

Managing metadata for Astra Gemini

V. A. Marshall

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

Contact | victoria.marshall@stfc.ac.uk

Introduction

For any system as complex as a high-power laser, there are considerable challenges in capturing the raw data, curating it for use by the scientists and laser operators, and providing the relevant metadata to enable more informed interpretation of the results in the lab or back at the office.

XML instrument catalogues

Interpretation of data has always depended upon a deeper understanding of the factors surrounding its generation. For the Astra Gemini project we started by creating an XML catalogue of the diagnostic instruments used.

Instruments were divided into two categories: catalogueable or generic. Catalogueable instruments were given a Gemini catalogue number and included Pockels cells, Grenouille (pulse-measurement devices) and spectrometers – distinct items, usually electronic, the configuration of which may have significant bearing on the experiment. Generic instruments included lenses, microscope objectives and polarisers – items of more general applicability and mutual interchangeability. Each instrument type was represented by an XML file capturing the different features of that type such as manufacturer and physical characteristics of a Grenouille, the diameter and retardance of a polariser, or objective diameter and focal length of a telescope.

Each instrument is represented by an instance XML file; Astra Gemini is however an on-going project with, inevitably, a very fluid data configuration. It was important that the XML reflected the on-going situation in the project without requiring the laser operators to continually hand-edit XML files to keep up. We therefore developed a system of ‘master’ XML templates for each type of instrument (also Assemblages and Channels; see later). The appropriate template plus instance XML file is used to generate a web-based user interface, see figure 1, which allows the operators to modify the parameters of the object then write them back to the instance XML file, but with the master’s structure. In effect, the editing suite allows the XML to be ‘self-modifying’ over time. A separate interface file (also in XML) provided an extensive number of named drop-down selection lists which could be used to constrain the values given as appropriate. For example:

```
<select_list_definition SelectListName="ShotType">
<option value="" display="-select-" />
<option value="GD" display="Gemini daily" />
<option value="GR" display="Gemini register" />
<option value="GS" display="Gemini shot" />
<option value="GT" display="Gemini target area" />
<option value="QS" display="Quantel shot" />
```

E. J. Divall

Central Laser Facility, STFC, Rutherford Appleton
Laboratory, HSIC, Didcot, Oxon OX11 0QX, UK

