The Consultative Committee for Space Data Systems

Draft Report Concerning
Space Data System Standards

PRODUCER-ARCHIVE INTERFACE SPECIFICATION (PAIS)—A TUTORIAL

DRAFT INFORMATIONAL REPORT

CCSDS 651.2-G-0

DRAFT GREEN BOOK

August 2016

AUTHORITY

Issue:	Draft Green Book, Issue 0
Date:	August 2016
Location:	Not Applicable

(WHEN THIS INFORMATIONAL REPORT IS FINALIZED, IT WILL CONTAIN THE FOLLOWING STATEMENT OF AUTHORITY:)

This document has been approved for publication by the Management Council of the Consultative Committee for Space Data Systems (CCSDS) and reflects the consensus of technical working group experts from CCSDS Member Agencies. The procedure for review and authorization of CCSDS Reports is detailed in *Organization and Processes for the Consultative Committee for Space Data Systems* (CCSDS A02.1-Y-4).

This document is published and maintained by:

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

FOREWORD

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

http://www.ccsds.org/

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

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

Member Agencies

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

Observer Agencies

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

DOCUMENT CONTROL

Document	Title	Date	Status
CCSDS 651.2-G-0	Producer-Archive Interface Specification (PAIS)—A Tutorial, Draft Informational Report, Issue 0	August 2016	Current draft

CONTENTS

CONTENTS (continued)

Section Page

Figure

CONTENTS (continued)

Figure Page

Table

CONTENTS (continued)

Table Page

1 INTRODUCTION

1.1 PURPOSE AND SCOPE

The purpose of this CCSDS report is to provide a tutorial for the Producer-Archive Interface Specification (PAIS) Recommended Standard (reference [1]).

1.2 DOCUMENT STRUCTURE

This document is organized as follows:

- Section 1 defines the purpose, scope, structure, definitions for terminology, and references to standards and external documents used in this CCSDS report.
- Section 2 introduces the concept of transfer, the terminology inherited from the PAIS standard and the OAIS standard (reference [4]).
- Section 3 provides an overview and guidelines for building a Model of Objects for Transfer (MOT) and associated Submission Information Packages (SIPs).
- Section 4 introduces the PAIS XML Descriptors and provides concrete examples covering all XML elements, and it provides best practices.
- Section 5 describes the PAIS abstract implementation of SIPs and their concrete generation, ingestion, and validation as either XFDU (reference [4]) or non-XFDU packages.
- Section 6 provides a series of use cases that enforce the understanding of the PAIS standard through concrete and complete examples.
- Section 7 introduces a series of existing software tools that may help implement the PAIS standard.
- Annexes:
  - Annex A addresses the specific Associated Descriptor Data Identifier.
  - Annexes B to E contain the Descriptors, SIP constraints, and extracts of SIP manifests from the four practical use cases:
    - ISEE—a typical use case;
    - ESA-SAFE—transfer of SAFE products;
    - CoRoT—end of mission bulk transfer; and
    - BnF and METS—a non XFDU SIP implementation.

It is good practice to define acronyms when first used.

1.3 DEFINITIONS

1.3.1 ACRONYMS AND ABBREVIATIONS

For the purposes of this document, the following acronyms and abbreviations apply.

AIP	Archival Information Package
ASCII	American Standard Code for Information Interchange
BnF	Bibliothèque nationale de France (French); French National Library (English)
CCSDS	Consultative Committee for Space Data Systems
CMC	CCSDS Management Council
CNES	Centre National d'Études Spatiales
CoRoT	COnvection ROtation et Transits planétaires (French); COnvection ROtation and planetary Transits (English)
DFDL	Data Format Description Language
DIP	Dissemination Information Package
EO	Earth Observation
ERS	European Remote Sensing Satellite
ESA	European Space Agency
FITS	Flexible Image Transport System
GB	Gigabyte
GUI	Graphical User Interface
ISEE	International Sun-Earth Explorer
KB	Kilobyte
MB	Megabyte
METS	Metadata Encoding and Transmission Standard
MOT	Model of Objects for Transfer
NASA	National Aeronautics and Space Administration
NSSDCA	NASA Space Science Data Coordinated Archive
OAIS	Open Archival Information System
PAIS	Producer Archive Interface Specification
PAIMAS	Producer-Archive Interface Methodology Abstract Standard
PDI	**Preservation Description Information**

