## Appendix F: Audit Requirements and Tests

# 1 Audit Principles

The COUNTER audit principles, procedures and tests set out in this Appendix seek to ensure that the usage reports for articles provided by vendors are in line with the COUNTER principles of credibility, consistency and comparability and follow uniform agreed procedures.

To this end, the audit seeks to provide assurance in the following key component areas:

CA1 - Formatting and logging of data
CA2 - Storage, filtering and processing of data
CA3 - Formatting and production of the three Article Reports – AR-1 (incl. AR-1j), AR-2 and AR-3 – set out in the COUNTER Code of Practice for Articles (CoP-A).

Each of these component areas has a design stage, where the audit gives assurance to COUNTER on the approach, systems and processes set up by the auditee, and an operation stage, where the auditee's systems are operationally tested by the auditor. This is to ensure that compliance is checked and issues resolved at the earliest possible opportunity, as well as to allow efficiencies in future audit work. Both stages need to be passed for COUNTER compliance.

Auditees are expected to fall into three main types:

- Repositories, for whom the audit generally will cover CA1
- Publishers and vendors, for whom the audit generally will cover CA1, CA2 and CA3
- Clearing houses, for whom the audit generally will cover CA2 and CA3.

| | CA1 | CA2 | CA3 |
|---|---|---|---|
| Repositories | Yes | ***occasionally*** | ■ |
| Publishers and vendors | Yes | Yes | Yes |
| Clearing houses | ***occasionally*** | Yes | Yes |

Hence, full compliance by a process involving multiple auditees is only achieved when ALL the boxes above have been ticked separately, e.g. by a single unique combination of repository and clearing house.

Prior to the audit, each auditee should discuss with the auditor and agree with COUNTER which components of the audit apply to them, so that the right audit tests can be carried out and the auditor's opinion is appropriately expressed to the COUNTER EC.

Any service or vendor providing reporting (CA3) must have already passed an audit proving their compliance with the Journal reports in Release 4 of the COUNTER Code of Practice for E-resources (CoP-E) before they are able to undergo an audit to the Code of Practice for Articles.

# 2 Description of audit tests

## 2.1 CA1 – Formatting and logging of data

The tests in this area seek to provide assurance in the effective design and operation of the auditee's gathering and storage of raw data for subsequent processing.

### *CA1.1 Identification of institutions and authors*

**Design:**
The auditee will be expected to show that the data format logs institutional and ORCID identifiers consistently and effectively. Furthermore, the auditee will need to explain and document the processes by which their data logging integrates with a valid ISNI institutional identifier provider (such as Ringgold) and ORCID to ensure that the identifiers logged are as up to date as possible.

**Operation:**
The auditor will generate usage from a variety of institutions for articles by a number of authors and check that the institutional and ORCID identifiers for each are recorded correctly.

### *CA1.2 Identification of articles*

**Design:**
The CoP-A applies to the measurement and reporting of only **articles,** as defined in Appendix A. Hence, the auditee will be expected to show that they have consistent and reliable means of identifying articles in the data as opposed to, for example, non-textual material or theses. Furthermore, the auditee will be expected to ensure that article Digital Object Identifiers (DOIs) are consistently and effectively logged.

**Operation:**
The auditor will generate usage for articles and, at their discretion, other resources, and check that article DOIs are recorded correctly in the available data or reports. The auditor may also check that articles without DOIs are not counted; this test will also assist in verifying CA2.1.

### *CA1.3 Compliant log format containing all required fields*

**Design:**
Data supporting at least the minimum required fields must be logged in a format that logs one set of fields per request. The following table summarises these minimum requirements and is laid out in the order of the fields in the template AR-1 report (page 10 of the CoP-A or Appendix E). Further guidance on server log fields is given in CoP-E Appendices C and D.

Note in particular that the last 6 fields in the table below are essential for compliant processing of data, even though they will never be displayed in any report.

| Field | Required? |
|---|---|
| **Journal (or book) title** | *No* |
| **Publisher** | *No* |
| **Platform** | *No* |
| **Journal (or book) DOI** | *No* |
| **Print ISSN** | **At least one** of these must be logged |
| **Online ISSN** | |
| **Article title** | *No* |
| **Article type** | *No* |
| **Article version** | *Yes* |
| **Article DOI** | *Yes* |
| **ORCID identifier** | *Yes* |
| **Online Publication Date** | *No* |
| **First successful request** | *Yes* |
| **Date (of current request)** | *Yes* |
| **IP address** | *Yes* |
| **User-Agent** | *Yes* |
| **HTTP Referrer** | *Yes* |
| **URL** | *Yes* |
| **HTTP Method** | *Yes* |
| **HTTP Status** | *Yes* |

The auditee will need to explain the methodology used for capturing data, whether server logs, browser-side analytics logs, tracker protocol, or OAI-PMH protocol. Finally, the auditee will need to explain any server configuration rules that are applied on an incoming request before it is logged (e.g. not logging requests with null DOIs, robotic user-agents or invalid HTTP status codes). While this is related to the filtering checks set out in CA2, CA1.3 specifically excludes subsequent post-processing of the data.

