

Draft Recommendation for
Space Data System Practices

|  |
| --- |
| Information Lifecycle framework |

PROPOSED Draft Recommended Practice

CCSDS 653.0-W-0.4

WHITE Book

* June 2015

AUTHORITY

|  |
| --- |
|  |
|  | Issue: | White Book, Issue 0.4 |  |
|  | Date: | June 2015 |  |
|  | Location: | Not Applicable |  |
|  |

**(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 | Current draft. Expanded descriptions of Lifecycle stages. Lifecycle activities section added. |
|  |  |  |  |
|  |  |  |  |

CONTENTS

Section Page

[DOCUMENT CONTROL v](#_Toc423470580)

[CONTENTS vi](#_Toc423470581)

[1 Introduction 1-1](#_Toc423470582)

[1.1 purpose and scope 1-1](#_Toc423470583)

[1.2 applicability 1-4](#_Toc423470584)

[1.3 rationale 1-4](#_Toc423470585)

[1.4 conformance 1-4](#_Toc423470586)

[1.5 document structure 1-5](#_Toc423470587)

[1.6 definitions 1-5](#_Toc423470588)

[1.6.1 acronyms and abbreviations 1-5](#_Toc423470589)

[1.6.2 terminology 1-7](#_Toc423470590)

[1.7 NOMENCLATURE 1-8](#_Toc423470591)

[1.7.1 NORMATIVE TEXT 1-8](#_Toc423470592)

[1.7.2 INFORMATIVE TEXT 1-9](#_Toc423470593)

[1.8 References 1-9](#_Toc423470594)

[2 Overview 2-1](#_Toc423470595)

[3 THE LIFECYCLE FRAMEWORK: THE MAIN STAGES, ACTIVITIES, AND DELIVERABLES 3-1](#_Toc423470596)

[3.1 Information Lifecycle stages 3-2](#_Toc423470597)

[3.2 Next SubSection 3-7](#_Toc423470598)

[4 LIfecycle Framework Activities detail 4-1](#_Toc423470599)

[4.1 Formulation Stage 4-1](#_Toc423470600)

[4.2 Implementation Stage 4-3](#_Toc423470601)

[4.3 Operation StagE 4-5](#_Toc423470602)

[4.4 exploitation Stage 4-6](#_Toc423470603)

[5 Long-Term Preservation Requirements 5-1](#_Toc423470604)

[6 Other Possible Requirements 6-1](#_Toc423470605)

[7 [SECTION TITLE] 7-1](#_Toc423470606)

# 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 the information lifecycle, from the proposal to the disposition of the resulting information objects, focusing on the activities needed at each stage which will help to ensure that the data can be optimally exploited over the long term. It should be applicable to both existing as well as future data and 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 the objectives, high-level activities and typical deliverables of each stage. Within each stage, this recommendation will identify concerns that must be addressed in order to ensure that the information needed to preserve and utilize information objects for the long-term is collected and/or created at the optimal time. It will identify standards, best practices and software tools that could be applied to address these concerns. Subsequent Recommendations will address aspects that should be addressed at each stage of the Information Lifecycle Framework in more detail. For example, future Recommendations or issues of this first Recommendation could address Data Management Plans, Risk Management issues, etc.

This framework considers 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. A number of high-level steps and associated activities will be identified that support curation and preservation at each stage of the lifecycle. Other concerns, such as data management plans, costing, risk management, metadata management, data formats, policies and workflow, value-adding and service architectures, are addressed at a high-level especially where they impact the curation and preservation. It is expected that these others concerns are identified in this recommendation but full treatment of the issues raised will require additional, more focused standards.

While this process is originating 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 different stages in the data/information lifecycle ;

– defines the objective of each of these stages, a high-level set of concerns that should be addressed during each of these stages, some possible actions to address those concerns and a typical set of expected deliverables (e.g., administrative, technical, contractual) at the end of a phase;

– forms a general methodological framework, which should be able to be applied and reused 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.

The purpose of this Recommended Practice is to provides a TBD process**.** This supports effective preservation and curation of information needed over the long-term.

This Recommended Standard fits into the context defined by:

* The *Reference Model for an Open Archival Information System (OAIS)* Recommended Practice (see reference [1]).
* The *Producer-Archive Interface Methodology Abstract Standard* (PAIMAS) Recommended Practice (see reference [4]).

