The Consultative Committee for Space Data Systems

Draft Recommendation for
Space Data System Practices

INFORMATION PREPARATION TO ENABLE LONG TERM USE

PROPOSED DRAFT RECOMMENDED PRACTICE

CCSDS 653.0-R-0

DRAFT RED BOOK

June 2020

AUTHORITY

Issue:	Red Book, Issue 0
Date:	June 2020
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 e-mail address below.

This document is published and maintained by:

CCSDS Secretariat
National Aeronautics and Space Administration
Washington, DC, USA
E-mail: secretariat@mailman.ccsds.org

STATEMENT OF INTENT

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

The Consultative Committee for Space Data Systems (CCSDS) is an organization officially established by the management of its members. The Committee meets periodically to address data systems problems that are common to all participants, and to formulate sound technical solutions to these problems. Inasmuch as participation in the CCSDS is completely voluntary, the results of Committee actions are termed **Recommendations** and are not in themselves considered binding on any Agency.

CCSDS Recommendations take two forms: **Recommended Standards** that are prescriptive and are the formal vehicles by which CCSDS Agencies create the standards that specify how elements of their space mission support infrastructure shall operate and interoperate with others; and **Recommended Practices** that are more descriptive in nature and are intended to provide general guidance about how to approach a particular problem associated with space mission support. This **Recommended Practice** is issued by, and represents the consensus of, the CCSDS members. Endorsement of this **Recommended Practice** is entirely voluntary and does not imply a commitment by any Agency or organization to implement its recommendations in a prescriptive sense.

No later than five years from its date of issuance, this **Recommended Practice** will be reviewed by the CCSDS to determine whether it should: (1) remain in effect without change; (2) be changed to reflect the impact of new technologies, new requirements, or new directions; or (3) be retired or canceled.

In those instances when a new version of a **Recommended Practice** is issued, existing CCSDS-related member Practices and implementations are not negated or deemed to be non-CCSDS compatible. It is the responsibility of each member to determine when such Practices or implementations are to be modified. Each member is, however, strongly encouraged to direct planning for its new Practices and implementations towards the later version of the Recommended Practice.

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 the *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 sent to the CCSDS Secretariat at the e-mail 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.
- Electronics and Telecommunications Research Institute (ETRI)/Korea.
- 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 'Red Book' status indicates that the CCSDS believes the document to be technically mature and has released it for formal review by appropriate technical organizations. As such, its technical contents are not stable, and several iterations of it may occur in response to comments received during the review process.

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 653.0-R-0	Information Preparation to Enable Long Term Use, Proposed Draft Recommended Practice, Issue 0	August 2017	Current proposed draft
CCSDS-653.0-R-0	Information Preparation to Enable Long Term Use, Proposed Draft Recommended Practice, Issue 1	November 2019	Updated draft to move information about PMBOK and DMBOK to Annex A
		January 2020	Add Life Science user case.
		February 2020	Add Library user case
		May 2020	Map use cases Phases to Collection Groups more clearly
		June 2020	Inserted new space mission preservation checklist (v2020-06-15) as Annex C a

CONTENTS

Section Page

1 INTRODUCTION

1.1 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 because the term "metadata" is too unspecific and its use can cause confusion) 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.

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.

1.2 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 Additional Information 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.

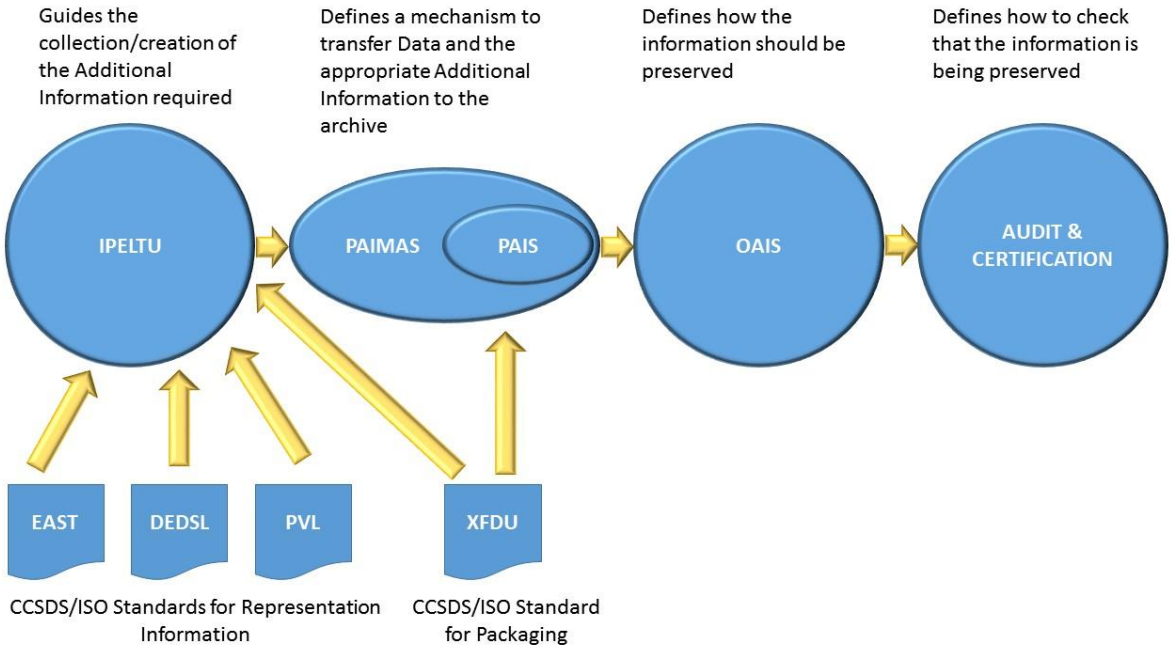

Figure 1-1 Relationship between CCSDS standards

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 negotiation is

required in order to ensure that the correct level of detail of description is produced. 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.

1.3 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.

1.4 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 change 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 the associated information which is important for preservation. But other types of Additional Information which may help on future long-term exploitation. All these 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 groups of activities, within larger project phases in a project where Additional Information should be collected (Collection Groups). These are typically where 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 help to 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.

1.5 CONFORMANCE

Conformance to this recommended practice requires that Additional Information is collected as described in Sections 4 and 5.

1.6 DOCUMENT STRUCTURE

Section 2 gives an overview of the document concepts and the way in which activities in projects can be grouped. These are expanded in the following sections. The overlapping activities that occur throughout a project or phases in projects 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 a Framework for the way in which that minimum useful information that should be captured may evolve through the project.

The Annexes provide supporting information. ANNEX A provides more details from PMBOK and DMBOK, from which a number of concepts are drawn. A brief comparison of other ways to break down projects is in ANNEX B. Checklists specific for Space projects are given in

ANNEX C. Examples of Frameworks, in less detail, are provided for a broader set of domains in ANNEX D. Security considerations are discussed in ANNEX E.

1.7 DEFINITIONS

1.7.1 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
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

1.7.2 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 capitalized, 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.

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 object composed of a set of bit sequences.

Information: Any type of knowledge that can be exchanged. In an exchange, it is represented by data.

