10 Sci2 FAQ

.

Installing Sci2

Windows

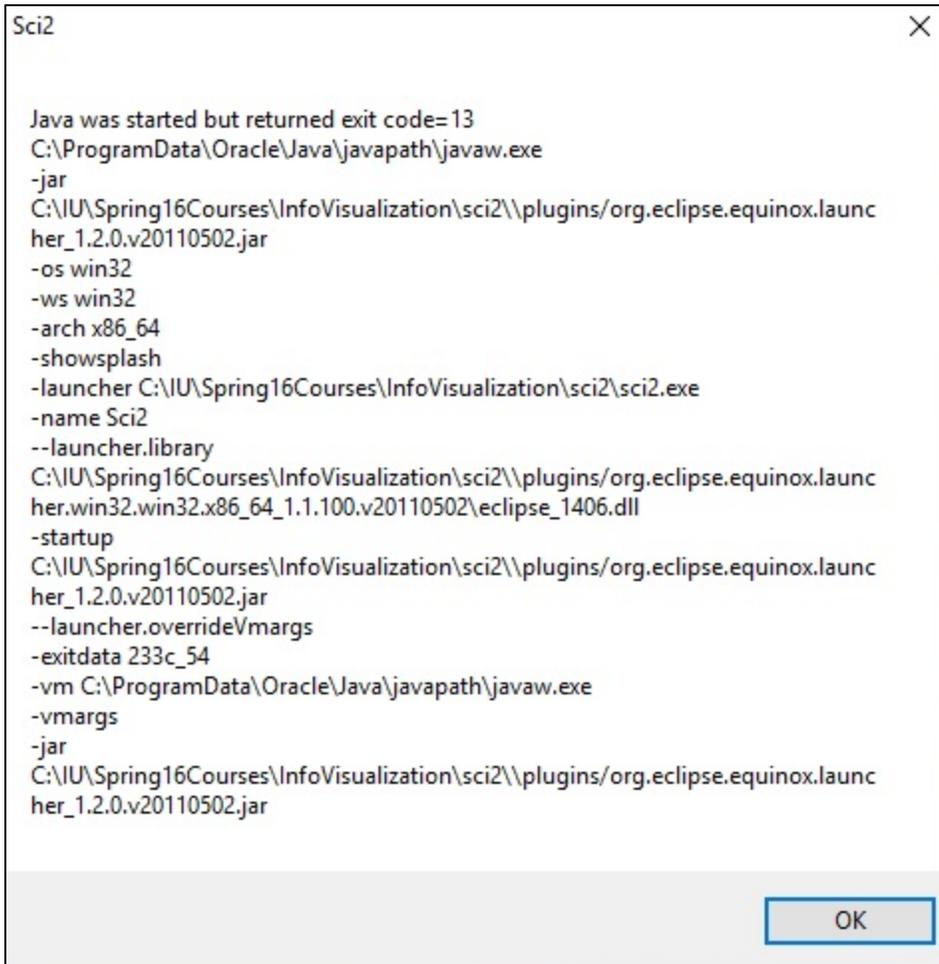**Q: I am having trouble installing and running Sci2 on a Windows PC. What should I do?**

A: One common problem when initially installing Sci2 is not having a current version of Java. Sci2 runs on Java 1.6 (32 bit) or newer. If you are having trouble installing Sci2, you may want to install the latest 32 bit version of Java. Make sure you install the 32 bit version, as Sci2 will only run with a 32 bit version. You can run 32 bit and 64 bit versions of Java simultaneously. If you have the latest version of Java installed, but is not being detected by Sci2, you must specify the correct path to be used by the tool to locate the correct version of Java. You will want to target the javaw.exe file – which is likely located in the C:\Program Files (x86)\Java\jre7\bin directory, assuming you have installed Java in the default place.

 Now, you will have to tell Sci2 to look in this directory. In your Sci2 installation, open sci2.ini (with Notepad or any other plain text editor), delete the contents, and replace them with this:

```
-vm
C:\Program Files (x86)\Java\jre7\bin\javaw.exe
-vmargs
-Xms15m
-Xmx350m
```

If you're installation of Java is located in another directory on your system, you'll need to replace the path to javaw.exe with the one you find on your system. Save the file in the Sci2 installation and you should be able to start the tool.

