

# Load NSF File Into Database

## Menu Path

File > Load... > Choose "NSF database"

## Description

Accepts NSF file as the input and produces an NSF database as the output. This is the only provided solution for loading an NSF file into a database.

As with our other database loaders, loading happens in two phases. In the first phase, the data is loaded from disk and processed so it can then be inserted into a database. In the second phase, a new database is created and the loaded data is inserted into it.

The progress Phase 1 is tracked in the "% Complete" column in the Scheduler. When it finishes, the "% Complete" is reset for Phase 2, which also tracks its progress there.

Upon the beginning of the first phase, the message "Beginning Phase 1 of loading the NSF file into a database." is printed to the Console. Upon the beginning of the second phase, the message "Beginning Phase 2 of loading the NSF file into a database." is printed to the Console.

## Usage Hints

Load a NSF file, then run this algorithm on it. This must be done as the first (set of) step(s) in the database pipeline. Currently we support both the "Comma separated" (CSV) & "Tab separated" (TSV) formats of NSF files. We do not have support for "Excel" format, yet.

**Note:** NSF format has pre-defined set of fields which are found in all the NSF files. They are as follows:

```
Title, Principal Investigator, Program(s), Organization Phone, State, Organization Zip, Program Element Code (s), NSF Directorate, NSF Organization, Field Of Application(s), Organization Street Address, Program Manager, Organization State, Expiration Date, Program Reference Code(s), PI Email Address, Organization, Award Instrument, Awarded Amount to Date, Last Amendment Date, Co-PI Name(s), Award Number, Organization City, Start Date, Abstract.
```

The database schema we use in this NSF loader and our other NSF database-related algorithms takes into account only these fields. Although any "arbitrary" fields found will merely be appended to the **AWARD** table and will not be reflected in the schema structure on a high level. Also many times due to CSV corruption we merge all the columns right to the abstract column that leads to "Abstract" being broken into multiple columns.

## Implementation Details

This loader first converts the NSF data into a table. It then transforms that table to a database. No data is lost or changed during either of these steps, but some linkages/metadata are created where possible.

The following is a list of NSF database tables that link to pages describing their fields and how they are parsed out of NSF datasets:

- [AWARD](#)
- [FIELD\\_OF\\_APPLICATION](#)
- [NSF\\_FILE](#)
- [ORGANIZATION](#)
- [PERSON](#)
- [INVESTIGATORS](#)
- [PROGRAM MANAGERS](#)
- [PROGRAM](#)
- [AWARD\\_OCCURRENCES](#)
- [AWARD\\_FIELD\\_OF\\_APPLICATIONS](#)
- [INVESTIGATOR\\_ORGANIZATIONS](#)
- [PROGRAM\\_NAME\\_AND\\_ELEMENT\\_CODES](#)
- [PROGRAM\\_REFERENCE\\_CODES](#)