The *Producer-Archive Interface Specification* Recommended Standard (see reference 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/DIS 20104:2015] .] Available from: http://public.ccsds.org/publications/archive/651x1b1.pdf

* ).
* The *Auditing and Certification of Trustworthy Digital Repositories* Recommended Practice (see reference [4]).
* The *Requirements for Bodies Providing Auditing and Certification of Candidate Trustworthy Digital Repositories* Recommended Practice (see reference [5]).

The OAIS Best Practice Standard (see reference [1]) 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 [4]) 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 [3]) 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 [4]) provides metrics for use in assessing the trustworthiness of digital repositories and the Guidelines for Bodies Recommended Practice (see reference [5]) provides procedures to be followed when conducting audits of digital repositories using those metrics.

## applicability

The considerations/processes defined in this document applies to any data producing activities. It is particularly 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.

These considerations/processes/practices are relevant in any data producing activities and can be informative regardless of any agreements that are in place or any shared implementations.

## rationale

TBD text

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

Regarding the TBD, this Recommended Standard should enable:

* the Producer to …;
* the Archive to …
* the scientist to …
* the user to …
* the sponsor to …
* etc

Digital Content requires active management throughout it’s entire period of use/data lifecycle.

## conformance

TBD text

## document structure

TBD text

## definitions

TBD text

### acronyms and abbreviations

Editor’s Note: The last column (Source) is included for our information during development. It will not be included in the final document.

|  |  |  |
| --- | --- | --- |
| **CCSDS** | Consultative Committee for Space Data Systems | OAIS, PAIMAS, PAIS |
| **OAIS** | Open Archival Information System | OAIS, PAIMAS, PAIS |
| **PAIMAS** | Producer-Archive Ingest Methodology Abstract Standard | PAIMAS, PAIS |
| **PAIS** | Producer-Archive Ingest Specification | PAIS |
| **XML** | eXtensible Markup Language | OAIS, PAIMAS, PAIS |
|  |  |  |
|  | Below are items from OAIS, PAIMAS, and PAIS (which are copied here in case we decide to use any of them in this document) |  |
|  |  |  |
| **AIC** | Archival Information Collection | OAIS |
| **AIP** | Archival Information Package | OAIS, PAIMAS, PAIS |
| **AIU** | Archival Information Unit | OAIS |
| **API** | Application Programming Interface | OAIS |
| **ASCII**  | American Standard Code for Information Interchange | OAIS, PAIMAS |
| **CCSDS CA** | CCSDS Control Authority | PAIS |
| **CD-ROM** | Compact Disk - Read Only Memory | OAIS |
| **CDO** | Content Data Object | OAIS |
| **CRC** | Cyclical Redundancy Check | OAIS, PAIS |
| **DBMS** | DataBase Management System | OAIS |
| **DED** | Data Entity Dictionary | PAIMAS, PAIS |
| **DEDSL** | Data Entity Dictionary Specification Language (reference [**Error! Reference source not found.**]) | PAIMAS |
| **DIF** | Directory Interchange Format | PAIMAS |
| **DIP** | Dissemination Information Package | OAIS |
| **DRM** | Digital Rights Management | OAIS |
| **DTD** | Document Type Definition | PAIMAS, PAIS |
| **EAD** | Encoded Archival Description | PAIMAS |
| **EAST** | Enhanced Ada SubseT | PAIMAS |
| **FGDC** | Federal Geographic Data Committee | PAIMAS |
| **FITS** | Flexible Information Transport System | OAIS |
| **FTP** | File Transport Protocol | OAIS, PAIS |
| **GB** | Gigabyte | PAIS |
| **HFMS** | Hierarchical File Management System | OAIS |
| **ICA** | International Council on Archives | PAIMAS |
| **ID** | Identifier | PAIS |
| **IEEE** | Institute of Electrical and Electronic Engineers | OAIS, PAIMAS |
| **IETF** | Internet Engineering Task Force | PAIS |
| **ISBN** | International Standard Book Number | OAIS |
| **ISO** | International Organization for Standardization | OAIS, PAIMAS, PAIS |
| **KB** | Kilobyte | PAIS |
| **MARC** | MAchine-Readable Cataloging | PAIMAS |
| **MB** | Megabyte | PAIS |
| **MIME** | Multipurpose Internet Mail Extensions | PAIS |
| **MOT** | Model of Objects for Transfer | PAIS |
| **MPEG** | Moving Picture Experts Group | OAIS |
| **PB** | Petabyte | PAIS |
| **PDF** | Portable Document Format | OAIS, PAIMAS |
| **PDI** | Preservation Description Information | OAIS, PAIMAS |
| **PVL** | Parameter Value Language | PAIMAS |
| **QA** | Quality Assurance | OAIS |
| **RM** | Reference Model | PAIMAS |
| **SGML** | Standard Generalized Markup Language | PAIMAS |
| **SIP** | Submission Information Package | OAIS, PAIMAS, PAIS |
| **TB** | Terabyte | PAIS |
| **TEI** | Text Encoding Initiative | PAIMAS |
| **TOTD** | Transfer Object Type Descriptor | PAIS |
| **UML**  | Unified Modeling Language | OAIS, PAIMAS |
| **UNICODE** | Universal Code | OAIS |
| **URL** | Universal Resource Locator | PAIS |
| **VHS** | Video Home System | OAIS |
| **WWW** | World Wide Web | OAIS |
| **XFDU** | XML Formatted Data Unit | OAIS, PAIS |
|  |  |  |

