

Network Working Group
Request For Comments: 1807
Obsoletes: 1357
Category: Informational

R. Lasher
Stanford
D. Cohen
Myricom
June 1995

A Format for Bibliographic Records

Status of this Memo

This memo provides information for the Internet community. This memo does not specify an Internet standard of any kind. Distribution of this memo is unlimited.

Abstract

This RFC defines a format for bibliographic records describing technical reports. This format is used by the Cornell University Dienst protocol and the Stanford University SIFT system. The original RFC (RFC 1357) was written by D. Cohen, ISI, July 1992. This is a revision of RFC 1357. New fields include handle, other_access, keyword, and withdraw.

Introduction

Many universities and other R&D organizations routinely announce new technical reports by mailing (via the postal services) the bibliographic records of these reports.

These mailings have non-trivial cost and delay. In addition, their recipients cannot conveniently file them, electronically, for later retrieval and searches.

Publishing organizations that wish to use e-mail or file transfer to obtain these announcements can do so by using the following format.

Organizations may automate to any degree (or not at all) both the creation of these records (about their own publications) and the handling of the records received from other organizations.

This format is designed to be simple, for people and for machines, to be easy to read ("human readable") and create without any special programs.

This RFC defines the format of bibliographic records, not how to process them.

This format is a "tagged" format with self-explaining alphabetic tags. It should be possible to prepare and to read bibliographic records using any text editor, without any special programs.

This RFC includes the CR-CATEGORY, a field useful for Computer Science publications. It is expected that similar fields will be added for other domains.

This format, as described in RFC 1357, was implemented as part of the Dienst system and has been in use by the five ARPA-funded computer science institutions to exchange bibliographic records (Cornell, SU, UC, MIT, and CMU). Programs have been written to map between this RFC and structured USMARC (format developed at the Library of Congress) cataloging records, also from USMARC to the RFC.

The focus of this ARPA-funded research has been into many aspects of digital libraries including searching and accessing techniques that do not necessarily use bibliographic records (for example, natural language processing, automatic and full-text indexing). However, the continued use of bibliographic records is expected to remain an important part of the library system environment of the future and its use is an important link between the physical world of scientific works and the on-line world of digital objects. The format described in this paper allows a link between these two worlds to be created.

This format was developed with considerable help and involvement of Computer Science and Library personnel from several organizations, including Carnegie Mellon University, Corporation for National Research Initiatives (CNRI), Cornell University, University of Southern California/Information Sciences Institute (ISI), Meridian (now called DynCorp), Massachusetts Institute of Technology, Stanford University, and the University of California. Key contributions were provided by Jerry Saltzer of MIT, and Larry Lannom of DynCorp. The initial draft was prepared by Danny Cohen and Larry Miller of ISI. The revision was done by Rebecca Lasher from Stanford with assistance from the CS-TR participants.

This RFC does not place any limitations on the dissemination of the bibliographic records. If there are limitations on the dissemination of the publication, it should be protected by some means such as passwords. This RFC does not address this protection.

The use of this format is encouraged. There are no limitations on its use.

The Information Fields

The various fields should follow the format described below.

<M> means Mandatory; a record without it is invalid.

<O> means Optional.

The tags (aka Field-IDs) are shown in upper case.

- <M> BIB-VERSION of this bibliographic records format
- <M> ID
- <M> ENTRY date
- <O> ORGANIZATION
- <O> TITLE
- <O> TYPE
- <O> REVISION
- <O> WITHDRAW
- <O> AUTHOR
- <O> CORP-AUTHOR
- <O> CONTACT for the author(s)
- <O> DATE of publication
- <O> PAGES count
- <O> COPYRIGHT, permissions and disclaimers
- <O> HANDLE
- <O> OTHER_ACCESS
- <O> RETRIEVAL
- <O> KEYWORD
- <O> CR-CATEGORY
- <O> PERIOD
- <O> SERIES
- <O> MONITORING organization(s)
- <O> FUNDING organization(s)
- <O> CONTRACT number(s)
- <O> GRANT number(s)
- <O> LANGUAGE name
- <O> NOTES
- <O> ABSTRACT
- <M> END

