

第 12 回 NOBUGS 国際会議 (NOBUGS2018) 報告

公益財団法人高輝度光科学研究センター
情報処理推進室 松本 崇博

1. はじめに

今回で 12 回目となる New Opportunities for Better User Software (以下、NOBUGS) の国際会議が 2018 年 10 月 21 日から 26 日にかけて、米国ニューヨーク州のブルックヘブン国立研究所 (BNL) にて開催された¹⁾。NOBUGS は、世界の放射光施設、中性子施設及びミュオン施設に関わる科学者と IT 技術者のコラボレーションを促進することを目的として開催されている。今回の NOBUGS2018 では以下に示すテーマが設定され、実験ユーザー向けソフトウェアに関する話題が中心として取り扱われた。

- Data Catalogs/Electronic Notebooks
- Data Streaming
- Use of (commercial) Cloud Systems
- Web Tools
- Workflow Engines & Tools
- Other Relevant Topics

本会議は参加者が 82 名、口頭発表が 23 件及びポスター発表が 26 件と小規模であったが、NSLS-II、European XFEL、ESRF、Diamond、MAX VI、APS、ISIS、ESS など、米国や欧州の主要な放射光施設や中性子施設から幅広い参加があった。日本からは SPring-8 から出席した筆者の他、J-PARC MLF (物質・生命科学実験施設) 関係から 4 名の参加者があった。NOBUGS2018 では少人数でパラレルセッションもないことから、会議の中で参加者と次第に打ち解けていきやすい雰囲気があった。またプロシーディングスの提出も必要ないことから堅苦しい雰囲気もなく、気軽に意見交換しやすい場である所がユニークな特徴であると感じた。筆者は SPring-8 においてオープンデータに向けた開発研究を進めているが、現在抱えている課題について他施設の方々と遠慮なく率直に

意見交換できる機会も得られ、貴重な経験となった。本会議では 3 つの基調講演も行われたが、全て IT 企業の方々の講演であった。Continuum Analytics の共同設立者の Peter Wang 氏による Python とデータ分析の講演もあり、科学者と IT 技術者の間で活発な交流が行われていた。

本会議は、2014 年 10 月から稼働を始めた新しい放射光施設 NSLS-II の近くにある Computational Science Initiative で開催された。BNL は SPring-8 と比べても広大であり、車での移動が前提となっている。筆者は BNL 内の宿舎に宿泊していたが、宿舎から NOBUGS2018 の会場まで徒歩 20 分程度かかった。車なしでは BNL から抜け出すことが困難であり、雑念にとらわれることなく会議に没頭できる環境にあった。BNL 施設内は消防署や警察署もあるなど施設全体が街を形成しているのが印象的であった。一方、施設内では七面鳥や鹿など野生動物に触れる機会もあり、豊かな自然を感じることもできた。



図 1 NSLS-II 施設の前で撮影した NOBUGS2018 参加者の集合写真

2. 会議報告

ここでは、本会議において筆者が印象に残った実験データ収集やデータ解析ソフトウェア、及びオープンデータに関する話題について紹介する。

世界の放射光施設や中性子施設では実験計測で用いるフレームワークの標準化が進められている。標準化されたフレームワークを施設内の複数のビームラインに横断的に導入することで、実験計測システムの高性能化が効率的に進められている。代表例としては Tango^[2]や EPICS^[3]といった制御フレームワークが挙げられる。これは SPring-8 における MADOCA^[4]と同様、大規模な制御システムにおいて、分散ネットワーク化にある機器やアプリケーションの制御を統一した手続きで簡易に行う枠組みである。

データ収集に関しては、このような標準化されたフレームワークとリンクして利用できるソフトウェアが数多く報告された。一例として ESRF で開発が進められている Bliss^[5]の報告があった。これはビームライン実験計測でよく使われている spec^[6]を改良したもので Python をベースに作られている。Tango と連携することもでき、Continuous Scan や Step Scan、データマネージメント等が柔軟に行えるように機能拡張されていた。ESRF では spec から Bliss へと全面的に更新することを決断し、2016 年より Bliss の導入を進めている。ESRF で準備が進められている次期放射光計画 EBS 移行時のシャットダウン中（2019～2020 年）にビームラインの半数を Bliss に置き換える予定とのことであった。また、NSLS-II において導入が進められている Python ベースのデータ収集フレームワーク bluesky^[7]の報告もあった。bluesky は EPICS と連携しており、GPFS をビームライン共通のファイルストレージとして用いている。Python を用いることにより、機器制御やデータ収集、及びデータマネージメントの高度な処理が簡易に行えるように工夫されていた。

