

Draft Recommendation for  
Space Data System Practices

|  |
| --- |
| Information Lifecycle And Long Term Usage |

PROPOSED Draft Recommended Practice

CCSDS 653.0-W-0.7

WHITE Book

* November 2015

AUTHORITY

|  |  |  |  |
| --- | --- | --- | --- |
|  | | | |
|  | Issue: | White Book, Issue 0.7 |  |
|  | Date: | November 2015 |  |
|  | Location: | Darmstadt, Germany |  |
|  | | | |

**(WHEN THIS RECOMMENDED PRACTICE IS FINALIZED, IT WILL CONTAIN THE FOLLOWING STATEMENT OF AUTHORITY:)**

This document has been approved for publication by the Management Council of the Consultative Committee for Space Data Systems (CCSDS) and represents the consensus technical agreement of the participating CCSDS Member Agencies. The procedure for review and authorization of CCSDS documents is detailed in *Organization and Processes for the Consultative Committee for Space Data Systems* (CCSDS A02.1-Y-4), and the record of Agency participation in the authorization of this document can be obtained from the CCSDS Secretariat at the address below.

This document is published and maintained by:

CCSDS Secretariat

Space Communications and Navigation Office, 7L70

Space Operations Mission Directorate

NASA Headquarters

Washington, DC 20546-0001, USA

FOREWORD

[Foreword text specific to this document goes here. The text below is boilerplate.]

Through the process of normal evolution, it is expected that expansion, deletion, or modification of this document may occur. This Recommended Practice is therefore subject to CCSDS document management and change control procedures, which are defined in *Organization and Processes for the Consultative Committee for Space Data Systems* (CCSDS A02.1-Y-4). Current versions of CCSDS documents are maintained at the CCSDS Web site:

http://www.ccsds.org/

Questions relating to the contents or status of this document should be addressed to the CCSDS Secretariat at the address indicated on page i.

At time of publication, the active Member and Observer Agencies of the CCSDS were:

Member Agencies

* Agenzia Spaziale Italiana (ASI)/Italy.
* Canadian Space Agency (CSA)/Canada.
* Centre National d’Etudes Spatiales (CNES)/France.
* China National Space Administration (CNSA)/People’s Republic of China.
* Deutsches Zentrum für Luft- und Raumfahrt (DLR)/Germany.
* European Space Agency (ESA)/Europe.
* Federal Space Agency (FSA)/Russian Federation.
* Instituto Nacional de Pesquisas Espaciais (INPE)/Brazil.
* Japan Aerospace Exploration Agency (JAXA)/Japan.
* National Aeronautics and Space Administration (NASA)/USA.
* UK Space Agency/United Kingdom.

Observer Agencies

* Austrian Space Agency (ASA)/Austria.
* Belgian Federal Science Policy Office (BFSPO)/Belgium.
* Central Research Institute of Machine Building (TsNIIMash)/Russian Federation.
* China Satellite Launch and Tracking Control General, Beijing Institute of Tracking and Telecommunications Technology (CLTC/BITTT)/China.
* Chinese Academy of Sciences (CAS)/China.
* Chinese Academy of Space Technology (CAST)/China.
* Commonwealth Scientific and Industrial Research Organization (CSIRO)/Australia.
* Danish National Space Center (DNSC)/Denmark.
* Departamento de Ciência e Tecnologia Aeroespacial (DCTA)/Brazil.
* European Organization for the Exploitation of Meteorological Satellites (EUMETSAT)/Europe.
* European Telecommunications Satellite Organization (EUTELSAT)/Europe.
* Geo-Informatics and Space Technology Development Agency (GISTDA)/Thailand.
* Hellenic National Space Committee (HNSC)/Greece.
* Indian Space Research Organization (ISRO)/India.
* Institute of Space Research (IKI)/Russian Federation.
* KFKI Research Institute for Particle & Nuclear Physics (KFKI)/Hungary.
* Korea Aerospace Research Institute (KARI)/Korea.
* Ministry of Communications (MOC)/Israel.
* National Institute of Information and Communications Technology (NICT)/Japan.
* National Oceanic and Atmospheric Administration (NOAA)/USA.
* National Space Agency of the Republic of Kazakhstan (NSARK)/Kazakhstan.
* National Space Organization (NSPO)/Chinese Taipei.
* Naval Center for Space Technology (NCST)/USA.
* Scientific and Technological Research Council of Turkey (TUBITAK)/Turkey.
* South African National Space Agency (SANSA)/Republic of South Africa.
* Space and Upper Atmosphere Research Commission (SUPARCO)/Pakistan.
* Swedish Space Corporation (SSC)/Sweden.
* Swiss Space Office (SSO)/Switzerland.
* United States Geological Survey (USGS)/USA.

PREFACE

This document is a draft CCSDS Recommended Practice. Its ‘White Book’ status indicates that its contents are not stable, and several iterations resulting in substantial technical changes are likely to occur before it is considered to be sufficiently mature to be released for review by the CCSDS Agencies.

Implementers are cautioned **not** to fabricate any final equipment in accordance with this document’s technical content.