### terminology

We expect to need definitions for:

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.

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)

**Metadata:** text TBD

**Data Record:** text TBD

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

**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/DIS 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>

[6] *Long Term Preservation of Earth Observation Space Data, Preservation Workflow*. Issue TBD. TBD Location.: CEOS/WGISS/DSIG/PW, December 2014. (under review) Available from: <http://tbd.ceos.org/>

[7] *CCSDS Glossary*. Washington, D.C.: CCSDS, Last updated TBD 2015. Available from: <http://www.sanaregistry.org/>

[8] *Reference Architecture for Space Data Systems*. Recommendation for Space Data System Practices, CCSDS 311.0-M-1. Magenta Book. Issue 1. Washington, D.C.: CCSDS, September 2008. [Equivalent to ISO 13537:2010.] Available from: <http://public.ccsds.org/publications/archive/311x0m1.pdf>

[9] *Reference Architecture for Space Data Systems*. Report for Space Data System Practices, CCSDS 312.0-G-1. Green Book. Issue 1. Washington, D.C.: CCSDS, November 2012. Available from: <http://public.ccsds.org/publications/archive/312x0g1.pdf> [10] *LTDP EO Data Stewardship Definitions*. Issue TBD. TBD Location.: LTDP, TBD 2014. Available from: <http://tbd.ltdp.org/>

 [11] *LTDP Common Guidelines*. Issue TBD. TBD Location.: LTDP, TBD 2014. Available from: <http://tbd.ltdp.org/>

[12] *Data Lifecycle Models and Concepts*. Issue TBD. TBD Location.: CEOS/WGISS, September 2011. Available from: <http://ceos.org/ourwork/workinggroups/wgiss/documents/> or http://www.pnamp.org/sites/default/files/data\_life\_cycle\_models\_and\_concepts.pdf

[13] *Life Cycle Models for Digital Stewardship*. Issue TBD. TBD Location.: US Library of Congress, February 2012. Available from: <http://blogs.loc.gov/digitalpresrvation/2012/02/life-cycle-models-for-digital-stewardship/>

[14] *Many Models*. Issue TBD. TBD Location.: TBD Source, TBD 20xx. Available from: <https://dl.dropboxusercontent.com/u/6959356/ICP/Many%20models.pptx>

[Editor note: References below are additional references suggested by MM. Again, only references to actual requirements can be included in this portion of the references. I don’t think any documents 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. If added to a reference list, complete, properly formatted citations for the items below need to be provided.]

[15] Aligning\_National\_Approaches\_to\_Digital\_Preservation, http://www.cni.org/wpcontent/uploads/2014/07/Aligning\_National\_Approaches\_to\_Digital\_Preservation.pdf

[16] APARSEN-Overview of Preservation Services, http://www.alliancepermanentaccess.org/wpcontent/uploads/downloads/2014/06/APARSEN-REP-D21\_1-01-2\_1\_incURN.pdf

[17] Towards the Preservation of Scientific Memory, http://www.ijdc.net/index.php/ijdc/article/view/10.1.196/392

[18] Digital Library Technology and Methodology Cookbook, https://www.coar-repositories.org/files/D3-4-Digital-Library-Technology-and-Methodology-Cookbook1.pdf