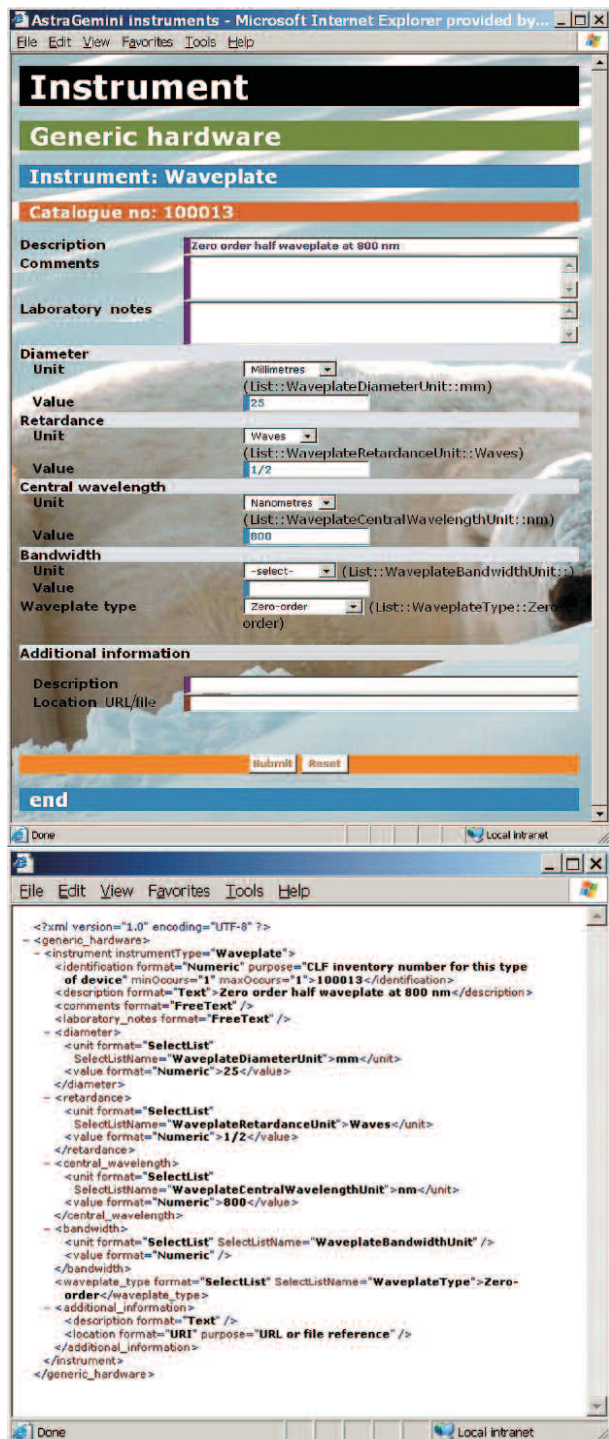


Figure 1. Screen dump of the interface to modify the parameters of a waveplate, and associated XML file.

```
<option value="PB" display="Polar bear configuration"/>
</select_list_definition>
```

Most of the electronic instruments are programmed to produce a stream of raw readings and calibration data which are written to a file and then copied to a central server. However, in order to interpret this data it is

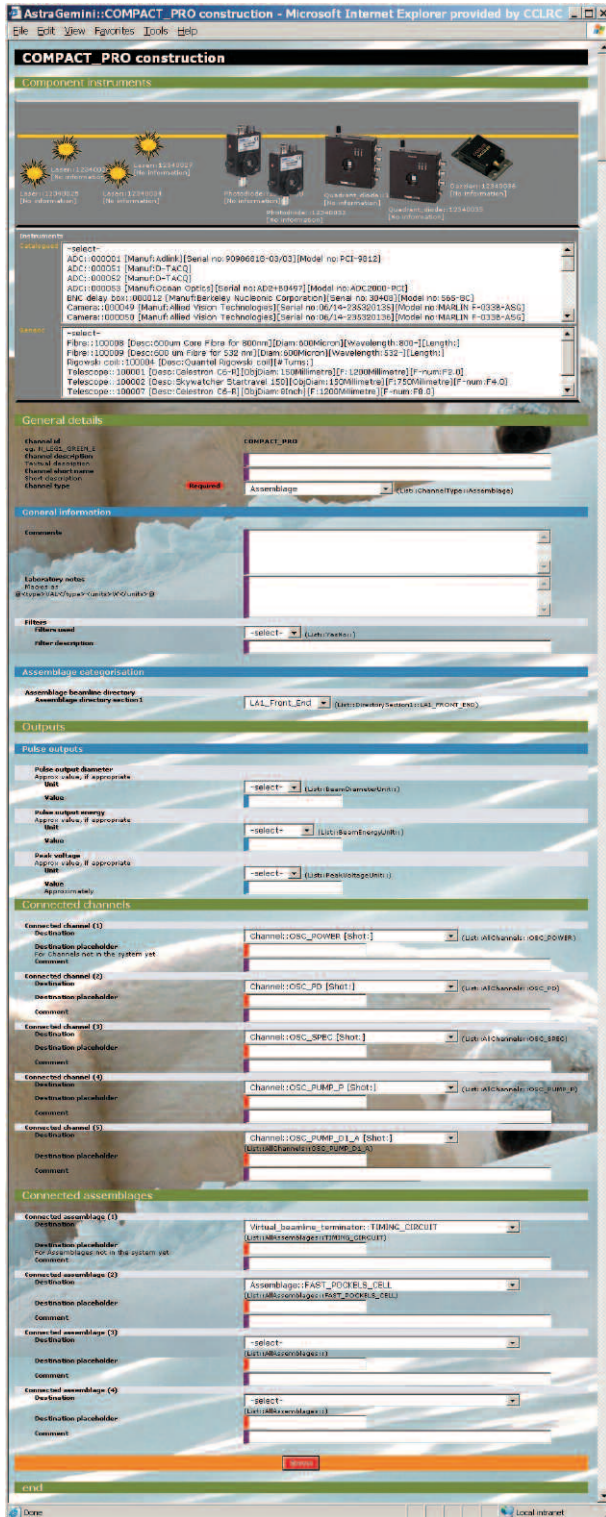


Figure 2. The XML editing suite for a Channel showing the instruments available, and the other Channels to which this Channel is linked.

important to understand how the instruments are used, which in turn depends on their purpose and the section of Gemini beamline in which they are situated; for example, a photodiode/oscilloscope combination could be used for timing, energy measurement or pre-pulse monitoring depending where they are situated.

