Management of Metadata and Automation for Mail-In Measurements with the APS 11-BM High-Throughput, High-Resolution Synchrotron Powder Diffractometer

Brian H. Toby[1], Yu Huang[1], Don Dohan[2], David Carroll[1,3], Xuesong Jiao[1], Lynn Ribaud[1], Jennifer A. Doebbler[1], Matthew R. Suchomel[1], Jun Wang[4], Curt Preissner[1], David Kline[1], and Tim M. Mooney[1]

1) Advanced Photon Source, Argonne National Laboratory, Argonne, IL 60439 (USA).
2) National Synchrotron Light Source-II, Brookhaven National Laboratory, Upton NY 11973 (USA).
3) Maplebrook Elementary School and Madison Junior High School, District 203, Naperville IL 60540 (USA).
4) National Synchrotron Light Source, Brookhaven National Laboratory, Upton NY 11973 (USA).

Supplementary materials

Document S1: Workflow for collection of data on the 11-BM instrument

The typical actions performed by users and beamline staff, from the submission of proposals, to obtaining access, through to data collection, as well as subsequent steps, is described here. Diagrams S1 and S2 show this workflow graphically.

Proposal Submission

Users submit requests to utilize 11-BM using the standard APS (GUP) submission system. Users may submit beam-time requests on a standard or rapid-access basis. Rapid-access allows users to request time in the current operational cycle, and thus data collection occurs in advance of proposal review; local policy restricts rapid access to a single 8-hour shift. Rapid-access requests are reviewed by beamline staff for feasibility and are usually accepted. The GUP system automatically sends users an e-mail message describing the rest of the access process and provides a link to a web address where bases for samples can be requested.

Sample Base Request

The 11-BM staff provide users with sample bases for their measurements. In this way, every sample base is associated with a user. Should an unidentified material arrive at the beamline, at least the source can be determined.

The sample-base request process works in several quick steps.

1. The user specifies a number of sample bases for their experiment via a web form. The only input at this stage is a GUP number, the quantity of requested sample bases, and an e-mail address, which should match one of the e-mail addresses on

the proposal. This information is stored by the external webserver into the **proposal_request** table of the local 11-BM database.

2. Twice an hour an internal server automatically checks for new sample base requests. The 11-BM software queries the GUP database for the proposal title, participant names and e-mail addresses, and the amount of beam-time allocated. It verifies that the number of bases requested is consistent with the amount of time that has been allocated, less what has already been used. An e-mail is then automatically sent out to the user's address(es) asking them to confirm certain information in the proposal, for example that the mailing address provided in the proposal is appropriate for shipment of the sample bases. This step helps insure that sample bases are sent to mailing addresses the user intends. The e-mail contains a unique URL to be used to confirm the request (see below). Should the user provide a different e-mail address in the sample base request than one(s) included in the proposal, an additional message is sent to that address to notify the user of the e-mail addresses that are used in the proposal, but this e-mail does not provide the confirmation URL. In this way a user is notified of potential incorrect e-mail addresses submitted in their proposal.

3. The user confirms the receipt of this e-mail by clicking on or copying to a browser the unique URL provided in the e-mail. When this URL is activated, the record in the **proposal_request** table is updated to indicate that the request is confirmed. Subsequent uses of this unique URL are ignored. The beamline staff are notified of the confirmed request via e-mail. If the proposal information must be updated, the user contacts the APS User Office with corrections and then

submits a new sample base request.

4. The notification e-mail to beamline staff provides a URL for an internal server that is used to process the request. Alternately, this link can be obtained through a web form to search for unsatisfied sample base requests. Either provides the staff member with a form containing information about the proposal and base request, including the shipping address. A barcode scanner is used to enter sample base numbers into the form. Submission of the form causes the base request to be noted as satisfied, creates new records in the **sample** database table (which are linked to e-mail addresses in the **user** database table) and generates an additional e-mail message to the user noting that the requested bases have been shipped. This step is referred to as "logging in" sample bases; each **sample** record is assigned a status (field *last_sample_event_type*) value of "logged in" at this stage. The beamline staff prepares a package for shipment to the user with materials and instructions for mounting the sample, and sample bases. "Cut and paste" copying can be used to move the mailing address from a web form to a shipping label.

