

## Scaffold Batch Getting Started Guide

Scaffold Batch 4.6.0, Revision 2016.06.03

This getting started guide is meant to help users who are experienced with Scaffold learn how to use Scaffold Batch for high throughput batch processing. Scaffold ships with a batch processing module that, when registered with a Scaffold Batch License, is activated for use. It is run for the command line, and gets its instructions for how to analyze data from an XML driver file called a SCAFML. See [this document](#) for a complete annotated version of the SCAFML driver file.

Scaffold Batch can be set up to run multiple experiments. It can generate SF3 files for collaboration using the [Scaffold Viewer](#), exports for spreadsheet analysis, protXMLs, and mzIdentMLs for further analysis. It's often used in a proteomics pipeline, and scripted for automated batch processing of large data sets or multiple file queues.

This guide will take you through a simple process to install, register, and set up an experiment for batch processing using Scaffold Batch.

### Requirements for This Guide

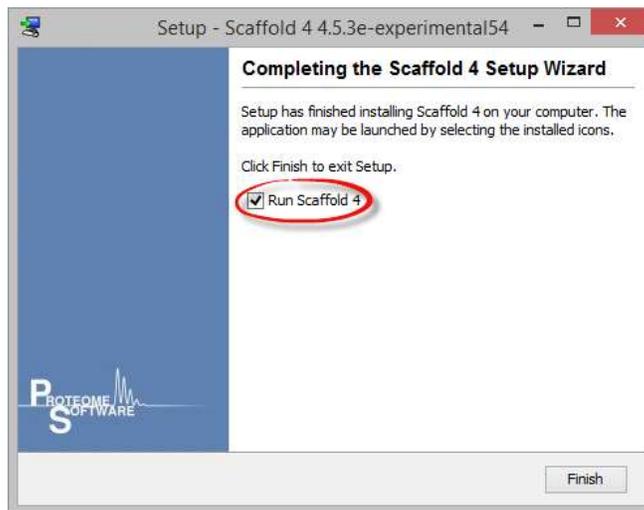
You will need a few things for this exercise:

1. An Internet connection to download Scaffold Batch and request a batch evaluation key (if you have a batch license, you can skip this).
2. Access to a browser and our [documentation](#) web page.
3. A text editor, such as [Notepad++](#) for editing your SCAFML text file. Windows Notepad or WordPad will also work.
4. Access to a terminal application like `cmd.exe` on Windows, or the terminal on Linux (this exercise will use the Windows command line tool, but the same steps can be followed on a Linux system with a terminal application).
5. The [data set and FASTA](#) provided on our web site (the same data that was used to generate our **Label-Free.sf3** demo file found in the Scaffold Desktop application).

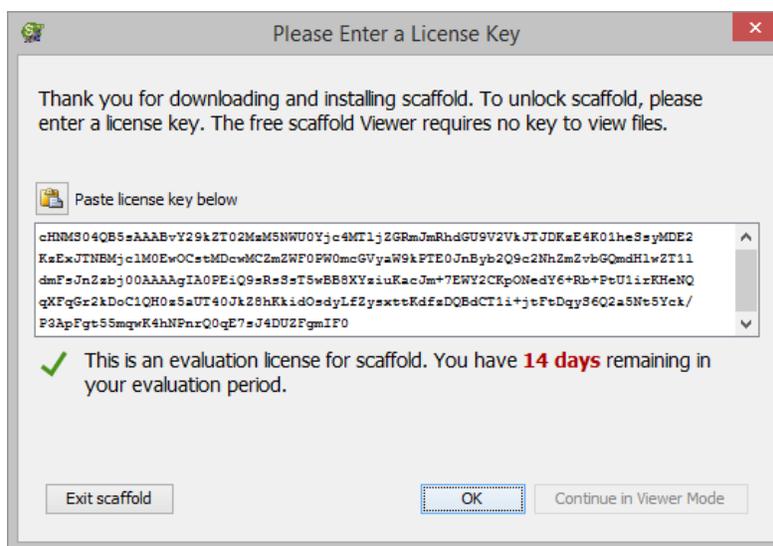
### Installation and License Registration

If you already have Scaffold installed, and it's licensed for Scaffold Batch, you can skip to the next section: [Setting Up Your Experiment](#).

