COUNTER Code of Practice for e-Resources

Release 4

Appendix D   **May 2015**

Guidelines for Implementation

Introduction

For ease of reference, the numbering used in this Appendix corresponds to that of the Code of Practice itself; where appropriate the relevant section of the Code of Practice text is quoted.

Section 4: Usage Reports

Nomenclature of vendor-specific reports derived from the COUNTER Usage Reports

In addition to providing the official COUNTER usage reports specified in the Code of Practice, vendors may, in addition, provide modified versions of these reports with, for example, additional data that customers may have requested. Vendors **must not**, however, use the COUNTER Report names in association with these modified reports, or include any part of the COUNTER report names in the titles of such reports. Vendors may, however, mention in a footnote to a modified report the COUNTER report from which the modified report is derived.

4.1: Example Usage Reports

Institutional Identifier

The ISNI (International Standard Name Identifier) should be used as the Institutional Identifier, where this is required in the COUNTER Usage Reports.

The International Standard Name Identifier (ISNI) is an ISO standard (ISO 27729) that identifies public identities of parties. The ISNI is not intended to provide direct access to comprehensive information about a Public Identity - instead, it is designed as a 'bridge identifier' to link between systems where comprehensive information is held.

For further information on ISNI and its application to institutions see: http://www.isni.org/how-it-works

 ISNI Syntax
 An ISNI consists of 16 numerical digits. Example: ISNI 1234 1234 1234 1234
Examples of ISNI numbers are as follows:

New York University:	0000 0001 2169 8901
Eidgenossische Technische Hochschule Zurich:	0000 0001 2156 2780
Shanghai University:	0000 0001 2323 5732
University of Tokyo:	0000 0001 2169 1048
University of Oxford:	0000 0001 2172 3089
International Business Machines Corp. (IBM):	0000 0001 2184 5342
Ministry of Health and Family Welfare, India:	0000 0001 0721 7403

Filing order of journal titles

For journal titles beginning with a definite or indefinite article (in any language) this definite or indefinite article should be disregarded in the filing order of the list of titles in the COUNTER usage reports.

Categories of content covered by the COUNTER usage reports

Release 4 of the COUNTER Code specifically covers journals, databases, books/reference works and multimedia content. Each of these important categories of content has its own group of usage reports, which are listed in Section 4 of the Code of Practice. It is recognized, however, that while these categories cover a large proportion of the online content purchased by librarians and library consortia, they do not cover everything. The COUNTER Usage Reports are, however, sufficiently flexible to allow other categories of content to be covered. Each case is judged on its own merits, but examples include:

- Newspaper articles: where a collection of full text articles includes articles from periodical publications, such as newspapers, that are not journals and which may not have an ISSN number, usage of such articles may be counted in the Journal Reports, as they are serial publications that constitute part of a package of content that has been purchased by a customer.
- Reports: reports that have neither an ISSN nor an ISBN may be part of a collection of online content that includes books and/or journals. Usage of such reports may be counted in the appropriate COUNTER journal or books reports (but **not** in both)
- Supplementary data sets, video clips, etc.: it is acknowledged that an online journal, for example, is more than a collection of articles and that a growing portion of the value of an online journal lies in the supplementary data and other features to which the user has access. To enable vendors to record the usage of such features, COUNTER has expanded the scope of Journal Report 3.

Changing access models/content structure and the COUNTER usage reports

COUNTER acknowledges that measures used to assess the value and impact of publications should take into account not only the evolving structure of content, but also changing access models, both of which affect how users interact with content. At the same time, in order to provide robust, comparative data, a metric must be stable over time. COUNTER tries to combine the need for both stability and flexibility by specifying, on the one hand, a set of core usage reports that

measure basics, such as successful requests for full-text articles (Journal Report 1) and on the other hand a number of optional additional usage reports that allow more flexible and more granular reporting of usage. Journal Report 3, which provides a framework for the reporting usage of a wider range of page types and activities (such as article sub-sections), supplementary data and usage on mobile devices, is a good example of such a report. Vendors may select the page type/activity that is most appropriate to their own content from the extensive list of options provided in Journal Report 3.

JR1 GOA: Gold Open Access (GOA) is defined as access, immediately upon publication and at no charge to the user (but usually supported financially by the author or the author's funding agency), of peer-reviewed, full-text articles that have been accepted for publication in a journal.   . In order to make allowances for administrative and system delays that may occur in recording GOA status, COUNTER allows 'immediately upon publication' to extend to no more than 7 days following online publication of an article.

Access Denied: Content Item Not LIcenced (JR2, DB2, BR3)

Examples of events that can be used as a proxy for counting an Access Denied; Content Not Licenced.include:
- Abstract View, when the user does not have the right to continue to view the corresponding full-text article
- A 'Not Entitled' message

Note: Access denied to a user by a vendor temporarily, due to suspected abuse, should not be counted in JR2, DB2 or BR3.

Vendors should contact COUNTER if they have queries about the categories of event that may be used as proxy for Access Denied; Content Item Not Licenced

Journal Report 1b

The purpose of Journal Report 1b is to enable vendors to consolidate usage of journals from different Platforms in one report. All Platforms included in Journal Report 1b must be COUNTER compliant.

