ANL/APS/TB-52

# 2006 XSD Scientific Software User Survey

Pete Jemian

August 26, 2006

In preparation for the 2006 XSD Scientific Software workshop, our committee sent a survey on June 16 to 100 users in the APS user community. This report contains the survey and the responses we received. The responses are presented in the order received.



## Responses

## 2006 XSD Scientific Software Workshop User Survey

June 13, 2006

Dear Colleague,

We have been asked by the XSD Division to organize a workshop to determine our fundamental needs and opportunities in scientific software systems for x-ray data reduction, analysis, modeling and simulation. The workshop has been scheduled for August 29, 2006 at the Advanced Photon Source.

In order to prepare for this workshop we would like your input on what you see as the needs and opportunities for scientific software development at the APS and in the X-ray community, as well as information that would support making a funding proposal for such resources. In particular:

1. What are the limitations of current tools for x-ray data reduction, analysis, modeling, and simulation?

2. What additional tools are needed?

3. How can the existing tools be improved?

4. What will most affect the scientific impact of your work?

We realize you have a busy schedule and appreciate your taking time to address these issues. We need the responses by Thursday, June 22. Please direct your responses to:

xrays at aps.anl.gov

Thank you,


2006 XSD Scientific Software Workshop Organizing Committee
Kenneth Evans, Jr. <evans at aps.anl.gov>
Francesco De Carlo <decarlo at aps.anl.gov>
Pete Jemian <jemian at anl.gov>
Jonathan Lang <lang at aps.anl.gov>
Ulrich Lienert <lienert at aps.anl.gov>
John Maclean <jfm at aps.anl.gov>
Matt Newville <newville at cars.uchicago.edu>
Brian James Tieman <tieman at aps.anl.gov>
Brian H. Toby <toby at anl.gov>
Michel A. Van Veenendaal <michel at aps.anl.gov>

# Response from Harald Sinn

Dear workshop organizers,

as far as the need for meV-IXS, I summarized
the need in the LDRD pre-proposal that I attach with
this mail.
In short, the idea is to develop our software for the
HERIX instrument withing the neutron DANSE
program package.

Harald

**A Unified Data Evaluation for Inelastic X-Ray and Neutron Scattering**

**PIs: Harald Sinn*** (APS, Sector30: HERIX), Ahmet Alatas (APS, Sector 3: Inelastic Spectrometer), Alexander Kolesnikov (IPNS: HRMECS), Chris Benmore (IPNS: GLAD and APS), Chun-Loong (IPNS: QENS)

Argonne National Laboratory, 9700 S. Cass Ave, Argonne, IL 60439

**Description:** Inelastic X-ray scattering (IXS) with meV-resolution has become a very successful technique at 3rd generation synchrotron sources around the world. Using hard X-rays of about 20 keV, an energy resolution better than 2 meV is now routinely achievable. IXS is conceptually similar to triple axes inelastic neutron spectroscopy, however, there are some important differences: Since IXS scatters off the electrons in the sample, the cross sections are quite different from those in neutron scattering. Therefore, for certain samples where the coherent cross section is very difficult or impossible to study with neutron scattering, such as biological samples, $He^3$, V or Pu, IXS can contribute significantly. In addition, the high luminosity of X-rays offer the possibility of micro-focusing, allowing studying small samples, e.g. small crystallites, nanostructures or samples under high pressure and high temperatures.

Despite its success in the synchrotron community, only few groups from the neutron scattering community use IXS as a regular part of their research program. One of the reasons is that data collection and analysis programs currently used in IXS are beamline specific and not as accessible as the software used for neutron scattering experiments.

With the new inelastic X-ray spectrometer HERIX coming up at the APS, offering an energy resolution down to 1 meV at nine analyzers simultaneously, Argonne National Laboratory will take a lead in the field of inelastic X-ray scattering. In order to connect to the well established user community of inelastic neutron scattering, we propose to develop data evaluation software for IXS instruments that is compatible with neutron scattering software, in particular **ISAW**, which is presently used at IPNS, and **DANSE**, a new software project for SNS, where currently four new inelastic neutron spectrometers are being built.

**Description of the effort:** We anticipate that one postdoc over a period of 2 years with a possible extension to a total of 3 years will work on the software projects described above. A bi-monthly meeting with the involved ANL instrument scientists will monitor the progress being made.
In addition to the effort in software development, this postdoc should also pursue an active research program using IXS and inelastic neutron scattering capabilities at **APS**, **IPNS** and elsewhere. Examples for areas of scientific interest are: Dynamics in confined liquids, liquids and glasses under high pressure, and phonons in crystalline materials under extreme conditions.

**Funding needed:** 1 postdoc position or ½ staff position over 3 years (80k$/year).

- sinn@aps.anl.gov , phone: 630 252 9137

## Response from Bruce Ravel

I think I come at this survey from a different perspective than many of your respondents. Many of my software needs are met by my own work and by my collaboration with one of the members of your committee. Given that background, the most serious shortcomings I see in the software landscape are infrastructural rather than end product. I see three big areas that need attention. I'll illustrate these by example from my own experience.

Fundamental data
================

One of my software efforts is a little tool based on the periodic table and on tables of x-ray absorption coefficients. Hephaestus (most of my programs have names that come from a minor obsession with mythology) tries to provide many of the small but useful calculations that one needs while preparing for an XAS experiment or while sitting at the beam-line. It does the additional interesting thing of incorporating five different tabulations of absorption coefficients, thus allowing the user to compare and contrast.

When I started writing Hephaestus, these various tabulations existed in a variety of inconvenient forms. One had been transcribed from a paper publication into a Fortran program. Three others were hidden behind web interfaces, each using a collection of data files, one per element, as the repository of the information. The most recently published was an oddly formatted flat text file available as the supplementary information from the journal in which it was published. What a mess!

I gathered these various resources together and wrote a bunch of one-off scripts to convert all of the data into a database. I then encapsulated these data behind a bit of OO code written in the same language as Hephaestus. That served my purpose very well -- I was able to access data from any resource with a common interface. My solution, however, serves other developers very poorly -- unless they happen to choose the same programming language I use.

The APS could serve its users very well by identifying information fundamental to the various experimental disciplines, encapsulating that data into a package using a low-level language, and providing a nice wrapping mechanism such that the information can be deployed in any context and using any programming language. The model I have in mind is a database + a wrapper written in C. SWIG could then be used with almost any language.

APIs to theory calculations
===========================

Theory is great! Without it, we would be stumped by most of what we measure at the APS. As great as theory is, it is all-too-often written by brilliant scientists with no training in computer science. The consequence is that many theory packages are clunky, monolithic black boxes that are difficult to deploy in context that their developers did not think of.

My main experience is with Feff, a popular, first-principles, absorption spectroscopy program. When version 5 appeared back in the early 90s, it was a hit and fueled a rapid development of the XAS technique. Sadly, Feff remains stuck in the early 90s from a user interface perspective. Information goes into Feff via a flat text input file and information comes out of Feff via inconsistently formatted flat text files. Subsequent software developed to use Feff as its theory engine must deal with a clumsy and fragile system of file-based IO and system calls.

