# Workshop Overview: Data Science and Computation for Rapid and Dynamic Compression Experiment Workflows at Experimental Facilities

Christine Sweeney, CCS-7

February 1, 2021

LA-UR-21-20754

# Motivation for Workshop

- Compression experiments important to LANL core mission.
- Scientific community sometimes frustrated at the pace of discovery.
- Data analytics for other experimental regimes are advancing.
- Now is the time as facilities are upgraded and coming online.
- Advances in data analytics and computational resources could push compression experiment discoveries to a new level.

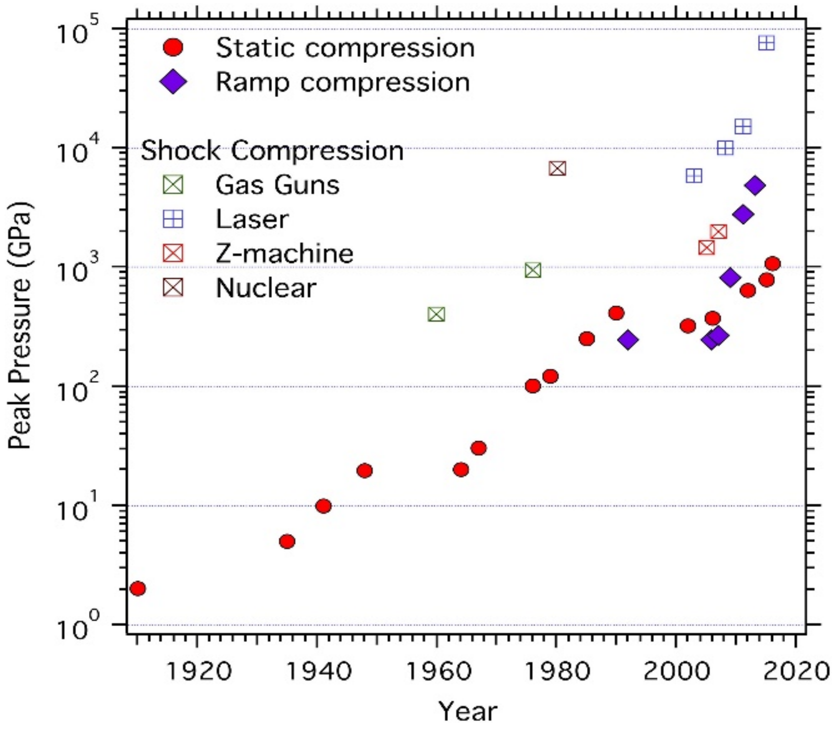
# High Pressure Applied to Materials Enables Discoveries

Science Drivers:

- Earth and planetary science
- Measure properties of materials
- Synthesize new materials
- Understand behavior of materials at extreme conditions

High Pressure Techniques:

- Diamond Anvil Cell (DAC)
- Dynamic DAC (dDAC)
- Laser
- Gun
- High-explosive drivers
- Pulsed power: Thor, Z Machine




*Progression of pressures achieved in the laboratory as a function of year for shock (open squares), static (red circles), and ramp-compression techniques (purple diamonds). T.S. Duffy and R.F. Smith, "Ultra-High Pressure Dynamic Compression Materials," Front. Earth. Sci. 7(23), doi:10.3389/feart.2019.00023 (2019).*

# Data Science Drivers

- Upgraded facilities enable more compression experiments and more diverse diagnostic techniques.
  - Eg. Higher brilliance X-rays in upgraded facilities enables more imaging techniques
- Short pulses enable time-resolved measurements.
- Increased pulse repetition rate and faster detectors enable data rates on order of MHz.
  - Eg. dDAC experiments in 48-72 hour beam time, >10,000 images, 400-900GB
  - Eg. 10 Hz shock studies will produce even larger datasets.
- Development of novel data science techniques such as machine learning (ML), statistics, imaging, visualization
- Leveraging HPC platforms and acceleration hardware such as graphics processing units (GPUs)