Journal Report 5

To enable Years of Publication (YOPs) that are part of a vendor/publisher archive to be distinguished from those that are not, an attribute is included in the XML to allow the archival YOPs to be flagged. This facilitates the automatic processing of JR5s by ERMs etc

Database Reports

Two new metrics, Result Clicks and Record Views, are included in the Release 4 database reports. Both terms are defined in Appendix A (Glossary of Terms). The purpose of these two new metrics is to provide customers with additional

insights into the usage and value of the databases they have purchased, beyond the 'Search' counts already reported.

Result Clicks simply report the number of times that users click on a set of search results, which is indicative of the value of such search results. The result click is counted irrespective of whether it takes the user to an internal record within the database searched, or to an external resource.

 Record Views count the number of times that textual records within the database(s) listed in a COUNTER Database Report are viewed (and not reported in other COUNTER Reports), irrespective of whether these records are reached from a set of search results or via browsing the database. Typical examples of Records to be counted in the Database Reports are: text abstracts, molecular structures, and chemical reactions. Full-text journal articles, book chapters, or multimedia full-content units are not treated by COUNTER as Records to be covered in the Database Reports and their usage should be reported in the relevant COUNTER Journal, Book or Multimedia reports.

Section 5: Data processing

'Only successful and valid requests should be counted. For webserver-logs successful requests are those with a specific return code. The standards for return codes are defined and maintained by NCSA.'
Requirement for Implementation:

Return codes that indicate a successful or valid request are specified in agreed, international web standards and protocols. The relevant governing document for hypertext protocols is RFC2068, which contains definitions for each Return Code number. There are five categories of return code numbers:

1xx (Information): this category provides information on a request, and often indicates that the user has come upon an experimental application.
2xx (Success): reserved for successful responses. This category of code is not usually seen by the user, but their browser will receive them and will know that whatever request was sent by the browser was received, understood and accepted.
3xx (Redirection): indicates the need for further action by the user's browser. User action may not be necessary, as the browser may deal with it automatically.
4xx (Client Error): this category of code is the one most frequently seen by the user and indicates an error.
5xx (Server Error): indicates where the server knows it has made an error, or is not capable of answering the request.

Categories **2xx** and **3xx** are relevant to Section 5a of the COUNTER Code of Practice, which deems that **only the following specific return codes indicate a successful or valid request:**

200 (OK) The request was successful and information was returned. This is, by far, the most common return code on the web.
304 (Not modified) In order to save bandwidth a browser may make a conditional request for resources. The conditional request contains an 'If-Modified-Since' field

4

and if the resource has not changed since that date the server will simply return the 304 code and the browser will use its cached copy of the resource.

Requests that result in any other return codes within the 2xx and 3xx categories must not be counted. This exclusion covers:

206 (Partial content) This indicates that the server has only filled part of a specific type of request.
301 (Moved permanently): The addressed resource has moved, and all future requests for that resource should be made to the new URL. Transfer to the new location may be automatic or may require manual intervention by the user. Either way, a successful request to the new location will generate a 200 return code.
302 (Moved temporarily) This indicates that the content has moved while the page requested still has the same URL. The page is, therefore, not retrieved and must not be counted.
303 (See other) The response to the browser's request can be found elsewhere. Automatic redirection may take place to the new location.

Full information on the five categories of http return codes and their definitions may be found at: http://www.w3.org/Protocols/rfc2068/rfc2068 goto: Chapter 10 (pp 53-64): Status Code Definitions. More summarised information may be found at: http://www.cknow.com/faqs/What/404andOtherHTTPReturnCode.html .

- **Guidelines for processing and filtering the raw usage data**

The filtering of the 'raw' usage data needs to go through a number of consecutive steps in order to meet the COUNTER requirements.

Step 1: Sorting the data file.

 The file to be used for reporting should be sorted chronologically by user.
 The following options for a user exist:

  1. Where only the IP address of a user is logged, that IP should be taken as the field to sort by.
  2. When a session-cookie is implemented and logged, the session-cookie should be used to sort by.
  3. When user-cookies are available and logged, the user-cookie should be used to sort by.
  4. When the username of a registered user is logged, this username should be used to sort by.

 Step 2: Remove all records with a return code other than '200' and '304'

 Step 3: Run the 'double-click-removal' script. (The time window for a double click may now be set at 30 seconds for all document types.)

The following example illustrates how this script should work:

A user requests the HTML format of one and the same item four times within the following time intervals:

Request 1:   9:51:10

Request 2:   9:51:39

Request 3:   9:52:12

Request 4:   9:52:34

Applying the double-click filter to the above example has the following results: comparing Requests 1 and 2 removes Request 1 and retains Request 2; next, comparing Request 2 with Request 3, retains both Request 2 and Request 3 as more than 30 seconds have elapsed between these two requests; next, comparing Request 3 with Request 4 removes Request 3 and retains Request 4, as less than 30 seconds have elapsed between Requests 3 and 4. Thus, applying the double-click filter to the above example results in two Successful Requests being recorded.