PREMIS	Preservation Metadata Implementation Strategies
SAFE	Standard Archive Format for Europe
SAR	Synthetic Aperture Radar
SIP	Submission Information Package
SLA	Service Level Agreement
SPAR	Scalable Preservation and Archiving Repository
TIFF	Tagged Image File Format
UML	Unified Modeling Language
XFDU	XML Formatted Data Unit
XML	Extensible Markup Language
LZW	Lempel-Ziv-Welch: compression algorithm

SANA

1.3.2 GLOSSARY OF TERMS

PAIS terminology, as defined in references [1] and [4], is used throughout this CCSDS Report. Only brief definitions are provided here.

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.

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.

Collection Descriptor: A set of attributes that describes a view of a single collection of data and that identifies the parent collection of which it is a part.

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

Data Object Type: A set of characteristics describing a Data Object (such as the size of this object and the description of its content). Typically there will be multiple Data Objects conforming to the same Data Object Type.

Descriptor: Either a Collection Descriptor or a Transfer Object Type Descriptor.

Descriptor Model: A model that defines the mandatory and optional attributes needed for a Collection Descriptor or a Transfer Object Type Descriptor.

Fixity Information: The information which documents the authentication mechanisms and provides authentication keys to ensure that the Content Information Object has not been altered in an undocumented manner.

Formal Definition Phase: 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.

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

model: A data entity described independently from any instance in a data product, and corresponding to a reusable data entity definition, from which other data entities may inherit the attributes and apply some specialization rules.

Model of Objects for Transfer, MOT: The set of all Descriptors for a given Producer-Archive Project. It is used jointly by the Producer and the Archive to provide a common and understandable hierarchical view of the Producer's Data Objects to be transferred and their organization into collections, and it supports possible additional relationships among them. The hierarchy may be viewed as a tree having leaf and non-leaf nodes. The Data Objects to be transferred, organized as 'Transfer Objects,' are represented by the leaves of the MOT. Thus the nodes of the MOT have a different meaning depending on whether they are leaves or not:

- A leaf node corresponds to a single Transfer Object Type and therefore one exists for each Transfer Object Type Descriptor.

- A non-leaf node corresponds to a collection view of Transfer Object Types, or of a collection of collections. A non-leaf node exists for each Collection Descriptor.

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

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

Producer-Archive Project: A Producer-Archive Project is a set of activities and the means used by the information Producer as well as the Archive to ingest a given set of information into the Archive.

Submission Information Package, SIP: An Information Package that is delivered by the Producer to the OAIS for use in the construction or update of one or more AIPs and/or the associated Descriptive Information.

SIP Content Type: A set of attributes describing the content of each type of SIP.

SIP Constraints: A set of attributes and values assigned to these attributes in order to characterize the constraints of content types or sequencing order.

SIP Sequencing Constraint Group: A set of attributes describing the delivery order applying to the SIP Content Types within the Producer-Archive Project.

Transfer Object: A set of one or more Transfer Object Groups, containing at least one Data Object, that are to be transferred to the Archive.

Transfer Object Group: A set of zero or more Data Objects and zero or more Transfer Object Groups.

Transfer Object Group Type: A set of characteristics describing a Transfer Object Group. Typically there can be multiple Transfer Object Groups conforming to the same Transfer Object Group Type.

Transfer Object Type: A set of characteristics describing a Transfer Object (such as the size of this object, the description of its content, and its makeup in terms of one or more Data Object Types). Typically there can be multiple Transfer Objects conforming to the same Transfer Object Type.