# Overview

- Workshop Title: **Data Science and Computation for Rapid and Dynamic Compression Experiment Workflows at Experimental Facilities**
- Workshop Dates: September 8-11, 2020 (3 days workshop, 1 day writing)
- Workshop Venue: Virtual (Webex)
- 95 registered scientists and analytics experts, 31 invited talks, 4 lightning
- 15 universities, 9 U.S. national laboratories, 1 industry, 5 U.S. and European X-ray light sources, LANSCE and NIF
- Institutions: LANL, SNL, LLNL, LBL, ANL/APS, BNL, ORNL, PNNL, SLAC, EuXFEL, DESY, Carnegie Mellon U., Washington State U., U.C. Davis, U.C. Santa Cruz, Minnesota, Brigham Young U., U. of Oxford, Georgia Tech U., Princeton U., U. of Michigan, U. of Illinois, Stanford U., Stony Brook U., U. of Edinburgh, U. of Utah

# Organizers

Christine Sweeney, Chair, Los Alamos National Laboratory

**Steering Committee:**
- Blake Sturtevant, Los Alamos National Laboratory
- Christopher Biwer, Los Alamos National Laboratory
- Cynthia Bolme, Los Alamos National Laboratory
- Rachel Huber, Los Alamos National Laboratory
- Larissa Huston, Los Alamos National Laboratory
- Emma McBride, SLAC National Accelerator Laboratory
- Lowell Miyagi, University of Utah
- Clemens Prescher, Deutschen Elektronen-Synchrotron
- Kyle Ramos, Los Alamos National Laboratory
- Jesse Smith, Advanced Photon Source, Argonne National Laboratory

**Institutional Host:** Richard Sheffield, Los Alamos National Laboratory

# Overview of Workshop Agenda

**Day 1**

- *Welcome:* A. Taylor, LANL
- *Keynote:* N. Velisavljevic, LLNL, APS, HPCAT
- *Plenary:* Materials Project (M. Horton, LBL)
- *Plenary:* BES Data Pilot (H. Krishnan, LBL)
- *Panel:* Common Data and Analytics
- *Session:* X-ray Diffraction (XRD) and Scattering Data Processing

**Day 2**

- *Session:* Time-dependent Measurement Using Static Apparatus
- *Session:* Image Processing
- *Lightning Talks*
- *Session:* Machine Learning
- *Session:* Visualization

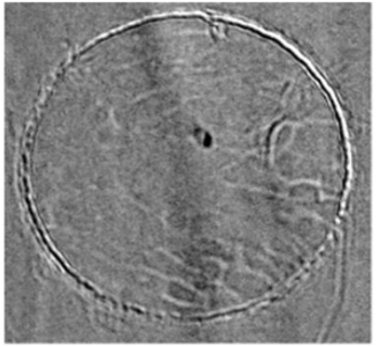
**Day 3**

- *Session:* Real-time Analytics Workflows
- *Session:* Post host Computational Tools
- *Plenary:* Detectors (Z. Wang, LANL)
- *Panel:* Facility Computational Resources
- *Wrap up*

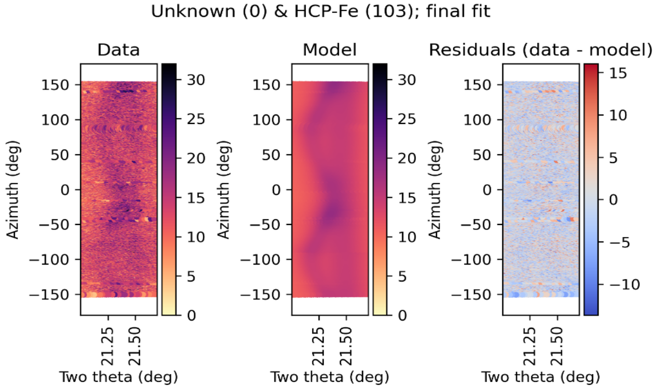
# Data Analysis Objectives and Tools

## Data Types and Processing for Compression Experiments

- Current data types and analytics:
  - X-ray diffraction (XRD), imaging, Velocity Interferometer System for any Reflector (VISAR), oscilloscope traces, pyrometry and other datatypes
  - Determine pressure reached, frame number of a phase change, nucleation, growth and melting
- Gaps:
  - Cannot process data in real-time to inform next exp.
  - Cannot process large volumes of data
  - Some tools too complex to use at beam time
  - Hard to correct for artifacts quickly and accurately
- Investments:
  - Portable/modular analysis packages, batch analysis
  - Forward modeling to help with non-unique data interp.
  - R/T analysis tools to support rapid assessment