Otherwise, contact our [technical support](#) department and request a Scaffold Batch evaluation key. Next, after you receive the key, go to the download page, and [download](#) the appropriate installer. Run the installer (choosing defaults is fine), and at the end of the installation wizard, check the box to run Scaffold.



A dialog box will come up, asking for your key. Paste the key you received into the box, and you should receive a message saying the registration was successful. Click **OK** and leave Scaffold running for the next step.



## Setting Up Your Experiment

Now you are ready to set up your experiment. First, go to the **data set and FASTA** link above, from number 5, and download the ZIP file to your computer. Unzip the file, and navigate into the folder. You should see these folders/files:

Data folder: **qe2\_10122012\_30\_10\_20**

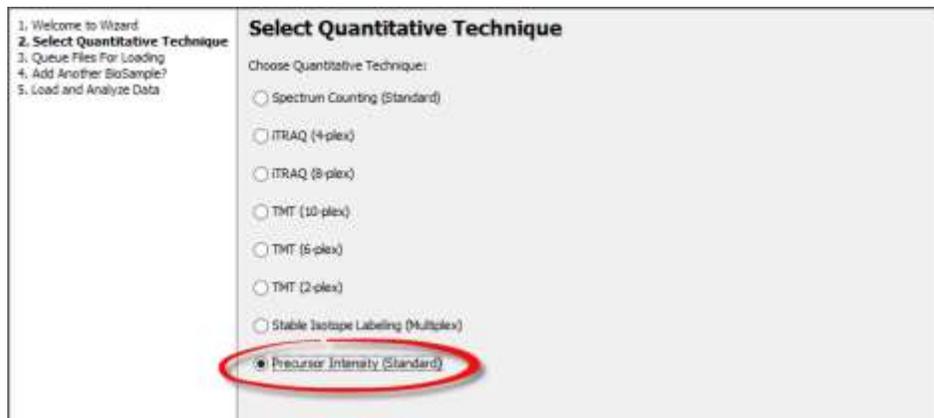
Data folder: **qe2\_10122012\_31\_10\_20**

FASTA file: **uniprot\_sprot.fasta**

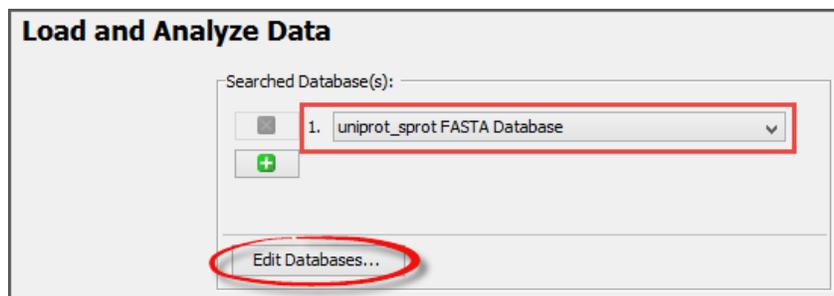
SCAFML file: **template.scafml** (you will generate your own SCAFML later on)

Go back to Scaffold, and you should see a **Welcome to Scaffold Q+S Evaluation** dialog box. Click the **New** button. If this dialog is not present, go to **File > New** to create a new experiment.

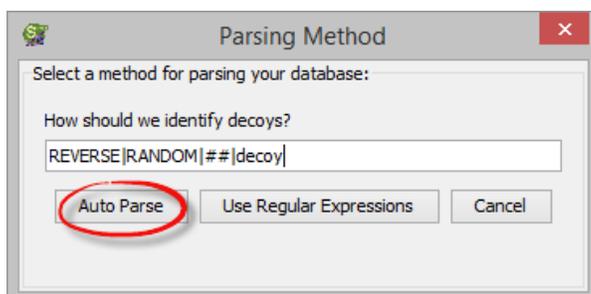
Select **Next >**, and then choose the radio button corresponding to **Precursor Intensity (Standard)**. Click **Next >** and fill in **Sample Name**, **Sample Category**, and check the **Condense data as it is loaded** checkbox. Choosing the defaults is OK.



