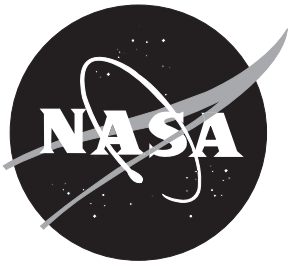
KSC LSDA Data Preparation Handbook



National Aeronautics and Space Administration

Kennedy Space Center, FL

June 1998

Chapter 1

Introduction

1.1 Purpose

This LSDA Data Preparation Handbook serves as a guide for the organization and preparation of data intended for submission to the Life Sciences Data Archive (LSDA). This document describes procedures to be used for preparing data to meet LSDA archive standards. It is intended that the use of this document will inform an investigator or flight project team of the exact procedures and formats needed for submission of data to LSDA.

1.2 Audience

This handbook is intended for scientist and engineers in the life sciences community who are in the process of, or will be, submitting experiment, payload, mission, or hardware data to LSDA. It is designed primarily for data producers that are new to LSDA. However, persons already familiar with LSDA data submission procedures may also find it useful as a reference.

LSDA has developed requirements and standards for data to help ensure that the data it provides to users in the science community are complete, accurate and easily accessible. LSDA works with flight projects during all phases of the mission; before launch to help flight personnel develop data acquisition plans, during the mission to archive any data that is downlinked, and after the mission to archive data produced through laboratory analysis. This handbook is intended for all types of data suppliers and developers working with LSDA.

1.3 Document Scope

The information included here describes the first version of KSC LSDA data preparation procedures for preparing data sets to meet the LSDA standard. These procedures set forth the practices of data submission to KSC's LSDA Project. Future versions of this documents will supersede information given herein.

1.4 Document Overview

This Handbook is organized by the steps involved in the data preparation and submission process. The requirements, standards, and procedures presented here reflect the most recent LSDA updates.

1.5 Related Documents

The LSDA "**Post Flight Document Preparation Guide**" should be used in conjunction with this handbook, for an explanation of the complete set of deliverables due to NASA upon completion of a flight experiment.

1.6 Overview of LSDA

LSDA was created by NASA to provide a means of archiving data that is produced from life sciences flight experiments. Sponsored by NASA's Life and Bioimedical Sciences and Applications Division (Code UL), LSDA provides a system of organization and support that is designed to stimulate research, facilitate data access, and support correlative

analysis. The LSDA's primary customer is the Life Sciences research community, but it is also used by students, educators and the general public.

The Life Sciences Data Archive (LSDA) Project is responsible for collecting and disseminating data of NASA funded Life Sciences space flight investigations. The data archived in the LSDA is produced by both intramural and extramural investigators funded to perform flight experiments through NASA grants. It is anticipated that the archive may grow to include data from intramural and extramural investigations which are completely ground based.

The LSDA is based upon a distributed archive structure with responsibilities distributed to LSDA Nodes at various NASA Centers responsible for flying experiments. There are the LSDA Project Nodes which are responsible for the actual collection and cataloging of data, and there is a LSDA Data Distribution Node which is responsible for dissemination of the data to the user community. The LSDA Project Nodes are located at Ames Research Center, Johnson Space Center and Kennedy Space Center. The Data Distribution Node is located at Johnson Space Center and can be accessed via the World Wide Web at http://lsda.jsc.nasa.gov/.

1.7 Public Access to Archived Data

Access to LSDA data is handled through the LSDA Data Distribution Node. Users enter the LSDA through the World Wide Web (WWW). They are provided a Master Catalog which is comprised of a relational database with a WWW forms interface. The catalog allows users to search and find data that meet their needs. Users are given various areas in which to search. Most users locate data by searching for a particular experiment and then viewing the information provided about the data. Data and documents also can be searched for via hardware, research subject, mission, personnel, sessions and biospecimen. Once data or documents of interest are located they can be downloaded or ordered through links in the catalog.

Chapter 2

Why Submit Data to LSDA?

As a NASA funded investigator your are obligated to turn over your data produced from a flight experiment, one year after flight. This requirement has been developed in order to ensure that this valuable and rare national resource is properly protected and is available to others. The LSDA was developed in order to make the submission and location of data easier. Scientist not immediately familiar with the archive system must be able to locate data sets of potential interest and be able to extract meaningful subsets of data for further study. With this in mind LSDA has established:

- Standards for identifying and organizing Life Sciences Data sets.
- Common nomenclature and data structures for cataloging and archiving life sciences data.
- A catalog system which allows users to serach for the location of data and documents and a distribution system which allows them to be easily downloaded or ordered.
- Data management expertise which is available to both Flight Projects and Principal Investigators for the development of archival protocols and procedures prior to flight.