XML definition of the logical structure of the beamline

This led to the need for a representation of the physical construction of the beamline in terms of discrete chunks: Assemblages and Channels. Assemblages form the backbone of the laser chain and contain the instruments used in its construction. The Channels represent the laser diagnostics along the laser chain. They are split off from the assemblages and contain the instruments to create that particular diagnostic. (The data produced from each channel was in turn called its datastreams i.e. a camera diagnostic could produce image and spot position datastreams.)

A similar approach was used to that described above, see figure 2, showing the interface to define the instruments and settings for each Channel and Assemblage. The resulting XML file details:

- The channel's ID and name (eg. AMP3_VAC_PR, Relay pressure)
- The section in which it occurs (eg. LA2_AMP_3/ VACUUM/PRES)
- Data collection method (manual, automatic, N/A)
- ADC and oscilloscope sockets used (if appropriate)
- Timing outputs and delays (in the timing circuit)
- Channels/Assemblages to which this is linked (if this is an Assemblage)
- Datastream names and type (if this is a Channel) (eg. AMP3_VAC_PR_VALUE)
- The diagnostic instruments used

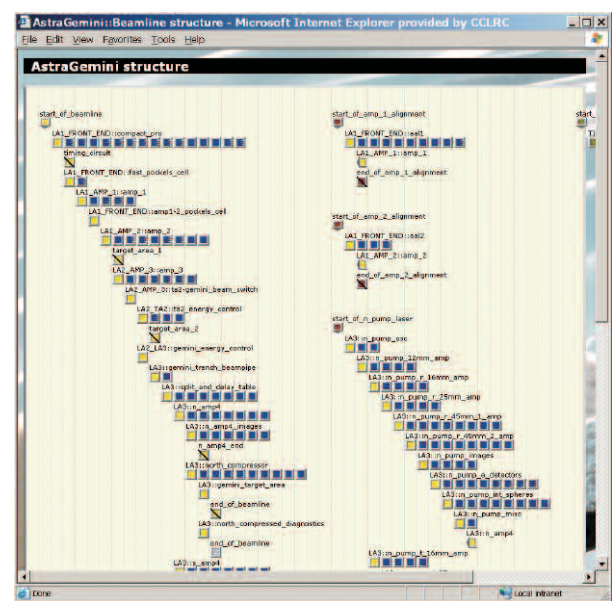


Figure 3. Schematic of the Astra Gemini project, generated directly from the beamline specification XML. The yellow squares represent assemblages and blue ones their data channels.



Figure 4. Actual PDA in use.

Some virtual Assemblages were also created to define the start and end of the beamline, its subassemblies and the timing circuit. By processing all of the Channel files it is therefore possible to compile an XML definition of the complete beamline in terms of what links to what, which subassemblies feed into it at which point, and which instruments are used where (or not used at all, or apparently used twice) see figure 3. Thus if each instrument 'knows' its datastream ID, it is possible to identify the context in which each stream of data is produced.

On the 'shop floor'

The XML metadata is entered by the user via a standard web browser on a desktop PC. There are occasions however on which modifications need to be made to the beamline on-the-spot, and these are done with a handheld PDA (figure 4). For example, most of the diagnostic instruments generate data automatically, but some require manual recording of their settings and configuration. E.g. one of the primary uses of the PDA is to manually record settings from any of the stand-alone diagnostics. Using the web based menu it is possible to select a datastream and enter any manual reading.