Meta Format

- * Keep It Simple.
- * One bibliographic record for each publication, where a "publication" is whatever the publishing institution defines as such.
- * A record contains several fields.
- * Each field starts with its tag (aka the field-ID) which is a reserved identifier (containing no separators) at the beginning of a new line with or without spaces before it), followed by two colons ("::"), followed by the field data.
- * Continuation lines: Lines are limited to 79 characters. When needed, fields may continue over several lines, with an implied space in between. In order to simplify the use no special marking is used to indicate continuation line. Hence, fields are terminated by a line that starts (apart from white space) with a word followed by two colons. Except for the "END::" that is terminated by the end of line.) For improved human readability it is suggested to start continuation lines with some spaces.
- * Several fields are mandatory and must appear in the record. All fields (unless specifically not permitted to) may be in any order and may be repeated as needed (e.g., the AUTHOR field). The order of the repeated fields is always preserved.
- * Only printable ASCII characters are to be used. The permissible characters are ASCII codes 040 (Space) through 176(~) and line breaks which are \012 (LF) or \012\015 (CRLF). Empty lines indicate paragraph break. \009 (tab) must be replaced by spaces. This specifically forbids tabs, null characters, DEL, backspaces, etc. (i.e., if used, the record is invalid.)

However full 8 bit ASCII may be used. WARNING: some electronic mailers cannot handle 8 bit ASCII and these records may need to be transported via other mechanisms.

Throughout this document the word "publisher" means the publishing organization of a report (e.g., a university or a department thereof), not necessarily an organization authorized to issue ISBN numbers.

EXAMPLE

```
-----
BIB-VERSION:: CS-TR-v2.1
            ID:: OUKS//CS-TR-91-123
            ENTRY:: January 15, 1992
ORGANIZATION:: Oceanview University, Kansas, Computer Science
            TYPE:: Technical Report
REVISION:: January 5, 1995; FTP access information added
            TITLE:: Scientific Communication must be timely
            AUTHOR:: Finnegan, James A.
            CONTACT:: Prof. J. A. Finnegan, CS Dept, Oceanview Univ,
                    Oceanview, KS 54321 Tel: 913-456-7890
                    <Finnegan@cs.ouks.edu>
            AUTHOR:: Pooh, Winnie The
            CONTACT:: 100 Aker Wood
            DATE:: December 1991
            PAGES:: 48
COPYRIGHT:: Copyright for the report (c) 1991, by J. A.
            Finnegan. All rights reserved. Permission is granted
            for any academic use of the report.
            HANDLE:: hdl:oceanview.electr/CS-TR-91-123
OTHER_ACCESS:: url:http://electr.oceanview.edu/CS-TR-91-123
OTHER_ACCESS:: url:ftp://electr.oceanview.edu/CS-TR-91-123
RETRIEVAL:: send email to Finnegan@cs.ouks.edu with fax number
            KEYWORD:: Scientific Communication
CR-CATEGORY:: D.0
CR-CATEGORY:: C.2.2 Computer Sys Org, Communication nets, Net
            Protocols
            SERIES:: Communication
            FUNDING:: FAS
            CONTRACT:: FAS-91-C-1234
MONITORING:: FNBO
            LANGUAGE:: English
            NOTES:: This report is the full version of the paper with
                    the same title in IEEE Trans ASSP Dec 1976
ABSTRACT::
```

Many alchemists in the country work on important fusion problems. All of them cooperate and interact with each other through the scientific literature. This scientific communication methodology has many advantages. Timeliness is not one of them.

```
END:: OUKS//CS-TR-91-123
```

```
----- End of Example -----
```

For reference, the above example has about 1,689 characters (184 words) including about 249 characters (36 words) in the abstract.

The Actual Format

The term "Open Ended Format" in the following means arbitrary text.

In the following double-quotes indicate complete strings. They are included only for grouping and are not expected to be used in the actual records.

The BIB-VERSION, ID, ENTRY, and END field must appear as the first, second, third, and last fields, and may not be repeated in the record. All other fields may be repeated as needed.

BIB-VERSION (M) -- This is the first field of any record. It is a mandatory field. It identifies the version of the format used to create this bibliographic record. This RFC defines BIB-Version TR-v2.1

BIB-VERSIONs that start with the letter X (case independent) are considered experimental. Bib-records sent with such a BIB-VERSION should NOT be incorporated in the permanent database of the recipient.

