

Draft Recommendation for
Space Data System Practices

|  |
| --- |
| Information Preparation to Enable Long Term Use |

PROPOSED Draft Recommended Practice

CCSDS 653.0-W-0.24

WHITE Book

 May 2017

AUTHORITY

|  |
| --- |
|  |
|  | Issue: | White Book, Issue 0.18 |  |
|  | Date: | May 2017 |  |
|  | Location: | Rome, Italy |  |
|  |

**(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 | Updates at CCSDS meeting. Remove detail section. |
| CCSDS 653.0-W-0.8 | Information Lifecycle and Long Term Usage, Proposed Draft Recommended Practice, Issue 0.8 | January 2016 | Updates following discussions at telecom |
| CCSDS 653.0-W-09 | Information Lifecycle and Long Term Usage, Proposed Draft Recommended Practice, Issue 0.8 | January 2016 | Updates following telecom |
| CCSDS 653.0-W-10 | Information Lifecycle and Long Term Usage, Proposed Draft Recommended Practice, Issue 0.10 | February 2016 | Updates based on telecons |
| CCSDS 653.0-W-11 | Information Preparation to Enable Long Term Use | March 2016 | Title changed to align with the scope and purpose of the document.Updates based on telecons in March 2016 |
| CCSDS 653.0-W-12 | Information Preparation to Enable Long Term Use | April 2016 | Add details to placeholder sections |
| CCSDS 653.0-W-13 | Information Preparation to Enable Long Term Use | May 2016 | Re-write using PMBOK as basis |
| CCSDS 653.0-W-14 | Information Preparation to Enable Long Term Use | June 2016 | Update based on comments from Mark Conrad and John Garrett and discussions during DAI meetings. |
| CCSDS 653.0-W-15 | Information Preparation to Enable Long Term Use | August 2016 | Updates based on inputs from DMBOK |
| CCSDS 653.0-W-16 | Information Preparation to Enable Long Term Use | Sept 2016 | Updates based on suggestions from DAI call 20160830 to clarify Collection Groups. |
| CCSDS 653.0-W-16JGG | Information Preparation to Enable Long Term Use | Sept 2016 | JGG Comments |
| CCSDS 653.0-W-17 | Information Preparation to Enable Long Term Use | Sept 2016 | Includes updates from the DAI WG meeting in Rome |
| CCSDS 653.0-W-18 | Information Preparation to Enable Long Term Use | December 2016 | Includes updates from the DAI WG discussion |
| CCSDS 653.0-W-19 | Information Preparation to Enable Long Term Use | December 2016 | Updates from the DAI WG discussion including integrating areas from PMBOK and DMBOK into OAIS areas and some tidying the annexes. |
| CCSDS 653.0-W-20 | Information Preparation to Enable Long Term Use | January 2017 | Add annex from M Kearney to replace NASA annex and corrected lots of small errors and inconsistencies which had arisen from multiple piecemeal edits. |
| CCSDS 653.0-W-21 | Information Preparation to Enable Long Term Use | Feb 2017 | Correct diagram in Section “6 MAPPING OF COLLECTION GROUPS TO OTHER PROJECT SCHEMES – change “Operations” in LTDP to grey. |
| CCSDS 653.0-W-22 | Information Preparation to Enable Long Term Use | April 2017 | Updates from comments from M Kearney |
| CCSDS-653.0-W-23 | Information Preparation to Enable Long Term Use | April 2017 | Corrections and updates from Mark Conrad |
| CCSDS-653.0-W-24 | Information Preparation to Enable Long Term Use | May 2017 | Corrections and updates from John Garrett |

CONTENTS

Section Page

[DOCUMENT CONTROL v](#_Toc479232680)

[CONTENTS viii](#_Toc479232681)

[1 Introduction 1](#_Toc479232682)

[1.1 purpose and scope 1](#_Toc479232683)

[1.2 Context 1](#_Toc479232684)

[1.3 Applicability 3](#_Toc479232685)

[1.4 Rationale 3](#_Toc479232686)

[1.5 conformance 4](#_Toc479232687)

[1.6 document structure 4](#_Toc479232688)

[1.7 definitions 4](#_Toc479232689)

[1.7.1 acronyms and abbreviations 4](#_Toc479232690)

[1.7.2 terminology 5](#_Toc479232691)

[1.8 NOMENCLATURE 6](#_Toc479232692)

[1.8.1 NORMATIVE TEXT 6](#_Toc479232693)

[1.8.2 INFORMATIVE TEXT 7](#_Toc479232694)

[1.9 References 7](#_Toc479232695)

[2 Overview of the Framework 9](#_Toc479232696)

[2.1 Groupings of Activities 9](#_Toc479232697)

[2.1.1 PMBOK Process Groups 9](#_Toc479232698)

[2.1.2 DMBOK Activity Groups 10](#_Toc479232699)

[2.1.3 Collection Groups 10](#_Toc479232700)

[2.2 Areas 10](#_Toc479232701)

[2.2.1 PMBOK Knowledge Areas 11](#_Toc479232702)

[2.2.2 DMBOK Functions 11](#_Toc479232703)

[2.2.3 Additional Information Areas 11](#_Toc479232704)

[3 Collection Groups 12](#_Toc479232705)

[3.1 Overview of Collection Groups 12](#_Toc479232706)

[3.2 Details of the Collection groups 12](#_Toc479232707)

[3.2.1 The Initiating Collection Group 12](#_Toc479232708)

[3.2.2 The Planning Collection Group 13](#_Toc479232709)

[3.2.3 the Executing Collection Group 13](#_Toc479232710)

[3.2.4 the Closing Collection Group 13](#_Toc479232711)

[3.2.5 the Control Collection Group 14](#_Toc479232712)

[4 Additional Information Areas – to Ensure Long-Term Usability 15](#_Toc479232713)

[4.1 Information Areas Derived from OAIS DEFINED Information Objects 16](#_Toc479232714)

[4.1.1 Content Information 16](#_Toc479232715)

[4.1.2 Preservation Description Information (PDI) 17](#_Toc479232716)

[4.1.3 Package Description 19](#_Toc479232717)

[4.1.4 Packaging Information 19](#_Toc479232718)

[4.2 Information Areas Derived from Issues Outside the Information Model 19](#_Toc479232719)

[4.2.1 Publications 19](#_Toc479232720)

[4.2.2 Related datasets 19](#_Toc479232721)

[4.2.3 Potential other uses of the data 20](#_Toc479232722)

