The COUNTER Code of Practice for Articles

Release 1

Appendix C: Specification for the Collection of Individual Item Usage Data

Data Gathering and Processing

There are three scenarios – with supporting protocols and standards - for the transmission of usage data and statistics:

Scenario (A): when a full-text article is downloaded, a message – raw usage data - is pushed out to a remote server
- protocol: 'tracker' – analogous to a server-side 'Google Analytics' for full-text article downloads
- standard: OpenURL key-value pair strings (URLs)
- candidate organisations: most repositories and some small publishers

Scenario (B): as full-text articles are downloaded, records of raw usage data events are stored locally and made available for harvesting by a remote server, on demand
- protocol: OAI-PMH – a protocol already familiar to repositories
- standard: OpenURL context objects (XML)
- candidate organisations : repositories

Scenario (C): as full-text articles are downloaded, records of raw usage data events are stored locally. Usage data is processed according to COUNTER rules, and made available for harvesting by a remote server, on demand
- protocol: SUSHI (Standardized Usage Statistics Harvesting Initiative Protocol) –familiar to publishers
- standard: proposed COUNTER-compliant AR1 report
- candidate organisations: publishers

A 'protocol' defines a set of rules to send and receive messages between computers on the internet; while a 'standard' defines the rules for the content of those messages.

The standards and protocols, central to the project are discussed in the following section.

A. Gathering Usage Events

1. COUNTER-compliant Publisher and COUNTER-compliant Repository usage events (Scenario C)

Article Report 1 (below) provides a standard format for the collection and consolidation of article-level usage data from COUNTER-compliant publishers and COUNTER-compliant repositories. PIRUS Reports must be provided in XML format, in accordance with the specified schema, in order to

facilitate the automated consolidation of article-level usage data from different sources. For visualisation purposes an example is provided below of Article Report 1.

Article Report 1: Format for the collection and consolidation of article-level usage data

Note:

1. Author Identifier may be the publisher's own author identifier
   (the ORCID Identifier will be the preferred option once it is implemented)
2. Article title data is highly recommended, but opional
3. Usage data should include:
   a. **Include:** successful full-text requests (HTML plus PDF)
   b. **Include:** Accepted Manuscript, Proof, Version of Record versions
   c. **Exclude**: Author's Original Manuscript and Submited Manuascript Under Review versions
   d. **Exclude:** any internal use by publisher and host, downloads from LOCKSS caches, and usage driven by robots

2. Non-COUNTER-compliant Repository usage events

Relying on the majority of repositories to gather usage data in the form of Article Report 1 (Scenario C) is not feasible, for the following reasons:

- The technology is unfamiliar to most repositories
- There are considerable auditing cost and data preservation implications in producing ready-made COUNTER-compliant reports

Scenarios A & B are, therefore, more applicable to repositories, where they can share raw usage data, while the audit and preservation responsibilities sit with the Central Clearing House.

With so many types of software available to run repositories, it is not possible to develop a common software plugin which will work for all types. However, in 'Appendix D – Protocols for Repositories', we do offer guidance on implementing the Tracker and OAI-PMH protocols to support scenarios A and B.