Q: I am having trouble running Sci2 on a Windows, I received an error message that starts with, "Java was started by returned exit code=13". What should I do?

```
Sci2                                                          ✕

Java was started but returned exit code=13
C:\ProgramData\Oracle\Java\javapath\javaw.exe
-jar
C:\IU\Spring16Courses\InfoVisualization\sci2\\plugins/org.eclipse.equinox.launc
her_1.2.0.v20110502.jar
-os win32
-ws win32
-arch x86_64
-showsplash
-launcher C:\IU\Spring16Courses\InfoVisualization\sci2\sci2.exe
-name Sci2
--launcher.library
C:\IU\Spring16Courses\InfoVisualization\sci2\\plugins/org.eclipse.equinox.launc
her.win32.win32.x86_64_1.1.100.v20110502\eclipse_1406.dll
-startup
C:\IU\Spring16Courses\InfoVisualization\sci2\\plugins/org.eclipse.equinox.launc
her_1.2.0.v20110502.jar
--launcher.overrideVmargs
-exitdata 233c_54
-vm C:\ProgramData\Oracle\Java\javapath\javaw.exe
-vmargs
-jar
C:\IU\Spring16Courses\InfoVisualization\sci2\\plugins/org.eclipse.equinox.launc
her_1.2.0.v20110502.jar

                                                    OK
```

A: The error message that you have received is caused by trying to run Sci2 with the incorrect version of Java required by the tool. For example, if you download the 64-bit version of Sci2, but only have a copy of 32-bit Java installed on your computer, Sci2 will try to run but it will produce this error. To solve this problem, you will need to verify the version of Java installed on your computer.

If Sci2 fails to run, and you find that Java is installed in the *Program Files (86x)* directory, then you have 32-bit Java installed and need to install 64-bit Java. Likewise, if Java is installed in the Program Files directory, hen you have 64-bit Java installed and need to install 32-bit Java.You will need to either try downloading the appropriate 32-bit or 64-bit version of Sci2, or install the appropriate version of Java and make sure that Sci2's instillation file is updated with the correct path to the new version of Java,

Macs

Q: I am having trouble installing and running Sci2 on a Mac OS X Mavericks or later. What should I do?
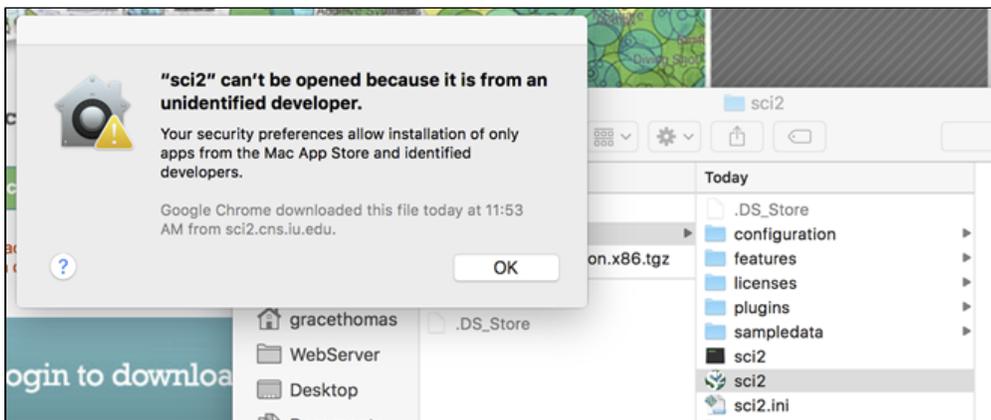
A: Recent updates to Apple's Mac OS X Mavericks have included security that prevent Sci2 from running on a machine if the program is not available from the Mac App store or an identified developer. To get Sci2 to run on your machine, you will have to change the system preferences. Instructions about how to update the Mac OS X's Gatekeeper are available at the Apple website (https://support.apple.com/en-us/HT202491).

Q: I am having trouble installing and running Sci2 on a Mac OS X Sierra. What should I do?

A: Sci2 users trying to install the tool on Mac OS X Sierra will first want to install the legacy edition of Java 6 for OS X 2015-001 (https://support.apple.com/kb/dl1572?locale=en_US).