[19] Digital Preservation and Permanent Access to Scientific Info - State of Practice, http://www.cendi.gov/publications/04-3dig\_preserv.pdf

[20] DLIB-D3.2b Digital Library Reference Model, https://www.coar-repositories.org/files/D3-2b-Digital-Library-Reference-Model.pdf

[21] Global Research Data Infrastructures: The GRDI2020 Vision, http://www.grdi2020.eu/Repository/FileScaricati/fc14b1f7-b8a3-41f8-9e1e-fd803d28ba76.pdf

[22] Harnessing the Power of Digital Data, https://www.nitrd.gov/about/harnessing\_power\_web.pdf

[23] Long-term Preservation of Product Lifecycle Metadata in OAIS Archives, Jorg Brunsmann, http://deposit.fernuni-hagen.de/2798/1/Diss\_Brunsmann.pdf

[24] NASA Earth Science Data Preservation Content Specification, https://earthdata.nasa.gov/standards/preservation-content-spec

[25] Preservation Metadata, http://www.dpconline.org/component/docman/doc\_download/894-dpctw13-03

[26] CASPAR Overall Component Architecture and Component Model, http://www.alliancepermanentaccess.org/filestore/CASPAR-deliverables/CASPAR-D1301-TN-0101-1\_1.pdf

[27] Towards the Preservation Infrastructure, Shirley Crompton, et al, http://users.ics.forth.gr/~marketak/download/SCIDIP-ES\_iPres2

[28] USGS Science Data Lifecycle Model, John Faundeen, et al, http://pubs.usgs.gov/of/2013/1265/pdf/of2013-1265.pdf

[29] PDS4 Data and System Architecture, TBD.

[30] PREMIS Data Dictionary, http://www.loc.gov/standards/premis/v2/premis-2-0.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.

Figure 2-1 below illustrates the relationship of a number of information lifecycle activity terms as used within this standard.



* Figure 2‑1: Relationship of Terminology

# THE LIFECYCLE FRAMEWORK: THE MAIN STAGES, ACTIVITIES, AND DELIVERABLES

[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 “project” to mean quite generally an activity “planned and designed to achieve a particular aim”. This aim may range from the creation of new data to the preservation of existing data with a particular preservation aim.



* Figure 3‑1: Information Lifecycle Framework

Figure 3-1 above identifies the stages of the information lifecycle.

1) The Formulation stage is responsible for proposing the project, which acquires or generates, manages or supports exploitation of data or information and then gaining approval and resources to carry out that activity.

2) The Implementation stage is responsible for preparing 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 Operation stage is responsible for carrying out the activity to acquire or generate, manage, and support exploitation of data or information from this activity

4) The Exploitation stage is responsible for keeping the data or information useable and adding value 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 Recommendation 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.

## Information Lifecycle stages

[Editor note: Original text below provided by MM. I think we agreed to keep a more extensive description to supplement what is above. However, I think we agreed to drop the Key Topics. We felt that they were too extensive to include in the current standard and come anywhere near completing this standard this year. My updates/comments are tracked.]

**Formulation Stage**

The participants in the formulation stage are Sponsors, Proposers and archives. Sponsors typically have a well-defined charter that identifies their domain of interest. Within the domain, programs are identified for specific research topics. Sponsors 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. Sponsors issue calls for proposals that will fit into the programs they have defined.

Proposers respond to these program calls by developing proposals. The proposals describe research projects that corresponds to the program objectives. The proposal will likely identify the cost of delivering useful archive products 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 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).

**NOTE: The following provides an list of the key topics that would be discussed for this stage. Each topic would likely take several pages of discussion. Why is it important, what forms does it take, applicability to some or all projects,**

Key Topics:

• Data Management Plan - This will include a discussion the various templates for Data management Plans that are used by various agencies and try to come up with a consolidated list of contents.

• Costing – This will discuss how to come up with cost estimates including the importance of shielding archive funds from being raided for other purposes. Products produced at the end of a project can suffer as project teams are let go.

• Risk Management - This will discuss risk management procedures.

• Information Models - This will discuss various options for creating and using an information model to define the structure and relationships of project data and metadata.

• Data Dictionary – This will discuss the importance of developing a data dictionary to describe mportant metadata elements for current and future use.

• Data Definition – This will discuss the various standards that are available for data and metadata formats.

