

Data Management Plan for CHESS Facility

Draft: 7-5-2017

Types of Data: All storage ring parameters are routinely recorded by the CESR control system. Data collected with beam instrumentation, including turn-by-turn and bunch-by-bunch beam position and bunch size monitors, and electron cloud detectors is archived, as are all relevant calibration constants.

The majority of research data produced at CHESS is in the form of electronic scans or electronic images that record x-ray interactions with specimens. Research data is created by staff scientists as well as facility visitors. In addition to research data, “machine” and “people” data are also produced by the CHESS technical and administrative staff. Managing the user programs creates data while scheduling and recording researcher access to CHESS, submission of proposals, review, equipment needs, scheduling of visits, recording hours received, and user feedback; these data are mostly entered electronically, stored and archived.

Data Standards: CESR data are stored in ASCII or, for large sets, binary formats. For a subset of measured quantities a “pass 1” processed version of the data are stored.. Data acquisition and processing software is documented to simplify future analysis.

X-ray data are recorded using a wide variety of software packages, some commercially produced and some created by staff and individual research groups. Commercial software uses file formats that are considered standards (i.e. SPEC, HKL, HDF), and many commercial and custom-built x-ray area detectors produce digital image formats that are standards (i.e. TIFF, JPEG). In addition, researchers utilize hardcopy notebooks (or electronic equivalents) to tie together metadata needed for indexing data files, specimen handling, and other pertinent conditions. People data are created and recorded by custom interactive databases.

General policies for Access to Data: CESR and CHESS data are stored in secure collection facilities until processed, analyzed, and selectively released by the owners of those data. Standard output channels include facility and community-wide topical web sites, reports in university publications, reports to review committees and NSF, etc.

CHESS data: CHESS considers itself owner of all data produced by staff supported by the NSF grant. In addition to general policies above, CHESS archives and distributes staff data in a variety of ways. CHESS staff honor routine requests for staff (and facility) owned data that may be used to clarify published results. CHESS staff members, similar to research scientists elsewhere, are often involved in collaborations with people in other laboratories. In these cases, primary responsibility for archiving the data may typically rests with the collaborator, using established procedures required by the relevant journal publishing the work. CHESS stores and archives all research data that is electronically produced on CHESS equipment by its staff and by visiting researchers. This data is stored redundantly on disk arrays for approximately six months, and then it is copied to magnetic tape for long-term archival storage. An index of the research data (metadata) is kept persistently on disk. Every dataset is archived twice to tape, and one of these copies is moved to an off-site storage facility. Experimental data that has been archived can be restored to disk upon request. Access to visiting researcher data is the responsibility of the researcher in accordance with the policies of their funding agencies.

Facility and people data: Facility and people data (proposal, beamtime requests, etc.) are always kept private and secure (no external access to raw data). Such information is summarized periodically for official reporting purposes only (i.e. NSF review committees, GPRA and NSF annual reports).

Policies for Re-Use, Re-distribution: The laboratory encourages staff and visitors to prepare and submit for publication all significant findings from work conducted at the facility. CHESS requests that all results from use of the facility acknowledge staff participants and facility funding sources. CESR and CHESS honor routine requests for laboratory owned data for appropriate re-use and publicity.

Archiving of Data and Preserving Access: All categories of data are maintained on redundant-hardware computers and copied to on-site and off-site time-stamped backup devices.