

An XML DTD for Counter-compliant web usage reports

Version 0.5: Discussion draft

Timo Hannay <t.hannay@nature.com>

19 May 2004

The following is a description of a draft DTD for holding Counter-compliant web usage reports. Please send feedback to the author or to the project director, Peter Shepherd <pshepherd@projectcounter.org>.

DTD URL AND NAMESPACE

- The definitive DTD resides at <http://www.projectcounter.org/dtd/2004/reports.dtd>
- The namespace for the DTD is <http://www.projectcounter.org/ns/2004/reports>. A suggested abbreviation for namespace prefixes is "counter".

THE ROOT ELEMENT

- **reports:** The root element, *reports*, can contain any number of journal and database reports (using the *journal_report* and *database_report* elements, respectively). It would normally be used to send one or more reports from a single vendor to a single customer but in principle it can be used to hold multiple reports relating to any number of different vendors and customers.

THE REPORT ELEMENTS

- **journal_report and database_report:** These elements always contain exactly one *header* element (which carries general information about the report, the vendor and the customer) and one *data* element (which carries the actual usage data). They both have a compulsory *id* attribute. They may also optionally include a single *cop_version* attribute and/or a single *cop_report* attribute.
 - **id:** This attribute must contain a unique identifier for the report. In order to maintain uniqueness of identifiers across vendors, the use of the following format is recommended: `<domain_name>_<local_identifier>`. The `<domain_name>` part should be an internet domain over which the vendor has control (e.g., example.com). The `<local_identifier>` part can be any locally unique XML-parsable string that the vendor chooses.
 - **cop_version:** This attribute is intended to hold information about the version number of the Counter Code of Practice to which the report corresponds. It is optional because reports other than those specified by Counter can be encoded with this DTD.

- **cop_report:** This attribute is intended to hold information about the Counter Code of Practice report number to which the report corresponds. It is optional because reports other than those specified by Counter can be encoded with this DTD.

THE HEADER SECTION ELEMENTS

- **header:** This element always contains exactly one of each of the *title*, *vendor* and *customer* elements.
 - **title:** This element is intended to hold a human-readable title for the report in question. It may contain any XML-parsable text.
 - **timestamp:** This element is intended to hold the date and time of the report's creation (thus allowing, for example, newer versions of the same report to supersede older ones). It may contain any XML-parsable text but timestamps corresponding to the long ISO8601 format (*YYYY-MM-DDThh:mm:ssZ*) are recommended. To avoid ambiguity, UTC times are recommended.
- **vendor:** This element must contain exactly one *vend_name* element. It may optionally contain one *vend_imprint* element, one *vend_logo* element and any number of *vend_site* and *vend_contact* elements.
 - **vend_name:** This element is intended to hold the name of the vendor. It may contain any XML-parsable text.
 - **vend_imprint:** This element is intended to hold the name of the relevant imprint or comparable vendor-owned brand name. It may contain any XML-parsable text.
 - **vend_site:** This element is intended to hold the full URL of the vendor's website. It may contain any XML-parsable text.
 - **vend_contact:** This element is intended to hold an email address or other contact information for the vendor. It may contain any XML-parsable text.
 - **vend_logo:** This is an empty (i.e., self-closing) element intended to hold information about a separate graphics file that contains the vendor's logo. The *url* attribute is required and gives the fully-qualified web ('http:') address of a graphics file containing the vendor's logo. The *format* attribute uses a value from the following list: *bmp*, *doc*, *eps*, *gif*, *jpeg*, *ole*, *pdf*, *pict*, *svg*, *tiff* and *wmf*. A value of "bmp" (bitmap) is assumed in the event that none is specified. There is an *other_type* attribute that can be used to explicitly state any type that is not in the list provided.

