U.S. DEPARTMENT OF ENERGY | Office of Science

Data and Communications in Basic Energy Sciences

Creating a Pathway for Scientific Discovery

A Workshop Co-sponsored by Basic Energy Sciences and Advanced Scientific Computing Research

Bethesda North Marriott Hotel & Conference Center

Bethesda, MD

October 24-25, 2011

Walt Polansky
Advanced Scientific Computing Research,
Office of Science

ASCR-BES Data Workshop

https://www.orau.gov/dataworkshop2011/default1.htm

Charge

- Review status, successes, and shortcomings of current data and communication pathways for scientific discovery in the basic energy sciences;
- Ascertain knowledge, methods and tools needed to mitigate present and projected data and communication shortcomings;
- Consider opportunities and challenges related to data and communications with the combination of techniques in single experiments;
- Identify research areas in data and communications needed to underpin advances in the basic energy sciences in the next ten years;
- Create the foundation for information exchanges and collaborations among ASCR-BES research and facilities communities

Co-Chairs

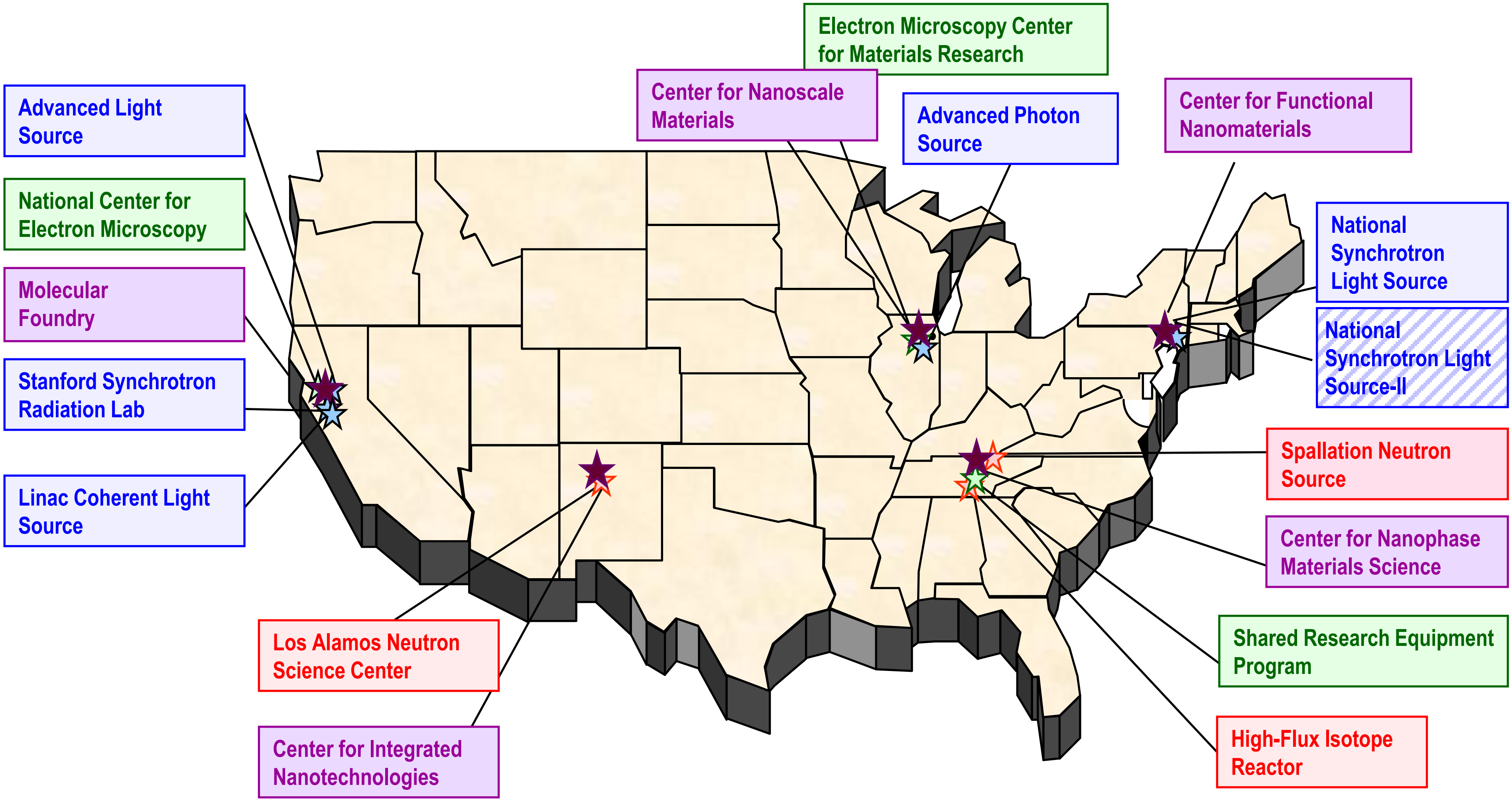
- Peter Nugent, LBNL (NERSC)
- J. Michael Simonson, ORNL (SNS)