実験データを保存するためのデータフォーマットも標準化の対象として重要であり、後で述べるオープンデータでも関わる項目である。欧州では PaNdata^[8]において放射光施設と中性子施設における実験データの取り扱い方針を定めているが、NeXus^[9]を共通のデータフォーマットとして使うことを推奨している。NeXus は日本では利用例が少ないが、海外では多くの施設で広く活用されている。NeXus は HDF5 データフォーマット^[10]をベースとして用いており、様々な型のデータやファイルを、階層構造を持たせてひとま

とめに行うことができる。HDF5 そのままでは自由度が高いため、NeXus ではさまざまな実験計測で共通に利用するための定義を設けている。

データ可視化やデータ解析ソフトウェアにおいても、NeXus とリンクして利用できるものが多く報告された。一例として DESY で開発が進められている 2 次元画像可視化ソフトウェア LaVue^[11]の報告があった。LaVue では NeXus を含む多様なファイルフォーマットに対応すると共に、HiDra^[12]と呼ばれる高速データ転送ソフトウェアや、2 次元検出器データ収集フレームワーク Lima^[13]など、多様なインターフェースと連携できるように工夫されていた。他には米国の研究機関が共同で開発を進めている XiCam^[14]の報告があった。XiCam ではローカルやリモートにある多様なフォーマットのデータにアクセスし、SAX/GISAX や XAS、Tomography 等さまざまなデータ解析をプラグインにより柔軟に行えるように実装されていた。

画像データ解析では大容量データ処理の手法も重要となる。Coherent Diffraction Imaging を高速処理するソフトウェア^[15]や、機械学習を活用し 2 次元検出器画像で類似した領域ごとに分割を行うソフトウェア SuRVoS^[16]の報告もあった。本会議で報告があったこれらデータ収集やデータ解析ソフトウェアは SPring-8 でも活用できそうなものも多く見られた。SPring-8/SACLA 内で広く情報共有し、積極的に利用検討を進めていく必要性を感じた。なお、報告があったソフトウェアのほぼ全てが GitHub や GitLab を用いてコード管理をしており、外部公開されていた。筆者も SPring-8 内で GitLab の運用を行っているが、今後このようなソフトウェア管理・公開の取り組みがより重要になると思われる。

オープンデータに関しては、本会議で最もアクティブなテーマであったように思う。各施設からの報告があったが ESRF での取り組みが最も進んでいるように感じられた。

オープンデータでは FAIR の概念が重要視されている。FAIR とは Findable (見つけられる)、Accessible (アクセスできる)、Interoperable (相互運用できる)、Re-usable (再利用できる) の頭文字をとったものであり、オープンデータで利用できるデータについて定義している。FAIR の条件を満たすためには、実験計

測におけるサンプル情報や実験条件などをメタデータとして定義し、メタデータカタログ(データベース)に登録する必要がある。このメタデータカタログを通して、必要なデータを検索しアクセスできるようになる。

メタデータカタログには ICAT^[17]を用いることが PaNdata の方針となっている。しかしながら、本会議では新しいメタデータカタログの開発例がいくつか報告された。これら新しいメタデータカタログでは、実験毎に異なるメタデータを柔軟に扱い、より現実に即した利用を行うため、スキーマレス DB を用いていた。例としては MAX VI や ESS などが開発されている MongoDB を用いた SciCat^[18]、BNL における Elasticsearch を用いた開発例の報告があった。筆者が SPring-8 において開発を進めている実験データ転送システム BENTEN^[19]でも Elasticsearch を用いたメタデータ管理を行っているが、今後、海外の動向も注視して開発を進めていく必要性を感じた。

研究データにデジタルオブジェクト識別子 (DOI)^[20]を付与する取り組みについての報告もあった。研究論文への DOI の付与は既に浸透しているが、研究データにも DOI を付与する動きがここ最近進んできている。研究データに DOI を付与するメリットとしては、研究論文によるデータ引用が挙げられる。データ公開者にとっても自分が公開したデータがどの研究論文で利用されているのか把握できるようになるメリットもあり、研究データへの DOI 付与により新しい文化が生まれることが期待されている。

