

# NSF format (.nsf)

## *What is it?*

The NSF file format contains results from the NSF (National Science Foundation) Award Database, which contains information on grants given out by the NSF, including how much money was awarded, and who are the beneficiaries of the grants.

## *How do I get data in this format?*

To get NSF award data, perform a query on [The NSF Award Database](#), download the CSV format by clicking the csv icon in the "Export Options" box at the bottom of the screen, and rename the extension to ".nsf". (Due to unusual issues with the NSF csv format, it is not possible to read in the raw .csv file.)

## *How is it used in Network Workbench?*

Basic statistics as well as the evolving co-PI networks as well as the geospatial coverage of certain research topics can be determined. To extract co-authorship (really co-investigatorship), choose the "All Investigators" column when utilizing the [Extract Co-Occurrence Network from Table](#) algorithm.

## *What should I know about how NWB handles this format?*

When an NSF csv file is read, the following actions are performed:

```
Convert NSF file into a table
Marks NSF file as "pre-2014" or "current" depending on format of headers
Handle duplicate columns
Replace '\"' with '\"' (more standard method of escaping quotes in csv
files)
Normalize CO-PI column (handle excessive spaces in names)
If it is a "current" format NSF:
    Normalize dollar amounts (go from strings to floats)
Else if it is a "pre-2014" format NSF:
    Normalize PI names (put it in "first name first" format with no
comma
Create "All Investigators" Column (merge CO-PI and Primary PI column)
```

## *What should I know about the format itself?*

"NSF" files are really comma-separated value files, but because they are horribly formatted, require special processing by Network Workbench. Thus, in order to get Network Workbench to treat them correctly, the extension must be changed to ".nsf".

## *What is the difference between the "current" and "pre-2014" NSF formats?*

A full listing of changes may be found on the [JIRA issue page](#) for this update. The most important thing for users to keep in mind is that **sets of data are now delimited by commas instead of pipes** in the current version. Also, when using aggregate function files, you will want to use specially labelled "pre-2014" properties files when working with those older data sets.

## *Links*

- [NSF Home Page](#)