Reports

- Draft- by December 9, 2011
- Final- by January 23, 2012

U.S. DEPARTMENT OF ENERGY

BES Scientific User Facilities: Resources for Energy Research



- 4 Synchrotron Radiation Light Sources
- Linac Coherent Light Source
- 3 Neutron Sources
- 3 Electron Beam Microcharacterization Centers
- 5 Nanoscale Science Research Centers

Provided to workshop participants by BES

Broader Context

- Other Agencies: Data Driven Science, Storage, Analysis, Simulation
- NSF Task Force on Data and Visualization
- Representatives here from NSF, NIST
- Other Programs in DOE: LHC, RHIC, Climate Research, Leadership Computing…..
- Interagency Working Group on Big Data
- Competes Act 2010: Working Group on Public Access
- Office of Science Working Group on Digital Data
- Data Play a Key Role in Materials Genome

Provided to workshop participants by BES

- **BES facilities have capability to produce TeraBytes of data per day from *single* beam lines**
- **LCLS, SNS, Synchrotron Light Sources, e$^-$ Microscopes are excellent examples**
- **Increasing use of time resolved & tomographic studies**
- **Increased need for analysis 'on the fly'**
- **Broad spectrum of BES data needs requires:**
  - New level of understanding as a result of sophisticated applied mathematics and computer science techniques
  - New science that extracts the most from our (i.e. BES) facilities

Provided to workshop participants by BES

ASCR-BES Data Workshop
- Participants, Observers & Speakers -

- **Number of Participants- 80**
  - National Laboratories, Universities, NIST, NSF & International

- **Observers- ASCR, BES, BER, HEP & SC-2**

- **Plenary Speakers**
  - Brent Fultz, CalTech- Workflow
  - Thomas Schulthess, ETH Swiss SC Center- Theory & Aglorithms
  - Dave Pugmire, ORNL, Visualization and Analysis
  - Quincey Koziol, University of Illinois, Data Processing & Management

- **Luncheon Speaker**
  - Adam Riess, 2011 Nobel Laureate in Physics
Professor, John Hopkins University
Scientist, Space Telescope Science Institute