LSDA not only seeks to preserve data that might otherwise be lost or poorly documented, but to make all data readily accessible to an ever-increasing community of potential users through the World Wide Web.

Chapter 3

What is a Data Set?

One of the objectives of the LSDA is to introduce consistency in the contents and organization of Life Sciences Experiment / Project Data Sets. The LSDA has established the an archive quality data set must include everything that is needed to understand and utilize the data. Towards this goal, the LSDA has developed this handbook.

The remainder of this handbook focuses on the process of preparing and submitting data to LSDA. To do this effectively, you must first understand the terminology used within the text to describe different levels and groupings of data, as well as, the terminology used to describe its physical organization.

3.2 What is Data ?

The data collected and archived by the LSDA consist of measurements of various parameters in spreadsheet form, electron-micrographs, echocardiographs, video tapes, RACAL tapes, analog tapes, physical specimens, microscope slides, photographs, and hard copy logbooks, lab books and other documents.

3.2 What is a Data Set?

The organization of Life Sciences Data can be quite complex. Producing an archive quality data set means including everything that is needed to understand and utilize data. The relationship between the data and everything else can be complex also.

To distinguish between different levels of data, groupings of data, and relationships between data and everything else, the following terminology will be used.

Data set - a data set most often refers to a collection of data organized in a scientifically relevant fashion. The LSDA organizes data elements into data sets by experiment or payload.

Data Element - a single piece of data such as an electronmicrograph or a spreadsheet.

Some concrete examples are:

- A graph showing summarized data, e.g. mean pooled data (n=7) of body mass fluctuations during the preflight, inflight and postflight periods of a mission

- A spreadsheet showing analyzed data, e.g. individual subject data of erythropoeitin levels preflight and postflight

- A spreadsheet showing raw data, e.g. individual organ weights from rodent dissections.

- An electronic data file of raw telemetered data.

Catalog information - descriptive information about a data element (e.g. mission on which the data were collected, time of data collection, biomedical parameters measured, description of the hardware or method used to collect the data).

Documents - textual material which describes the data set. This can include documents which outline the methodology and/or processes used to collect the data, and can include references to science publications and other text useful for interpreting the data.

3.3 How Data Sets are Physically Organized

The data set components identified in Section 3.1 must be organized onto digital media for deep archiving and distribution. Data sets are organized via experiment. Since the main mechanism for distribution of data is via on-line downloading, the data is downloaded through the catalog.

Chapter 4

The Data Preparation Process

There are six steps to the delivery of data sets to the LSDA.
1) Orientation - Learning what LSDA expects
2) Archive Planning - deciding what to archive, when and generally how
3) Archive Design - Learning the details of putting an archive data set together
4) Data set Assembly and Validation - pulling the pieces together
5) Data Set Reviews - review by LSDA and yourself to determine the completeness and quality of data set
6) Delivery - passing the final complete data set over to LSDA

Subsequent chapters of this handbook provide detailed information for completing each of these steps. These steps to submit Life Sciences data to the LSDA are similar whether the data are being submitted from a Principal Investigator or a Flight Project.

Chapter 5

Orientation

This section describes the first phase of the archive process. During this phase contact is established between LSDA and the Principal Investigator and the Payload Mission Manager.

5.1 Establish Contact with LSDA

The first point of contact with LSDA will be through the LSDA Manager at KSC. The LSDA Manager coordinates support for all aspects of archiving active flight project data and past flight projects for KSC funded investigations.

5.2 LSDA Orientation Material

During the Science or Payload Verification Test (SVT or PVT) for your Payload you will meet with the LSDA Manager who will explain the archive process and discuss NASA's archive requirements. During this meeting you will be provided several LSDA documents to help in your understanding of the LSDA requirements. These documents are:

- Understanding the LSDA (An Investigators Guide)
- LSDA Brochure
- KSC LSDA Data Preparation Handbook
- KSC LSDA Postflight Document Preparation Guide

At this point the you should name an"Experiment Archive Coordinator". This person will act as the lead for preparing documentation for the archive and will work with the LSDA representative to properly archive the experiment.