DOCUMENT CONTROL

|  |  |  |  |
| --- | --- | --- | --- |
| **Document** | **Title and Issue** | **Date** | **Status** |
| CCSDS 000.0-W-0.1 | Information Curation Process, Proposed Draft Recommended Practice, Issue 0.1 | April 2014 | Original proposed draft |
| CCSDS 000.0-W-0.1a | Information Curation Process, Proposed Draft Recommended Practice, Issue 0.1a | June 2014 | Added Scope and Purpose Text, Import Abbreviations and Terminology |
| CCSDS 000.0-W-0.2 | Information Curation Process, Proposed Draft Recommended Practice, Issue 0.2 | September 2014 | Reworked Abbreviations from other standards. Still need to incorporate Purpose and Scope and Terminology from other documents. |
| CCSDS 653.0-W-0.3 | Information Lifecycle Framework, Proposed Draft Recommended Practice, Issue 0.3 | June 2015 | Renamed document, Entire document reworked to include only material from the agreed project description document. |
| CCSDS 653.0-W-0.4 | Information Lifecycle Framework, Proposed Draft Recommended Practice, Issue 0.4 | June 2015 | Expanded descriptions of Lifecycle stages. Lifecycle activities section added. |
| CCSDS 653.0-W-0.4 | Information Lifecycle Framework, Proposed Draft Recommended Practice, Issue 0.4-DG | August 2015 | Current draft. Entire document reworked. Activities removed, Topics added. |
| CCSDS 653.0-W-0.5 | Information Lifecycle and Long Term Usage, Proposed Draft Recommended Practice, Issue 0.5 | August 2015 | Renamed document. Merged 2 V0.4 version as agreed at telecon. |
| CCSDS 653.0-W-0.6 | Information Lifecycle and Long Term Usage, Proposed Draft Recommended Practice, Issue 0.6 | October 2015 | Updated Activities, Updated list of topics |
|  |  |  |  |
| CCSDS 653.0-W-0.7 | Information Lifecycle and Long Term Usage, Proposed Draft Recommended Practice, Issue 0.7 | November 2015 | Current draft. Updates at CCSDS meeting. Remove detail section. |
|  |  |  |  |
|  |  |  |  |

CONTENTS

Section Page

[DOCUMENT CONTROL v](#_Toc434976366)

[CONTENTS vii](#_Toc434976367)

[1 Introduction 1-1](#_Toc434976368)

[1.1 purpose and scope 1-1](#_Toc434976369)

[1.2 applicability 1-3](#_Toc434976370)

[1.3 rationale 1-3](#_Toc434976371)

[1.4 conformance 1-4](#_Toc434976372)

[1.5 document structure 1-4](#_Toc434976373)

[1.6 definitions 1-4](#_Toc434976374)

[1.6.1 acronyms and abbreviations 1-4](#_Toc434976375)

[1.6.2 terminology 1-5](#_Toc434976376)

[1.7 NOMENCLATURE 1-6](#_Toc434976377)

[1.7.1 NORMATIVE TEXT 1-6](#_Toc434976378)

[1.7.2 INFORMATIVE TEXT 1-6](#_Toc434976379)

[1.8 References 1-7](#_Toc434976380)

[2 Overview 2-1](#_Toc434976381)

[3 INFORMATION TOPICS of Interest for Long-Term Perservation 3-1](#_Toc434976382)

[3.1 Information topiCS TABLE 3-1](#_Toc434976383)

[3.2 topiCS High Level Description 3-2](#_Toc434976384)

[3.2.1 Content Data 3-2](#_Toc434976385)

[3.2.2 Representation Information 3-3](#_Toc434976386)

[3.2.3 PDI – REFeFerence INformation 3-3](#_Toc434976387)

[3.2.4 PDI – Provenance Information 3-3](#_Toc434976388)

[3.2.5 PDI – Context Information 3-3](#_Toc434976389)

[3.2.6 PDI – Fixity Information 3-3](#_Toc434976390)

[3.2.7 PDI – ACCESS Rights INformation 3-3](#_Toc434976391)

[3.2.8 DESCRIPTIVE INFORMATION 3-3](#_Toc434976392)

[3.2.9 PACKAGING INFORMATION 3-3](#_Toc434976393)

[3.2.10 – Non Information Model Issues 3-3](#_Toc434976394)

[4 THE LIFECYCLE FRAMEWORK: THE MAIN STAGES and Information to be gathered 4-1](#_Toc434976395)

[4.1 Information Lifecycle stages 4-1](#_Toc434976396)

[4.1.1 Formulate Stage 4-2](#_Toc434976397)

[4.1.2 Implement Stage 4-2](#_Toc434976398)

[4.1.3 Operate Stage 4-2](#_Toc434976399)

[4.1.4 Exploit Stage 4-3](#_Toc434976400)

[5 LIfecycle Framework - Activities detail 5-4](#_Toc434976401)

[5.1 Formulate Stage 5-4](#_Toc434976402)

[5.1.1 Funder Develop Mandate Roadmap 5-4](#_Toc434976403)

[5.1.2 Call for Proposal Preparation and Submission 5-4](#_Toc434976404)

[5.1.3 Proposal Preparation and Submission 5-5](#_Toc434976405)

[5.1.4 Proposal ReView 5-6](#_Toc434976406)

[5.1.5 Initial Cost and Risk Assessment 5-6](#_Toc434976407)

[5.1.6 Data Set Appraisal 5-7](#_Toc434976408)

[5.1.7 Definition of Designated Community and Definition of Preservation Objective 5-8](#_Toc434976409)

[5.2 Implement Stage 5-9](#_Toc434976410)

[5.2.1 Specification of Preservation and Curation Requirements 5-9](#_Toc434976411)

[5.2.2 Tailoring of the Consolidation Process 5-10](#_Toc434976412)

[5.2.3 Tailoring of the Preserved Data Set Content and Filling of the corresponding Inventory Table 5-10](#_Toc434976413)

[5.2.4 Consultation and Agreement with Designated Community 5-11](#_Toc434976414)

[5.2.5 Implementation of Required Systems 5-12](#_Toc434976415)

[5.2.6 Test of Implementation prior to Operatios 5-13](#_Toc434976416)

[5.2.7 Put in Place Trained Staff 5-14](#_Toc434976417)

[5.2.8 Cost and Risk Assessment (Pre-Operation) 5-14](#_Toc434976418)

[5.3 Operate Stage 5-15](#_Toc434976419)

[5.3.1 Gathering of Missing PDSC ITems and Updateof the PDSC Table 5-15](#_Toc434976420)

[5.3.2 Data Ingestion – Collect (or Create) Data 5-15](#_Toc434976421)

[5.3.3 Data Ingestion – Process Collected Data 5-17](#_Toc434976422)

[5.3.4 Data Ingestion – Validate Collected and Processed Data 5-17](#_Toc434976423)