[4.2.4 Suggestions about the appropriate Knowledge Base for the Designated Community 20](#_Toc479232723)

[4.2.5 Suggested Transformational Information Properties 20](#_Toc479232724)

[5 Framework - Activities detail 21](#_Toc479232725)

[6 Mapping of Collection Groups to Other Project Schemes 26](#_Toc479232726)

[7 ANNEX POTENTIAL USE Case (Informative) 27](#_Toc479232727)

[7.1 Small research project 27](#_Toc479232728)

[7.2 Large research project 27](#_Toc479232729)

[7.3 Aircraft Manufacture 27](#_Toc479232730)

[8 Informative References (Informative) 29](#_Toc479232731)

[9 Security Considerations (Informative) 30](#_Toc479232732)

[9.1 Introduction 30](#_Toc479232733)

[9.2 security concerns with respect to the CCSDS document 30](#_Toc479232734)

[9.3 Data privacy 30](#_Toc479232735)

[9.4 Data integrity 30](#_Toc479232736)

[9.5 Authentication of communicating entities 30](#_Toc479232737)

[9.6 Control of access to resources 30](#_Toc479232738)

[9.7 Availability of resources 30](#_Toc479232739)

[9.8 Auditing of resource usage 30](#_Toc479232740)

[9.9 Potential threats and attack scenarios 30](#_Toc479232741)

[9.10 Consequences of not applying security to the technology 30](#_Toc479232742)

# Introduction

## purpose and scope

There is a well-recognized need to capture digital information associated with a great variety of endeavours in virtually all areas of society. However it is widely recognized that many such endeavours are not able, for one reason or another, to leave a sufficient legacy of information so others can reuse and fully leverage the effort that has gone into the endeavour. The purpose of this Recommended Practice is provide guidance for projects about the metadata (the term Additional Information is used below) that needs to be captured and/or generated, and retained in order to ensure that the information created by the project, either as part of its main objectives or as a by-product of achieving those objectives, can be exploited over the long term.

This Recommended Practice deals with the aspects of a project, in particular the terminology used. Many of these terms are already used with various definitions within the target communities for this standard – e.g. space, science, records management and archival communities. It is expected that other communities can easily map this terminology to the terminology used within those communities. The *Reference Model for an Open Archival Information System (OAIS)* provided a starting point and inputs from a variety of other sources were used to arrive at the terms used within this standard.

This Recommended Practice accomplishes the following:

* identifies the Additional Information to be collected or improved at various points;
* forms a basis for the specification of Data Management Plans
* forms a basis for the identification and/or development of additional standards and implementation guides including those that address particular concerns in more detail;

## Context

This Recommended Practice extends the approach taken by the widely used *Project Managers Book of Knowledge (PMBOK) Guide* [1] and the related *Data Management Body of Knowledge (DMBOK)*[2][3].

The PMBOK defines a project as an endeavour which is temporary, i.e. having a beginning and an end, undertaken to create a unique product, service or result and focusses on the information and techniques required to manage the project so that it achieves its objectives. The DMBOK focuses on all aspects of data management within such an activity, while noting that “*Data, and information created from data, are now widely recognised as enterprise assets*”, and furthermore “*Data has value only when it is actually used, or can be useful in the future*”.

This document is focussed on the information (called Additional Information in this document) that needs to be captured and/or generated and retained in order to ensure that the information created by the project, either as part of its main objectives or as a by-product of achieving those objectives, can be exploited over the short, medium and long term. It is expected that, by ensuring this Additional Information is collected as fully as possible, projects can significantly improve their information legacy to the benefit of the wider community.

At various times in the project, and for various reasons, data is captured or created. There is Additional Information associated with this data that also needs to be captured. The types of Additional Information are informed by the Reference Model for an Open Archival Information System [4], referred to as OAIS below, that provides a conceptual view of long term information preservation in an archive.

This Recommended Practice fits into the overall context defined by a number of other standards. Some relationships between the documents are illustrated in Figure 1-1.



Figure ‑ Relationship between CCSDS standards

OAIS 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.

The *Producer-Archive Interface Methodology Abstract Standard (PAIMAS)* [5] 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 *Producer-Archive Interface Specification (PAIS)* [6] 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 *Audit and Certification of Trustworthy Digital Repositories* Recommended Practice [7] provides metrics for use in assessing the trustworthiness of digital repositories or archives.

In addition there are other CCSDS/ISO standards that may be used to create Representation Information (the *Parameter Value Language (PVL)* [8], *the Data Description Language EAST Specification* [9] and the *Data Entity Dictionary Specification Language (DEDSL)* [10]) and also to package information the *XML Formatted Data Unit (XFDU)* [11]). There are many other techniques for creating Additional Information but these are outside the scope of this document.

In addition, the archival community has an existing, well established, set of concepts and terminology. The relationship with these and the OAIS concepts that underpin this document is described in **XXXXX**.

## Applicability

While this recommendation 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 records management and archival communities. It is applicable to information created by an individual, in an individual project or by an organisation as a whole. It is applicable to projects where the data already exists as well as projects where data is to be created in the future. It is also applicable to projects where the associated data is not the main focus of the project.

This document should be of use to funders and information creators, a role which may be played by multiple actors such as researchers or manufacturers, archive managers and end-users. It will be of use by helping to increase the effectiveness of preservation activities and the exploitation of information and by informing the practices and standards these users define in their communities.

This guidance can form the basis on which plans, including Data Management Plans, can be constructed, updated and monitored, to achieve the objectives noted above.

## Rationale

Data that is collected or created must have Additional Information associated with it if it is to be independently understandable, usable and trusted as being authentic. The amount and content of the 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 (PDI) for this associated information. It must be accumulated over the life of the project. For example Provenance Information (part of PDI) should originate at data creation and will accumulate over time, recording the things that have happened to the data.

In the case of information created by individual projects, 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. This standard encourages the active management of these plans to continue to address the communities’ needs and uses for the data. These Data Management Plans can also be captured as part of the Additional Information.

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

There are a small number of generally applicable processes in a project where, typically, the responsibility is handed on from one individual or team to another. Each of those individuals or teams has specific knowledge about the information which subsequent individuals or teams may not possess. There is a need to specify the information to be captured within and at the interfaces between each of those Project Phases. Improvements or changes to the Additional Information must be considered as the work proceeds. Therefore there is a need for guidance as to what Additional Information should be captured or improved through the various Project Phases.