オープンデータの開発は各施設において急ピッチで進められているが、ESRF では研究データのオープンアクセスのための Web ポータルを立ち上げ、運用を開始したとのことであった。先に述べた研究データに DOI を付与するためのインターフェースの整備も進んでいた。オープンデータにおいてはデータアクセスのためのデータポリシーも重要となるが、欧州ではエンバーゴの期限 3 年を過ぎたのちはデータが一般公開されることになっている。

他、クラウド利用もオープンデータの話題でよく議題になる項目であるが、本会議では J-PARC MLF においてデータ配送及びデータ解析ファームで試験的に利用している事例のみ報告があった。J-PARC の担

当者に伺ったところ、J-PARC MLF では JAEA が運用していることからセキュリティ対策に非常に注意を払っており、オープンデータのため外部から施設内の計算機に直接アクセスするのは避けたい、という意向があるようであった。このため、クラウド上にデータ転送しオープンアクセスする手段について積極的に試験を進めているようである。クラウドでデータ利用する際にはデータダウンロード時に課金が発生するが、現在は扱うデータ量が少ないこともあり施設側で負担しているとのことであった。今後、大容量データのストレージやデータ解析でのコスト低減のためクラウド利用がより注目されると考えられるが、実際の利用が進むにつれ運用手法がより確立していくだろう。

会議の最終日には NSLS-II の施設見学があり、ビームラインや制御室を見学した。NSLS-II は運用を始めたばかりであるため、施設は新しくとても綺麗であった。中央制御室では情報掲示用に大型のタッチパネルディスプレイが整備されていた。施設内には 24 のビームラインが稼働しているが、BNL で開発された Python ベースのデータ収集フレームワーク bluesky が各ビームラインに導入され、実際に便利に活用されているところを拝見することもでき、印象に残った。



図2 施設見学で訪れた NSLS-II ビームライン

3. おわりに

会議後半 10 月 25 日、26 日は NOBUGS2018 のサテライト会議があり、NeXus の International Advisory Committee (NIAC)^[21]にオブザーバーとして参加する機会を得た。NIAC では NeXus データフォ

ーマットに関する規則を定めたり、NeXus 利用を促進するなどの活動を進めている。会議では NeXus で現在課題になっている項目が挙げられ、今後どのような方針で進めていくのか議論がされていた。SPring-8 では、まだ NeXus の利用事例はないが、NIAC で議論されていた項目は今後データフォーマットをどのように定義して利用するのか考えていく際にも有用であり大変参考になった。筆者はオープンデータの取り組みを進めていることから NIAC に興味を持ち、NOBUGS2018 の後、NIAC のメンバーとして正式に承認されることになった。今後、NOBUGS2018 で顔なじみになった他施設関係者の方々と連携し、NeXus の利用も念頭に置きながらオープンデータやその他の開発研究を精力的に進めていきたい。

次回の NOBUGS は DESY がホストし、European XFEL があるドイツ・ハンブルグにおいて開催される予定である。

参考文献

- [1] <https://www.bnl.gov/nobugs2018/>
- [2] <http://www.tango-controls.org/>
- [3] <https://epics.anl.gov/>
- [4] 古川行人 他:SPring-8/SACLA 利用者情報 **19**(2014) 392-395.
- [5] <http://www.esrf.eu/computing/bliss/doc/bliss/>
- [6] <https://certif.com/spec.html>
- [7] <https://nsls-ii.github.io/bluesky/>
- [8] <http://pan-data.eu/>
- [9] <https://www.nexusformat.org/>
- [10] <https://support.hdfgroup.org/HDF5/>
- [11] <https://confluence.desy.de/display/FSEC/LaVue+-+Live+Image+Viewer>
- [12] <https://stash.desy.de/projects/HIDRA/repos/hidra/>
- [13] <http://lima.blissgarden.org/>
- [14] <https://www.camera.lbl.gov/xi-cam-interface>
- [15] <https://github.com/advancedPhotonSource/cdi>
- [16] <https://diamondlightsource.github.io/SuRVoS/>
- [17] <https://icatproject.org/>
- [18] <https://github.com/SciCatProject/>
- [19] T. Matsumoto *et al.*, “Experimental Data Collection and Data Access Software Through Internet at SPring-8”, AIP conference proceedings (SRI2018)
- [20] <https://www.doi.org/>
- [21] <https://www.nexusformat.org/NIAC.html>

松本 崇博 MATSUMOTO Takahiro

(公財) 高輝度光科学研究センター 情報処理推進室
〒679-5198 兵庫県佐用郡佐用町光都 1-1-1
TEL : 0791-58-0802 ext 3270
e-mail : matumot@spring8.or.jp