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

Information Package: A logical container composed of optional Content Information and optional associated Preservation Description Information. Associated with this Information Package is Packaging Information used to delimit and identify the Content Information and Package Description information used to facilitate searches for the Content Information

Physical Object: An object (such as a moon rock, bio-specimen, microscope slide) with physically observable properties that represent information that is considered suitable for being adequately documented for preservation, distribution, and independent usage.

Preservation Description Information (PDI): The information which is necessary for adequate preservation of the Content Information and which can be categorized as Provenance, Reference, Fixity, Context, and Access Rights Information.

Representation Information: The information that maps a Data Object into more meaningful concepts.

Activity: A distinct, scheduled portion of work performed during 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.

Collection Groups: types of Activities where Additional Information may be collected 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, at the end of the project, turning the information over to the long-term preservation organization.
- **Control -** track, review, and orchestrate the progress and performance of the activities.

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 endeavor 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).

1.8 NOMENCLATURE

1.8.1 NORMATIVE TEXT

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

a) the words ‘shall’ and ‘must’ imply a binding and verifiable specification;

b) the word ‘should’ implies an optional, but desirable, specification;

c) the word ‘may’ implies an optional specification;

d) 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.

1.9 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/, final version available from https://www.amazon.co.uk/DAMA-DMBOK-Data-Management-Body-Knowledge/dp/1634622340

[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

[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

[9] *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

[10] *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

[11] *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

2 OVERVIEW

Following the PMBOK [1] terminology, a **Project** may be divided into any number of **Project Phases** (see for example ANNEX A). 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 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, PMBOK uses Knowledge Areas, and DMBOK uses Functions). In this document the Areas are about collections of activities for which information is needed in order 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↓	Activity 1	Activity 2			Activity n
Area 1	Expected activity of type "Activity 1" in "Area 1"	…	…	…	…
Area 2	…	…	…	…	…
Area …	…	…	…	…	…
Area m	…	…	---	---	Expected activity of type "Activity n" in "Area m"

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

2.1 GROUPINGS OF ACTIVITIES

PMBOK and DMBOK use slightly different groupings and terminology, the former uses the term Process Groups, described in section A1 while the latter uses the term Activity Groups,

described in sectionA2. In this document we use the term Collection Groups, which best fits the requirements of this document and are described in section 2.1.1.

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

2.1.1 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 specialization 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, at the end of the project, 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.

2.2 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,** described in A3.While DMBOK currently uses the term **Functions,** described in A4 . This document uses the term **Additional Information Areas**, described in section 2.2.1.

2.2.1 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 finer granularity by using the concepts provided by the OAIS Information Model. OAIS defines the information 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

- Content 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 uses of the data
    - Potential Designated Community/Knowledge Base
    - Potential Transformational Information Properties

These are discussed in more detail in section 4.

3 COLLECTION GROUPS

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

3.1 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.

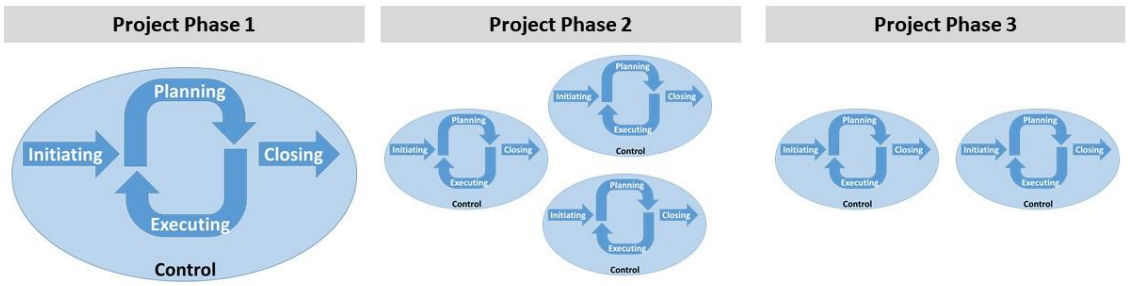

Figure 3-1 Example Collection Groups within a project with three phases

3.2 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.

3.2.1 THE INITIATING COLLECTION GROUP

The Initiating Collection Group consists of processes performed to justify 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.

3.2.2 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.

3.2.3 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 such as processing algorithms, validation and qualification plans, qualification matrixes, testing results, and project logs.

3.2.4 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 signed acceptances, procurement documents, associated data, and related publication.

3.2.5 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.

4 ADDITIONAL INFORMATION AREAS TO ENSURE LONG-TERM USABILITY

An archive must create Archival Information Packages (AIP) 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.

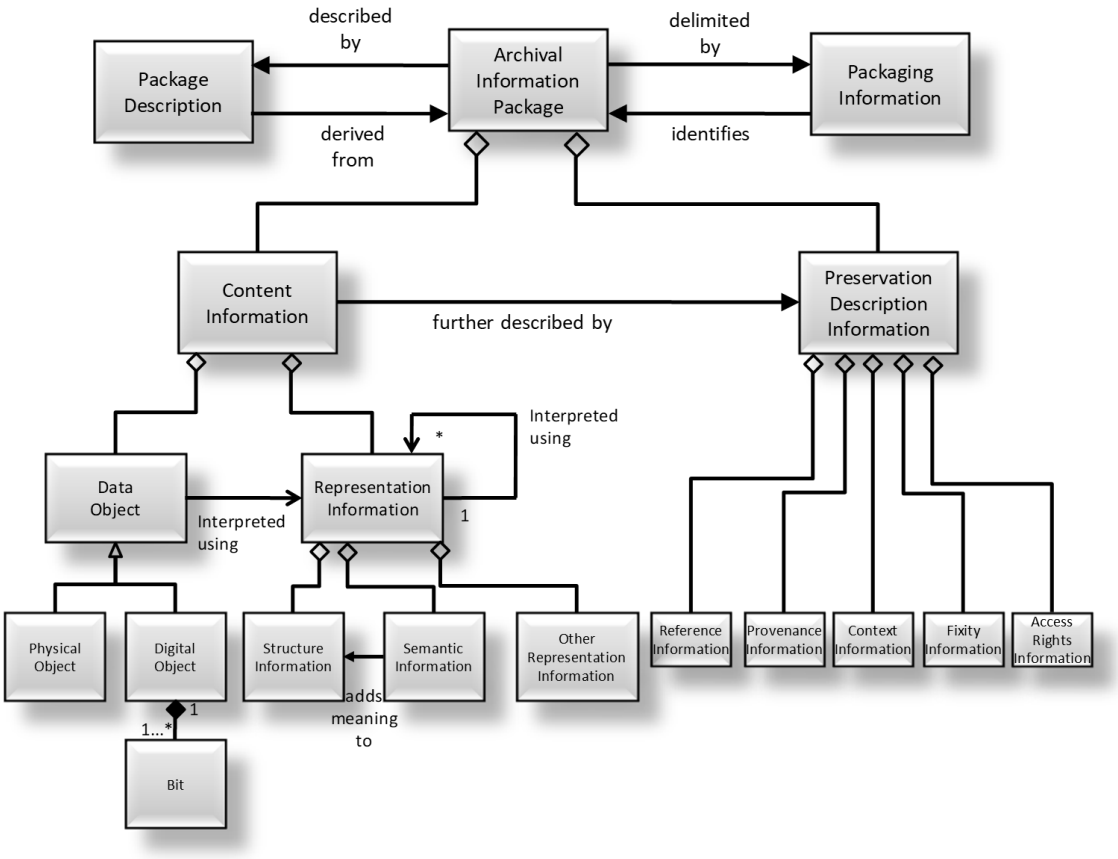

Figure 4-1 Archival Information Package (Detailed View)

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.