Sample Registration

Before users are allowed to send samples to Argonne for use at 11-BM, information about the sample contents and hazards must be provided. This process is initiated using a URL in the e-mail message generated in the sample login. Alternately, the user can search for unregistered samples logged in against their e-mail address using a web form. Both approaches return a form with a list of samples that have not been registered and provide a mechanism for entry of the required sample information. During this process, the user can select parameters for how data will be collected, which at present are limited

to data collection temperature – a process referred to as "sample registration." The form will not allow registration of samples that do not meet the hazard exclusions for mail-in use of 11-BM. (For example, radioactive and biohazardous materials are not accepted.) This information is stored in the *safety_codes* field in the appropriate **sample** table record. Also, records are added to the **run_request** table for each measurement to be performed on each sample. The sample status is updated to "registered" at this point.

Users may submit a web form to indicate they wish to have their samples returned. If this is not requested, samples are discarded after data collection. This information is saved in the **msds** database table. Due to documentation requirements for international shipments, mailed-in samples are not returned to users outside the United States.

Twice a day, the 11-BM database is queried for newly registered samples. The registration information provides all the hazard information required for the ESAF review. An ESAF form for each proposal listing the newly registered samples is automatically generated and submitted to the APS for review. When the ESAF is submitted, it is assigned a number. This number is stored in the appropriate **sample** record(s).

Sample Receipt

When the samples arrive at Argonne, beamline staff typically scan the sample base barcode numbers into a web form to note receipt. Submission of the receipt form initiates an e-mail to the user(s) acknowledging the arrival of their samples and causes the status flag for the sample to be updated from "registered" to "received." Optionally, a storage

location may be noted on the receipt form, which will also be transferred to the **sample** record in the local database. Should a sample be received prior to registration of sample information, the user is sent a different e-mail message informing them that they are out of compliance with policy and directing them to rectify this immediately.

As samples are received, the web page flags samples with hazards that have special storage requirements (for example, flammable samples may not be stored with oxidizing materials). This allows the bases for these samples to be specially marked.

Data Collection

Data collection is performed in several steps. First, samples are loaded onto individually numbered mounting points that are indexed for loading by the robot. Samples may be loaded individually or in magazines. As is described in greater detail below, the data collection program is instructed to use the robot to pick up and scan the barcodes for samples in pre-selected locations. A schedule for collection of data for these samples that are ready is displayed in the data collection software GUI; the operator can modify this schedule, if desired. Finally, the operator initiates data collection. As the data collection program cycles through the schedule, data files are collected and added to the **run_data** table in the database. The status fields in the **run_request** and **sample** tables are also updated to "data collected." As each file is collected, its name is added to a queue for subsequent processing.

On occasion, samples are taken directly to the diffractometer, skipping the previously described receipt step. In this case, the beamline automation software will note that the

status of the sample has not been set to "received," and will perform the receipt process.

Automatic Calibration

Every 20 minutes the queue of unprocessed data files is automatically scanned for data sets collected using a mixture of NIST Si/Al_2O_3 calibration standards. When such a dataset is collected, it is automatically converted as input for Rietveld fitting using the GSAS software package. A refinement using individual datasets from each detector is then started that determines three calibration constants for each of the 12 detectors: their relative sensitivity, a zero offset, and the wavelength. If the refinement converges to a reasonable χ^2 value, this information is saved in a calibration (.calib) file and the name of the calibration file is also saved in the **run_data** table.

Data Reduction

While most data collection on 11-BM provides results where each detector is equivalent, except for the differing scan ranges and the aforementioned calibration corrections, occasional problems are noted. These are discovered visually inspecting the data from each detector for a few selected peaks, after corrections have been applied for detector sensitivity, zero offset, and wavelength. Discrepancies between successive measurement of the same data range by different detectors across the pattern demonstrates problems such as: samples that decay or evolve in the beam, formation of ice crystals, or irregular intensities when diffraction is dominated by a small number of large crystallites. In the absence of problems, the data are merged to create a pattern that appears as if it had been collected in a single scan, but with the signal to noise enhanced by the repeated observations. When the data merge is completed an e-mail is sent to the user notifying

them that their data have been collected and are ready for downloading. As part of the data merging process, beamline staff have the option to include comments concerning particular issues related to the user's sample and/or data in this e-mail. If datasets are not acceptable, the operator can elect that the status field in the **run_request** record be reset, so that data can be recollected. The names of the data files created in this merge process are stored in the **process_data** table. An instrument operator can cycle through the queue of collected data files quickly with a software tool written for this task.