This document should enable:

* the Producer 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 (e.g. re-use or secondary use) of the information;
* the user to re-use 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.

## conformance

Conformance to this recommended practice requires that Additional Information is collected as described in Section 5.

## document structure

Section 2 gives an overview of the document concepts which are expanded in the following sections. The Process Groups in the project are described in more detail in Section 3. Section 4 defines the areas about which information should be collected and identifies the major pieces of information related to eventual re-use and exploitation which need to be collected. Section 5 shows the way in which that information may evolve through the project.

## definitions

### acronyms and abbreviations

|  |  |
| --- | --- |
| **AIP** | Archival Information Package |
|  |  |
| **CCSDS** | Consultative Committee for Space Data Systems |
| **CRC** | Cyclic(al) Redundancy Check |
| **CRIS** | Current Research Information System |
| **DAMA** | Data Management Association International |
|  |  |
| **DEDSL** | Data Entity Dictionary Specification Language |
| **DMBOK** | Data Management Body of Knowledge |
| **DMP** | Data Management Plan |
|  |  |
| **EO** | Earth Observation |
| **ESDIS** | Earth Science Data and Information System |
| **FITS** | Flexible Image Transport System |
|  |  |
|  |  |
| **JPEG** |  |
| **LTDP** | Long-Term Data Preservation |
| **OAIS** | Open Archival Information System |
| **PAIMAS** | Producer-Archive Ingest Methodology Abstract Standard |
| **PAIS** | Producer-Archive Ingest Specification |
|  |  |
| **PDI** | Preservation Description Information |
|  |  |
| **PMBOK** | Project Management Book of Knowledge |
| **PVL** | Parameter Value Language |
| **RIN** | Representation Information Network |
|  |  |
| **SIP** | Submission Information Package |
|  |  |
|  |  |
| **XFDU** | XML Formatted Data Unit |
| **XML** | eXtensible Markup Language |

### terminology

There are many terms which are used in this document which need to have well defined meanings. These terms are defined in this subsection. When first used in the text, they are shown in bold and are capitalized. Subsequent use employs capitalization only. They should eventually be available online at http://www.sanaregistry.org/r/terms/terms.html.

Apart from the extra terms below, the definitions provided by the Reference Model for an Open Archival Information System (OAIS) [Ref. 4] and the other standards described in section 1.2 are used; these terms are normally capitalised, following the OAIS convention. It is assumed that the reader has some familiarity with OAIS.

Note: For convenience a number of selected definitions from OAIS are included here::

**Content Information:** A set of information that is the original target of preservation or that includes part or all of that information. It is an Information Object composed of its Content Data Object and its Representation Information.

**Data:** A reinterpretable representation of information in a formalized manner suitable for communication, interpretation, or processing. Examples of data include a sequence of bits, a table of numbers, the characters on a page, the recording of sounds made by a person speaking, or a moon rock specimen.

**Data Object**:Either a Physical Object or a Digital Object.

**Designated Community**: An identified group of potential Consumers who should be able to understand a particular set of information. The Designated Community may be composed of multiple user communities. A Designated Community is defined by the Archive and this definition may change over time.

**Digital Object**:An objectcomposed of a set of bit sequences.

**Information**:Any type of knowledge that can be exchanged. In an exchange, it is represented by data. An example is a string of bits (the data) accompanied by a description of how to interpret the string of bits as numbers representing temperature observations measured in degrees Celsius (the Representation Information).

**Information Object**:A Data Object together with its Representation Information.

**Preservation Descriptive Information** (PDI):

**Representation Information**:The information that maps a Data Object into more meaningful concepts. An example of Representation Information for a bit sequence which is a FITS file might consist of the FITS standard which defines the format plus a dictionary which defines the meaning in the file of keywords which are not part of the standard. Another example is JPEG software which is used to render a JPEG file; rendering the JPEG file as bits is not very meaningful to humans but the software, which embodies an understanding of the JPEG standard, maps the bits into pixels which can then be rendered as an image for human viewing.

**Activity:** A distinct, scheduledportion of work performedduring the course of a project . (from PMBOK)

**Additional Information**: The information which should accompany Data to ensure that it can be preserved and exploited. This will include Representation Information and Preservation Description Information (PDI), as defined by OAIS.

**Additional Information Area**: A complete set of concepts, terms and activities that make up the Additional Information that is needed to support long-term exploitation of data.

**Data Management Plan**: A document that describes how Data will be handled throughout the project and what will happen to it when the project ends.

**Deliverable**: Any unique and verifiable product, result, or capability to perform a service that is required to be produced to complete a process, phase or project. (from PMBOK)

**Project**: A temporary endeavour undertaken to create a unique product, service or result. (from PMBOK)

**Project Phase**: A collection of logically related project activities that culminates in the completion of one or more outputs. (from PMBOK)

## NOMENCLATURE

### NORMATIVE TEXT

The following conventions apply for the normative specifications in this document:

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 1 to 5), 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.

1. A Guide to the Project Management Body of Knowledge (PMBOK® Guide)—Fifth Edition, 2013, see <http://www.pmi.org/pmbok-guide-and-standards/pmbok-guide.aspx>
2. The DAMA Guide to the Data Management Body of Knowledge (DMBOK Guide) First Edition, 2009, <https://www.dama.org/content/body-knowledge>
3. DMBOK Version 2 see [http://dama-dach.org/dmbok2-DMBOK-version-2/](http://dama-dach.org/dmbok2-dama-dmbok-version-2/)
4. *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: <https://public.ccsds.org/Pubs/650x0m2.pdf>
5. *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: [https://public.ccsds.org/Pubs /651x0m1.pdf](https://public.ccsds.org/Pubs%20/651x0m1.pdf)
6. *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: <https://public.ccsds.org/Pubs/651x1b1.pdf>
7. *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: <https://public.ccsds.org/Pubs/652x0m1.pdf>
8. *Parameter Value Language Specification (CCSD0006 and CCSD0008)*. Blue Book. Issue 2. Washington, D.C.: CCSDS, June 2000. [Equivalent to ISO 21962:2003.]

Available from: <https://public.ccsds.org/Pubs/641x0b2.pdf>

1. *The Data Description Language EAST Specification (CCSD0010)*. Blue Book. Issue 3. Washington, D.C.: CCSDS, June 2010. [Equivalent to ISO 15889:2011.]

Available from <https://public.ccsds.org/Pubs/644x0b3.pdf>

1. *Data Entity Dictionary Specification Language (DEDSL)—XML/DTD Syntax (CCSD0013)*. Blue Book. Issue 1. Washington, D.C.: CCSDS, January 2002. [Equivalent to ISO 15889:2011.]

