

Report on the ACA Best Practices for Data Analysis and Archiving SIG meeting, Jun 13 2023

The officers of the SIG are:

Chair 2023: Aaron Brewster

Chair-elect 2023/Chair 2024: Christine Beavers

Secretary-Treasurer 2022-2023: Charles Hespen

Chair 2022: Lee Daniels

During our discussion, a few broad themes were developed, summarized below and elaborated in the minutes:

1. We want to more easily capture and deposit the full path from raw data to figures in a paper. This aligns with our charter, “to further the advancement of all aspects of the application of Best Practices for Data Analysis and Archiving to structural science and to promote communication between persons interested in such techniques and results.” This was developed into a new session, adapting the previous session on Validating Models from the Data
2. We specifically want to be inclusive of X-ray, neutron, and electron data in our sessions and planning
3. We want to encourage new SIG members to join! To this end, we will be doing more recruiting during the ACA proper

Attendance:

- Aaron Brewster
- Frances and Herbert Bernstein
- Alexandra Giglia
- Lee Daniels
- Brandon Mercado
- Joseph Ferrara
- Jeanette Krause
- John Helliwell
- Nicholas Sauter
- Yaohua liu
- Marian Szebenyi
- Cora Lind-Kovacs
- Sarah Bowman
- Peter Mueller
- Andy Howard
- Steven Sheriff

Minutes

- Aaron Brewster: introduced the session and outlined the agenda
 - Read out the bit of the charter cited above

- For the past years, our task has been to bring people together to treat difficult data. It has been great to see these ideas produce valuable sessions and discussion.
 - 2022 minutes were approved by the attendees
- Alexandra Giglia, Journal Manager of Structural Dynamics: presented new scope for Structural Dynamics, that it's more of a home for the ACA, as now articles can contain structure or dynamics instead of requiring both. Open special topics right now include:
 - The Advent of Ultrafast X-ray Absorption Spectroscopy
 - Enabling Methods in Ultrafast Electron Diffraction, Scattering and Imaging
 - Imaging ultrafast structural dynamics in attosecond and strong-field spectroscopy
 - A decade of biology with XFELs: The 10th Annual BioXFEL International
 - Celebrating the work and achievements of Keith Moffat
- Brandon Mercado, planning committee: thanked SIG members. Noted that the Google form for new sessions should be submitted by June 30th.
- AB: nominations for new leadership positions. We need 2025 chair-elect and the 2024-2025 secretary/treasurer
 - No nominations/volunteers
 - Nick Sauter proposed a nomination committee
 - Nomination committee
 - Nick Sauter
 - Joseph Ferrera
 - Christine Beavers
- AB: sponsored sessions for 2023:
 - **1.1.2 Data Analysis Software**, July 8th, 8:30am-11:30am, Kent A-C
Session Chair(s): Herbert Bernstein & Marian Szebenyi
(Sponsor)
 - **1.1.5 Crystal Growth & Optimization**, July 8th, 8:30am-11:30am, Waterview AB (Co-sponsor)
Session Chair(s): Gabby Budziszewski & Xu Liu
 - **2.1.2 Structures from Artificial Intelligence**, July 9th, 8:30am-11:30am, Laurel CD (Sponsor)
Session Chair(s): Melanie Vollmar & Jennifer Wierman
 - **2.2.5 Serial Crystallography**, July 9th, 2pm-5pm, Kent A-C (Sponsor)
Session Chair(s): Iris Young, Petra Fromme
 - **3.1.2 Quantum Crystallography I**, July 10th, 8:30am-11:30am, Kent A-C (Co-sponsor)
Session Chair(s): Yu-Sheng Chen, Florian Kleemiss & Krzysztof Wozniak
 - **3.1.6 Validating Models from the Data, Other Data, and Theory**, July 10th, 8:30am-11:30am, Laurel AB (Sponsor)
Session Chair(s): Joseph Ferrara & John Rose

