Consultative Committee for Space Data Systems

DRAFT RECOMMENDATION FOR SPACE

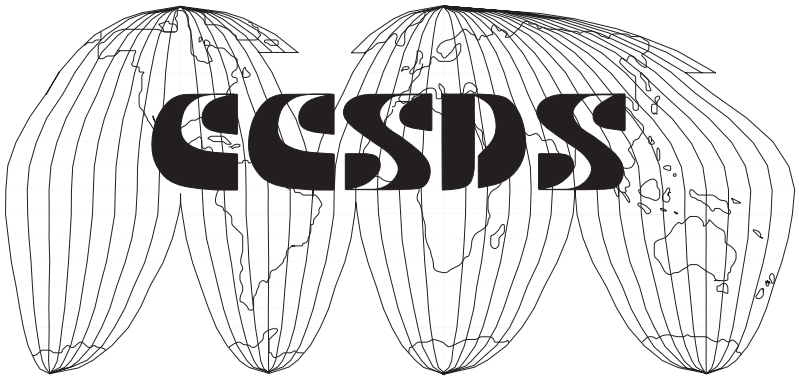
DATA SYSTEM STANDARDS

Producer-Archive Interface Methodology Abstract Standard

CCSDS –651.0-W-2

WHITE BOOK

June 15, 2002



This document has been prepared in the framework of the CCSDS/ISO activities.

For any comment, please contact

claude.huc@cnes.fr, daniele.boucon@cnes.fr

CONTENTS

CONTENTS (CONTINUED)

Figure

Table

1. INTRODUCTION

1.1 PURPOSE AND SCOPE

The object of this document is to regulate the relationships and interactions between an information Producer and an Archive. It defines the methodology to allow all the actions to be structured that are required to apply from the time of contact being made between the Producer and the Archive until the objects of information are received by the Archive. These actions imply the first stage of the Ingest Process as defined in the Reference Model OAIS [1]. It describes part of the functional entities: Administration ("Negotiate Submission Agreement") and Ingest ("Receive Submission" and "Quality Assurance").

This methodology document:

- Identifies the different phases in the process of transferring information between a Producer and an Archive,
- Defines the objective of each of these phases, the actions that must be carried out during these phases and the expected results (administrative, technical, contractual, etc.) at the end of a phase,
- Forms a general methodological framework, which should be able to be applied and reused in those processes that relate to the Producer-OAIS Archive interface. This general framework should also provide sufficient flexibility for each particular case,
- Forms a basis for the identification and/or development of standards and implementation guides, in the domain in question,
- Forms a basis for identification and/or development of a set of software tools that will assist the development, operation and checking of the different stages in the process of information transfer between the Producer and the Archive.

The term "Archive" means an Archive that is in compliance with the OAIS Reference Model. The vocabulary used is in accordance with the glossary defined in the Reference Model.

1.2 APPLICABILITY

The methodology defined in this document applies both to the information Producer and to the Archives to which this information must be transmitted: such Archives being conformant to the Reference Model.

This methodology could also be of interest and be fully or partially applied to Archives that are not conformant to the Reference Model.

This methodology is of interest for archives in which the holdings are physical as well as digital objects.

1.3 RATIONALE

Relationships between Archives and the Producers are rarely simple and easy. There are serious difficulties with the management of the Producer-Archive Interface in all the contexts which have been analyzed in preparation of this document (traditional archives, libraries, Scientific Data Centers, business archives, etc.).

These difficulties generally lead to an additional workload and may have negative consequences on the quality of the archived information. They can create a difficult relationship between the Archive and the Producer.

These potential problems are of several kinds, and include the following:

1. The digital objects received do not conform to what the Archive expects.

2. What the Producer delivers has not been clearly defined.

3. The ingestion schedule is not fulfilled by the Producer.

4. Errors in transfers are detected late by the Archive, or are not detected prior to use.

Within this context, the development of a standard methodology in this domain should aid in reducing such problems.

1.4 CONFORMANCE

This standard provides an abstract general methodology framework applicable to the interface between the Archive and the Producer. The aim is to create a 'Submission Agreement' then to transfer Submission Information Packages (SIPs) to the Archive, and finally to check these Submission Information Packages (SIPs). In order for this standard to be easily applicable to a particular community, a specific standard or "community standard" should be created in order to take into account all the specific features of the community in question.

This community standard will be considered to conform to this abstract standard if:

- All the actions have been considered and incorporated appropriately.

- The methodology for creating the community standard has addressed the various work phases defined in the chapter 'Creating an Archive-Producer interface methodology for a particular community' (Chapter 4).

In the case that this abstract standard is directly used by a Producer and an Archive within the framework of a certain Archive Project, the methodology applied will be deemed as conform to the abstract standard if:

- All the actions have been considered and implemented as appropriate within the context of that project.

1.5 DOCUMENT STRUCTURE

1.5.1 HOW TO READ THIS DOCUMENT

All readers should study the Purpose and Scope (1.1), Applicability (1.2), and Conformance (1.4) sections to obtain a view on the objectives and applicability of the document.

Those who want just an overview of the methodology should also read 'an overall view of the producer-archive interface methodology' (section 2).

Those who will use the methodology should read the entire document.

Good previous knowledge of the concepts and vocabulary defined in the OAIS Reference Model [1] is required in order to understand this document. Nevertheless, annex A is a targeted overview of the OAIS RM [1] dedicated to the Methodology Abstract Standard.

1.5.2 ORGANIZATION BY SECTION

Section 1 defines the purpose, scope, applicability and definition sections. This also provides help in the work undertaken to define this methodology. Finally, it specifies what conformance with this standard consists of.

Section 2 gives a general overview of the methodology, the players involved, their relationships and the activity phases that should be organized to manage an Archive Project.

Section 3 analyses in detail each of the four phases defined in the methodology for all Archive

Projects: The preliminary definition phase, the formal phase, the transfer phase and the validation phase.

Section 4 describes the work stages that enable a methodological community standard to be created in conformance with this abstract standard.

The annexes are not part of the Methodology and are provided for the convenience of the reader:

– annex A is a targeted overview of the OAIS RM [1] dedicated to the Methodology Abstract Standard;

– annex B provides a table showing the correspondence between the preliminary phase and formal phase.

1.6 DEFINITIONS

1.6.1 ACRONYMS AND ABBREVIATIONS

AIP	Archival Information Package
ASCII	American Standard Code for Information Interchange
CCSDS	Consultative Committee for Space Data Systems
DIF	Directory Interchange Format
DTD	Document Type Declaration
EAD	Encoded Archival Description
IEEE	Institute of Electrical and Electronic Engineers
ICA	International Council on Archives
ISO	International Organization for Standardization
OAIS	Open Archival Information System
PDI	Preservation Description Information

PDF	Portable Document Format
SGML	Standard Generalized Markup language
SIP	Submission Information Package
TEI	Text Encoding Initiative
UML	Unified Modeling Language
XML	eXtensible Markup Language

1.6.2 TERMINOLOGY

The terminology used is mainly defined in the OAIS RM. The reader is invited to refer to this document. Only a brief description is given here. This terminology does not seek to replace already existing terminology in the various domains related to archiving. Each domain should be able to apply this methodology while retaining their specific terminology.

Here follows a short glossary of the OAIS terminology indispensable for this document. The definitions printed in italics are related to terms that are not defined in the OAIS glossary

Access: The OAIS entity that contains the services and functions which make the archival information holdings and related services visible to Consumers.

Archival Information Package (AIP): An Information package, consisting of the Content Information and the associated Preservation Description Information (PDI), which is preserved within an OAIS.

Archive: An organization that intends to preserve information for access and use by a Designated Community.

Archive Project: *An Archive Project is the set of activities and means used by the information Provider as well as the Archive for the ingestion of a given set of information into the Archive.*

Consumer: The role played by those persons, or client systems, who interact with OAIS services to find preserved information of interest and to access that information in detail. This can include other OAISs, as well as internal OAIS persons or systems.

Content Data Object: The Data Object, that together with associated Representation Information, is the original target of preservation.

Content Information: The set of information that is the primary target for preservation. It is an Information Object comprised of its Content Data Object and its Representation Information. An example of Content Information could be a single table of numbers representing, and understandable as, temperatures, but excluding the documentation that would explain its history and origin, how it relates to other observations, etc.

Data Dictionary: A formal repository of terms used to describe data.

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

Data Submission Session: A delivered set of media or a single telecommunications session that provides Data to an OAIS. The Data Submission Session format/contents is based on a data model negotiated between the OAIS and the Producer in the Submission Agreement. This data model identifies the logical constructs used by the Producer and how they are represented on each media delivery or in the telecommunication session.

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

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

Ingest: The OAIS entity that contains the services and functions that accept Submission Information Packages from Producers, prepares Archival Information Packages for storage and ensures that Archival Information Packages and their supporting Descriptive Information become established within the OAIS.

Meta-data*: Data about the content, the quality, condition and other characteristics of the data (from FGDC Standards Reference Model, March 1996).*

Packaging Information: The information that is used to bind and identify the components of an Information Package. For example, it may be the ISO 9660 volume and directory information used on a CD-ROM to provide the content of several files containing Content Information and Preservation Description Information.

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, and Context information.

Producer: The role played by those persons or client systems, who provide the information to be preserved. This can include other OAISs or internal OAIS persons or systems

Representation Information: The information that maps a Data Object into more meaningful concepts. An example is the ASCII definition that describes how a sequence of bits (i.e., a Data Object) is mapped into a symbol.

Submission Agreement: The agreement reached between an OAIS and the Producer that specifies a data model for the Data Submission Session. This data model identifies format/contents and the logical constructs used by the Producer and how they are represented on each media delivery or in a telecommunication session.

Submission Information Package (SIP): An Information Package that is delivered by the Producer to the OAIS for use in the construction of one or AIPs.

Transfer*: The act involved in a change of physical custody of SIPs (definition derived from the Dictionary on Archival Terminology of the ICA).*

The terms **'class'**, **'association', and 'aggregation'** refer to UML terminology.

1.7 APPLICABLE REFERENCES

[1] Reference *Model for an Open Archival Information System (OAIS).* Draft Recommendation for Space Data System Standards, CCSDS 650.0-R-2. Red Book. Issue 2. Washington, D.C.: CCSDS, July 2001.

[2] *Data Entity Dictionary Specification Language (DEDSL) Abstract Syntax (CCSD0011).* Recommendation for Space Data System Standards, CCSDS 647.1-B-1. Blue Book. Issue 1. Washington, D.C.: CCSDS, June 2001.

1.8 INFORMATION REFERENCES

[B1] *Unified Modeling Language*. Version 1.1. Cupertino, CA: Rational Software Corporation, September 1, 1997. <http://www.rational.com/uml/resources>.

[B2] *Data Entity Dictionary Specification Language (DEDSL) - XML/DTD Syntax (CCSD0013).* CCSDS 647.3-R-1. Red Book. Issue 1. June 2001.

[B3] *Data Entity Dictionary Specification Language (DEDSL) PVL Syntax (CCSD0012)*. Recommendation for Space Data System Standards, CCSDS 647.2-B-1. Blue Book. Issue 1. June 2001.

2. AN OVERALL VIEW OF THE PRODUCER-ARCHIVE INTERFACE METHODOLOGY

2.1 THOSE INVOLVED AND THEIR RELATIONSHIPS

2.1.1 THE PRODUCER

In conformance with the definition given in the Reference Model, the term 'Producer' designates the persons and systems who supply the Archive with information to be preserved. Due to the fact that the Submission Agreement is entered into between persons and not between systems, it is considered here, that the Producer can always be represented by a person.

The term 'Producer' thus covers a wide variety of situations: the Producer can be an editor, a scientific team, a laboratory, a company department, a Ministry, an administrative body, a private individual, etc.

The Producer's activities can be multiple and varied and they may require the involvement of a whole group of people with different skills and professions.

For the purpose of this methodology, it is assumed that the Producer is represented by a single entity that has the responsibility for all the activities related to a phase, and for each of the phases identified in this methodology.

The Producer has his own management. This management defines the objectives, the responsibilities of the Producer's activity and provides him with the necessary resources. This management may be different from or the same as the Producer. In this document, the Producer and the Producer's management are differentiated and considered to be two different functions even if they are assumed by the same person.

2.1.2 THE ARCHIVE

The Archive is an OAIS Archive. The main responsibility of an Archive is to preserve a whole set of information and to make this available in an intelligible and useable form to a defined Designated Community.