Available from <https://public.ccsds.org/Pubs/647x3b1.pdf>

1. *XML Formatted Data Unit (XFDU) Structure and Construction Rules*. Blue Book. Issue 1. Washington, D.C.: CCSDS, September 2008. [Equivalent to ISO 13527:2010.]

Available from <https://public.ccsds.org/Pubs/661x0b1.pdf>

# Overview of the Framework

Following the PMBOK [1] terminology, a **Project** may be divided into any number of **Project Phases** (see for example section 6). A Project Phase is a collection of logically related project **Activities** that culminates in the completion of one or more **Deliverables**. The Project Phases may be sequential or overlapping.

There are many system lifecycle descriptions [I-2] which describe the stages or phases of systems. The PMBOK [1] and the DMBOK [2][3] instead use a more flexible approach of describing groups of activities which appear repeatedly in these various stages or phases.

The basic approach of PMBOK, DMBOK and this document is to identify a framework in the form of a matrix of groupings of activities/processes for a number of areas (in this document we use the term Additional Information Areas, in PMBOK uses Knowledge Areas, and DMBOK uses Functions).In this document the areas are about which information is needed to support long-term exploitation of data

The aim is to fill in the cells in the table below to show that for a given area there should be some specific Activity. This will act as a checklist to help to ensure that data is useable over the long term.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Activities→Areas↓ |  |  |  |  |  |
|  |  |  |  |  |  |
|  |  |  |  |  |  |

Section 2.1 describes the groupings of activities while section 2.2 describes the areas.

## Groupings of Activities

PMBOK and DMBOK use slightly different groupings and terminology, the former uses the term Process Groups while the latter uses the term Activity Groups. In this document we use the term Collection Groups, which best fits the requirements of this document. The next sub-sections describe these groupings.

### PMBOK Process Groups

PMBOK describes five Process Groups:

* ***Initiating*** *- consisting of processes performed to define a new project or new phase of an existing project by obtaining authorization to start the project or phase*
* ***Planning*** *– consisting of those processes performed to establish the total scope of the effort, define and refine the objectives, and develop the course of action required to attain those objectives*
* ***Executing*** *– consists of those processes performed to complete the work defined in the project management plan to satisfy the specifications.*
* ***Closing*** *– consists of those processes performed to conclude all activities across all Project Management Process Groups to formally complete the project, phase or contractual obligations.*
* ***Control*** *– consists of those processes required to track, review, and orchestrate the progress and performance of the project.*

*The Project Management Process Groups are linked by the outputs they produce. The Process Groups are seldom either discrete or one-time events; they are overlapping activities that occur throughout the project. The output of one process generally becomes an input to another process or is a deliverable of the project, subproject or project phase.*

### DMBOK Activity Groups

DMBOK uses four Activity Groups which have obvious overlaps with PMBOK Process Groups:

* **Planning** – activities that set the strategic and tactical course for other data management activities. Planning Activities may be performed on a recurring basis.
* **Development** – activities undertaken within implementation projects and recognized as part of the systems development lifecycle (SDLC), creating data deliverables through analysis, design, building, testing, preparation, and deployment.
* **Control** – supervisory activities performed on an on-going basis.
* **Operational** – service and support activities performed on an on-going basis.

Each data management activity fits into one or more data management activity groups.

### Collection Groups

This document uses the term Collection Groups because the concern is to collect Additional Information about the data being created. The Collection Groups identify types of Activities where Additional Information may be collected and is a specialisation of the PMBOK terminology, in particular including “Closing” because when the process to create data closes, steps must be taken to ensure its usability after the end of that process. The Collection Groups are:

* **Initiating** – justification for creating the data and initial definition of the data project
* **Planning** – planning for the data creation and encoding
* **Executing** – creating/collecting/encoding the data. At each point there may be deviations from the planned results, including instrument effects and unexpected influences.
* **Closing** – completing the data creation/collection/encoding to satisfy the requirements of the project, phase or contractual obligations, and turning the information over to the long-term preservation organization.
* **Control** - track, review, and orchestrate the progress and performance of the activities.

These will be discussed in detail in section 3.

## Areas

The areas identify general headings which are or should be used in most projects most of the time. PMBOK and DMBOK provide lists of these areas in general terms. PMBOK uses the term **Knowledge Areas** while DMBOK currently uses the term **Functions**. This document uses the term **Additional Information Areas**.

### PMBOK Knowledge Areas

The PMBOK identifies ten Knowledge Areas:

* Integration Management
* Scope Management
* Time Management
* Cost Management
* Quality Management
* Human Resource Management
* Communications Management
* Risk Management
* Procurement Management
* Stakeholder Management

These each represent a “*complete set of concepts, terms, and activities that make up a professional field, project management field, or area of specialisation. These ten areas Knowledge Areas are used in most projects most of the time.”*

### DMBOK Functions

DMBOK identifies ten Functions:

* Data Governance
* Data Architecture Management
* Data Development (to be renamed in the next version [3] of DMBOK to Modelling and Design)
* Data Operations Management (to be renamed Data Storage and Operations)
* Data Security Management
* Reference and Master Data Management
* Data Warehousing and Business Intelligence Management
* Document and Content Management
* Meta-data Management
* Data Quality Management

The next version of DMBOK will add Data Integration and Interoperability and renames Functions to “Knowledge Areas”.

### Additional Information Areas

The Additional Information Areas of course have a strong relationship to the PMBOK and DMBOK Knowledge Areas/Functions. PMBOK and DMBOK deal with all aspects of a project but this document is concerned only with those aspects which ensure long term usability of the data.

The DMBOK Metadata Management function clearly is of particular significance but for our purposes it is important to provide greater granularity by using the concepts provided by the OAIS Information Model. OAIS defines the information objects required for Long-Term Preservation. All or part of the Information required by these Information Objects must therefore be created/collected through the project. These are

 largely the areas of information which are required to create an Archival Information Package, supplemented by a number of other areas which are not covered by OAIS.

The Additional Information Areas are

* Contest Information
	+ Content Data Object
	+ Representation Information
* Preservation Description Information (PDI)
	+ Reference Information
	+ Provenance Information
	+ Context Information
	+ Fixity Information
	+ Access Rights Information
* Package Description
* Packaging Information
* Issues Outside OAIS Information Model
	+ Publications
	+ Related Data Set
	+ Potential Other Uses
	+ Designated Community/Knowledge Base
	+ Transformational Information Properties

These are discussed in more detail in section 4.