After installation of the appropriate version of Java, You will need to update the Sci2.ini file

```
-startup
plugins/org.eclipse.equinox.launcher_1.2.0.v20110502.jar
--launcher.library
plugins/org.eclipse.equinox.launcher.carbon.macosx_1.1.100.v20110502
-vmargs
-Xms256m
-Xmx1024m
```

After making the modification to the Sci2.ini file, to run the tool, you will need to use the unix-executable file to launch Sci2. The Sci2 Unix Executable file is found by selecting Control on the keyboard and then click Sci2 icon, which will show a drop down menu. In the menu select "Show Package Contents". Next, in the new directory, navigate to Contents>MacOS and select the file labeled "Sci2." Selecting this file will launch a terminal window and a separate tool window.
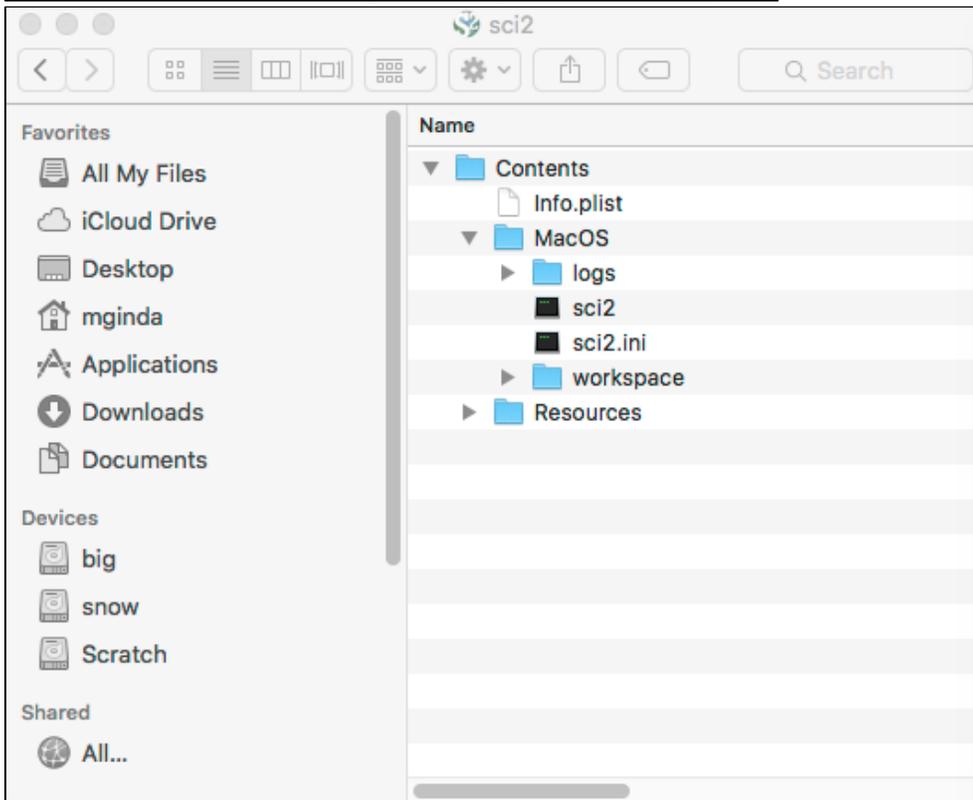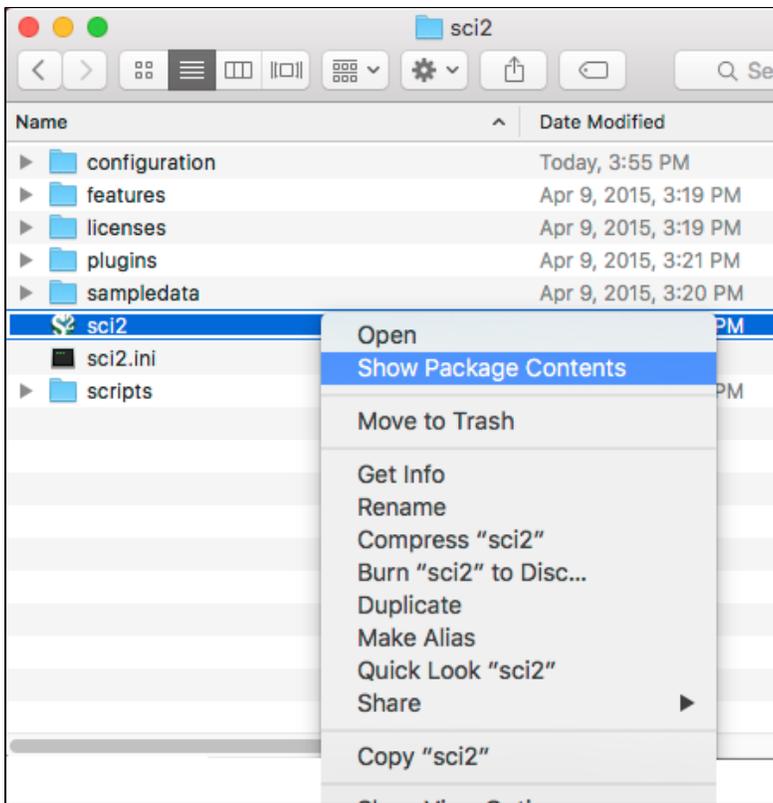
Sci2 Locale Issues