[5.3.5 Disseminate Data and MetadatA 5-18](#_Toc434976424)

[5.3.6 Analyze Collected and Processed Data 5-19](#_Toc434976425)

[5.3.7 Publish Data and Results 5-19](#_Toc434976426)

[5.3.8 Disseminate to Long-Term Archives 5-20](#_Toc434976427)

[5.3.9 Operations and Maintenance 5-20](#_Toc434976428)

[5.3.10 Cost and Risk Assessments (Continuing Operations and Maintenance) 5-21](#_Toc434976429)

[5.4 Exploit Stage 5-22](#_Toc434976430)

[5.4.1 Receive SIPs 5-22](#_Toc434976431)

[5.4.2 Validate SIPs 5-23](#_Toc434976432)

[5.4.3 Incorporate SIP data into Archival Holdings 5-23](#_Toc434976433)

[5.4.4 Reprocessing of the PReserved Data to Add Provenance in Accordance with Project PRoposal 5-24](#_Toc434976434)

[5.4.5 Disseminate Data and Metadata 5-25](#_Toc434976435)

[5.4.6 Long-Term Preservation According to (Possibly Updated) Preservation Plan 5-25](#_Toc434976436)

[5.4.7 Cost and Risk Assessment (Long-Term PreserVation) 5-26](#_Toc434976437)

[5.4.8 De-Accession If Appropriate 5-27](#_Toc434976438)

[6 [SECTION TITLE] 6-1](#_Toc434976439)

# Introduction

[Insert introductory subsections such as PURPOSE, SCOPE, APPLICABILITY, RATIONALE, etc. See CCSDS A20.0-Y-4, *CCSDS Publications Manual* (Yellow Book, Issue 4, April 2014) for the contents of section 1.]

## purpose and scope

The purpose of this Recommended Practice is to define a framework for gathering information throughout the information lifecycle, from the proposal to the long-term re-use of the resulting information, focusing on the issues that need to be addressed at each stage which will help to ensure that the data can be optimally exploited over the long term. It can form the basis on which Data Management Plans can be constructed. It should be applicable to projects where the data is existant as well as where data is to be created in the future. It should be of use to funders, researchers, archive managers and end-users by helping to reduce the effort and increase the efficacy of preservation and exploitation of data.

The Recommendation does not cover all aspects of the lifecycle and aspects of the activities it does specify do not have to be carried out strictly sequentially, and indeed some may be revisited and improved at several of the stages.

It will describe stages of the information lifecycle and within each stage, this recommendation identifies the information which should be collected, created or improved in order to be able to preserve and utilize information objects for the long-term. Within the framework, standards, best practices and software tools that could be used are identified.

This framework takes the view that curation and preservation are not separate activities to be considered at the end of an information production project, but as a set of actions that must be conducted throughout the information lifecycle.

Other aspects, such as costing, risk management, metadata management, data formats, policies and workflow, provenance-adding and service architectures, which are clearly of interest in preservation and curation, are not addressed except at a high-level, to provide context. It is expected that full treatment of these issues will require additional, more focused, standards.

While this process originates in the space community, it is being designed in a generic way and should be applicable to any science domain and to the wider library and archival communities.

This Recommended Practice accomplishes the following:

* identifies the main stages in the information lifecycle;
* forms a general methodological framework, which should be applicable and usable in any information stewardship, curation or preservation context (this general framework should provide sufficient flexibility to be applied to individual user’s situations);
* forms a basis for the identification and/or development of additional standards and implementation guides including those that address particular concerns in more detail;
* forms a basis for identification and/or development of a set of software tools that will assist the development, operation and checking of the different stages of the lifecycle.

This Recommended Standard fits into the context defined by:

* The *Reference Model for an Open Archival Information System (OAIS)* Recommended Practice (see reference **Error! Reference source not found.**).
* The *Producer-Archive Interface Methodology Abstract Standard* (PAIMAS) Recommended Practice (see reference **Error! Reference source not found.**).
* The *Producer-Archive Interface Specification* Recommended Standard (see reference **Error! Reference source not found.**)
* The *Auditing and Certification of Trustworthy Digital Repositories* Recommended Practice (see reference **Error! Reference source not found.**).
* The *Requirements for Bodies Providing Auditing and Certification of Candidate Trustworthy Digital Repositories* Recommended Practice (see reference **Error! Reference source not found.**).

The OAIS Best Practice Standard (see reference **Error! Reference source not found.**) is one of the most widely recognized and applied archival standards available today. An OAIS is an archive, consisting of an organization of people and systems, that has accepted the responsibility to preserve information and make it available for a Designated Community. It meets a set of such responsibilities as defined in this document. The model provides a framework for the understanding and increased awareness of archival concepts needed for long-term digital information preservation and access, and for describing and comparing architectures and operations of existing and future archives. It also guides the identification and production of OAIS related standards.

The PAIMAS Best Practice Standard (see reference **Error! Reference source not found.**) defines a methodology for transferring data from an Information Producer to an Archives based on the four following phases: Preliminary, Formal Definition, Transfer, Validation. Required activities during each phase are identified.

The PAIS Recommended Standard (see reference **Error! Reference source not found.**) provides the abstract syntax and an XML implementation of descriptions of data to be sent to an archive. These descriptions are negotiated agreements between the data Producer and the Archive and facilitate production of agreed data by the Producer and validation of received data by the Archive. The Recommended Standard includes an abstract syntax and one possible concrete implementation for the packages.

The Auditing and Certification Recommended Practice (see reference **Error! Reference source not found.**) provides metrics for use in assessing the trustworthiness of digital repositories and the Guidelines for Bodies Recommended Practice (see reference **Error! Reference source not found.**) provides procedures to be followed when conducting audits of digital repositories using those metrics.

## applicability

The considerations/processes defined in this document apply to any activities producing data which is (or may be) re-used and preserved for significant periods. It is applicable to those organizations and individuals who create information that may need long-term preservation and to organizations making information available for the long term.