Data Retrieval

Users can obtain their collected data either through e-mail or via anonymous FTP through the web system. The retrieval process is implemented with several steps to prevent unauthorized access and service disruptions. The user can obtain a form with a list of files that have been collected for a specified e-mail address. That form can be sent by e-mail to that address or displayed in the browser. In the latter case, sample description information is not included in the form to prevent snooping. The user selects the data files they would like to obtain from the form, chooses the data format(s) to be used, and submits the form. The form allows users to specify if they want files e-mailed or posted to an anonymous FTP server. Files are grouped and compressed into Zip archives for transmission. Users can also specify the largest size Zip file they wish to have sent or made available for download. This causes a data file request to be recorded in the **send_data_request** and **send_data_files** tables and generates an e-mail to the registered e-mail address. This e-mail contains a unique URL for the request. When this URL is loaded to a browser, it causes the status of the request to be updated as "released."

Every half-hour the beamline server scans the database for data file requests with status "released." It prepares one or more Zip archives with the requested data and either sends the Zip file(s) as an attachment in one or more e-mails, or copies the files to the FTP server and sends an e-mail with URL links to the files. When data files are sent as an attachment, an additional summary e-mail without any attachments is also sent. This is done because in some cases, e-mail messages with data file attachments have been blocked without notice. Once the files are prepared and are sent or posted, the send request is assigned a status of "completed."

Note that once the request status has been set to "released", executing the unique URL again has no effect. Files will be sent only once in response to that URL, regardless of the number of times the URL is executed. Likewise, once the request is marked as completed, the URL corresponding to that request is ignored. It is anticipated that data files will be retained on the beamline server for years, so that users may retrieve data files again at a later time by submitting a new retrieval request.

Post-Collection Sample Storage

After samples have been used for data collection, they are stored on-site for a minimum of a month. Laboratory safety rules require segregatation of incompatible materials to prevent hazards. For 11-BM, five storage classes are used: no segragation required, nanoscale, flammable, corrosive, and oxidizing materials. As samples are removed from the experimental hutch, they are scanned into a web form and then are loaded into compact trays for storage. An identity code for the tray and its intended hazard class is also entered into the form and that tray label is saved in the **sample** record. When the web

form is submitted, any incompatibilities are flagged. A separate interactive software tool, interfaced to a hand-held barcode scanner, immediately displays sample information, including hazard codes, when the sample is scanned.

Sample Disposal/Return

Samples will be discarded no sooner than one month after data collection, unless a request has been made for sample return. Samples must be segregated into certain disposal classifications, such as nano-particulate materials and those containing metals that are regulated in the EPA Resource Conservation and Recovery Act (As, Ba, Cd, Cr, Pb, Hg, Se, or Ag.) In addition, samples that would pose hazards if mixed (flammable, corrosive, and oxidizing materials) must also be separated.

Sample disposal is facilitated with the aid of an interactive program that interfaces with a barcode scanner. The beamline operator designates a disposal container for a sample and then selects samples to add to that container. As the sample is scanned, the program determines if the sample is appropriate for the selected disposal container and is ready for disposal. If so, a label is printed out for the sample and the sample is added to the inventory for the container. If the sample is to be returned to the user, a mailing label is prepared instead. While disposal rules do not require identification of individual samples for disposal, attaching a label to a capillary does not lengthen the process significantly, and thus is worthwhile. The containers for sample disposal are tracked in the **disposal_container** table while the samples are tracked in the **disposal_log** table.

Project Tracking

One of the anticipated challenges for a high-throughput beamline that services a large pool of users is accumulation and reporting of usage statistics. The most significant milestone is the publication of results. However, gaining information on the utilization of experimental data and the problems that prevent users from completing analysis (particularly if there is an instrumental deficiency) is also of value. As mentioned, there is a status field associated with each sample in the **sample** table. Users are asked to update this field to specify the status of their work with that sample using an easy-to-complete web page. If the user indicates that analysis work has been abandoned, they are asked to provide a quick explanation, with some of the more common reasons, such as sample purity, available as check-off boxes. Citation information is requested for when users indicate that work has been published or is in press, and this or problem information is stored in the **pubinfo** table. Every four months, all users are provided with a form listing all their samples where data was collected more than six months prior, until users update the sample status to "abandoned" or "published."