**Crystallization of Ga from Pioneers in Cameras and Optoelectronics (PCO) high-speed camera; image processing as part of a DESY/LLNL collaboration**

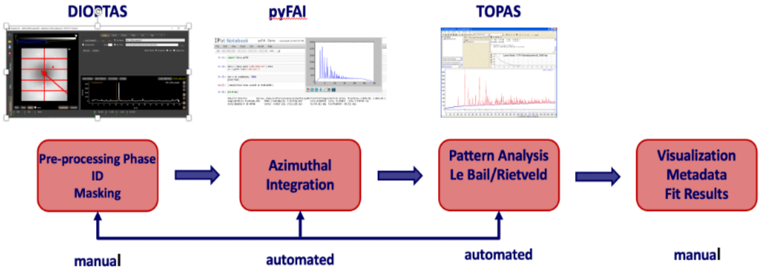



**The diffraction of X-rays at specific angles produces peaks like those in the data (left). Fitting these peaks with improved models (model fit, center; residuals, right) advances our understanding of how matter behaves. The model we proposed here uses Fourier series to fit the data in a single step rather than using the multistep process. Images courtesy of Simon Hunt and Danielle Fenech.**

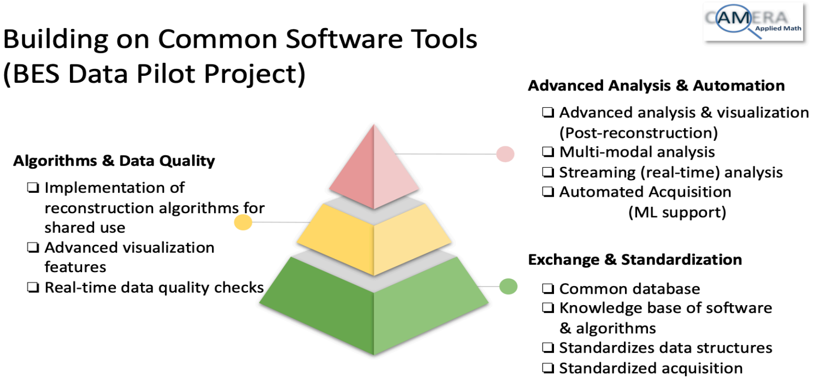
# Data Analysis Objectives and Tools

## X-Ray Diffraction Analysis Tools

- Tool capabilities:
  - Optimized for technical capabilities (not speed or usability) due to lack of good data quality
  - Packages perform image integration and simple fits or Rietveld refinement
- Gaps:
  - Must switch between various tools
  - Algorithms too slow for automated processing
  - Metadata not available
- Investments:
  - Machine learning (ML) as a route to fast approximate analysis
  - Protocols for interoperable tools




Data analysis approach for reaction of feldspars. Lars Ehm.

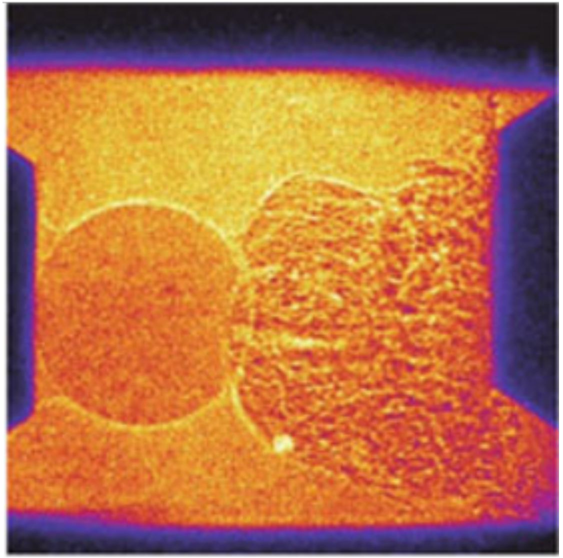



BES Data Pilot Project diagram. Image courtesy of Hari Krishnan (Lawrence Berkeley Laboratory).