Key Points:

Costs and other considerations for delivering useful archive products need to be identified in proposal phase.

The target archive system should be identified and involved in the proposal process. The archive can provide support for costing, sample data management plans,

Components: Strategic Plans, Program Goals and Objectives, Calls for Proposals, Proposals, Data Management Plans, Information Model, Data Dictionary, Product Specification, Archive Support Plan.

Standards: DPA Active Data Management Plan (In process). Others TBD

**NOTE: The following is just a set of rough placeholder descriptions of the other stages. Each would have a detailed discussion of all the activities carried out in the stage. Each would have a set of Special Topics with discussions of how each relates to the long term usability and preservation of the data.**

**Implementation Stage**

The implementation 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).

Key Topics

• Cyberinfrastructures – Off the shelf hardware and software components to build data processing systems.

• Common Core – Recommendations for common components for system design, development and operations.

• Data Management – This topic will discuss the data management environment of a project and how it may drive the collection and distribution of project metadata and provide access to both the data and metadata collections.

• Service Architecture – Implications of implementing project data processing systems as services rather than monolithic applications.

• Calibration – What needs to be done to extract meaningful values from the collected data.

• System Testing – How is the operational system tested to assure it will be able to achieve its goals.

**Operation Stage**

The operation stage is performed by the project team possibly with support from the archive. It operates the acquisition systems to gather data. The projet 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. ItThe project team prepares Submission Information Packages (SIPs) for delivery to the long-term archive and participates in peer reviews.

Key Topics

• Preservation Description Information – What needs to be captured to provide provenance, fixity, etc.

• Ancilliary and event data – What additional descriptive information will be needed to support interpretation of the data.

• Data Record Improvement - What techniques can be used to • Validation – How the project team verifies the validity of the data collection process and the data processing.

• Analysis Techniques – What needs to be captured to understand how the data were processed and displayed by the project team.

• Publication – How the publication process should support replication of results and continuing analysis of the data.

• Information Packaging – How should the data and metadata be structured for optimal transfer to the archive.

**Exploitation Stage**

The exploitation 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.

Key Topics

• Selection - This topic will discuss strategies for identifying, evaluating and selecting candidate projects for inclusion in the archive.

• Organization – How to integrate the Submission Packages into the archive.

• Presentation – Develop mechanism to allow the designated community to access the collection.

• Value adding - This will discuss the range of options for making the archive collection more useful to the designated community.

• Consolidation - This will discuss mechanisms for collecting and integrating related collections to build more comprehensive collections.

• Validation - This will discuss procedures for assuring that the submitted packages will meet the needs of the designated community.

• Delivery mechanisms – This will discuss mechanisms for delivering large archives.

• Hardware and Software Acquisition - Sometimes it may be desirable to transfer entire hardware and software systems from the project to the archive.

• Resident Archives – This topic will discuss situations where collections continue to be maintained outside the archive after the end of project to form a federated archive.

• Preservation – This topic will discuss the array of options (refreshment, migration, etc.) available for maintaining the usability, accessibility and integrity of the data collection.

• Certification – This will discuss the importance of self or external certification to assess and improve the performance of the archive.

## Next SubSection

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

# 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.]

## Formulation Stage

|  |
| --- |
| Proposal preparation and submission |
| TBD description |
| Input | TBD |
| Output | TBD |

|  |
| --- |
| Call for Proposal preparation and release |
| TBD description |
| Input | TBD |
| Output | TBD |

|  |
| --- |
| Proposal review |
| TBD description |
| Input | TBD |
| Output | TBD |