Feff would be much more powerful with some kind of API. One part of the code uses previously-computed atomic potentials to calculate the contribution from a multiple scattering geometry given a set of atomic coordinates. As a part of data analysis package, it would be attractive to be able to move atom positions and recompute the contribution on the fly. Currently, such a thing requires file IO and system calls. The capacity to do this via an API would allow the development of different, interesting analysis software.

I fully understand that Feff is a poor example of something that the APS could work on. Feff is the research project of a particular university research group and is encumbered with a restrictive license. The point of this example was to illustrate a way in which a dedicated synchrotron software group could serve its constituency by making extablished theory available for novel uses.

## Documentation of file formats

When I collect imaging data from 2ID-D, I come home with maps in the mda file format. Fortunately, I get to use Stefan Vogt's MAPS to import these files and to export individual pixel spectra in a easier-to-use format.

But ... what's an mda file? Perhaps its contents are documented somewhere, but that place isn't easy to find. Perhaps I don't google cleverly enough, but I was never able to find any documentation on its contents. It may have been helpful to me to be able to examine the data independently of MAPS, but I never had that option. I am not opposed to big, binary blobs in principle. But without documentation and a clean API, they are not particularly useful.

Good documentation of file formats used on the floor and a central, easy-to-locate respository of information would be very useful indeed.

There is a theme to this long-winded email. None of my concerns are about end-user software products. Each is about the things that are used to build end-user software products. The APS draws lots of clever, motivated people to its doors. Were those clever, motivated people given nice tools, they could build great things.

Hope that helps,
B

--

Bruce Ravel ------------------------------------ bravel at anl.gov

Molecular Environmental Science Group, Building 203, Room E-165
MRCAT, Sector 10, Advanced Photon Source, Building 433, Room B007

Argonne National Laboratory phone and voice mail: (1) 630 252 5033
Argonne IL 60439, USA fax: (1) 630 252 9793

My homepage: http://cars9.uchicago.edu/~ravel
EXAFS software: http://cars9.uchicago.edu/~ravel/software/exafs/

# Response from Yong Chu

Dear Colleagues,

The software and data handling/storage are of significant importance to me and my collaborators. We routinely produce over 10-50 Gbytes of data per day, when we are conscious about saving only the useful data. The following issues are important to me.

1) fast saving of data for the acquisition so that the bottlenect of the data collection is not on the saving time.
2) backup and safe storage of large amount of data (100~200 Gbytes over per week).
3) implementation of the imaging process software (that we developed in IGOR) into a better platform (such as Java or others) so that the software can be distributed easily without copy right issues.
4) find a way to run the imaging processing software on faster machine with a close tie with the data acquisition (i.e. as soon as the data are collected, they are processed and the results can be viewed).

Please let me know how I can give my input to the Scientific Software Workshop.
Sinerely,

Yong Chu
630-252-0150

## Response from Jan Ilavsky

>
>*1. What are the limitations of current tools for*
> *x-ray data reduction, analysis, modeling, and simulation?*

There is no system to available tools - missing is:
1. central data base with description, examples, capabilities
2. central download server with backup of older versions

Finding necessary tool is currently detective work of searching papers, asking colleagues, etc... Does not guarantee, that one will actually find all (or most) of available appropriate tools...

In more general terms, the software available is patchwork of programs developed on different systems, by different groups, with different GUIs and file formats, etc. Very difficult to get them work correctly in more complicated problem.

>*2. What additional tools are needed?*

More than what particular tools - system is needed. System, where one can use modules appropriate for particular step in data processing/analysis...

>*3. How can the existing tools be improved?*

Have central support - what I mean to have way of finding help in case I need to use existing software which I do not understand. May be have database of "local experts"? If someone has problem with Fit2D (for example), who is local here with "expert" knowledge - and willing to help me?

Oh yes, and having manuals would be useful... But that is another problem.

>*4. What will most affect the scientific impact of your work?*

If someone else will develop tools my users need :-D

More realistically, if I can be sure that others will be developing tools compatible to what I develop. Or may be the other way around - if I could be developing tools in an environment (=system), in which I would be sure they can be used with software from others... We duplicate our efforts too much.

Jan

*****************************************************************
Dr. Jan Ilavsky, physicist

X-Ray Science Division
Advanced Photon Source
Argonne National Laboratory
9700 S. Cass Avenue, bldg 438E
Argonne, IL 60439, USA
business mobile (630) 774 8814
office (630) 252-0866
(630) 252-0862 (Fax)
ilavsky at aps.anl.gov
SMS: 6307748814 at messaging.nextel.com

web page:
www.uni.aps.anl.gov/~ilavsky
Indra 2 web page: www.uni.aps.anl.gov/~ilavsky/indra_2.html
Irena 2 web page: http://www.uni.aps.anl.gov/~ilavsky/irena.html
SAS web page:   www.uni.aps.anl.gov/~ilavsky/sas.html
USAXS web page: http://www.uni.aps.anl.gov/usaxs/
APS small-angle page: http://small-angle.aps.anl.gov/

# Response from Gene Ice

> *1. What are the limitations of current tools for*
> *x-ray data reduction, analysis, modeling, and simulation?*

Software for polychromatic microdiffraction is one-of-a-kind and primitive in its ability to pass information from one part of the experiment to another. There is a need to simplify the user framework without losing flexibility. Jon Tischler has lots of ideas on how this can be accomplished but little time. This is a worldwide problem with new instruments of the same class being developed around the world. We believe an international collaboration-which we are fostering- will allow for more user friendly, powerful and robust data analysis. Right now, only a few scientists can analyze data after it is collected. This is a tragedy and should not be the situation. The data should be analyzed as it is taken.

>
> *2. What additional tools are needed?*

1. Automated fit to elastic strain with uncertainties
2. Automated uncertainties for orientation
3. Automated fit to dislocations
4. Automated fit to simultaneous elastic and plastic deformation
5. Real-time data analysis
6. Energy scan automated phase determination (in collaboration with high pressure community)
7. Build in the ability to accept and parse data at much higher frame rates (~100 Hz).

>
> *3. How can the existing tools be improved?*
>

1. Build a framework that is robust and has standalone modules that can be worked on by worldwide community

> *4. What will most affect the scientific impact of your work?*
>

Polychromatic microdiffraction already provides unique information. With the improvements above we believe we can revolutionize our understanding of materials. Our facility is incredibly productive as is, but with these improvements it will become flooded with proposals and with groundbreaking scientists and will be much more useful for outside users. These improvements should increase productivity by a factor of 2 to 3!

# Response from Ward Smith

In regard to Scientific Software Workshop I would make the following points,

Common format for raw data (i.e. images); an initiative by the American Crystallographic Assoc has begun and will include as many different types of raw data as possible, not only crystallographic data. See for example WK.02 at
http://www.xray.chem.ufl.edu/aca2006/private/FebCFP.pdf

Data archiving - a major issue for some of us with large format CCD detectors - up to 36Mbytes every few seconds up to 200+ Gbytes per 24 hours.