In that context, the term "information" is used with the meaning defined by the OAIS Reference Model [1], section 2.2.1 "Information definition". This definition is also available in section 1.6.2 "Terminology" of this document, and the understanding of the OAIS framework is synthesized in Annex A.

The responsibilities of the Archive (which information to archive, which Designated Community, etc.) are defined by the OAIS Management.

2.1.3 GENERAL FRAMEWORK OF PRODUCER-ARCHIVE RELATIONSHIPS

There are a wide range of relationships and context situations that can exist between a Producer and an OAIS Archive. The following items are some of these:

- They can have the same management: this is the situation in a company, in which a department is entrusted to archive the information produced by the other departments,

- They can have different management, the transfer of data to be archived is, none the less, of an obligatory nature: this is the case for government Archives and Legal Deposit Libraries, whose tasks are defined by regulations or law.

- They can merely be on a voluntary basis when there is no obligation for the Producer to co-operate with the Archive. These Archives are called collecting Archives. Collecting Archives often specialize in one type of records such as the as labor union records, business records, commercial broadcasting records, or immigration records.Labor Archives at Wayne State University or the Urban Archives at Temple University. This is also the case for publications being archived by a public library.

- They can be on a purely contractual basis: this is the case of 'commercial Archives', i.e. companies specialized in archiving and who ensure the preservation of information for other companies.

- (add 2 bullet points calling out ….(1) Producer as separate role in Archive (2) no producer role set up and one relies on the charter for the archive) (incorporate deleted paragraph)

In some cases there is no relationship established between the Archive and the Producer, e.g. this is the case when an institutional library is entrusted to archive all electronic publications (CD-ROM) and due to the great number of editors or to their non co-operation, there is no relationship – and thus no negotiation - between the Producer and the Archive. In this case, the library should decide to create a department, within its own structure, to collect electronic publications to be archived and prepare the SIPs. This department plays the role of a Producer with respect to the Archive department, both departments being in the same library.

2.1.4 NEGOTIATION FOR AN AGREEMENT

The conditions under which negotiation takes place between the Producer and the Archive depend on the nature of the relationship between the Producer and the Archive and whether the archiving is of an obligatory nature or not.

This negotiation can be of an iterative nature. Negotiations should result in a 'Submission Agreement'. This agreement defines in a precise and thorough manner: the different Data Objects which are to be transmitted to the Archive, the means used to transfer this data, the transfer schedule, etc.

Under certain conditions this agreement may be subject to revisions.

In the absence of a relationship between the Producer and the Archive, as discussed previously, there is no negotiation with the actual Producer. For example, the Archive may collect information from certain Web sites. In essence the Archive establishes a virtual Submission Agreement with the actual Producer without any negotiation beyond that involved in conformance to Web protocols. Virtual Submission Agreement is understood in the sense defined in the section 2.3.2 "Producer interaction" of the Reference Model [1].

Whatever the Archive/Producer relationships may be, experience shows that negotiations are easier, when they take placethey are initiated very early on in the information creation process. It is always easier to agree on a data format before, rather than after, such data is produced.

2.2 THE ARCHIVE PROJECT

An Archive Project is the set of activities and means used by the information Provider as well as the OAIS Archive for the ingestion of a given set of information into the Archive.

The agreement between the Producer and the Archive covers the provision by the Producer of a set of information defined in the framework of an Archive Project

Within this set of information, the primary information that must be preserved must be clearly identified by the Archive.

The complementary information, which is necessary for the Archival Information Packages (AIPs) to be made up, could be the following:

- Delivered by the Producer within the context of the Archive Project in question.
- Delivered by the same Producer within the context of the previous Archive Project.
- Delivered by another institution (for standards, for instance).
- Produced by the Archive itself (reference, fixity of AIPs).

Periodic updates of the agreement may be required because additional data is collected, or the scope of data provided has been expanded to include additional area of information. Technological changes or new standards may also imply agreement updates (see also the section "Change management during the life of an Archive Project").

2.3 THE PHASES

2.3.1 GENERAL DESCRIPTION

The Producer-Archive Interactions consist of four different phases:

- The Preliminary Phase, also known as a pre-ingest or pre-accessioning phase, includes the initial contacts between the Producer and the Archive and any resulting feasibility studies, preliminary definition of the scope of the project, a draft of the SIP definition and finally a draft Submission Agreement.
- The Formal Definition Phase includes completing the SIP design with precise definitions of the digital objects to be delivered, completing the Submission Agreement with precise contractual transfer conditions such as restrictions on access and establishing the delivery schedule.
- The Transfer Phase performs the actual Transfer of the SIP from the Producer to the Archive and the preliminary processing of the SIP by the Archive, as it is defined in the agreement.
- The Validation Phase includes the actual validation processing of the SIP by the Archive and any required follow-up action with the Producer. Different systematic or in-depth levels of validation may be defined. Validations may be performed after each delivery, or later, depending on the validation constraints.

Each phase is divided into one sub-phase level. This sub-phase level has to be considered in a chronological order.

Each of these sub-phases is made up of set of actions. In the same set, the actions can be carried out in any order.

While the phases are chronologically different there can be a significant lapse in time between the formal definition phase and the actual transfer phase. Within the archives the transfer phase and the validation phase can take place concurrently if the actual transfer phase occurs over an extended length of time.

2.3.2 RELATIONSHIPS BETWEEN THE PHASES

Figure 1 gives a more precise description of the relationships between these phases. In each text box on the left hand side of the diagram there is a brief indication of the aims of each phase. On the right hand side, the outputs between each phase, are articulated as follows:

- The preliminary phase leads to a summary document that decides on the feasibility of the Archive Project and approves proceeding to the formal phase (or stopping the project).

- This document is the basis on which the formal phase is developed. The formal phase leads to the Submission Agreement being drawn up, which summarizes all the aspects of the formal phase. This agreement refers to a dictionary or a formal model. All of these elements are needed in order to proceed with the transfer phase.

- The outputs of the transfer phase are data objects that are input to the validation phase. As previously mentioned, validation can start before all the elements of the model have been delivered. The transfer and validation phases are often carried out partially in parallel, as there is iteration when all the information to be submitted is not submitted at once.

- The Archive sends the Producer its validation report for the objects received or forms reporting the anomalies found (the Archive may also acknowledge receipt of SIPs after accession, and only notify the Producer if there were anomaly forms or invalid data).

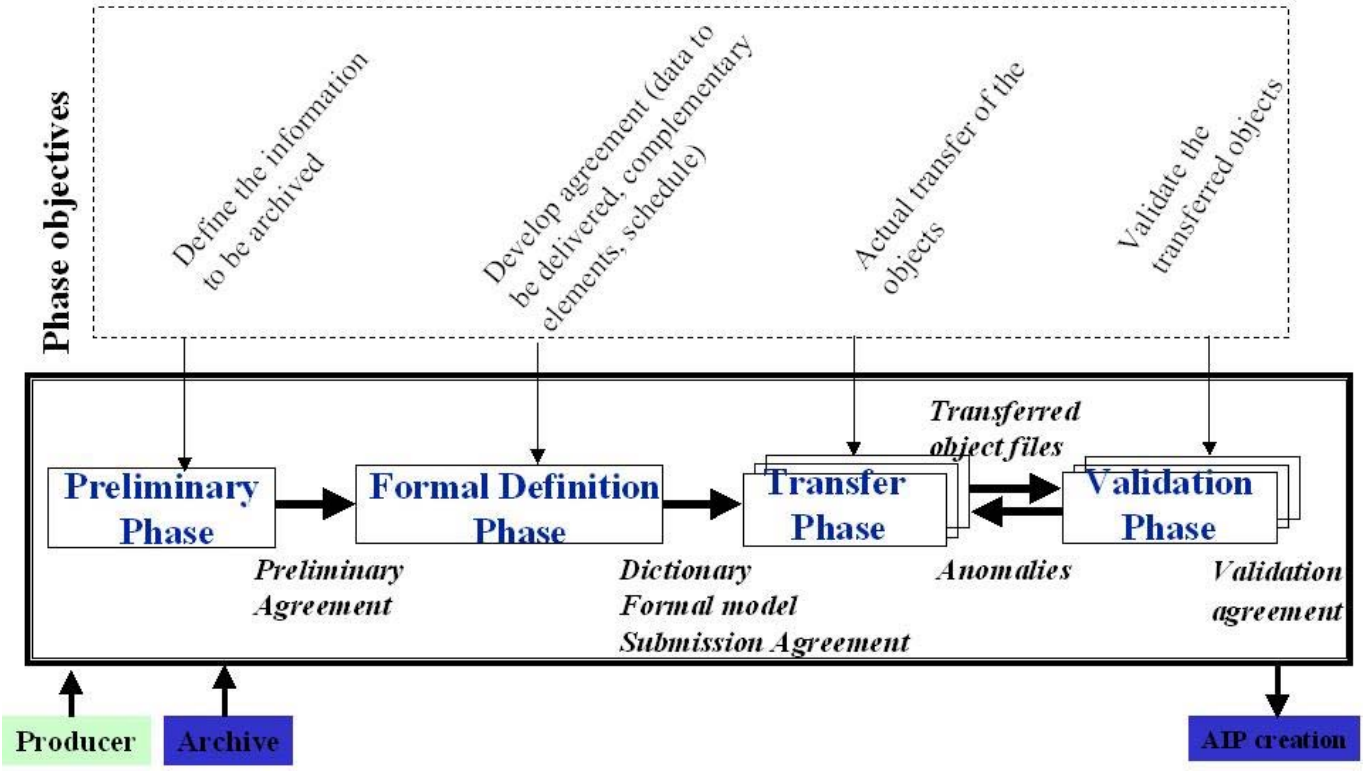


Figure 1: main phase objectives and outputs

3. DETAILED DESCRIPTION OF PHASES

Each of the four phases detailed below is divided into sub-phases. These sub-phases are identified in a Summary-table at the head of each related section. The sub-phases are characterized by the actions listed in the Action-table. These actions may be mandatory, conditional or optional:

- Mandatory (M): always required;
- Conditional (C): required under specified conditions;
- Optional (O): allowed but not required.

An obligation column indicates whether an action is mandatory (M), conditional (C), optional (O).

3.1 PRELIMINARY PHASE

Aims of the phase:

- **To identify the primary information which the Archive must preserve.**
- **To establish a preliminary definition of the different Data Objects that will be transmitted to the Archive by the Producer.**
- **To analyze all aspects of feasibility.**
- **To decide on the feasibility of the operation, as much from the Producer's as from the Archive's point of view.**
- **To make an estimate of the required resources.**
- **In conclusion, to draw up a preliminary Submission Agreement**.

This phase is fundamental. It establishes the foundation on which an Archive Project can be built in the best way. Whenever possible, this preliminary phase **should** be carried out very early on, even before the information to be archived has been produced.

This observation is based on practical experience.

The preliminary phase is itself made up of 3 sub-phases:

- First contact.
- Preliminary definition, feasibility study and assessment of the Archive Project (the different subjects discussed in this phase are listed in the table below. They are detailed in the corresponding section).
- Drawing up of the preliminary agreement.

These are accomplished within the context of the standards, guides and tools available for this phase.

Summary of the preliminary phase	
First contact.	
	Information to be archived.
	Digital objects and standards applied to these objects.
	Quantification.

Preliminary definition, feasibility and assessment	Object references.
	Security conditions.
	Legal and contractual aspects.
	Transfer operations.
	Validation.
	Schedule.
	Permanent impact on the Archive.
	Summary of costs, risks.
	Critical points.
Establishment of a preliminary agreement.	

Summary table 3.1-1: Preliminary phase

3.1.1 FIRST CONTACT

Preliminary phase: first contact	**Involves**
Identify the Contact Persons and work organization	Producer and Archive
Exchange general information	Producer and Archive

Action table 3.1.1-1: Preliminary phase: first contact

The first contact between the OAIS Archive and the Producer can be made on the initiative of the Archive, the Producer, the Archive management or even by an external entity.

- **Identification of the Contact Persons and work organization**

This is the stage to agree in principle on how to proceed with the preliminary phase **in conformity with this methodology** and to identify the main contact person, both on the Producer's side and on the Archive's side. Complementary contact persons for specific questions (technical, administrative, etc.) can also be identified and their role should be defined. These persons may also ask for help from experts depending on the point examined (standards, legal questions, etc.). The list of potential contacts includes appropriate subject matter specialists from the archives.

The organization and work division for this phase should also be defined at this point.