## rationale

Data that is collected or created needs to have additional information associated with it if it is to be independently understandable, usable and trusted as being authentic. That additional information changes over time, as hardware, software, the general environment and users’ tacit knowledge changes. OAIS uses the terms Representation Information and Preservation Description Information for this associated information. Itmust be accumulated over the lifecycle of the data. For example Provenance Information will accumulate over time, recording the things which have happened to the data.

Funders are increasingly asking that Data Management Plans accompany any request for project funding, however these tend not to evolve with the project and are difficult to monitor.

Many data lifecycles have been proposed. However they do not focus on the activities needed at each stage which will help to ensure that the data can be optimally exploited over the long term.

There are a small number of generally applicable stages in the information lifecycle where, typically, the responsibility is handed on from one individual or team to another. Each of those teams has specific knowledge about the information which subsequent teams may not possess. Therefore there is a need to specify the information to be captured at each of those stages. Improvements may be needed to, for example, the Representation Information, which was recorded in an earlier stage; this may arise when the better understood or reformatted or re-processed in later stages.

Therefore there is a need for a lifecycle framework to provide guidance as to what additional information should be captured or improved through the various stages of the lifecycle.

This document should enable:

* the Producer (including for example scientists who create the data) to capture and record the relevant information in a timely manner;
* the Archive to be assured that it will receive adequate information to enable it to perform preservation activities and support exploitation of the information
* the user to re-use unfamiliar information more easily
* the funder/sponsor to be assured that the resources that they contribute to the creation of the information will have suitable pay-back
* etc

Digital Content requires active management throughout its entire life.

## conformance

TBD text

## document structure

TBD text

## definitions

TBD text

### acronyms and abbreviations

|  |  |
| --- | --- |
| **ADMP** | Active Data Management Plan |
| **CCSDS** | Consultative Committee for Space Data Systems |
| **DMP** | Data Management Plan |
| **OAIS** | Open Archival Information System |
| **PAIMAS** | Producer-Archive Ingest Methodology Abstract Standard |
| **PAIS** | Producer-Archive Ingest Specification |
| **XML** | eXtensible Markup Language |
|  |  |

### terminology

Note: In common usage there is confusion/overlap between many of these terms. To avoid confusion and to allow precision, we will need specific definitions that will apply within the recommendation. Hopefully these definitions will resonate with others and will gain wider usage. However we expect that existing domains will continue with their own usage, but we expect that they will be easily able to map their usages to the usage within the recommendation.

SEE TABLE 3-1

We need definitions compatible with diagram in following section:

**Preservation**: (does not include Consolidation)

**Curation:** (wider concept that includes Preservation and Consolidation)

**Stewardship** (wider concept that includes Curation)

**Consolidation**  (Includes Collection and Integration)

**Preservation**: (does not include Consolidation) (Includes Integrity, Usability, and Accessibility)

**Curation:** (wider concept that includes Preservation and Consolidation) (Includes Selection, Consolidation, Organization, Presentation and Preservation of Data/Information and Data Record Improvement)

**Stewardship** (wider concept that includes Curation) (Includes Funding, Management, Planning and Certification of Data/Information Projects and Curation and Valuing Adding for Data/Information)

**Metadata:** data about data. OAIS identifies several categories of metadata including Representation Information (including Structure, Semantic, and Other Representation Information), Preservation Description Information (including Reference Information, Context Information, Provenance Information, Fixity Information, and Access Rights Information), Packaging Information, and Descriptive Information.

**Data Record:** text TBD

**Preserved Data Set Content (PDSC):** text TBD

**Active Data Management Plan :** An evolving record of the metadata needed to be confident of the authenticity and re-usability of the information which has been created.

**Curation Project :** an activity planned and designed to achieve a particular aim ranging from the creation of new data to the preservation of existing data with a particular preservation aim.

**Data Management Plan** (Wikipedia): A data management plan or DMP is a formal document that outlines how you will handle your data both during your research, and after the project is completed. The goal of a data management plan is to consider the many aspects of data management, metadata generation, data preservation, and analysis before the project begins. This ensures that data are well-managed in the present, and prepared for preservation in the future.

## NOMENCLATURE

### NORMATIVE TEXT

The following conventions apply for the normative specifications in this Recommended Standard:

1. the words ‘shall’ and ‘must’ imply a binding and verifiable specification;
2. the word ‘should’ implies an optional, but desirable, specification;
3. the word ‘may’ implies an optional specification;
4. the words ‘is’, ‘are’, and ‘will’ imply statements of fact.

NOTE – These conventions do not imply constraints on diction in text that is clearly informative in nature.

### INFORMATIVE TEXT

In the normative sections of this document (sections **Error! Reference source not found.**-**Error! Reference source not found.**), informative text is set off from the normative specifications either in notes or under one of the following subsection headings:

* Overview;
* Background;
* Rationale;
* Discussion.

## References

The following publications contain provisions which, through reference in this text, constitute provisions of this document. At the time of publication, the editions indicated were valid. All publications are subject to revision, and users of this document are encouraged to investigate the possibility of applying the most recent editions of the publications indicated below. The CCSDS Secretariat maintains a register of currently valid CCSDS publications.

[Only references required as part of the specification are listed in the References subsection. See CCSDS A20.0-Y-3, *CCSDS Publications Manual* (Yellow Book, Issue 3, December 2011) for additional information on this subsection.]

[Editor note: References below are from the project description document. I don’t think any of those except for the CCSDS Recommendations will actually be referenced. I dimmed the items that I don’t expect to be in the final reference list. It is also possible to add an information only reference list.]