|  |
| --- |
| Data set appraisal |
| An appraisal of the data set will provide an initial conception of whether the data set should be preserved and kept accessible and usable for the long term. Topics to be considered include mission relevance, economic considerations, temporal and geographical coverage, size, storage media and archiving format. The United States Geological Survey (USGS) provides helpful information for assessing the 'preservation value' of a data set (see link below). |
| Input | Example: http://eros.usgs.gov/government/ratool/ |
| Output | Data set appraisal (document), addressing at a minimum the aspects of the following topics, as proposed by the USGS:* Mission alignment with its own mandate, significance
* General characteristics (including coverage, time span, completeness)
* Access & distribution characteristics (including users, legal constraints, IP)
* Physical characteristics (including media, volume, formats, processing level)
* Metadata characteristics (including mission, sensor, calibration, processing information)
* Economic characteristics (including preservation costs estimate, cost-benefit analysis)
 |
|  |
| Definition of designated community and Definition of preservation objective |
| Defining the designated community will help taking decisions during the preservation planning process. Data formats and access infrastructures may be adapted to the skills, resources and knowledge base that a community has access to. The community should be wide enough to allow for different levels of knowledge, applications and evolving user needs. The challenge lies in foreseeing a future user community and future uses of the data set. The designated community therefore should be re-assessed periodically, e.g. every ten years, to account for any changes in e.g. community composition or data use. The user community should be defined with sufficient detail to permit meaningful decisions to be made, regarding the composition of the data set to be preserved, and to allow derivation of requirements for effective re-use of the data. The preservation objective can be derived from a dialog with the user community. It should define the level of use that an archive wishes to maintain for the Designated community. It may address topics such data discovery and access, or the provision of visualization, processing and analysis tools and infrastructure. |
| Input | Data set appraisal (document) |
| Output | Designated community definition (document)Preservation objective specification (document) addressing e.g.:* Intended use
* Temporal scope
* Data discovery and access
* Visualization, processing, and analysis tools and infrastructure.
 |
| Implementation Stage |
| Specification of preservation and curation requirements |
| The preservation objective is translated into preservation requirements. These are more specific and may be based on user scenarios and use cases, possibly including detailed system requirements. Requirements for data value adding, across mission data set alignment, access, re-processing, or exploitation may also be included. |
| Input | *EO Data preservation guidelines[[1]](#footnote-1)*Designated community definition (document) Preservation objective specification (document) |
| Output | Preservation and curation requirements specification (document) |
|  |
| Definition (or Tailoring) of the consolidation process |
| The consolidation process produces, from the input data records (L0 and auxiliary data), the corresponding, consolidated and validated data records, devoid of corrupted and duplicate files, aligned to the same naming convention and file format, and associated quality indicators. This process also impacts the services and functions which make the archival information holdings accessible to users, i.e. data search, discovery, retrieval, and use.The *Generic EO Data Set Consolidation Process* document helps define a tailored procedure for the specific data records at hand. The tailored consolidation process will be applied to the data records during the consolidation phase. |
| Input | Preservation requirements specification (document)*Generic EO Data Set Consolidation Process* |
| Output | Tailored consolidation process (document) addressing the following topics, as specified in the *Generic EO Data Set Consolidation Process*:* Data collection
* Cleaning and pre-processing
* Completeness analysis
* Processing and re-processing
 |
|  |
| Tailoring of preserved data set content and filling of the corresponding inventory table |
| The *preserved data set content* document describes which data records and associated knowledge should be preserved in order to ensure long-term usability of the data set. The composition of the PDSC varies by sensor category and needs to be tailored for the specific data set at hand, taking into consideration the designated community, the preservation objective, requirements and dependencies, if any.The data manager should generate and fill a preserved data set content inventory table to assess which data records, information, and software is available and should be preserved. The table facilitates the assessment of completeness against the tailored preserved data set content document. For the data records, the information should be collected at both collection level and at scene/pass level. The detailed scene/pass-based data inventory will be used in assessing spatial and temporal gaps in the data records. Only items listed in this table will be preserved. The tailored and completed preserved data set content inventory table is therefore a critical document in the preservation process.  |
| Input | Designated community definition (document)Preservation objective specification (document)Preservation requirements specification (document)*Preserved data set content* *document*Preserved data set content inventory table |
| Output | Tailored preserved data set content (document) - draftTailored completed preserved data set content inventory (table) - draft |
|  |
| Consultation and agreement with designated community |
| The data set specific tailored PDSC should be discussed and agreed upon with the designated user community. The PDSC inventory table lists all the data set specific data records, information and software which are to be preserved for the future. Items not listed in this table will not be preserved. Hence, acceptance by the user community should be sought before continuing any further preservation activities. |
| Input | Designated community definition (document)Preservation objective specification (document)Tailored preserved data set content (document) - draftTailored completed preserved data set content inventory (table) - draft |
| Output | Final preserved data set content (document) Revised tailored completed preserved data set content inventory (table) |
|  |
| Cost and risk assessment |
| A cost and risk assessment should accompany the entire preservation process. Periodical re-assessment of both costs and risks helps identify and mitigate upcoming changes and hazards. Risks assessed should include at minimum semantic risks, technical risks, organizational risks, resource risks and IPR related risks. An assessment of probability and severity/impact, together with a mitigation plan, should be prepared for each risk. The initial planning should extend at least 20 years into the future and be updated regularly, e.g. every ten years. Since the temporal scope of the preservation activity extends over several decades, risks may change considerably.  |
| Input | Data set appraisal (document) Preservation objective specification (document)Revised tailored completed preserved data set content inventory (table) |
| Output | Cost assessment (document) addressing at minimum the following issues:* Updated preservation cost estimate from appraisal
* Resource planning (personnel, investments, operating expenses)

