A Web-based Interface for Collaborative Multi-User Data

Analysis in a Scientific Research Environment

Nathan Knauf

The Harley School

Brighton, NY

Advisor: Richard Kidder

Laboratory for Laser Energetics

University of Rochester

Rochester, NY

January 2016

1. Abstract

A web-based interface was developed for scientific data analysis that allows data to be intuitively viewed and interpreted in a collaborative online environment. The interface improves the ability of scientists and engineers to analyze and compare results from different experiments and diagnostic types. It allows users to quickly view and investigate laser shot results without the need to manually download or manipulate data for specialized routines and to analyze shot results with both preexisting and user-written Python routines. The interface was demonstrated using data from experiments on the OMEGA and OMEGA EP lasers. The work also provides a framework for future networked utilities in which users both on and off site could collaboratively analyze data in web-based live sessions.

2. Introduction

At LLE, many diagnostic systems are in place on both the OMEGA and OMEGA EP laser systems for use during laser shots. Principal Investigators (PIs) require access to the data from these diagnostics to better understand the results of their experiments. Currently, there is a page on the LLE website, illustrated in **Figure 1**, through which PIs can search for laser shots on the basis of their shot identification numbers, view logistical information about the shots, and download data files from the different diagnostics used in the shots. The current page is not designed to pair PIs with data from their own experiments, instead relying entirely on a search function that finds shots on the basis of their identification number, so PIs must recover these numbers from physical records outside of the website in order to access their data. When PIs view the diagnostic data from a laser shot, they are presented with basic summary information and preview images of the data, but the PIs are ultimately limited in their usage of this data, because they cannot access any

information beyond a few summary statistics of each diagnostic, they cannot compare data between multiple diagnostics, and in some cases they cannot even preview the entire data set from a single diagnostic.

Shot	Date	Shot Type	RID	Campaign	Target ID(s)	# of PCUs	
78453	25-Aug-2015 13:48:38	Target High Yield	53879	LBS	ISE-3Q15-02-22	215	Download Images
78452	25-Aug-2015 12:27:14	Target High Yield	53878	LBS	ISE-3Q15-02-15	215	Download Images
78451	25-Aug-2015 11:35:20	Target High Yield	53867	LBS	ISE-4Q15-01-05	215	Download Images
78450	25-Aug-2015 11:23:54	Target Low Yield	50349	Other	PCRYO_ANY	0	Download Images
78449	25-Aug-2015 10:40:47	Target High Yield	53865	LBS	ISE-4Q15-01-04	215	Download Images
78448	25-Aug-2015 09:37:47	Target High Yield	53864	LBS	ISE-4Q15-01-02	215	Download Images
78447	25-Aug-2015 09:23:12	Non-Prop				1	Download Images
78446	25-Aug-2015 08:52:42	Target High Yield	53841	LBS	ISE-4Q15-01-01	215	Download Images
78445	25-Aug-2015 08:40:25	Non-Prop				0	Download Images
78444	25-Aug-2015 08:28:39	Non-Prop				213	Download Images

Figure 1: The current LLE search page, where PIs can look up laser shot records by their Shot ID numbers

In this research, a new, streamlined, and user-friendly interface was designed and developed for PIs to access and interpret all of their data on the website both intuitively and efficiently. Existing information from the records of laser shots about which PIs are associated with each shot was incorporated into the new web page so that when PIs use the web page, they are immediately presented with shots they are associated with so that they do not have to rely on external records of laser shot numbers to find their data. PIs may search for laser shots on the basis of which other PIs worked on them and also on the basis of shot identification number, just as in the original interface. The new interface implements new web-based technologies so that PIs can interact with the data and its component layers in a more thorough manner. Most importantly, the

new web page allows PIs to perform detailed analysis within their browser, with both a wide assortment of pre-made analysis routines and the ability to write and execute custom Python[1] code, which is of special value to both external and internal PIs who may need to work with their data from a computer that is not equipped with specialized analysis software.