[1] *Reference Model for an Open Archival Information System (OAIS)*. Recommendation for Space Data System Practices, CCSDS 650.0-M-2. Blue Book. Issue 1. Washington, D.C.: CCSDS, June 2012. [Equivalent to ISO 14721:2012.] Available from: <http://public.ccsds.org/publications/archive/650x0m2.pdf>

[2] *Producer-Archive Interface Methodology Abstract Standard*. Recommendation for Space Data System Practices, CCSDS 651.0-M-1. Magenta Book. Issue 1. Washington, D.C.: CCSDS, May 2004. [Equivalent to ISO 20652:2006.] Available from: <http://public.ccsds.org/publications/archive/651x0m1.pdf>

[3] *Producer-Archive Ingest Specifications*. Recommendation for Space Data System Standards, CCSDS 651.1-B-1. Blue Book. Issue 1. Washington, D.C.: CCSDS, February 2014. [Equivalent to ISO 20104:2015] Available from: <http://public.ccsds.org/publications/archive/651x1b1.pdf>

[4] *Audit and Certification of Trustworthy Digital Repositories*. Recommendation for Space Data System Practices, CCSDS 652.0-M-1. Magenta Book. Issue 1. Washington, D.C.: CCSDS, September 2011. [Equivalent to ISO 16363:2012.] Available from: <http://public.ccsds.org/publications/archive/652x0m1.pdf>

[5] *Requirements for Bodies Providing Audit and Certification of Candidate Trustworthy Digital Repositories*. Recommendation for Space Data System Practices, CCSDS 652.1-M-2. Magenta Book. Issue 2. Washington, D.C.: CCSDS, March 2014. [Equivalent to ISO 16919:2014.] Available from: <http://public.ccsds.org/publications/archive/652x1m2.pdf>

# Overview

[Non-normative overview text appears in section 2. See CCSDS A20.0-Y-4, *CCSDS Publications Manual* (Yellow Book, Issue 4, April 2014) for the contents of section 2.]

This standard deals with the entire information and data lifecycle. An important aspect of this standard is the terminology used within this standard. Many of these terms are already used with various definitions within the target communities for this standard – e.g. space, science, library and archival communities. The terminology defined will be applicable within this standard, but we expect that the target communities can easily map our terminology to the terminology used within their own communities. The OAIS Reference Model provided a starting point and inputs from a variety of other sources were used to arrive at the terms used within this standard.

# INFORMATION TOPICS of Interest for Long-Term Perservation

[Normative specifications appear in sections 3 through *n*. See CCSDS A20.0-Y-4, *CCSDS Publications Manual* (Yellow Book, Issue 4, April 2014).

In the following we use the term “curation project” to mean quite generally an activity “planned and designed to achieve a particular aim ranging from the creation of new data to the preservation of existing data with a particular preservation aim.

Section 4.1 outlines the stages of the lifecycle, section **Error! Reference source not found.** identifies the major pieces of information related to curation which need to be collected and section **Error! Reference source not found.** shows the way in which that information evolves through the lifecycle.

## Information topiCS TABLE

**Error! Reference source not found.** indicates the topics under which the curation information is grouped.

Table 3‑1 Curation information to be collected

|  |  |
| --- | --- |
| **Topic** | **Brief description** |
| **Content Data** | The data created or collected which may be re-used and preserved |
| **Representation Information** | See OAIS Representation Information |
| **Reference Information** | See OAIS Reference Information |
| **Provenance Information** | See OAIS Provenance |
| **Context Information** | See OAIS Representation Information |
| **Fixity Information** | See OAIS Fixity |
| **Access Rights Information** | See OAIS Access Rights |
| **Packaging Information** | See OAIS Packaging |
| **Descriptive Information** | See OAIS Descriptive Information |
| **Issues Outside the Information Model** | Issues that do not fit cleanly into the OAIS Information Model |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |

## topiCS High Level Description

The Topics covered in this document are organized around the Archival Information Package (AIP) Information Model Components. OAIS defines four major categories of information related to an AIP – Content Information, Preservation Descriptive Information (PDI), Packaging Information and Descriptive Information. An AIP consists of Content Information and PDI. Content Information is composed of Content Data and Representation Information. PDI is composed of Reference Information, Provenance Information, Context Information, Fixity Information, and Access Rights Information. Packaging Information delimits the Information Package and Descriptive Information is used for creating Access Aids to find and access AIPs.

### Content Data

Content Information includes the Data Objects as well as Structure Information, Semantic Information and Other Representation Information.

The project proposal will identify the types of data products that the project intends to produce in general terms (tables, images, maps).  The Project Data Management Plan will provide more detail, including actual format specifications where available.  If an existing standard is used then the documentation requirements are minimized, as the organization, standard name and version and URL can be referenced.  This will satisfy the need for Structure Information but not the need for Semantic Information.  If a standard format is not used then the project will need to provide documentation which describes the Structure Information.  There are data description languages (DFDL, EAST) and registration schemes (SFDU, XFDU) that can be used to define the explicit structure of data objects.  [NOTE:  I would like to talk about format registries, but I am unclear which ones are active (UFDR) and which are dead (GDFR)or comatose (CCSDS)].  A Software Interface Specification (SIS) document is often used to provide an explicit description of the byte by byte structure as well as the interpretation of the values in the data object.  There are also data definition standards (DEDSL, ISO11179) which can be used to describe the meaning of data values in a standard way.  The use of these standards will promote interoperability between information systems.  The category Other Representation Information includes software, algorithms, encryption and written instructions.  These items should all be described in the SIS document.  If software is to be included as a deliverable it should follow the guidelines for submission to a public software repository (e.g. GITHUB).  Note that some archives don't accept software, so such a public repository may be the only way to provide software to future data users.

The data objects should be the same objects that are provided to the project team for analysis and should be delivered in the same format as used by the project. Thus it is important that the project understand the archive format requirements in the design stage to avoid having to transform the data prior to delivery. Such transformations are extremely risky and require an extra validation step. The combination of data objects and representation information should provide the capability to recreate any results cited in publications. If not then those products should also be delivered to the archive.

Guidelines for data formats.

1. Use the established format standards of the designated community throughout the data collection, processing and analysis activities. Wherever possible, use existing community and commercial tools to access and analyze data objects.

2. Use open, registered, formally-documented formats with defined mime-types and standard file extensions that can be inspected with widely available tools.