# Collection Groups

Section 3.1 outlines the Collection Groups. Section 3.2 provides a brief description of each of the Collection Groups.

## Overview of Collection Groups

The Collection Groups, following in the style of PMBOK’s Knowledge Areas, are illustrated as follows. The groups are linked by the outputs they produce; they are overlapping activities that occur throughout a project or phases of a project. They are not expected to be either one time or discrete events.

Figure 3‑1 illustrates these Collection Groups in a project with three phases.



Figure ‑ Example Collection Area application within a project with three phases

## Details of the Collection groups

In the following sections the focus is on the Information created by the project – including the project management information – which may need to be retained as Additional Information so that the data (created or collected by the project) will remain understandable and usable.

### The Initiating Collection Group

The Initiating Collection Group consists of processes performed to justivy the data collection and to define a new project, or new phase of an existing project, by obtaining authorization to start the project or phase.

This could include proposing the project/phase, perhaps responding to solicitations and funding information available. It would be reasonable to expect the following types of information to be created:

* the aims of the project to be clear enough to justify the data collection and its resources;
* the way in which data would be collected and the kind of data to be collected would be known in general terms;
* the initial exploitation of the data would be outlined

These are likely to be important pieces of Additional Information that should be preserved as documentation of the project. The participants in this group of processes will almost certainly include sponsors and proposers and may also include data managers and archivists. Examples of documents to begin managing during project initiation include the list of project participants and organizations represented, the criteria for data collection, privacy and data protection, the criteria for repositories where the project data and documentation will be preserved, agreements among participants regarding authorship ownership of intellectual property produced by the project as well as relevant policies of participating organizations regarding such rights.

### The Planning Collection Group

The Planning Collection Group consists of those processes performed to establish the total scope of the effort, define and refine the objectives, and develop the course of action required to attain those objectives.

In the Planning Collection Group the preparations are made to collect or create data. This could include:

* the design and assembly of the components of the information system;
* the development or update of hardware and/or software systems;
* the development of the associated procedures for data collection, privacy and protection;
* the establishment of a data dictionary

These are likely to be important pieces of Additional Information that should be preserved as documentation of the project. Examples of documents to be managed during project planning include the project mission statement, the project management plan, the communication plan, the risk management plan, assignments for roles and responsibilities of team members, the list of project deliverables, the list of candidate repositories and how they meet the established criteria for managing data and documents produced by the project.

### the Executing Collection Group

The Executing Collection Group consists of those processes performed to complete the work defined in the project/phase plan to satisfy the specifications.

Activities are carried out which:

* create or collect the data;
* process and analyse data.

These processes will produce data that needs to be preserved for the long term either as a product or by-product. Examples of documents to be managed during project execution include signed contracts and approvals received from stakeholders or other authorities, data access policies and processes, and project logs, .

### the Closing Collection Group

The Closing Collection Group consists of those processes performed to conclude all activities across all Collection Groups to formally complete the project phase, or the entire project.

The data which may be part of the legacy of the project and which can be exploited in various ways include:

* publication of research findings;
* generation of income
* exchange of social information
* predictions
* scientific and social advancements

There may also be ideas for exploitation in future.

The Closing Collection Group is performed by the project/phase team to use/re-use and exploit the information and, if appropriate, prepare it for handing over for long-term preservation, re-use and exploitation. Examples of documents to be managed during project closing include invoices, signed acceptances, procurement documents, evidence of payments received, associated data, and related publication.

### the Control Collection Group

The Control Collection Group consists of those processes performed to ensure the project is on track or to identify areas which need attention. This process group provides information needed to manage the other process groups. The information collected during the controlling processes is part of the legacy of the project and therefore may need long term preservation.

This could include:

* Programmatic changes
* Configuration management materials
* Changes in development or execution schedules
* Program or design review materials
* Changes in scope
* Test results

Examples of documents to be managed during project monitoring and controlling include Configuration Change Requests and other documents describing proposed changes, and documented decisions of the Change Management Board or other decision bodies, test procedures and logs.

# Additional Information Areas – to Ensure Long-Term Usability

An archive must create AIPs as part of the preservation process. Many of the components that form the AIPs may only be known by the participants in the project. This Recommended Practice provides guidance for the project participants to help ensure that the information is captured, as part of the required Additional Information.

The Additional Information Areas covered in this document are organized around the OAIS information model concepts, in particular the Archival Information Package (AIP) Information Model Components. An AIP should contain all the information required for long term usability and therefore this information must be collected in a timely way throughout the project.

The OAIS standard contains the following diagram to show the various components of an AIP.



Figure ‑ Archival Information Package (Detailed View)

There are other pieces of information that are not covered by the scope of the OAIS Information Model but may be useful for those preparing to archive the information. These include

• the total volume of data – planned, current and actual

• ideas about the Designated Community – previous, current, and future

• ideas about the ways in which the data may be exploited – planned, previous, and current

The Additional Information Areas are discussed in more detail in the following sections.

## Information Areas Derived from OAIS DEFINED Information Objects

OAIS defines several major categories of information that make up the Archival Information Package (AIP): Content Information and Preservation Description Information (PDI). Packaging Information needs to be available to clearly identify and delimit what makes up the AIP. The OAIS standard also defines the Package Description, which is needed to provide visibility and access into the contents of an Archive..

The next sections provide additional information about each of these Information Areas.

### Content Information

Content Information includes the Data Objects as well as the Representation Information needed to understand and use the Data Objects. Representation Information is classified as Structure Information, Semantic Information and Other Representation Information. In broad terms Structure Information describes the physical layout of the Data Objects, Semantic Information describes the meaning of the values in the Data Object and Other Representation Information identifies other dependencies that need to be understood to use the Data Objects including software.

#### Data Objects

Data Objects are the data which will become the primary focus of preservation. Additional Information could include:

* Planned and actual data rates
* Planned and actual volumes of data
* Quality tests which may be performed on the data and test results
* Information Properties which may be of use e.g. accuracy of the data values

#### Representation Information

The Representation Information includes

* structure,
* semantics including the relationship between data elements
* other Representation Information such as analysis and display software.