Q: I am having trouble loading data into Sci2 ***and having trouble converting PostScript files to PDF files.***

A: If you are outside of the United States, problems loading data, saving files from the data manager, converting files types, and running the Temporal Bar Graph algorithm, due to locale issues. If you are working on a Windows machine (for Mac and Linux refer to the Changing System Locale guide) and you want to change the locale, go to the Start Menu and then select Control Panel and then select Region and Language, then change the format to English (United States):
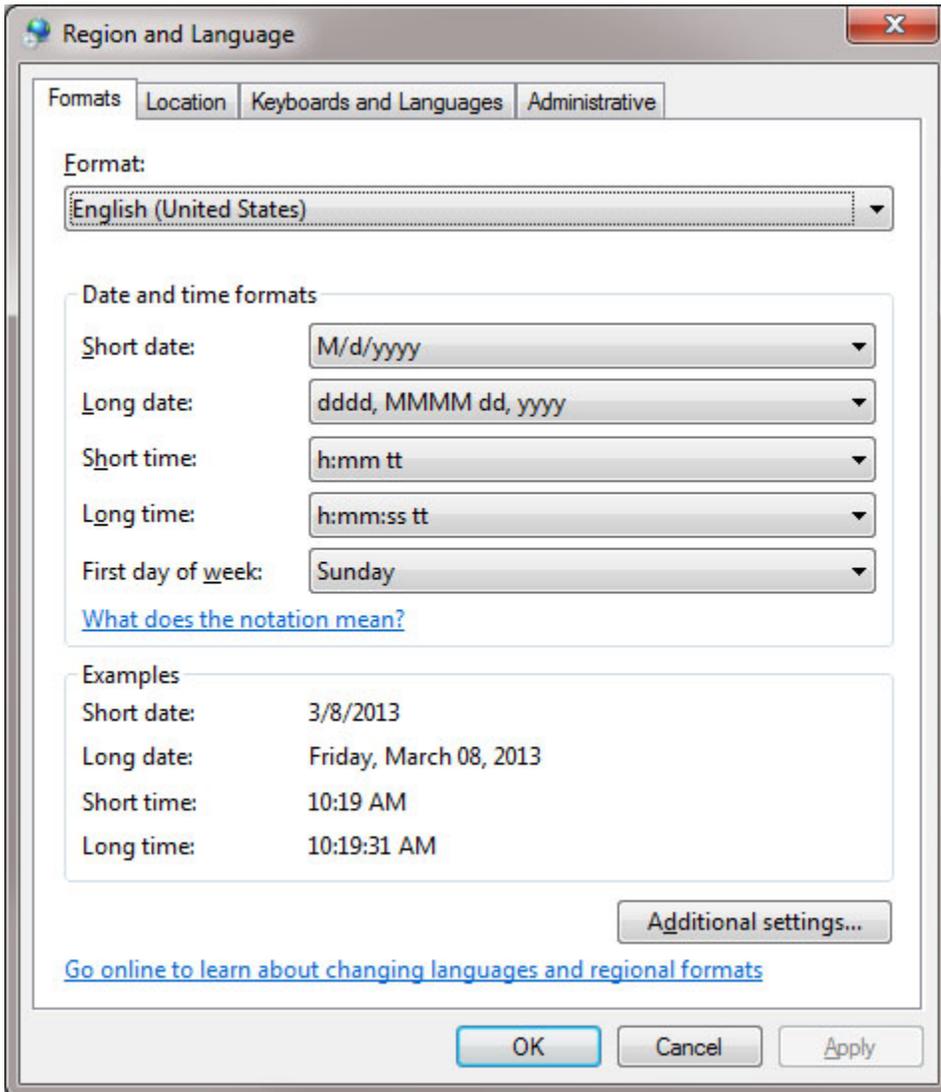
If you are working on a Mac, open System Preferences and choose Language and Text. Then go to the Formats tab and set the Region to the United States.

Loading Data

General Data Loading

Q: Does it matter that my data file has the same file format extension used in Sci2 documentation or may I just use the .txt and .csv extensions?