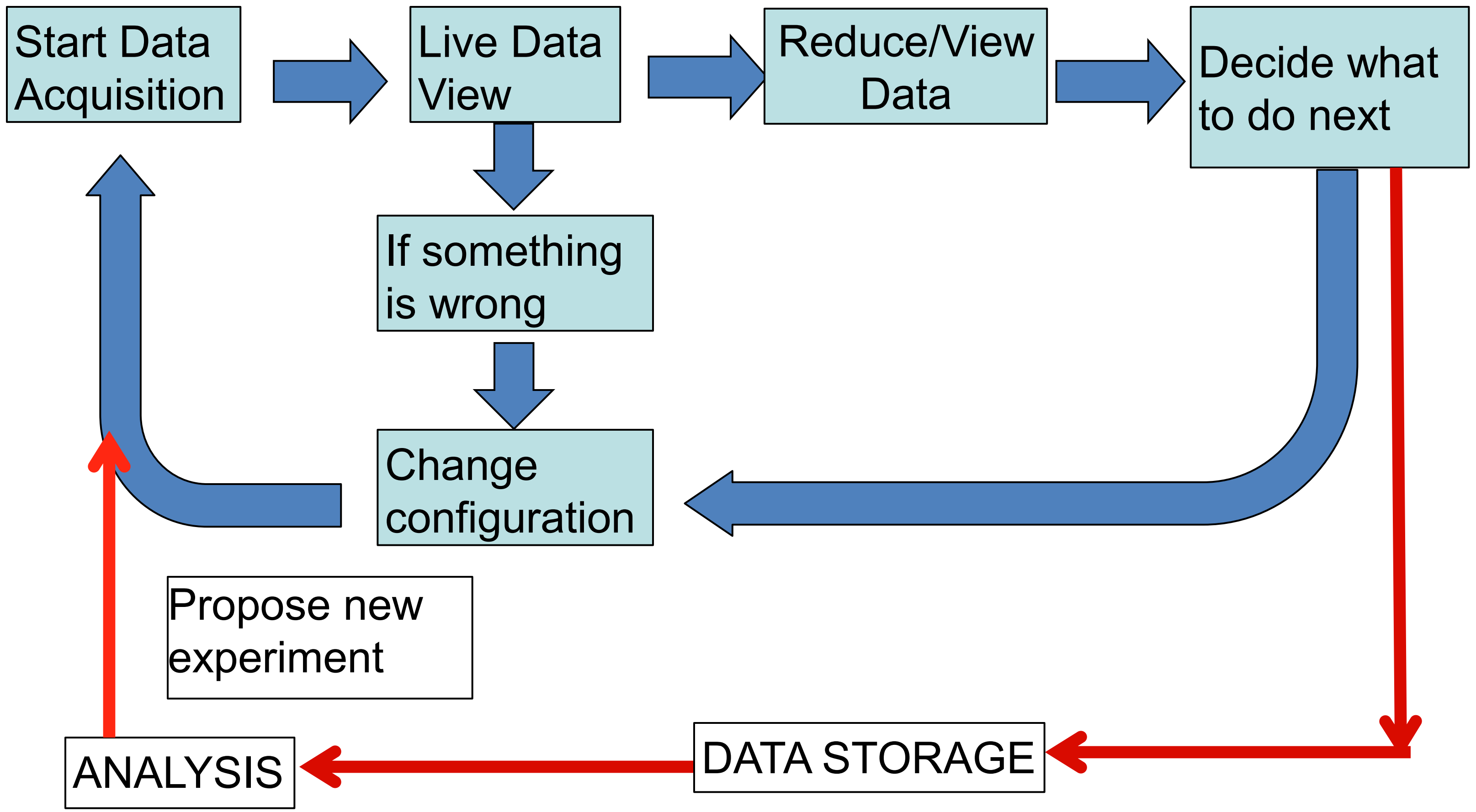
Break-out Sessions

- **Workflow management: Experiment to Science**
  - Identifying and managing the data path from experiment to publication.
- **Theory and algorithms**
  - Recognizing the need for new tools for computation at scale, supporting large data sets and realistic theoretical models.
- **Visualization and Analysis**
  - Supporting near-real-time feedback for experiment optimization and new ways to extract and communicate critical information from large data sets.
- **Data Processing and Management**
  - Outlining needs in computational and communication approaches and infrastructure needed to handle unprecedented data volume and information content.

Poster Session

- “High Performance Computing in Accelerator Science”
- “SciDAC’s Scientific Data Management Center”
- “ESnet: A Partner in Data Intensive Science”
- “Challenges and Opportunities in Data Systems at the Spallation Neutron Source”
- “Data Challenges at Current and Future Light Sources”
- “Linac Coherent Light Source”
- “Data Needs from BES Nanoscale Research Centers: Examples from the Center for Nanophase Materials Sciences”

Workflow, Processing & Viz.



Theory & Algorithms

- **Sustained support for interdisciplinary team consisting of domain scientists (theory and experiment), applied mathematicians, computer scientists working together to meet the theory/algorithm challenges for facilities' data**
  - **Pilot studies**
  - **One team for each end station/facility**
- **Inverse problems and solution algorithms (near term 1-3 years), but there is a need for long term R&D**
- **Feature extraction, image analysis (near term 1-3 years), include model based constraint (longer term)**
- **Combine multiple data sources and imaging techniques to provide more reliable solutions (mid-term 5 years)**
- **Ab initio theory guided experiments for data triage/reduction (long term 5-10 years)**
- **Computational endstation that couples virtual and real experiments (long term 5-10 years)**

Enabling Transformative Science

Detector and source advancements will enable transformative science within BES facilities. Current systems are producing a tremendous amount of data. Future systems will overwhelm current analysis pipelines.

- **Integrate theory and analysis components seamlessly within experimental workflow.**
- **Move analysis to closer to experiment.**
- **Match data management access and capabilities with advancements in detectors & sources.**

Integrate theory and analysis…

… components seamlessly within experimental workflow.

- Coupled simulation and experiment
- Theory guidance for experimental design
- Analysis feedback to steer experiment
- Common data formats
- Common community toolsets for analysis and workflow
- Apply ASCR's investment in visualization and analysis tools (invest in adaptation specific to experiments)

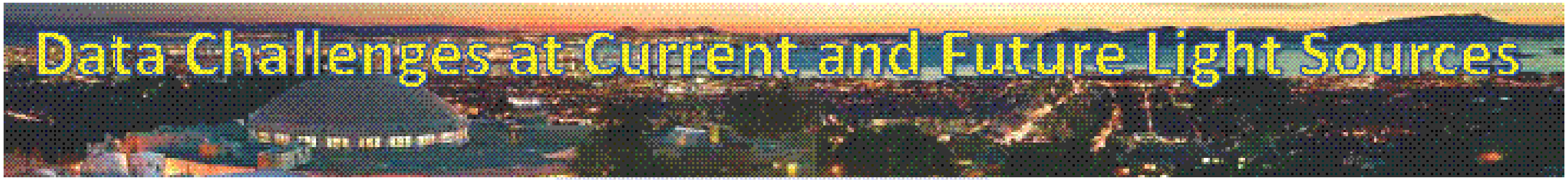
Move analysis to closer to experiment

- Real-time (in-situ), streaming analysis at beamline.
- Local data reduction capabilities
  - 0 suppression
  - Hierarchical filtering
  - Baseline and background subtraction
- Live visualization of experiment
- Improve the efficiency of the experiment
- Increase data-quality
- Improve off-line analysis

Match data management access and capabilities…

… with advancements in detectors & sources.