3. Development

The layout and design of the new interface were made as intuitive and usable as possible, while remaining consistent with the aesthetic and organization of the rest of the LLE website. Work was done to implement a system that would run user-selected data analysis routines on data on the LLE servers, display both the data and analysis results on the user's computer in an efficient and convenient manner, and allow users to write their own custom Python analysis algorithms for a more adaptable analysis.

Python scripts were used on LLE's servers to process shot data according to user input independent of the computing power of the user's workstation. Thanks to its free, open-source, high-level, and easy to use nature, Python has established itself as a widely popular language among the scientific community. Since it is so universally known as well as structurally similar to many other common high-level languages (such as MatLab, Mathematica, Ruby, etc.), Python was determined to be the best suited for use in the software.

Different graphics libraries were considered to allow the shot data and analysis output to be viewed in interactive JavaScript widgets on the analysis page. After investigating Bokeh, MPLD3,[2] and a custom HTML5 framework, MPLD3 was chosen for its lightweight nature and smaller processing power need, which allowed it to perform well on users' possibly less powerful personal workstations or laptops as well as on specialized and powerful lab computers.

Different open-source Python notebook projects, such as Sage and Jupyter[3] (previously known as IPython), were investigated to allow users to write and execute code in Python inside of a web browser so as to run custom analysis routines on their shot data. Jupyter was determined to be the optimal notebook framework, as it supported over fifty other languages besides Python, had a much larger and more active development team, and had recently received a large research grant of $6,000,000 to continue its development.[3]
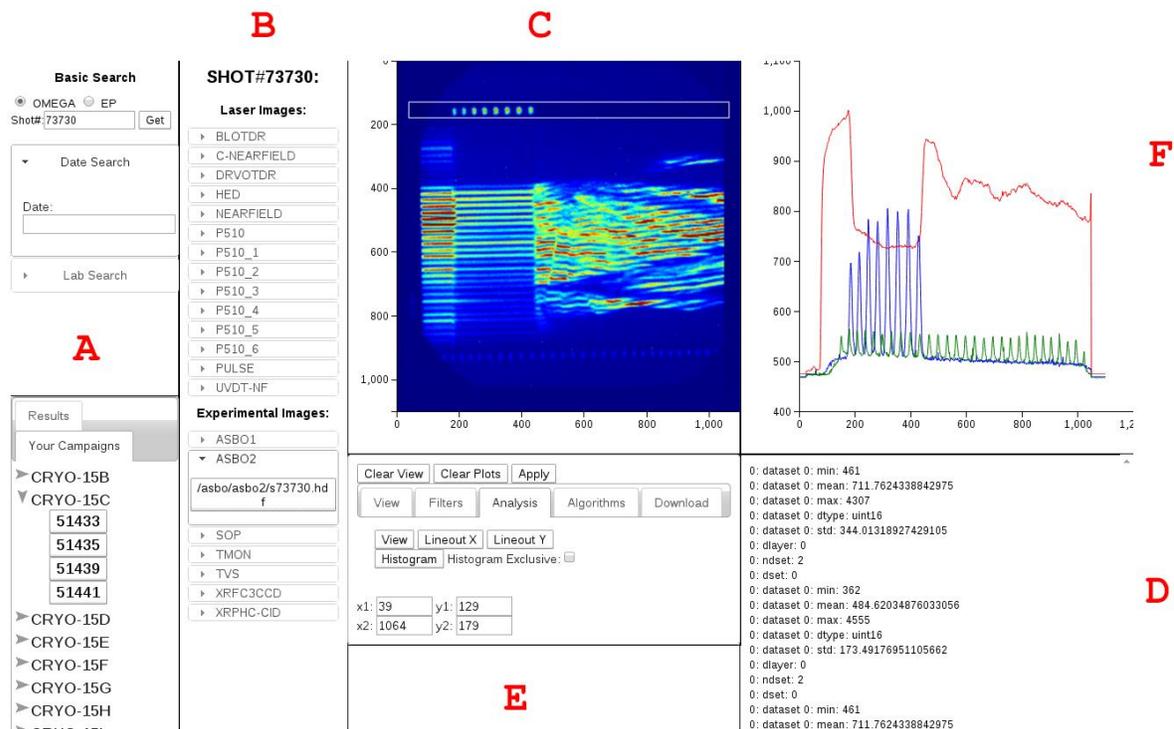