# Data Science Techniques

## Image Processing

- Observations supported by imaging now:
  - Void collapse, complex wave interactions, melting/crystallization, crack propagation, nucleation, density
  - Can use GPUs for fast image processing
- Gaps:
  - Challenges posed by phase wraps, spatial resolution, large data volumes, sparse datasets
  - Many codes not available to larger community
- Investments
  - Feature tracking via image analysis,
  - Phase retrieval for complex index of refraction and density
  - Physics-based models to interpret sparse datasets
  - Lensless reconstruction techniques and ML techniques
  - Algorithm acceleration
  - Multimodal measurements (eg. phase contrast imaging w/ XRD)



**Dynamic multi-frame X-ray phase contrast imaging (XPCI) consisting of a steel projectile impacting three BS spheres at a velocity of 0.237 km/s. The times relative to impact are shown in the false color, where black represents complete absorption of the X-rays. Experiment at APS. B.J. Jensen et al.**

# Data Science Techniques

## Advanced Analysis, Modeling and Machine Learning

- Current techniques:
  - Analysis and simulation of XRD, (new) crystal structure refinements being added, mixed phases
  - Combining forward models, emulators and uncertainties
  - ML for classifying crystal structures, individual grains, size distribution, image reconstruction
- Gaps:
  - Obtaining data for advance analytics and ML
  - Non-unique predictions from models
  - Cannot analyze complex structures
- Investments:
  - More development and integration of models, uncertainty quantification, ML
  - Multimodal data analysis
  - Inclusion of mixed phases and mixed crystallographic states

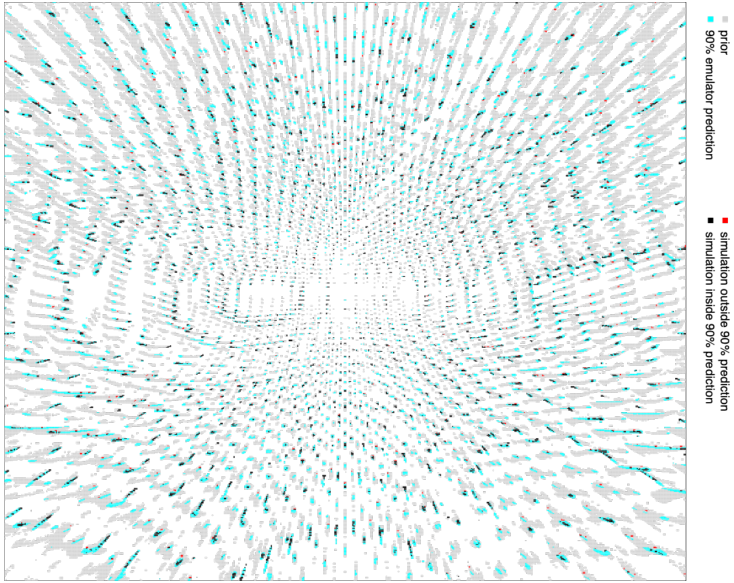
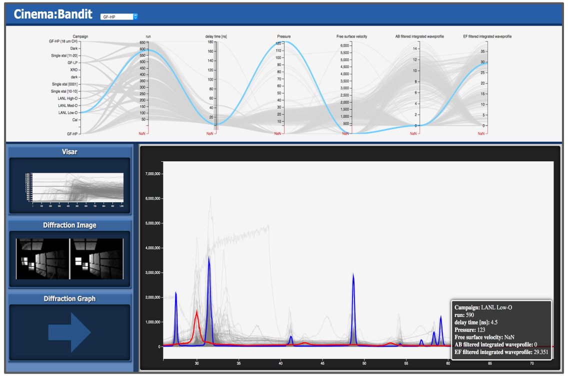



**Figure 7. Diffraction image holdout prediction shows an accurate prediction by Flag with the DiscoFlux hydrocode emulator and BarberShop XRD simulator. D. Francom et al.**

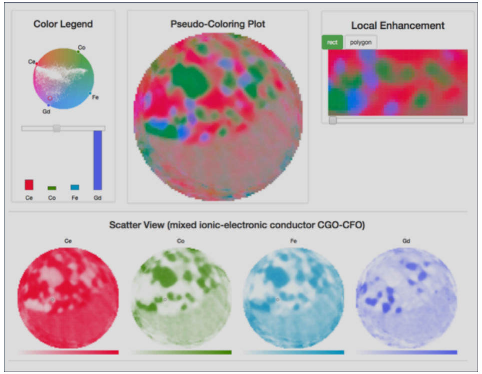
# Data Science Techniques

## Visualization

- Techniques used:
  - Exploring high-dimensional data and finding anomalies
  - Incorporating ML for feature detection
  - Data-driven color mapping for X-ray images
- Gaps:
  - Many tools not in common use
  - Data output streams not effectively input to tools
- Investments:
  - Interdisciplinary teams to come up with novel solutions
  - Development of component-based modular tools for wide range of workflows
  - Quantitative visual comparison of experiment and model data




The Cinema:Bandit visualization tool links the data across the various view panels. The main view screen (bottom right) can view any of the three sets of data (left column of views): VISAR, diffraction images, or diffraction graph. D. Orban et al.

