

MAPPING DEVELOPMENTS AT DIAMOND

R. Walton, A.W. Ashton, M. Basham, P.C.Y. Chang, T. Cobb, S da Graca, A.J. Dent, J. Filik
M. Gerring, C. Mita, J. Mudd, C.M. Palmer, U. Pedersen, P.D. Quinn, N. Rees
Diamond Light Source, Oxfordshire, UK

Abstract

Many synchrotron beamlines offer some form of continuous scanning for either energy scanning or sample mapping. However, this is normally done on an ad-hoc beamline by beamline basis. Diamond has recently embarked on an ambitious project to define how to implement continuous scanning as the standard way of doing virtually all mapping tasks on beamlines. The project is split into four main areas: 1) User interfaces to describe the mapping process in a scientifically relevant way, generating a scan description that can be used later; 2) The physical process of scanning and coordinating hardware motion and detector data capture across the beamline; 3) Capture of the detector data and all the associated meta-data to disk, deciding and describing the layout of the file (or files) for the main use cases; 4) Display and analysis of live data and display of processed data. In order to achieve this common approach across beamlines, the standard software used throughout the facility (Delta Tau motor controllers, EPICS, GDA and DAWN), has been built on.

INTRODUCTION

With the increase of flux obtained with most modern 3rd generation synchrotrons, collecting maps of surfaces is becoming increasingly popular. This process involves moving a sample in front of a focussed x-ray beam, then recording various interactions of this beam with the illuminated part of the sample. Interactions include, but are not limited to, fluorescence, diffraction, absorption and scattering, with each requiring specific detectors to be available on a beamline. Scans are often performed continuously as detectors are triggered and maps containing millions of individual measurements are routinely collected at specialist beamlines. The high flux available means that these scans can be performed increasingly quickly, providing significant technical challenges in all areas, from motion control and detector readout to processing and visualisation.

At Diamond Light Source there are many beamlines which can perform mapping experiments. Dealing with these challenges on a case by case basis has become unmanageable. At the time of writing we are six months into a cross beamline, cross support group project to standardise and modularise as many elements of this process as possible. As well as providing improved functionality and user experience on these beamlines a key aim is to make the solution supportable with as little effort as possible. Many aspects of the solution are applicable to techniques other than mapping (a line scan

is just single row map scan) and we expect the impact of this project over the coming years to reach many of Diamond's beamlines.

To give the best chance of producing a general solution, the project targeted five beamlines with similar but unique mapping requirements. These include traditional XRF (X-ray Fluorescence) mapping [1], XRD (X-Ray Diffraction) mapping, Absorption mapping, Coherent Diffraction mapping [2] and ARPES (Angle Resolved Photoemission Spectroscopy) mapping beamlines.

At the heart of this project is a commitment to standardise the way data is handled with a new standard data flow, standardise the way continuous scans are controlled with a new EPICS control layer, and to improve the user interface and speed of collection. The remainder of this paper highlights these high level design decisions before presenting further details of the system and reporting progress.

STANDARD DATA FLOW

To simplify post processing and analysis it is important that the multidimensional and in some cases multimodal data, along with all the appropriate metadata such as experimental conditions, are stored in a standard format. Diamond uses the NeXus format [3] backed by HDF5 [4]. HDF5 performs very well, especially with the Lustre and GPFS file systems used at Diamond. Unfortunately, HDF5 cannot be read as it is written and so in order to provide live data processing and visualisation separate data paths have been maintained. To simplify this situation the HDF group was funded by Diamond, ESRF and Dectris to extend HDF5 to allow access to data as it is written. This extension is known as SWMR which stands for Single Write Multiple Read [5].

The access to live descriptive data SWMR affords greatly streamlines processing and visualisation as the same software used offline can, with some work, be used online. Further, because data can be read almost anywhere, it can be processed using tools running on a cluster which are available for post processing, allowing for sophisticated, and computationally expensive processing to be achieved in close to real time. Processed data will also be written with SWMR, providing the visualisation software live access to both processed and raw data.

MALCOLM EPICS CONTROL LAYER

The GDA data acquisition software has good support for dynamically conducting arbitrary scans, but when performance is important, such as with mapping

experiments, hardware control must be pushed down to the motion controller. Historically this has been done with a variety of solutions, but Diamond is converging on a solution that supplements Delta Tau motion controllers with a Zebra triggering box [6, 7]. Currently Diamond's GDA server [8] communicates with this range of hardware directly or via EPICS [9]. Although these solutions work they are expensive to maintain, and not that scalable.