3. Avoid the use of machine or platform dependent data types, interleaving of logical objects and compression or encoding schemes.

4. Use formats that contain embedded structural information required to view the data object as well as semantic information which identifies the format name and version as well as attributes necessary to interpret the object.

5. Text-based formats (XML, JSON, CSV) for tabular data and simple binary arrays of 8 or 16 bit integers allow data inspection with common utilities.

### Representation Information

### PDI – REFeFerence INformation

### PDI – Provenance Information

### PDI – Context Information

### PDI – Fixity Information

### PDI – ACCESS Rights INformation

### DESCRIPTIVE INFORMATION

### PACKAGING INFORMATION

### – Issues Outside the Information Model

\*\*\*

# THE LIFECYCLE FRAMEWORK: THE MAIN STAGES and Information to be gathered

[Normative specifications appear in sections 3 through *n*. See CCSDS A20.0-Y-4, *CCSDS Publications Manual* (Yellow Book, Issue 4, April 2014).

In the following we use the term “curation project” to mean quite generally an activity “planned and designed to achieve a particular aim ranging from the creation of new data to the preservation of existing data with a particular preservation aim.

Section 4.1 outlines the stages of the lifecycle. Subsections 4.2 to 4.5 provide a brief description of each stage. The following subsections provide details of activities identified for each stage and relate them to the information topics identified in Section 3.

## Information Lifecycle stages

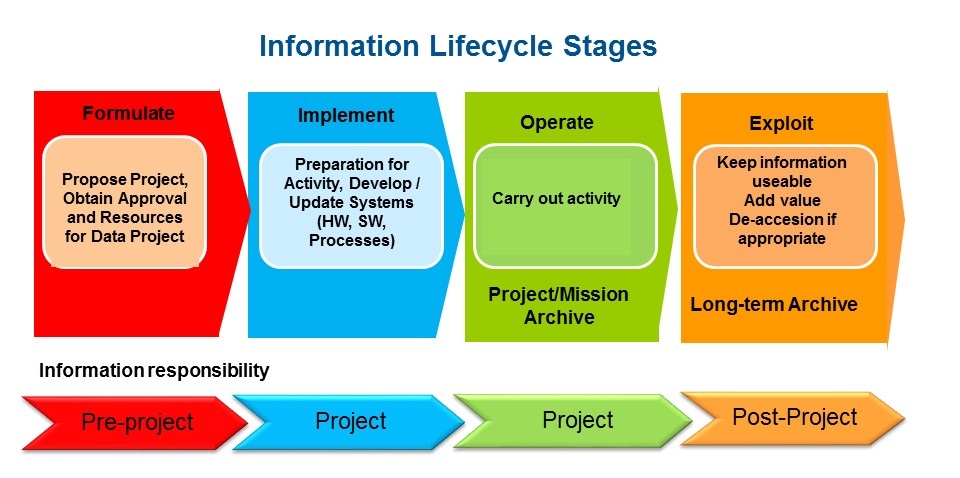


Figure 4‑4‑1 Stages of the Information Curation Lifecycle

Figure 4‑4‑1 above identifies the stages of the information lifecycle:

1. The Formulate stage is where those responsible for domains call for, review and chose among proposed projects to acquire or generate, manage or support exploitation of data or information gain approval and resources to carry out that activity and where those responsible for proposing the projects gain approval and resources to carry out that activity.

2) The Implement stage is where those responsible prepare to carry out the activity. The project develops (or updates) systems (hardware, software, data and processes to meet the needs of the proposed activity.

3) The Operate stage is where those responsible carry out the activity to acquire or generate, manage, and support exploitation of data or information from this activity

4) The Exploit stage is where those responsible maintain the data or information in a useable form and add Provenance if possible and proposed as part of activity. It also handles de-accessioning of the data if appropriate.

The next subsection will provide a few additional paragraphs describing the high-level actors, activities, and deliverables. Within each of these stages, we will then provide a more detailed breakdown of the high-level activities and typical deliverables. Once this framework has been defined, we will be able to provide guidelines for activities that need to be performed during these stages to address a number of different data and information stewardship concerns. Within this first Recommendation we will focus on preservation. We intend to identify and provide guidelines for preservation concerns and activities that would ideally take place at each stage. Subsequent Recommendations will address additional aspects that should be addressed at each stage of the Information Lifecycle Framework. For example, future Recommendations or issues of this first Recommendation could address Data Management Plans, Risk Management issues, etc.

### Formulate Stage

The participants in the Formulate stage will certainly include Sponsors and Proposers and may also include Data Managers and Archives.

Sponsors typically have a well-defined charter that identifies their area of interest, within which, for example, programs are identified for specific research topics. Sponsors may bring in domain experts from existing projects and interest groups to develop strategic plans and objectives for the programs as well as long term schedules to fulfill the plans and objectives.

Proposers develop proposals that describe potential projects The proposals are likely to provide estimates of the cost and a preliminary risk assessment. The proposer may also prepare a data management plan that identifies the information gathering system, data processing system, and the types of products that will be produced. The data management plan will include estimates of data volumes and a schedule of activities and deliveries.

The proposer will often have to work with an archive to understand the standards in place and the mechanisms for delivering information objects. The archive may be a co-signatory for the proposal or at least the Data Management Plan. The archive uses the data management plan to develop a support plan that identifies the cost and resources that will be required to support the Project. These resource requirements need to be integrated into the project’s and the archive's long term budget and schedule.

Some projects may be able to use the Proposal and Data Management Plan as their system design. But for large and complex projects the selection is only the beginning of a substantial design process. We identify a System Design document to represent what may actually represent a sequence of requirements and design documents. The initial project design includes all elements of the acquisition, transfer, processing, distribution and analysis systems. For large projects there may be substantial negotiation and iteration between the sponsor and the project during the design process.