- **Exchange of general information**

  - The Producer provides the Archive with a set of general information that concerns the aim of the information to be preserved, its context, its schedule and its constraints. Producer may also provide information on requirements of Designated Community.

- The Archive presents its role, its general mode of operation, the standards that it generally applies, the tools that may be used in the Producer-Archive Interface, etc.

- The Archive submits the **Producer-Archive methodology** to the Producer:

  - The method, its advantages and constraints (dictionary, models of the data to be delivered).

  - The service aids for the method's application (existing data dictionaries, tools for creating dictionaries, tools for creating formal models, which identify the objects to be delivered, service aids for creating descriptors, etc).

At this point, each of the two partners can supply all possibly useful information (for the project): general documents, reference documents, documentary references, Internet site references, etc.

3.1.2 PRELIMINARY DEFINITION, FEASIBILITY AND ASSESSMENT

This is the focus point of the preliminary phase. It should result in the following:

- Identification of the information to be preserved by the Archive, and description of the main features or properties of that information.

- Establishment of a preliminary definition of the Data Objects, Data sets and sub-sets, description of the main features or properties of those objects, sets and sub-sets, that the Producer is expected to transmit to the Archive.

- Each party will have drawn up a feasibility report of the project from their own point of view. Feasibility covers all the aspects that could put the project in question: technical, financial, legal, etc.

- Both archive and producer create an assessment of the project cost.

- All the elements required to draw up a Submission Agreement have been collected.

The text below deals with a whole group of subjects that should or must be analyzed as part of the preliminary phase. The depth of the analysis needed to reach the goal is not, a priori, defined. This depends on the context, the information to be archived and those involved. Definition of the required depth of analysis point by point is thus the responsibility of the Producer and the Archive.

In the paragraphs below the subjects are approached in the form of actions to be carried out, by the Archive, the Producer or both parties depending on the context. There is often interdependence between these subjects.

Most of these subjects can be approached and treated at the same time, e.g. information and standards, while respecting the dependency (e.g. digital objects must be identified before considering quantification).

The Producer and Archive should ask the following questions for each of the subjects examined:

- Does the subject concern the Archive Project?

- What level of definition should be reached in the preliminary phase?

- Is the subject critical for the Archive Project?

Some subjects can be completely covered in this phase, whereas other subjects should be further developed in the formal definition phase (these should be specified and noted in the summary document).

3.1.2.1 INFORMATION TO BE ARCHIVED

Preliminary phase: information to be archived	**Involves**
Identify the Content Information to be preserved and clearly define the limits of the Archive Project	Producer and Archive
Identify the complementary information to be preserved: Representation Information, PDI. Check that they exist or it is planned to create them	Producer and Archive
Identify the Designated Community	Producer and Archive
Define user access to this information	Producer and/or Archive
Assess the planned duration of the archiving of this data	Producer and/or Archive
Assess the feasibility and cost induced by the previous points	Producer and/or Archive

Action table 3.1.2-1: Preliminary phase: information to be archived

The following points are inter-dependent.

- Identify the information to be preserved: this is the primary starting point and it is important at this stage to clearly define and delimit the information which constitutes the primary object of the Archive Project. If there are still some open options, this is the time to make these explicit. The preliminary phase cannot be completed until this has been accomplished.

- Identify the complementary information to archive: Representation Information, Preservation Description Information (PDI). Draw up an inventory of the available data and information and those which must be created and possibly establish priority levels for these elements.

Example of complementary information in a space mission

A space mission is composed of experiments, an experiment produces data sets (main, auxiliary data, images), a data set is a set of homogeneous files.

- ❖ Descriptive files for the mission and the experiments are the PDI (context, source -the laboratory, the reference -file names). The data sets (and its data files) are the Data Objects. A data set is described by a DIF file (Directory Interchange Format).

- ❖ The data files are described by Representation Information:

  * An EAST (ISO language for data description) structure file, giving the exact structure bit per bit of the data files (syntactic representation).

  * DED file (Data Entity Dictionary): describes the semantics of the data files. The DED file is a sample of the Concept Data Entity Dictionary.

- Identify the Designated Community: how and by whom the data will be used, e.g. whether for the general public or for researchers. This point has an influence on the required level of information (high or low) and thus on the previous point and on access (e.g., research by key word, by author, by time-related or geographic criteria) and thus on the following point. However, it should be noted that for some institutional and/or governmental Archives, the Producer has no precise idea of how the documents archived by historians will be used. Even with scientific observation archives, 10 years after data production, scientific data is used in ways that the Producers could not even imagine.

- Define user access (also see section "Security conditions" of the preliminary phase):
  - ❖ E.g., unrestricted or limited access.
  - ❖ Free or paid access.
  - ❖ Availability and access authorization over time (retaining time before being made available).
  - ❖ Required service level: speed, performance, type of access (interactive server, data transfer by network or on a digital support, etc.), typical selection criteria and requested volumes of data dissemination expected, research aids, etc.
  - ❖ This all leads to a pre-identification of the Descriptive Information.
- Estimate the lifetime of the Archive.
- Assess the cost induced by the previous points. If this cost reveals clear non-feasibility, stop the work at this stage and possibly restart on a new basis. This remark is valid for the other subjects examined during the preliminary phase, as follows.

3.1.2.2 DIGITAL OBJECTS AND STANDARDS APPLIED TO THESE OBJECTS

Preliminary phase: digital objects and standards applied to these objects	**Involves**
Make a preliminary identification of the Data Objects related to the different categories of information to be archived	Producer and Archive
Define the rules and standards related to these objects that are accepted by the Archive	Archive
Describe the aid tools for the application of the rules and standards known by the Archive	Archive
Provide the rules and standards applied to Data Objects by the Producer	Producer
Describe the help tools for application of the rules and standards known by the Producer	Producer
Assess the compatibility and study solutions	Producer and Archive
Assess the efforts to be made and the costs	Producer and Archive

Action table 3.1.2-2: Preliminary phase: digital objects and standards applied to these objects

- Preliminary identification of these Data Objects enables a first list of object categories to be drawn up: Content Data Objects (which contain the primary information to be preserved), Data Objects containing Representation Information on the primary Data Objects, Data Objects describing the context and source of the primary information, etc.

For each of these object categories, priority being given to the Content Data Objects and their associated Representation Information, the Archive and Producer should attempt to reach an agreement on what the Producer will create and what the Archive will receive.

- The following paragraphs cover actions concerning discussion of rules, standards and tools:

❖ Standards applicable to Content Data Objects: data files in ASCII or binary, the form of which is defined by a specific application, particular standards applicable to the geographic representation of information or the representation of time and dates, standards related to a profession, sound, image, video files, SGML or XML files conforming to a DTD or a predefined schema, PDF files, etc.

❖ Standards applicable to Data Objects containing the Representation Information of Content Information: simple reference to a standard that should also be archived or use of a syntactic description of data language (e.g. EAST), semantic description language (SGML, XML), etc.

❖ Standards applicable to meta-data levels: ISO/TC211 standards for the description of geographic data, MARC for libraries, DIF for scientific data, DTD EAD for the archivists, etc.

If the standards accepted by the Archive do not correspond to those used by the Producer, it is possible that the availability of aid tools for the use of these standards could enable the partners to find common ground. Possible solutions should be analyzed in terms of technical feasibility and cost. If the objects already exist, what are the necessary migration efforts? Otherwise, what would the creation effort be, to satisfy the requirements?

- Assess the compatibility between the rules, standards and tools already in place and those that should be used. Carry out a study of the possible solutions.

- Deduce from the previous study what resources must be deployed and the relevant costs.

3.1.2.3 OBJECT REFERENCES

Preliminary phase: object references	**Involves**
Draw up an inventory of the information on the existing identification rules or nomenclature within the domain, legal provisions, and standards.	Producer and Archive
Define the rules that could or should be applied within the context of the Archive Project	Producer and Archive
Assess the associated costs	Producer

Action table 3.1.2-3: Preliminary phase: object references

- The Archive provides the Producer with information on:

  ❖ The existing identification rules or nomenclature, (bibliographic description, namespaces, etc.).

  ❖ Any possibly imposed legal provisions.

  ❖ The standards used.

- The Producer and Archive negotiate the pertinent rules to be applied to the Archive Project.

- The Producer evaluates the cost of these constraints.

3.1.2.4 QUANTIFICATION

Preliminary phase: quantification	**Involves**
Estimate the data volume to be transmitted to the Archive	Producer
Assess the permanent data volume to store	Archive
Assess the storage capability need for the ingest process	Archive
Assess the associated costs	Archive

Action table 3.1.2-4: Preliminary phase: quantification

- The Producer must estimate the volumes to be transmitted in the short, medium and long term (global volume, minimum, average and maximum planned size of files, number of files), as well as the frequency of the transfer sessions. These elements have an influence on the technique used for the transfer.

- The Archive must estimate the permanent global data volume to store with the above elements provided by the Producer. This estimate implies an associated cost for the Archive. This cost is evaluated in the section "Permanent impact on the Archive".

- The Archive must assess the storage need for the ingest process (data storage before transformation to AIP and transfer to OAIS storage function).

It should be noted that this point is not independent from the choices made for the standards applicable to transmitted Data Objects: for the Data Objects containing scientific observations it has frequently been noticed that there is a factor 2 between the volume of data coded in the form of IEEE floating numbers and the volume of the same data coded in ASCII. In much the same way, the size of a file structured in XML can be much larger than the same file in pure text.

- The Archive must assess the cost associated with the storage needs.

3.1.2.5 SECURITY CONDITIONS

Preliminary phase: security conditions	**Involves**
Identify the requirements for confidentiality of the information and for authentication of the source of the information in the transfer between the Producer and the Archive.	Producer and Archive
Identify the requirements for security of the holdings at the archives including storage vaults, limiting physical access, separation of master and copy, etc.	Archive
Identify the requirements for confidentiality of the information and for authentication of the source of the information in the transfer between the Archive and the Consumer.	Producer and Archive
Identify the standards and tools that could be used.	Producer and Archive

Assess the associated costs.	Producer and Archive

Action table 3.1.2-5: Preliminary phase: security conditions

- Confidentiality and Authenticity (Producer/Archive):
  - ❖ Confidentiality of the information in the transfer between the Producer and the Archive: this means that the Archive and Producer must implement the required measures such as the following: encryption of the information held by the Producer, using secure transfer techniques.
  - ❖ Authenticity of the information in the transfer between the Producer and the Archive: this may imply the establishment of encoding and signature mechanisms – at a digital object transmission level – in order to guarantee this authenticity. In-depth validation and particular attention to documentation are also important aspects.
- Implementation of specific measures for security of the holdings may be required by the Archive:
  - ❖ Specific storage measures for the archive.
  - ❖ This also implies the subsequent establishment of techniques to guarantee the integrity of preserved objects. Documenting the preservation process and maintaining an untouched set of the data in archival storage are also important aspects.
  - ❖ The Archive must take into account the change of technology in the long term.
- Confidentiality of the information in the transfer between the Archive and the Consumer and authenticity of the information in the transfer between the Archive and the Consumer: see the impact on the interface Archive/Producer in the first bullet (same paragraph). Furthermore, numerous Consumers on different sites may access the same Archive. This could mean an impact on the techniques used.
- For each subject examined, the following should be made explicit: identification of the applicable regulations, specification of standards and tools that could be used.
- Assessment of the associated costs to cover these aspects.

3.1.2.6 LEGAL AND CONTRACTUAL ASPECTS

Preliminary phase: legal and contractual aspects	**Involves**
Define the nature of the relationships between the Archive and the Producer.	Producer and Archive
Assess the problem of intellectual property.	Producer and Archive
Define the conditions for access to data	Producer and Archive
Archive Certification	Archive
Provide the standards and tools used.	Producer and Archive
Assess the associated costs.	Producer and Archive

Action table 3.1.2-6: Preliminary phase: legal and contractual aspects

In this paragraph, all the aspects that involve legal consideration are looked at. These aspects depend to a large extent on the nature of the relationships between the Archive and the Producer that should thus be made explicit.