- **customer:** This element must contain exactly one *cust_name* element and can optionally contain one *cust_ref* element as well as any number of *cust_ip*, *cust_username* and *cust_criteria* elements.
 - **cust_name:** This element is intended to hold the name of the customer and may contain any XML-parsable text.
 - **cust_ref:** This element is intended to hold a reference used by the vendor to uniquely identify the customer. It may contain any XML-parsable text.
 - **cust_ip:** This element is intended to hold a information about the IP address range to which this report corresponds (particularly where access is provided on the basis of IP address). There can be any number of *cust_ip* elements because several non-contiguous IP address ranges may have to be listed. There are several ways of defining IP address ranges. The particular method used must be specified in the *type* attribute using the appropriate term given in italics below:
 - Single IP address (*single*): 123.124.45.0
 - Explicit range (*range*): 123.124.45.0-123.124.45.255
 - Wildcard (*wildcard*): 123.124.45.*
 - Subnet mask (*subnet*): 123.124.45.0:255.255.255.0
 - Classless Inter-Domain Routing (*cidr*): 123.124.45.0/8
 - **cust_username:** This element is intended to hold the vendor-, service-, database- or journal-specific username to which this report corresponds (particularly where access is provided on the basis of a username). There can be any number of *cust_username* elements because several usernames may have to be listed. It may contain any XML-parsable text.
 - **cust_criteria:** This element is included in anticipation of Release 2 of the Code of Practice. It is intended to contain information about, for example, the organizational level (institution, department, etc.) to which the usage report refers. The DTD requires only that it contain XML-parsable text.

THE DATA SECTION ELEMENTS

- **journal_data:** This element may contain an optional *totals* element plus any number of *journal* or *service* elements.
- **database_data:** This element may contain an optional *totals* element plus any number of *database* or *service* elements.

- **totals:** This element is intended to hold the sum total data for any given report. It may contain any of the following elements: *requests*, *turnaways*, *searches* and *sessions*.
- **journal:** This element is intended to hold data associated with a particular journal. It may contain any of the following elements: *requests*, *turnaways*, *searches* and *sessions*. It also has a compulsory attribute that holds the journal's name as well as optional attributes to hold the print and online ISSNs and the session timeout period (in minutes). In line with the Counter Code of Practice, the DTD specifies a default timeout period of 30 minutes for XML documents that don't explicitly state a value.
- **service:** This element is intended to hold data associated with a particular journal or database service. It may contain any of the following elements: *requests*, *turnaways*, *searches* and *sessions*. It also has a compulsory attribute that holds the name of the service as well as an optional attribute to hold the session timeout period (in minutes). In line with the Counter Code of Practice, the timeout period is assumed to be 30 minutes for XML documents that don't explicitly state a value.
- **database:** This element is intended to hold data associated with a particular database. It may contain any of the following elements: *requests*, *turnaways*, *searches* and *sessions*. It also has a compulsory attribute that holds the name of the database and an optional attribute that holds the session timeout period (in minutes). In line with the Counter Code of Practice, the DTD specifies a default timeout period of 30 minutes for XML documents that don't explicitly state a value.

THE DATA ELEMENTS

- **requests:** This element is intended to hold the number of successful page requests (as defined by the Counter Code of Practice) served in a specified period of time. The period of time is defined by the compulsory *start* and *end* attributes. These will usually be stated to the nearest day, in which case the short ISO8601 format (*YYYY-MM-DD*) is recommended and the period is assumed to be inclusive of the dates given. In principle, *start* and *end* values can include times as well as dates, in which case the long ISO8601 format (*YYYY-MM-DDThh:mm:ssZ*) is recommended. There is also an optional attribute, *type*, to define the type of page, which is required for some Counter reports. The value must come from the following list: *toc*, *abs*, *refs*, *ft_ps*, *ft_html*, *ft_pdf*, *ft_total* and *other*. If a value of *other* is used then the additional attribute *other_type* should be used to give the full name of this page type.
- **turnaways:** This element is intended to hold the number of turnaways (as defined by the Counter Code of Practice). One attribute, *reason*, is included in anticipation of Release 2 of the Code of Practice and is intended to capture

information about why a user was turned away. Valid options are *exceeded_user_limit*, *no_licence*, *system_error* and *other*. Where the *other* option is used, the type should be explicitly stated using the *other_reason* attribute. Where a value is not specified, a default of *exceeded_user_limit* is assumed (since this corresponds to the definition of turnaways in Release 1 of the Counter Code of Practice). In every other way the use of this element is identical to the *requests* element above.

- **searches:** This element is intended to hold the number of searches run (as defined by the Counter Code of Practice) during a specific period of time. The use of the *start* and *end* attributes is identical to that for the *requests* element above.
- **sessions:** This element is intended to hold the number of sessions (as defined by the Counter Code of Practice) during a specific period of time. The use of the *start* and *end* attributes is identical to that for the *requests* element above.