Other Capabilities

One other feature offered to users is the ability to obtain a report on all samples associated with their e-mail address in the 11-BM database. This report can be an HTML table or can be provided as a comma-separated variable (.csv) spreadsheet file. Since the report contains details about samples registered to a user, the report can be obtained only by e-mail to a registered e-mail address.

Beamline staff have access to a more sophisticated set of web interfaces, where database entries may be searched, reported, and edited. For example, the staff may change the data collection parameters for a scan, or even add additional scans for a sample. Beamline staff may also accociate additional e-mail addresses with a sample or update a user's e-mail to a new address. Logs of all data files collected in a time period may be displayed, or the inventories for the sample disposal containers may be printed.

Document S2. Steps Performed for Data Collection.

The data collection program checks the instrument status, to ensure that the robot and x-ray beam are available for data collection. Then the following steps are performed for each run in the schedule:

1. The beam is blocked with an absorber, to minimize beam damage to samples.

2. The diffractometer 2θ axis is moved to the starting position for the next scan; subsequent steps (3-10, below) are performed while this motor is moving.

3. Unless the sample to be run is already loaded on the diffractometer, the Cryostream is moved out of the sample position.

4. The sample translation is moved to the robot home position and the sample spinner is stopped, if running.

5. Unless this is the first sample to be run, or a repeated run on the last loaded sample, the previously loaded sample is removed from the diffractometer.

6. A sample to be run is picked up and loaded on the diffractometer; the barcode is confirmed to match what was found when the schedule was created.

7. The sample spinner is started and the translation is changed to the data collection home position.

8. If the current run request calls for ambient data collection, the Cryostream set point is set to the next required temperature setting.

9. (or) If the current run request calls for non-ambient data collection:

   a. Data collection pauses until the Cryostream set point is reached, if needed.

   b. The Cryostream device is inserted to control the sample temperature.

   c. A short delay (3 min) is initiated to allow the sample to equilibrate.

10. It is confirmed that beam has been delivered for a minimum period of time to ensure thermal stability of the beamline optics; if not, a delay is initiated.

11. A delay is made, if needed, at this point for the diffractometer 2θ axis to reach the starting angle.

12. The beam absorber is removed from the beam path and data collection is initiated. Once data collection is complete the beam is again blocked by the absorber.

13. When the data collection is complete, it is confirmed that the x-ray beam was delivered for the entire run; if not, the run is repeated.

14. Entries are placed in a queue to add status information about the collected file to the local database. When appropriate, the **run_request** status is updated. Also, a record of the new datafile is added to the data processing queue.

Table S1. The fields contained in each 11-BM sample database table. Subsequent

columns contain the field type and a short explanation of the use of the field.

a) Proposal_Request

Field Name	Data Type	Description
proposal_id	Integer	Auto-generated primary key
proposal_number	Integer	Pointer to entry in GUP database
samples_requested	Integer	Number of sample bases requested by user
request_email	Text	E-mail address on request form
request_status	Integer	Indicates if sample base request has been completed
samples_allowed	Integer	Number of samples allowed based on *hours_allocated* and prior usage
hours_allocated	Integer	Hours of beam-time allocated for proposal in GUP database
local_pickup	Integer	Flag to indicate if sample bases should be shipped to user or held for on-site pickup
request_ip	Text	IP address that supplied sample base request
request_event_time	Timestamp	Time of sample base request change
sent_event_time	Timestamp	Time when sample bases are shipped
request_hashcode	Integer	Unique code for this request

b) Sample

Field Name	Data Type	Description
sample_id	Integer	Auto-generated primary key
bar_code_id	Text	Sample base serial number
proposal_number	Integer	Pointer to entry in GUP database
ESAF_number	Text	Pointer to entry in ESAF database
msds_id	Integer	Pointer to entry in **MSDS** table
chemical_formula	Text	User-supplied chemical formula
chemical_name	Text	User-supplied chemical name
sample_name	Text	Optional user-supplied sample identifier
safety_codes	Text	Sample hazard codes as a concatenated string
safety_text	Text	Other hazard-related user-supplied comments
storage_location	Text	Storage location of sample
last_sample_event_type	Integer	Code for status of sample in registration/measurement process