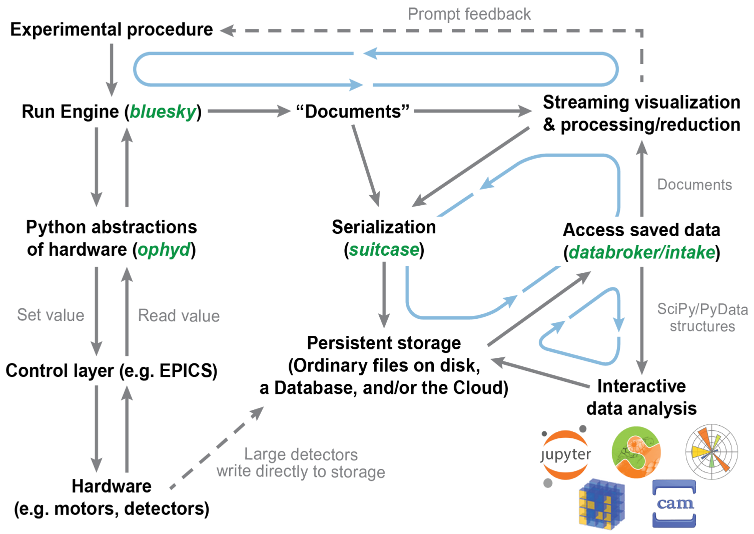



The automatic colorization using ColorMapND for a multi-channel fluorescence image of a mixed ionic-electronic conductor sample -- from from S. Cheng, W. Xu, K. Mueller,

# Computational Workflows and Facility Support

## Real-Time Analytics Workflows and Computation

- Current Workflows:
  - A few frameworks available at beamline facilities for data acquisition, organizing metadata, monitoring data quality.
  - Emerging area in adaptive logic that steers experiments
- Gaps:
  - Little support for remote access during experiments
  - Not enough assistance to control experiment
- Investments:
  - Shared data acquisition or data broker library across facilities
  - Interdisciplinary work with computational and experimental scientists toward R/T frameworks




BlueSky infrastructure for experimental data collection and data management. T. Caswell et al.

# Computational Workflows and Facility Support

## Facility Workflows

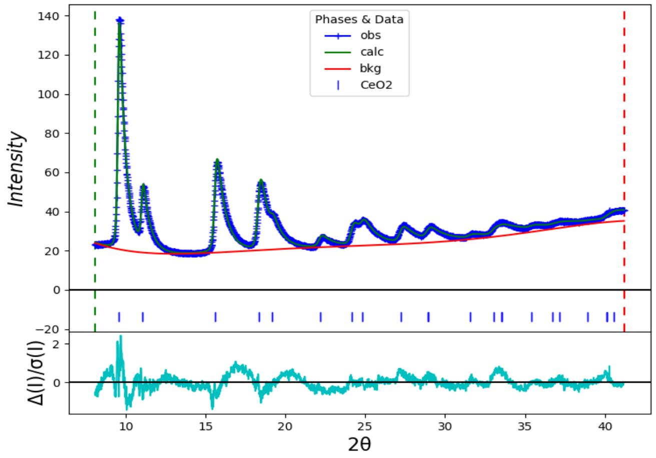
- Facility capabilities:
  - Superfacility project (LBL) underway
  - Data reduction planning by LCLS
  - Cloud computing poised
  - Data hosting (Materials Project)
- Gaps
  - Computing policy, authentication, federation, humans steps thwart automation
  - Speed of analysis and mobility lacking
- Investments:
  - Workflows, data management, and user engagement.


Descriptions of the Superfacility Initiative. Image courtesy of Debbie Bard.

