

WebPlotDigitizer User Manual

Version 3.4

Ankit Rohatgi*

September 22, 2014

Contents

1	Introduction	3
1.1	History	4
1.2	User Manual and Tutorials	4
1.3	License	4
1.4	Source Code	4
1.5	Availability	4
1.6	Supported Browsers	5
1.7	Citing WebPlotDigitizer	5
1.8	Reporting Issues	5
1.9	Data Privacy	5
1.10	Funding	5
2	Loading Plots	6
2.1	Supported Image Formats	6
3	Define Axes	6
3.1	2D (X-Y) Plot	7
3.1.1	Format of Calibration Values	7
3.2	Polar Diagram	8
3.3	Ternary Diagram	8
3.4	Map With Scale Bar	8
3.5	Image (Align to Image Pixels)	10
3.6	Transformation Equations	10
4	Acquire Data	10
4.1	Manual Mode	11
4.2	Automatic Mode	11
4.3	Digitization Algorithms	13

*E-Mail: ankitrohatgi@hotmail.com

4.3.1	Averaging Window	13
4.3.2	X Step with Interpolation	14
4.3.3	X Step	14
4.3.4	Blob Detector	14
5	Multiple Datasets	15
6	Handling Digitized Data	15
6.1	Sort Data	15
6.2	Formatting Dates	15
6.3	Number Formatting	17
6.4	Export to .CSV File	17
7	Distance and Angle Measurements	17
8	Import and Export JSON	18

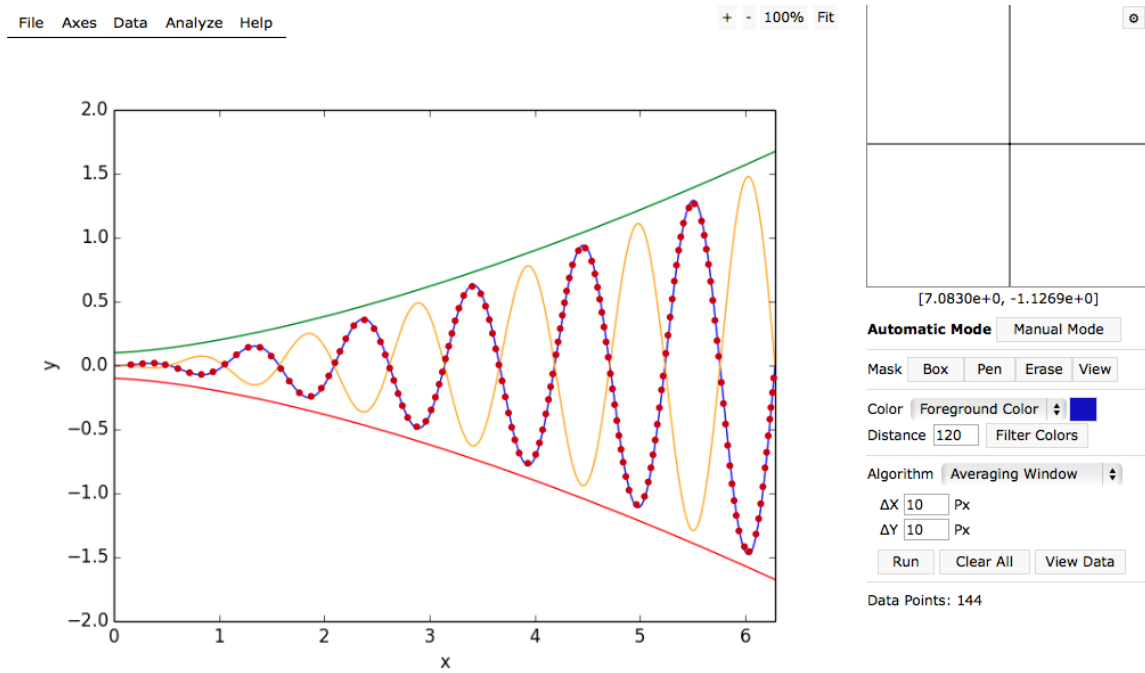


Figure 1: Screenshot of WebPlotDigitizer showing the data points recovered on a plot via automatic detection.

1 Introduction

A large quantity of useful information is available only as plotted data points in an image. In these images, it is easy to determine the relationship between the variables involved, but recovering the exact numerical values of the data is usually a tedious and error prone process. To aid this time consuming task of data recovery, many digitization softwares have been developed over the years. However, even with the abundance of free and commercial softwares, this task remains daunting and prone to errors. Many of the existing softwares are either designed to work only on specific operating systems or work with a limited variety of plots. Some are just difficult to use or prone to errors. Finally, many require a paid license which prevents their widespread use by students, independent researchers or organizations with limited resources.