Sci2 documentation often references data formats through the use of file extensions (.isi, .nwb, .mat, .enw), which indicate the type of data being loaded into Sci2 for parsing. Yet often data are exported by other tools and and databases as files with the file extension *.txt*, which can indicate one of two possibilities: that the data are formatted as a tab separated columns; or that the data is for written in a machine readable format. For data saved as a Tab Separated format, you will need to convert these tables into CSV formats before you load these files into Sci2.

However, if the data is written in a Machine readable format, such as Web of Science (*.isi*) files and Endnote Export format file (*.enw*), Sci2 will parse data loaded with the *.txt* extension, provided that you know the correct data formatting used in the file. For example, when you export data

from EndNote Web or Endnote desktop as the Endnote Export format, files are given the *.txt* extension. You may maintain this extension or change it to *.enw*. Changing the extension allows Sci2 to automatically read in the data without selecting the data format parameter. Whereas, by keeping the .txt file extension, you will have to select the parameter "Refer Scholarly format" when loading the data into Sci2.

ISI Format Loading questions

Q: I am having trouble loading an ISI file that I downloaded from Web of Science. What should I do?

A: Web of Science recently changed the format for their data. Try updating your version of Sci2 to the newest, v1.0 alpha. Also, see the page about the ISI format, which includes a work-around for loading ISI data in older versions of the tool.

Q: I want to import some ISI data into Sci2 but I can only download files in increments of 500 records from the Web of Science. How can I merge and load an ISI dataset larger than 500 records into Sci2?

A: It is typically best to perform this type of merge before you load the data into Sci2. Download the data from the Web of Science as a plain text file and then open them in the text editor of your choice. It is usually to open the files in a more advanced text editor, such as WordPad or Notepad ++ because they will be formatted in a way that easier to read. However, do not open the files in Word. You will want to copy and paste the records into one file. Make sure to paste the contents of the files after the ER (End Record) tag of the last record and before the EF (End File) tag that closes the ISI file. Once you have merged the files, you will want to shave it as a text (.txt) file. Then you should change the extension from .txt to .isi. If you are having trouble seeing file extensions right-click on the file and select "Properties" (Windows) and then you should be able to see the file name and extension. Simply delete the old extension and replace it with the desired one. On a Mac, you will need to right-click and select "Get Info" and then choose the "Name & Extension" tab to change the file extension.

Now you are ready to load the file into Sci2. Make sure to load the file as an "ISI flat format" file. If you are having trouble loading the ISI file, see the previous question in this section. For ISI data, Sci2 should automatically detect and remove duplicate records. However, you still may need to check over the data, as Web of Science data can sometimes be inconsistent.

Q: The data I am working with is really messy. How much of the data cleaning can I do in Sci2?

A: It depends on the data format, but there are some options for tidying up your data under the Preprocessing menu in Sci2. If you are working with a large dataset in CSV format, you may need to do a lot of the data cleaning prior to loading the file into Sci2. You can do a lot of this data cleaning in Excel or with Google Refine, an extremely useful program for cleaning data.

Q: The ISI data I downloaded was not formatted in the same way as those used in your workflows?

Web of Science provides researchers with two ways to gather citation data: *Citation Lists* are detailed article records that may be generated through the search interface or the marked list tool in; *Citation Reports* simplified citation information and a citation statistics for the articles selected by a user that can be exported into tabular formats. *Citation List* text files are the format used in our network analysis workflows; in particular, this format can be acquired with cited reference lists that are important creating citation networks and document co-occurrence networks. These types of records can be loaded into a Sci2 directly and they are reformatted by the Sci2 program into CSV files for further analysis. Instructions on how to download Web of Science Citation List records in this format can be found in the Data Acquisition portion of the manual.

Network Extraction

Q: How are networks extracted from my data?