Risk assessment (document) addressing at minimum the following issues:* Risks: semantic, technical, organizational, resource, IPR related
* For each risk: probability, impact, severity, mitigation plan
 |

|  |
| --- |
| Implementation of consolidation process |
| The tailored consolidation process, defined earlier, is being implemented. As specified, the set of activities will be applied to the data records to be preserved. |
| Input | Tailored consolidation process (document) |
| Output | Consolidated, clean data records, ready for preservation and user access |
| Operation StagE |
| Gathering of missing PDSC items and update of the PDSC table |
| Compiling the knowledge associated with the data set to be preserved, i.e. information and tools, may continue and should be completed during the consolidation phase. The data set specific PDSC inventory table shall be finalized. |
| Input | Final preserved data set content (document) Revised tailored completed preserved data set content inventory (table) |
| Output | Final tailored and completed preserved data set content inventory (table) |

|  |
| --- |
| Data ingestion, master inventory generation, and catalogue population |
| The data set to be preserved, i.e. the consolidated data records and the associated knowledge, are being ingested into the respective repositories. A master inventory should be generated and the catalogue should be populated in preparation for data dissemination. Ideally, these are done automatically during ingestion. |
| Input | Consolidated data recordsAssociated information (as specified in revised tailored completed preserved data set content inventory)Associated tools (as specified in revised tailored completed preserved data set content inventor) |
| Output | Complete, consolidated data set (data records and associated knowledge) ingested into sustainable repositoriesMaster inventoryPreserved data set catalogue |
|  |
| Dissemination  |
| The data set, i.e. the data records and (specific) associated knowledge, is being made available to users, for discovery and retrieval. Providing tools for visualization, analysis, processing and/or corresponding exploitation infrastructure may be provided, as outlined in the Preservation Requirements Specifications. |
| Input | Consolidated data recordsPreserved data set cataloguePreservation objective specification (document)Preservation requirements specification (document) |
| Output | Online discovery and retrieval (download, ordering) of the data records and (selected) associated knowledge |

|  |
| --- |
| Operations and Maintenance |
| The data sets, catalogue, and management inventories are being attended to. The archive, access, and management infrastructure is being operated, i.e. monitored for errors with corrective action taken in case of problems. In accordance with the *EO Data Preservation Guidelines*, the infrastructure is being updated and migration activities are performed as required. In response to reprocessing requirements, e.g. resulting from a processing algorithm update, the preservation workflow may be re-initialized.As the end of the preservation period, defined in the initialization phase, is approaching, a re-assessment of the preservation planning should be done in order to adjust preservation objectives and priorities. |
| Input | Data and inventoriesArchive, access, and management infrastructure |
| Output | Sustainable data preservation and access |

## exploitation Stage

|  |
| --- |
| Stewardship |
| In order to add value to or to improve accessibility and usability of the preserved data set, curation activities should be conducted. These may include an alignment to generate an across-mission time series, improving data citation and discovery by introducing persistent identifiers, or augmenting the metadata to facilitate content-based image retrieval and data mining. These activities, however, are outside the scope of the preservation workflow. |
| Input | Data and inventoriesArchive, access, and management infrastructure |
| Output | Improved accessibility and usability |

# Long-Term Preservation Requirements

[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.]

# Other Possible Requirements

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

Sponsors should have a well-defined charter that identifies their domain of interest.

Proposal calls should include directives for the publication of results.

Proposal calls should include directives for the disposition of resulting information objects, including requirements to deliver them to an archive.

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

# [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. [ANNEX TITLE]

[EITHER Normative or 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
1. Documents in *italics* are available as CEOS best practice documents on www.ceos.org [↑](#footnote-ref-1)