4.1 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.

4.1.1 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.

4.1.1.1 Data Objects

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

- Raw data, for example from scientific instruments. Space missions often label data produced by successive stage of processing including
  - Level 0
  - Level 1 … n
- The data which encodes other Information Objects that are to be preserved, including
  - Provenance
  - Representation Information
  - Software of various kinds including data processing software
  - Auxiliary 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

4.1.1.2 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
- Annotation systems used with the data (if any)
- 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).

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.

4.1.2 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.

4.1.2.1 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
- Reasons for selecting a particular referencing convention

4.1.2.2 Provenance Information

Provenance Information provides information including

- specific aspects of the project origins and history,
    - Mission documentation including
        - Mission architecture documents describing purpose, scope and performances of the mission and of the on-board instruments, information relevant orbits, platform position, attitude, ground coverage (acquisition footprint), head-roll-pitch.

        - Documents describing data and products formats specification.
        - Documents describing measurement requirements and/or measurement performances (theoretical models). Documents drawing instruments characteristics, performances and instrument description (physical implementations).
        - Documents describing models and/or algorithms needed (used) to obtain mission data and products including specific/special cases, known errors and configuration necessities. In other words, all documents covering conceptual environment, its implementation and its operations.
        - Reports concerned with measurement trends, failures, changes of performances, and out of service for any reason.
        - Documents related to the process of data qualification: precision, numerical representations, formats, uncertainties, errors, adjustment/correction methods (e.g. Cal/Val procedures and documents).
    - from what it was derived i.e. previously collected data
    - processing software
    - 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
- Defence against hacking
- Which anti-virus checks performed
- Roles of people e.g. who can change/delete

4.1.2.3 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. Examples include:

- 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
- research publications based on the data
- publications containing the data

4.1.2.4 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

4.1.2.5 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
- Designated Community
- Legal framework(s)
- Licensing offers
- Specifications for rights enforcement measures applied at
- Dissemination time
- Pointers to Fixity and Provenance Information (e.g., digital signatures, and rights holders)

4.1.3 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.

4.1.4 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.

4.2 INFORMATION AREAS DERIVED FROM ISSUES OUTSIDE THE INFORMATION MODEL

4.2.1 PUBLICATIONS

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

- documents about the data – some of these documents may also be Representation Information

- Scientific publications based on the data exploitation or relevant to them (properly linked to the data) and outreach material.
- Reports and outcomes from events like congresses, studies, communities and investigators concerned with models' review, algorithm changes and Cal/Val changes affecting data processing chains.
- Community tagging e.g. quality tags held by 3rd parties

4.2.2 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

4.2.3 POTENTIAL 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. Alternatively, the data may be a text document about a particular topic or an image which represents a concept or an audio recording of some activity.

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. A text document may have other uses as an object for text mining; an image may be used to analyze the use of colors; an audio recording may be harmonically analyzed to extract other interesting or important aspects of the recording.

Some or all of these may be used as tests of preservation i.e. can the digital objects continue to be used in these ways in the future.

4.2.4 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.

4.2.5 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.

5 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.

Collection Group / **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 • Processing logs • 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)
Package Description			• 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 Additional Information in Collection Groups

ANNEX A

PMBOK AND DMBOK

(INFORMATIVE)

A1 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.

A2 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.

A3 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 specialization. These ten Knowledge Areas are used in most projects most of the time.*"

A4 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
- Metadata Management
- Data Quality Management 

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

ANNEX B

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:

A) Long Term Preservation(LTDP), see http://ceos.org/document_management/Working_Groups/WGISS/Documents/WGISS_DSIG-Data-Lifecycle-Models-and-Concepts-v8_Sep2011.docx also https://earth.esa.int/documents/1656065/1681917/LTDP_PDSC_4.0.pdf and https://www.earthobservations.org/documents/se/130_GEO_ltdp_guidelines.pdf

B) The NASA Mission Life Cycle as defined in NASA Procedural Requirements NPR 7120.5E available from https://nodis3.gsfc.nasa.gov/npg_img/N_PR_7120_005E_/N_PR_7120_005E_.pdf . 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.

C) PMBOK, the Project Management Body of Knowledge, discussed earlier in this document.

D) A Survey on Data Lifecycle Models: Discussions toward the 6Vs Challenges, 2015, see https://www.ac.upc.edu/app/research-reports/html/RR/2015/18.pdf

The Controlling group processes are 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.

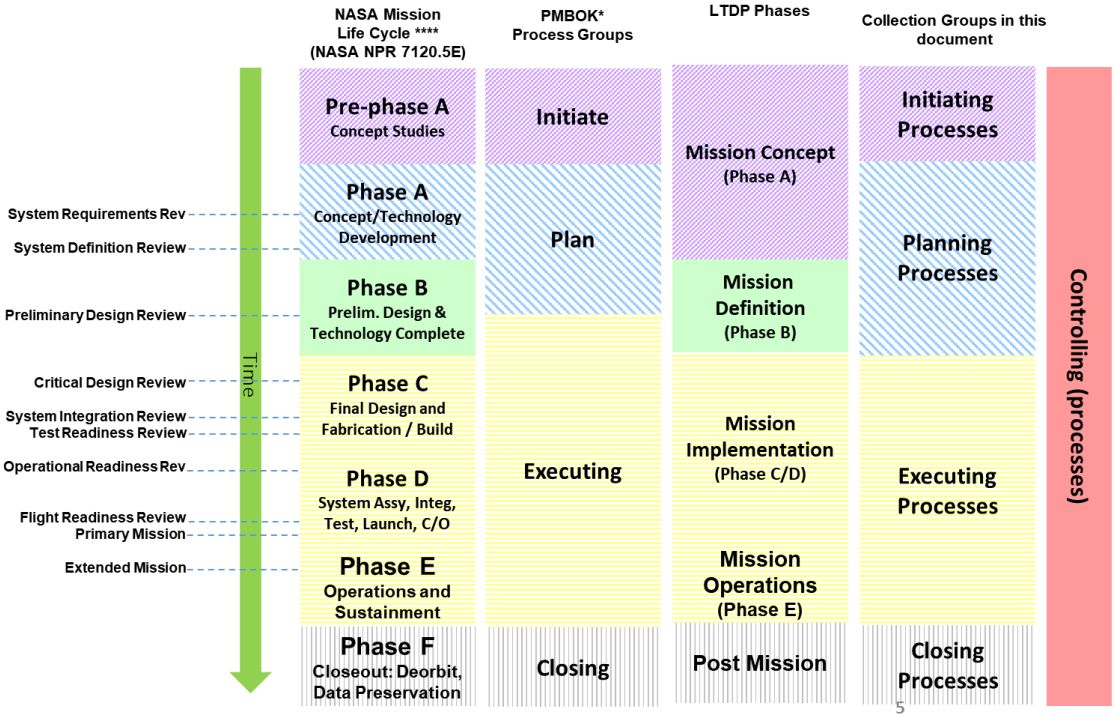

Figure 5-1 Mapping Collection Groups to commonly used phases

ANNEX C: SPACE MISSION DIGITAL TARGET OF PRESERVATION PROFORMA (DTOPP) CHECKLIST (INFORMATIVE)