# Priority Gaps

- Analytics not ready for increasing experiment data volumes and data rates
- Diverse software available but not integrated/interoperable
- Data noise, phase wraps, resolution, sparseness, artifacts
- Difficulty obtaining training and simulation data necessary for doing data science
- Lack of forward models and model shortcomings
- Lack of metadata generation and management mechanisms
- Lack of tools to drive remote experiments

## Highlight: Filling a Gap

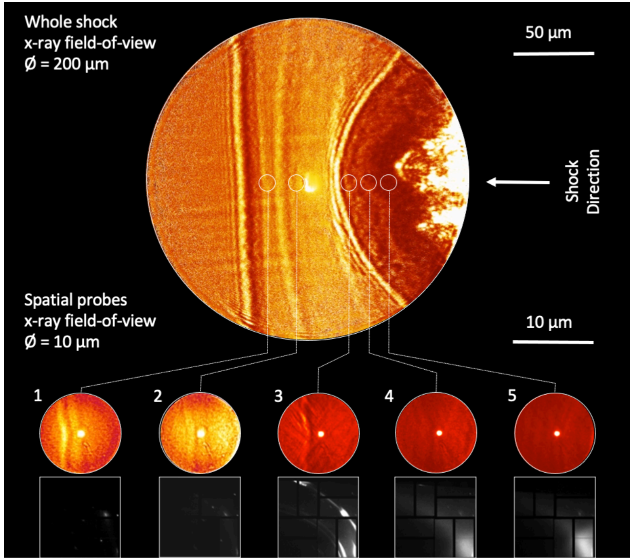



A profile function for pink beam diffraction used in single peak fits and in a Rietveld refinement for $CeO_2$ . Added to GSAS-II tool. R. B. Von Dreele.

# Priority Investments

- Algorithm development for analytics
- More universally applicable/accessible software tools
- Integration of multiple streams of data
- Need for metadata to access data and for tool development
- User engagement by computing facilities
- Data processing and reduction for increased data volumes
- Forward model development and combination with analytics
- Interfacility or facility-to-remote-laptop workflows

## Highlight: Imaging




Research on high resolution imaging and density measurement of shock waves, defects and phase transformations. Here we see direct imaging of ultrafast lattice dynamics.

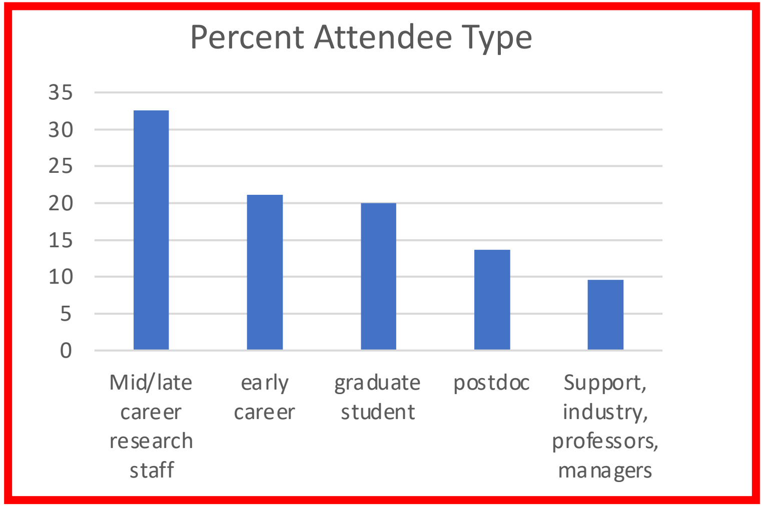
*S. B. Brown, et al. Science Advances, vol. 5, no. 3, 2019.*