Figure 2: The Diagnostic Analysis page developed in this research, where PIs can search for, review, and analyze shot data

4. New Interface Capabilities

The new interface is centered on the new Diagnostic Analysis page shown in Figure 2. This page was designed so that PIs are immediately presented with data records from campaigns and laser shots they are associated with. PIs also have access to a search functionality, depicted in Section A of **Figure 2**, which allows them to search for different laser shots based on shot identification numbers, the dates the shots occurred, and the laboratories and researchers

associated with the shot. Once they have selected a laser shot, PIs are shown a list of all diagnostics utilized in the shot and all the data files created from the shot, depicted in Section B of **Figure 2**.

Upon the selection of any data file, the analysis page sends an asynchronous JavaScript (AJAX) request to the LLE web server, with identifying information about the diagnostic the PI wishes to view. After verifying that the user has permissions to view the diagnostic data, the web server pulls the data files (generally in .HDF and .DAT formats) from the shot records and uses Python and MPLD3 routines to create a .JSON file containing the diagnostic data and necessary information about how the data should be viewed, such as the type of the data, the dimensions of the dataset, or whether it should be interpreted as a line graph or an image. After the .JSON file is sent back to the client analysis page, it is interpreted by the MPLD3 JavaScript package to create an interactive JavaScript widget on the page, depicted in Section C of **Figure 2**, through which PIs can scale, transform, and view the data in detail. Summary information and details about the .JSON file are printed in a log depicted in Section D of **Figure 2**.

On the page, depicted in Section E of **Figure 2**, are a variety of settings for how the diagnostic data is to be viewed. PIs can manipulate these settings to view the data in different ways, such as by viewing different layers and datasets, by entering coordinates to designate which regions of the data to view, by changing the color map used to represent the data as images, by selecting the level of quality used to render data images, or by stretching, shrinking, or rotating data images. When PIs alter any of these view settings, new AJAX requests are sent from the analysis page to the web server, and the web server replies with .JSON files containing the appropriate selections and interpretations of data.

Alongside the view settings, also depicted in Section E of **Figure 2**, PIs have access to an array of analysis options for their data and are able to view the output from their chosen analysis

routines in a secondary graph widget, depicted in Section F of **Figure 2**, so that they can view their interpreted data in comparison with the original data. PIs can use the interactive JavaScript widget holding the diagnostic data to intuitively drag selection boxes around regions of interest in the diagnostic data, automatically calculating appropriate coordinates and designating these regions for analysis. PIs are then able to analyze their data in a variety of ways, by viewing histograms and distributions of the values of a data file, by exploring the differences in values between the layers and datasets of a multi-dimensional data file, and by producing specialized graphs of regions of interest such as images with data values normalized or averaged over the horizontal or vertical axes. Also in Section E of **Figure 2** are tools to download the images and plots of both the original and output data, with options for graph format and scale.