- Remove the bottlenecks
- Apply existing data transport and mobility toolsets
- Apply forefront mathematical techniques to more efficiently extract science from the experiment.
- Interoperability across different facilities/beamlines
- Combine multiple data sets
- Expandable
- Incorporate legacy data
- Integrated teams of engineers, scientists and computer scientists to solve these problems



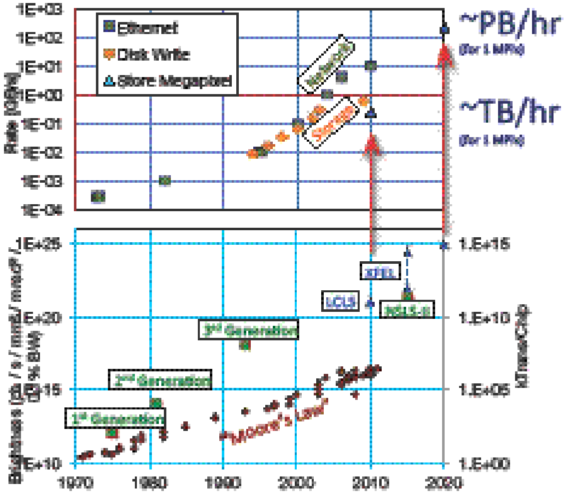
Data Challenges at Current and Future Light Sources

Data Rate ~ Brightness x N_{BL}

100s of Megapixel/second detectors are starting to be used at light sources (and EM centers). ALS, for example, will soon be generating >10 Tb/hr. Requires:

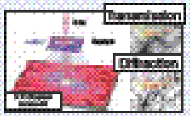
- Implementation of uniform data transfer and management approaches
- Realtime data reduction
- Information extraction

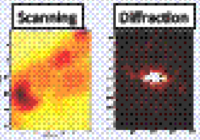
Needs will grow faster than Moore's law – next generation sources will produce ~TB/s



Today – Examples:

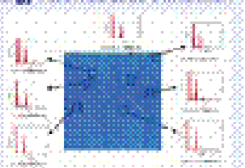
Ptychography



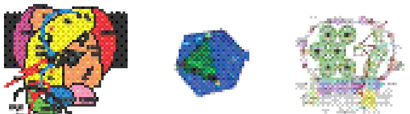


- Tomography
- Scanning X-ray microscopies
- Correlation spectroscopy
- Photoemission microscopies
- ...

Energy-resolved Laue Diffraction



Computational Methods for Determining the Structure of Energy-Cycle Biomachines Through Diffractive Imaging



Now (LCLS): Microcrystals   Future (NGLS): Heterogeneous Single Particles

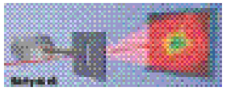
- Engineered bacterial microcompartments as building blocks for artificial solar fuel production systems – their structures cannot be determined by conventional X-ray diffraction

Algorithm and Theory Tasks:

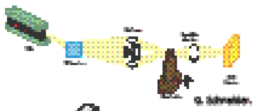
- Image formation theory
- Preprocessing
- Iterative reconstruction and structure refinement
  - Dimension reduction
  - Noise characterization
  - Compressive sensing
    - Nanocrystallography
    - Phase contrast tomography
    - Phase retrieval
- Postprocessing

Tomorrow – Examples:

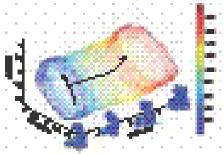
Chemically-Specific Nanoscale Imaging

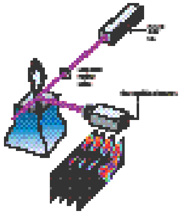


Ultrafast Nanoscale Spin Dynamics



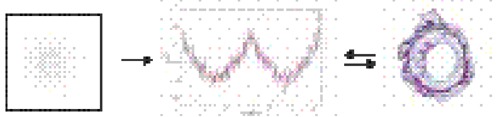
Giga-shot Imaging of Heterogeneous Ensembles





Cinematic Imaging of Reacting Flows

Fluctuation X-ray Scattering



Motivation: Speckle patterns from fs solution X-ray scattering allows more accurate shape determinations of macromolecules in solution. The experiment is performed on an ensemble of particles in solution.
Challenges: How to perform data processing at kHz to MHz rate? How to efficiently solve the 'inverse' problem from autocorrelations to structure?