- The Archive and the Producer should first define the nature of their relationship. They should examine and answer the following questions:
  - ❖ Does the Archive Project enter into the context of an obligatory government archiving? What are the consequences of this aspect of the project?
  - ❖ If the relationship between the Archive and the Producer are of a contractual type, what is the aim of the contract and how are the responsibilities for the Archive defined within this contract?
  - ❖ What are the specific responsibilities implied by their relationships ?
- Is the data to be archived subject to intellectual property rights? What are the consequences for the Archive? The Archive must, of course, already be familiar, or become familiar, with the national or international legislation on copyrights. Does the transfer of data between the Producer and the Archive imply a transfer of these rights?
  - ❖ If so, what documents should be provided in order to legalize this transfer?
  - ❖ If not, what obligations does the Archive have with respect to this data?

  In negotiating intellectual property rights the archive should distinguish between preservation and access.

  It may be necessary to secure an agreement to preserve, although no-one will be granted access.

  This may be the only way to prevent loss of historically important material,

  as the original medium and technology are unlikely to survive long enough for copyright expiry -- currently 50 years for digital material.
- What obligation does the Archive have with respect to information protection and access to this information? Define the rules which govern these conditions (authorized persons, immediate access or authorized after a legal lapse of time, etc.).
- The different issues brought up here may also imply that the Archive should be certified with respect to an Archive certification baseline, if this in fact exists.
- For each subject examined, the following should be made explicit: identification of the applicable regulations, specification of the standards and tools that could be used.
- Assessment of the associated costs to cover these aspects.

These aspects should be included in the Submission Agreement.

3.1.2.7 TRANSFER OPERATIONS

Preliminary phase: transfer operations	**Involves**

Make a preliminary definition of the Submission Information Packages (SIPs)	Producer and Archive
Producer and Archive exchange the requirements and constraints with respect to the transfer of Data Objects. Identify possible solutions.	Producer and Archive
Assess costs	Producer and Archive

Action table 3.1.2-7: Preliminary phase: transfer operations

- The Producer and the Archive should together study the possible solutions as regards Submission Information Package (SIP). More precisely, it is important to study the Packaging of the different Data Objects for their transmission to the Archive.

- The Producer and Archive exchange their transfer constraints and requirements: network or support (e.g. Compact Disc). Identification of communication protocol and the tools which could be used (e.g. ftp, http) and adapted (depending on the frequency and volumes). It could be necessary to envisage an automated transfer, a securitized transfer for which the required level of security should be defined (also see section "Legal and contractual aspects" of the "Preliminary phase"), etc. Identification of the possible solution(s).

- Cost related to these operations.

3.1.2.8 VALIDATION

Preliminary phase: validation	**Involves**
Supply the Producer with information on the SIP validation procedures, the reject procedures and the tools that are applied by the Archive	Archive
Study the development or modification of the validation tools required	Archive
Study the implementation of quality methods (and tools) to answer needs	Producer
Assess costs	Producer and Archive

Action table 3.1.2-8: Preliminary phase: validation

- The Archive provides the Producer with a certain amount of general information:

  - ❖ On the validation procedures for the SIPs that it uses. It is important to distinguish, on the one hand, the validation methods for the reception of an SIP/ conformity to the model and on the other hand, the validation methods that concern the content of SIP objects.

  - ❖ On the reject procedures in the event of an anomaly.

  - ❖ On the validation tools. Some of these tools may be supplied to the Producer for validation at this end before transfer.

- The Archive may need to modify existing tools, or even to develop some in order to adapt to the context.

- The Producer makes an independent study of the actions to be considered in order to fulfil the quality and validation requirements of the Archive.

- The Producer assesses the cost associated to these actions.

3.1.2.9 SCHEDULE

Preliminary phase: schedule	**Involves**
Define a preliminary schedule	Producer and Archive

Action table 3.1.2-9: Preliminary phase: schedule

- The Archive and the Producer must negotiate a preliminary schedule: data production, transfer, validation, data archiving and data availability for the Designated Community.

3.1.2.10 PERMANENT IMPACT ON THE ARCHIVE

Preliminary phase: permanent impact on the Archive	**Involves**
Assess the permanent impact and the associated costs on the archive.	Archive

Action table 3.1.2-10: Preliminary phase: permanent impact on the Archive

These points are the Archive's responsibility. They concern an assessment of any possible future impact on archiving the data in question, beyond the ingest operation time. This impact and the associated costs take in account:

- The permanent data volume to store, which is estimated in the section "Quantification" of the preliminary phase. This volume may imply an increase in the number of storage Archive volumes, or changes in the support type and an associated cost.

- Take into consideration the necessary long-term migration (support renewing, duplication, re-packaging, transformation of information). Long-term migration should also include plans for transfer of information to another archive in the case of closure of the archive.

- Establishment of specific precautionary measures to avoid the loss of data (destruction, alteration, etc.), e.g. copying to another Archive. In the event of loss or alteration of data, the Archive will inform the Archive Management, the Producer (if it is still available) and the Designated Community, of any measures taken.

- The security requirements (also see section "Security conditions" of the preliminary phase).

It is important that the Archive defines and maintains a cost model to be able to estimate the cost of the archive, within a context when the speed of technological changes is not known in advance.

3.1.2.11 SUMMARY OF COSTS, RISKS

Preliminary phase: summary of costs, risks	**Involves**
Carry out a cost summary, estimate risks	Producer and Archive

Action table 3.1.2-11: Preliminary phase: summary of costs, risks

- Producer and Archive should make a summary of the different costs assessed in the above paragraphs on a short, medium and long term basis: each side should assess the costs that may be implied for them. The following aspects should be taken into account:
  - ❖ Possible changes either on the side of the Producer or Archive, which would require new investment in the end (new data collection, technical changes, etc.).
  - ❖ Available resources and means (human and material).
  - ❖ Risks on either the side of the Archive or Producer.
  - ❖ Available budgets (possibly readjust them)

This summary could lead to numerous negotiations that in turn could lead to an agreement on both sides.

3.1.2.12 CRITICAL POINTS

Preliminary phase: critical points	**Involves**
Assess the critical points.	Producer and Archive

Action table 3.1.2-12: Preliminary phase: critical points

- The Producer and Archive must assess, from among all the points that have already been raised, which ones may cause serious problems and could imply a complete or partial failure risk for the Archive Project.

3.1.3 ESTABLISHMENT OF A PRELIMINARY AGREEMENT

Preliminary phase: establishment of a preliminary agreement	**Involves**
Draw up a synthesis document that summarizes the preliminary phase, with a feasibility assessment and an agreement on proceeding with the formal phase (or stopping it).	Producer and/or Archive
Take the decision to proceed to the next phase	Producer and/or Archive

Action table 3.1.3-1: Preliminary phase: establishment of a preliminary agreement

- **Summary of the preliminary phase**

  - This is the conclusive phase of the preliminary study examined above. The end of this phase is the approval of the synthesis document by the Producer and the Archive. How the drafting of the document is divided up must be decided between the two parties.

    This document provides a basis on which the feasibility of the project can be decided and also shows which are the critical points of the project:

    - ❖ Evident non-feasibility of the project: stopping the project or a search for solutions (financing, etc.).

    - ❖ Evident feasibility that leads to a preliminary agreement. At this stage, this is not the final Submission Agreement (which is made at the end of the formal phase), but a preliminary agreement to proceed with the next phase, which is the formal phase.

    This agreement contains the first elements:

    - ❖ The SIP content (Content Information, PDI, Packaging Information, Descriptive Information), and the data model,

    - ❖ A first submission timetable.

    - ❖ Data access restrictions.

    - ❖ Validation procedures.

    - ❖ Revision, and renegotiation clauses.

- **Decision to proceed to the next phase**

  Depending on the context, this decision is taken in a more or less formal manner. In some cases, this decision is signified by a simple end-of-preliminary-phase report. In other situations, there must be an official decision taken, on the one hand, by the Producer and by the Archive, on the other.

3.2 FORMAL DEFINITION PHASE

Objective: the negotiation of the "Submission Agreement" that includes a complete and precise definition of:

- **The data to be delivered to the Archive by the Producer.**
- **The contractual and legal aspects.**
- **The complementary elements required to define the transfer and validation process.**
- **The schedule.**

The formal definition phase is itself made up of 3 sub-phases:

- Setting up of the organization.
- Formal definition (the different subjects discussed in this phase are listed in the Summary table below. They are detailed in the corresponding section).
- Drawing-up of the Submission Agreement and it being signed by the Producer and the Archive.

This is accomplished in the context of the standards, guides and tools available for this phase.

The above subjects are dealt with in a more precise way in the following paragraphs in the form of lists of actions to be carried out. They require negotiation between the Archive and the Producer. Most of these subjects can be examined and dealt with at the same time, while respecting the inter-dependencies (e.g., the information must be identified before creating the dictionary).

<table>
<tr><th colspan="3">Summary of the formal definition phase</th></tr>
<tr><td colspan="3">Setting up of the organization.</td></tr>
<tr><td rowspan="6">Formal definition</td><td rowspan="3">Information to be preserved and Model of Data Objects to be delivered.</td><td>General project context and definition of Information Objects</td></tr>
<tr><td>Creation of a dictionary</td></tr>
<tr><td>Construction of a formal model</td></tr>
<tr><td colspan="2">Formalization of contractual and legal aspects</td></tr>
<tr><td colspan="2">Definition of transfer conditions.</td></tr>
<tr><td colspan="2">Validation definition.</td></tr>
<tr><td colspan="2">Delivery schedule.</td></tr>
<tr><td colspan="2">Change management.</td></tr>
</table>

	Feasibility and assessment.
Submission Agreement.	

Summary table 3.2-1: Formal definition phase

The table in Annex B shows the relation between the stages of the preliminary phase and those of the formal definition phase.

The actions identified in the preliminary phase are treated in a formal way in this phase. Certain sections of the formal definition phase are new. The following should be taken into account:

- The section “Quantification” of the preliminary phase broaches numerous aspects which are partly drawn up in the sections “General project context and definition of Information Objects”, “Definition of transfer conditions” and “Feasibility and assessment” (regarding the cost aspect).

- The actions in the section “Permanent impact on the Archive” should be reassessed regarding their costs in the section “ Feasibility and assessment”.

- The critical points have no direct relation with any section of the formal phase, as the different points identified by the partners in the preliminary phase must be dealt with separately in the section related to the formal phase.

3.2.1 SETTING UP THE ORGANIZATION

Formal definition phase: setting up the organization	**Involves**
Setting up of the formal phase organization.	Producer and/or Archive
Specify the points previously raised and to be made explicit in the formal phase.	Producer and Archive

Action table 3.2.1-1: Formal definition phase: setting up the organization

- The Archive and the Producer must negotiate the organization of the formal phase, as well as the definition of their individual roles and responsibilities:
  - ❖ Plan the different archiving stages (production, transfer, ingestion), the key points and technical approval (validation: planning of the validation phase),
  - ❖ Define the documents to be produced and who is responsible for them.
- The Archive and the Producer must specify the points in the preliminary phase that need to be examined in greater depth.

3.2.2 FORMAL DEFINITION

3.2.2.1 INFORMATION TO BE PRESERVED AND MODEL OF DATA OBJECTS TO BE DELIVERED

This concerns the precise definition of the information to be transferred from the Producer to the Archive. This definition is a formal Model of objects to be delivered. This Model contains a definition of the objects to be delivered that is as precise and non-ambiguous as possible.

3 main work stages are required to create this model:

1. Description of the general objectives and project context, definition of all the Information Objects, definition of the coding, format, Information Object identifiers, in the form of a text document. All of these points have already been studied in the preliminary phase.

2. Definition of the object classes associated with the aforementioned Information Objects, and creation of an associated dictionary to list these definitions.

3. Construction of the formal model of the Archive Project.

It may possibly be necessary to add an examination of the question of the legal and contractual form of the project to these three work stages.

3.2.2.1.1 General project context and definition of Information Objects

Formal definition phase: General project context and definition of Information Objects	**Involves**
Define the general project context as well as the list and contents of the information elements to be delivered.	Producer and Archive
Define the formats, coding rules, standards to be applied for the objects to be delivered.	Producer and Archive
Define the volume indicators.	Producer
Define the references for the objects to be delivered.	Producer and Archive
Choose the tools on the Producer's side.	Producer and Archive
Write a description of the Information Objects referring to a dictionary and a model (part of the final agreement).	Producer and/or Archive

Action table 3.2.2-1: Formal definition phase: General project context and definition of Information Objects

- At this stage the Producer and Archive must agree on all the information elements to be archived and on the contents:

  - Content Information: Data Object and Representation Information (syntactic and semantic).

- ❖ Preservation Description Information (provenance, context, reference, fixity).
- ❖ Descriptive Information.