Ward

## Response from Paul G. Evans

We've been spending a fair amount of time working between the slices of reciprocal space acquired in CCD detectors and useful reciprocal space maps.  This is a problem that everyone using CCDs to do thin film diffraction faces, and something that I've seen limiting peoples' experiment.

Basically, the idea is that the stack of CCD images covers a volume of reciprocal space - which is really useful.  But the two axes of each image are often not particularly useful scientifically and the most useful maps have to be made by integrating across several images.  I'll find a picture of what I'm talking about.

A tool to integrate across several images in a general way would be fantastic.

Paul

# Response from Jon Tischler

The point that could have the greatest impact would be to get the written is a manner that is uniform across the facility. This way, analyzing data, particularly the immediate analysis during an experiment could happen more efficiently. Right now there are multiple programs that take data, all in their own fashion, and even for one program, there is no consensus on how to use it to save data (e.g. spec is used multiple ways here).

Simple things, such as naming the channels in spec (is incident beam intensity called, Io, I0, iO, io, ion, ionc, ...?). Is the scattered beam detector (scint, det, bicron, Bicron, detector, oxford, signal, NaI, ...?) Is the detector a photon counter (which needs a dead time correction) or is it an integrating detector? Even little things like this can bring some sanity to the analysis. It is not enough to have a description of the detectors, but a dependable naming (or some means of identification) is needed so that at analysis time the computer will know what it is. People will spend weeks writing routines to operate an experiment, and only minutes on a way to write the data This lack becomes evident when it is time to analyze the results; especially when something is added or changed in the data file. I've even heard that some people are still using two column files without any headers, because it was once easy to write. I guess they deserve what they get.

Continued support for EPICS is essential, it is one of the few things that can be depended upon to work in more than one experiment.

EPICS is good for controlling things, but is no good at taking data. Should the issue of data collection be revisited?

Jon

## Response from Lyle Levine

The two beamlines that I use heavily are 33ID and 34ID.

USAXS on 33ID: This is a perfect model for how x-ray data reduction, analysis, modeling, and simulation should be done. A complete set of user-friendly analysis and modeling tools are available to all users and good documentation was given a high priority. Many (probably most) users are not experts in X-ray physics and such software tools greatly enhance the usefulness of the beamline.

DAXM on 34ID: The experimental design of this beamline is excellent, but the data reduction and modeling tools are virtually inaccessible to users. The problems have 2 main components:

1) Data sets are extremely large, making access difficult. Raw data can only be provided to users on many DVD's.
2) Even if the users take their data with them to their home institutions, the only software for interpreting the data exists at the APS, in the form of user-unfriendly routines that require an expert to run them and that cannot readily be used elsewhere.

Suggested solution:

1) User-friendly, documented analysis programs are critically needed.
2) Users require access to both theses analysis programs and to their data, either through the web or at their home institutions. If web access is unfeasible, then the analysis programs must be written to run on the most generic computer systems possible such as Window XP boxes and be provided as executables.

Regards,
Lyle Levine

## Response from Randy Alkire

To the committee:

*1. What are the limitations of current tools for
x-ray data reduction, analysis, modeling, and simulation?

Speaking for myself and not SBC directly, it is my opinion that most of the protein crystallographic beamlines are using software developed by third parties. We have a collaboration with HKL3000 developers to work on data collection, data processing and data reduction. Because we are a beta test site, most of the cutting edge processing and reduction software is not released to our users until fully tested. This gives us an advantage in the short term which we use to improve our operations. Because the requirements for developing this kind of software are highly specialized, it is unlikely APS personnel would be able to add anything meaningful to the existing packages we already use.

*2. What additional tools are needed?

Due to the age of our beamline we need to completely re-work the motor hardware and software. Our components are end-of-life and we would benefit from any resources available from the APS that aids this lengthy integration step, including new tools and techniques for integrating scalers and fast scanning techniques.

*4. What will most affect the scientific impact of your work?

We would benefit greatly by improved data collection and optics monitoring hardware, including capabilities that would allow real-time monitoring of the beam location, reduced source vibration and high-precision goniometer synchronization. We would also greatly benefit by an improved cold-stream design that allows real-time monitoring of nitrogen flow and temperature.

If this is just about software, we could benefit from improved fluorescence scanning and monitoring, including an embedded MCA. We would benefit by having scanning software developed that would allow multiple pass scans, have well integrated plotting software and allow high-speed scanning. Linkage to high-speed scanning may be related back to the motor operations.

Sometimes advice is all that we need. However, if there were a set of existing beamline controls that could be adapted to help setup the beamline in a rapid time frame or maintain its optimal alignment, say after a mode change, this would be useful to all our users.

Randy

## Response from Daniel Haskel

1. What are the limitations of current tools for x-ray data reduction, analysis, modeling, and simulation?

Perhaps a distinction between needs during the experiment and after it is in place. I believe there is a need for a flexible package for data manipulation for the beamlines to be used to evaluate the quality of data and perhaps some preliminary analysis. Some of this is too specific to implement in a broad sense, like XAFS and powder diffraction, and there already exists good packages for that (ATHENA, ARTHEMIS, IFEFFIT etc...). I see need for a package that can read in multiple files, operate on columns within a single file or between files, performing a large number of mathematical operations (log, derivative, smoothing, integrals). This should be accompanied by a nice GUI with good plotting capabilities, allowing plotting of data columns from multiple files, results of operations, zooming, cross-hairs, etc.

A number of packages do some of this, such as Newplot or the data processing capability within XOP (XPLOT). Both of these handle SPEC files, they can plot columns from more than one file, but they can't operate between columns from different files. A very simple example is wanting to calculate the rms variation in some data using 5 consecutive measurements.  Another is to take the difference between two data sets taken with opposite applied field directions. Yet another is to fit a polynomial background to a data set, subtract the background from the data, and save/plot the background-corrected data. These are all "trivial" things that we all do with our favorites IDL routines, C-PLOT, spreadsheets, or whatever it may be, but I don't believe that we have here at the APS a GUI/data manipulation software for the beamlines that is generic enough that it can be effectively used by all. Newplot seems to be a good beginning, but it is far from being very useful.

My personal opinion is that good beamline-based data-manipulation software is very important to make the user's experience the best it can be. This largely falls on the beamline scientists, resulting in various sector-specific tools which make it harder for an experimenter to move from sector to sector. A generic powerful package for data manipulation will be a huge improvement to all beamlines.

2. What additional tools are needed?

See no. 1

3. How can the existing tools be improved?

See no. 1

4. What will most affect the scientific impact of your work?

Some users, especially those new to the synchrotron radiation arena, need lots of help with data analysis. Since the beamline scientist time is limited, this delays the turn around time between experiment and publication a lot. It might be worth considering having a couple of people in the theory group that do modeling/data analysis work in a couple of major areas such as scattering and spectroscopy. These people will be strong in modeling and could work with users that need help in modeling their data. I'm not talking about exoteric theory but simply help with analyzing data. Outreach activities such as summer schools, etc, to train people in the use of existing/new packages is also a great complementary way to pursue this problem.