C1.1 INTRODUCTION

This document provides a checklist for a space mission, program, or project (hereafter referred to as simply "mission") to document the policy which is implemented by that mission's program/project management concerning what data is important enough to be preserved in the long term[1], past the "event horizon" of system/software obsolescence. Proforma means the conventional business communications definition; an example to show how other documents of the same type should be written or prepared.

This DTOPP Checklist is an example intended to be used generally by an organization to tell the mission participants (stakeholders, customers, contractors, subcontractors, etc.) which digital data under this space mission program/project should be prepared for long-term preservation. For example, when a certain data type is identified as a preservation target, subcontractors will know that they shall preserve adequate metadata or executable application software so that the data will be retrievable and understandable in the long term. Basically, this checklist is a coordination tool to ensure that ***failures*** to preserve mission products and information are made by a conscious choice of the mission management rather than by oversight.

The intention is that program management for the mission, in concert with their legal counsel and records manager, will establish this list during the pre-phase A (earliest) stage of the program, and will indicate by checkmarks the types of data that they expect to be preserved throughout the mission, and after mission termination. Concurrently, program management will then know what funding and resources to establish for the gathering of metadata, establishment of migration strategy, or arrangements for long-term hosting of applications as long as the object data is intended to be retrievable. It will also cause management to address funding and planning strategies for turnover of long-term preservation object data to post-mission establishments (organization CIO, national archives, etc.) for long term preservation of that object data after mission termination.

This DTOPP Checklist is formatted in a manner and in a sufficient level of detail that missions can use as attachments for contracts and other program/project management vehicles. If a mission finds portions of the checklist unsuitable, the CCSDS Data Archive Interoperability (DAI) Working Group (WG) welcomes participation and inputs to improve the Space Mission DTOPP Checklist for later use and for other missions. This entire section, including this introductory material, should be included in mission documentation so that mission participants will understand the purpose, motivation and value of the DTOPP Checklist. A word-processing document (docx) is available on the CCSDS website, with easily "checkable"

[1] Definition of "Long Term" from OAIS (CCSDS standard **650.0-M-2**): A period of time long enough for there to be concern about the impacts of changing technologies, including support for new media and data formats, and of a changing Designated Community or changes to the Designated Community's Knowledge Base, on the information being held in an OAIS. This period extends into the indefinite future."

items; it can be used as a template for your specific mission. Once management decides what to preserve, it should be distributed to the team in write-protected format and included as an attachment to contracts.

C1.2 SPACE MISSION DTOPP CHECKLIST FORM

Annex to *Information Preservation to Enable Long Term Use* (CCSDS 6NN…)

Mission Name: _____
Company/Agency: _____
Contract (if applicable): _____

C1.3 LEVEL 1 STATEMENT (CHOSE ONE):

☐ This organization/mission chooses to enact long-term digital preservation for the valuable products of this mission as described below in the level 2 and 3 statements.
Proceed to Level 2.

☐ This organization/mission chooses to not enact long-term digital preservation for any data associated with this mission because no products of this mission will be of value to stakeholders, the public, or future mission developers after this mission terminates. No further completion of this form is needed. NOTE: Please consult legal counsel and records management before checking this box.

C1.4 LEVEL 2 STATEMENT: (CHOOSE ALL APPLICABLE)

	Data Type
☐	C1.5.1 Spacecraft-originated Science Telemetry
☐	C1.5.2 Other Science Data Products
☐	C1.5.3 Ground-originated Science Data
☐	0 Spacecraft originated Systems Telemetry
☐	0

	Data Type
☐	Raw Telemetry Data
☐	Level 0 Telemetry Products

☐	Associated major/minor frame and channel structure definitions
☐	
☐	
☐	

☐	Level 1 Telemetry Products

☐	Position, altitude and spin phase of the spacecraft

		☐	Command history and comments
		☐	Calibration of the spacecraft clock
		☐	
		☐	
	☐	Level 2 Telemetry Products	
		☐	Calibration algorithms for all parameters
		☐	
		☐	
	☐	Command history and comments	
	☐	Ancillary data	
	☐	Data Dictionaries, XML Schema, Schematron validation files (or equivalent)	
	☐	Human language terms and definitions	
	☐	Representative onboard computer dumps	
	☐	Select spacecraft telemetry system and instrument design information	

Ground-originated Systems Data

☐ 0

	Data Type
☐	Uplink data, including commands and command sequences that are ground originated.
☐	Command history in database format
☐	Spacecraft observations from ground assets
☐	Telemetry and data from ground communications system that provide information about the status of the transport of the spacecraft data to the archive.
☐	Context Information: Additional data from ***ground sources*** about the environment within which the data were collected.
☐	Onboard systems design/descriptions as needed to interpret the systems telemetry.
☐	Onboard systems design documentation
☐	Provenance information - observing logs, hardware descriptions, archive plans, etc.
☐	High-level introductory documentation - Documents that support the scientific use of the science data.
☐	Specifications and Standards (including version number) that the science data was intended to comply with.
☐	Other Documentation: Information about understanding how the data were produced or are to be used.
☐	
☐	
☐	
☐	
☐	
☐	

C1.4.1 Spacecraft Operations Data

These data types are operational products that can be valuable for organizations to analyse to improve future missions or to analyse products from this mission, when not already covered above.

	Data Type
☐	Flight rules setting boundaries for safely and effectively operating the spacecraft
☐	Procedures for operating, maintaining and troubleshooting everything aboard
☐	Lists of in-flight anomalies, their causes and their solutions
☐	Logistics: materials, their properties and their locations. Logistics process descriptions.
☐	Lists and descriptions authorizing configuration changes
☐	Resource plans
☐	Activity schedules
☐	Robotic timelines
☐	Ground-operated onboard activity schedules
☐	Ground activity schedules
☐	Inter-team communications covering the negotiations through the above topics
☐	Commercial Proprietary data that is protected and the methods (keys) to decrypt it.
☐	
☐	
☐	

C1.4.2 Human-Crewed Vehicle Unique Data

Discussions of the International Space Station program illuminated some special needs for a crewed vehicle. In particular, for a ***long-term*** crewed vehicle, these items may be important to mission planners that are looking forward to surface colonies, etc.

