GRID ENABLING THE ASTRA GEMINI LASER DATA

E. Divall, K. Hayrapetyan, A. Kidd, D. Neely, M. Notley, Central Laser Facility,
V. Marshall, Corporate Information and Communications Technology,
S. Nagella, A. Pakhira, L. Sastry, e-Science Department,
Science & Technologies Facilities Council, Rutherford Appleton Laboratory, Harwell Science and Innovation Campus, Didcot, OX11 0QX, UK

M. Gleaves, L. Lerusse, e-Science Department, Science & Technologies Facilities Council, Daresbury Laboratory, Warrington Cheshire, WA4 4AD. UK

Abstract

Astra Gemini is a dual beam Ti:Sapphire laser capable of delivering up to 0.5 PetaWatt in each of its two beams [1]. The system can fire once every 20 seconds (producing over 1000 shots per day) and has over 200 diagnostic channels, including spectra, pulse length, traces, near and far-field images. This combination of multiple diagnostics and high shot rate leads to an unprecedented amount of performance and diagnostic data to save and analyse. To cope with this demand a system has been developed to automatically capture and analysis laser data on every shot, store it in an Oracle database and retrieve it on demand. A graphical user interface has been written to extract, sort and display the data in a tabular form. Powerful functions have been implemented to allow any parameters to be selected and plotted against one another to analyse performance trends and fluctuations. Metadata about each diagnostic can also be input to build a holistic picture of the laser system and help with future analysis. To increase the value it is planned to incorporate the target area experimental diagnostics into the system and make the data available to participating experimenters anywhere around the world [2].

INTRODUCTION

The eCLF project was started with the aim of Grid-enabling the data taken within the CLF. This was seen as an opportunity to add value to the existing data by making sure it is all easily accessible with powerful tools available for analysis and visualisation, rather than standalone applications hidden away on various diagnostics PC.

So far the project has primarily aimed at addressing data capture / visualisation on the Gemini Laser, with the Gemini target area, Vulcan and Target Fabrication areas due to follow when the Gemini system is fully up and running. It must be noted that due to the similarity of the data in the target areas and Vulcan, much of the foundation work made for Gemini would be applicable for all groups. (Gemini was chosen because at the start of the project there were several months expected before the start of laser operations and the diagnostic suites were still to be defined and could be tailored to match the needs of the eCLF project.)

PROJECT REQUIREMENTS AND DATA FLOW

Firstly the project requirements were assessed by a series of discussions with members of Gemini, Vulcan and Experimental Science groups and a set of requirements drawn up describing the data to be captured. This included:-

Laser: Configuration:- The way the laser is constructed, what hardware is used and how it fits together. This involves building an inventory of laser kit and a catalogue of how it fits together in such a way that the laser could be recreated in the same way in the future and anomalies in data diagnosed.

Laser: Low power end:- Many performance parameters are measured continuously throughout the day, i.e. front end laser performance, pressures, temperatures etc.. These need to be logged from the outset.

Laser: High power:- This is the actual laser data taken on a shot and is recorded by Gemini diagnostic system.

Data Storage:- It was agreed that the Atlas data store would be utilised for long term data storage and a CLF SRB would be set up.

Data Ingestion and Format:- It was decided to store the data using a Laser adaptation of the NeXus format. NeXus is a common data format for neutron, x-ray, and muon science. It is an international standard in order to facilitate greater cooperation in the analysis and visualization of Neutron, X-ray, and muon data.

Data Analysis / Viewer:- A data viewer was necessary to extract, view and perform basic analysis both locally and away from the lab, the key requirements were assessed.

Each of these sections is a project in its own right, but it is important that they bond together as smoothly as possible. i.e. the viewer should be linked into all the previous sections to extract and display date. To achieve this, a number of standard methods, ideas and data formats were evolved, e.g.:-

Terminology:- To describe the data a number of new terms were evolved. Each source of data was called a channel. These were subdivided into data streams containing the individual images, traces etc. that it produces. To avoid confusion each data stream was given a unique name.