--------------------------------------------------------------------

Dr. Daniel Haskel
Physicist, Magnetic Materials group

Advanced Photon Source, Bldg.431/E008
Argonne National Laboratory
9700 S. Cass Ave.
Argonne, IL 60439, USA

Tel: (630) 252-7758
Fax; (630) 252-7392

http://www.aps.anl.gov/xfd/people/haskel/dani.html

## Response from Wenjun Liu

Glad to know that there will be an XSD scientific software Workshop soon.
Here are some of my thoughts on improving data processing capability for analyzing large sets of synchrotron three-dimensional imaging data at 34-ID-E microdiffraction beamline.

The major research activities in the beamline 34-ID-E involve development of new micro/nano-diffraction techniques for characterization and microscopy for condensed matter physics and material sciences. Three-dimensional micro-diffraction provides structural information of materials with better than 1 micron spatial resolution in all three dimensions enabling detailed studies of fundamental deformation processes, basic grain-growth behavior, and small scale structures.

The unique capabilities of 3D micro-diffraction probe will continue to be improved to meet user demands, including enhanced capabilities of fast data collection and process and more user-friendly analyzing software improvement.

1. Enhanced capabilities of fast data collection and process:

Current commercially available CCD detectors appropriate for 3D micro-diffraction needs collect data at speed of >10MB/sec. To process the image data at these rates and enable real time in-situ measurements requires enhanced computing capability. A cluster computer with high-speed network connections will be a solution. To take the advantage of parallel computing on cluster, supporting software development and integration into existing analyzing software are essential.

2. User-friendly analyzing software improvement:

The current 3D micro-diffraction analyzing software package was developed by Oak Ridge National Lab using IDL language. ORNL, as partner user of the beamline working together with APS beamline staff, will continue developing the existing software to make it more user friendly and efficient for fast data processing.
Recent plan includes modification of existing modules and possible redesign of the software architecture to meet parallel computing needs.

We believe that all these will continue the excellent scientific research and development and provide dramatically increased user throughput not for 34-ID-E beamline only, but also for the growing community of 3-D x-ray microdiffraction world wide.

Thanks,

Wenjun

========================================

Wenjun Liu, Ph. D.
Beamline Scientist
Building 438D

Advanced Photon Source
Argonne National Laboratory
9700 South Cass Avenue
Argonne, IL 60439
E-mail: wjliu at anl.gov
Tel:(630)252-0890 (office)
(630)252-1834 (beamline)
Fax:(630)252-0862
http://www.uni.aps.anl.gov/microdiff/

## Response from Jan P. Hessler

I hope the attached covers our previous discussion.

As always, any questions, just call.

Jan

Jan P. Hessler, Ph.D.

Chemistry Division

Argonne National Laboratory

9700 South Cass Avenue

Argonne, Illinois 60439-4831

Voice 1-630-252-3717

FAX: 1-630-252-4470

To: XSD Scientific Software Workshop Organizing Committee
From: Jan P. Hessler, Chemistry Division
Subject: Needs and Opportunities for Scientific Software Development

These comments stem from my recent experiences with Small-Angle X-ray Scattering,
SAXS, and my long interest in data reduction.

1. During an experiment it is essential that the scientists have the capability to view their results as quickly as possible. For SAXS this involves more than just looking at a plot of intensity (arbitrary units) vs. transferred momentum (1/nm). The arbitrary units have to be converted into absolute units and some initial data reduction must be available. The more detailed the data reduction the better. Basically, the scientist must be able to determine that (1) the data is publishable and (2) the data answers the question the experiment was designed to address. Unfortunately, these questions must be answered quickly so the scientist can, if needed, modify the experiment.

2. When a user completes a series of experiments at the APS the files of data must be complete, easily understood, and ready for further data reduction. By complete I mean that every important parameter is available and that calibrations, for example of the detectors, are traceable. Documentation must be available that details how all calculations were performed. Finally, data reduction requires realistic estimates of the experimental uncertainties. Finally, the user must be able to compare and contrast data taken at the APS with data taken at other synchrotron facilities and, perhaps, neutron scattering facilities. I know that I am stating the obvious. However, I also know that at least one user is dissatisfied with their experience at the APS and will not return unless these conditions are meet.

3. Users must be able to reduce their data with many different software tools from many different sources. For example, at a recent conference I presented my data, another speaker presented data taken at the ESRF, and a third data taken at a neutron scattering facility. When we reduced our data, we all used different approaches. Of course, we all got different answers and drew different conclusions. At the end one participant commented that he was not going to take any of our results seriously until we had used the same data reduction techniques. This is a fair criticism and it demonstrates the need to have many different software packages available. After the conference we exchanged our data, but we should be able to address these issues before a conference, not after.

4. State-of-the-art data reduction techniques must be made available to researchers during the experiment and after it is completed. For example, one of the best ways to estimate the uncertainties in the parameters of a model is to perform a Monte Carlo analysis by rerunning the data reduction many times. This is best done with a cluster of processors where each processor performs a single reduction. Therefore, we should have a cluster of at least 1024 processors. This cluster should be available to all users both during their experiment and from their home site.

5. As our understanding in certain areas progresses, simulations are going to be relied upon to design experiments. These simulations must

include not only the different models that are being tested, but also a realistic estimate of the noise in an experiment. At least two questions must be answered: (1) Will this experiment distinguish between one or more different models? (2) Is the anticipated signal-to-noise ratio sufficient to provide an unambiguous answer?

# Response from Henning Friis Poulsen

In this mail I'll summarise my recommendation for an APS software strategy within the field of hard x-ray materials imaging. My background is 15 years of synchrotron instrumentation and development mainly in Europe. In particular our group has collaborated closely with ESRF on algorithm development and automation at the Materials Science beamline ID11. My connection with APS is mainly via a 3 year long development project at Sector 1-ID.

The thoughts on implementation below originate to a large degree from Andy Götz at ESRF and from Søren Schmidt at Risø.

Relevance

Imaging of materials (in the broadest sense of the word including diffraction, fluorescence, phase ... contrast and including use of tomographic reconstruction as well as scanning principles) is one of the great success stories of third generations synchrotrons. With new capabilities like nano-characterisation and with increased awareness in the materials science community, the user group is bound to increase.

Put schematically, the field faces three challenges:

* High throughput. The many new users and the natural drive towards matrix studies with many samples rather than just one-sample feasibility studies makes automation and standardisation a clear issue. Most users would also demand that the time it takes to perform data analysis at home is reduced from ~1 year to 1 week or preferably handled on-line (similar to characterisation with e.g. electron microscopy)
* In situ studies. Combined sectioning/electron microscopy machines are now on the market, with a spatial resolution of 50 nm or better. This emphasise that the main role of the synchrotron will be in dynamics studies, as these by definition cannot be made using invasive tools. For in situ studies on line visualisation is crucial, in part to steer the experiment ("having seen this global map I now want to focus on this part of the sample"), in part to learn whether the experiment was successful ("did anything nucleate in this sample during the heat treatment?")
* Data handling. The combination of more powerful detectors/computers and the wish for fast in situ dynamic studies implies that we should expect >1 Tb of data per day. Storing, reducing and analysing this data on line is beyond the reach of any user and requires an APS strategy

In one sentence, it is widely recognised that the main limitations of hard x-ray imaging today are software and detectors.