5.3 Provide General Information to LSDA

In order to begin the process of archiving your experiment certain information will be important for you to provide to the LSDA after this meeting these include:

- Experiment description {objectives and approach},
- Types of data to be collected {measurements in spreadsheets, electron micrographs, light microscope slides, DNA analysis, etc..}
- Volume of data expected (i.e., 1 gigabit of spreadsheets, 30 Electron Micrographs)

Chapter 6

Archive Planning Checklist

❒ Prepare a Data Transfer Plan (DTP)

❒ Plan for Updates to the DTP

❒ Keep LSDA Informed

❒ Prepare Preliminary Accomplishments Report (60 days Postflight)

❒ Prepare Final Research Report (1 year Post Flight)

❒ Prepare Data Submission Package (1 year Post Flight)

Chapter 6

Archive Planning

For Active flight projects, archive planning consists of identifying the data to be archived, developing a detailed archiving schedule and defining data flow. Part of this planning also defines roles and responsibilities of the various people involved in producing the final archive products.

6.1 Prepare an Data Transfer Plan (DTP)

The Data Transfer Plan (DTP) provides a detailed description of the production and delivery plans for archive products for a project.

The contents of a DTP include:

1. A list of proposed data elements for archive, including a general description of the science content of each data element.
2. The storage method, storage medium, count, total size, processing level, etc., of each proposed data set.
3. Proposed data transfer mechanism (disks, harddrives, FTP).
4. LSDA will also provide assistance in developing the DTP at the request of the project.

6.2 Plans for updates to the DTP

It is inevitable that changes will occur that will effect the DTP. In particular, detailed data set lists and schedules found in the DTP appendices often develop over time, making these appendices "working" guidelines for archive planning. The LSDA Project Manager should be notified of changes to these plans as they occur, and document revisions should be scheduled periodically. Changes include additions, and/or deletions of products, changes in schedule and changes in quantity, product content or format.

Chapter 7

Archive Design

Archive Design consists of:

- a review of LSDA standards
- data product design
- data product development
- catalog development

This chapter describes each of these activities. These tasks are not meant to be sequential. In many cases there may be several iteration between the various steps.

7.1 Review LSDA Standards

When developing your data for archiving you must keep in mind that it needs to accessible and usable to the Space Life Sciences community. LSDA has developed standards to help you in organizing and describing your data. This handbook should provide you with a basic understanding of what the LSDA standards are and how to implement them with your data.

For active flight project a Data Preparation Workshop will be held with you during one of your trips to hanger L. This workshop will focus on what the LSDA standards are etc....

The LSDA standards for different types of data are as follows:

Spreadsheets - Microsoft Excel

Documents / Text - Microsoft Word

Images - photographs (will be scanned) or Digital images - Tiff image format

Log / Lab Books - Xerox copy of original

Light Microsope slides

7.2 Data Product Design

This activity includes the determination of the contents, the file format and the catalog records for the data element.

Each data product must contain information to make it usable to the general sceintific community. To do this it is required that the data be in the LSDA required formats (see section 7.1) and be labled clearly. For example if the data object being submitted to the LSDA is a spreadsheet, it is required that it contains a title, each column be clearly labled and it has a legend that describes what each column header means and what was being measured. (EXAMPLE)