In some projects the Representation Information may be captured in a number of formal documents. In others, especially those which extend over many years or even decades, there are likely to be a number of pieces of Representation Information which are not formally captured. For example there may be information which “everyone knows” such as:

* modelling and designs
* the way in which software libraries are named or organized
* the meaning of comments e.g. “will run on Cray-like machines” – may actually mean the software must be built on machines which use double-precision floating point numbers by default.
* Compiler bugs which must be worked-around
* The meaning of elements of the data header (if any)
* The location of documentation for proprietary systems
* Quality flags and magic values (care needed when transformed) or special values representing NULL or missing values

For long-term preservation all the pieces of information that “everyone knows” should be captured in as much detail as possible.

Each piece of Representation Information will consist of a Data Object and its Representation Information; each piece of this Representation Information will have its own Data Object and possibly its own Representation Information, and so on. OAIS describes this as a Representation (Information) Network (RIN).

Figure 4-2 provides an example of what this RIN could look like.



Figure ‑ Example of a RIN

The amount of Representation Information which the archive will eventually require will depend upon the Designated Community which the archive serves. It may be useful to work with the archive to draft the RIN as early and in as much detail as possible.

### Preservation Description Information (PDI)

The PDI is information that is necessary to preserve the Content Information. It includes Reference Information, Provenance Information, Context Information, Fixity Information and Access Rights Information.

#### Reference Information

Reference Information provides a unique identification for each product. Useful Additional Information may include:

* Identifiers used in publications
* Naming conventions used in internal systems
	+ How versions/editions are dealt with e.g. numerical or time tagged versions
* Resons for selecting a particular referencing convention

#### Provenance Information

Provenance Information provides information including

* specific aspects of the project origins and history,
	+ from what it was derived i.e. previously collected data
	+ what data is related
* data custody – who was in control of the data at various points in the project,
* version control – what, if any, version control was used for the data,
* calibration and test
* data products from which this information was derived, or example Level 0, Level 1 etc
* processing hardware/software
* processing logs
* how the quality of the information may be checked
* Migration management
* Management of copies of the data
* Synchronisation policy of copies
* Defense against hacking
* Which anti-virus checks performed
* Roles of people e.g. who can change/delete

#### Context Information

Context Information identifies or captures the knowledge that is needed to fully understand and interpret the project results. It includes background, publications and relationships. Provenance Information is a type of Context Information, but there is additional contextual information that is not also Provenance Information.

* Broader aspects of the project origins and history
* The scope of the information collection and any changes in scope which may have occurred during the project
* Funders
* Current Research Information Systems (CRIS) information
* Cultural heritage context

#### Fixity Information

Fixity Information allows verification of the integrity of Data Objects and could include:

* Digests and Checksums – how they were calculated and where they are kept
* Description of how the digests are safeguarded - where they are kept and who can change them..
* Logs of fixity checks and any problems detected

#### Access Rights Information

Access Rights Information including

* ownership,
* copyright and licensing or access restrictions and documents authorizing use
* confidentiality/sensitivity/security constraints
* Embargoes on data publication
* Legal implications if data is released
* Licences used to create, use, distribute information

### Package Description

The Package Description is used to provide a search capability to identify collections or products of interest. It includes finding aids and browse data. The archive must create appropriate Package Description Information. The project could provide information to allow the archive to do this.

### Packaging Information

The Packaging Information is the information that is used to bind and identify the components of an Information Package.

The archive creates the AIPs and it is unlikely that the project will provide information to help in this unless the archive and the project have a close relation and the archive has chosen to maintain an AIP structure that maps directly to the project data structure.

## Information Areas Derived from Issues Outside the Information Model

### Publications

There may be many publications associated with the Data Objects including:

* research publications based on the data
* publications containing the data
* documents about the data – some of these documents may also be Representation Information
* Annotation systems used with the data (if any)
* Community tagging e.g. quality tags held by 3rd parties

### Related datasets

There may be many other data instances which may be related to the Data Objects (with its Additional Information) and which may aid in exploiting the Data Objects, for example

* data in the same discipline, for example astronomical data
* data in a complementary discipline, for example atomic spectral databases and astronomical data
* data about the same object, for example data measured at different wavelengths about a particular star

### Potential other uses of the data

The Data Objects may have been created for a particular purpose, for example a particular research study or as a record of a step in a manufacturing process. The initial exploitation of the data may then be to produce a research paper or to prove the quality of manufacture.

The project may only be interested in, or may only have funding for, exploiting the Data Objects in those ways.

However the project members may recognize that the Data Objects may have potential other uses. For example the Data Objects may have been overhead imagery captured to monitor changes in infrastructure to aid mapping functions and another data project may be able to make use of that same imagery for weather domain cloud cover studies.

### Suggestions about the appropriate Knowledge Base for the Designated Community

The project may have some specific ideas about what Knowledge Base would be needed to understand and use the Data Objects, given the Representation Information which the project provides. For example there may be a general area of scientific expertise or a type of manufacturing process. This information could be useful for any archive which wishes to preserve and facilitate the exploitation of the Data Objects, given the Representation Information provided by the project.

### Suggested Transformational Information Properties

A Transformational Information Property is an Information Property the preservation of the value of which is regarded as being necessary but not sufficient to verify that any Non-Reversible Transformation has adequately preserved information content. This could be important as contributing to evidence about Authenticity. Such an Information Property is dependent upon specific Representation Information, including Semantic Information, to denote how it is encoded and what it means. (The term ‘significant property’, which has various definitions in the literature, is sometimes used in a way that is consistent with its being a Transformational Information Property). Examples include:

* The precision (i.e. number of significant figures) which must be the same when one compares data before and after transformation in a numerical dataset.
* The colour variation allowed between a pre- and post- transformation image.
* Pagination
* Line numbering (for example in legal documents)

Note a decision needs to be made whether any particular Informational Property is a Transformational Information Property, i.e. whether the value for that Information Property need to be maintained for the long-term preservation to be considered successful.

# Framework - Activities detail

The table below indicates the minimum useful status of information capture for each of the areas in each of the Collection Groups. The Control Collection Group is not included in the table because those processes would ensure that the information is captured.

Typically information to address each issue and to document the decisions made in regard to each of these areas will begin to be accumulated early in the project. 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 project, 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 to ensure that the data remains understandable as the Designated Community’s Knowledge Base changes.