Because of the above limitations in current digitizing softwares, WebPlotDigitizer was developed to facilitate easy and accurate data extraction from a variety of plot types and also maps. This program has been built using HTML5 which allows it to run within most popular web browsers and does not require an installation process that is performed by the user. This is distributed free of charge as an opensource software. A screenshot of a typical session of the software is shown in Figure 1.

1.1 History

WebPlotDigitizer was initially developed while working on my graduate studies at the University of Notre Dame. Having to extract data from many publications for comparing and contrasting my own findings in the lab was a time consuming task. The search for a tool to aid this process usually ended in realizing that most of the existing softwares for this purpose did not fulfill many of the requirements. I was faced with a similar task as an undergraduate student. Back then, just writing a simple Java based code for picking a few points by clicking on them manually was sufficient. My advisor at Notre Dame also had a Matlab code to do this job for two dimensional XY plots, but the points still had to be picked manually. None of these options were very convenient.

Some of the experimental work in the lab required me to learn some basic image processing techniques which eventually formed the basis of the automatic detection algorithms used here. Image processing knowledge along with some interest in learning the very popular HTML5 APIs were a perfect match to create a web based data extraction tool like this.

WebPlotDigitizer is now used by hundreds of people everyday and because of this interest, I have continued to refine the software in my spare time.

1.2 User Manual and Tutorials

This user manual describes the various capabilities of the software and aims to help the user in making an effective use of the software. This manual may be updated continuously to match the latest deployed version of the software. A few video tutorials for the previous version of the software are available at <http://arohatgi.info/WebPlotDigitizer>. Other resources to find technical information about the software are also being considered.

1.3 License

WebPlotDigitizer is distributed under GNU General Public License version 3 by Ankit Rohatgi. For complete terms and conditions, please refer to <http://www.gnu.org/copyleft/gpl.html>

1.4 Source Code

WebPlotDigitizer is an open source software (see above). The source code can be obtained from GitHub (<https://github.com/ankitrohatgi/WebPlotDigitizer/>). Feel free to contact via email if you wish to contribute to this project.

1.5 Availability

The latest released version of the software can be used directly from the website <http://arohatgi.info/WebPlotDigitizer>. For the Google Chrome web browser, an *app* pointing to the online software is also available at the Chrome App Store (<https://chrome.google.com/webstore/category/apps>).

1.6 Supported Browsers

Version 3.4 was ensured to work without major issues on the following browsers:

- Safari 7.0.6 on Mac OS 10.9.4
- Google Chrome 37 on Mac OS 10.9.4
- Firefox 32.0 on Mac OS 10.9.4
- Google Chrome on Windows 7 32-bit
- Firefox 32.0 on Windows 7 32-bit
- Internet Explorer 11 on Windows 7 32-bit

It is expected that browsers similar in functionality and support for the HTML5 API should not have any major problems executing the version 3.4.

1.7 Citing WebPlotDigitizer

If you wish to cite WebPlotDigitizer in any of your works, then please use the following information:

Author	Ankit Rohatgi
Title	WebPlotDigitizer
Website	http://arohatgi.info/WebPlotDigitizer
Version	3.4
Date	September, 2014
E-Mail	ankitrohatgi@hotmail.com

1.8 Reporting Issues

In case of issues with the data recovery, access to the software or general technical questions, feel free to contact via e-mail. Issues specific to bugs in the software can also be reported on the issues page on GitHub: <https://github.com/ankitrohatgi/WebPlotDigitizer/issues>

1.9 Data Privacy

WebPlotDigitizer's image analysis code runs entirely on the user's computer and does not store the loaded images or data on to any server. When *Graph in Plotly* option is selected, the digitized data is transmitted to Plotly (<http://plot.ly>) servers.

1.10 Funding

WebPlotDigitizer is not funded by any organization. PayPal donations from the website and my personal time and effort are responsible for keeping this project afloat.

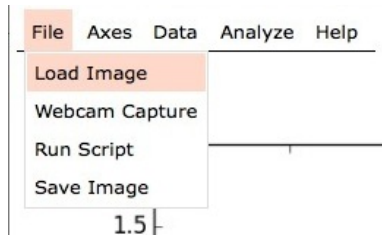


Figure 2: The top row of buttons in WebPlotDigitizer.