Table S1, continued

c) MSDS

Field Name	Data Type	Description
msds_id	Integer	Auto-generated primary key
msds_file	Text	Intended for use for upload of sample MSDS listings; not in current use
date_submitted	Timestamp	Documents date of request
Approved	Integer	Not in current use
Comment	Text	Not in current use

d) User

Field Name	Data Type	Description
user_id	Integer	Auto-generated primary key
email	Text	User e-mail address
mailing_address	Text	Not in current use
user_text	Text	Not in current use
name	Text	Not in current use

e) Sample__User

Field Name	Data Type	Description
sample__user_id	Integer	Auto-generated primary key
sample_id	Integer	Pointer to entry in **sample** table
user_id	Integer	Pointer to entry in **user** table

f) Sample_Event

Field Name	Data Type	Description
sample_event_id	Integer	Auto-generated primary key
sample_id	Integer	Pointer to entry in **sample** table
sample_event_type_id	Integer	Code for status of sample at time of change
sample_event_time	Timestamp	Time of status change

g) Run_Request

Field Name	Data Type	Description
run_request_id	Integer	Auto-generated primary key
sample_id	Integer	Pointer to entry in **sample** table
no_of_scans	Integer	Number of times scan should be repeated
start_angle	Real	Starting value for 1st detector during diffraction scan
end_angle	Real	Ending value for 1st detector during diffraction scan
step_size	Real	Angular range covered in each step
count_time	Real	Total time in sec to count each data point; divided by no_of_scans
z_start	Real	Position of sample for data collection
z_increment	Real	Amount to shift sample (in mm) after each data collection
collection_temp	Real	Set-point for Cryostream
status	Integer	Indicates if a run request is active or has been completed
run_request_time	Timestamp	Date run request was entered
comment	text	For use by beamline staff

h) Run_Data

Field Name	Data Type	Description
run_data_id	Integer	Auto-generated primary key
data_file_name	Text	Name of collected data file or calibration result
scan_complete	Integer	Flag to indicate if this scan has been interrupted by a beam dump
run_request_id	Integer	Pointer to entry in **run_request** table
processing_flag	Integer	Indicates if this entry is for a data file or calibration result
run_start	Timestamp	Start time for data collection
run_end_time	Timestamp	End time for data collection

Table S1, continued

i) Process_data

Field Name	Data Type	Description
process_data_id	Integer	Auto-generated primary key
data_file_name	text	File name for reduced data file
run_data_id	Integer	Pointer to entry in **run_data** table
run_request_id	Integer	Pointer to entry in **run_request** table
file_type_code	Integer	Code number to indicate type of processed data file
process_time	Timestamp	Time of data reduction

j) Send_data

Field Name	Data Type	Description
send_id	Integer	Auto-generated primary key
send_hashcode	Integer	Unique code for this request
send_type_code	Integer	Method to be used to convey data (e-mail or FTP)
send_status	Integer	Status of request: placed, confirmed or done
user_id	Integer	Pointer to e-mail address in **user** table
request_ip	Text	IP address that originated request
max_file_size	Integer	Maximum allowed file size in Mb
request_event_time	Timestamp	Time file retrieval request was made

k) Send_Data_Files

Field Name	Data Type	Description
send_file_id	Integer	Auto-generated primary key
send_id	Integer	Pointer to request in **send_data** table
run_data_id	Integer	Pointer to raw data file in **run_data** table
process_data_id	Integer	Pointer to processed file in **processed_data** table

l) Disposal_Container

Field Name	Data Type	Description
container_id	Integer	Auto-generated primary key
container_label	Text	Name assigned to container
container_type	Integer	Hazard class for the container
container_status	Integer	Set when the container has been submitted for disposal and no more sample should be added

