

3.8 Header Map Files

When dealing with various data formats in Sci2, column header names sometimes need to be standardized to a generic Sci2 format before algorithms can be run on the data. This process is called header standardization, and can be achieved using the [Convert to Generic Publication Algorithm](#). This algorithm uses a specially formatted file, ending with the **.hmap** extension, to replace preexisting headers with the generic Sci2 equivalents.

Syntax

The syntax of a **.hmap** file is identical to that of a Java **.properties** file. Line comments can be included, and should start with a **#**. The purpose of the file is to map out preexisting headers with their replacement headers. This is done using a key-value format, where the old headers are the keys, and the new generic headers are the values. Each line should contain a key, followed by a **=**, then a value. It is important to note that any spaces in a key (old header) should be represented by the Unicode space value - **\u0020**. Normal spaces can be used in the value (replacement headers), however. The reason for this is that, when the **.hmap** file is read in by Java (using **.properties** syntax interpretation), it will mistake any spaces before the equal sign as a key-value separator. Therefore, Unicode space values must be used to escape any spaces before the equal sign.

Example

It may be hard to visualize what a **.hmap** file should look like without an example, so one is provided below. This is a sample **.hmap** file mapping SCOPUS format headers to their generic Sci2 equivalents:

scopus.hmap

```
# Keys (to be replaced) are on the left side of the equal sign.
# Values (used as replacement) are on the right side of the equal sign.
Authors=Authors
Title=Title
Year=Publication Year
Source\u0020title=Journal Title (Full)
Volume=Volume
Issue=Issue
Art.\u0020No.=New Article Number
Page\u0020start=Beginning Page
Page\u0020end=Ending Page
Abstract=Abstract
Author\u0020Keywords=Original Keywords
Index\u0020Keywords=New ISI Keywords
References=Cited References
Editors=Editors
Publisher=Publisher
Conference\u0020name=Conference Title
Conference\u0020date=Conference Dates
Conference\u0020location=Conference Location
ISSN=ISSN
ISBN=ISBN
DOI=DOI
Language\u0020of\u0020Original\u0020Document=Language
Document\u0020Type=Document Type
Self\u0020Reference=Cite Me As
```