2 Loading Plots

The image file containing the figure to be analyzed can be loaded into the software in the following ways:

1. **Drag & Drop Operation:** Image can be dragged and dropped from the file browser on to the image viewing area of the application.
2. **File Menu → Load Image:** Browse for a file on the hard disk to load.
3. **Copy-Paste from Clipboard:** This is only supported in Google Chrome web browser. An image selected by copying in a PDF or an image viewer can be pasted on to the software via a simple copy-paste operation.
4. **File Menu → Webcam Capture:** A snapshot taken from the webcam can also be used. For best results, the webcam should be pointed directly along the normal to the plot surface. In the future, some image transformation features might be added to WebPlotDigitizer to compensate for the distortions.

2.1 Supported Image Formats

WebPlotDigitizer relies on the image formats supported by the HTML5 *canvas* element. Most browsers support common image formats such as JPEG, PNG, BMP and GIF. Since the support for an image format depends on the browser used to access the software, please refer to your browser's manual for details. For popular browsers, you can also refer to Wikipedia (http://en.wikipedia.org/wiki/Comparison_of_web_browsers#Image_format_support).

3 Define Axes

After loading the desired image, you should specify the type of axes that is used in the plot. This step is required for the software to correctly map the image pixels to possible data points in the image. Depending on the plot type, you will have to select a few known points on the axes. On clicking the *Axes → Define Axes* menu item (see Figure 2), you should be presented with the menu shown in Figure 3.

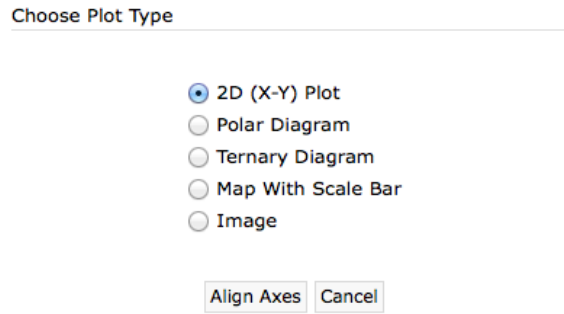


Figure 3: Popup with plot types that are supported in the software.

3.1 2D (X-Y) Plot

Most plots that are on a two dimensional cartesian coordinate system fall under this category. Plots with non-orthogonal axes or plots rotated by an angle are also supported.

On selecting this option, you should be presented with a popup window which asks you to click on two points on the horizontal axis and two points on the vertical axis. After clicking *Proceed* on that popup, click on two points on one of the two axes (x_1, x_2) and two on the other (y_1, y_2) . For better accuracy during the digitization process, pick the points that are as far away from each other as possible. Also, note down the data values at (x_1, x_2) and (y_1, y_2) positions as you will be required to enter those once the four points have been clicked on.

You can also make fine adjustments to the location of a calibration point that is highlighted in green using keyboard cursor keys. Pressing the *Shift*-key along with the cursor keys will move the point at a faster rate. To highlight a point to be adjusted, you can also click on the point after selecting the four required points. Zooming-in the image can help improve the accuracy.

After the four points have been clicked on, press *Continue* and another popup window will appear where you will be required to enter the values at these points. This helps the software map the image pixels corresponding to data points to their actual values when the image is digitized. To view the mapping equations, select the *Transformation Equations* option from the *Axes* menu.

3.1.1 Format of Calibration Values

Like most computer programs, WebPlotDigitizer accepts integers (e.g. 1, 2, 3 etc.) or floating point numbers (e.g. 3.14159). Some extra things to keep in mind are as follows:

1. Fractions (e.g. $1/2$) are not computed as numbers.
2. For exponentials, the caret symbol (^) is not recognized and the values have to be entered as 1.45e-10 for 1.45×10^{-10} (for example).
3. **Dates:** This is enabled only for 2D (X-Y) Plots. At the time of calibration, the dates have to be entered in the format shown below. With the final digitized data, however, results can be formatted in many different ways (see section 6.2).