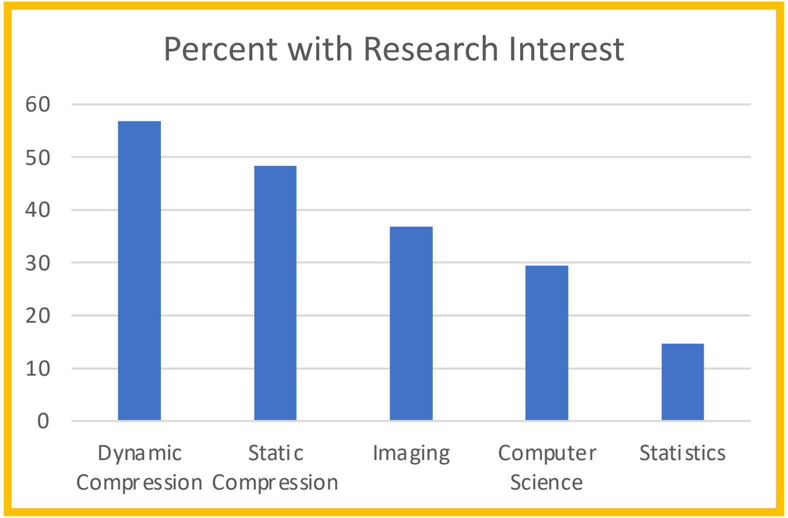
# Suggested Activities

Community seems ready to collaborate and tackle identified challenges

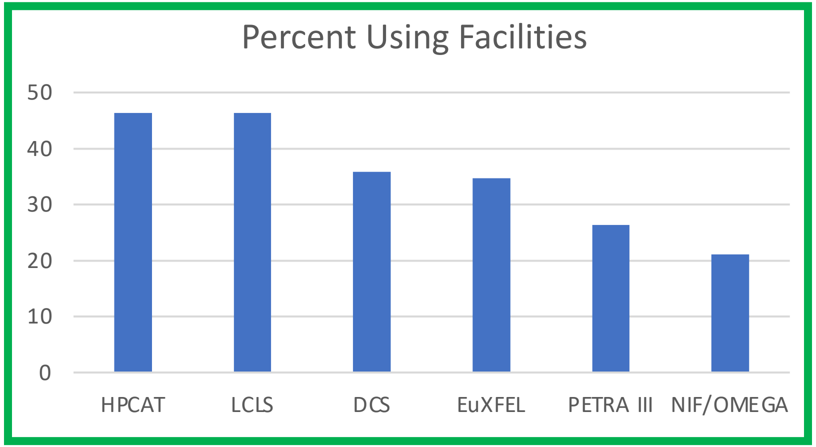
- Increase awareness of community data analytics needs through:
  - Collaboration, especially interdisciplinary
  - Experimental AND computing user facility meetings, professional society meetings or other venues
- Keep momentum from this workshop, distribute workshop report widely
- Determine avenues of funding appropriate for priority areas of investment