To take best advantage of the Delta Tau with Zebra solution, a new EPICS control layer between the high level acquisition server and the low level EPICS software has been created. Malcolm aims to abstract some of the complexity involved with controlling multi-purpose EPICS drivers into simple devices that have parameters that would make sense to an end user [10]. For instance, an AreaDetector chain of 1 driver and 7 plugins controlling a detector would be modelled in Malcolm as 8 low level devices, and one top level device that captures the attributes that are relevant for a particular mode of operation, and orchestrates the low level devices. A set of motors, a trigger box, and a spiral scan would all be Malcolm devices, all of which implement a subset of the same configure-run state machine shown in Figure 1. This allows nested scans to be easily created, as all devices obey the same interface.

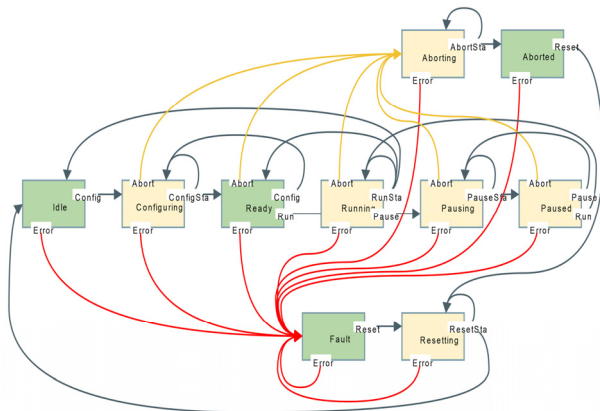


Figure 1: Malcom state machine.

The prototype system is written in Python and uses a JSON presentation layer over ZeroMQ or websocket transport. At the time of writing, work is progressing on device definitions for Diamond's I05 ARPES branchline.

When the architecture is proven, pvData over pvAccess will be added to the Python server, and a C++ port written to allow AreaDetector drivers and plugins to provide this interface natively. The hoped final outcomes are EPICS V4 transport and presentation layers, and device definitions for a large range of detectors, trigger boxes, motors, and complex scan types.

STEPS DURING AN EXPERIMENT

The software will help the user explore a sample's surface interactively by building up and visualising data from a number of scans and microscope images. Each

scan might use different techniques, detectors or processing, but all will be visualised in the same framework. Figure 2 shows an example client.

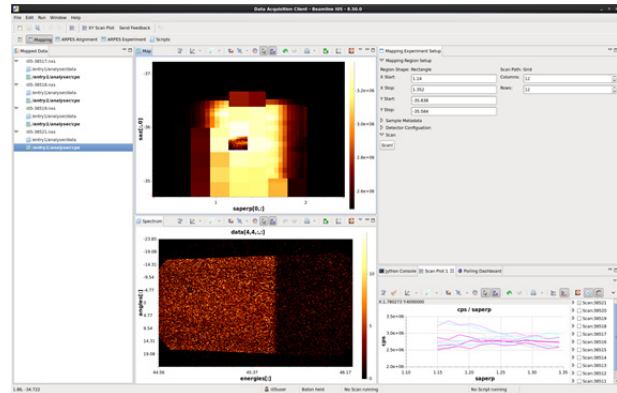


Figure 2: GDA client showing the new *mapped data view* and experiment setup pane on the I05 ARPES beamline. The *Mapped Data View* shows all data which is available for plotting, in this case 4 different resolution maps of the sample. The *Map View* shows all available maps overlaid, and new regions can be selected for scanning visually in this section. The *Spectrum View* shows the spectrum or image collected at the selected point on the Map View. Finally, the *Mapping Experiment Setup View* is an auto-generated view based on the experimental requirements of the beamline.

In most cases a user will start a session by taking a visual image of the surface of a newly mounted sample. This image will then form the basis for defining regions to scan over. A scan comprises a series of events, most of which are performed concurrently as data is collected. These events, with the component responsible for each indicated in *italic*, are:

1. **User defines scan parameters.** The user uses the *scan parameter editor* in concert with the *mapped data views* to describe a scan including detectors and, in many cases, a region over which to map and then requests that the *acquisition server* orchestrates a scan. Scans can also be performed using a command line interface suitable for scripting.
2. **Trigger and coordinate scan.** Having received (a possibly queued) request to perform a scan the *acquisition server* component connects to the *processing* and mapped data view components to inform them a scan will start and to await progress updates. It then starts the collection. The acquisition server is built into Diamond's GDA software, but could be easily deployed within other Java products.
3. **Sequence motion and trigger detectors.** Simple step scans will be sequenced by the acquisition server directly. Scans requiring continuous motion will use

the new *Malcom middle layer* software described above which provides a layer above *Epics AreaDetector*, Diamond's *Zebra* box used for triggering, and a motor controller.