It is presumed that the Designated Community and the access have already been identified during the preliminary phase. This has an impact on the level of complementary information to be archived with the Data Objects as well as the Descriptive Information.

The Producer and the Archive must agree on the contents of the documentary elements. Several levels can be established, e.g. a standard document model (with a table of contents model), or specifications that define the elements that must be present in a compulsory manner, the recommended elements, the optional elements, which also provides the quality level for the elements archived.

- They must then choose the format, the coding rules, the standards to be applied, for each of the above-mentioned defined objects, drawing on the elements already provided during the preliminary phase. Some objects already exist, others do not. If the format of existing objects does not correspond to the specified format, the Producer and the Archive must reach an agreement (e.g. migrations).
- The Producer provides the Archive with information on the volume measurements (global volume to archive and also granular information on the volume of Content Data, mean and maximum volume of a file, etc.).
- Producer and Archive define the references of the information elements, drawing on the results of the preliminary phase.
- Producer and Archive define the tools to be installed by the Producer (aid with data production, aid with the production of descriptors, aid with document production, etc.).
- Writing of a description of the elements previously negotiated by the Archive or the Producer. This description will be part of the final folder of the Submission Agreement. This description refers to the dictionary and the formal model (defined below).

Note: The Packaging Information is defined in the transfer stage (see section "Definition of transfer conditions" in the "Formal definition phase").

3.2.2.1.2 Creation of a dictionary

Formal definition phase: creation of a dictionary	**Involves**
Define the object classes and their attributes, set up the associated dictionary.	Producer and Archive
Code the dictionary.	Producer or Archive

Action table 3.2.2-2: Formal definition phase: creation of a dictionary

- From the information already provided, the Producer and the Archive define the classes of the objects associated with all the defined information and their attributes. These classes could be subject to change (see section "change management").
- The complete, formal and precise definition of the different classes of Data Objects to be delivered, constitute the project dictionary. This dictionary is conform to document [2]. Its

implementation could conform to the information documents [B2] or [B3] or be subject to specific implementation.

It is recommended to draw on already existing dictionaries.

3.2.2.1.3 Construction of a formal model

Formal definition phase: construction of a formal model	**Involves**
Define the model of the data to be delivered.	Producer and Archive
Draw up a Model representation, completed if necessary by a text document.	Producer or Archive

Action table 3.2.2-3: Formal definition phase: construction of a formal model

- The formal model identifies the different instances of Data Objects that will be delivered. This model defines the nature of the relationships between these different instances. It also provides a logical and coherent overall view of the whole set of objects. The model must be created depending on the transfer possibilities (objects delivered in a separate manner or not). The granularity of the model will enable definition of the Data Objects or set of Data Objects that may be delivered independently. This data or set of Data Objects are the basis for the definition of the SIPs. There is no single, unique model; moreover, this model may be subject to change (section "change management").

- It is recommended to define this model using a UML type of representation, e.g. (see [B1]). A text document may accompany the model, if this is useful, particularly for complex models.

3.2.2.2 FORMALIZATION OF CONTRACTUAL AND LEGAL ASPECTS

Formal definition phase: formalization of contractual and legal aspects	**Involves**
Drawing up legal and contractual agreements between the Archive and the Producer concerning the data (part of the final agreement).	Producer and Archive

Action table 3.2.2-4: Formal definition phase: formalization of contractual and legal aspects

- This concerns formalizing all the points already stated in the preliminary phase and reaching an agreement on this matter by the Archive and the Producer. In particular, if a transfer of intellectual property must take place, the conditions and the date of this transfer must be defined at this level.

3.2.2.3 DEFINITION OF TRANSFER CONDITIONS

Formal definition phase: definition of transfer conditions	**Involves**
Define the communication procedures (digital network, protocols, supports, etc.).	Producer and Archive
Define the Packaging Information of delivered objects (in what form the data is delivered).	Producer and Archive
Define a transfer session (functional and time-related structure of the transfer of digital objects).	Producer and Archive
Define the initial transfer test.	Producer and Archive
Identify the tools that may be used during the transfer phase.	Producer and Archive
Write a description of the transfer procedures (from the above 4 elements).	Producer and/or Archive

Action table 3.2.2-5: Formal definition phase: definition of transfer conditions

The first 5 points are developed in the following paragraphs.

The last point is the writing of a description of the transfer procedures defined between the Archive and the Producer. This description will be part of the final folder of the Submission Agreement.

3.2.2.3.1 Communication procedures

The Archive and the Producer must precisely define the communication procedure – type of transfer and type of support used for the transfer of objects -, drawing on the elements in the preliminary phase, and taking account elements which have an impact on the calibration of transfer and reception operations, such as data volume and frequency, maximum number of objects delivered by session, maximum and mean object size. The volume of the data delivered by session has firstly been estimated in the section "Quantification" of the preliminary phase.

Several scenarios may occur for the transfer of data from the Producer to the Archive: transfer via a physical media, via a network, sometimes the Archive fetches data from a predefined site. The communication procedures may involve the particular means used in order to ensure the security conditions identified in the preliminary phase (see section "Security conditions" of the "Preliminary phase"): authenticity, integrity and/or confidentiality of the data.

3.2.2.3.2 Packaging

The Archive and the Producer must agree on the technical choices concerning Packaging Information and already looked at in the preliminary phase.

Producer and Archive must define how the objects or set of Data Objects of the formal model will be packaged. E.g. a set of attributes about a data file might be expressed using XML and be combined with the data file bytes using a standard packaging approach such as ISO 12175 (1999).

3.2.2.3.3 Data Submission Session

The actual transfer of Data Objects is divided into successive sessions. The time notion of sequence structures the data transfer in successive stages. This is a logical concept regardless of the physical resources used.

A Submission Session is a term defined in OAIS. It is an operation that enables data transfer from the Producer to the Archive to be carried out. A transfer session thus corresponds to the set of objects that are delivered:

- By transmission on a private or public (Internet) network, by ftp, E-mail, http, etc.
- By delivering a package of one or more physical media.

The Archive and the Producer must define:

- On the one hand, the functional structure of a session. A session may be a homogeneous package of objects (set of documentation, file packet of scientific data, etc.), or a retransfer of data following non-conformities or an update.
- On the other hand, the structure with respect to time. In fact, very often, all the instances of the Data Objects of a model are not delivered simultaneously, but in several sessions (depending on the data production, the means of transfer, etc.). This process can be spread over several months, or several years or be ongoing.

The characteristics of the session must take into account the two previous aspects: identifier, date, version, start and end date in the case of an ongoing process, etc. This could also be a descriptive file provided simultaneously.

Lastly, the Archive and the Producer must establish a procedure for sending/receiving messages (forms, e-mails, acknowledgement of receipt, etc.), depending on needs. The Archive must have precise information on the contents of a session and, in turn, inform the Producer of the correct reception of the objects. E.g., in order to acknowledge session reception, the Archive may send an e-mail to the Producer indicating the date and contents of the reception.

A systematic validation is carried out, as described in the following chapter “Validation definition”.

3.2.2.3.4 Define the initial transfer test

The Producer and Archive must:

- Define the test SIPs.
- Identify the various kinds of tests, the aim of which is to check the following:
  - ❖ On the one hand the nominal functioning of the transfer: tests at the utmost limit (maximum volume of a file, maximum number of files), and then test performance. Test of the integrity of the objects received.
  - ❖ On the other hand the procedures in the event of breakdown (for example in the case of the transfer being interrupted).

3.2.2.3.5 Tools for the transfer

The Producer and Archive identify the software to be used by each other to manage the transfer. It can have an impact on the description of the transfer procedures.

3.2.2.4 VALIDATION DEFINITION

Formal definition phase: validation definition	Involves
Define a systematic validation plan.	Archive
Define an in-depth validation plan.	Archive
Define the procedures for rejection, re-transfer, object acceptance (forms, anomaly forms, technical approvals, reviews, etc.).	Producer and Archive
Define the initial validation test	Producer and Archive
Identify the validation tools.	Archive
Write a description of the validation procedures (part of the final agreement).	Archive

Action table 3.2.2-6: Formal definition phase: validation definition

- Systematic validations are carried out in a systematic way at the time of object reception. In this case, errors lead to immediate rejection,
- A more in-depth level of validation, which depends on the quality required by the Archive, may be carried out later. In this case, a classification of non-conformities must be established.
- In each of these two previous cases, the agreement or rejection procedures must be defined and approved by the Archive and the Producer,
- On the one hand these tests enable the validation means to be checked, on the other hand, the compliance of the objects delivered by the producer.
- This is an informative point for the Producer. The Archive identifies the tools to be used for the validation. Then Archive and Producer discuss the possibility for the Producer to re-use these tools.
- The last point concerns the writing of a description of validation as defined by the Archive and the Producer. This description will be part of the final folder of the Submission Agreement.

This chapter draws on the elements in the preliminary phase.

These points are developed in the paragraphs below.

3.2.2.4.1 Definition of the systematic validation plan

The Archive informs the Producer on the systematic validation carried out after reception. The following items are some important points to consider:

- Completeness (all the objects in the session have been correctly received).
- Integrity (the objects have not undergone any deterioration: checking with indicators such as volume).
- Conformity to the formal model. The objects delivered must correspond to the objects already identified in the model and they must conform to the dictionary (attributes).

3.2.2.4.2 Definition of the in-depth validation plan

Contrary to systematic validation, this is a more in-depth validation of the SIPs, such as, e.g., checking the coherence of the syntactic description of a file with respect to a described file or checking the contents of text documents.

The Archive informs the Producer of the desired validation level, the necessary validation time (and the conditions for this validation to take place, in particular, the elements which must be present in a compulsory manner). These checks can concern objects delivered in different transfer sessions. The Archive can establish a validation classification.

The checks automatically carried out should be distinguished from those that are carried out manually by people. These checks can be carried out in an complete manner or random sampling:

- Automatic checks, such as:
  - ❖ Checking the structure of a document (table of contents, conformity to a DTD for an XML document, for instance). This structure was defined during the Information Object definition phase.

Checking the structure of a data file with its syntactic description (e.g. EAST descriptor for a scientific data file)

- Manual checks:
  - ❖ Checking the intelligibility of document contents by partially or fully rereading (under no circumstances can the relevance and clarity of the semantic description of a file containing scientific observations be checked automatically).
  - ❖ Lastly, validation by experts representing the Designated Community should be considered. This point already concerns the AIP. However, the feedback can reveal insufficiency in the data model and thus lead to changes. It is essential to ensure that all the information delivered, possibly completed by other information already held by the Archive, enable the AIP to be created containing all the required qualities from a user point of view. The comprehensiveness and relevance of the information can only be determined by a peer review composed of experts and representatives of the Designated Community. The archivists may, if they consider it appropriate, invite the data Producer to this peer review.

3.2.2.4.3 Definition of the initial validation test

The Archive and Producer must:

- Define the test SIPs.
- Identify the kinds of test:
  - ❖ Test the validation means (tools, procedures),
  - ❖ Test conformity to the test SIPs received. In the case of anolalies of the objects, the Archive alerts non-conformities to the Producer. Then, the latter must correct the anomalies before the actual start-up of the deliveries.

3.2.2.4.4 Procedures and tools

The Archive and the Producer agree on the (total or partial) acceptance or (total or partial) rejection procedures of the session in the event of non-conformity with previous elements (anomaly forms, other forms, etc.). They also decide on the re-transfer procedures (and the deadlines). A technical receipt can close this phase. After these validations, the Archive can, for example, ask for modification of certain objects or complementary information.

Define the tools and validation procedures to be installed on both sides (some tools can be installed on the premises of the Producer so that validation can be carried out at that end. For example, a tool enabling a check of the compliance of an XML document with its DTD).

3.2.2.5 DELIVERY SCHEDULE

Formal definition phase: delivery schedule	**Involves**
Define a reference delivery schedule (part of the final agreement).	Producer and Archive
Define the procedures to implement in the event of the schedule not being followed	Producer and Archive

Action table 3.2.2-7: Formal definition phase: delivery schedule

- Define a reference delivery schedule with respect to the different objects or sets of objects that will be transferred. This schedule is an updated and completed version of the preliminary phase one (type of elements delivered-data files, descriptive files-, timetable, key dates, etc.).

- The schedule must be regularly revised and the reasons for any divergence must be analyzed. The Producer and the Archive must specify the procedure to follow in the event of divergence.

