eCLF project progress

E. J. Divall, K. Hayrapetyan, A. Kidd, D. Neely and M. Notley
Central Laser Facility, STFC, Rutherford Appleton Laboratory, HSIC, Didcot, Oxon OX11 0QX, UK

M. Gleaves and L. Lerusse
e-Science Department, STFC, Daresbury Laboratory, Warrington, Cheshire WA4 4AD, UK

V. Marshall
CICT, STFC, Rutherford Appleton Laboratory, HSIC, Didcot, Oxon OX11 0QX, UK

L. Sastry, A. Pakhira and S. Nagella
eScience Centre, STFC, Rutherford Appleton Laboratory, HSIC, Didcot, Oxon OX11 0QX, UK

Contact | **edwin.divall@stfc.ac.uk**

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.

The first year of project was primarily aimed at addressing data capture/visualisation from the Gemini Laser, with the Gemini target area, Vulcan and Target Fabrication due to follow on in years 2/3. However 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 there was several months until the laser was expected to be operational and the diagnostic suites still had 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 each laser group and the 'Data Management Requirements' [1] and 'Data Analysis Requirements' [2] documents written and published. These cover the whole experiment process from the initial experiment proposal and scheduling to the final analysis and publication as shown in figure 1.

With the deadline of Gemini's first shot less than a year away, it was decided to concentrate on the Laser and Data Analysis sections shown in figure 1. 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 [3] in such away 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 [4].
- 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 NeXμs format. NeXμs is a common data format for neutron, x-ray, and mμon science. It is an international standard in order to facilitate greater cooperation in the analysis and visualization of Neutron, X-ray, and mμon data [5].
- 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 [2].

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

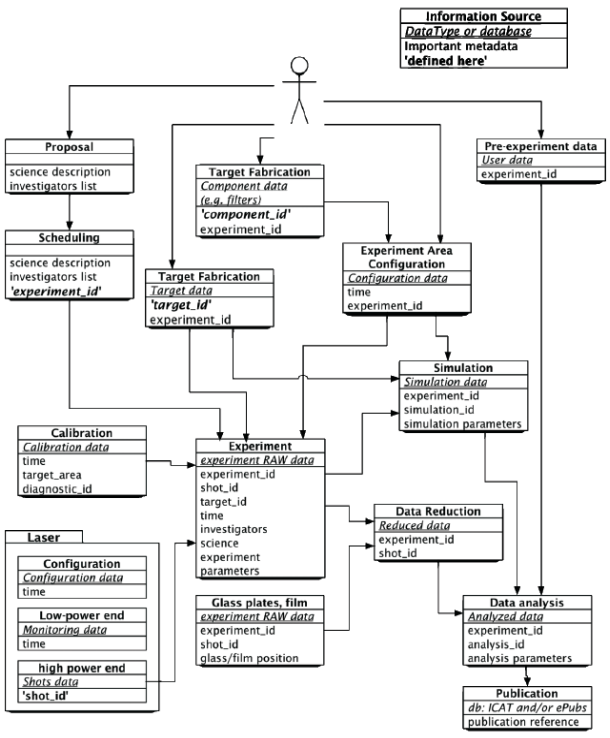

Figure 1. Class diagram for CLF data.

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 datastreams containing the individual images, traces etc that it produces. To avoid confusion each data stream was given a unique name.
- *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 format* – the NeXμs format was agreed upon as described above.

Project progress

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.

The ingestion software runs on this machine taking the raw data and creating the NeXμs files, see below. Also running in the background is a 'thumbnailer' creating small 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.

SRB (Storage Resource Broker)[9]

SRB acts as a virtual hierarchy file system for data distributed across multiple storage systems. This is used within the CLF infrastructure to migrate data securely from the disk ATLAS tape store.

A Dell PowerEdge R200 Windows 2003 server was also purchased to host the system catalogue and create a windows gateway between the two networks.

Data capture
Software has been written to capture data range of sources including: oscilloscopes, fibre optic energy monitor, Spectrometers, Rogowski coils, etc.[2,5] 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 away that it can easily be added to future programs.

As well as the data format was decided 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 PCs.

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.

Data ingestion
By its very nature the CLF laser data is different from ISIS and Diamond 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 NeXμs format enabling standard ingestion/extraction routines to be used and reducing the work required for creating a custom format. The NeXμs data files can then be examined using the freely available HDFView software[6].

The data is split up with one NeXμs 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.

System catalogue (Polar Bear Suite)
To more fully describe the laser system a system catalogue has been was created[3]. It builds up a picture of the possible data sources and equipment used to create Gemini. It also specifies exactly which scope, diode type etc are used in each data channel.

Alongside the catalogue an electronic logbook is being developed to replace the excel spread sheet and paper books around Astra[3]. Operators will be able to enter manual data settings, comments, set-up images from any part of the laser system through a webpage or wireless pad.

Both of these applications have a web based interface and can be used throughout the laser system.

User interface (eCLF analysis tool – eCAT)
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 datastream. Individual datastreams 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 [7]. 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.

Future work

A road map has been drawn up detailing the future work [8]. Key areas include role out across the facility and creation of an external server to provide external access and a full iCAT to allow metadata to be more easily browsed, see figure 2. The work also includes additions to the viewer to complete the basic analysis/functionality and allow data to be analysed in MatLab.

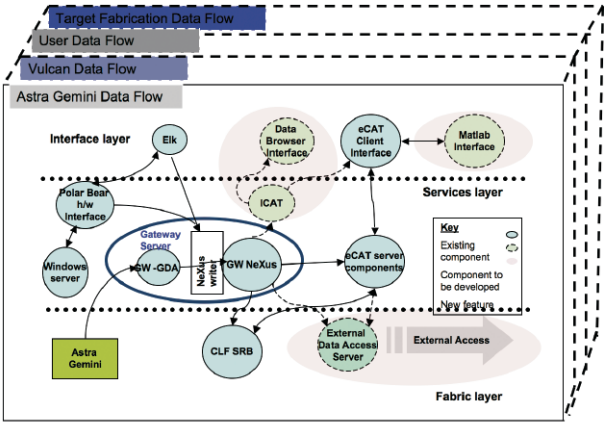

Figure 2. Overview of generic architecture for Astra Gemini – potentially to be extended across the facility.

References

1. Management requirements*
2. Data Analysis requirements*
3. Managing metadata for Astra Gemini. CLF Annual Report 2007-8, p216.
4. Gemini Diagnostics. CLF Annual Report 2007-8, p223.
5. eScience-CLF Data Acquisition System. CLF Annual Report 2007-8, p236.
6. http://hdf.ncsa.uiuc.edu/hdf-java-html/hdfview/
7. Laser Performance Data Analysis Tool. CLF Annual Report 2007-8, p220.
8. CLF Road Map*
9. http://www.sdsc.edu/srb/index.php/Main_Page

*Repository:- http://www.e-science.stfc.ac.uk/projects/CLFeInfrastructure/