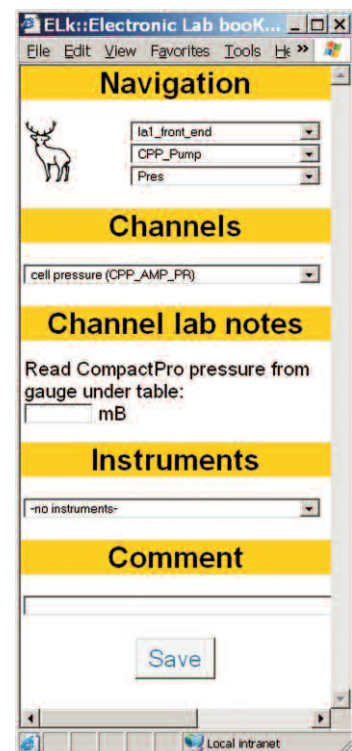
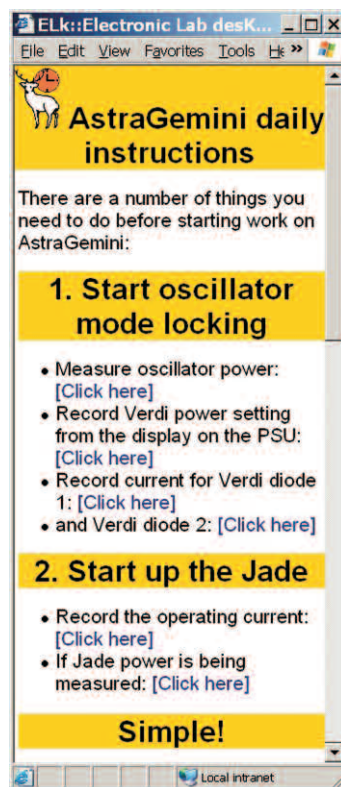
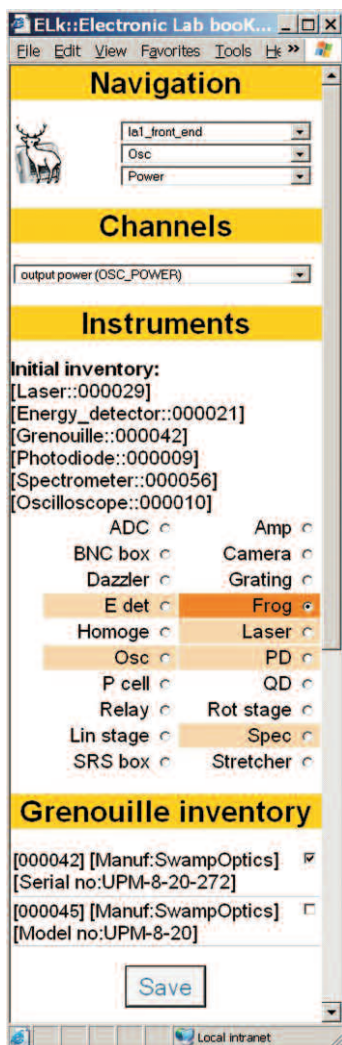


Figure 5 (left). Simulated PDA screen showing the instrument inventory for the Channel OSC_POWER

Figure 6 (middle). Simulated PDA screen of daily start-up instructions.

Figure 7 (right). Lab notes setup to record CompactPro pressure.

It is also likely that during the lifetime of the project, some instruments will require replacement, and this too needs to be captured in the experiment's metadata. Laser operators on the 'shop floor' already use ruggedized PDAs on a private WiFi network to enable them to control the laser shots, so it made sense to re-utilise these (figures 4 and 5). There were a number of issues to address in doing so however: the limited screen size and lack of keyboard are the most obvious issues, see figures 5, 6, and 7 for screen layouts. Also because the PDAs would often be used when the laser was present, the operators would be wearing amber safety goggles which presented severe restrictions on the use of colour and font size.

a. Instrument inventory

The PDA software uses the beamline specification XML to present a highly-tailored interface to enable the operators to walk round the beamline and catalogue the instruments used within each section of it. Once this has been done, the base metadata can be updated and then uploaded to the central server for ingestion into the experiment's NeXus files.

b. Daily start-up

Because a web browser is used to present information on the PDA, it is very easy to set up a series of HTML pages detailing the various maintenance procedures for Astra Gemini. See figures 6 and 7 showing list of manual readings that should be taken when the laser is set up. The procedures include links to the relevant beamline section and aide-mémoire lab notes to remind the operator what needs to be done and providing them with the means to enter the values or upload the relevant file.

c. Ad hoc events

The PDAs can also be used to record events such as instrument replacements or when the laser was switched on/off. These events are stored in a database for longer-term future use such as statistics on equipment reliability, or service performance metrics.

Footnote: The metadata part of this work has become known within the project as the 'polar bear suite' after the non-laser person involved affectionately (if inaccurately) referred to 'polarisers' as 'polar bears'. This whimsical theme was continued onto the PDAs, which have become known as 'ELks' or 'Electronic Lab books' in addition to 'pads'.