Transfer Object Type Descriptor: A set of attributes that describes a Transfer Object Type and that identifies the parent collection of which it is a part.

Transfer Phase: 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.

Validation Phase: 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.

1.4 CONVENTIONS

This document uses the UML convention for the figures. The object relationships are described in the UML diagrams below. This is extracted from annex C of reference [4].

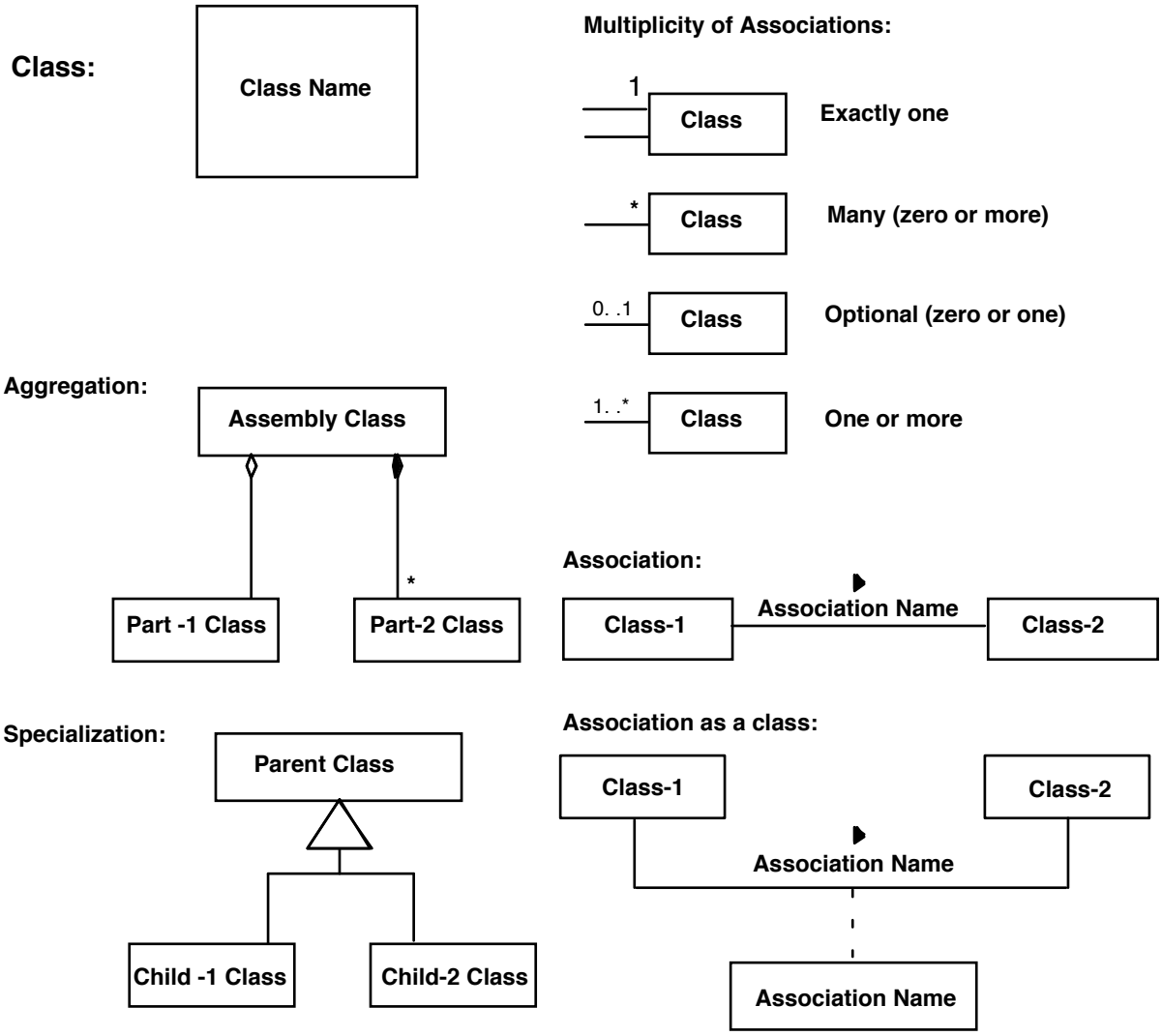

Figure 1-1: Key to UML Relationships

A Class is indicated by a rectangle containing the Class name. The UML representation of a class is a three-compartment rectangle with name in the top compartment attributes in the second compartment and methods in the lowest compartment. In this document the attributes and operations compartments are always empty and UML states empty compartments can be suppressed.

Classes of objects are related to one another through Associations, and there are various multiplicities that may be attached to these associations as shown. The multiplicity refers to the number of instances, or objects, of that class that are involved in the relationship.

A solid line connecting two classes indicates the general association, among two classes. The line is labeled with an association name, indicating the nature of the association, and a solid arrowhead indicating the direction that the relationship should be read. The multiplicity of each class is shown next to the class near the association line. If the association forms a class that may have its own attributes or methods, that association class is shown as a rectangle connected to the solid line by a dashed line. The multiplicity may be omitted if the association is one to one.

There are two particular associations that are commonly used, aggregation and specialization, and these have particular symbols to indicate them.

An Aggregation association is one where a class is considered to be a part of another class. In UML, a diamond connecting the aggregation association to the aggregated class shows association. There are two types of aggregation defined by UML. Composition (sometimes referred to as strong aggregation), where the part classes are physically stored as part of the aggregated class, is shown with a solid diamond. In a Composition, if the aggregated class is destroyed, the child classes are also destroyed.