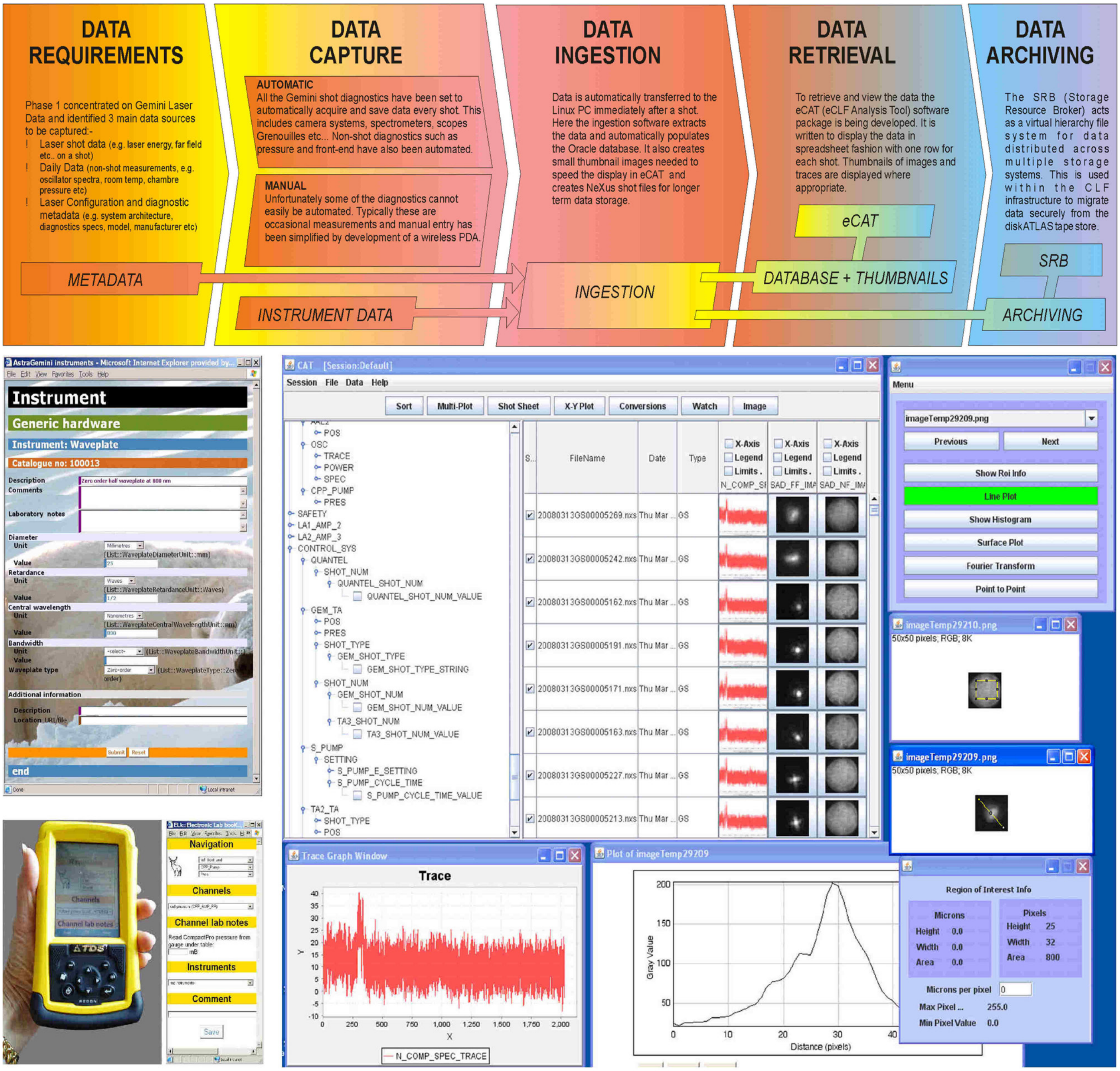

Figure 1: Top – diagram showing the process of data flow; middle left – metadata input form; lower left – PDA for mobile data entry; lower right – example of the eCAT data display.