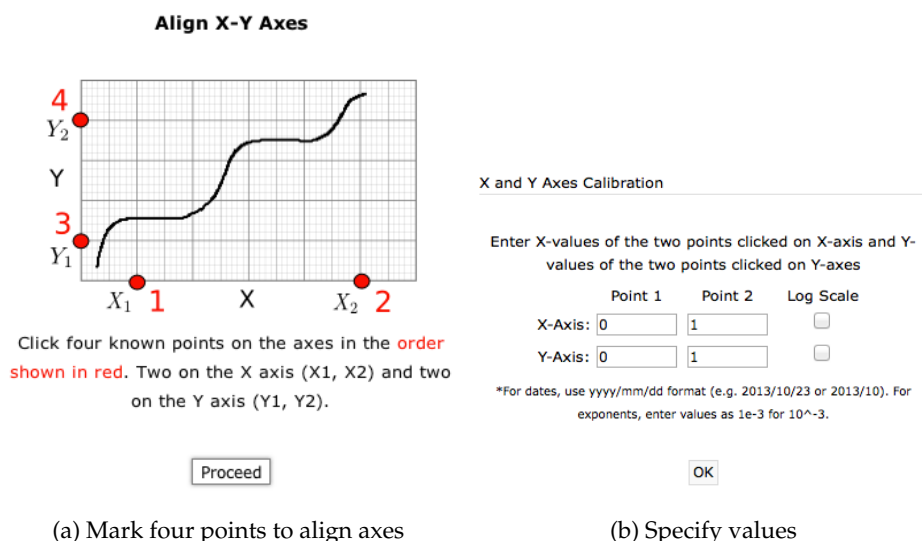


Figure 4: Alignment for 2D (X-Y) Plot.

Date	Format	Examples
Year, Month and Date	YYYY/MM/DD	2012/10/23, 2012/10/5 or 2012/10/05
Year, Month	YYYY/MM	2012/10 or 1989/5
Year	YYYY	2012 (treated as any integer)

3.2 Polar Diagram

Select this option if the data points in the image are plotted on a polar axes. On selecting this, you will be required to click on three known points including the center of the polar diagram (Figure 5). After clicking on 3 points, you can also select the axes orientation and select Degrees or Radians for the angle. The values entered here also have to follow the format similar to 2D (X-Y) Plots. Dates are not parsed here.

3.3 Ternary Diagram

Ternary phase diagrams are typically harder to read than simple two dimensional cartesian or polar plots. Using this software to recover data makes the process of data recovery extremely straightforward and thus reduces the possibility of misinterpreting the data. For this type of plot, simply mark the three corners as shown in the instructions and then specify the range of variables and orientation of the diagram (Figure 6).

3.4 Map With Scale Bar

This plot type is similar to 2D (X-Y) Plots and is provided only as a convenience for images that only have scale information (e.g. microscope images or maps). To calibrate this plot type, simply click on the two ends of the scale bar and enter the scale value. You can also specify the label to be used for

Align Polar Axes

Click on the center, followed by two known points.

Point 1

R₁:

Θ₁:

Point 2

R₂:

Θ₂:

☒ Degrees
 ☐ Radians

☐ Clockwise

(a) Mark three points to align axes

(b) Specify values

Figure 5: Alignment for Polar Diagram.

Align Ternary Axes

Click on the three corners in the order shown above.

Select Range of Variables

Axes Orientation

☒ Normal
 ☐ Reverse

Range of Variables

☒ 0 to 1
 ☐ 0 to 100

(a) Mark three points to align axes

(b) Specify values

Figure 6: Alignment for Ternary Diagram.

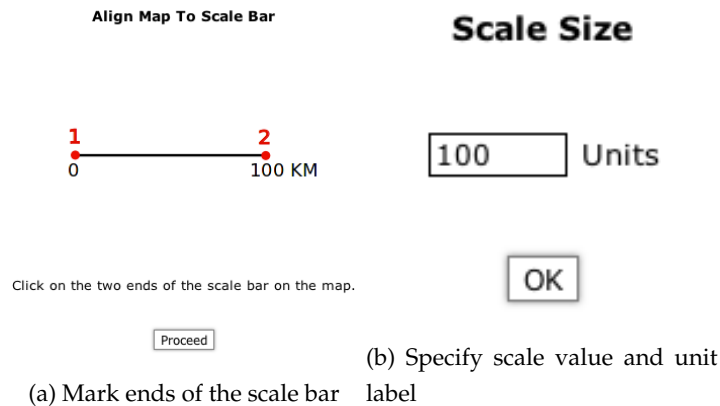


Figure 7: Alignment for maps and microscope images.

the unit (Figure 7). The coordinates reported by the software assume the origin to be located at the bottom left of the image. The (x,y) values that are generated are scaled by the scaled factor entered during calibration.

3.5 Image (Align to Image Pixels)

This plot type is similar to the Map plot type and is also provided as a convenience over the generic 2D (X-Y) plots. If you select this type, no calibration information is required as the software will calibrate the on-screen pixels to the pixels on the image automatically. This plot type is useful for looking up exact pixel location of a feature in an image.