A: First a data set is loaded into Sci2. This tool can handle a variety of different data formats, to see all the data formats supported by Sci2 check the data formats section of our documentation. Now, once a data set has been loaded, it can be viewed as a comma separated value (CSV) file. Simply right click on the table icon in the data manager of Sci2 and view the table in the spreadsheet program of your choice.
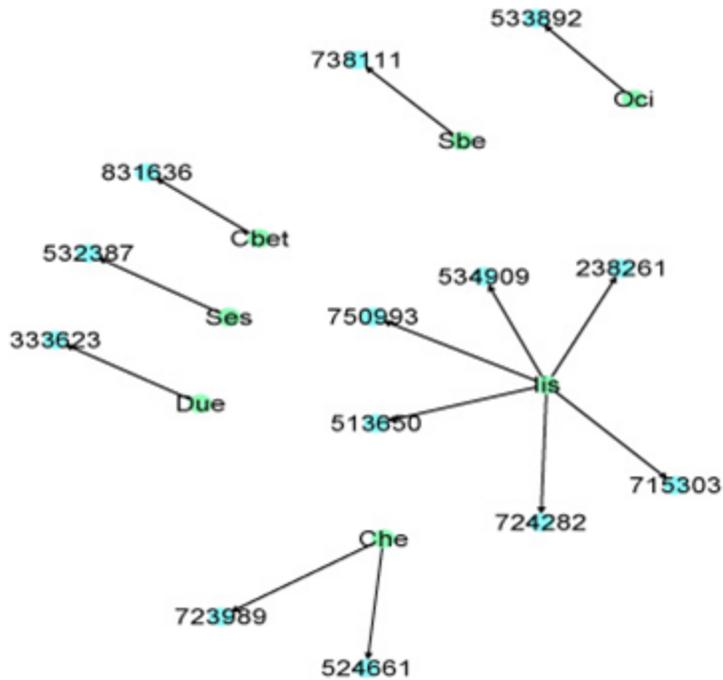
	A	B	C	D	E	F	G
1	Award Number	Title	NSF Organization	Program(s)	Start Date	Last Amendment Date	Principal Investigator
2	238261	CAREER: Visualizing Knowledge Domai	IIS	INFO INTEGRATIC	1-Sep-03	27-Jun-07	Katy Borner
3	333623	Project ENABLE: Learning through Assc	DUE	NATIONAL SMETE	1-Sep-03	13-Sep-06	Javed Mostafa
4	513650	NetWorkBench: A Large-Scale Networl	IIS	ITR-INFORMATIO	1-Sep-05	12-Jul-05	Katy Borner
5	524661	Mapping Chemistry	CHE	PROJECTS	1-Aug-05	14-Jul-06	Katy Borner
6	532387	May 2006 International Workshop and	SES	MATHEMATICAL S	1-Oct-05	26-Aug-05	Stanley Wasserman
7	533892	SCI: Workshop: The Role of Social N	OCI	SPECIAL PROJECT	15-Aug-05	16-Mar-06	Noshir Contractor
8	534909	Collaborative Research: Social Networ	IIS	INFO INTEGRATIC	15-Mar-06	24-Mar-08	Katy Borner
9	715303	Creative Metaphors to Stimulate New	IIS	ITR-CreativeIT	1-Jun-07	2-Jun-09	Katy Borner
10	723989	Mapping Science Exhibit at the 233rd N	CHE	PROJECTS	15-Mar-07	9-Mar-07	Katy Borner
11	724282	III: "Visualizing Network Dynamics" --	IIS	INFO INTEGRATIC	1-Apr-07	6-Apr-07	Katy Borner
12	738111	TLS: Towards a Macroscope for Scienc	SBE	SCIENCE OF SCIEN	1-Jan-08	25-Mar-08	Katy Borner
13	750993	NSF Workshop on Knowledge Manage	IIS	INFO INTEGRATIC	1-Jan-08	7-Sep-07	Katy Borner
14	831636	SGER: Collaborative Research: Mappin	CBET	ENERGY FOR SUST	1-Aug-08	28-Jul-08	Katy Borner

Once loaded into Sci2, all data assume the table format. In the CSV file pictured above, each column head represents a different attribute of the network and each row represents a node in the network. For example, if a directed network were to be extracted from the table above with a target column of *Award Number* and a source column of *NSF Organization*

