Funded by the National Science Foundation, Grant #2320373

CyberInfrastructure Training and Education for Synchrotron X-Ray Science (X-CITE)

Introduction to X-CITE Training Program

Anirban Mandal, Erik Scott, Sajith Sasidharan (RENCI, UNC Chapel Hill)

Ewa Deelman, Karan Vahi, Mats Rynge (ISI, USC)

Matthew Miller, Werner Sun, Peter Ko, Kelly Nygren, Keara Soloway, Rolf Verberg (CHESS, Cornell)

Brandon Sorge (IUPUI)

NSF OAC Award Numbers: 2320373, 2320375, 2320374

Funded by the National Science Foundation, Grant #2320373

What is X-CITE and who are we ?

- X-CITE is an NSF funded project with the goal of developing a training program targeted to CHESS users that
  - reduces barriers to the use of CHESS Cyberinfrastructure (CI) - instruments, computing, data, and tools;
  - democratizes access to NSF CI resources and services; and
  - accelerates X-ray science for a broad user community.
- We are a group with expertise in CI as well as X-ray and related domain sciences - [RENCI/UNC, USC, CHESS]

Anirban Mandal, RENCI (Lead PI) | Ewa Deelman, USC (co-PI) | Matthew Miller, CHESS (co-PI)

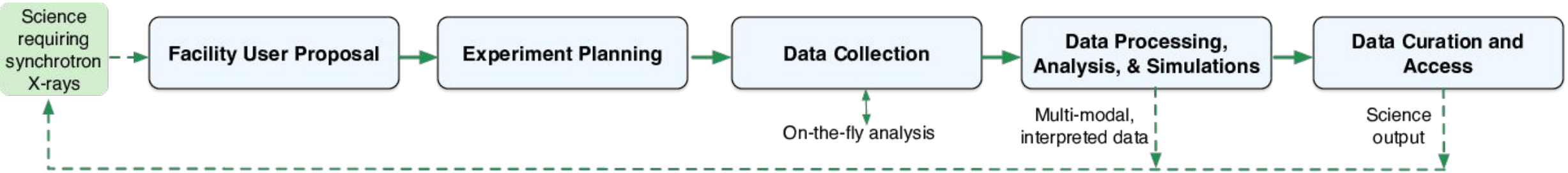
Erik Scott, RENCI | Sajith Sasidharan, RENCI | Werner Sun, CHESS | Peter Ko, CHESS

Kelly Nygren, CHESS | Karan Vahi, USC | Brandon Sorge, IUPUI (Evaluator)