Click **Next >** again, and then click the button **Queue Files For Loading**. Navigate to the where you unzipped the file from above, and select one of the data folders. Click **Next >**, and then **Add Another BioSample**. Repeat the process for adding the second data folder. Click **Next >** until you get to the **Load and Analyze Data** step.



If you haven't configured any FASTA databases yet, click the **Edit Databases...** button, and configure one now. The one provided in the ZIP file above is called: **uniprot\_sprot.fasta**. For this exercise, use the **Auto Parse** option. Make sure the database you want to use is selected from the drop down menu.



We are not actually going to process the data right now, so you can click **Done** when you are finished.

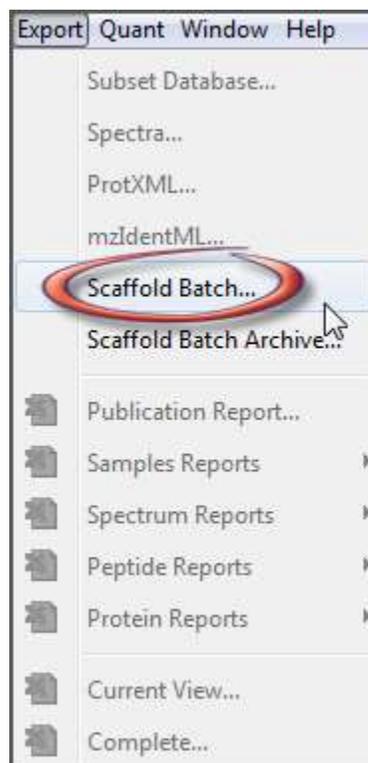
## Export the SCAFML

To create a SCAFML that we can work with in the next section, go to **Export > Scaffold Batch**, and select a location on your computer to save the file. Name it something like **demo.scafml**.

After you have saved the file, open it with your text editor (eg, Notepad++). Take a look at the structure of this file.

This is an XML file, and you will probably recognize some of the XML elements and attributes. Some of these attributes are set to the defaults, but some are set according to those settings that were set when you exported the SCAFML.

**HINT:** *While it is perfectly fine to edit the SCAFML directly (as we'll do below), an easy way to set certain parameters is to set up the experiment in Scaffold Desktop first, and then export the SCAFML as we've done here.*



## Running Scaffold Batch from the Command Line

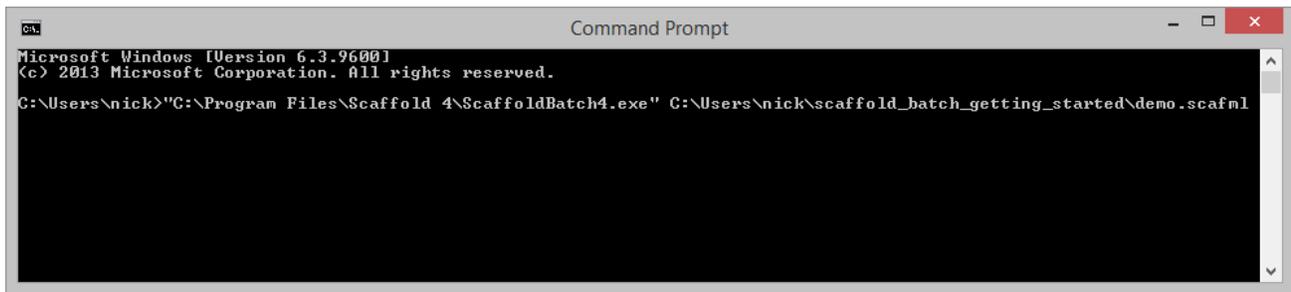
At this point, you could technically run this file and generate a valid SF3 that can be opened with the Scaffold Viewer. To test this presumption, let's give it a try. Since Scaffold and Scaffold Batch are Java programs, the JVM they use relies on RAM to process data. Therefore, for optimal performance, you must tell Scaffold Batch how much memory to use. If you only use Scaffold GUI, this can be set by going to **Edit > Preferences > Memory**. For Batch, this can be set by editing the first line in the **ScaffoldBatch4.vmoptions** file in the Scaffold installation directory (this will tell Batch to use 5 GB of RAM, if needed):

```
-Xmx5000m
```