“Information produced during this stage provides a snapshot of the scientific and technical framework in which the project was born. Project … requirements, assessment studies, technology readiness reviews and cost analysis are performed during this stage. Preserving this information – both for approved and not approved projects – would allow future users to have reference material for new missions evaluation and definition. Traceability of this information is also useful to compare initial expectations to what was actually achieved by the mission and to understand which changes occurred between the pre-mission and the next stages.” (LTDP PDSC document).

### Implement Stage

The implement stage is performed by the project team. Once the project proposal is approved, the project implementation team develops the information acquisition systems or provides access to existing data sources. This could be as complicated as building a spacecraft or as simple as accessing an online database. The project implementation team may provide command and control systems to interact with the information collection systems. The project implementation team develops processing software and scripts. It provides data processing and distribution systems which include analysis and access software. It documents the information content of all these systems via interface specifications, information models and data dictionaries.

“This stage produces the entire project and data detailed definition documents. It includes Sensor/Instrument requirements, characteristics, calibration methods, etc. Preserving this information is fundamental to understand changes that may have occurred over time while in operation. It is also needed to understand procedural impacts relative to instrument, algorithm and product implementation. Data acquired during the calibration and validation campaigns of the instrument under construction (e.g. in a laboratory or dedicated campaigns) is of critical importance as a reference for the future use of the data.” (LTDP PDSC document).

### Operate Stage

The operate stage is performed by the project team possibly with support from the archive. It operates the acquisition systems to gather data. The project team receives and processes the data to support both command and control and data analysis. The project team maintains a record of all acquisition and processing activities that might be useful for understanding and interpreting the data. It performs data analysis and publishes results. The project team operates the implemented distribution mechanisms for transferring data to the project team and other users. The project team prepares Submission Information Packages (SIPs) for delivery to the long-term archive and participates in peer reviews.

### Exploit Stage

The exploit stage is performed by the long-term archive possibly with support from the project. The archive receives and validates SIP deliveries from the project. It performs any conversion or consolidation necessary to integrate the data and metadata content of the SIPs into the archive collection. It transfers the data contents of the SIPs to the archive repository. It extracts metadata from the SIPs and loads it into the archive registry or database. It provides an access capability for users to be able to search, select and download both metadata and data from the archive collection. It provides for long term preservation by insuring the integrity, usability and accessibility of the data.

The differences between the operation and the exploitation stages are in the source of funding, the designated community and maturity of the products. The project and the archive may have completely separate funding sources. The project users may be limited to a small team whereas the archive may represent a large community. Special hardware and software capabilities may be provided to project teams but not to archive users. The archive may provide capabilities for exploiting the project data in concert with other data products.

# LIfecycle Framework - Activities detail

[Editor note: Original activity text below were extracted from EU EOS LTDP Framework document. The activities were placed in the sections identified in the mapping by JGG provided prior to the previous telecon. My updates/comments are tracked.]

Figure 5-1 below indicates the status of information capture for each of the topical issues at each of the lifecycle stages. Typically information to address the issue and to document the decisions made in regard to each of these topical issues will begin to be accumulated early in the lifecycle. Then as time goes on more information is gained until the needed information is complete. In the case where new information about a topical issue will continue to be generated, then by late in the information lifecycle, the collected information should be up to date. And even once complete, maintenance efforts and periodic reviews should be made to ensure that the information remains up to date.