Weak aggregation, where the part classes are referred to by the aggregated class, is shown with an empty diamond. In a weak aggregation, if the aggregated class is destroyed, the part classes are not destroyed and may be aggregated into other new classes. Composition can be thought of as aggregation by value, while weak aggregation can be thought of as aggregation by reference.

In figure 1-1, the aggregation association says that the Assembly class contains exactly one Part-1 class instance and zero or more Part-2 class instances. Also if an instance Assembly is destroyed the Part-1 instance will continue to exist but all the Part-2 instances will be destroyed.

A Specialization association is one where a child class inherits attributes and methods from the parent class. In UML, a broad triangle connecting the aggregation association to the parent class shows specialization. An instance of a child class contains all the attributes and methods contained by its parent class, so an instance of the child class can be used in any operation where an instance of the parent class would be valid. However, the child class may add any number of new attributes or methods, so an instance of a parent class is not necessarily a valid replacement for the child class. In the figure, the specialization association says that the Parent class attributes and methods are inherited by the Child-1 class and the Child-2 class.

1.5 REFERENCES

The following documents are referenced in this Report. At the time of publication, the editions indicated were valid. All documents are subject to revision, and users of this Report are encouraged to investigate the possibility of applying the most recent editions of the documents indicated below. The CCSDS Secretariat maintains a register of currently valid CCSDS documents.

[1] *Producer-Archive Interface Specification (PAIS)*. Issue 1. Recommendation for Space Data System Standards (Blue Book), CCSDS 651.1-B-1. Washington, D.C.: CCSDS, February 2014. [Equivalent to ISO 20104:2015.]

[2] *XML Formatted Data Unit (XFDU) Structure and Construction Rules*. Issue 1. Recommendation for Space Data System Standards (Blue Book), CCSDS 661.0-B-1. Washington, D.C.: CCSDS, September 2008. [Equivalent to ISO 13527:2010.]

[3] *Producer-Archive Interface Methodology Abstract Standard*. Issue 1. Recommendation for Space Data System Practices (Magenta Book), CCSDS 651.0-M-1. Washington, D.C.: CCSDS, May 2004. [Equivalent to ISO 20652:2006.]