Using this version of this format, this field is always:

Format: BIB-VERSION:: CS-TR-v2.1

ID (M) -- This is the second field of any record. It is also a mandatory field. The ID field identifies the bibliographic record and is used in management of these records. Its format is "ID:: XXX//YYY", where XXX is the publisher-ID (the controlled symbol of the publisher) and YYY is the ID (e.g., report number) of the publication as assigned by the publisher. This ID is typically printed on the cover, and may contain slashes.

The organization symbols "DUMMY" and "TEST" (case independent) are reserved for test records that should NOT be incorporated in the permanent database of the recipients.

Format: ID:: <publisher-ID>//<free-text>

Example: ID:: OUKS//CS-TR-91-123

**** See the note at the end regarding the ****
**** controlled symbols of the publishers *****

ENTRY (M) -- This is a mandatory field. It is the date of creating this bibliographic record.

The format for ENTRY date is "Month Day, Year". The month must be alphabetic (spelled out). The "Day" is a 1- or 2-digit number. The "Year" is a 4-digit number.

Format: ENTRY:: <date>

Example: ENTRY:: January 15, 1992

ORGANIZATION (O) -- It is the full name spelled out (no acronyms, please) of the publishing organization. The use of this name is controlled together with the controlled symbol of the publisher (as discussed above for the ID field).

Avoid acronyms because there are many common acronyms, such as ISI and USC. Please provide it in ascending order, such as "X University, Y Department" (not "Y Department, X University").

Format: ORGANIZATION:: <free-text>

Example: ORGANIZATION:: Stanford University, Department of
Computer Science

TITLE (O) -- This is the title of the work as assigned by the author. This field should include the complete title with all the subtitles, if any.

Format: TITLE:: <free-text>

Example: TITLE:: The Computerization of Oceanview with
High Speed Fiber Optics Communication

TYPE (O) -- Indicates the type of publication (summary, final project report, etc.) as assigned by the issuing organization.

Format: TYPE:: <free-text>

Example: TYPE:: Technical Report

REVISION (O) -- Indicates that the current bibliographic record is a revision of a previously issued record and is intended

to replace it. Revision information consists of a date and/or followed by a semicolon and by text in an open ended format. The revised bibliographic record should contain a complete record for the publication, not just a list of changes to the old record. If revision is omitted, the record is assumed to be a new record and not a revision. If the revision date is specified as 0, this is assumed to be January 1, 1900 (the previous RFC, used revision data of 0, 1, 2, 3, etc. this specification is for programs that might process records from RFC1357).

The text before the semicolon in this field is a date of the form month day, year. Any record with a more recent revision date replaces completely any record with an earlier revision date (supplied either explicitly or by default). Use the text to describe the revision. Reasons to send out a revised record include an error in the original, or change in the access information.

Format: REVISION:: January 1, 1995; <free-text>

Example: REVISION:: January 1, 1995; FTP information added

WITHDRAW (O) Withdraw means the document is no longer available. Some Institutions choose to delete the record others remove some of the fields. It is up to each institution to decide how to process withdraw records.

A withdraw record has all of the mandatory fields plus the withdraw field and a mandatory revision field. The Withdraw field should indicate the reason for the withdraw in free text.

Example for withdrawing a bibliographic record::

```
BIB-VERSION:: CS-TR-v2.1
ID::         OUKS//CS-TR-91-123
ENTRY::      January 21, 1995
ORGANIZATION:: Oceanview University, Kansas, Computer
               Science
TITLE::      The Computerization of Oceanview with
               High Speed Fiber Optics Communication
REVISION::   January 21, 1995
WITHDRAW::   Withdrawn, found to be irrelevant
END::        OUKS//CS-TR-91-123
```


AUTHOR (O) -- Personal names only. Normal last name first inversion. Editors should be listed here as well, identified with the usual "(ed.)" as shown below in the last example.

If the report was not authored by a person (e.g., it was authored by a committee or a panel) use CORP-AUTHOR (see below) instead of AUTHOR.

Multiple authors are entered by using multiple lines, each in the form of "AUTHOR:: <free-text>".

The system preserves the order of the authors.

Format: AUTHOR:: <free-text>

Example: AUTHOR:: Finnegan, James A.
AUTHOR:: Pooh, Winnie The
AUTHOR:: Lastname, Firstname (ed.)

CORP-AUTHOR (O) -- The corporate author (e.g., a committee or a panel) that authored the report, which may be different from the ORGANIZATION issuing the report.

In entering the corporate name please omit initial "the" or "a". If it is really part of the name, please invert it.