	A	B	C	D	E	F	G
1	Award Number	Title	NSF Organization	Program(s)	Start Date	Last Amendment Date	Principal Investigator
2	238261	CAREER: Visualizing Knowledge Domai	IIS	INFO INTEGRATIC	1-Sep-03	27-Jun-07	Katy Borner
3	333623	Project ENABLE: Learning through Assc	DUE	NATIONAL SMETE	1-Sep-03	13-Sep-06	Javed Mostafa
4	513650	NetWorkBench: A Large-Scale Networl	IIS	ITR-INFORMATIO	1-Sep-05	12-Jul-05	Katy Borner
5	524661	Mapping Chemistry	CHE	PROJECTS	1-Aug-05	14-Jul-06	Katy Borner
6	532387	May 2006 International Workshop and	SES	MATHEMATICAL S	1-Oct-05	26-Aug-05	Stanley Wasserman
7	533892	SCI: Workshop: The Role of Social N	OCI	SPECIAL PROJECT	15-Aug-05	16-Mar-06	Noshir Contractor
8	534909	Collaborative Research: Social Networ	IIS	INFO INTEGRATIC	15-Mar-06	24-Mar-08	Katy Borner
9	715303	Creative Metaphors to Stimulate New	IIS	ITR-CreativeIT	1-Jun-07	2-Jun-09	Katy Borner
10	723989	Mapping Science Exhibit at the 233rd N	CHE	PROJECTS	15-Mar-07	9-Mar-07	Katy Borner
11	724282	III: "Visualizing Network Dynamics" --	IIS	INFO INTEGRATIC	1-Apr-07	6-Apr-07	Katy Borner
12	738111	TLS: Towards a Macroscope for Scienc	SBE	SCIENCE OF SCIEN	1-Jan-08	25-Mar-08	Katy Borner
13	750993	NSF Workshop on Knowledge Manage	IIS	INFO INTEGRATIC	1-Jan-08	7-Sep-07	Katy Borner
14	831636	SGER: Collaborative Research: Mappin	CBET	ENERGY FOR SUST	1-Aug-08	28-Jul-08	Katy Borner

Then the resulting network visualization will look like this:

533892

738111

Oci

Sbe

831636

Cbet

532387

Ses

534909   238261

750993

333623

Iis

Due

513650

715303

724282

Che

723989

524661

Note, the above network was generated in Sci2 and then opened in Gephi, where the color and labels were added. Sci2 v1.0 alpha supports a bridge to Gephi

In the network above, the green nodes represent the NSF organizations in this data set and the blue nodes represent the awards (labeled by award number) associated with those NSF organizations. You will notice that since the NSF organization was selected as the source node, each organization is represented in the network only once, while it may appear multiple times in the original data set. This is because each row in the data set is associated with a specific award, the target in this extraction. In some cases this is a one-to-multiple relationship and in others a one-to-one, depending on how many awards are associated with a particular organization. This is an example of just one type of network that can be extracted from data loaded into Sci2. You can create your own data set in Excel or any other spreadsheet program, save the data as a CSV (*.csv) file, load it into Sci2, and perform a variety of analyses and visualizations.

You can also practice with the variety of sample data sets located here. It is highly recommended that you follow the sample workflows provided on this documentation wiki.

Q: I am trying to extract a network by using more than two attributes of the data set as input parameters in a network extraction algorithm. Is this possible?

A: Yes, often times you will want to have multiple attributes available in your network extraction. For example, you may want to extract a directed network from NSF principle investigators (PIs) to their awards and size the award nodes by amount awarded. However, you will notice the Extract Directed Network algorithm only allows for two input parameters: source column and target column. More attributes can be added to the network by using property files (aggregate function files). These are simple text files that can be added to many network extractions to aggregate data, allowing you to have more than two attributes in a network with one extraction. More information, including how to create your own property files, can be found here.

Q: While trying to extract a network from a CSV file and using a comma "," in the text delimiter field I got an error.

A: The parameter field will not accept a comma. This is due, in part, to the Prefuse library that was created back in 2006 and the CSV parser is still not able to handle a comma in the same column. You can simply use a ";" or a "|" instead. If your data set contains many commas consider opening the file in Excel and using a macro to replace the commas with another acceptable symbol.

Database Functionality

Q: Is there database functionality for Sci2 v1.0 alpha?

A: Currently, there are ISI, Medline, and Scopus databases for Sci2 v1.0. However, the database plugin not ready for release yet. We need another sprint to get them ready. However, the plugin that allows files to be loaded as databases is available for Sci2 v0.5.2 alpha or older. Be sure to check the Sci2 news page for updated information on current and upcoming releases of the tool. https://sci2.cns.iu.edu/user/news.php

Global Level Studies - Macro

Q: I am interested in visualizing some of my geospatial data, but I only see a world map and a U.S. map. Are there country-specific maps on which I can overlay a country-specific geospatial network?

A: Unfortunately the U.S. map and the world map are currently the only maps we have available. However, if you would like to submit a feature request it can be done by visiting the Sci2 development page. https://sci2.cns.iu.edu/user/developer.php

Saving and Visualizing Networks

Q: GUESS is taking a very long time to visualize my network. Is there any way to speed up this process?