3.6 Transformation Equations

After defining the axes, you can select the *Transform Equations* option in the *Axes* menu to view the relationship that is constructed by WebPlotDigitizer to convert image pixels to data. This feature allows users to use this mapping in other graphics codes or softwares where it may be required to convert image pixel locations to corresponding data or vice versa.

4 Acquire Data

Once the plot axes have been calibrated, you can begin selecting data points on the image. Also note that the numbers below the zoom window reflect actual data coordinates corresponding to your mouse position on the image. If you see incorrect numbers here, then perhaps incorrect calibration values were entered. You must repeat the axes calibration in this situation.

WebPlotDigitizer should also show a side panel with the data acquisition controls (Figure 8) when the axes are aligned. This sidebar can also be brought up by selecting the *Acquire Data* option in the *Data* menu. The data acquisition can be performed in either manual or automatic mode. You can alternate between the two modes at any time. In the manual mode you can add, adjust or remove data points by manually clicking at the desired locations. In the automatic mode, you can set up

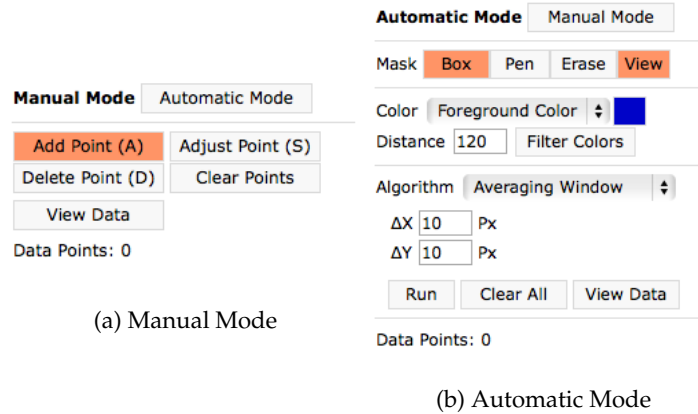


Figure 8: Data acquisition controls.

and execute an extraction algorithm that can differentiate between the data points and the image background and identify several data points in a short time.

4.1 Manual Mode

The controls available in this mode are as follows:

1. **Dataset:** Pick the dataset from the list to specify the data series the data points are to be added to (or modified). See section () for further details.
2. **Automatic Mode:** Switch to automatic extraction mode.
3. **Add Point (A):** After clicking this button or pressing the 'A' key, you can click in the image area to add data points. Any point added on the image will automatically be converted from on-screen pixels to data values utilizing the axes calibration.
4. **Adjust Point (S):** After clicking this or pressing the 'S' key, you can click on an existing data point to select it. The selected point can then be repositioned using the cursor keys on the keyboard. Press Shift+Cursor key for a faster rate of movement.
5. **Delete Point (D):** After clicking this button or pressing the 'D' key, you can click on a previously added data point to delete it.
6. **Clear Points:** This deletes all data points added on the image. This does not clear the axes calibration.
7. **View Data:** This launches a popup dialog where the collected data can be viewed, exported to a CSV file or Plotly.

4.2 Automatic Mode

The controls available in the automatic mode are used for selecting the appropriate algorithm and providing the necessary inputs required for automated extraction of data points.

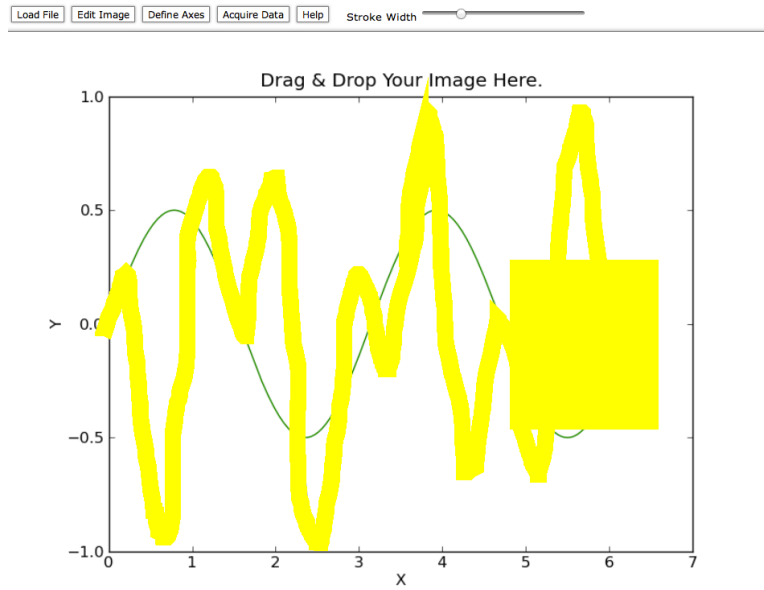


Figure 9: Use the *Box*, *Pen* and *Erase* tools to mark the region containing the required data (automatic extraction).