The terms used in this section could be mapped to equivalent terms used in the local environment. For examples of this type of mapping see the Annexes.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Additional Information Area** | **Initiating** | **Planning** | **Executing** | **Closing** |
| **Data Object** | * Estimate of volume of data to be produced
* Ideas of the potential value of the data
 | * Update Additional Information from Initiating based on more detailed plans
* Identify types of data (raw, processed, etc.) which should be preserved
* Identify types of data e.g. images, tables – and any generic interfaces
* Quality constraints
* Planned rate of data production
* Expand and add detail
 | * Update Additional Information from Planning based on what really happens
 | * Finalise Additional Information from Executing
* Inventory of data produced which should be preserved
* Volume that would require preservation
* Collect quality checks which may be performed on the data by non-experts
* Define Information Properties which may be useful
* Checks for (and logs of) any missing data
 |
| **Representation Information** | * Standards planned to be used
* Information Model
 | * Update Additional Information from Initiating based on more detailed plans
* Review applicable standards
* Refine Information Model
* Choice of data format
* Identify Hardware and Software Dependencies
* Relationships between data items
 | * Update Additional Information from Planning based on what really happens
* Collect Semantics of the data elements e.g. data dictionaries and other semantics
* Collect Format definitions and formal descriptions
* Create Other Data Documentation
* Calibration and system test tools and system test data that will be delivered
 | * Finalise Additional Information from Executing
* Finalise Representation Information Networks to reasonable level
* Identify other software which may be used on the data
* Create suggestions for the Designated Community and Representation Information needed
 |
| **Reference Information** | * Identify standards which will be used to identify and reference the data and metadata
 | * Update Additional Information from Initiating based on more detailed plans
* Identify which unique identifiers should be used (e.g. DOI or other)
 | * Update Additional Information from Planning based on what really happens
* Rules, methods, tools for referencing data
* Generate references to data as it is being created/captured
 | * Finalise Additional Information from Executing
* Identify what may be used in future to identify the Information
* Checks for (and logs of) missing references and logs of any
 |
| **Provenance Information** | * Record of origins of the project e.g. in a Current Research Information System (CRI)
 | * Update Additional Information from Initiating based on more detailed plans
* Define Processing workflow, Processing inputs and Processing parameters
* Define System Testing required
* Documents from system development milestones
 | * Update Additional Information from Planning based on what really happens
* Documentation about the hardware and software used to create the data, including a history of the changes in these over time
* Update Documentation of Processing workflow, Processing inputs and Processing parameters
* Record who was responsible for each stage of processing
* Record when each stage was performed
* Record of any special hardware needed
* Record Calibration
* Record checking of Fixity
 | * Finalise Additional Information from Executing
* Finalise Provenance handover
 |
| **Context Information** | * Outline of background concepts needed to understand the project
 | * Update Additional Information from Initiating based on more detailed plans
 | * Update Additional Information from Planning based on what really happens
* Collect publications related to the data or the processing system
* Potential Value of the data and likely business case for sustainability
 | * Finalise Additional Information from Executing
* Identify related data which may in the future be combined with this data
 |
| **Fixity Information** |  | * Fixity mechanism (e.g. CRC or digest) of data which may be preserved
 | * Update Additional Information from Planning based on what really happens
* Identify any special validation procedures that should be carried out.
 | * Finalise Additional Information from Executing
* Identify how do we verify that all files are intact
 |
| **Access Rights Information** |  | * What are the restrictions on access in the long term
* Clear identification of Intellectual Property Rights
* Owners of the data – who can authorize hand-over
 | * Update Additional Information from Planning based on what really happens
 | * Finalise Additional Information from Executing
* Licenses involved
* The owner, and the restrictions on access (licenses), and the intellectual property rights
 |
| **Packaging Information** |  |  |  | * Details of the way components are packaged together for delivery to a repository
* Definition of mechanisms for transferring information to next element in the workflow or next in the chain of preservation (e.g. definitions of SIPs)
 |
| **Descriptive Information** |  |  | * Identification of methods for exploration/ quick-look at the data
 | * Finalise Additional Information from Executing
* Create browse/query data if needed
 |
| **Issues Outside the Information Model** | * Estimated Cost of the project
 | * The budget for archiving.
* The schedule for major project milestones and deliveries to the archive.
* Identification of archives which are likely to be able to host the data
 | * Update Additional Information from Planning based on what really happens
 | * Finalise Additional Information from Executing
* Schedule of deliveries
* Pointers to the components to be transferred to the next element in the workflow or next in the chain of preservation
* Potential preservation aims for the information created
* Potential risks to preservation and exploitation of the data
* Define the mechanism for communication between project and archive.
* Define suggested Transformational Information Properties
 |

Table 5-1: Status of Information Capture for Topical Issues in Process Groups

# Annex A: Mapping of Collection Groups to Other Project Schemes (Informative)

This informative annex provides context by providing a graphic which notionally maps the Collection Groups defined by this document to some commonly used project phases, on the basis that Phases tend to have a focus on one or other Collection Group. The Collection groups are compared to:

1. The NASA Mission Life Cycle as defined in NASA Procedural Requirements 7120.5E. While this is technically NASA-specific, the general breakdown of phases (A, B, etc.) are used by most space agencies. Typical project milestones are indicated as well.
2. PMBOK, the Project Management Body Of Knowledge, discussed earlier in this document.
3. LTDP

Items to note:

* The Controlling Group or Processes is normally associated with the operational phase of a project or mission, but closer examination reveals that some controlling processes are performed during all other phases. Hence for the purposes of this comparison, it is shown as an active phase throughout a project’s timeline.
* Controlling applied by all approaches in all phases of the projects. This “Controlling” phase is a Process Group in PMBOK and IPELTU.



# ANNEX B: Example USE Cases (Informative)

## Small research project

An individual researcher applies for funding to perform an experiment. The researcher is successful and sets up the experiment and data collection system and writes a Data Management Plan as required by the funders. He/she performs the experiment, thereby producing data. The researcher analyses the data using software he/she has created and publishes results.

He/she adds appropriate Representation Information and Preservation Description Information and transfers the information to an archive. The archive has defined a Designated Community for this information and ensures that it has adequate Representation Information. The archive may add Representation Information and Descriptive Information in order that the data can be found, accessed and understood by a wider community beyond the Designated Community, in order to further exploit the data.

## Large research project

A scientific instrument is required by a space agency for a satellite which is to study the Sun. Several multi-national consortia submit proposals, which include appropriate Data Management Plans. The proposal from one consortium is selected. The funding is obtained for the various consortium members from the various national funders.

The various consortium members undertake various tasks to build the instrument and the data collection system and associated software. Over a period of 10 years the instrument is built and integrated into the satellite. The satellite is launched and the instrument is deployed and collects data.