7.3 Data Product Development

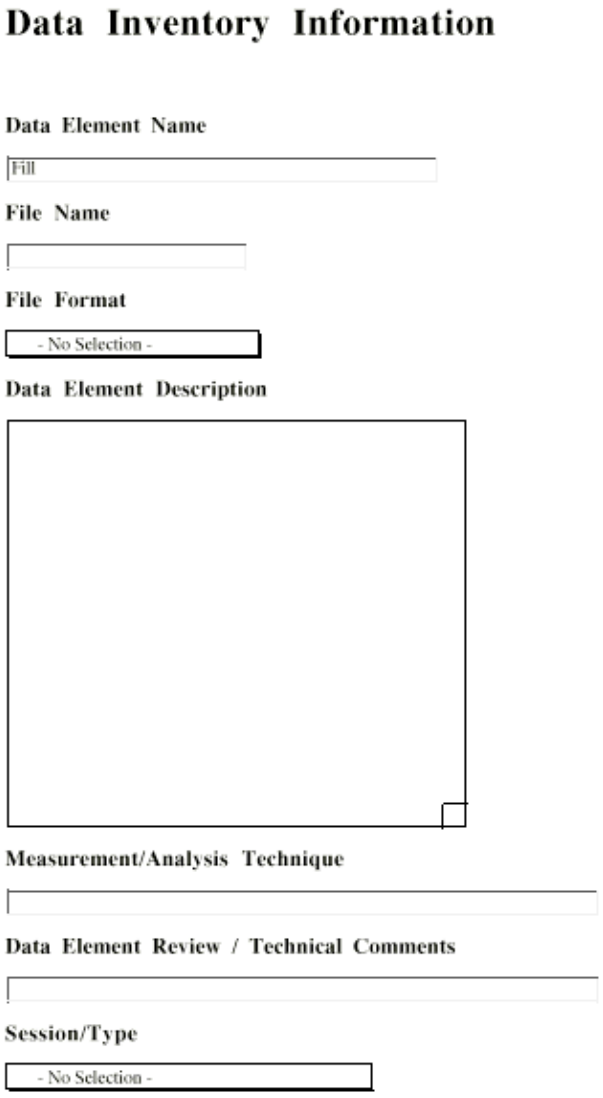
7.4 Data Inventory Catalog

To make a complete submission package a catalog record needs to be produced that clearly describles the data. The LSDA uses a Data Inventory Catalog to do this. The data inventory catalog contains fields with the information LSDA needs to adequatly describe the spreadsheet.

To produce the catalog record you would use the World Wide Web to access a data inventory catalog provided by the LSDA for your experiment. The caltalog is a database with a WWW forms interface (Figure 7.4.1) . The form contains fields that you fill out to describle the data.

This catalog is also a useful tool for you to track the data being produced in your labortory as it is being produced. By utilizing the catalog as a labortory tracking tool it not only organizes your data to make it more useful to your lab, but it also makes the final delivery of data to the LSDA less cumbersome and time consumming.

Figure 7.4.1



Parameters Measured

Acetylene concentration (environmental) Acid phosphatase content Actin cytoskeleton Actin depolymerization Actomyosin based locomotion Adenosine triphosphate (ATP) content ADP glucose pyrophosphorylase AGPase activity Alkaline phosphatase (AP) content Amino Acid content Amylase activity Amyloplast content (or count) Amylose concentration Anionic peroxidases Anticlinal division Auxin content Auxin transport Axis polarity: embryo B-glucanases Bacterial contamination Bacterial culture Bacterial density Behavior Beta Particles Binucleated cells Blastulae Calcitonin Calcium Calcium, 48 isotope-label Calcium content Calcium, ionized Calcium mediated signal transduction Cambial zone cAMP Carbohydrate Carbohydrate metabolism Carbohydrate partioning Carotenoid content cDNA Cell count Cell density Cell division Cell formation Cell function Cell proliferation Cell ultrastructure Cell volume Cell wall formation Cellular second messengers Cellulose Cellulose microfibril cGMP Chitinases Chloride content Chlorophyll content Chromosome breakage Chromosome damage Chromosome damage, bridges Chromosome damage, deletions Chromosome damage, laggards Chromosome partitioning Circadian rhythm Circumnutation Citrate synthase (CS) content Clinorotation CO2 content CO2 usage Colony forming units (CFU) Compression wood formation Cortical granule exocytosis Cotyledons cRNA content Cross-section area Cyclic AMP (cAMP) Cytogenesis Cytokinesis Cytology Cytoplasm Cytoskeletal development Development Developmental abnormalities Differentiation DNA DNA biosynthesis Double nuclei E-caffeic acid E-cinamic acid E-coniferyl acid E-p-cinamic acid E-p-coumaric acid Early polar development Ecdysone levels Electron transport Elemental composition Embryo development Embryo initiation Embryogenesis Embryonic growth Embryonic structures Enzyme activity Enzyme, substrates Esterase Ethylene Expansin gene expression Expansins Extracellular fluid volume (ECF) Extracellular matrix (ECM) Fatty acid content Feeding behavior Fertilization Fibronectin Flower formation Food consumption Food vacuoles Fructose Fungal contamination Fungal culture Fungal density Gamma rays Gastrulae Gene expression Gene regulation Genetic abnormalities Genetic defects Geotaxis Geotropism Germination GH3 promoter-GUS Glucose Glutamate Glutaminase Glycine Glycogen Gravireceptors Graviresponse Gravisensing Gravitaxis Gravitropism Gravity signal perception Gravity signal transduction Growth rate GUS reporter gene Heartwood Hemicellulose Hemolymph amino acid High gradient magnetic fields Histology Hook formation Humidity, ambient Hydration Hydrogen peroxide concentration Hypergravity Hypocotyls Hypogravity HZE particles Iron Leaf movement Life span Lignin content Lipase Lipid accumulation Locomotion Magnesium Mating behavior Meiosis Meiotic chromosomes Membrane diffusing capacity Membrane fluidity Membrane structure Mesophyll cells Metabolism Microfilament formation Micronuclei Microtubule cytoskeleton Microtubule depolymerization Microtubule-based locomotion Mineral loss Mitochondria number Mitochondrial volume Mitotic chromosomes Mitotic forms Mitotic index Monolignols Monomeric lignin Morphological development Morphology Mortality mRNA Mutation rate Neutron dosage Nitrogen balance Nitrogen content Nuclear division Nuclear migration Nucleus density Nucleus volume Nutrient solution Orientation Oxygen content Pathogen ingress Pectins Peg formation Periclinial division Peroxidase activity Phenolic polymer deposition Phenolic polymers Phenylalanine Phenylalanine ammonia lypase Phospholipase D Phosphorus content Photography, time-lapse Photosynthesis Photosynthetic characteristics Phototropism Phytoalexin biosynthesis Pigmentation Plant respiration Plastids Pluteus Pollen fertility Pollen production Pollen viability Population size Potassium content Potassium iodine (KI) concentration Proembryo Protein content Protein pattern Protonemata PSI PSII Radiation dosage Radiation injuries (MeSH) Radiation protection (MeSH) Radiation tracks Radiographs Reproduction Ribonucleic acid (RNA) content Root growth Root rot SAUR promoter-GUS Seed production Sexual differentiation Shoot growth Signal transduction Sodium content Somatic embryos Spermatogenesis Spindle structures Spore count Starch concentration Statolith count Stem water potential Suberin Sucrose content Sucrose phosphate synthase Sulfur content Survival rate Swimming Swimming behavior TAGES - Transgenic Arabidopsis Gene Expression System Taxol Temperature Thylakoid membrane stacking Tissue elemental composition Totipotent cell count Transgenic Transpiration Trehalose Ultrastructure Waste production Water Weight Zinc content