4. **Record data and metadata to file.** The acquisition server writes a NeXus file that contains links to the HDF5 files written by *Epics AreaDetector* and *Zebra*:
 - a. The acquisition server's *NeXus file writing* component writes the NeXus file using SWMR compliant HDF5 libraries.
 - b. *Epics AreaDetector* and *Zebra* must write data using SWMR in any arbitrary order given knowledge of the scan's shape and path.
 - c. *Remote Dataset* components make possibly down-sampled slices of data from disk available offsite and from workstations which, because SWMR works only across POSIX compatible file systems, cannot access the data directly.
5. **Process data.** The live processing component optionally creates a new NeXus file to store processed data; this will include a link to the original NeXus file containing only the raw data.
6. **Visualise data.** The user will see the raw data, and optionally the processed data, appearing live in the mapped data view. This is the same view available offline in Diamond's DAWN software and provides tools for interactive analysis.

COMPONENT DETAILS AND PROGRESS

At the start of the project a feasibility study showed that SWMR would work well in this. The HDF Group is currently in advanced stages of merging and testing SWMR along with other new features into the trunk of the HDF5 code base and appears to be on schedule for the HDF5 1.10 release by spring 2016. Despite the HDF5 SWMR feature only being available in a beta prototype form, HDF5 language bindings for Python (h5py) have already been extended to support this new feature.

Details of the components introduced above follow.

Mapped Data Views

The large and complex data produced by mapping experiments will be investigated and visualised with a bespoke user interface (a perspective in Eclipse RCP technology). This user interface, an example of which is shown in Figure 2, allows multiple maps, with raw and processed data, from multiple detectors, to be viewed at the same time. Due to the modular structure of RCP the same user interface will be available in DAWN [11] for offline analysis, and in a GDA client, for live data visualisation and experimental configuration.

Scan Parameter Editor

The GDA client will include an easy-to-use GUI which allows a user to draw a region on a map or camera image, set up other experimental parameters as required, and then trigger a new map scan of the defined region. When a region is drawn on a map the region coordinates are added to a Java bean. A simple auto-generated GUI for the bean allows other fields to be edited, and there is a scan button which passes the bean to the acquisition server to trigger a scan. The GUI is auto-generated using the Metawidget framework.

Epics AreaDetector & Zebra

Epics AreaDetector and *Zebra* write data that is linked to by the scan's NeXus file. Both must be updated to write SWMR compatible HDF5 files to allow live access to their data. Both are also being updated to write data frames into multidimensional datasets in arbitrary orders.

This latter feature will provide support for complex or more efficient scan mechanisms, from bi-directional raster scans to spiral or even random sampling scans. During these scans data must be written in multiple dimensions to match the dimensionality and path of the scan. For example, during a bi-directional mapping scan where each second line is reversed, *AreaDetector* must write detector data as nodes in a two dimensional dataset in the order that mirrors the path the x-ray beam traverses across the surface.

To achieve this, *AreaDetector* attaches the global position of each frame in a scan as meta-data immediately after the frame has been acquired from the detector. The file-writer then makes use of this meta-data when storing the frame in the multi-dimensional dataset. At the time of writing, this work is well advanced and has been rolled out to a non-operational beamline for beta-testing.

NeXus File Writing

At the time of writing, the GDA acquisition server includes an early implementation of the new SWMR features built using the Java bindings for the HDF5 library.

Remote Dataset Component

Remote Dataset is a technology available at Diamond which allows data situated on a remote disk to be accessed by a client. It exposes to the remote client something called a *Lazy Dataset* with which data can be optionally sliced and optionally down sampled then returned to be visualized on the client. It also allows MJPEG streams to be connected and plotted live in the plotting system with functioning visual tools on the remote live data. *Remote Dataset* works in Eclipse RCP using an OSGi service [12] but it also works with non-Java clients because the data is requested and returned by web sockets.

Live Processing Component

With all the data available through SWMR, it is possible for processing to be performed by any software

ISBN 978-3-95450-148-9

supporting this new protocol. As standard though the DAWN processing system will be used, as it has been used successfully for post processing in many fields such as diffraction imaging and ARPES analysis, integrates with GDA, and can easily be extended to make full use of SWMR.