	Data Type
☐	State of the vehicle as it changes over time (crew- and ground-initiated changes)
☐	Software configuration lessons learned for non-vehicle devices (tablets, etc.)
☐	Activity Schedules
☐	Crew timelines (an addition to the uncrewed activity schedules in prior table)
☐	Medical Privacy data that is protected, and the methods (keys) to decrypt it.
☐	Scientific results of experiments in long-duration human spaceflight.
☐	Management lessons learned in long-duration human spaceflight.
☐	
☐	
☐	

	☐
	☐

C1.4.3 Other data types not yet expanded

There are a number of other data types that Mission Managers/Planners should consider as targets of long-term preservation. However, this SDO has not yet found volunteer authors with real mission experience to expand them yet. One check box is provided for that data type, so managers can indicate a general plan to reserve that data type. A future version of this document will hopefully expand them each of these data types to the section numbers indicated, with inputs from mission users.

☐	**Errore. L'origine riferimento non è stata trovata. Errore. L'origine riferimento non è stata trovata.**
☐	**Errore. L'origine riferimento non è stata trovata. Errore. L'origine riferimento non è stata trovata.**
☐	C1.5.6 Spacecraft Operations Data
☐	**Errore. L'origine riferimento non è stata trovata. Errore. L'origine riferimento non è stata trovata.**
☐	**Errore. L'origine riferimento non è stata trovata. Errore. L'origine riferimento non è stata trovata.**

C1.5 LEVEL 3 STATEMENT (CHOOSE ALL APPLICABLE)

To reiterate, this proforma checklist is intended to be an example. It is essential that program and project management for space missions should clearly identify what digital assets and data are necessary to preserve the mission products in the long term, after the operational mission ground systems are obsolete. This form can be modified by a program/project to utilize program-specific terms and definitions. However, as a reminder, those program-specific terms and definitions need to be documented in order for later personnel (not involved in the mission) and later systems (newly developed systems) to recover and use the mission products.

C1.5.1 Spacecraft-originated Science Telemetry

Interpretation of the below list requires understanding of conventional definitions of Level 0, 1 and 2 telemetry processing. For the purposes of this generalized list, we have adopted these definitions:

- In level 0 processing, duplicate data are removed from the data stream, data are time ordered, and data quality and accounting summaries are appended.
- In level 1 processing, the data are separated out by instrument and each instrument data set is formatted to meet the requirements of that data set and team.
- Level 2 processing includes such operations as application of calibration data and detector response maps, organization of data into appropriate energy and time bins, and application of ancillary data.

Note that Systems telemetry addressed below in 0 may include the science telemetry if it is archived as the original intact telemetry stream. It is broken out here separately because some science facilities may only archive the science telemetry after separation from systems telemetry.

Spacecraft originated Systems Telemetry

	Data Type	
☐	Raw Telemetry Data recorded as a stream (requires products below for interpretation)	
☐	Level 0 Telemetry Products	
	☐	Science Telemetry through Level 2 processing
	☐	Associated major/minor frame and channel structure definitions
	☐	
	☐	
☐	Level 1 Telemetry Products	
	☐	Science Telemetry through Level 1 processing
	☐	Position, altitude and spin phase of the spacecraft
	☐	Command history and comments
	☐	Calibration of the spacecraft clock
	☐	
	☐	
☐	Level 2 Telemetry Products	
	☐	Calibration algorithms for all parameters

	☐	
	☐	
☐	Ancillary data	
☐	Data Dictionaries, XML Schema, Schematron validation files (or equivalent)	
☐	Human language terms and definitions	
☐	Representative onboard computer dumps	
☐	Select spacecraft telemetry system and instrument design information that informs interpretation of the telemetry	

C1.5.2 Other Science Data Products

Interpretation of the below list requires understanding of Science Data Processing of raw data in analytical formats. The conventional definition for raw data is reconstructed, unprocessed instrument and payload data at full resolution, with any and all communications artifacts (e.g., synchronization frames, communications headers, duplicate data) removed.

	Data Type	
☐	Science data: The data to be preserved, for example an image from a camera or a time series from a magnetometer.	
	☐	Raw data (in analytical formats)
	☐	Calibrated data (reversible and non-reversible)
	☐	Higher-order results (photometry, maps, shape models, production rates, etc.)
☐	Calibration data: The data (flat fields, dark current, sky images, crosstalk etc.) needed to perform the calibration of the science data.	
☐	Calibration documentation: The documentation which describes the process and algorithms applied during the calibration of the science data.	
☐	Geometry data: The data needed to orient the science data.	
	☐	Attitude and pointing
	☐	Target ephemerides
	☐	Correlation of data structure to pointing (e.g., FITS WCS)
☐	Context Information: Additional data from the ***spacecraft*** about the spacecraft's environment within which the data were collected.	

C1.5.3 Ground-originated Science Data products

This data type originates on the ground. It may be a science data product, or other ground-originated data required to interpret the telemetry from the spacecraft.

	Data Type

☐	Uplink data, including commands and command sequences that are ground originated.
☐	Command history in database format
☐	Context Information: Additional data from ***ground sources*** about the environment within which the data were collected.
☐	Investigation - The mission or project managing the collection of the science data
☐	Observing System - The spacecraft and/or other platforms on which the instrument was mounted.
☐	Instrument - The identification of the instrument used to collect the science data
☐	Target - The object(s) from or for which the science data were collected.
☐	Facility - A site or institution involved in the collection of the science data or institutions that could have made decisions that affected the quality of the data and documentation being deposited
☐	Provenance information - observing logs, hardware descriptions, archive plans, etc.
☐	High-level introductory documentation - Documents that support the scientific use of the science data.
☐	Specifications and Standards (including version number) that the science data was intended to comply with.
☐	Other Documentation: Information about understanding how the data were produced or are to be used.

C1.5.4 Spacecraft originated Systems Telemetry

These are the data products received from the spacecraft, primarily traditional telemetry, but could also include "session traffic" for internet protocol transactions. This may also include the science telemetry, which is also included in the first section above, if they are archived in an intact stream as received from the spacecraft.

	Data Type	
☐	Raw Telemetry Data	
☐	Level 0 Telemetry Products	
	☐	Associated major/minor frame and channel structure definitions
	☐	
	☐	
	☐	
☐	Level 1 Telemetry Products	
	☐	Position, altitude and spin phase of the spacecraft
	☐	Command history and comments
	☐	Calibration of the spacecraft clock
	☐	
	☐	

☐	Level 2 Telemetry Products	
	☐	Calibration algorithms for all parameters
	☐	
	☐	
☐	Command history and comments	
☐	Ancillary data	
☐	Data Dictionaries, XML Schema, Schematron validation files (or equivalent)	
☐	Human language terms and definitions	
☐	Representative onboard computer dumps	
☐	Select spacecraft telemetry system and instrument design information	

C1.5.5 Ground-originated Systems Data

This data type originates on the ground but is required to interpret the products from the spacecraft as well as the performance and history of the spacecraft design and operations.