Data Collection Sites

- No Selection -

Collection Phase

- No Selection -

Collection Day

Collection Date

Protocol/Approach

Data Medium

- No Selection -

Research Specimens/Subjects Group

Start Time

Stop Time

Level of Processing

- No Selection -

7.4.1 Catalog Fields and Valid Values

The following fields of information are used in the inventory catalog to describe the data. Soce fields have values that have been specified by the LSDA project. These values are in pull down menus or check boxes on the inventroy catalog web page.

Catalog Number
Unique identification number, supplied by the appropriate project, that identifies each data element. This fields values are automatically entered into the catalog when a record is created.

Collection day
Day or range of days on which the data set element was collected

L-x
lx
R+x

Ranges: L-x to R+x
or Report as a list
L-5
FD2
FD3
R+5

Collection phase
Flight phase in which the data were collected

Preflight
Inflight
Postflight
Not applicable
Not reported

Data collection site
Site where the raw data were collected

Kennedy Space Center (Life Sciences Supoport Facility - Hanger L)
Investigator laboratory
Shuttle/Spacelab
Dryden Flight Research Facility (DFRF)

Data element description
Should describe the contents of each data element, including subjects involved, main parameters measured as well as the timeframe in which they are measured.

Data element name
Name given to the archived data element; the name should contain enough information to differentiate it from other similar data elements.

Data element review_technical remarks
Textual assessments of the data element's quality; should describe any extraneous factors which have adversely influenced the accuracy of data (hardware problems, subject illness, etc.)

Excellent (if there are no problems with the data)
Otherwise, a brief paragraph describing the problem is necessary.

Data medium
The medium on which the experiment data are available.

16 mm film
35 mm film
Ambient temperature
Beta video
Cassette tape
Exabyte tape
Microfilm
Microscope slide
Printed medium
RACAL tape
TEAC tape
Temperature storage -70°C
Temperature storage +4°C
VHS video

File format
Format of the archived data; includes version number of software-specific formats. The following softwares and version numbers are current standards of the LSDA.

Adobe Acrobat
ASCII
Claris FileMaker Pro 2.1
Microsoft Word 5.0
Microsoft Excel 4.0
Not applicable (if document is only available in hardcopy form)

Level of processing
Reflects the degree of data reduction and refinement of the data element

Analyzed
Processed
Raw
Summarized

Measurement_analysis technique

Parameters measured
Biomedical parameters measured for a particular experiment

see Appendix 1

Protocol_approach
Methods, tests, or other approaches used during data collection
- Asynchronous control
- Synchronous control
- OES Control

Research Subject ID
Unique identifier for a specific research subject or group of subjects; ID is assigned by LSDA Level IIIs

Session_type
Field refers to data groupings. For KSC, this field refers to the type of experiment group the subjects belong to
- Delayed Flight Profile Test (DFPT)
- Flight
- Ground control

Start time
Year, followed by start time of data element collection
- Valid values don't apply to this field
- Use Not available if no start time is listed

Stop time
Year, followed by start time of data element collection

File name
File name for documents available in electronic form

Chapter 8

Data Set Assembly and Validation

Chapter 9

Data Set Reviews

Chapter 10

Delivery