\*\*\*

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Topic** | **Issue** | **Formulation** | **Implementation** | **Operation** | **Exploitation** |
|  |  |  |  |  |  |
| Content Data | Inventory of data produced/ expected | Rough idea | Increasingly detailed | Becoming complete | Complete |
| Types of data (raw, processed, etc.) which should be preserved? | Rough idea | Increasingly detailed | Becoming complete | Complete |
| Type of data e.g. images, tables – which generic interfaces? | Rough idea | Increasingly detailed | Becoming complete | Complete |
| Volume that would require preservation | Rough idea | Increasingly detailed | Becoming complete | Complete |
| Quality constraints | Rough idea | Increasingly detailed | Becoming complete | Complete |
| Quality checks which may be performed on the data by non-experts | Rough idea | Increasingly detailed | Increasingly detailed | Complete |
| Representation Information | Choice of data format | Rough idea | Increasingly detailed | Becoming complete | Complete |
| Format definitions and formal descriptions | Rough idea | Increasingly detailed | Becoming complete | Up to date and accumulating |
| Semantics of the data elements | Rough idea | Increasingly detailed | Becoming complete | Almost complete |
| Data dictionaries and other semantics | Rough idea | Increasingly detailed | Becoming complete | Up to date and accumulating |
| Information Model | Rough idea | Increasingly detailed | Becoming complete | Complete |
| Other Data Documentation | Rough idea | Increasingly detailed | Becoming complete | Up to date and accumulating |
| Applicable standards | Rough idea | Increasingly detailed | Becoming complete | Complete |
| Hardware and Software Dependencies | Rough idea | Increasingly detailed | Becoming complete | Up to date and accumulating |
| Other software which may be used on the data |  | Increasingly detailed | Increasingly detailed | Growing |
| What calibration and system test tools and system test data will be delivered. | Rough idea | Increasingly detailed | Becoming complete | Up to date and accumulating |
| Relationships between data items | Rough idea | Increasingly detailed | Complete | Complete |
| Reference Information | DOI or other unique identifiers | Rough idea | Becoming complete | Up to date and accumulating | Up to date and accumulating; New methods could be introduced |
| Rules, methods, tools for referencing data | Rough idea | Becoming complete | Up to date and accumulating | Up to date and accumulating; New methods could be introduced |
| What standards will be used to ~~format,~~ identify and reference the data and metadata | Rough idea | Becoming complete | Up to date and accumulating | Up to date and accumulating; New methods could be introduced |
| What may be used in future to identify the data | Fairly firm | Increasingly detailed | Increasingly detailed | Evolving |
|  |  |  |  |  |
| Provenance Information | Record of origins of the project e.g. in a CRIS system | Fairly firm | Complete | Completed | Complete |
| Documentation about the hardware and software used to create the data, including a history of the changes in these over time |  | Rough Idea then Increasingly detailed | Becoming complete | Up to date and accumulating |
| Processing workflow | Rough idea | Increasingly detailed | Becoming complete | Complete |
| Processing inputs |  | Rough Idea then Increasingly detailed | Becoming complete | Complete |
| Processing parameters | Rough idea | Increasingly detailed | Becoming complete | Complete |
| Who was responsible for each stage of processing |  | Increasingly detailed | Becoming complete | Complete |
| When each stage was performed |  | Increasingly detailed | Becoming complete | Complete |
| Record of any special hardware needed (JGG for what?) | Rough idea | Increasingly detailed | Becoming complete | Complete |
| Calibration | Rough idea | Becoming complete | Complete | Complete |
| System Testing | Rough idea | Becoming complete | Up to date and accumulating | Up to date and accumulating; New methods could be introduced |
| Resident Archives |  |  | Rough idea | Becoming complete |
| Who was responsible for each stage of processing (Fixity) |  | Up to date and accumulating | Up to date and accumulating | Up to date and accumulating |
| Context | Outline of background concepts needed to understand the project | Rough idea | Increasingly detailed | Becoming complete | Up to date and accumulating |
| Publications related to the data (Risk) |  |  | Rough idea | Evolving |
| Publications related to the data (publications) | Rough idea | Increasingly detailed | Up to date and accumulating | Up to date and accumulating |
| Related data which may in the future be combined with this data |  | Increasingly detailed | Increasingly detailed | Growing |
| Potential Value of the data and likely business case for sustainability | Rough idea | Rough idea | Developing | Developing |
| Identification of archives which are likely to be able to host the data | Rough idea | Increasingly detailed | Increasingly detailed | Complete |
| Provide a bibliography of related publications | Rough idea | Increasingly detailed | Up to date and accumulating | Up to date and accumulating |
|  |  |  |  |  |
| Fixity | Fixity (e.g. CRC or digest) of data which may be preserved |  | Complete | Complete, but may Evolve | Complete, but may Evolve |
| How do we verify that all files are intact |  | Complete | Complete, but may Evolve | Complete, but may Evolve |
| Identify any special validation procedures that should be carried out. |  | Complete | Complete, but may Evolve | Complete, but may Evolve |
| Access Rights Information | What are the restrictions on access in the long term |  | Complete | Complete, but may Evolve | Complete, but may Evolve |
| Clear identification of Intellectual Property Rights |  | Complete | Complete, but may Evolve | Complete, but may Evolve |
| Licenses involved | Rough idea | Complete | Complete | Complete |
| Owners of the data – who can authorize hand-over | Rough idea | Complete | Complete | Complete |
| Who is the owner, what are the restrictions on access (licenses), what are intellectual property rights | Rough idea | Complete | Complete, but may Evolve | Complete, but may Evolve |
|  |  |  |  |  |
| Packaging Information | Details of the way components are packaged together for delivery to a repository |  | Increasingly detailed | Complete | Complete |
| Definition of SIPs |  | Developing | Complete | Complete |
|  |  |  |  |  |
| Descriptive Information | Methods for exploration/ quick-look at the data | Fairly firm | Increasingly detailed | Increasingly detailed | Evolving |
| Is browse data needed? | Fairly firm | Complete | Complete, but may Evolve | Complete, but may Evolve |
|  |  |  |  |  |
| Issues Outside the Information Model | Schedule of deliveries | Fairly firm | Increasingly detailed | Complete |  |
| Cost | Fairly firm | Increasingly detailed | Complete, but may Evolve | Complete, but may Evolve |
| Pointers to the components to be transferred to the archive |  | Fairly firm | Complete | Complete, but may Evolve |
| Potential preservation aims of the archive | Rough idea | Increasingly detailed | Increasingly detailed | Complete |
| Potential risks to preservation and exploitation of the data | Fairly firm | Increasingly detailed | Complete, but may Evolve | Complete, but may Evolve |
| What are the target archives and designated community for the solicitation. | Fairly firm | Complete | Complete, but may Evolve | Complete, but may Evolve |
| What is the budget for archiving. | Fairly firm | Complete | Complete, but may Evolve | Complete, but may Evolve |
| What is the schedule for major project milestones and deliveries to the archive. | Fairly firm | Complete | Complete, but may Evolve | Complete, but may Evolve |
| Change Management |  | Complete | Complete, but may Evolve | Complete, but may Evolve |
| What is the mechanism for communication between project and archive. | Fairly firm | Complete | Complete, but may Evolve | Complete, but may Evolve |
|  |  |  |  |  |

Table 5-1: Status of Information Capture for Topical Issues at Lifecycle Stages

## TBD







































































































































# [SECTION TITLE]

[Normative specifications appear in sections 3 through *n*. See CCSDS A20.0-Y-4, *CCSDS Publications Manual* (Yellow Book, Issue 4, April 2014).

All sections and annexes should be separated by Word continuous section breaks.]

1. Mapping to LTDP Workflow  
     
   (Informative)

[Annexes contain ancillary information. Normative annexes precede informative annexes. Informative references are placed in an informative annex. See CCSDS A20.0-Y-4, *CCSDS Publications Manual* (Yellow Book, Issue 4, April 2014) for discussion of the kinds of material contained in annexes.]

1. Mapping to TBD (Others besides LTDP)  
     
   (Informative)

[Annexes contain ancillary information. Normative annexes precede informative annexes. Informative references are placed in an informative annex. See CCSDS A20.0-Y-4, *CCSDS Publications Manual* (Yellow Book, Issue 4, April 2014) for discussion of the kinds of material contained in annexes.]

1. Security Considerations  
     
   (Informative)
   1. Introduction
   2. security concerns with respect to the CCSDS document
      1. Data privacy
      2. Data integrity
      3. Authentication of communicating entities
      4. Control of access to resources
      5. Availability of resources
      6. Auditing of resource usage
   3. Potential threats and attack scenarios
   4. Consequences of not applying security to the technology