Thomson Reuters – ResearchSoft

"RIS" Format Documentation

Adding a "Direct Export" Button to Your Web Page or Web Application

This document is addressed to information providers who deliver bibliographic information to customers on the web. Examples of such providers include organizations such as the National Library of Medicine (NLM®) and commercial vendors like ISI®, Ovid®, and OCLC®. Typically, providers maintain websites that allow a user to search a database of bibliographic references and obtain lists of matching references that they can view or save in varying formats.

Users of these websites who also use bibliographic software such as EndNote® or Reference Manager® are accustomed to saving these lists as text files and then importing them back into their bibliographic software program. This process involves several steps: choosing the correct listing format, saving the listing, selecting an appropriate import option, and importing that into the bibliographic program.

An "Export to Bibliographic Software" button makes the process one step. From the user's point of view, once a search has been completed and a set of references marked, they simply click the "Export" button and their bibliographic program pops up and imports the references directly. It makes it look like the web page of the information provider and the bibliographic program are integrated.

In addition to these instructions, our technical support team is available to answer any questions about providing direct export functionality on a web page. In addition, we provide trial versions of our desktop products which may be useful in creating and testing direct export options.

EndNote® Trial - http://www.endnote.com/endemo.asp

Reference Manager® Trial - http://www.refman.com/rmdemo.asp

More information about our products and additional technical support are available on our web site:
http://scientific.thomsonreuters.com/rs/

THOMSON REUTERS

HOW DOES AN EXPORT BUTTON WORK?

SPECIFYING THE CONTENT TYPE

Most information providers are using a cgi or similar program to feed responses to searches through their web server software. The cgi simply sends http responses to the web server. The first part of this http response is the response header, and one of the parameters in that header is the Content type. For a regular web page, the Content type is "text/html". If, instead of "text/html", the cgi program lists something else, the user's web browser will react to the response by running whatever helper application has been assigned to that Content type.

The information provider can add an "Export to bibliographic software" button by programming the cgi, in response to a click on that button, to produce the desired output file, but rather than sending it as another web page (Content type text/html), it sends the file, with MIME type **"application/x-research-info-systems "** in the response header.

THE "APPLICATION/X-RESEARCH-INFO-SYSTEMS" MIME

The "application/x-research-info-systems" MIME or media type does not define a specific file format. Rather, it specifies a method of including metadata about the included reference data so that a desktop bibliography program can correctly interpret the data. The user's browser, when set up correctly, will be able to recognize the MIME and handle it appropriately. After the cgi is set up to send the "application/x-research-info-systems" MIME in the http response, the helper application will look for the corresponding metadata. The metadata consist of a simple document header. Two line breaks (cr/lf) separate the document header from the actual reference data.

Note that "header" in the following discussion refers to text within the file, not the http response header generated by the web server (although both the document header and the http header use the format described in RFC 822). Once the cgi is programmed to include the "application/x-research-info-systems" MIME in its header, the following information should be included in the body of the document.

DOCUMENT HEADER INFORMATION

After the http response header, signified by the two line breaks, the document header information should be included. The header for a file consists of these entries:

Provider: The name of the information provider (required)

Database: The name of the database (optional)

Tagformat: Name of the tag format used identify fields (optional)

Content: media type for the body of the file (required)

In practice, the "Provider:" and "Database:" fields will be used by the desktop bibliography software to identify the correct import option to process the data within the body of the file. Even if you are using the RIS format, you may still want to include this information.