Sections:- Over 200 potential data sources (channels) were identified. This required someway of categorising them rather than trawling through all 200. This was done by introducing a three level hierarchy (know as sections) by breaking down the channel's location in the laser chain and type of data it will produce. i.e. section LA3/S_COMP/ TRACE will hold a trace taken in LA3 on the south compressed beam. This data is entered through the Laser Configuration interface and saved to be used by both the ingestion and viewing software.

Diagnostic Data Format:- There are essentially three types of data that can taken from each diagnostic: images, traces and singular/string data. A standard format was agreed for each of these so it could be interpreted by the ingestion software.

DATA CAPTURE

Software has been written to capture data range of sources including: oscilloscopes, fibre optic energy monitor, Spectrometers, Rogowski coils, etc.. This included generic code to save the data in the correct format to be understood by the ingestion software. This was written in both VB and LabVIEW in such a way that it can easily be added to future programs.

As well as the data format it was decided to standardise on a directory structure to hold the data with a new directory being created every day. When data is saved a log file is also updated with the filenames of all the data saved that day. An automatic file transfer program then works its way through the log files copying all the data over to the Linux computer immediately after each shot. This program was standardised and run on all the diagnostic purchase.

A simple UDP receiver application was also written to receive singular value data / metadata and convert them to the correct format to be copied to the Linux database. This was especially useful in removing workload form the main control PC.

HARDWARE INFRASTRUCTURE

A Dell PowerEdge 1950II, dual, quad-core Linux server with 500Gb RAID 1 storage was purchased to host the raw and ingested data. The 500Gb is sufficient to hold several months of data, although a 2 Tb SCSI array is about to be installed.

It is configured with shared hard drives so that individual diagnostic PCs can copy data directly to a predefined directory for ingestion. Dual Gigabit Ethernet ports were also specified so that it can exist on both the private diagnostic and Lab networks simultaneously. The time server function was also implemented to ensure all the diagnostic PCs were correctly synchronised.

DATA INGESTION

By its very nature the CLF laser data is different from the ISIS and Diamond facilities which run as "continuous" sources rather than single shot machines. The software to ingest the data has been written specifically for the CLF to run in this "single" shot regime. Data from the diagnostic PCs is picked up and converted in to the standard NeXus format enabling standard ingestion / extraction routines to be used and reducing the work required for creating a custom format. The NeXus data files can then be examined using the freely available HDFView software.

The data is split up with one NeXus file being used for each shot thus all the data from each shot is kept together. Separate files are created for Daily (Low power) Data.

The ingestion software also extracts and stores singular values and paths to the images in an Oracle database. A “thumbnailer” program running in the background creates thumbnail images of all the image and trace data to speed up display to the user. It is intended to move as much of the visualisation / analysis onto this server as it is closest to the data and will help reduce network traffic.

eCAT (eCLF ANALYSIS TOOL)

Discussions with potential users regarding the data retrieval interface agreed that this would take the format of a spreadsheet with one column for each possible data stream. Individual data streams should be able to be selected (made visible) and plotted against one another, i.e. plot electron yield vs. energy, spectral width, pulse length etc.. Thumbnail images of traces and images should be shown on the spreadsheet with the ability to select images and analyse them for integrated intensity, spot size etc.. Continually measured parameters, e.g. pressures, temperatures would also be captured and displayed.

The first prototype of eCAT has been written and deployed for testing. It is written in Java making it portable across operating systems. To minimize processing thumbnail images are created for all the image and trace data which are downloaded for instant display in the spread sheet. Image analyse tools have been added and using public domain software such as ImageJ.

CONCLUSION

The eCAT system is proving to be a vital part of the data gathering and visualisation tools in Gemini operations and its future enhancement and deployment to other CLF lasers such as Vulcan, is anticipated in the near future.

REFERENCES

[1] http://www.clf.rl.ac.uk/reports/2007-2008/pdfs/s7/ar07-08_s7_commisioning_the_south_beam.pdf.

[2] http://www.clf.rl.ac.uk/reports/2007-2008/pdfs/s7/ar07-08_s7_eclf_project_progress.pdf.