 A: Unfortunately GUESS only uses one core in the CPU instead of multiple and this means visualizing large networks can take some time. Since GUESS is a third party visualization tool supported by Sci2 we don't have control over how many cores are used in processing. However, the latest release, Sci2 v1.0 alpha, supports Gephi (http://gephi.org/) a visualization tool that handles large networks well. For more information see our documentation on Gephi.

Q: I saved my visualization as a Postscript file, but I am having trouble converting the Postscript file to a PDF so it can be viewed.

A: Converting Postscript files can be tricky. If you have a version of the Adobe suite (or the Pro version of Adobe Acrobat) installed on your machine, the converter will come pre-installed. However, there are also some free options available for postscript conversion: Ghostscript and GSviewer. Ghostscript is a PostScript interpreter, and GSview is a graphical interface for Ghostscript; you can download both from Ghostscript, Ghostview and GSview. Both are freeware. The GSview/Ghostscript combination is capable of opening PostScript level two files, which are the type generated by Sci2. See the following web page for a useful tutorial on how to install Ghostscript: http://pages.cs.wisc.edu/~ghost/doc/gnu/7.05/Install.htm

Q: I am having trouble entering commands in the Interpreter.

 A: Typing commands into the Interpreter can be tricky, but once you learn it can be an extremely powerful way to manipulate a network visualization. Here a few common mistakes to avoid:

1. After typing a ":" you will need to hit "enter" and "tab" to start the new line, followed by another "enter" to execute the command. This is most common in the following scenario:

```
>for n in g.nodes:
     n.strokecolor = n.color
```

2. When you type a command in the interpreter you do not need to type ">" to begin, simply begin with the text of your command. To learn more about how to manipulate network visualizations see our documentation on GUESS.

Q: I load the network into GUESS but no node is shown or the network is loaded without Graph Modifier panel.

 A: There are some attribute names are reserved in GUESS. The same attribute name in the data could cause conflict in GUESS. Rename the following attribute names in your data if exist. Then save the data and reload into Sci2.

- count
- size
- type

The following show sample data issues in NWB file format. Simply rename the attribute to size_d, type_d, count_d would solve the problem. See our documentation on GUESS to learn more about GUESS.

```
*Nodes
id*int label*string size*float type*string count*int
```

```
1          "Amber" 4.0 "ellipse" 4
2          "Azizha" 4.0 "ellipse" 5
3          "Brend" 4.0 "ellipse" 6
```

Third-Party Plugins and Licenses

Q: What are the third-party plugins used by Sci2 and what are their terms of use?

A: Sci2 has been released under the Apache 2.0 license. Licenses for the various components used by Sci2 can be found under the "licenses" folder in your Sci2 installation directory.

The Sci2 v1.0 build released as of September 2013 will include additional third-party plugins :

- Google Scholar Citation Reader
- Bing Geocoder
- Flickr Reader
- Facebook
- Twitter Reader

Each plugin as it's own data usage policy as dictated by it's license and terms of use:

- Facebook plugin only allows users download their own data. For more information, see http://www.givememydata.com/.
- Twitter Reader plugin allows users to download the tweets of any twitter user, however the user and the application have to first be registered and authenticated by the API before performing any queries. For more information please see: https://dev.twitter.com/terms/api-terms
- Google Scholar Citation Reader will fetch data indexed by the Google Scholar search engine. For data usage policies please review the Google license.
- Bing Geocoder plugin will require an authentication key (Basic Key or Enterprise Key). The usage policy depends on the type of key. For more information, see http://www.microsoft.com/maps/
- Flickr Reader plugin downloads image URLs through REST service with API key without authentication. For more information, see http://www.flickr.com/services/api/

Please note that if required all plugins can be uninstalled separately.

Q: Does Sci2 perform any encryption or utilize any encryption processes?

A : Sci2 by itself does not perform or use encryption, however plugins such as the Facebook plugin and the Twitter Reader plugin use encryption for user authentication and data downloads.

Q: Does Sci2 have ECCN?

A : No, Sci2 does not have a ECCN

Don't see what you are looking for in Sci2?

Development services can be performed by our team. You can submit a feature request at the Sci2 by visiting the Sci2 web site: https://sci2.cns.iu.edu/user/developer.php. Also, development services can be performed for a fee, prices vary according to the scope of work. Previous paid development/services performed by our center have included evaluation and monitoring, data compilation, visual interfaces for community data, VIVO researcher networking, and consulting/training services. To get an estimate for your project contact our center: http://cns.slis.indiana.edu/.