THE "PROVIDER:" FIELD

The provider field contains the name for the organization responsible for generating the file. This does not need to be a standardized name, but should be recognizable and the same for all databases maintained by the provider. The provider name can be a legal name or a common name, for example:

Provider: Ovid Technologies, Inc. ®
Provider: National Agricultural Library

THE "DATABASE:" FIELD

The database field contains the name or names of the databases from which the references within the body of the file were generated. Some providers maintain only one database; if so, the "Database:" field may be omitted from the header. If the provider maintains multiple databases, but all references in the body of the file were drawn from a single database, the name of the database should be included. For example:

Provider: National Library of Medicine
Database: PubMed®

Commercial information providers often maintain several databases, and allow searches across databases. Information from different databases, even when presented in a common output format, may be different enough that the desktop bibliography software may not be able to parse the output correctly. This specification makes no provision for output from multiple databases. Providers who wish to implement direct output from multiple databases should contact Thomson Reuters ResearchSoftware® for additional information. It may also be possible for the provider to parse information from multiple databases all into RIS format for direct export. In that case the bibliographic software would recognize the Thomson Reuters ResearchSoftware® tagged output format defined later in this document.

THE "TAGFORMAT:" FIELD

Most database providers support several output formats that use different tags to identify fields within references. This field allows those providers to specify which tag format is used in the file, allowing the import filter to parse the data

accordingly. Tagformat is optional, and is unnecessary for databases providing content in only one format. If several tag formats are possible it may be desirable to identify the one used in the reference data. For example:

Provider: Ovid Technologies, Inc. [®]
Database: INSPEC [®]
Tagformat: reprint
Content:text/plain; charset="us-ascii"

See: http://www.iana.org/assignments/media-types/index.html for information about media types

THE "CONTENT:" FIELD

This is where information about the type of character encoding is included. EndNote may not necessarily be able to detect encoding automatically, but this information should be included if possible.

http://www.iana.org/assignments/character-sets for information about character set specifiers.

CHOOSING AN OUTPUT FORMAT

An information provider wishing to implement an "Export to bibliographic software" feature for its website can use its own tagged output format. You should inform Thomson Reuters ResearchSoftware[®] as to the output format, provider and database name(s) you intend to use. If Thomson Reuters ResearchSoftware[®] has already developed import filters for your import format, we may suggest provider and database names that correspond to those filters. If new filters need to be created, we will need samples of your format. If, however, an information provider does not have its own tagged output format, we recommend the following format for seamless integration.

Last Update: May, 2009

"RIS" TAGGED OUTPUT FORMAT

Commonly referred to as "RIS" format, this tagging system can be read by any Thomson Reuters ResearchSoftware® Product. Each reference is composed of a variable number of fields; and each field is preceded by a six-character label or "tag". Some tags are specific only to certain reference types.

Each six-character tag must be in the following format:

<upper-case letter><upper-case letter or number><space><space><dash><space>

Character 1	Uppercase alphabetic character, specific to each field type. (These are described in the following sections.)
Character 2	Uppercase alphabetic or numeric character, specific to each field type. (These are described in the following sections.)
Characters 3 and 4	Spaces (ANSI 32)
Character 5	Dash (ANSI 45)
Character 6	Space (ANSI 32)

USE SEPARATE LINES FOR EACH TAG

Each tag and its contents must be on a separate line, preceded by a "carriage return/line feed" (ANSI 13 10).

TAG ORDER

Except for the first tag of each reference, which must be "TY   - " and the last tag of each reference, which must be "ER   - ," the tags within each reference can be in any order.

CHARACTERS ALLOWED IN FIELDS

The characters allowed in the reference ID fields can be in the set "0" through "9," or "A" through "Z." The characters allowed in all other fields can be in the set from "space" (character 32) to character 255 in the ANSI Character Set. Note, however, that the asterisk (character 42) is not allowed in the author, keywords or periodical name fields.

HOW TO HANDLE LONG FIELDS

If the information following any one tag is more than 70 characters long, it is allowable (though not necessary) to insert a carriage return/line feed at the end of 70 characters and continue on the next line.

REPEATING FIELDS

Most tagged fields may occur once per reference using the syntax below. However, the author fields and the keyword field may be repeated so that they can be delimited correctly by the program. Each author should be tagged separately as should each keyword. Each reference can include an unlimited number of authors and keywords.

TAG AND FIELD DEFINITIONS

This section details the syntax and definitions required for all possible tags and reference types in the RIS format. Note that the required tags are "TY" which must be the first tag in the reference, and "ER" which must be the last tag in the reference. Other tags can be in any order. These are described below (Note: quotation marks should not be included in the actual text file).