Format: CORP-AUTHOR:: <free-text>

Example: CORP-AUTHOR:: Committee on long-range computing

CONTACT (O) -- The contact for the author(s).
Open-ended, most likely E-mail and postal addresses.

A CONTACT field for each author should be provided, separately, or for all the AUTHOR fields.
E-mail addresses should always be in "pointy brackets" (as in the example below).

Format: CONTACT:: <free-text>

Example: CONTACT:: Prof. J. A. Finnegan, CS Dept,
Oceanview Univ., Oceanview, Kansas, 54321
Tel: 913-456-7890 <Finnegan@cs.ouks.edu>

DATE (O) -- The publication date. The formats are "Month Year" and "Month Day, Year". The month must be alphabetic (spelled out). The "Day" is a 1- or 2-digit number. The "Year" is a 4- digit number.

Format: DATE:: <date>

Example: DATE:: January 1992

Example: DATE:: January 15, 1992

PAGES (O) -- Total number of pages, without being too picky about it. Final numbered page is actually preferred, if it is a reasonable approximation to the total number of pages.

Format: PAGES:: <number>

Example: PAGES:: 48

COPYRIGHT (O) -- Copyright information. Open ended format. The COPYRIGHT field applies to the cited report, rather than to the current bibliographic record.

Format: COPYRIGHT:: <free-text>

Example: COPYRIGHT:: Copyright for the report (c) 1991,
by J. A. Finnegan. All rights
reserved.
Permission is granted for any academic
use of the report.

HANDLE (O) -- Handles are unique permanent identifiers that are used in the Handle Management System to retrieve location data. A handle is a printable string which when given to a handle server returns the location of the data.

Handles are used to identify digital objects stored within a digital library. If the technical report is available in electronic form, the Handle MUST be supplied in the bibliographic record.

Format is "HANDLE:: hdl:<naming authority>/string of characters". The string of characters can be the report number of the technical report as assigned by the publisher. For more information on handles and handle servers see the CNRI WEB page at

<http://www.cnri.reston.va.us>.

**** NOTE: White space in HANDLE due to line wrap is ignored.

Format: HANDLE:: hdl:<naming authority>/string of
characters

Example: HANDLE:: hdl:oceanview.electr/CS-TR-91-123

OTHER_ACCESS (0) -- For URLs, URNs, and other yet to be invented
formatted retrieval systems.

Only one URL or URN per occurrence of the field.

URL and URN information is available in the internet
drafts from the IETF (Internet Engineering Task Force).
The most recent drafts can be found on the CNRI WEB page
at <http://www.cnri.reston.va.us>.

**** NOTE: White space in a URL or URN due to line wrap is ignored.

Format: OTHER_ACCESS:: URL:<URL>
OTHER_ACCESS:: URN:<URN>

Example: OTHER_ACCESS:: URL:<http://elib.stanford.edu/Document/STANFORD.CS:CS-TN-94-1>

Example: OTHER_ACCESS:: URL:<ftp://JUPITER.CS.OUKS.EDU/PUBS/computerization.txt>.

When the URN standard is finalized naming authorities will
be registered and URNs will be viable unique identifiers.
Until then this is a place holder. For the latest URN
drafts see CNRI WEB page at <http://www.cnri.reston.va.us>.

RETRIEVAL (0) -- Open-ended format describing how to get
a copy of the full text. This is an optional, repeatable
field.

No limitations are placed on the dissemination of the
bibliographic records. If there are limitations on the
dissemination of the publication, it should be protected
by some means such as passwords. This format does not
address this protection.

Format: RETRIEVAL:: <free-text>

RETRIEVAL:: for full text with color pictures
send a self-addressed stamped envelope to
Prof. J.A. Finnegan, CS Dept,
Oceanview University, Oceanview, KS 54321

KEYWORD (O) -- Specify any keywords, controlled or uncontrolled.
This is an optional, repeatable field. Multiple keywords
are entered using multiple lines in the form of
"KEYWORD:: <free-text>.

Format: KEYWORD:: <free-text>

Example: KEYWORD:: Scientific Communication
KEYWORD:: Communication Theory

CR-CATEGORY (O) -- Specify the CR-category. The CR-category (the
Computer Reviews Category) index (e.g., "B.3") should
always be included, optionally followed by the name of that
category. If the name is specified it should be fully
specified with parent levels as needed to clarify it, as in
the second example below. Use multiple lines for multiple
categories.

