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?