	Data Type
☐	Uplink data, including commands and command sequences that are ground originated.
☐	Command history in database format
☐	Spacecraft observations from ground assets
☐	Telemetry and data from ground communications system that provide information about the status of the transport of the spacecraft data to the archive.
☐	Context Information: Additional data from ***ground sources*** about the environment within which the data were collected.
☐	Onboard systems design/descriptions as needed to interpret the systems telemetry.
☐	Onboard systems design documentation
☐	Provenance information - observing logs, hardware descriptions, archive plans, etc.
☐	High-level introductory documentation - Documents that support the scientific use of the science data.
☐	Specifications and Standards (including version number) that the science data was intended to comply with.
☐	Other Documentation: Information about understanding how the data were produced or are to be used.
☐	
☐	
☐	
☐	
☐	
☐	

C1.5.6 Spacecraft Operations Data

These data types are operational products that can be valuable for organizations to analyse to improve future missions or to analyse products from this mission, when not already covered above.

	Data Type
☐	Flight rules setting boundaries for safely and effectively operating the spacecraft
☐	Procedures for operating, maintaining and troubleshooting everything aboard
☐	Lists of in-flight anomalies, their causes and their solutions
☐	Logistics: materials, their properties and their locations. Logistics process descriptions.
☐	Lists and descriptions authorizing configuration changes
☐	Resource plans
☐	Activity schedules
	☐ Robotic timelines
	☐ Ground-operated onboard activity schedules
	☐ Ground activity schedules
☐	Inter-team communications covering the negotiations through the above topics
☐	Commercial Proprietary data that is protected and the methods (keys) to decrypt it.
☐	
☐	
☐	

C1.5.7 Human-Crewed Vehicle Unique Data

Discussions of the International Space Station program illuminated some special needs for a crewed vehicle. In particular, for a ***long-term*** crewed vehicle, these items may be important to mission planners that are looking forward to surface colonies, etc.

	Data Type
☐	State of the vehicle as it changes over time (crew- and ground-initiated changes)
☐	Software configuration lessons learned for non-vehicle devices (tablets, etc.)
☐	Activity Schedules
	☐ Crew timelines (an addition to the uncrewed activity schedules in prior table)
☐	Medical Privacy data that is protected, and the methods (keys) to decrypt it.
☐	Scientific results of experiments in long-duration human spaceflight.
☐	Management lessons learned in long-duration human spaceflight.
☐	
☐	
☐	
	☐
	☐

C1.5.8 Other data types not yet expanded

There are a number of other data types that Mission Managers/Planners should consider as targets of long-term preservation. However, this SDO has not yet found volunteer authors with real mission experience to expand them yet. One check box is provided for that data type, so managers can indicate a general plan to reserve that data type. A future version of this document will hopefully expand them each of these data types to the section numbers indicated, with inputs from mission users.

☐	**1.1.1**

5.1.	Data Type
☐	Uplink data, including commands and command sequences that are ground originated.
☐	Command history in database format
☐	Spacecraft observations from ground assets
☐	Telemetry and data from ground communications system that provide information about the status of the transport of the spacecraft data to the archive.
☐	Context Information: Additional data from ***ground sources*** about the environment within which the data were collected.
☐	Onboard systems design/descriptions as needed to interpret the systems telemetry.
☐	Onboard systems design documentation
☐	Provenance information - observing logs, hardware descriptions, archive plans, etc.
☐	High-level introductory documentation - Documents that support the scientific use of the science data.
☐	Specifications and Standards (including version number) that the science data was intended to comply with.
☐	Other Documentation: Information about understanding how the data were produced or are to be used.
☐	
☐	
☐	
☐	
☐	
☐	

C1.5.9 Spacecraft Operations Data

These data types are operational products that can be valuable for organizations to analyse to improve future missions or to analyse products from this mission, when not already covered above.

	Data Type
☐	Flight rules setting boundaries for safely and effectively operating the spacecraft
☐	Procedures for operating, maintaining and troubleshooting everything aboard
☐	Lists of in-flight anomalies, their causes and their solutions
☐	Logistics: materials, their properties and their locations. Logistics process descriptions.
☐	Lists and descriptions authorizing configuration changes
☐	Resource plans
☐	Activity schedules
☐	Robotic timelines
☐	Ground-operated onboard activity schedules
☐	Ground activity schedules
☐	Inter-team communications covering the negotiations through the above topics
☐	Commercial Proprietary data that is protected and the methods (keys) to decrypt it.
☐	
☐	
☐	

C1.5.10 Human-Crewed Vehicle Unique Data

Discussions of the International Space Station program illuminated some special needs for a crewed vehicle. In particular, for a ***long-term*** crewed vehicle, these items may be important to mission planners that are looking forward to surface colonies, etc.

	Data Type
☐	State of the vehicle as it changes over time (crew- and ground-initiated changes)
☐	Software configuration lessons learned for non-vehicle devices (tablets, etc.)
☐	Activity Schedules
☐	Crew timelines (an addition to the uncrewed activity schedules in prior table)
☐	Medical Privacy data that is protected, and the methods (keys) to decrypt it.
☐	Scientific results of experiments in long-duration human spaceflight.
☐	Management lessons learned in long-duration human spaceflight.
☐	
☐	
☐	
☐	
☐	

	C1.5.11 Other data types not yet expanded There are a number of other data types that Mission Managers/Planners should consider as targets of long-term preservation. However, this SDO has not yet found volunteer authors with real mission experience to expand them yet. One check box is provided for that data type, so managers can indicate a general plan to reserve that data type. A future version of this document will hopefully expand them each of these data types to the section numbers indicated, with inputs from mission users. **1.1.2**
☐	**1.1.3** ERRORE. L'ORIGINE RIFERIMENTO NON È STATA TROVATA.
☐	**1.1.4** ERRORE. L'ORIGINE RIFERIMENTO NON È STATA TROVATA.
☐	**1.1.5** ERRORE. L'ORIGINE RIFERIMENTO NON È STATA TROVATA.
☐	**1.1.6** ERRORE. L'ORIGINE RIFERIMENTO NON È STATA TROVATA.

Additional examples of information to be collected are provided in the Annexes of *Long Term Data Preservation Earth Observation Preserved Data Set Content LTDP/PDSC* (see reference A) of ANNEX B)

ANNEX D

EXAMPLE USE CASES

(INFORMATIVE)

D1 EXAMPLE SMALL RESEARCH PROJECT

D1.1 MOTIVATION FOR PROJECT

An individual researcher wishes to perform an experiment and publish results that are used and cited by others.

D1.2 EXAMPLE PHASES AND COLLECTION GROUPS

Phase	**Collection Groups**			
	Initiating	**Planning**	**Executing**	**Closing**
Application for funding to perform the experiment.	Record idea for the research	Application for funding	Collect CRIS data and best guess at data format and semantics	Application success
writes a Data Management Plan (DMP) as required by the funders	Requirements for DMP	Strategy for writing DMP	Collect DMP as written and collect modifications as time goes on	
sets up the experiment and data collection system	Collect requirements and rationale for experiment setup	Plan details of experiment setup	Collect details of the experiment and data system as created	Record status of experiment readiness and initial calibration
he/she performs the experiment, thereby producing data	Initiate experiment	Plan the details of the experiment	Run experiment and collect data plus calibration of the data and other Representation Information	Close experiment
the researcher analyses the data using software he/she has created and publishes results	Gather ideas on how to data might be used. Estimate data volumes to be published	Update existing Additional Information.	Researcher, and perhaps the archive, adds Representation Information to be sure the Designated Community (and perhaps others) can use the data and PDI.	Finalise Additional Information and hand-over.