### Operation:

The auditor will seek to confirm that logging continues to be operationally compliant across all 24 months of data required for processing. Log files will be checked throughout the audit period to verify that all required fields are logged and that any incoming filtering operates as designed. The auditor may also ask for descriptions of, and test, any changes to previously seen log formats.

## *CA1.4 Data feeds to third parties*

### Design:

The auditee will need to explain how reliability of data supply to third party data processors or aggregators (e.g. publishers or clearing houses) is assured. This may include policies on server maintenance and redundancy, data upload and transfer procedures, log rotation, server capacity, maximum filesize limits and similar measures to mitigate risks around data supply.

### Operation:

The auditor will seek to confirm that data feeds to third parties continue to be reliable by reviewing records of data transfer (such as availability and size of files) throughout the audit period. The auditor may ask for evidence of how any issues with the data supply process were found and resolved.

## 2.2 CA2 – Storage, filtering and processing of data

The tests in this area seek to provide assurance that the processing of the auditee's data meets the many specific requirements throughout the CoP-A (in the summary list below), and that the auditee has stored and filtered the 24 months of data required to produce COUNTER Articles reports.

### *CA2.1 Post-processing of data*

Records considered compliant for measurement under the CoP-A must have had the following filters applied. This table explains the reasoning behind each filter and the fields that the filter applies to:

| **Filtering Requirement** | **Field(s) for filtering checks** |
|---|---|
| **Correct versions only are recorded OR** | *Article Version* |
| **Only usage of articles is recorded** | *Item Type or DOI* |
| **Articles lacking both print ISSN and online ISSN are removed** | *At least one of print and online ISSN must be populated* |
| **Only requests with HTTP status 200 or 304 may be counted** | *HTTP Status* |
| **Only requests for content (GET or POST) may be counted** | *HTTP Method* |
| **The first of a pair of requests by an identified "User" for the same item within the double-click filter time window must be removed** | *Date+Time*<br>*User identifier (e.g. cookie, IP+User-Agent)* |
| **Article must be associated with a valid ORCID identifier** | *ORCID identifier* |
| **Article must have a DOI** | *DOI* |
| **Robotic User-Agents must be excluded** | *User-Agent* |
| **Federated searches must be excluded** | *Identifier of federated search client (typically one or both of IP and User-Agent)* |
| **Outliers and "gaming traffic" must be removed** | *Multiple – may include User identifier, IP, User-Agent, Date+Time, URL, Referrer* |
| ***[QUESTION: Internal traffic e.g. institutional repositories?]*** | ***[this may be necessary but may also be dealt with by the gaming protocol; advice sought from COUNTER]*** |

#### Design:

The auditee will explain the processes to remove non-compliant records from the raw data, in particular the flow of filtering and the steps undertaken to reconcile totals and identify possible errors. The auditee must also provide the policies on removing suspected "gaming" traffic as defined in the CoP-A.

#### Operation:

If audit component CA3 (reporting) is within the scope of the audit, the auditor will test filtering effectiveness by making requests that do not meet the filtering requirements, and testing that these requests are not reported.

If CA3 is not being checked, the auditor will need to request sample log files to verify that the filtering continues to operate as designed. Hence, this aspect of the audit is made more efficient by testing CA2 in conjunction with CA3.

### *CA2.2 Reliability of incoming data and storage of data*

#### Design:

The auditee will need to explain the checks put in place to ensure consistent receipt and filtering of incoming data, as well as storage of the post-processed data. This may include policies on server maintenance and redundancy, re-processing of data in the event of errors being found further downstream (such as in report generation), data transfer procedures, and verification of each stage of filtering.

**Operation:**
The auditor will confirm that incoming data is received and stored reliably by testing the processes described in the Design stage, and may review any reported incidents of missing or reduced data, as well as relevant follow up actions taken by the auditee.

## 2.3 CA3 – Formatting and production of reports

This part of the audit sets out to verify compliance of the reports generated by the auditee with the formatting and file access requirements of the CoP-A.

### *CA3.1 General report design and accessibility*

**Design:**
The auditee will need to provide specimen reports to be reviewed against the templates shown in the CoP-A (sections 4.5 and 4.6 as well as Appendix E). Specimens of ALL reports (AR-1, AR-1j, AR-2 and AR-3) must be delivered as specified by the Code of Practice (as separate XML and Excel-compatible files).

The auditee must also explain the process by which correctly formatted reports can be retrieved via SUSHI. Auditees should ensure that different (sample) hosts can be shown in their sample AR-2 and different article sources (journals, books etc.) in their sample AR-3. The auditor will at this stage seek to verify how many different hosts feed into the auditee's AR-2 report so as to plan for the operational testing below.

