

**PHYSICS REQUIREMENT
DOCUMENT (PRD)****Doc. No.
SP-391-000-06 R0****LUSI SUB-SYSTEM
Controls and Data System****Physics Requirements for the
LUSI Data Management System**Yiping Feng
LUSI Scientist, Author

Signature

Date

David Fritz
LUSI Scientist

Signature

Date

Aymeric Robert
LUSI Scientist

Signature

Date

Sebastien Boutet
LUSI Scientist

Signature

Date

Niels van Bakel
LUSI Scientist

Signature

Date

Steffen Luitz
LUSI Data Management Manager

Signature

Date

Nadine Kurita
LUSI Chief Engineer

Signature

Date

Darren Marsh
LCLS QA Manager

Signature

Date

Jerome Hastings
LUSI Project Director

Signature

Date

Revision	Date	Description of Changes	Approved
R0	4/8/2008	Initial release	

DCR

1. Overview

Experimental instruments that use the Linac Coherent Light Source (LCLS) in its full capability will produce large amounts of data in a very short amount of time. As such, a data management system is required to help organize and archive the data, and make it possible to retrieve it efficiently for post-processing. This document describes the physics requirements for this system.

For front end data acquisition, the physics specifications are described in a separate document "Physics Requirements for the LUSI Controls and Data System". The interface between the backend data management system described in this document and the data acquisition system will be covered by an interface control document, where specific interface specifications are defined.

The nomenclatures used in this document are defined in "Controls-Data-Systems-Dictionary".

2. Requirements

2.1. System Users

The data management system shall be employed by the following:

- 2.1.1. Users retrieving batches of science data with associated metadata after an experiment.
- 2.1.2. Users or data management personnel retrieving science data from individual experimental measurements with associated metadata during or after an experiment.
- 2.1.3. Automated quality control systems retrieving science data from individual measurements with associated metadata during or after an experiment.
- 2.1.4. Automated data analysis systems retrieving science data from individual measurements with associated metadata during or after an experiment.

2.2. Functional Requirements

- 2.2.1. The data management system shall perform management functionalities defined below on data collected by the LUSI Controls/Data system described in sp39100003 at the beginning of LCLS operation in July 2009.
- 2.2.2. The data management in its initial implementation will be designed to only perform long term data archiving and management functions, not data processing.
- 2.2.3. The data management shall be designed and implemented in a scalable manner to accommodate growth in data rates and volume as the LCLS operation improves its uptime, repetition rate, and overall efficiency after the LCLS commissioning in July 2009.
- 2.2.4. The data management system shall define standard data formats for all science and metadata.
- 2.2.5. The standard data format shall include both science data and metadata describing the experiment measurements, and shall be in a common format and self-describing.
- 2.2.6. The standard data format shall be expandable to allow inclusion of more parameters when required.
- 2.2.7. The data management system shall allow retrieving science data based on criteria including

but not limited to the instrument ID, run ID, measurement ID, LCLS X-ray pulse ID, or alternatively the instrument ID, date, and time.

- 2.2.8. The data management shall perform transfer of the science and metadata to long term storage for archiving and subsequent retrieval.
- 2.2.9. The data management system shall allow retrieving data either from the online or the long term storage to designated computer nodes. .
- 2.2.10. The data management system shall implement methods to stream science data and metadata directly to designated computer nodes for data processing.
- 2.2.11. The data management system shall be capable of interworking with common analysis tools including but not limited to MatLab for data analysis. .
- 2.2.12. The data management system shall allow transfer of data to offsite destinations on a best-effort basis.
- 2.2.13. The data management system shall provide adequate protection of the data integrity.
- 2.2.14. The data management system shall provide adequate functionality to prevent the loss of science data and associated metadata.
- 2.2.15. The data management system shall implement access control to prevent unauthorized users from accessing science data and associated metadata.
- 2.2.16. The data management system shall comply with all the computer security policies instituted by the appropriate SLAC management.

2.3. Performance, Capacity and Reliability Requirements

- 2.3.1. The data management system shall be operational at the start of the LCLS operation, which is scheduled to begin in July of 2009.
- 2.3.2. The long term storage shall provide a total capacity of 300 TB initially in July of 2009, and shall be expandable to at least 5 PB.
- 2.3.3. The science data shall have a lifetime of at least 1 year from the time of collection to the time of being purged from the storage systems.
- 2.3.4. Data retrieval does not need to support simultaneous retrieval of the same data set by more than one user at the beginning of the LCLS operation in July of 2009.
- 2.3.5. Data retrieval does not need to support simultaneous retrieval of different data sets by more than 3 users at the beginning of the LCLS operation in July of 2009.
- 2.3.6. The rate of data retrieval shall be comparable to that of data storage.
- 2.3.7. Data transfer to offsite destinations shall be provided on a best-effort basis at a rate consistent with the storage infrastructure at SLAC and the network infrastructure between the SLAC and destination institution.
- 2.3.8. The data management system shall provide reliability consistent with the complexity of the system at the beginning of LCLS operation in July 2009. In other words, the data management software shall not be the limiting factor in the systems' operation.
- 2.3.9. The data management system shall make adequate efforts to prevent the loss of more than 1 day worth of science data and associated metadata.