Automatic data extraction relies on separating the color of the data points or curves from the background in the image. The extraction algorithms can work in two modes of color extraction: Foreground mode and Background mode. In the foreground mode, the algorithms look for the foreground color specified for the data and ignore everything else. In the background mode, the algorithms include everything except the background color as potential data points. If the data points or curves of interest are uniformly colored (approximately), then the foreground mode may be more suitable. Otherwise if the background is uniformly colored (approximately) and the curve or data points are not then the background mode may be more suitable.

The extraction algorithms also need to know the region of the image to be searched for the specified colors. The software does not search the entire image as in many cases, the data point or the curve colors may be repeated in non-data parts of the image. To specify the region of interest, use the *Box*, *Pen* and *Erase* tools (described below) to paint over a yellow mask over the data part of the image as shown in Figure 9. The extraction algorithms will look for data only under the yellow colored region. The various controls available in this mode are as follows:

1. **Dataset:** Pick the dataset from the list to specify the data series the data points are to be added to (or modified). See section () for further details.
2. **Manual Mode:** Switch to manual mode. You can switch between these modes at any time.
3. **Mask Controls:** These controls are to be used to mark the search region for the extraction algorithms. Use the following drawing tools to highlight the approximate region containing data. You should try to exclude all features that are similar in color to the desired data points.
 - (a) **Box:** This is used to mark a rectangular region to be used during the search for data points.

- (b) **Pen:** This is used for marking areas of the image to be included in the search for data points via free hand drawing.
 - (c) **Erase:** This is used to unmark the areas marked using the *Box* or *Pen* tools.
 - (d) **View:** This is used to simply view the marked search region.
4. **Color Controls:** Use these controls to specify the color of the data points and an approximate RGB distance from this color that is acceptable. You either need to specify the foreground color of the data or the color of the background. In most cases, the distance value will require some hit-and-trial based on the results generated by the extraction algorithm.
 - (a) **Foreground/Background Color Selection:** Select the appropriate option from the drop-down menu and click on the colored button on the right to show a popup for color selection. The popup dialog also shows a list of most dominant colors at the bottom that can be selected.
 - (b) **Distance:** Change this value to adjust the range of colors around the selected color to be accepted as foreground or background.
 - (c) **Filter Colors:** Test the color selection and the specified distance value. The area highlighted in yellow is the region that will finally be used by the automatic extraction algorithms.
 5. **Algorithm Selection:** Select the appropriate algorithm depending on the application. Each algorithm is described in detail below.
 6. **Run:** Start the auto-detection algorithm. After this is completed, the detected points should appear over the image. If necessary, adjust the parameters of the extraction algorithm, mask or color settings and run the algorithm until you are satisfied. You can also switch to manual mode and edit these data points.
 7. **Clear All:** Clear all data points in the current dataset.
 8. **View Data:** This launches a popup dialog where the data can be viewed and exported to a CSV file or Plotly.

4.3 Digitization Algorithms

Four different digitization algorithms are available in WebPlotDigitizer as shown in Figure 10. The *Averaging Window* algorithm is set as the default as it is usually suitable in most simple cases.

4.3.1 Averaging Window

As mentioned above, this is probably the most useful algorithm and is useful across many plot types. The data points detected by this algorithm are determined by calculating the average locations of pixels with acceptable color that lie in a small regions that are ΔX pixels (on-screen) wide and ΔY pixels (on-screen) tall. As a user, you should increase the size of this window for thick lines or large data points and decrease it for thin lines. If you see multiple points incorrectly detected across the width of a thick data curve, then you need to increase the numbers specified here. The fact that this

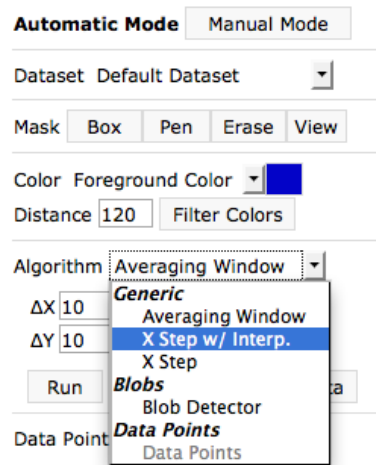


Figure 10: Four automatic extraction algorithms are available. These are subject to change in the future releases.

requires on-screen pixels may be changed in the future so that the values in actual units in the current axes can be specified.