**Operation:**
The auditor will confirm that reports are accessible, correctly formatted and available in the file types required, by downloading reports within 3 months of calendar year end. The auditee must therefore supply access credentials to their reporting area in time for this check to be done. The auditor will also verify that usage data going back 24 months is available. This may mean that reports generated for other institutions may be checked if the auditor has not been able to generate usage across the 24 months themselves.

### *CA3.2 Accurate reporting of usage*

**Design:**
The auditee will explain the processes by which reports are generated from compliant data and demonstrate that these are stable and clearly documented. The auditee will also describe the logins to the system to generate the operational usage activity covered below.

**Operation:**
The auditor will check that the usage statistics reported by the vendor accurately record the activity carried out by the auditor, behaving as an institution or institutions with an ISNI identifier, during the 24 month period. This includes checking that the vendor provides consistent usage statistics when its reports are accessed using the latest versions of the three different browsers required by COUNTER (currently Google Chrome, Internet Explorer and Mozilla Firefox).

This testing can be carried out as an additional facet of the COUNTER JR-1 testing set out in Appendix E of the COUNTER E-Resources Code of Practice. However, given the longer period of generation of COUNTER Articles reports (24 months' data required as opposed to a single month) the opinion may not be listed at the same time.

# 3 Audit frequency and timing

## 3.1 Audit frequency

Every organisation that wishes to remain compliant with one or more of the areas of the COUNTER Articles audit for a given year must pass an audit for that year.

The **first audit** of an audit component (CA1, CA2 or CA3) must include both design and operational tests.

Since the COUNTER Articles reports are produced yearly and require processing of 24 months of data, elements of the audit may overlap with a previous audit if one has taken place.

Failure to meet these audit requirements will result in a vendor losing its listing on the COUNTER Articles Register of Compliant Vendors.

### *3.1.1 Design tests*

These tests must be carried out in full in the first audit. In following years, the auditee must inform the auditor of any changes to the systems under audit; any such changes may necessitate additional design tests.

### *3.1.2 Operational tests*

Operational tests must take place **yearly** as a minimum. Operational tests in a component area can only be started once the design of that area has been validated.

## 3.2 Expected sequence of audit tests over time

| Component | Design testing | Operational testing |
|---|---|---|
| **CA1 (formatting and logging of data)** | Before the year of reporting | Before or during the year of reporting |
| **CA2 (storage, filtering and processing of data)** | Before the year of reporting | A few times during the year of reporting |
| **CA3 (formatting and production of reports)** | Before the year of reporting | A few times during the year of reporting |

# 4 Required audit outputs

## 4.1 Audit results

Once each stage of an audit component (CA1, CA2, CA3) has been completed, the auditor will report the results as PASS, QUALIFIED PASS or FAIL.

| | |
|---|---|
| **PASS** | The auditor is satisfied that the aspects of the COUNTER Articles Code of Practice being tested **are in full compliance** with the requirements of the Code.<br>No further action is required by the auditee as a result of the audit. In some cases the auditor may add Observations to the audit report to help the auditee improve its processes. |
| **QUALIFIED PASS** | The auditor deems the vendor or service to have passed the audit, but has raised a Minor Issue requiring further action to maintain a valid listing in the COUNTER Articles Register. A Minor Issue does not affect the reported figures or necessitate further testing, but should be resolved within 3 months of the audit to COUNTER's satisfaction. An example of a Minor Issue is where a report format does not conform to the COUNTER Articles specifications. |
| **FAIL** | The auditor has identified an issue that **must be resolved** for the vendor or service to maintain a valid listing in the COUNTER Articles Register. This typically means either:<br>1. The processes in place are not sufficient to assure compliance with the requirements of the Code, and **require major changes** within 3 months of the FAIL being issued to become compliant. Further audit work will be required to validate the changes, and if this work continues to identify a FAIL the auditee will be removed from the COUNTER Articles Register.<br>Or<br>2. An unresolvable issue (e.g. missing data or log fields) has been found that will invalidate all reporting until new and improved systems are put in place for the next reporting period. The auditee will be removed from the COUNTER Articles Register until the next audit has been passed. |

## 4.2 Reporting to COUNTER

When results are issued, the auditor will provide the auditee and COUNTER with a summary report providing the following information:

- The name of the auditee
- The component area(s) and stage(s) of the audit undertaken (e.g. CA1 Design, CA2 Design)
- The date
- A summary of any significant issues noted in the audit, and recommendations and timelines for resolution
- A clear indication of the overall outcome for the audit (PASS, QUALIFIED PASS, or FAIL)

- Any other comments relating to the audit worthy of consideration by the COUNTER EC.

For an auditee to achieve a listing in the COUNTER Articles Register of Vendors, **both design and operational testing must PASS**. The auditee must pass subsequent annual audits to maintain their valid listing in the COUNTER Articles Register.