A significant advantage of this new interface is that PIs are able to view data from multiple diagnostics as well as from different shots simultaneously in order to quickly visualize the differences between diagnostic datasets. **Figure 3** depicts four datasets being plotted individually in the old interface on the left, but together in the new interface on the right, showing how the new interface can make data visualizations more meaningful by allowing PIs to view data in the context of other data.
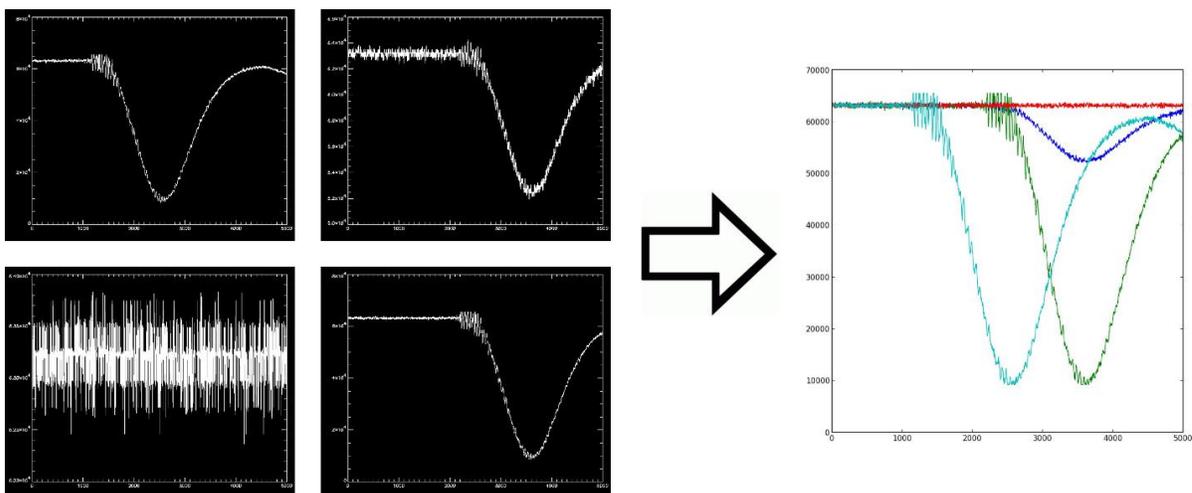

Figure 2: Multiple datasets viewed simultaneously in the new interface

As well as having access to a large number of predefined analysis routines, PIs now have the ability to write their own custom routines, in Python and other languages, through the newly implemented Jupyter notebooks on the LLE server. PIs are able to create notebook files in their own secure directories, where they can write and execute code in their browser to analyze diagnostic data in further depth and specialization than is possible in the analysis page, as well as use the text editing features of the notebook to accompany their mathematical interpretations of the data with written explanations of their methods. PIs may export their notebooks as HTML files or upload them to GitHub, an industry standard free source code repository, so that other researchers, scientists, and colleagues may view the PIs' work in a collaborative online environment, where they can assist PIs by sharing and discussing their experimental and analytical techniques.

Figure 4 shows an example of a Jupyter notebook being used to analyze shot data. **Figure 4** depicts a portion towards the end of a longer notebook of Python code that analyzes a shot data file from the LLE server, specifically from a P510 Diagnostic that records data in the form of an image of parallel streak lines. The full code breaks the image down into different regions containing the distinct streak lines and then plots distinct curves, one for each streak line, depicting the sum of values along the vertical columns of data in each region of the image. Through this kind of analysis, a PI might observe the differences between the streak lines, such as variations in intensities, lengths, or timings. Since code written in the Jupyter notebooks is processed by LLE servers, the use of Jupyter Notebooks allows PIs to use custom code to pursue detailed analysis of their data from their personal workstations or laptops, which might not have enough processing power to run the analysis code locally.
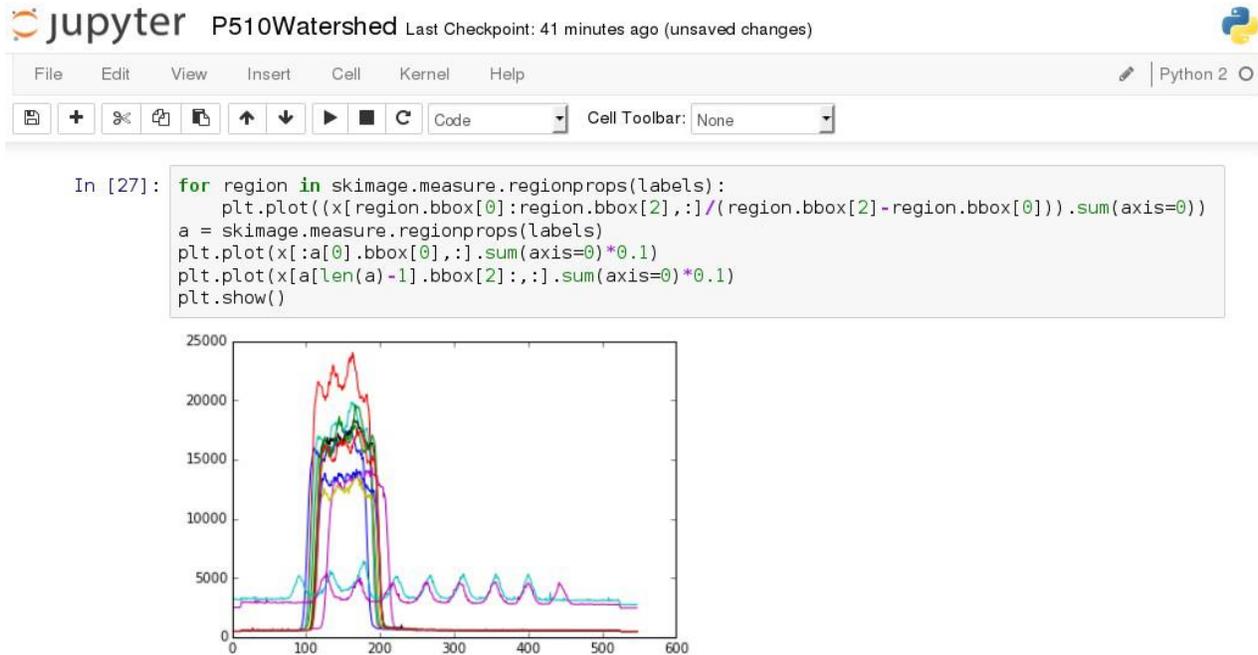
Figure 5: An example of a graph from a Jupyter Notebook being utilized to display lineouts from P150 streak data