Open the command line tool. In Windows this program is called **cmd.exe**, and can be found by searching for that name in the search tool for Windows. Once you have it open, run this command (assuming the work is done in your user directory):

```
"C:\Program Files\Scaffold 4\ScaffoldBatch4.exe" C:\Users\nick\scaffold_batch_getting_started\demo.scafml
```

The call to the executable is in quotes because the path to the Scaffold installation has a space in it. This is Windows way of making sure the correct path is used.



```
Command Prompt
Microsoft Windows [Version 6.3.9600]
(c) 2013 Microsoft Corporation. All rights reserved.
C:\Users\nick>"C:\Program Files\Scaffold 4\ScaffoldBatch4.exe" C:\Users\nick\scaffold_batch_getting_started\demo.scafml
```

This will run for several minutes, and generate a file called **My Experiment.sf3** in the same directory as the data and SCAFML. Feel free to open this with Scaffold Desktop (or the free [Scaffold Viewer](#)).

### Fine-tuning Your SCAFML

This last section explores some of the ways you can fine tune your SCAFML so it makes more sense, and is clearer for others viewing it. First thing to note is that Scaffold Batch creates an SF3 called **My Experiment.sf3**. This is not the most descriptive name, so let's change that. With the SCAFML open in your text editor, find the **name** attribute in the **Experiment** element (`<Experiment name="My Experiment" . . . >`). Change this to something more useful:

```
<Experiment name="batch_demo" . . . >
```

This will generate an SF3 file called **batch\_demo.sf3**. Next, let's remove some of the extra attributes that are not needed. As mentioned above, some of these were exported because they were set in the GUI, but some are defaults. Included in the ZIP folder from our web site is a **template.scafml** file that shows the minimum elements and attributes needed to process data with Scaffold Batch.

Using the template and the SCAFML you generated, try and clean up your **demo.scafml**. For example, remove those attributes that are not in the template. Maybe rename the BioSample names or add descriptions:

```
<BiologicalSample name="sample_1" . . . >
```

```
. . .
```

```
<QuantitativeSample description="Treatment Sample" name="treatment_1" . . . />
```

Perhaps add an export for an experiment report:

```
<Export thresholds="thresh" type="experiment-report"
  experimentDisplayType="Exclusive Unique Spectrum Count" />
```

Here is an example of a working SCAFML that was cleaned up:

```
<?xml version="1.0" encoding="UTF-8"?>
<Scaffold pathsep="/" version="exp45">
  <Experiment name="batch_demo">
    <FastaDatabase id="db0" path="uniprot_sprot.fasta"/>
    <MultiplexExperimentalDesign analysisType="Intensity-based"
      experimentType="independent-groups"
      quantitativeTechnique="Precursor Intensity">
      <ReferenceAlignmentGroup id="0">
        <Category id="Reference">
          <QuantitativeSample description="" msID="1"
            name="Control" reporter="Precursor Intensity Sample"/>
        </Category>
        <Category id="Treatment">
          <QuantitativeSample description="" msID="2"
            name="Treatment" reporter="Precursor Intensity Sample"/>
        </Category>
      </ReferenceAlignmentGroup>
    </MultiplexExperimentalDesign>
    <BiologicalSample category="Control" name="T1"
      database="db0" quantitativeTechnique="Precursor Intensity">
      <InputFile>qe2_10122012_31_10_20\qe2_10122012_31.xml</InputFile>
    </BiologicalSample>
    <BiologicalSample category="Treatment" name="T2"
      database="db0" quantitativeTechnique="Precursor Intensity">
      <InputFile>qe2_10122012_30_10_20\qe2_10122012_30_10_20.xml</InputFile>
    </BiologicalSample>
    <DisplayThresholds id="thresh" />
    <Export thresholds="thresh" type="sf3" />
    <Export thresholds="thresh" type="experiment-report"
      experimentDisplayType="Exclusive Unique Spectrum Count" />
  </Experiment>
</Scaffold>
```