4.3.2 X Step with Interpolation

This algorithm is available only for non log-scale 2D (X-Y) axes plots. In the future, this will be extended to other axes types. This algorithm can identify data points at regular intervals on the X-axis that fall between X_{min} and X_{max} and Y_{min} and Y_{max} . The data points are spaced at an interval ΔX units apart. This algorithm interpolates over missing data using cubic splines and is therefore suitable even for curves with dotted lines or a series with just data points. The *Smoothing* value can be increased from zero (for example, try 0.5) to average over a larger neighborhood around the data points. This is useful for reducing noise in the captured data. (Also see: http://arohatgi.info/WebPlotDigitizer/blog/posts/discontinuous_data.html)

4.3.3 X Step

This is a simplified version of the *X Step with Interpolation* algorithm. this does not use cubic splines to interpolate over data. The *Line Width* parameter is the y-direction thickness of the curve in on-screen pixels (In future, this will be updated to use the appropriate units in Y).

4.3.4 Blob Detector

This algorithm is useful for counting number of objects, determining the location of their centroids, calculating areas and also first moments of continuous objects. All object that are within the *Min Diameter* and *Max Diameter* range are added as data points. (Also see: http://arohatgi.info/WebPlotDigitizer/blog/posts/blob_detection.html)

5 Multiple Datasets

For plots that have multiple sets of data, you can capture data for one dataset, export the CSV file, then clear all the data points and start over to collect data for the next set. This may be acceptable in many cases, but this is can be somewhat inconvenient if you wish to return to an old dataset and modify some data points. Also, this way, you will not be able to store data for all the datasets in an exported JSON file (see later).

A better option is to create a separate dataset for each series you wish to capture by selecting the *Add* option in the *Manage Dataset* popup that can be accessed from the *Data* menu. From this popup, you can also delete or rename a dataset. After adding multiple sets, you can then switch between these from the dataset menus available in the manual and automatic extraction controls and also in the *View Data* window. By default, one dataset called "Default Dataset" is available.

6 Handling Digitized Data

Once the required data points are marked on the image using the manual mode, automatic mode or a combination of both, the digitized values can be seen by clicking the *Create .CSV* button. This presents a popup window as shown in Figure 11. Here, the digitized values can be sorted by the variable or in order of the distance between the points (Nearest Neighbor). The values can also be copied and used in common data analysis softwares. Recently, an option to send these values over to another cloud based data analysis and graphing software called Plotly (<http://plot.ly>) has also been added.

6.1 Sort Data

The digitized data can be left unsorted (Raw Output) or by one of the axes variables in ascending or descending order. The Nearest Neighbor option sorts the data depending on the distance of the points from each other.

6.2 Formatting Dates

If one or both of the axes in a 2D (X-Y) plot contain dates then fields to specify the output format of the values are also shown (Figure 12). In these fields, the following pieces of text are replaced with the corresponding part of the date to format the text (case insensitive):

Text	Replaced With	Example
YYYY or yyyy	Year, all digits	2012
YY or yy	Year, last two digits	98 for 1998
MMMM or mmmm	Month, full name	January
MMM or mmm	Month, short name	Jan
MM or mm	Month, numeric	10
DD or dd	Date	23

A few examples are shown below:

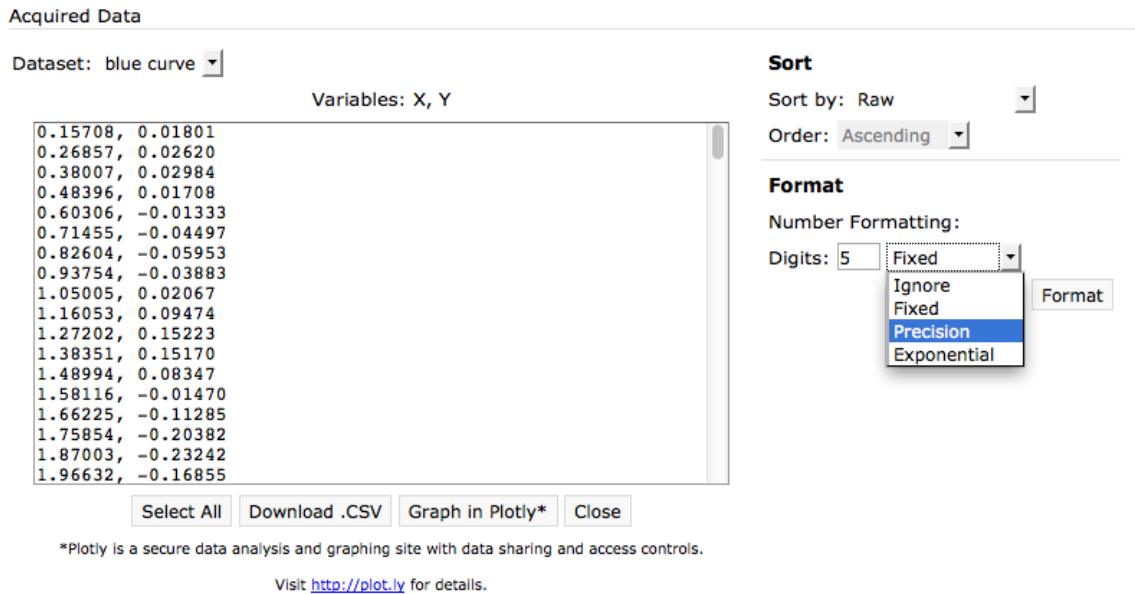


Figure 11: CSV formatted digitized data. This can be sorted, pasted to a .CSV file or graphed in Plotly.

rt by: Raw Output ▾ Order: Ascending ▾

Date Formatting: X mmm ~~yyy~~ Format

Variables: x, y

Figure 12: Field to specify formatting of dates. This appears on X, Y or both axes depending on which variables contained dates at the time of axes calibration.

Format Field Text	Date Shown (Example)
dd-mm-yyyy	23-10-2012
mmm-yyyy	Oct-2012
yyyy abc mmmm	2012 abc October
mmm 'yy	Oct '12

6.3 Number Formatting

The number formatting of the digitized data columns can be changed by specifying the number of digits and the formatting style from the menu. Three formatting styles are available at the moment:

1. **Fixed:** Use this to display the desired number of digits after the decimal.
2. **Precision:** Use this to set the number of significant digits.
3. **Exponential:** Use scientific notation with the desired number of digits after the decimal.

The *Ignore* option disables the number formatting.

6.4 Export to .CSV File

Comma-Separated Values (CSV) format files are simple text files containing tabular data. Each row of line corresponds to a table row and the column values are separated by a character or a string (usually just a comma)¹. Due to its simplicity, CSV format is supported in most data analysis softwares like Microsoft Excel, Matlab etc.

To save the values from WebPlotDigitizer into a CSV file, all you have to do is open your favorite text editor (e.g. Notepad on Microsoft Windows) and Copy-Paste the values shown in the results popup window. Save the file with a .CSV file extension. You can now use this file in your favorite data analysis software.

7 Distance and Angle Measurements

Besides extracting data points from plots, WebPlotDigitizer can also be used to make some simple distance and angle measurements on images. In the future, many more advanced analysis tools will be added. To make distance or angle measurements, choose the appropriate option from the *Analysis* menu.

For distance measurements, if the chosen plot type is a *Map* axes, then the distance is converted to the units specified during calibration. Otherwise distance values are shown in pixels on the original image (not same as on-screen pixels!). Angle measurements are not dependent on any axes. In each case, you can make a single measurement after pressing the *Add* button (or pressing 'A' key) and then clicking on the image. After the measurement is made, you can adjust the highlighted (green) point using the cursor keys on the keyboard (use Shift for faster movement). To change the highlighted point, click on the desired point. A measurement can be deleted after clicking the *Delete* button or pressing the key 'D'. *Clear All* will delete all measurement data in that mode. *View Data* will show a popup with all measurements that can be exported to *Plotly* or a CSV file.

¹For more information on CSV format, check out http://en.wikipedia.org/wiki/Comma-separated_values

8 Import and Export JSON

If you wish to save the axes calibration, data points in all the datasets and the measurements to a file for future use, then you can select the *Export JSON* option in the *File* menu. Note that this file does not contain the image and so you will also have to keep the plot image along with this JSON file. To use the information stored in the JSON file, simply use the *Import JSON* option in the *File* menu. The Export and Import feature can also be used to reuse axes calibrations across multiple plot images which are plotted on the same axes.