Keara Soloway, Rolf Verberg

Funded by the National Science Foundation, Grant #2320373

CI Training for CHESS: Motivation

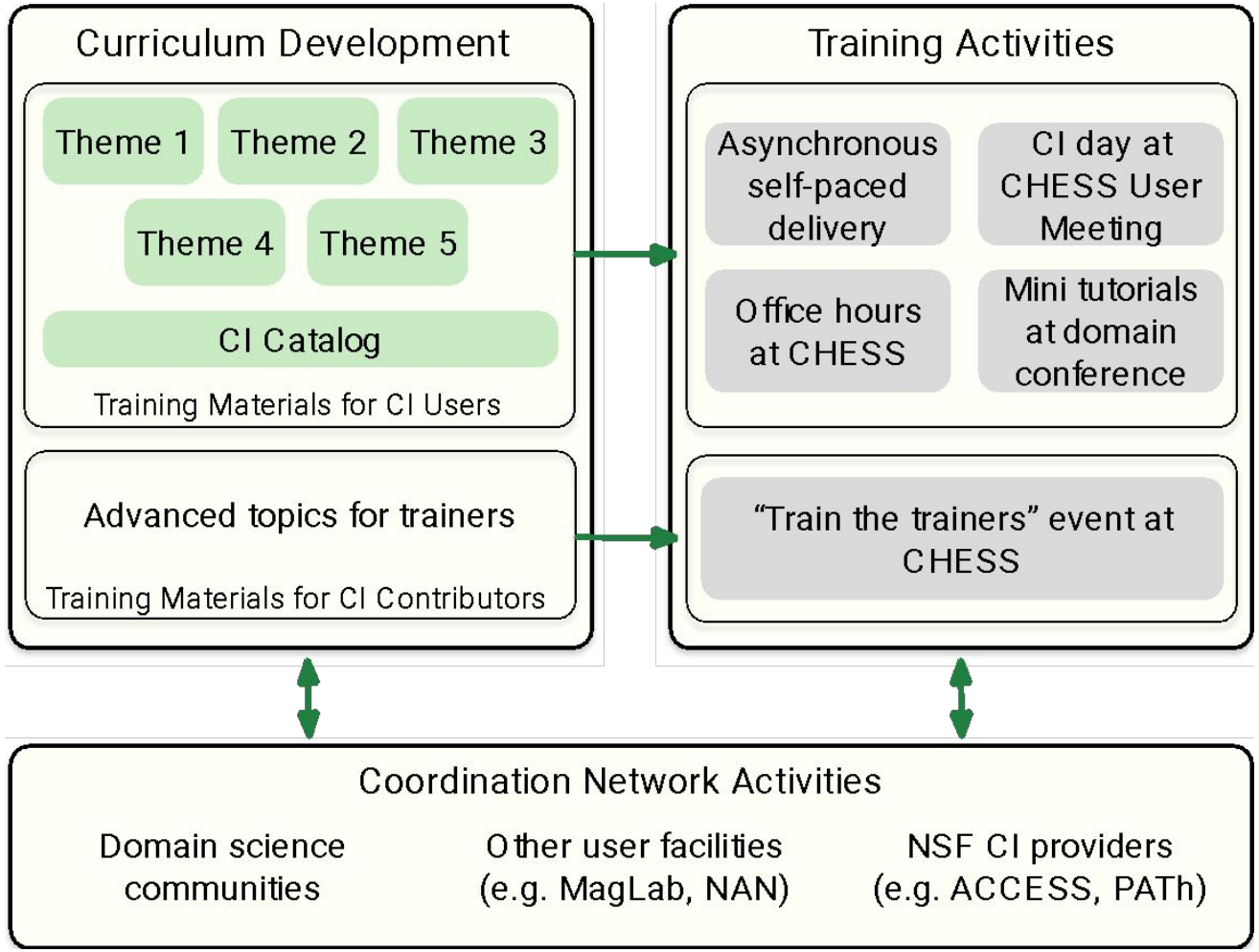


- Proposal and planning stage
  - Proposal plans for processing and analyzing synchrotron X-ray data need some prior CI knowledge in all areas above.
- Data collection stage
  - Need for basic computer systems knowledge (command line) and (on-the-fly) data analysis; Best practices for collecting data
- Data processing stage
  - Processing data (reduction, analysis, simulation, interpretation), handling large data sets, leveraging existing software and CI.
- Data curation stage
  - Metadata management, Open Science/FAIR and data curation.

Funded by the National Science Foundation, Grant #2320373

X-CITE Training Program Overview

- Curriculum development
  - Targeted to the Users
    - Programming Essentials
    - Systems fundamentals
    - Distributed computing
    - X-ray science software
    - Data curation and FAIR
    - CI Catalog
  - Targeted to trainers
    - Advanced CI uses and topics
- Training activities
  - Asynchronous materials
  - Training events
    - CHESS User meeting
    - Tutorials at domain conferences
- Coordination network activities
  - Domain sciences, other facilities, NSF CI providers (ACCESS, etc.)