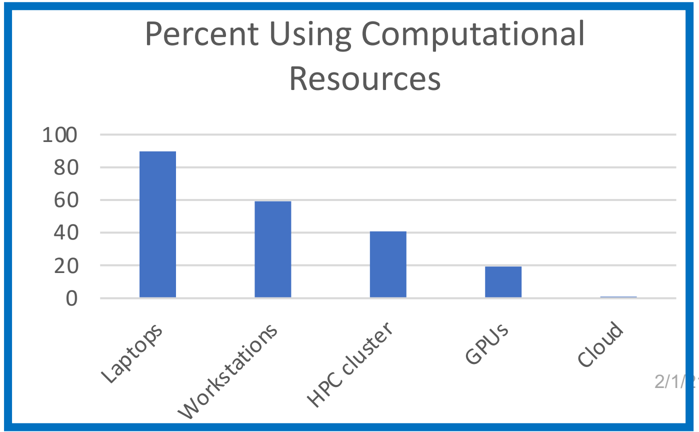
# Workshop Participant Survey Results



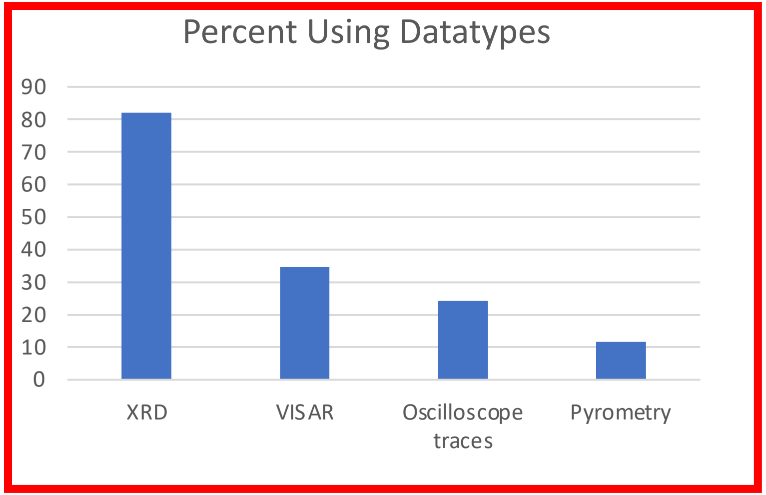



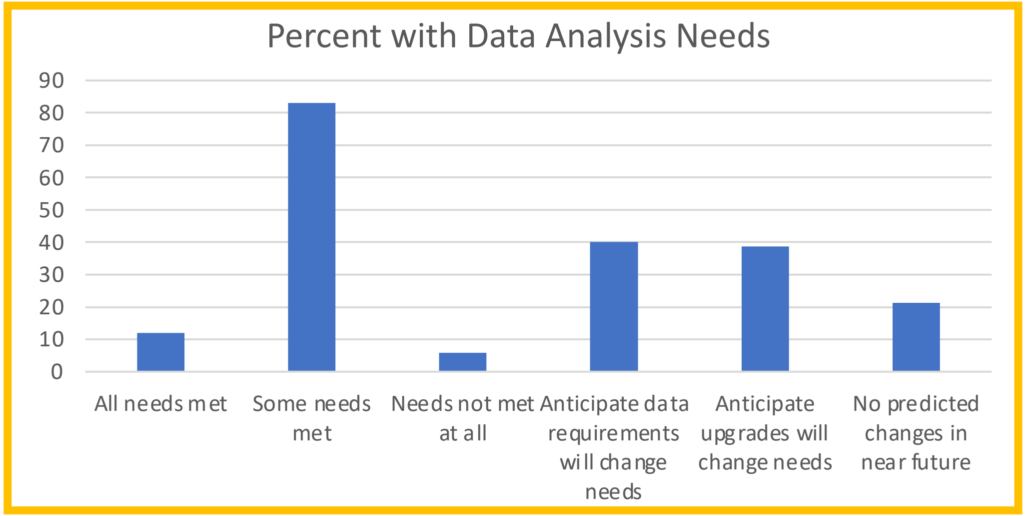

*Survey courtesy L. Huston (LANL)*

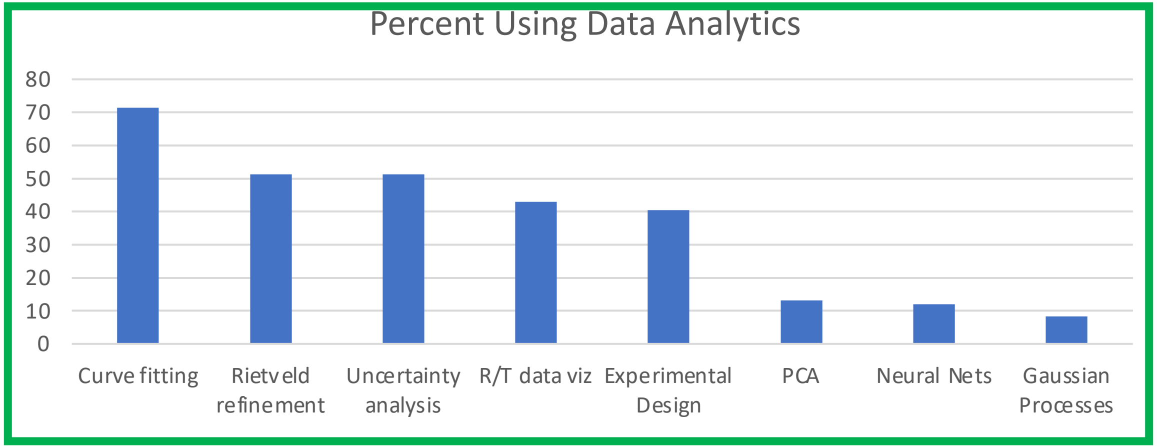


# Workshop Participant Survey Results (cont.)



# Workshop Report Available

- Workshop web site: https://www.lanl.gov/conferences/compress/
- Also available from Los Alamos Authors LA-UR-20-24310

- Previous workshop (2018): Gap Analysis: Materials Discovery through Data Science at Advanced User Light Sources
- https://www.lanl.gov/2018gapanalysis
- Also available from Los Alamos Authors LA-UR-19-21342

# Questions?