Suggested strategy

One may define 3 levels of operation:

1. Basic interfacing:

There should be a standard way of interfacing all hardware: detectors, furnaces, stress rigs, encoders. My suggestion would be via platform independent device servers, such as the TACOs developed by Andy Gotz at ESRF.

There should be one generally accepted program for basic control of motors, interfacing to the device servers, data handling (directories) and logging. There are a number of programs, which all can do the job, but I suggest spec, as I believe this is the one that most people are familiar with. (It may be that a lot of the spec functionality is superseded later on by more high-level programs - see below - but one needs from day one something robust for these key features.)

For commissioning, debugging etc. APS should continue to rely on EPICS. However, the regular user should not have to deal with this low level of control.

The architecture should allow all data to be handled immediately by a cluster, if needed.

Each imaging beamline should be associated with one "interfacing" expert. The aim should be that users simply should not worry about interfacing at all !

2. Image analysis and data reduction

There are a number of image analysis operations which are regularly performed on images. These include: visualisation, new data format, rotations/transformations, background subtraction, corrections for spatial distortion and flat field, deconvolution with point-spread-function, normalization of intensity, screening, polar transform (so-called caking), and identification and characterisation of areas-of-interest such as diffraction spots.

All of these operations can be handled by an existing program like FIT2D, which however is not well suited for high throughput. Instead these operations should be coded in machine language for maximum speed. The aim is that such operations per default is performed automatically and that users most of the time do not take home raw data.

To be able to keep up with the input data - 2D detectors can read out in milliseconds - long-term it is probably needed to include distributing images over a set of nodes in a cluster. However, a slower analysis would be acceptable as a start (it is much better to be able to see 10% of the data on line than 0%).

The user would need to tell which operations to include. This could be done by a small primitive GUI or perhaps just in spec.

3. Data analysis

Data analysis in imaging is characterised by application of large mathematically-heavy programs like reconstruction, segmentation and grain mapping algorithms. The output are 3D renderings of the materials investigated. Following that science specific quantification programs are applied which measure geometry, frequency, correlations etc.

Developing these algorithms is very much an on-going task which furthermore requires substantial resources and mathematical insight. Some users will be interested in "standard products" like a filtered backprojection reconstruction in conventional tomography, while others will arrive with programs of their own or wish for other algorithms, which for their particular case might provide better resolution or be faster. Hence, I believe APS should have a dual aim:

- maximum ease/flexibility for experienced users to bring own software and get that fully integrated in the on line analysis chain

- identification of some standard tools for routine use and optimisation of these with help from the relevant expert users.

At ESRF, Andy Goetz has suggested a "plug-and-play" solution with a GUI based on Eclipse. This seems very powerful. In particular:

- The GUI itself is already made, is independent of the scientific modules and includes 2D and 3D plots. This e.g. implies that one should be able to make first use of this program with a relatively modest effort.

- It should be able to combine spec commands with the image analysis operations identified in 2 above and with calls to user functions of relevance to 3.

- It allows for direct interfacing to MATLAB.  MATLAB is in my view an excellent tool for "on-the-fly" analysis.

- Each user can have their own GUI, such that the information they get on the screen is customised to their needs and is the "same as last time". First-time users can get a "standard GUI".

In the 5-10 year perspective I foresee that the shear data-load will be so heavy that one needs to approach the issue of data analysis and reduction in the same way as it is done today in particle physics. Søren Schmidt and others at Risø has worked for the last 3 years on the backbone of such a system, known as FABLE.

Finally I recommend that all software is open source, such that one can take full advantage of specialised modules written by users and parallel efforts at other synchrotrons.

I believe the above recommendations would also apply to beamlines specialising in diffraction, including crystallography.

Status

Being a user with a long-term proposal running at both sector 1 at APS and beamline ID11 at ESRF makes a direct comparison natural for me:

ID11 at ESRF completed phase 1 above >5 years ago and have nearly completed phase 2. Work has just begun on phase 3. In contrast, Sector 1 at APS is presently still working on phase 1.

(To be fair the group at APS has the edge on ESRF in other aspects, such as optics.)

# Response from Stefan Vogt

Dear all,

here are my comments regarding your email on scientific software. I would like to point out that they are from the perspective of what I am most interested in, x-ray fluorescence microscopy.

Before addressing specifically the points that are raised below, I would like to make a more general, but possibly somewhat to the side comment.
It is appreciated that efforts are being put into development of 'scientific' software. However, I would like to point out that there is 'room for improvement' in the software that is currently being used to run the beamlines, in terms of the efficiency of setting up scans, as well as the ability to view acquired data. It seems ironic that one can set up a 5 dimensional scan, but that (to the best of my knowledge) no software exists to actually view the acquire data. Or that I can find a very irregular structure that I wish to scan, but a scan that I will have to set up, ends up scanning mostly empty space, and wasting beamtime.

Both have direct impacts on the efficiency of operation.

> *In order to prepare for this workshop we would like your input on*
> *what you see as the needs and opportunities for scientific*
> *software development at the APS and in the X-ray community, as*

Personally, I believe that as the x-ray communities matures, and the science applications are done in a fairly routine user mode, we have to not only have an optimized beamline, but must also support users in data analysis, so that user's go home with as much pre-analysed data as possible. Emphasis should also be placed on useability. I believe both are particularily true for users from communities that do NOT have a physics background, etc (e.g., users from the life sciences).

> *well as information that would support making a funding proposal*
> *for such resources.  In particular:*
>
> *1. What are the limitations of current tools for*
>   *x-ray data reduction, analysis, modeling, and simulation?*

we acquire 'raw' data as spectra with counts/s, what our user's want in a first approximation is quantified data, reduced to elemental concentration (e.g., mmol Zn). Then, one should be able to analyse large quantities of acquired data semi-automatically, and report content for specific regions or areas across different scanned samples.

This involves per pixel fitting of spectra, possibly multivariate analysis, cluster analysis, comparison to a claibration curve, etc. Currently, I am not aware of any tool that would allow one to do all of this, in particular not across the different beamlines.
For sector 2, I have written an IDL based software to do some of this, but it is limited in that it requires the use of certain standards, and is not routinely useable at other beamlines.

> *2. What additional tools are needed?*

APS has a fairly large number of microfluorescence beamlines, either present, or in planning/contruction. It would seem a very good idea if there was a supported set of software tools available / developped, that would make it possible to perform the tasks outlined above, and convert the acquired, raw data into 'useable', pre-analysed data for users.