Funded by the National Science Foundation, Grant #2320373

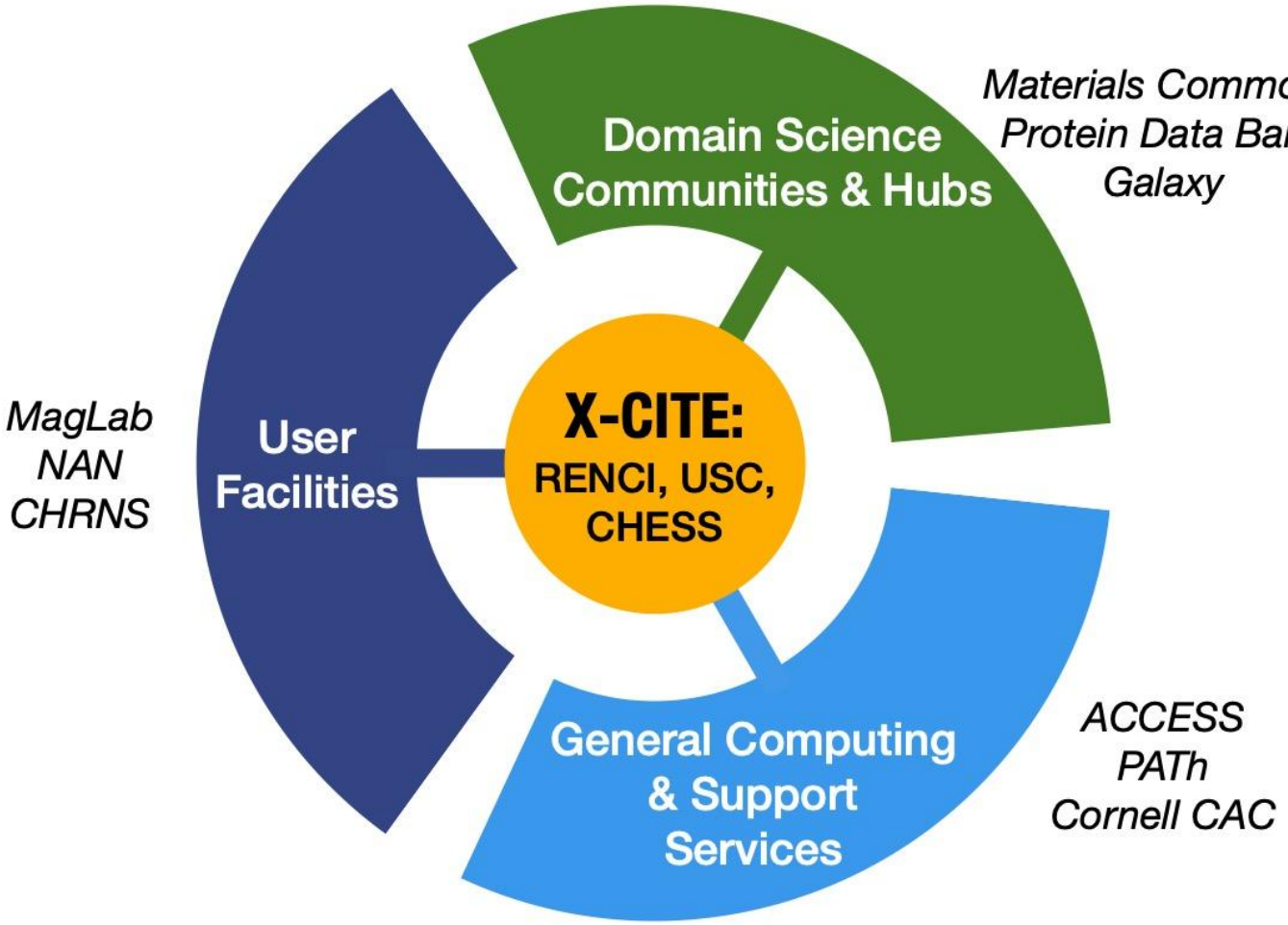
X-CITE Training Program Overview

X-ray Facility Operational Workflow	Facility User Proposal	Experiment Planning	Data Collection	Data Processing, Analysis, & Simulations	Data Curation and Access
Artifacts	Experiment and data management plan	Metadata design based on domain standards Results of user literacy survey User-specific CI requirements analysis	Raw data annotated with correct metadata Data in local CHESS repository or file-system in appropriate file formats On-the-fly analyzed data, annotated	Analyzed/processed data and final results annotated and pushed to local repository and linked to raw data Codes and workflows pushed to repositories CI configurations and Notebooks	DOI applied to raw and processed data and other artifacts Final artifacts in external domain-specific repository Publications linked to data DOI
Training Materials	CF 101, CI Catalog	CF 100, DC 200, CI Catalog	CF 102, XS 100-101, SF 100, PE 100	CF 102, XS 100-102, PE 100-103, SF 100-101, SF 200-201, DC 100-102, DC 200	CF 100-102, CF 200

Artifacts produced during operational workflow and supporting training materials.

