involved with a number of new hosting platform builds and migrations, as well as workflow builds, and has now brought that experience to bear in her role at Semantico.



COUNTER wishes to express its gratitude to **AAP** ASSOCIATION OF AMERICAN PUBLISHERS / Professional & Scholarly Publishing, whose generosity has helped to make this guide possible and to **semantico**  for supporting the creation of the text.

COUNTER: 25 Egbert Road, Winchester, SO25 7EB
lorraine.estelle@counterusage.org



This work is licensed under a Creative Commons Attribution-NonCommercial 3.0 Unported License