As a next step, for each type of experiment, one should have a more specialised set of tools that allows one to further process the pre-analysed data, e.g., segment each acquired scan into specific (similar) areas, and automatically compare different areas for different sample conditions/treatments. This could be a cluster analysis or neural network type method that would be able to work on scans of cells, and separate them into different characteristic subunits (e.g., background, cytoplama, nucleus), and compare these across different treatments.
For example, a typical user might be coming in and ask, O.K., I have the two cell lines, how is their elemental content influenced by, and how do they react differently to, a treatment with this therapeutic drug. The user will probably not care about that one needs to fit the data in the first place, compare it to a calibration curve, find corresponding regions of the cell, extract that data, and then compare the quantified data from one treatment to the next. All the user is really interested in is in the result.

Additionally, tools that allow simulation of data from specimens, to investigate feasiblity of certain studies without the need to spend actual beamtime on it, should be available.

> *3. How can the existing tools be improved?*
methods and software pieces for most of the required functionality is available somewhere, but it is not unified in any way. I believe one should instead have one COMMON piece of software that is useable across the different beamlines that have similar needs. A user should be able to use the same software tool (though possibly with different options), for data that was acquired, no matter whether it was acquired in sector 2, 18 or 20; or, to go one step further, at APS, ESRF, or elsewhere.

>
> *4. What will most affect the scientific impact of your work?*

accurate analysis of acquired data is a must if we want to remain competitive. As the throughput of our beamlines increases, and the quality of the data improves, the analysis part is more and more of a bottleneck. As a first step, I believe that (semi)-automated data (pre)-analysis will be crucial (raw data -> pre-analysed).
The next step is then haveing more sophisticated routines that allows further processing with as little as possible user interaction, to tackle large datasets (e.g., compare cytoplasmatic vs nuclear elemental content of 100 cells that were imaged during the last run).

Cheers,
Stefan

--
Dr. Stefan Vogt,
Experimental Facilities Division
Advanced Photon Source
Argonne National Laboratory
phone: (630) 252-3071; beamline: -3711
fax: (630) 252-0140
e-mail: vogt at aps.anl.gov
http://www.stefan.vogt.net/

## Response from Andrew Allen

June 21, 2006

Fundamental needs and opportunities in scientific software systems for x-ray data reduction, analysis, modeling and simulation at the APS and in the X-ray community:

1. What are the limitations of current tools for x-ray data reduction, analysis, modeling, and simulation?

Many things are piecemeal and left to the user to sort out.
At some beam lines the basic reduction software is glitchy and labor intensive to use.
There is no house style (hence standard) for x-ray data reduction.
For data analysis the situation is frequently worse. I am aware of very few data analysis packages, readily available to the user, that are home-grown at APS.
Frequently, the best course of action is to go to other external web-sites, frequently to ESRF - where there seems to be a large body of analysis software built up specifically for ESRF users.
Clearly ESRF is very different in this regard; so are the US neutron user facilities (at least some of them!). There are also some bright spots at APS, e.g., sector 33.

2. What additional tools are needed?

The cutting edge has to be with the high-Gb data-producing beam lines such as 34-ID, 2-BM, etc:
From the many discs of data obtained over a few days, how does one extract the key results and salient conclusions within a few weeks?
This will require fast machines, large amounts of memory, and very sophisticated techniques to "mine" the data and pull together the relatively small proportion that satisfies the "necessary" and "sufficient" conditions for a successful result.
But NASA and NOAA have to do this all the time. The APS should consider setting up a partnership with these agencies (and others) in order to bring sufficient resources to bear on this increasingly daunting challenge. Having put into place some strategies and resources needed to meet the challenge, it will then be essential to provide the users with easy remote access to the data and to the processing system. [I think this item could form the basis of a budget proposal.]

3. How can the existing tools be improved?

Major manpower effort is required to write reduction and analysis software tools, together with appropriate manuals, that will guide the user through to getting publishable results.
This means individual personnel taking responsibility for the packages produced, and interactively improving/advancing the software tools with an ongoing interaction with the user community.
Very little can be just written and left (orphaned!); there has to be an ongoing interaction in the various experiment techniques areas.
While software specialists may be needed, it is essential that the beam line scientists adopt intellectual ownership of what is provided, but

in such a way that it serves the user community, not just their own science.

4. What will most affect the scientific impact of your work?

Clearly, all of the above are relevant here, but I would add that some more thought could be given to how data are moved around from instrument to instrument and even between techniques.
As a crystallographic journal editor, I am becoming aware of an increasing number of software submissions that are more sophisticated in handling data from different places, in different forms, across platforms, etc. It might be worth the APS as a whole deciding which program languages will be supported, and then converting as much of the data reduction and analysis software as possible to run on many or all of the supported languages and platforms.

I realize that the above sets out some daunting tasks. However, to take some or all of these issues on in a serious coordinated way would put the APS at the center of x-ray science on the world stage to a greater degree than at present. Clearly, the ESRF has done this in Europe, very successfully. The fact that many APS users look to the ESRF for their data analysis tools says rather a lot! Obviously, the APS could not do the things I suggest above without a very active and ongoing collaboration with its users.

I hope something here is of use!

Andrew Allen

-------------------------------------------------

Dr. Andrew J. Allen, Physicist
U.S. Department of Commerce
National Institute of Standards and Technology
Building 223, Room B206
100 Bureau Drive, Stop 8520
Gaithersburg, MD 20899-8520
Phone: 301-975-5982
FAX: 301-975-5334
E-mail: andrew.allen at nist.gov

## Response from Sergey Stepanov

```
> In order to prepare for this workshop we would like your input on
 > what you see as the needs and opportunities for scientific
 > software development at the APS and in the X-ray community, as
 > well as information that would support making a funding proposal
 > for such resources.

The need for scientific software can be proven by:

-- the statistics of X-Ray Server (http://sergey.gmca.aps.anl.gov) that
    I have been running at the ANL since 1997. Although the Server
    provides scientific software for a small subset of diverse X-ray
    studies carried at the APS (namely for high-resolution diffraction
    and scattering in the field of material science), it has got close
    to 130,000 calculations requests from about 5,000 researches with
    about 1,500 regular users who submitted ten or more jobs.

-- the existence of scientific software projects at similar facilities,
    for example:
    (a) the 9-member Scientific Software Group at the ESRF
        (www.esrf.fr/UsersAndScience/Experiments/TBS/SciSoft/Members/).
        The group is authoring widely used software packages like XOP
        and FI2D.
    (b) the DANSE (http://wiki.cacr.caltech.edu/danse/) scientific
        software project at the Spallation Neutron Source for which
ORNL
        requested $15M five-year grant and has already received some of
        those money.
    (c) the popular software project at the LBL Center for X-ray Optics
        (CXRO), see: http://www.cxro.lbl.gov/optical_constants/
    (d) the CCP4 (Collaborative Computational Project Number 4) in
        Protein Crystallography at Daresbury laboratory in UK
        (http://www.ccp4.ac.uk).

-- existence of many other scattered resources for scientific
computing,
    e.g. NIST and LLNL databases, SHADOW at the University of
Wisconsin,
    BioSAXS software at EMBL Hamburg, and  the attempts of the
    International Union for Crystallography to systemize them; see very
    long lists maintained by the IUCr at
    http://www.iucr.org/sincris-top/logiciel/
    and
    http://www.iucr.org/iucr-top/data/



 > In particular:
 >
 > 1. What are the limitations of current tools for
 >   x-ray data reduction, analysis, modeling, and simulation?
 > 2. What additional tools are needed?
 >
 > 3. How can the existing tools be improved?
```

>

Here I would suggest to draw a rough distinction between mostly software projects and those were the physical model is the dominant part in the development.