The data is collected at a ground station and sent to the researchers who are part of the instrument consortium. Modifications are made from time to time to the on-board software and the data processing software.

The Data Objects and Additional Information are sent to an archive for preservation and re-use by other researchers.

The Additional Information sent to the archive includes Representation Information such as the data format, semantics and processing software, which have been created by various members of the consortium. However many members of the consortium have moved on to other projects or have retired and some relevant information has been lost. This may substantially reduce the value of the data over the long term as it limits who will be able to re-use it, and may even call into question the Authenticity of the data.

The archive adds Descriptive Information and Representation Information to help other communities, including some suggested by the instrument consortium, to exploit the information in other disciplines.

## Aircraft Manufacture

A large aircraft manufacturer wishes to create and sell a new type of aircraft. The initial design team creates a design which is tested and improved by a number of other specialist teams. A number of sub-systems, such as engines and wings, are sub-contracted to other specialist manufacturers. Over a period of 20 years the aircraft design goes through many stages. A great deal of information is collected to provide evidence for such things as safety and air-worthiness certificates.

The aircraft goes into production and is sold world-wide for the next 40 years.

The information that has been collected, including the design and the evidence about certification are legally required to be kept for 50 years beyond the time of manufacture of this model of aircraft. In addition the information can be used by the manufacturer to develop variants of the aircraft and also entirely new types of aircraft.

#  ANNEX C: Informative References (Informative)

1. ESDIS Project. 2013. "NASA Earth Science Data Preservation Content Specification." NASA/GSFC. <https://earthdata.nasa.gov/files/423-SPEC-001_NASA%20ESD_Preservation_Spec_OriginalCh01_0.pdf>
2. Data Lifecycle Models and Concepts by CEOS, 2012, see <http://www.ceos.org/images/DSIG/Data%20Lifecycle%20Models%20and%20Concepts%20v13.docx>
3. [EO Preserved Data Set Content](http://ceos.org/document_management/Working_Groups/WGISS/Interest_Groups/Data_Stewardship/Recommendations/EO%20Preserved%20Data%20Set%20Content_v1.0.pdf) available from <http://ceos.org/document_management/Working_Groups/WGISS/Interest_Groups/Data_Stewardship/Recommendations/EO%20Preserved%20Data%20Set%20Content_v1.0.pdf>
4. [EO Data Preservation Guidelines](http://ceos.org/document_management/Working_Groups/WGISS/Interest_Groups/Data_Stewardship/Recommendations/EO%20Data%20Preservation%20Guidelines_v1.1.pdf) available from <http://ceos.org/document_management/Working_Groups/WGISS/Interest_Groups/Data_Stewardship/Recommendations/EO%20Data%20Preservation%20Guidelines_v1.1.pdf>
5. [Preview Image Principle](http://ceos.org/document_management/Working_Groups/WGISS/Interest_Groups/Data_Stewardship/Best_Practices/WGISS_DSIG_Preview-Image-Principle-v2_Jan2013.pdf) available from <http://ceos.org/document_management/Working_Groups/WGISS/Interest_Groups/Data_Stewardship/Best_Practices/WGISS_DSIG_Preview-Image-Principle-v2_Jan2013.pdf>
6. [Data Management Statement](http://ceos.org/document_management/Working_Groups/WGISS/Interest_Groups/Data_Stewardship/Best_Practices/WGISS_DSIG_Data-Management-Statement-Apr2012.docx) available from <http://ceos.org/document_management/Working_Groups/WGISS/Interest_Groups/Data_Stewardship/Best_Practices/WGISS_DSIG_Data-Management-Statement-Apr2012.docx>
7. [CEOS Persistent Identifier Best Practices](http://ceos.org/document_management/Working_Groups/WGISS/Interest_Groups/Data_Stewardship/Best_Practices/CEOS%20Persistent%20Identifier%20Best%20Practices_v1.0.pdf) available from [http://ceos.org/document\_management/Working\_Groups/WGISS/Interest\_Groups/Data\_Stewardship/Best\_Practices/CEOS Persistent Identifier Best Practices\_v1.0.pdf](http://ceos.org/document_management/Working_Groups/WGISS/Interest_Groups/Data_Stewardship/Best_Practices/CEOS%20Persistent%20Identifier%20Best%20Practices_v1.0.pdf)
8. [Generic Earth Observation Data Set Consolidation Process](http://ceos.org/document_management/Working_Groups/WGISS/Interest_Groups/Data_Stewardship/Best_Practices/GenericEarthObservationDataSetConsolidationProcess_v1.0.pdf) available from <http://ceos.org/document_management/Working_Groups/WGISS/Interest_Groups/Data_Stewardship/Best_Practices/GenericEarthObservationDataSetConsolidationProcess_v1.0.pdf>
9. [Long Term Preservation of Earth Observation Space Data:  Preservation Workflow](http://ceos.org/document_management/Working_Groups/WGISS/Interest_Groups/Data_Stewardship/Best_Practices/Preservation%20Workflow_v1.0.pdf) available from <http://ceos.org/document_management/Working_Groups/WGISS/Interest_Groups/Data_Stewardship/Best_Practices/Preservation%20Workflow_v1.0.pdf>

#  Security Considerations (Informative)

## Introduction

The use of this Recommended Practice has a potential area of security concern, namely that in the case of data which should be confidential and its use restricted to a specific community, information is collected which allows that data to be found and used.

## security concerns with respect to the CCSDS document

This document provides guidance on Additional Information to be collected.

## Data privacy

The Additional Information may itself need to be subject to similar or different privacy considerations as the data being preserved and exploited.

## Data integrity

The Additional Information should itself be subject to the same consideration concerning preservation and authenticity as the data being preserved and exploited.

## Authentication of communicating entities

Authentication of communicating entities must be the responsibility of the individuals and organisations responsible to the data holdings and is not covered by this recommended practice.

## Control of access to resources

Control of access to resources must be the responsibility of the individuals and organisations responsible to the data holdings and is not covered by this recommended practice.

## Availability of resources

Availability of resources must be the responsibility of the individuals and organisations responsible to the data holdings and is not covered by this recommended practice.

## Auditing of resource usage

Auditing of resource usage must be the responsibility of the individuals and organisations responsible to the data holdings and is not covered by this recommended practice.

## Potential threats and attack scenarios

None.

## Consequences of not applying security to the technology

Consequences of not applying security to the data to which this recommended practice is applied will depend upon the sensitivity of the data being created/preserved.