The same procedure applies to all formats of an item, including HTML, PDF, video clips, etc.

Guidelines for counting usage of individual sub-sections of an article

In JR-1, JR-1GOA, JR-1a, and JR5 only usage of full-text articles may be reported. Where a full-text article is downloadable in sub-sections, only the first successful request for a sub-section of a full-text article may be counted in JR-1, JR-1GOA, JR1-a and JR-5. Subsequent requests for other sub-sections of the same article within 30 seconds of each other must be filtered out. Article abstracts may not be reported in any of the above COUNTER Reports.

> Usage of individual sub-sections of full-text articles, as well as abstracts, may be reported in the optional reports JR-3, JR-3M, TR-3 and TR3-M.

- **Requirements for recording and reporting usage in Database Report 1 (DB-1) and Platform Report 1 (PR-1) when databases may be searched individually or as a group, as part of a Platform**

Many databases are offered by vendors as part of a platform that typically covers a number of databases.

Database Report 1 (DB-1) should report usage of each database, and Platform Report 1 (PR-1) should report usage of each platform, as illustrated in the following example:

A vendor offers a set of 5 databases (A, B, C, D and E) as part of Platform X. The user may search each database individually, or may specify searches across a group of 2 or more databases. Usage should be reported as follows in DB-1 and PR-1 for the scenarios below:

Scenario 1:
- o   User makes 1 search across 'All Databases'.
- o   Vendor should report:
    - ▪   1 search in PR-1 and **also**
    - ▪   1 search in DB-1 for **each** database (A, B, C, D and E)

Scenario 2:
- o   User makes 1 search in Database E only.
- o   Vendor should report:
    - ▪   1 search in PR-1 and **also**
    - ▪   1 search in DB-1 for Database E

Scenario 3:
- o   User makes 1 search, selecting two Databases (C and D) to search in.
- o   Vendor should report:
    - ▪   1 search in PR-1 and **also**
    - ▪   1 search for Database C in DB-1, and 1 search for Database D in DB-1

Results:

If usage is reported correctly for all three of the scenarios described above, then the vendor's DB-1 and PR-1 reports should report the following:

- PR-1 reports 3 searches in total for Platform X
- DB-1 reports 8 searches in total –
    - o   1 for Database A (from the search in Scenario 1)
    - o   1 for Database B (from the search in Scenario 1)
    - o   2 for Database C (from the searches in Scenarios 1 and 3)
    - o   2 for Database D (from the searches in Scenarios 1 and 3)
    - o   2 for Database E (from the searches in Scenarios 1 and 2)

- **Guidelines for identifying abnormal spikes in usage**

The Guidelines below are designed to provide simple, easily implementable, criteria that may be used to flag unusual spikes in usage, for further investigation. They may be applied to usage data in any of the monthly COUNTER usage reports.

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 one hundred units of measurement greater than 300% (three hundred percent) above the previous twelve month average.

<u>Negative Spike in Usage</u>: 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 percent) of the previous twelve month average usage. (Note: the average usage of that product in the previous twelve months should be at least twenty units of measurement).

The above approaches will provide only 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. It should be noted that negative spikes may occur as a result of relatively high usage in a different month which may or may not fulfil the requirements for a positive spike.

- **Protocol for ReadCube, Mendeley, and other tools for downloading and managing PDF files**

ReadCube and Mendeley are examples of tools designed to help researchers download and organize articles that are of interest to them. ReadCube, for example, incorporates a tool that allows users to download PDF versions of articles in a single step. When the user clicks on the PDF download button for that document he is presented with a choice of continuing to the regular PDF, or viewing the enhanced PDF. **Irrespective of which choice they make, the view is counted as a successful PDF request in the COUNTER reports**. The downloaded article can be saved as a regular PDF for later reading in Acrobat, or as an enhanced PDF to their desktop instance of ReadCube. **Any further usage is not recorded in the COUNTER reports. This protocol should also be followed for Mendeley and other similar tools.**

Appendix H (Optional Additional Usage Reports)

- **Protocol for detecting a mobile device**

The following optional additional reports enable usage on mobile devices to be reported separately:

<u>Journal Report 3 Mobile: Number of Successful Item Requests by Month, Journal and Page Type for usage on a Mobile Device</u>

<u>Title Report 1 Mobile: Number of Successful Requests for Journal Full-text Articles and Book Sections by Month and Ttle (formatted for normal browsers/delivered to mobile devices AND formatted for mobile devices/delivered to mobile devices)</u>

<u>Title Report 3 Mobile: Number of Successful Requests by Month, Title and Page Type (formatted for normal browsers/delivered to mobile devices AND formatted for mobile devices/delivered to mobile devices)</u>

COUNTER will recognize as usage on a mobile device, which may be reported in the above reports, any usage that meets one of the following criteria:

- useragents that are included in the WURFL list. WURFL is the **W**ireless **U**niversal **R**esource **FiL**e, a database containing the profile of mobile devices; this database may be found at: http://wurfl.sourceforge.net/
- usage via a proprietary mobile App provided by the publisher/content provider

Vendors must indicate on their sites, preferably in the usage reporting area, which of the above criteria they are using to identify mobile devices.