The first group would comprise software tools and databases that are based on well established algorithms and models. The examples could be some scientific visualization and analysis software like GRACE or FIT2D, databases of X-ray scattering factors, many macromolecular crystallography packages, and etc. The limitations in this group are: sometimes lack of good interfaces, installation difficulties on different computer platforms, poor documentation and the need of remote access in some cases. The improvements in this area would mostly require software engineering effort.

To the second group I would refer modelling and data reduction software based on recent or ongoing research projects. This is very challenging part in terms of making such software available to the APS community (see below about the difficulties), but no doubts that building such a pipeline between the most recent theoretical research and the experimental community at the APS would be the great way to improve the productivity of experiments at the APS.

Questions 1 to 3 in the survey mostly apply to the first group for which they are certainly important. With the second group those questions cannot be answered because the "tools" are not known yet -- they may be just emerging or may appear only tomorrow and no one can list what has not been discovered.

The most important question for the second group is how to work out a framework for quick interfacing new emerging scientific software tools. Some attempts of that kind have been made within the DANSE project. Namely, they suggested to wrap pieces of data analysis software written by different researches into Python scripts and that way to link them together even between different computer systems. Thus, the original data analysis code would not have to be rewritten from original language (e.g. Fortran or C) or ported from original operating system (e.g. Unix, or Windows, or Mac). This sounds attractive, but the procedure seems to presume that the scientific algorithm is something fully settled and the DANSE team only needs to replace the I/O interface. However, most of data analysis software is based on physical models and approximations that have their limitations and those limitations are not always clear until one starts getting some weird results. This is a critical distinction from software implementing pure mathematical algorithms (e.g. in crystallography). So, the programmer who modifies the original code will never be able to figure out what is wrong: he may see that the calculated reflection coefficient is e.g. 25, but it would not tell him anything because the "formula" was programmed correctly!

Therefore, I would suggest creating a framework with a closer involvement of original developers into providing a common interface to their software. The Scientific Software group that should be created at the APS needs to consist of both software engineers and X-ray theorists. The latter are needed in order to understand the models.

That group could develop some scientific software on its own, but it should also closely work with each scientific software provider on individual basis helping him/her to adapt his/her software to a common interface that needs to be defined within the framework. The ultimate goal should be to preserve the link between the original developer and the modified code, so that he/her would still have it under complete control, be able to monitor usage/bugs and refine the code by himself. I have very strong proofs based on my long term experience that preserving these links is vitally important for making the scientific software project efficient.

Thanks,

Sergey Stepanov

--
Sergey Stepanov, Ph.D.
Staff Scientist, GM/CA Collaborative Access Team,
Advanced Photon Source, Argonne National Laboratory
Email: sstepanov at anl.gov Voice: +1(630)252-0664
http://sergey.gmca.aps.anl.gov Fax: +1(630)252-0667

## Response from Dean Haeffner

1 & 2. There are so many limitations, that space will not permit anything approaching a complete list. First, I don't think data acquisition should be left out of this list. Many of our problems have to do with coordinating macros, detectors, sample environments, motions, etc. for collecting data, and this is by no means at a satisfactory state. Most beamlines do not log data in a satisfactory way, even though the tools basically exist. In many ways, it is much more important than the four listed, because the beamlines must perform that task, while the others can be off loaded to users to some degree.

The lack of a widely accepted data format is a persistent problem that has been the subject of talk, but little action for at least 12 years. Standard analysis tools (e.g., Fit2D) have largely been taken from other facilities (the ESRF in particular), and as such, are usually not tailored to the APS in way. A suite of robust analysis tools should be available, though some users will always prefer their own.
I think modeling and simulation is largely in the domain of the user, and the APS cannot hope to do much to make more than a few users happy. There may be a couple of exceptions to this, but generally I think we should focus on acquisition, reduction, and basic analysis.

3. A centralized effort is needed, presumably from the BCDA group. Tools that are developed need to be better publicized, and documentation (especially the webpage) must be much improved. No effort will please everyone, but some programs such as SPEC are so widely used that it is possible to help a lot of people. For the case of SPEC, many macros exist around the APS, and much could be gained from developing a way of sharing this software. Small focus groups on particular software packages might be very helpful. A few pizzas can go a long way to further communication. Also, to strongly matrix the BCDA group to beamlines would help. Clearly more personnel are needed before this can ever really be implemented well. A good goal would to be 0.5 FTE per sector.

4. More people doing software & hardware development directly working with beamline staff and users.

Dean

## Response from Matt Newville

I see two main areas that the APS could focus on to really help the Users need for scientific software:

First, the APS could help develop and encourage use of a common set of tools for staff and Users to write software in. I wouldn't exactly call it a Framework, but that might be close to the idea. The primary needs here are visualization and data handling, and so probably ARE closely related to data collection. I think the APS should not try to dictate a "Total Solution" by itself, but provide basic tools (and support and documentation for them) and to foster the development of a community for x-ray software.

Personally, I think the APS should encourage and promote the use of Python and C for this work, but that bridges to other languages must be available (or fairly easy for someone else to make). That is, I think Java (or .NET) would be poor choices for getting community involvement from User scientists. I think we should be looking beyond the x-ray and neutron communities as well (astronomy and computational chemistry, bio-informatics, and geographic information systems come to mind), as many of the tools needed are fairly common to many scientific fields. As a first priority, I think that embracing the challenge of a good data visualization package for 1-D, 2-D, and image data would be the right goal.

Reasonable starting points exist and are actively developed and supported in Python:
  http://scipy.org/
  http://matplotlib.sourceforge.net/

Second, and completely unrelated to the discussion above, the APS could support a "User Facility" level interface into modern computational chemistry and physics codes, such as for DFT calculations. Much of the work done at the APS could greatly benefit from having a computational component.

Several groups are doing such work now, but the barrier to these DFT and other codes can be pretty high. I'd guess that most people including such calculations in their work are collaborating with computational chemists. It would be easy to say this is how it should be, but just as the APS provides support labs to users for sample preparation and some detectors and other equipment for general use, it might be worth considering having a supported User Computational center for such high performance computations.

--Matt Newville <newville at cars.uchicago.edu>

## Response from Paul Zschack

Thanks for the opportunity for input.

Generally, it seems to me that area detectors which handle crystallographic data will become much more widely used. Not only the large CCD detectors, but smaller devices that mount to a goniometer, and will need to be integrated both for ease of use and data analysis. So, I suspect one of the more significant impacts one could have on data reduction/analysis would address and anticipate these needs.