Acquisition Server and Inter-process Communication

The acquisition server coordinates the actions of the Malcolm, processing and client-side processes during a scan. Figure 3 shows an early design of how these processes will fit together.

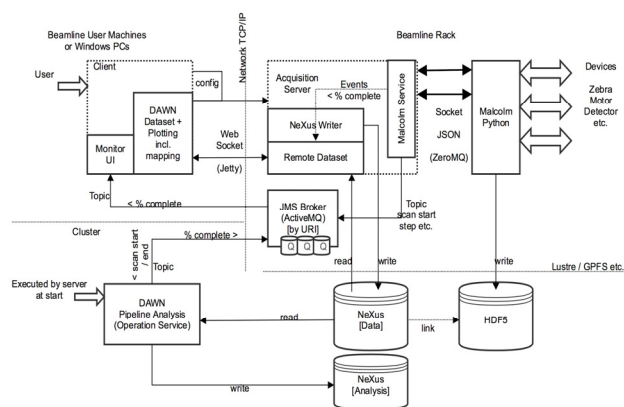


Figure 3: An early design showing the acquisition server at the middle of the system.

CONCLUSION

Diamond has targeted five beamlines with similar but unique mapping requirements. This variety, combined with a coordinated approach from the centralised controls, acquisition and analysis groups should ensure the result is modular, maintainable and expandable. The prototype Malcolm EPICS layer provides a promising API for continuous scanning; and the use of HDF5 SWMR to streamline data flow simplifies many components in the system and has almost removed the line between live and post processing and visualisation. The project is progressing well and it likely that over the coming years its impact will spread wider than its initial target of five beamlines.

AUTHOR CONTRIBUTIONS

RW, MB, AA, PCYC, TC, JF, MG, CM, JM, CP, UP and SG collaboratively developed the software packages involved. RW and MB jointly prepared the text for this article, with contributions from TC, JF, MG, CMP and UP. RW, AWA, AJD and NR directed the development effort and the writing of this article, and PQ represented the beamline staff's requirements and use cases.

REFERENCES

- [1] Mosselmans, J. Frederick W., Paul D. Quinn, Andrew J. Dent, Stuart A. Cavill, Sofia Diaz Moreno, Andrew Peach, Peter J. Leicester et al. "I18-the microfocus spectroscopy beamline at the Diamond Light Source." *Journal of synchrotron radiation* 16, no. 6 (2009), 818-824.
- [2] Rau, Christoph, Ulrich Wagner, Zoran Pešić, and Alberto De Fanis. "Coherent imaging at the Diamond beamline I13." *physica status solidi (a)* 208, no. 11 (2011), 2522-2525.
- [3] Könecke, Mark, Frederick A. Akeroyd, Herbert J. Bernstein, Aaron S. Brewster, Stuart I. Campbell, Björn Clausen, Stephen Cottrell et al. "The NeXus data format." *Journal of applied crystallography* 48, no. 1 (2015), 301-305.
- [4] The HDF Group. Hierarchical Data Format, version 5, 1997-2015. <http://www.hdfgroup.org/HDF5/>
- [5] HDF5 SWMR, <https://www.hdfgroup.org/HDF5/docNewFeatures/NewFeaturesSwmrDocs.html>
- [6] T. Cobb, "Zebra: A flexible solution for controlling scanning experiments", ICALEPCS 2013, San Francisco, USA.
- [7] Malcolm <http://malcolm.readthedocs.org>
- [8] Generic Data Acquisition, www.opengda.org
- [9] Dalesio, L. R., Kraimer, M.R., Kozubal, A. J., "EPICS Architecture," in *Proceedings of International Conference on Accelerator and Large Experimental Physics Control Systems*, C. O. Pac, S. Kurokawa and T. Katoh, Eds. (ICALEPCS, KEK, Tsukuba, Japan, 1991), pp. 278-282.
- [10] I.S. Uzun, et al, "PandA Motion Project - A Collaboration Between SOLEIL and Diamond to Upgrade Their Position and Acquisition Processing Platform", *Proc. of ICALEPCS'2015, Melbourne, Australia*.
- [11] Basham, M., Filik, J., Wharmby, M. T., Chang, P. C., El Kassaby, B., Gerring, M., & Ashton, A. W. (2015). "Data Analysis Workbench (DAWN)". *Journal of synchrotron radiation*, 22(3), 0-0.
- [12] M. Gerring, "Open Source Contributions and Using OSGi Bundles at Diamond Light Source", ICALEPCS 2015, Melbourne, Australia.