3.2.2.6 CHANGE MANAGEMENT DURING THE LIFE OF AN ARCHIVE PROJECT[1]

Formal definition phase: change management during the life of an Archive Project	**Involves**
Assess the causes for the change.	Producer and Archive
Define how to manage changes in information already archived (updates, migrations) and versions.	Producer and Archive
Assess the limits of the Submission Agreement with respect to changes (when should a new agreement be re-negotiated).	Producer and Archive
Write a description of the change management (part of the final agreement).	Producer and/or Archive

Action table 3.2.2-8: Formal definition phase: change management during the life of an Archive Project

- The causes for the change: it is important to emphasize that even in the perspective of long-term archiving, the formalism proposed for the Producer-Archive Interface is not static. It is however necessary to adapt to a context that is permanently changing: definition of new Data Objects to deliver, modification of the characteristics of objects already defined, new formats, transfer method changes, validation procedure changes, etc.

❖ The model of the instances to be delivered and the dictionary are the two indispensable elements for a complete and non-ambiguous definition of the objects that must be delivered. Whatever the case may be, this definition is neither static nor fixed in a permanent manner. A certain number of mechanisms must be planned to complete or modify this formal definition during the process of Ingest itself.

❖ Changes in the volume of data, the frequency, the development of new technologies may imply changes in the transfer procedure.

❖ The above changes may imply a change in the transfer sessions or the validation procedures.

- The information already archived may change in time due to updates or migrations. The Producer and the Archive must define how to handle these changes and the versions that may result from such changes.

- Some changes are minor, others may require re-negotiation of the Submission Agreement. The Producer and the Archive must negotiate what the limits of the agreement are.

- The last point is the writing of a description of the change management. This description will be part of folder of the final agreement.

[1] The changes that may appear necessary after the end of an Archive Project (for example errors detected in data files long after) are not handled in this paragraph, because they imply a re-negociation of the Submission Agreement.

3.2.2.7 FEASIBILITY AND ASSESSMENT

Formal definition phase: **feasibility and assessment**	**Involves**
Assess the costs for the Archive and the Producer.	Producer and Archive
Validate the project's feasibility.	Producer and Archive

Action table 3.2.2-9: Formal definition phase: feasibility and assessment

- This concerns the validation of the feasibility of the project, assessed in the preliminary phase.
- The Archive and the Producer must re-assess their costs separately (internal documents).

At this stage, the Archive must reexamine the points that only concern the Archive (see section "Permanent impact on the Archive" of the preliminary phase, and all tasks related to data ingestion, see also section "Quantification" of the preliminary phase).

3.2.3 SUBMISSION AGREEMENT

All the element resulting from this formal definition (dictionary, model, etc.) must be approved and signed jointly by the Producer and the Archive.

Formal definition phase: Submission Agreement	**Involves**
Draw up the Submission Agreement	Producer and/or Archive

Action table 3.2.3-1: Formal definition phase: Submission Agreement

The formal phase is concluded by drawing up the Submission Agreement. This document is the result of all the preceding negotiations. It is a folder that regroups all the textual descriptions for each and all the paragraphs that make up the formal phase:

- Information to be transferred (SIP contents, SIP packaging, data models, Designated Community, legal and contractual aspects, etc.).
- Transfer definition (specification of the Data Submission Sessions, etc.).
- Validation definition.
- Change management (conditions for modification of the agreement, for breaking the agreement, etc.).
- Schedule (submission timetable).

In some cases, there can be several 'Submission Agreements' between a Producer and an Archive, and these different agreements cover different and independent sets of information. As a matter of fact, the Producer may not be able to agree on all planned data sets, but on sets or sub-sets of information, due to constraints linked to long term Data production (for example the lack of resources may imply changes in data production).

3.3 TRANSFER PHASE

Aim: actual transfer of the Data Objects between the Producer and the Archive.

During a Data Submission Session, one or more SIPs are delivered. The SIP is, in turn, composed of one or more digital Data Objects, the characteristics of which are described in the dictionary.

Each object delivered is in reference to an object that has been previously identified with respect to a data model.

The subjects of the transfer phase are dealt with in a more precise way in the following paragraphs in the form of lists of actions to be carried out.

Summary of the transfer phase
Carry out the transfer test.
Manage the transfer.

Summary table 3.3-1: Transfer phase

3.3.1 CARRY OUT THE TRANSFER TEST

Transfer phase: carry out the transfer test	**Involves**
Initial transfer test.	Producer and Archive

Action table 3.3.1-1: Transfer phase: carry out the transfer test

- Initial test: to ensure full agreement on both sides, some initial submissions should be performed on the "test data" before the beginning of the data delivery. These tests must be carried out in compliance with that defined in the formal definition phase. After these tests have been carried out, the anomalies arising must be corrected and the operating parameters of the transfer must be adjusted. It can be seen whether the differences between the performance shown and the expected performance require a review of the agreement or the schedule.

  A test transfer may not be necessary for each new Submission Agreement. The Archives may not require a test transfer from a Producer with which the Archive has a good working relationship and has had no prior transfer or data validation problems.

All these tests must be carried out before the start-up of the actual transfer operations.

3.3.2 MANAGE THE TRANSFER

Transfer phase: manage the transfer	**Involves**
Ensure the good execution of the data transfer operation from both the Producer and Archive sides.	Producer and Archive

Action table 3.3.2-1: Transfer phase: manage the transfer

- This phase consists of ensuring that the data transfer takes place correctly, both on the side of the Producer and the Archive:

  - ❖ Respecting the schedule for the Data Submission Sessions (transfer within planned time periods). This implies handling a timetable, for transmissions from the Producer, and for receptions by the Archive (-progress indicators, etc.).

  - ❖ The establishment and respect of procedures defined in the formal phase (session contents, packaging, supports, etc.).

  - ❖ Making sure that the operation runs well technically: e.g. good network transmission (no cut-off, no transfer problems, etc.). This implies establishing a maintenance service to ensure the correct operation of the communication networks and that carries out appropriate actions in the event of failure.

  - ❖ In the case of media transfers: making sure that the media sent by the Producer has been received by the Archive, that it has not been damaged and that it is readable.

  - ❖ Management of transmission anomalies, re-transfers.

  - ❖ Sending acknowledgements of receipt per session by the Archive.

In this phase, for the transfer the Archive and Producer should use the tools identified in the formal phase.

3.4 VALIDATION PHASE

Aim: to carry out the validation of delivered objects, manage the anomalies detected, and accept all the objects transferred.

The subjects of the validation phase are handled in a more precise manner in the following paragraphs in the form of lists of actions to be carried out.

Summary of the validation phase
Carry out the validation test.
Manage the validation.

Summary table 3.4-1: Validation phase

3.4.1 CARRY OUT THE VALIDATION TEST

Validation phase: carry out the validation test	**Involves**
Initial validation test.	Producer and Archive

Action table 3.4.1-1: Validation phase: carry out the validation test

- The tests must be carried out in compliance with that defined in the formal definition phase:

  - ❖ Initial test: to ensure full agreement on both sides, the systematic validation plan should be performed on “test data” before the beginning of the data delivery.

  - ❖ It should be taken into account that the validation tests are related to the types of information on which they are applied. They must be performed prior to the first deliveries of this information,

and thus may be spread out in time, according to the arranged schedule. In addition, the test phases may reappear in the course of time if new information categories are defined.

3.4.2 MANAGE THE VALIDATION

Validation phase: manage the validation	**Involves**
Apply the validations.	Archive
Manage the results of the validation.	Producer and Archive

Action table 3.4.2-1: Validation phase: manage the validation

In this phase, the Archive and Producer should use the validation tools identified in the formal phase.

- Check the conformity of the delivered objects with respect to the model of objects to be delivered and validate their contents. Two validation plans identified in the formal phase may be distinguished:

  ❖ Systematic validation:

  These validations are carried out after each transfer session.

  At this stage, the Archive implements the systematic validation plan defined in the formal phase. In order to do this; the Archive must have already installed the required tools.

  All non-conformity, at this stage, implies rejection of the delivered objects during the session, and an anomaly form is sent to the Producer. The non-conformity is dealt with by both the Archive and Producer.

  ❖ In-depth validation:

  These validations are not necessarily carried out in every session. They can be carried out when there is a coherent package of information, or even at the end of the Archive Project when all the Data Objects are present. Some checks, in fact, require the presence of several files that are not necessarily delivered at the same time.

  At this stage, the Archive carries out the checks defined in the in-depth validation plan in the formal phase.

  The Archive must have already installed the required tools for the automatic checks.

- Managing the results means:

  ❖ The Archive identifies and sends out diagnostic and/or irregularity forms in accordance with the procedure defined in the formal phase.

  ❖ The Archive and the Producer manage the anomaly forms.

The Archive agrees on the transferred objects: the Archive sends the Producer its agreement to specify that the Data Objects it has received have been validated and accepted (there may be a first level and then a second level agreement).

4. CREATING AN ARCHIVE-PRODUCER INTERFACE METHODOLOGY FOR A PARTICULAR COMMUNITY

4.1 PURPOSE

This document provides a general framework applicable to the interface between the Archive and the Producer of information in order to reach a '**Submission Agreement**', then to transfer the SIPs, and finally to validate these SIPs by the Archive.

In order for this standard to be really useful and applicable to a particular community, this standard should be tailored to take into account the specific features of the community in question.

The term community is used here in a very broad and open sense: it could be a huge set such as that of the archive of scientific data files or documents files for libraries. On the other hand, it could be limited to just one Archive and to the community of the information Producers related to this Archive.

Taking into account the specific features of the community must give rise to a new standard, which is called community standard. From this standard, when a large community is addressed, further tailoring could be used to create specific standards for sub-communities.

The aim of this section is to define the stages required for creating a community standard by taking the Abstract standard as an initial working base with which the community standard must be in conformance.

4.2 THE RESPONSIBILITY OF CREATING THE COMMUNITY STANDARD

It is the more or less wide range of the community that enables us to know who can undertake the task of creating a community standard:

According to the community range, this could be the following:

- National and international standardization bodies, which are usually organized and structured by grouping the players of a certain problem. This may be, for example, the International Standardization Organization (ISO),

- National and international organizations of the community itself: This could be an organization with the role of coordinating activities of the community itself and that is responsible for the tasks of a regulatory kind. For example, the International Council on Archives.

- It could merely be an Archive that organizes drawing up the implementation standard to be proposed to its information Producers.

The list shown above is merely an example and the aim of this list is to show the different contexts in which a community standard may be created.

4.3 THE WORK STAGES FOR THIS CREATION

4.3.1 TERMINOLOGY

The abstract standard has been drawn up with a neutral vocabulary defined for basic purposes in the Reference Model OAIS (see [1]).

In order for the community standard to be used and easily understood by the players in the community, the vocabulary of the community itself must be used.

Therefore, the terminology of the community that will be used should be defined or referenced in this standard. It is also advisable, but not indispensable, to provide an equivalence table between the vocabulary of the standard and the vocabulary of the community, as an annex.

4.3.2 INFORMATION MODEL OF THE COMMUNITY

The terminology must enable, in particular, the main information objects of the community and the general attributes of the relevant data objects to be defined.

In addition to this terminology, the kind of relationship between these object categories must also be defined.

This community modelisation work is extremely useful for creating the data dictionary and the formal model related to an Archive Project.

4.3.3 STUDY OF THE ACTIONS DEFINED IN THE ABSTRACT STANDARD

- Each action defined in the abstract standard has to be analyzed in the context of the community, considering the following:
  - ❖ The action conforms to the context and can be applied as it is.
  - ❖ The action needs to be specialized: that means precision is added to conform to the context of the community. The actions related to known difficult points must be analyzed and described with the appropriate level of detail.
  - ❖ The action is out of context, and needs to be suppressed.
- Additional actions specific to the community may be added.

4.4 COMMUNITY STANDARDS

The general standards, the standards related to the community, and, more generally, all the documents, forms, applicable procedure manuals must be identified and referenced.

4.5 COMMUNITY TOOLS

The community tools that may or must be used are identified regarding each of the phases in the process.

ANNEX A

TARGETED OVERVIEW OF THE OAIS RM DEDICATED TO THE METHODOLOGY ABSTRACT STANDARD

(This annex is **not** part of the Methodology.)

(version 3.0 - June 22, 2002)

The purpose of this appendix is to provide a brief overview of the important terms and concepts, as defined in the "Reference Model for an Open Archival Information System (OAIS)", needed to understand this document. Readers are urged to read the full OAIS document to fully understand the concepts.

