

ISI (*.isi)

What is it?

The ISI format is an output format for the Thompson Scientific/ISI publication database, most notably the Web of Science database. For Web of Science, ISI files contain author and citation information, as well as full abstracts. ISI files are generated as the results of search queries on ISI databases.

How do I get data in this format?

In order to access ISI databases, you or your institution must have obtained a license. Once a license is obtained, you may download ISI Web of Science query results with the following steps...

1. Go to <http://www.isiknowledge.com>.
2. Click the "Web of Science" tab. (Only web of science contains citation information).
3. Enter your search query in the provided fields.
4. Click "Search".
5. Select the papers you want to download, either by manually checking them, or selecting a range of papers, using the options in the "Step 1" box in the bottom left-hand corner of the page.
6. Select how much information for each paper you want to download (do you want abstracts? citations?) in the "Step 2" box.
7. Select "Save to Other Reference Software" in the drop-down list in the "Step 3" box.
8. Click Save in the "Step 3" box.
9. Wait while ISI processes your request.
10. Download the file (a download dialog box should have appeared).
11. Rename the file to have the extension ".isi".

From here, you can load the ISI file as normal in [Sci2](#) or [NWB](#), and perform the operations noted in the section below.

How is it used in Network Workbench?

Since ISI files contain both citation and author information, you can extract co-citation and co-authorship networks using the [Extract Co-Occurrence Network From Table](#) algorithm. You may also perform [Burst Detection](#).

NWB also supports combining the results of multiple ISI queries, by copying and pasting multiple files into a single file. When loaded a file which is created in this manner, either load it using the "Load and Clean ISI File" option, or load it normally and run the [ISI Duplicate Remover](#) Algorithm.

What should I know about how NWB handles this format?

Network workbench performs the following special processing on the raw ISI data: Reads data into a table

Normalizes author names to be separated by "|"

Attempts to normalize journal name column so that they match the journal names used in the reference column (uses a heuristic algorithm)

Adds a self-reference column (a column containing info on how the paper should be referenced)

Are there different versions of the ISI file format?

Yes. Different sources will give you ISI files containing different types of data. For instance, on the Web of Knowledge web site, you can choose whether to include citation information when you download a file.

Files downloaded at different times may also contain different information, due to ISI/Web of Knowledge making changes to their file format. For example, files downloaded after September, 2011 may include information about which author is reachable at which address, while files downloaded before then are not as specific. Sci2 does not yet have the capability to interpret this new information, but we hope to.

Why can't I load newly-downloaded ISI files?

Web of Science made a change to their output format in September, 2011. Older versions of Sci2 tool may refuse to load these new files, with an error like "Invalid ISI format file selected."

Workaround

Sci2 solution

If you are using older Sci2 tool, you can download the [WOS-plugins.zip](#) file and uncompress the JAR files into your sci2/plugins/ directory. Restart Sci2 to activate the fixes. You can now load the downloaded ISI files into the Sci2 without any additional step.

NWB and Sci2 solution

If you are using the old Sci2 tool or the NWB tool, you will need to follow the guidelines below before you can load the new WOS format file into the tool.

You can fix this problem for individual files by opening them in Notepad (or your favorite text editor). The file will start with the words:

```
FN Thomson Reuters Web of Knowledge
```

Just add the word ISI:

And then Save the file. Now it should load.

Fix

The fix to allow these new files to load is already in the [nightly builds](#) of NWB and Sci2. As of September 16, 2011, we expect to release a version that can read these files, very soon.

What should I know about the format itself?

The subsequent text was copied from <http://wos.isitrial.com/help/helpprn.html>

ISI's export file format is illustrated in the following sample record. You will see bibliographic information such as this for each record on your marked list. The [two-character tags](#) (also see field tags at [here](#)) identify each data element in the record. Records are separated by an ER (end of record) tag.

Each export file begins with two lines that identify the file type (FN) and version number (VR) of the export format.

Example File

```

FN      ISI Generic Export Format
VR      1.0
PT      J
AU      Foster, JD
        Hunter, N
        Williams, A
        Mylne, MJA
        McKelvey, WAC
        Hope, J
        Fraser, H
        Bostock, C
TI      Observations on the transmission of scrapie in experiments          using embryo transfer
SO      VETERINARY RECORD
LA      English
DT      Article
NR      15
SN      0042-4900
PU      BRITISH VETERINARY ASSOC
CI      INST ANIM HLTH, OGSTON BLDG, W MAINS RD, EDINBURGH EH9 3JF,
        MIDLOTHIAN, SCOTLAND.MRC, NEUROPATHOGENESIS UNIT,
        EDINBURGH EH9 3JF, MIDLOTHIAN, SCOTLAND.SCOTTISH AGR COLL,
        VET SERV DIV, PENICUIK, MIDLOTHIAN, SCOTLAND.INST ANIM HLTH,
        COMPTON LAB, READING RG16 0NN, BERKS, ENGLAND.
ID      FIBRIL PROTEIN PRP; SIP GENE; SHEEP
AB      This investigation studied the maternal transmission of scrapie in sheep ...
CR      DICKINSON AG, 1974, V84, P19, J COMP PATHOL
        DICKINSON AG, 1988, P63, NOVEL INFECTIOUS AGE
        DICKINSON AG, 1976, P209, SLOW VIRUS DISEASES FOSTER JD, 1993, P229, ...
TC      0
BP      559
EP      562
PG      6
PY      1996
PD      JUN 8
VL      138
IS      23
GA      UR372
PI      LONDON
RP      Foster JD
ER

```

Why are some values for records in the ISI file appearing as -1?

When a file is loaded into Sci2 the values will be set to -1 for records anytime the values have not been provided by the original ISI file. For example, in the FourNetSciResearchers.isi file in the Sci2 sample data, the "Total Times Cited" column does not have any values associated with it and has been set to -1 for all records.

Relevant Links

- [ISI Home Page](#)