			Create Descriptive Information to help users find the Information. Publish and archive the Information.	

D2 LARGE RESEARCH PROJECT

D2.1 MOTIVATION FOR PROJECT

A scientific instrument is required by a space agency for a satellite which is to study the Sun.

D2.2 EXAMPLE PHASES AND COLLECTION GROUPS

Phase	Collection Groups			
	Initiating	**Planning**	**Executing**	**Closing**
Several multi-national consortia submit proposals, which include appropriate Data Management Plans.	Decide to submit proposals	Go through initial plans to decide on budget needed and technologies required. Collect the decisions made.	Create the proposal for the bid including hardware and software requirements and initial designs and feasibility studies. Collect Additional Information including the design justifications and compromises and data processing requirements.	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.	Agree on distribution of work	Each consortium member makes plans for performing their work.	Perform the work including detailed design, construction, testing and calibrating the data collection instruments. Collect calibration information, including data on imperfections and non-conformities, processing software and build systems.	Deliver the instrument(s)

Over a period of 10 years the instrument is built and integrated into the satellite.	Agree integration procedures	Detailed planning of integration steps.	Integrate and test hardware, software and communications. Collect Additional Information including calibration of instruments in the satellite and communications setup e.g. packages and encodings used in communications.	Ready for launch
The satellite is launched, and the instrument is deployed and collects data.	Launch satellite	Schedule operations. Including commissioning, making observations and calibrations.	Carry out the operations of the satellite. Collect Additional Information including data produced by the instruments plus engineering data e.g. temperatures, voltages, particle radiation etc.	End when satellite operations cease.
The data is collected at a ground station and sent to the researchers who are part of the instrument consortium.	Set up data collection system (probably in parallel with satellite preparation)	Schedule data transmission and plan project data storage, processing and distribution.	Collect, store, process and distribute data. Collect Additional Information including data definitions, including data semantics, software source code and build systems and processing system dependencies.	Hand over to long term archive.
Modifications are made from time to time to the on-board software and the data processing software.	These will be initiated as required, in parallel with the phases above.	Plan the changes to integrate as smoothly as possible into the operations of the satellite, including testing.	Carry out the changes as scheduled. Collect Additional Information about the changes including testing results, recalibrations and changes required to data processing system.	Continue until not required.
The Data Objects and Additional Information are sent to an archive for preservation and re-use by other researchers.	Agree on long term archive or at least next in the chain of preservation.	Plan handover of all required information to the archive.	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.	

D3 AIRCRAFT MANUFACTURE

D3.1 MOTIVATION FOR PROJECT

An aircraft manufacturer wishes to create and sell a new type of aircraft, following international regulations.

D3.2 EXAMPLE PHASES AND COLLECTION GROUPS

Phase	Collection Groups			
	Initiating	**Planning**	**Executing**	**Closing**
A large aircraft manufacturer wishes to create and sell a new type of aircraft.	Company makes decision to proceed.	Create plans and budget for new aircraft.	Carry out plans for next steps in development. Collect Additional Information including budgets, aircraft	

			requirements and technologies.	
The initial design team creates a design which is tested and improved by a number of other specialist teams.	Create initial design.	Schedule the specialist teams to review design.	Specialist teams revise designs. Collect Additional Information about designs and the changes to the designs.	Sign-off the production designs.
A number of sub-systems, such as engines and wings, are sub-contracted to other specialist manufacturers.	Agree with sub-contractors to carry out the work.	Schedule the work with sub-contractors.	Carry out the scheduled work and construct the aircraft. Collect Additional Information including designs, software design system and details of systems on which it depends.	End of aircraft production.
The aircraft goes into production and is sold world-wide for the next 40 years.	Commitment to budget and operational requirements.	Schedule production, sales and finance interdependencies.	Produce, sell and maintain aircraft. Collect Additional Information including budgets, technical details, operational and maintenance details.	End with end of aircraft operations.
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.	Initiate changes as required by, for example, commercial, technical or regulatory demands.	Plan changes to fit in with safety, production and operational requirements.	Carry out the changes. Collect Additional Information including design, software, technical,	End of aircraft production.

	These will be carried out in parallel with the previous phase.		regulatory, safety and certification requirements.	
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.	Initiate long term archiving of information.	Schedule the hand-over of all relevant information and resources.	Handover the information agreed and carry out preservation activities.	Carry on as long as required.
In addition, the information can be used by the manufacturer to develop variants of the aircraft and also entirely new types of aircraft.	When required in new designs, seek to learn from previous designs.	Plan retrieval and re-use of the aircraft information (including Additional Information)	Re-use appropriate parts of the information about the aircraft. Collect Additional Information about variants and new designs.	Continue as long as required.

D4 LIFE SCIENCES

D4.1 MOTIVATION FOR PROJECT

Life Sciences companies are generating huge volume of datasets by collecting raw data, documents, images, multimedia, files, records, dossiers from internal R&D laboratories and departments, Manufacturing, Regulatory Affairs, Legal and external partners or investigators (CROs, hospitals, Universities, Research Centers, …). During the product lifecycle, the involved organizations are identifying the relevant processes and events for the data patrimony curation and preservation in the long-term.

Pharmaceutical, Biotech and Medical Devices companies have common denominators for managing the data patrimony as a company digital resource. A key focus is on meeting regulations and GxP compliance, protecting reputation, reducing clinical trials archiving costs, keeping records safe and secure, ensuring confidentiality, distributing controlled information packages to the designated communities. New regulations and compliance constraints require a continuous control and monitoring between business requirements, compliance requirements and applicable metrics.

Biotech and Medical Devices companies operate in the market with compliance constraints, obligations and risks as for Pharma industry, but the small and medium sizes of their organizations determine a big effort and difficulties in managing data patrimony governance in the long-term in a cost-effective manner with a service approach. Data preservation services should be based on a sustainable information model with secure and controlled access of sensitive information.

D4.2 EXAMPLE PHASES AND COLLECTION GROUPS

In Biotech companies, researchers use and produce huge amount of data deriving from the different stages of biomedical science research lifecycle and from a diversity of sources and scientific communities. The usage of interdisciplinary bioinformatic methods and technologies determines an increasing umbrella of new data formats subjected to obsolescence in the medium and long-time. The biotech researcher performs the pipeline studies, starting from biological data sources, measuring biomarkers and thereby producing big volume of data. The researcher compares and analyses complex data sequences using laboratories dedicated tools and disseminates the results (publications and raw data) in secure collaboration network where the researcher's communities can access and contribute to the development of the research results.

Medical devices manufacturers or distributors have a to meet compliance to the Medical Devices Regulation (MDR), ensuring safety and quality of the devices for patients and tracking the data objects since the project initiation phase, until the execution of devices testing and the validation of results to ensure regulations compliance.