Every year, the January issue of CR has the full list
of these categories, with a detailed discussion of the
CR Classification System, and a full index. Typically the
full index appears in every January issue, and the top two
levels in every issue.

Format: CR-CATEGORY:: <free-text>

Example: CR-CATEGORY:: D.1

Example: CR-CATEGORY:: B.3 Hardware, Memory Structures

PERIOD (O) -- Time period covered (date range). Applicable
primarily to progress reports, etc. Any format is
acceptable, as long as the two dates are separated with
" to " (the word "to" surrounded by spaces) and each date
is in the format allowed for dates, as described above for
the date field.

Format: PERIOD:: <date> to <date>

Example: PERIOD:: January 1990 to March 1990

SERIES (O) -- Series title, including volume number within series.
Open-ended format, with producing institution strongly
encouraged to be internally consistent.

Format: SERIES:: <free-text>

Example: SERIES:: Communication

FUNDING (O) -- The name(s) of the funding organization(s).

Format: FUNDING:: <free-text>

Example: FUNDING:: ARPA

MONITORING (O) -- The name(s) of the monitoring organization(s).

Format: MONITORING:: <free-text>

Example: MONITORING:: ONR

CONTRACT (O) -- The contract number(s).

Format: CONTRACT:: <free-text>

Example: CONTRACT:: MMA-90-23-456

GRANT (O) -- The grant number(s).

Format: GRANT:: <free-text>

Example: GRANT:: NASA-91-2345

LANGUAGE (O) -- The language in which the report is written.
Please use the full English name of that language.

Please include the Abstract in English, if possible.

If the language is not specified, English is assumed.

Format: LANGUAGE:: <free-text>

Example: LANGUAGE:: English

Example: LANGUAGE:: French

NOTES (O) -- Miscellaneous free text.

Format: NOTES:: <free-text>

Example: NOTES:: This report is the full version of the
paper with the same title in IEEE Trans ASSP
Dec 1976

ABSTRACT (O) -- Highly recommended, but not mandatory. Even though no limit is defined for its length, it is suggested not to expect applications to be able to handle more than 10,000 characters.

The ABSTRACT is expected to be used for subject searching since titles are not enough. Even if the report is not in English, an English ABSTRACT is preferable. If no formal abstract appears on document, the producers of the bibliographic records are encouraged to use pieces of the introduction, first paragraph, etc.

Format: ABSTRACT:: xxxx xxxxxxxx
xxxx xxxxxxxx

xxxx xxxxxxxx
xxxx xxxxxxxx

END (M) -- This is a mandatory field. It must be the last entry of a record, identifying the record that it ends, by stating the same ID that was used at the beginning of the records, in its "ID::".

Format: END:: XXX//YYY

Example: END:: OUKS//CS-TR-91-123

>>>>>> [END OF FORMAT DEFINITION] <<<<<<

A Note Regarding the Controlled Symbols of the Publishers

In order to avoid conflicts among the symbols of the publishing organizations (the XXX part of the "ID:: XXX//YYY") it is suggested that the various organizations that publish reports (such as universities, departments, and laboratories) register their <publisher-ID> symbols and names, in a way similar to the registration of other key parameters and names in the Internet.

Rebecca Lasher (RLASHER@Forsythe.stanford.edu), of Stanford working with CNRI has agreed to coordinate this registration with the IANA for the publishers of Computer Science technical reports. It is suggested that before using this format the publishing organizations would coordinate with her (by e-mail) their symbols and the names of their organizations.

In order to help automated handling of the received bibliographic records, it is expected that the producers of bibliographic records will always use the same name, exactly, in the ORGANIZATION field.

Security Considerations

Security issues are not discussed in this memo.

Acknowledgements

This work was supported by the Advanced Research Projects Agency under Grant No. MDA-972-92-J-1029 with the Corporation for National Research Initiatives (CNRI). Its content does not necessarily reflect the position or the policy of the Government or CNRI, and no official endorsement should be inferred.

Authors' Addresses

Rebecca Lasher
Mathematical and Computer Sciences Library
M.S. 2125
Stanford University
Stanford, CA, USA 94305

Phone: +1 415 723 0864
EMail: rlasher@forsythe.stanford.edu

Danny Cohen
Myricom
325 N. Santa Anita Ave.
Arcadia, CA 91006
USA

Phone: +1 818 821 5555
EMail: Cohen@myri.com