Funded by the National Science Foundation, Grant #2320373

X-CITE Community Activities



- Other User Facilities
  - Generalize training materials
  - E.g. MagLab, NAN, CHRNS
- Different Domain Sciences
  - Embed X-CITE materials in domain-specific training
  - E.g. Materials Commons
- Broader computing/CI services
  - NSF CI providers (ACCESS, PATh etc.), Cornell CAC

Funded by the National Science Foundation, Grant #2320373

X-CITE Training Curriculum for Facility Users

Year 1

Theme 1: Programming Essentials (PE)	Theme 2: Systems Fundamentals (SF)	Theme 3: Distributed Computing in the CI Ecosystem (DC)	Theme 4: X-ray Science S&E Software (XS)	Theme 5: Data Curation and FAIR (CF)
PE 100: Python Programming and Jupyter notebooks	**SF 100:** Intro to Linux/ commandline/ scripting	**DC 100:** Distributed computing concepts	**XS 100:** Data collection, preparing input parameters, SPEC and CLI	**CF 100:** Intro to domain metadata standards, formats and repositories
PE 101: Using Python packages & libraries, Conda	**SF 101:** Containers and virtualization	**DC 101:** Scientific Workflow management	**XS 101:** Basic / on-the-fly data analysis, viewing detector images	**CF 101:** Best practices for developing DMP
PE 102: Numerical data analysis with Python	**SF 200:** Parallel computing concepts	**DC 102:** Using science gateways with Open OnDemand example	**XS 102:** Large-scale data analysis: from images to science parameters to interpretation	**CF 102:** Metadata annotation and DOI
PE 103: Software version control, testing and debugging	**SF 201:** Batch Systems and compute farms with CHESS example	**DC 200:** Computing with CI ecosystem - ACCESS, PATh, Campus, Cloud	**XS 200:** Metadata for data fidelity and systematic checks	**CF 200:** Curating data, code, workflows, and publishing

renci USC Viterbi School of Engineering Information Sciences Institute CHESS CORNELL HIGH ENERGY SYNCHROTRON SOURCE

Funded by the National Science Foundation, Grant #2320373

X-CITE Workshop Agenda

Time	Content
8:30 - 8:50 am	**Introduction to X-CITE training program**
8:50 - 9:40 am	**Overview of CHESS research workflow and CI (data collection software, data analysis software, Compute Farm, preparing for beamtime)**
9:40 - 09:55 am	**Break**
9:55 - 10:55 am	**Programming modules (Python and Jupyter, packages/libraries Conda, numerical analysis with Python)**
10:55 - 11:20 am	**Systems Fundamentals (Linux, command line, scripting)**
11:20 - 11:40 am	**CI resources at CHESS and beyond (job submission, CHESS systems, other NSF resources (ACCESS))**
11:40 - 12:15 pm	**Town Hall on CHESS User Experiences and Challenges**
12:15 - 1:15 pm	**Lunch**
1:15 - 3:15 pm	**Hands-on session with complete end-to-end data analysis examples**
3:15 - 3:30 pm	**User survey and feedback**
3:30 - 4:30 pm	**Office hours**

Funded by the National Science Foundation, Grant #2320373

Thanks

Project website: https://sites.google.com/view/x-cite-nsf/home

Training materials: https://xcitecourse.org

National Science Foundation

NSF Cybertraining Award # 2320373

renci

USC Viterbi School of Engineering Information Sciences Institute

CHESS CORNELL HIGH ENERGY SYNCHROTRON SOURCE