Phase	Collection Groups			
	Initiating	**Planning**	**Executing**	**Closing**
In Pharmaceutical industry, a new drug goes in Phase III.	Set up budget for trial	Plan study protocols for clinical trials, constrained by regulations. Collect Additional Information about regulations in operation and study protocols.	Carry out trials and collect information and create Trial Master File. Exchange information between Sponsor and CRO using e-clinical trial platforms. Create submission dossiers and market authorization of the drug. Data integrity and 9 ALCOA+ principles must be ensured in a time frame of over 25 years.	Present results and close trial.

			Collect additional information such as (perhaps anonymized) participant data, trial results including appropriate Representation Information for data files, including Provenance and Fixity Information to provide evidence for data integrity.	
Preserve the information packages for the long-term in a trusted and secure digital archive, for GxP compliance, Intellectual Property and legal reasons,		Create data management and preservation plan to support the whole data curation and long-term preservation and identify and certify data objects, additional information, relevant events, applicable procedures (for transfer, appraisal, ingestion, archival storage, access) and services that the archive must ensure. Collect Additional Information about what needs to be preserved and plans to undertake the preservation.	Preserve the Information. Collect Additional Information includes Representation Information such as the data format, semantics and processing software, which have been created by various internal team and external teams according to the submission agreement and related contract clauses. When the contracts with CROs expires, some relevant information must be provided to the Pharmaceutical company digital archive instead to be lost.	Continue preservation activities

For patient centric R&D lifecycle, the data curation and preservation are mandatory not only to ensure compliance but to facilitate the researchers developing new drug variations depending on the live data coming directly from the patients and the associations of the patients, and on a huge volume of historical data preserved in the digital archive.

To improve the experience of the designated communities, during the data exploitation the digital archive adds descriptive information and representation information that facilitate the understanding of data over time.

D5 LIBRARIES

D5.1 MOTIVATION FOR PROJECT

Since the beginning of human history, libraries have been a foundation for culture, education, and social participation. Libraries key role consists in ensuring in the centuries access and preservation of historical and actual materials and treasures physically stored into it. The digital transformation has strong impacts in archiving and disseminating the information concerning the history of humanity. Libraries enforced their role of competent digital incubators of culture, acting as a continuous producer of knowledge, transforming the operational processes, adopting new organizational model, innovative spaces and technologies, building digital platforms to provide new services and facilitate a broad engagement of users.

To manage this transformation, digitalization programmes based on a sustainable financial plan are defined, whose objectives are the rethinking of the information model, the digitization of physical material in internal labs or co-labs with other partners, the widespread and interoperable access between Libraries, Museum, Archives, Universities, and companies in the territory and in the international ecosystem, and the continuous exchange of cultural assets in order to provide knowledge in new forms depending on the user's requests. High volume of data and several files formats requires specific procedures to ensure the life of data against technological obsolescence.

Digitalization is supported by Digital Asset Management systems that introduce new means of classifying, searching, upload collection of digital objects from sources (scanners, OCR, ...), performing quality checks, verifying the confidentiality / adding rules for protection or rendition before publishing on the digital library access platform. Digital libraries platforms may be viewed as a new form of information institution or as an extension of the services libraries currently provide. Libraries can support a diverse range of activities and develop new relationships across user communities.

The purpose of library's digitization strategy is a broad access and the long-term preservation of materials stored into it. Project plan shall prevent loss or deterioration of historical data of relevance to the user community. The strategy shall consider several issues like target audience, nature of materials, technology to be used, standards and requirements.

In a digitization project shall be consider the reasons for digitizing that can include, for example, improving access to rare and/or unique materials, to protect fragile or heavily used materials, to create educational resources and new target audience. Another essential element is the nature of file's format. It is necessary organize and manage every kind of file format depending on whether they are visual materials, printed text, audio, video or physical objects.

After selecting resources to be digitized, it is necessary to establish that collections have a clearly defined ownership, like rights and permissions for electronic distribution that must be secured or securable. Every object should be in an appropriate format and physical condition for scanning, that can be handled, nondegraded, with enough clarity of detail.

For every collection selected for digitization it is necessary to consider if the materials are in the public domain or if the Library can get permission from the rights holder to make them available online, if they are unique and not already online. Every digital object put online must be followed by information describing the item's copyright status and the provenance. Shall also be determined which are appropriate equipment and skills for creating a good-quality version of materials. In fact, it is necessary to proceed to physical digitization of materials through use of appropriate technology like scanner, software, external hard drive, workstation.

Once established copyright status statement and physical condition of materials, it is necessary to collect and organize the collection through item-level data (e.g. date, creator, title, etc.). Libraries use Archivists' tools to classify the materials according to the applicable standard (e.g. ISAD, ISAAR) and to enrich the descriptive information with relevant information for the users.

Data shall be used to locate or manage information resources by abstracting or classifying those resources or by capturing information related to them. In general, is good practice organize data into distinct categories and relies on conventions to establish the values for each category.

To improve the experience of the user communities, during the data discovery and exploitation the Libraries add descriptive information and representation information that facilitate the understanding of collections over time.

The digitalization shall also be considering the target audience through the implementation of appropriate new services of use. Digital collection shall expand access to document, enhance services, optimize resources, to enrich user experience. The access must however be based on a system of assignment of permission based on user's role. After the publication, the digital objects are searchable and accessible from the discovery tool integrated in the library Online public access catalogue (OPAC) of from the digital library website.

D5.2 EXAMPLE PHASES AND COLLECTION GROUPS

Phase	Collection Groups			
	Initiating	Planning	Executing	Closing

Digitize documents	Obtain agreement for digitization	Plan digitization activities and archival of resulting digital information.	Carry out the digitization. Collect Additional Information including details of method of digitization and equipment used.	Complete digitization
Archive digital documents	Initiated as part of the overall digitization project.	Agree Submission Agreements	Transfer the digital images to the Archive and Preserve them. Collect Additional Information including that needed to create the AIPs, definition of the Designated Community, Transformational Information Properties.	Continue preserving the information.
Make information available	Initiated as part of the overall digitization project.	Plan access using International Image Interoperability Framework (IIIF)	Implement the framework which consists in a set of APIs for authentication, searching, viewing, comparing and annotating digital objects. IIIF enables an effective exchange of books, maps, manuscripts, archival material, musical scores, ancient newspapers images between different user communities around the world. Open IIIF Viewers are promoted by the community for a shared user experience. Collect Additional Information including details of IIIF, software implementations and search parameters.	

The LTDP Preserved Data Set Content (PDSC)[1]

[1]

ANNEX E

SECURITY CONSIDERATIONS

(INFORMATIVE)

E1 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.

E2 SECURITY CONCERNS WITH RESPECT TO THE CCSDS DOCUMENT

This document provides guidance on Additional Information to be collected.

E2.1 DATA PRIVACY

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

E2.2 DATA INTEGRITY

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

E2.3 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.

E2.4 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.

E2.5 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.

E2.6 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.

E3 POTENTIAL THREATS AND ATTACK SCENARIOS

Inappropriate access and/or changes to the Additional Information collected, physical or virtual attacks on data stores.

E4 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.