The wide variety of commercially available data analysis software tools is a major problem. Too many different users have their way of doing things and this diversity is a distraction. If APS would select a single product platform and offer support for common tasks, and distribute common routines, this would become the platform to which all would eventually migrate (at least to some extent). But I see you've already recognized this.

In the area of general purpose and surface/interface scattering, the experimental needs are quite diverse. So, tools that allow easy access to and visualization of multi-dimensional (MCAs, areas or volumes) data in reciprocal space are important. This is generally not image data, so additional processing and reduction is required. For a user to know that the data collected is correct, the routines for visualization and comparison with modeling data are also essential.

Successful integration of data collection, extraction, and comparison to models in a straightforward approach that doesn't require specific expertise would make a significant impact to the productivity in our area.

With regards,
Paul

# Response from Brian H. Toby

> *1. What are the limitations of current tools for*
> *x-ray data reduction, analysis, modeling, and simulation?*

Powder diffraction (Rietveld) software is quite well developed and sophisticated, but it has been written with experts as intended users, so it tends not to do much input validation and does not do much "hand holding". It has not been well integrated into data collection on user instruments, so we frequently see the question "how do I convert data from instrument X to format Y."

> *2. What additional tools are needed?*

The ability to fit models against multiple techniques (EXAFS, NMR, Electron microscopy, quantum,...).

>
> *3. How can the existing tools be improved?*
>

The programs themselves are monolithic and this limits science. We now want modular software so that analysis can be knit together as needed by the experiment. Further, existing codes are written in Fortran-77, which is very hard now to maintain or expand. It is time to think about starting a new analysis package from the ground up.

> *4. What will most affect the scientific impact of your work?*

Impacts should be very significant. New codes that designed to be used by non-experts will open the technique up to a very wide range of users. Ability to work on problems where diffraction data alone are inconclusive, but where other physical characterization data can be modeled will allow for much new science to be done on very complex, on poorly ordered or small crystallites problems that are frequently intractable now.

Brian

********************************************************************
Brian H. Toby, Ph.D. office: 630-252-5488
Materials Characterization Group Leader, Advanced Photon Source
9700 S. Cass Ave, Bldg. 433/D003 work cell: 630-327-8426
Argonne National Laboratory secretary (Kristy): 630-252-5453
Argonne, IL 60439-4856 e-mail: brian dot toby at anl dot gov

## Response from Ersan Ustundag

Dear Pete and Ken,

I hope you can still accept my response (see below). I have been in Turkey over the past few weeks for a somewhat urgent family trip and did not get a chance to communicate with you in more detail. Just before I left, we had some concrete ideas about specific steps we can take to demonstrate the feasibility of a seed (pilot) project. I hope to discuss those with you upon my return to the US next week.

By the way, I am definitely planning on attending the workshop. All the best...

Ersan

--- Ersan Ustundag ---

---

*1. What are the limitations of current tools for x-ray data reduction, analysis, modeling, and simulation?*

The current tools are mostly ad-hoc solutions to immediate (and sometimes not-directly related) problems and cannot be easily integrated for efficient work.

*2. What additional tools are needed?*

Basically, we need an integrated software that will allow COMPLETE experiment simulation (including realistic x-ray optics and sample contributions), fast and real-time data analysis and more importantly integration of advanced materials models.

*3. How can the existing tools be improved?*

We have to use a common platform (such as Eclipse or Pyre) to bind existing tools and then add new code for further analysis and modeling.

*4. What will most affect the scientific impact of your work?*

Sophisticated integration of models to data analysis would be the most crucial step. This should be done in a way to guide data collection via modeling.

## Response from Andy Howard

Hello survey organizers,
> *We have been asked by the XSD Division to organize a workshop to*
> *determine our fundamental needs and opportunities in scientific*
> *software systems for x-ray data reduction, analysis, modeling and*
> *simulation. The workshop has been scheduled for August 29, 2006*
> *at the Advanced Photon Source.*
>
> *In order to prepare for this workshop we would like your input on*
> *what you see as the needs and opportunities for scientific*
> *software development at the APS and in the X-ray community, as*
> *well as information that would support making a funding proposal*
> *for such resources. In particular:*
>
> *We realize you have a busy schedule and appreciate your taking*
> *time to address these issues. We need the responses by*
> *Thursday, June 22. Please direct your responses to:*

I'm sorry I didn't respond sooner. I was caught up in preparations for the ACA Summer School (acaschool.iit.edu). I spoke briefly with Pete Jemian on Wednesday, and he said that some responses would be useful even at this late date. Prof. Wang had also asked that we at SER-CAT pool our recommendations, and my non-response was partly a matter of an inability to get together with the other potential responders from SER-CAT.

> *1. What are the limitations of current tools for*
> *x-ray data reduction, analysis, modeling, and simulation?*

Within macromolecular crystallography the tools for these purposes are actually fairly mature, although of course there are opportunities to make them better. As a developer of crystallographic data processing software I have a fairly major stake in improving what's out there. Outside of crystallography I wouldn't pretend to any particular expertise, apart from a bit of hard-earned experience with beamline control systems.

Some of the most serious problems are integrative rather than algorithmic. Getting data obtained in one context (e.g. fluorescence-scan information) into another context (e.g. the conduct of a multiwavelength anomalous diffraction experiment) is often cumbersome, and the routes required vary from one beamline to another or even from one experimental situation to another. I'm pessimistic that any group could enforce standardization of such integrative tools in a way that would be widely adopted even within one synchrotron, much less in a community-wide way.

Another issue is commercialization. Many of the most useful tools for crystallographic data manipulation are commercial, and licensing is often cumbersome and expensive. For each sector or beamline to have to negotiate contracts for functional software is consumptive of time and funds. If there were a way that the entire facility could take charge of such contractural issues, the resident users would be able to devote their human and fiscal resources toward meeting other needs.

> *2. What additional tools are needed?*

Integrative tools (see above); tools for unbiasedly comparing the results derived from one crystallographic facility with those obtained at another.

> *3. How can the existing tools be improved?*
Site licenses (see above); more conversations between developers and users to determine optimal upgrade paths.

> *4. What will most affect the scientific impact of your work?*
The flip answer I could give is "more rapid adoption by other beamlines of my software," but a better answer would be "clear-eyed evaluation of data quality for crystallographic data sets derived from multiple beamlines."

Warmly,
Andy Howard


```
  _____________________________________________________
/ Andrew J.Howard, Assoc. Professor of Biology, CSRRI \
| Biological, Chemical, & Physical Sciences Department |
| Coll.of Sci.&Lett., Illinois Institute of Technology |
| 3101 South Dearborn Street, Chicago IL 60616 USA     |
| Co-director, IIT Masters in Health Physics program   |
| phone: 312-567-5881;     fax: 312-567-3576           |
| e-mail: howard at iit.edu;  web: csrri.iit.edu/~howard/ |
  \_____________________________________________________/
```