FIELD TYPES WITH SPECIAL SYNTAX

TY····	Type of reference. This must contain the abbreviation for the reference type as found in the next section. This will determine how all other fields are interpreted.
ER····	End of reference. Must be the last tag in a reference.
AU···· A2···· A3···· A4····	Authors, Editors, Translators. Each author must be on a separate line, preceded by the tag that corresponds to the author role (see individual ref type matrix for role definitions). Each reference can contain unlimited author fields, and can contain up to 255 characters for each field. The author name must be in the following syntax: Lastname, Firstname, Suffix For Firstname, you can use full names, initials, or both. The format for the author's first name is as follows: Phillips, A.J Phillips, Albert John Phillips, Albert Lastname = Any string of letters, spaces, and hyphens Firstname = Any string of letters, spaces, and hyphens Initial = Any single letter followed by a period Full Name = Any string of letters, spaces, and hyphens Suffix = Jr/Sr/II/III/MD etc. (Phillips,A.J.,Sr.); use of the suffix is optional
PY····	This is the publication year. It must be four numberic

	characters. Dates prior to 1000 should use "0" in the positions before the intended date. For example, the year 765 would be represented as 0765.
DA····	Dates must be in the following format: YYYY/MM/DD/other info The year, month and day fields are all numeric. The other info field can be any string of letters, spaces and hyphens. Note that each specific date information is optional, however the slashes ("/") are not. For example, if you just had the \<year\> and \<other info\>, then the output would look like: "1993///Spring." Date information should be set forth in the following format: YYYY or YYYY/MM or YYYY/MM/DD or YYYY/MM/DD/other info
KW····	Keywords. Each keyword or phrase must be on its own line, preceded by this tag. A keyword can consist of multiple words (phrases) and can be up to 255 characters long. There can be unlimited keywords in a reference.
RP····	Reprint status. This optional field can contain one of three status notes. Each must be in uppercase, and the date after "ON REQUEST" must be in USA format, in parentheses (MM/DD/YY). If this field is blank the Import function assumes the reprint status is "NOT IN FILE." The three options are: IN FILE – The data provider has a corresponding physical copy for the reference. NOT IN FILE – The data provider does not have a corresponding physical copy. ON REQUEST (mm/dd/yy) - This means that the data provider has requested a reprint of the reference; the date is the date on which the reprint was requested (in mm/dd/yy format).
J2····	Periodical name: standard abbreviation. This is the abbreviation of the periodical in which the article is published. If possible, periodical names should be abbreviated in the Index Medicus® style, with periods after the abbreviations. This field is mapped to the full journal name in T2 and is used as the journal abbreviation in output

	styles.
UR····	Web/URL. There is no practical length limit to this field. URL addresses can be entered individually, one per tag, or multiple addresses can be entered on one line using a semi-colon as a separator.

REFERENCE TYPES

The following describes the valid reference type field names that can be used with the TY tag to indicate the type of reference. If a reference type is not recognized on import it will be labeled as Generic.

FIELD NAME	REFERENCE TYPE
GEN	Generic
ABST	Abstract
AGGR	Aggregated Database
ANCIENT	Ancient Text
ART	Artwork
ADVS	Audiovisual Material
BILL	Bill
BLOG	Blog
BOOK	Book
CHAP	Book Section
CASE	Case
CTLG	Catalog
CHART	Chart
CLSWK	Classical Work
COMP	Computer Program
CPAPER	Conference Paper
CONF	Conference Proceeding
DATA	Dataset
DICT	Dictionary
EDBOOK	Edited Book
EBOOK	Electronic Book
ECHAP	Electronic Book Section
EJOUR	Electronic Article
ENCYC	Encyclopedia
EQUA	Equation
FIGURE	Figure
MPCT	Film or Broadcast
JFULL	Full Journal
GOVDOC	Government Document
GRNT	Grant