5. Future Plans

In earlier work by Smith,[4] an online tool was researched and developed to view experimental and diagnostic configurations of the OMEGA laser system. This work demonstrated the feasibility and benefits of web-based live sessions, in which PIs can work on the same configuration simultaneously in a collaborative online environment. In the future, this research can be implemented into both the Diagnostic Analysis page and the Jupyter notebook pages to allow PIs from different laboratories to more effectively communicate and to collaboratively analyze data from laser shots in the OMEGA and OMEGA EP systems.

The analysis page can also be expanded to include further features, such as a custom graphic interface, the ability to directly compare data values between different diagnostic data files by subtracting or overlapping their data, and a larger selection of specialized analysis routines designed for specific diagnostics.

6. Conclusion

A new web-based interface has been developed to allow PIs to analyze diagnostic data from laser shots on the OMEGA and OMEGA EP laser systems. New analytical tools have made the workflow of PIs significantly more efficient by providing a streamlined and intuitive user interface. The new tools are advantageous to PIs because they allow diagnostic data to be viewed quickly and intuitively, they facilitate the comparison and interpretation of data from different diagnostics and shots, and they provide an online environment in which PIs can collaborate and easily share their findings. The new interface serves as a basis for future tools and software to allow LLE scientists as well as researchers from around the world to work efficiently and collaboratively in online environments.

7. Acknowledgements

This research could not have been accomplished without the assistance of Mr. Richard Kidder, my advisor, who guided the overall direction of the project while giving me the opportunity to research and develop the new analysis tools independently. I also received significant support from my coworkers, Andrew Zeller and Michael Charissis, who helped me to investigate the current technologies in use at LLE and to implement the newly developed tools into the existing LLE website. Without Dr. Stephen Craxton's hard work managing and advocating LLE's Summer Research Program for High School Juniors, none of this research would have been possible.

8. References

1. "Python" Python Software Foundation. Web. Jan 2016.

   https://www.python.org/

2. "MPLD3" MPLD3 Developers. Web. Jan 2016.

   hhttp://mpld3.github.io/

3. "Jupyter" Project Jupyter. Web. Jan 2016.

   http://jupyter.org/

4. Liam Smith, "Evaluation of a Collaborative Networking Environment for Experimental Configurations" Laboratory for Laser Energetics High School Summer Research Program (2014).