- **3.2.1 Quantum Crystallography II**, July 10th, 2pm-5pm, Kent A-C (Co-sponsor)
 - Session Chair(s): Krzysztof Wozniak, Florian Kleemiss & Yu-Sheng Chen
- **4.2.3 Computational techniques for SAS**, July 11th, 2pm-5pm, Waterview CD (Co-sponsor)
 - Session Chair(s): Thomas Weiss & Steve Meisburger
- NS: discussed post publication peer review, and the difficulty of reproducing figures, which is a continuing concern. Data is deposited to the pdb, but it doesn't accept the workflow information.
- AB: noticed trends: in his early SIG meetings, conversation was along the lines of how to create data (images/coordinates) that is FAIR. Now though things may be in a better place: for example, NeXus adopted by Dectris. Now problems are 1) processing at speed and manage large datasets 2) multi-modal experiments and difficult to treat data, where it's hard to write down the path from the raw diffraction data to a figure in the paper
- John Helliwell: UK responsibility for reproducibility rests with the facility. Raw, processed, and derived. US: responsibility rests with the experimenter. LCLS was setting up a working group (Mike Dunne) to look into these sorts of issues, and Andy Gertz (ESRF) is giving a keynote at IUCr.
- Andy Howard, three words that can help define the problem/guide discussion:
 - Provenance: do we know that the data have been securely captured where they are not being manipulated
 - Workflow: how the data moves through the processing software packages
 - Metadata: beyond the diffraction data to crystallization and storage and so forth
- AB: the data to figure route is often long and involved (kind of the provenance idea)
- Brandon Mercado: new session along Joe and John's session. Possibly a workshop. Present case studies, how data was archived securely, workflow, metadata, and figure reproduction
- AB: spectroscopy in XFEL as a multi-modal data example. Another example is jupyter notebooks published alongside papers to reproduce figures.
- AB: suggest general session of data validation and archival issues, second part as several detailed use cases. Use Andy's words: Provenance, Workflow, Metadata.
- AB: do we want to propose all of the above sessions this year?
- Joseph Ferrara: Small Molecule SIG is looking at another Quantum Crystallography session Brian Patrick working on organizing that
- BM: asked for 2023 sponsor vs. co-sponsored
- AB: added that info to slides

- AB: proposed specifically that the AI session continues to be sponsored by the Best Practices SIG, so that the best practices ideas continue to be part of the discussion there
- AB: asked JF and John Rose if they thought that the new full data session would supplement, change, or be in addition to the Validating Models from the Data Session
- JH: discusses who is responsible for making these data available, and even further, just who owns the data. Is a joint responsibility. The central facility challenge is to link the subsequent calculations by the user back to the facility. Trend now is to do as much as possible at the facility. Also noted that electrons and neutrons are missing from these conversations.
- AB: that list bit goes to recruiting. Will encourage session chairs to add 5 minutes of recruiting to SIGs in general and Best Practices SIGS specifically to the start of their sessions.
- JH: Best Practices may be a neutral actually so neutrons and electrons might actually be included
- AB: metadata is quite different in some senses, and people only now starting to get a handle and rationalize the differences
- BM: back to new session, full day, with session descriptions broken up. Energy source could be encapsulated in session description (electron, neutrons, photons). Responsibility of where data is archived could be useful to discuss. Maybe a panel discussion would be useful.
- NS: "Provenance, workflow, and responsibility for X-rays, neutrons, and electrons", possible title. Experimental side in the morning, facility side in the afternoon.
- AB: word responsibility is great because it is so nebulous, particularly JH's comments seem relevant. I always assumed it was 100% my responsibility without any support from the facility
- JH: complication is that data becomes open after 3 years in Europe. JH wants to take responsibility, but in the end it's the facilities decision.
- JR: one day session: invite IUCr validation of small molecules to the full day session. JR to JH, what about UK?
- JH: UK has different policy but may move to UK's policy
- Lee Daniels: need more young SIG members! Maybe we need more advertising for SIGs in general. Need to lower the average age for SIG membership
- BM: maybe can encourage Etter student award winners to join? BM will take this to the planning committee
- LD: it's so easy to join!
- BM: specifically reach out to the Best Practices Etter award winner to join
- AB: will send note to session chairs to get them to advertise the SIG
- Peter Mueller motioned to adjourn, but second didn't show up in zoom

Additional chat comments:

- AH: Tom Irving, who runs the SAS facility at the APS, is about 15 meters from me.
- JF: I think the merger plan is a good idea to create a full day session.
- JH: Just who does own research data?
<https://forums.iucr.org/viewtopic.php?f=39&t=445>
- AH: It's worth recognizing that the user-responsibility concept in the US was driven partly by the participation of the pharma companies. They (supposedly) don't want the data to be open.
- JH: Also in Europe the pharma do get rights because of payment. ie confidentiality.