The OAIS Reference Model is a framework for understanding and applying concepts needed for long-term digital information preservation (where long-term is long enough to be concerned about changing technologies). It is also a starting point for a model addressing non-digital information. It does not specify any implementation.

Open Archival Information System:

What is meant by an “Open Archival Information System?”

‘Open’ is simply referring to the fact that this standard was developed in an open forum and is freely available.

The “Information” part is more difficult and can have subtle ramifications. For now, information is simply any type of knowledge that can be exchanged, and that data refers to the way this knowledge is represented in the exchange. This will be expanded upon later.

The phrase “Archival Information System” is used to refer not only to the hardware and software, but also the people who are involved in acquiring information, preserving it, and making it available to those needing the information.

There are many terms that need to be used in well defined ways in order to construct a reference model. The OAIS has a glossary of such terms, and a few of the more important of these are defined below when they are needed.

Environment Model:

The modeling starts by giving a view, in Figure A-1, of the OAIS as a box with three primary interfaces.

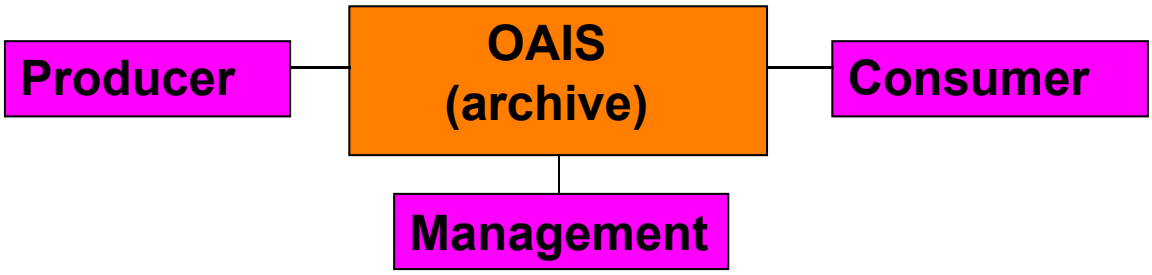


Figure A-1: OAIS Environment Model

– Producers play the role of those who provide the information to be preserved.

– Management plays the role of those who set overall OAIS policy where the OAIS is only one of their concerns. Day-to-day administration of the OAIS is handled by an Administration function within the OAIS box.

– Consumers play the role of those who interact with the OAIS services to find information of interest and to access this information.

Later, the OAIS box will be expanded into six functional areas. Although not described here, the OAIS Reference Model also identifies a minimum set of responsibilities that must be discharged for an Archive to call itself an OAIS Archive.

Information Modeling:

As mentioned above, information is expressed by some type of data. It is the interpretation of the data, using additional representation information, that yields the information desired. This is shown in Figure A-2 schematically. Consider a simple example to clarify the relationships.

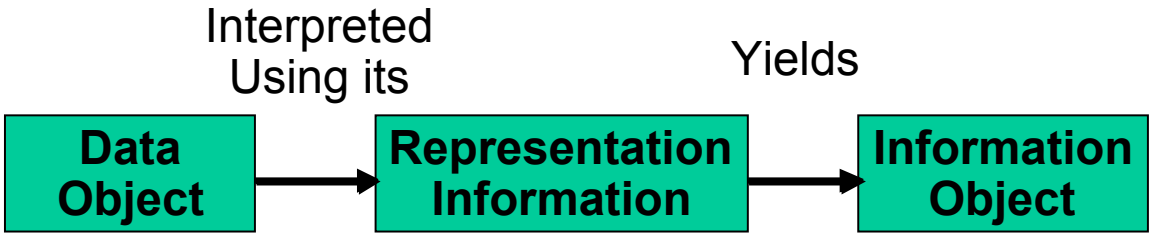


Figure A-2: An Information Object

Consider a data object to be a particular string of 128 bits in a file. Given the information that these bits are to be interpreted by applying the ASCII standard, an understanding the data (bit string) as a sequence of ASCII characters is obtained. This process has converted the data object (bit string), using the ASCII standard (Representation Information), into an Information object that is more meaningful than the original bit string. Note that in order to preserve the information object, it is necessary to preserve not only the bit string, but also the ASCII representation information and the association between the two.

Of course the Representation Information may be much more complex than the ASCII standard, and so the Information Object may be much more complex than a sequence of characters.

A key information-modeling concept in the OAIS is the Information Package. Think of it as a container, as shown in Figure A-3, which holds two types of information, called Content Information and Preservation Description Information.

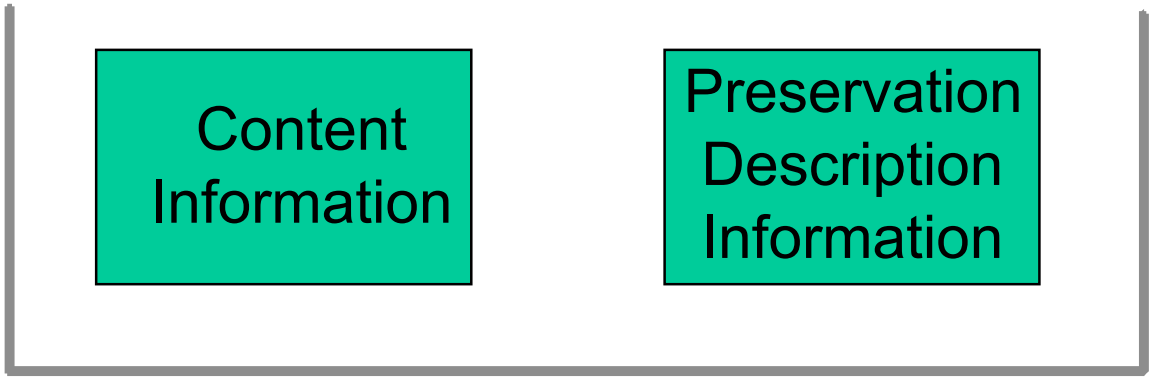


Figure A-3: Information Package Definition

Note that each of these is an Information Object and thus will have its own Data Object and Representation Information. The Content Information's Data Object is referred to as the Content Data Object.

The Content Information is defined to be that information that is the original target of preservation. For example, suppose the objective it to preserve the content of a book in electronic form. It could be decided that the Content Information is all the information that allows a re-creation of a view of the book, from its cover through all the pages, including figures, etc. This could be constructed as, or received as, a single data file in Adobe's PDF format. This would be called the Content Data Object. The associated Representation Information, needed to provide the end view of the book, would be contained in the Adobe software as it has the information to map the bits of the file into the view that is to be preserved.

Alternatively, it might be that the book is really just text organized into chapters. It can be adequately represented simply as a text file with no need to use PDF or other complex formatting. Just what constitutes the Content Information to be preserved is not always obvious, and may need to be negotiated with the Producer.

Note that in the general case, the Content Data Object doesn't have to be a digital object. It could be a physical object, such as moon rock or a piece of film. The Representation Information would be used to add meaning about what was being preserved.

In addition to the Content Information, and Information Package may also contain a type of information called Preservation Description Information. The purpose of this information is to assist in preserving the Content Information, and it is broken down into four sub-categories.

First, the Reference Information is used to provide one or more systems of identifiers by which to identify the Content Information. For example, this might include bibliographic attributes and/or a Digital Object Identifier.

Second, the Provenance Information describes the history of the Content Information, including the chain of custody, so that Consumers can better judge how much to trust the information.

Third, the Context Information relates the Content Information to other information outside the Information Package. This provides Consumers with an understanding of how the information being preserved relates to a wider environment.

Finally, the Fixity Information is used to help ensure that the Content Information is not altered in an undocumented manner. For example, this might include checksums and digital signatures.

The Preservation Description Information is an essential part of the Information Package used by the OAIS for its preservation function.

While an Information Package typically contains two types of information, Content Information and Preservation Description Information, there are also three variants of the Information Package depending on where the package is being used in the OAIS environment.

The first of these is the Submission Information Package, used to provide information to the OAIS by the Producer. Typically it is subject to negotiation between the two.

The second of these is the Archival Information Package. It is used by the OAIS to hold the Content Information and Preservation Description Information as it performs its preservation function. Note that it may take several Submission Information Packages to form a single Archival Information Package, or one Submission Information Package may result in several Archival information Packages.

The third of these is the Dissemination Information Package. It is used to provide requested information to the Consumer. Note that it may contain only a part, or all, of one or more Archival information Packages as determined by the OAIS in response to requests.

The use of the three variants of an Information Package are shown in Figure A-4.

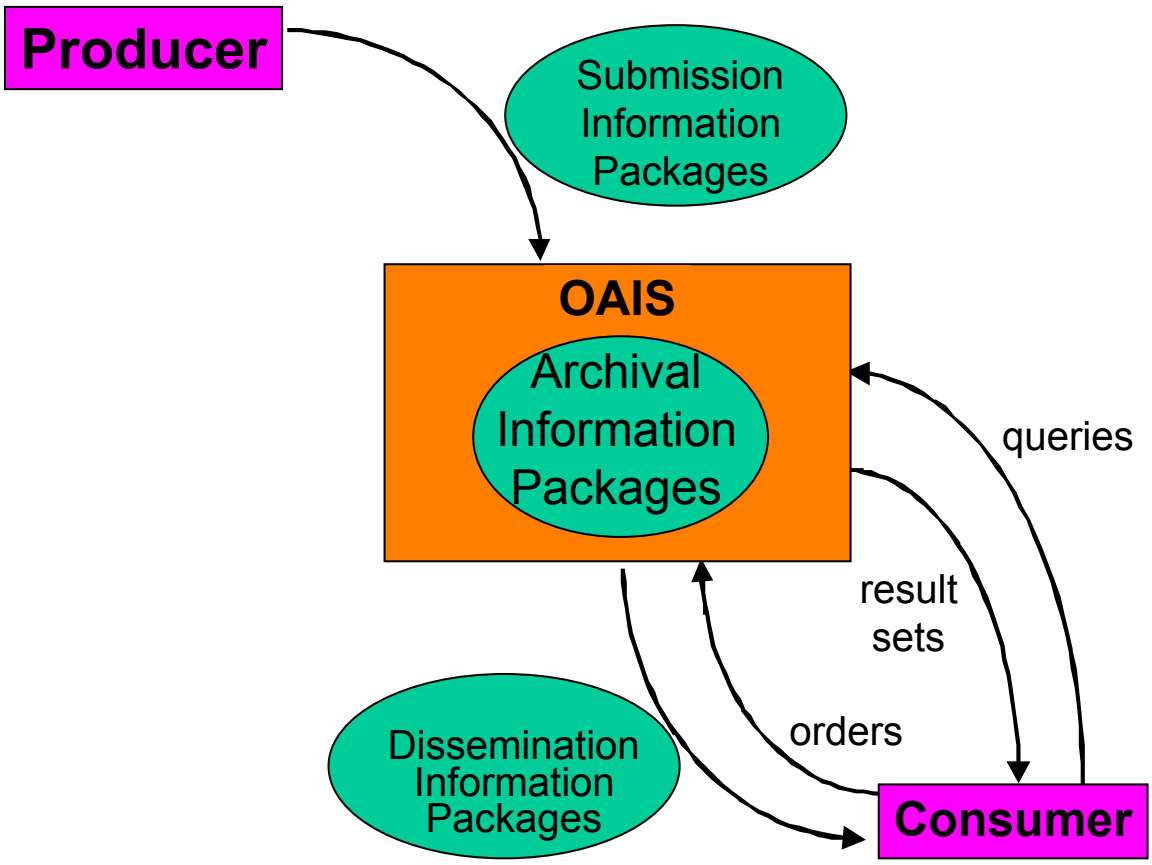


Figure A-4: External Data Flow View

The Submission Information Package is submitted to the OAIS by a Producer. The OAIS holds and preserves the information using Archival Information Packages. In response to Consumer queries and resulting orders, Dissemination Information Packages are returned.

The OAIS reference model goes into additional detail regarding the modeling of an Archival Information Package. It would not be appropriate to present all of this detail here, but some additional modeling is needed and is shown in Figure A-5.

Figure A-5 is an example of the more formal modeling, using the Unified Modeling Language, of information in the OAIS as applied to the Archival Information Package (AIP).