HEAR	Hearing
INPR	In Press Article
ICOMM	Internet Communication
JOUR	Journal Article
LEGAL	Legal Rule
MGZN	Magazine Article
MANSCPT	Manuscript
MAP	Map
MUSIC	Music
NEWS	Newspaper Article
DBASE	Online Database
MULTI	Online Multimedia
PAMP	Pamphlet
PAT	Patent
PCOMM	Personal Communication
RPRT	Report
SER	Serial
SLIDE	Slide
SOUND	Sound Recording
STAND	Standard
STAT	Statute
THES	Thesis
UNBILL	Unenacted Bill
UNPD	Unpublished Work
VIDEO	Video Recording
ELEC	Web Page

FIELD DEFINITIONS

Here you'll find the tags and definitions grouped by worksheet for each unique reference type. (NOTE: This is an embedded speadsheet)

TY -	`GEN`
A2 -	Secondary Author
A3 -	Tertiary Author
A4 -	Subsidiary Author
AB -	Abstract
AD -	Author Address
AN -	Accession Number
AU -	Author
C1 -	Custom 1
C2 -	Custom 2
C3 -	Custom 3
C4 -	Custom 4
C5 -	Custom 5
C6 -	Custom 6
C7 -	Custom 7
C8 -	Custom 8
CA -	Caption
CN -	Call Number
CY -	Place Published
DA -	Date
DB -	Name of Database
DO -	DOI
DP -	Database Provider
ET -	Edition
J2 -	Alternate Title
KW -	Keywords
L1 -	File Attachments
L4 -	Figure
LA -	Language
LB -	Label
IS -	Number
M3 -	Type of Work
N1 -	Notes
NV -	Number of Volumes
OP -	Original Publication
PB -	Publisher
PY -	Year

SAMPLE REFERENCES

```
TY  - JOUR
AU  - Baldwin,S.A.
AU  - Fugaccia,I.
AU  - Brown,D.R.
AU  - Brown,L.V.
AU  - Scheff,S.W.
TI  - Blood-brain barrier breach following cortical contusion in the rat
T2  - Journal of Neurosurgery
PY  - 1996
VL  - 85
IS  - 4
SP  - 476-481
SN  - 0022-3085
AB  - Adult Fisher 344 rats were subjected to a unilateral impact to the dorsal cortex above the hippocampus at 3.5
      m/sec with a 2 mm cortical depression. This caused severe cortical damage and neuronal loss in
      hippocampus subfields CAU, CA3 and hilus.
KW  - cortical contusion
KW  - blood-brain barrier
KW  - horseradish peroxidase
KW  - hippocampus
KW  - rat
DO  - DOI:10xxxxxxxx
ER  -


TY  - BOOK
AU  - Behrens, J.
TI  - History of the CDC
PY  - 1999
CY  - Chicago
PB  - Parity Press
SP  - 144
VL  - 2nd
KW  - Epidemiology
KW  - U.S. Gov't
DO  - DOI: 10.xxxxxxxxx
ER  -


TY  - CHAP
AU  - Franks,L.M.
TI  - Preface by an AIDS Victim
PY  - 1991
VL  - 3
T2  - Cancer, HIV and AIDS
T3  - Cancer Studies Series
A2  - Jackson, H.
A2  - Weisl, R.
SP  - vii- viii
CY  - Berkeley, CA
PB  - Berkeley Press
KW  - HIV
```

```
KW  - AIDS
DO  - DOI: 10.xxxxxxxxx
SN  - 0-679-40110-5
ER  -
```

FOR MORE INFORMATION:

For more information contact Thomson Reuters Scientific®:

SALES

2141 Palomar Airport Road, Suite 350 Carlsbad, CA 92011
Phone: 760.438.5526 | 800.722.1227

CUSTOMER TECHNICAL SUPPORT

425 Market Street, 6th floor San Francisco, CA 94105
Phone: 1 800 336 4474 press 4 at the prompt
http://scientific.thomsonreuters.com/techsupport/

Last Update: May, 2009