Table S1, continued

m) Disposal_Log

Field Name	Data Type	Description
disposal_id	Integer	Auto-generated primary key
sample_id	Integer	Pointer to entry in **sample** table
container_id	Integer	Pointer to entry in **disposal_container** table
disp_event_time	Timestamp	Time sample is added to container

n) Pubinfo

Field Name	Data Type	Description
pubinfo_id	Integer	Auto-generated primary key
citation	Text	User-supplied publication information
dataproblem	Text	User-supplied information on problems with sample or dataset
pub_ip	Text	IP address that supplied publication information
pub_event_time	Timestamp	Time information was supplied

o) Sample__Pubinfo

Field Name	Data Type	Description
sample_pubinfo_id	Integer	Auto-generated primary key
sample_id	Integer	Pointer to entry in **sample** table
pubinfo_id	Integer	Pointer to entry in **pubinfo** table
sample_status_type	Integer	Status of sample when information is supplied

Fig. S1: Work flow diagram for management of the beam-time and sample receipt process. The diagram begins with the central circle and then progresses counter-clockwise from the upper left.
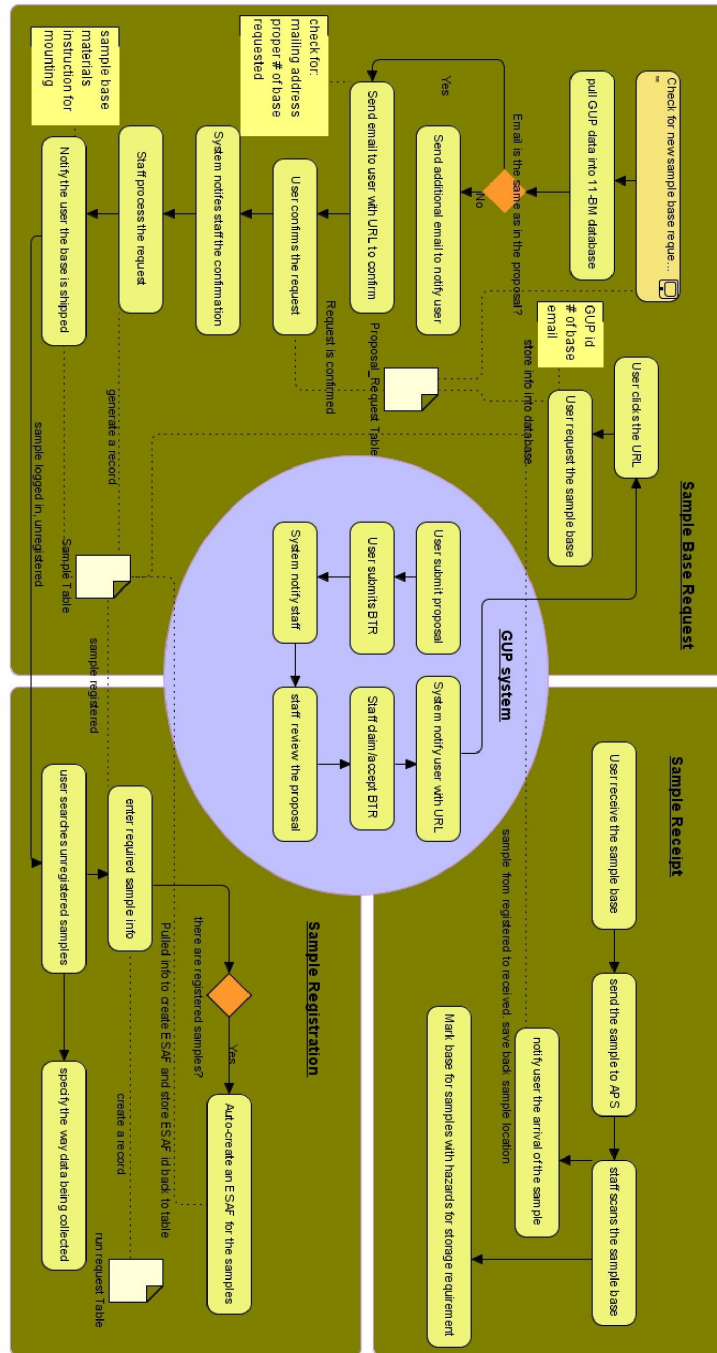
Fig. S2: Work flow diagram for management of the data collection, reduction, and retrieval process as well as collection of user progress information. The diagram begins in the upper left and then progresses counter-clockwise. The final step is in the central circle.