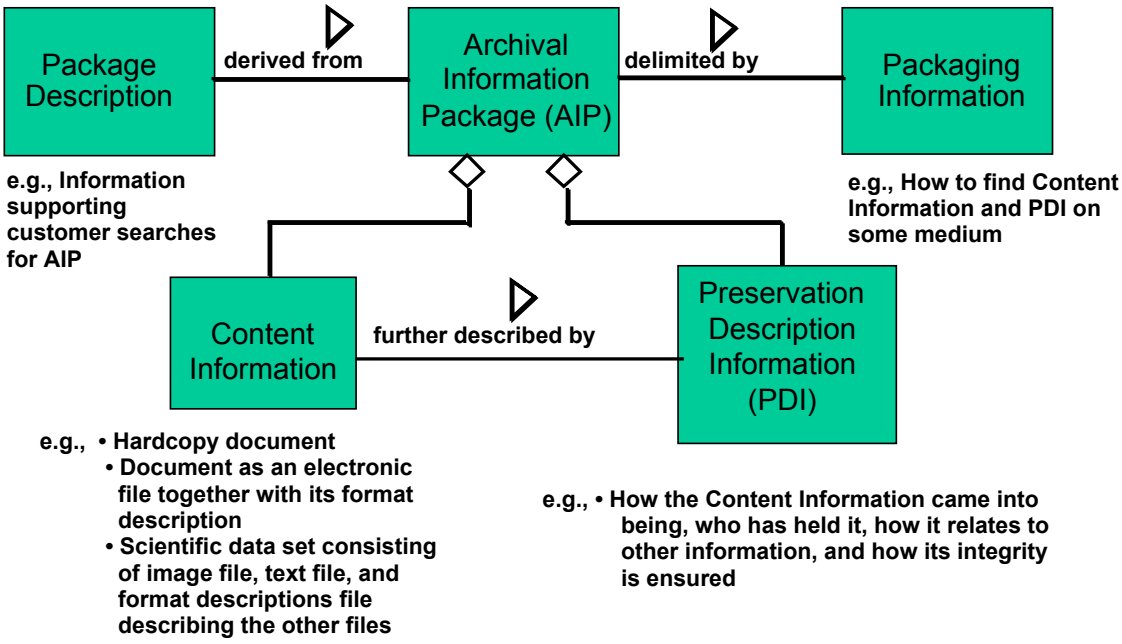


Figure A-5: Archival Information Package

The diamonds under the AIP box indicate that the AIP is a container holding two types of information - the Content Information and the Preservation Description Information. Examples of these types of information are given in the text below each of the boxes.

For example, the Content Information may be a hardcopy document, an electronic document with its Representation Information, or a set of files corresponding to a scientific data set with its Representation Information. Note that the Representation Information will include a format description, and may include additional semantic information such as that provided by a Data Dictionary. It is important for the OAIS to ensure that the Content Information and Preservation Description Information are understandable to the expected Consumer community. Such a community is referred to as the Designated Community for the given Archival Information Package.

What is new in this expanded view of an AIP are two additional types of associated information. The one on the right is called Packaging Information and it is used to bind the Content and PDI. The one on the left is called Package Description and it is used to support searching for the Content Information.

Packaging Information is the information that is used to logically, or actually, bind the Content Information and Preservation Description Information into a recognizable package with its constituent parts. It allows one to actually find the constituent parts on some media. It might be implemented using file systems, directory structures, pointers, and generic languages like XML.

The Package Description is used to hold the type of information needed by access aids, to support a Consumer's search for and retrieval of desired Content Information. It is most likely to be implemented in databases, and it is viewed as information that is most likely to be updated over time. A card catalogue is an example. It is not critical for preservation because it can be regenerated, in principle, if needed.

Having looked at the information modeling aspects of the OAIS reference model, it is time to take a brief look at the modeling of archive functions.

Functional Modeling:

Six primary functions have been identified, as previously noted.

Ingest is the first, and this entity provides the major interface between the OAIS and the Producer. It accepts Submission Information Packages from Producers during a Data Submission Session. This session may be comprised of a delivered set of media, or it may be a single telecommunications session. The Submission Information Packages will conform to agreements reached between the Producer and the OAIS as defined in the Submission Agreement. Ingest prepares Archival Information Packages and Package Descriptions for storage and subsequent access.

Archival Storage is the second, and this entity accepts Archival Information Packages, maintains them, and provides them upon request.

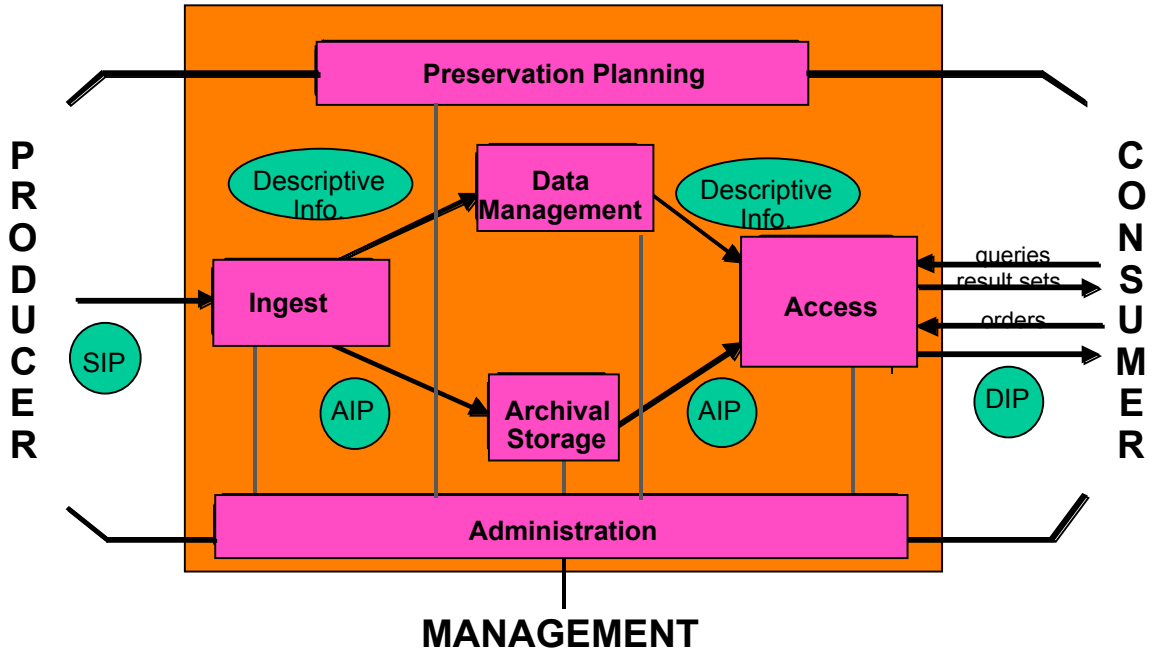
Data Management is the third, and this entity accepts Package Descriptions from the Ingest function and other types of metadata needed to support overall OAIS operations.

Administration is the fourth, and this entity is responsible for managing the overall operation of the OAIS on a day-to-day basis.

Preservation Planning is the fifth, and this entity is responsible for monitoring technology evolution and the needs of the Designated Communities, and for forming preservation strategies and techniques to support the OAIS preservation function.

Access is the last function, and this entity supports Consumers in identifying, locating, and accessing the information of interest.

The conceptual relationships of the six functional areas, along with the three variations of information packages, are shown in Figure A-6.



SIP = Submission Information Package
AIP = Archival Information Package
DIP = Dissemination Information Package

Figure A-6: OAIS Functional Entities

This Figure may be understood as follows:

Conceptually, a Submission Information Package is provided by a Producer to the Ingest entity. An AIP is created and delivered to Archival Storage. Related Descriptive Information is provided to Data Management. A Consumer searches for, and requests, information using appropriate Descriptive Information and access aids. The appropriate AIP is retrieved from Archival Storage and transformed by the Access entity into the appropriate Dissemination Information Package for delivery to the Consumer. This is all under the guidance of the Administration entity. Preservation strategies and techniques are recommended by Preservation Planning and put in place by the Administration entity.

Within the OAIS the functional entities are broken into sub-functions. The purpose is to more clearly identify the types of functions involved, not to promote a specific implementation. The readers should consult to the OAIS Reference Model for these details.

To summarize, the OAIS reference model is applicable to all digital Archives, their Producers and Consumers.

It established common terms and concepts for comparing archival concepts and implementations, but it does not specify a particular implementation.

It identifies a minimum set of responsibilities that must be discharged for an Archive to call itself an OAIS Archive.

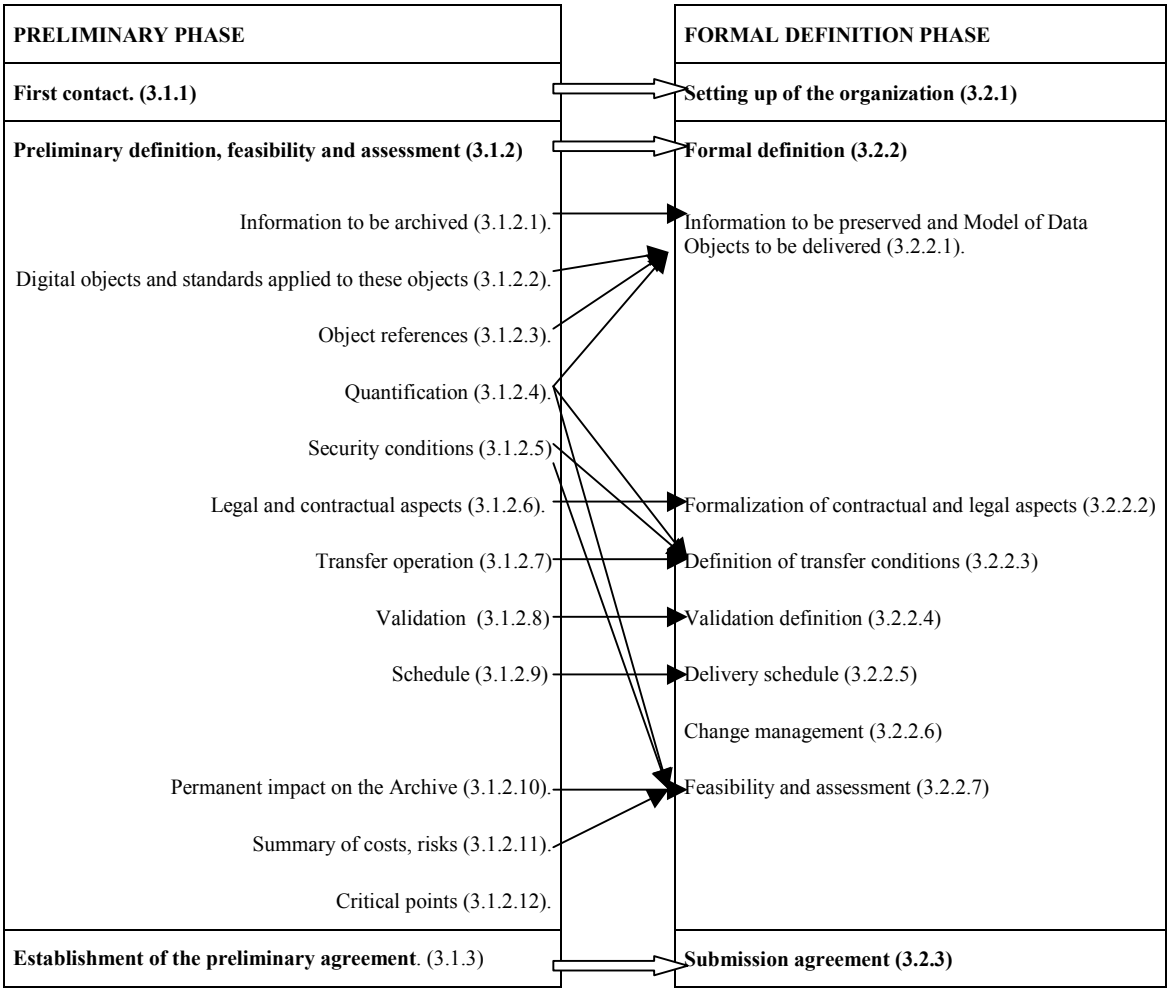
It provides detailed models for archival function and for the information associated with Archives.

Although not discussed in this paper, it also provides perspectives on migration, emulation and interoperability among OAISs.

ANNEX B

LINKS BETWEEN PRELIMINARY AND FORMAL DEFINITION PHASE

(This annex is **not** part of the Methodology.)



Summary table B-1: Correspondence between preliminary and formal definition phases

In this table,

- The big arrows describe the links between sub-phases levels,
- The fine arrows describe the links between groups of actions in a sub-phase.