[4] *Reference Model for an Open Archival Information System (OAIS)*. Issue 2. Recommendation for Space Data System Practices (Magenta Book), CCSDS 650.0-M-2. Washington, D.C.: CCSDS, June 2012. [Equivalent to ISO 14721:2012.]

[5] *Organization and Processes for the Consultative Committee for Space Data Systems*. Issue 4. CCSDS Record (Yellow Book), CCSDS A02.1-Y-4. Washington, D.C.: CCSDS, April 2014.

[6] *CCSDS Publications Manual*. Issue 4. CCSDS Record (Yellow Book), CCSDS A20.0-Y-4. Washington, D.C.: CCSDS, April 2014.

[7] *Standard Archive Format for Europe (SAFE)—Control Book—Volume 1—Core Specifications*. Issue 2. PGSI-GSEG-EOPG-FS-05-0001. Frascati: ESA/ESRIN, 2015.

Include RMP, SANA, and CCSDS URN references

2 PAIS AT A GLANCE

The purpose of the PAIS standard is to provide a standard method for formally defining the digital information objects to be transferred by an information Producer to an Archive and for effectively packaging these objects in the form of SIPs. This supports effective transfer and validation of SIP data. The PAIS standard provides an XML implementation to formally define the information to be transferred, and this formal definition can be used to support transfer and validation of that information.

PAIS will enable the Producer to share with the Archive a precise and unambiguous formal definition of the Digital Objects to be produced and transferred. It does this by means of a model and its instantiation as XML files. It will also enable a precise definition of the packaging of these objects in the form of SIPs, including the ability to specify the order in which they should be transferred.

As seen by the PAIS standard, a transfer, is the movement of Data Objects from a Producer to an Archive. The Data Objects are not transferred as independent plain items, but rather grouped and encapsulated in higher level objects known as SIPs, thereby providing better control in term of SIP Content Types, Fixity Information, inter-relationships, and sequencing as outlined in the following figure 2-1.

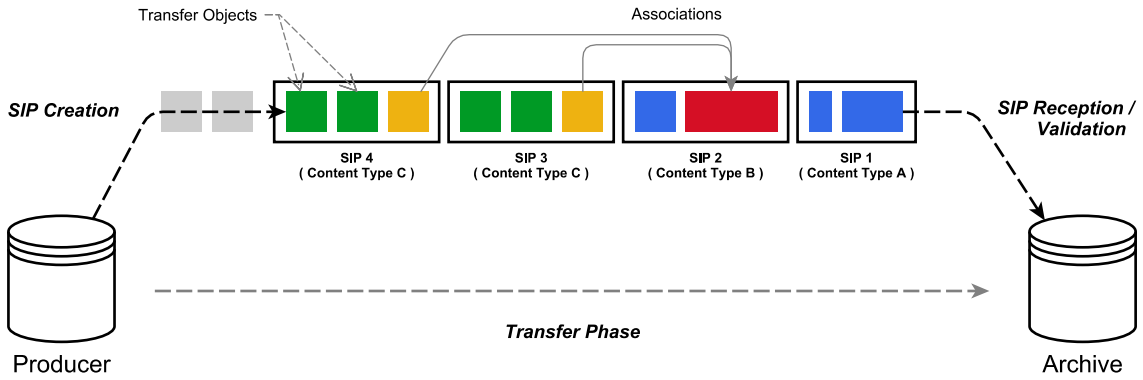

Figure 2-1: Example of Transfer

The Producer is responsible for the creation of SIPs according to content types agreed with the Archive and for their submission in a sequencing order that may also have been negotiated with the receiving Archive. In the example above, the Producer has generated and submitted four SIPs, one of SIP Content Type A, the second of SIP Content Type B, and the remainders of SIP Content Type C. As suggested by their names, the SIP Content Types govern the actual content allowed for a SIP in term of structure and data format. The PAIS standard supports the control of the transfer through the use of the SIP Content Types.

According to the PAIS standard the contents of the SIPs are decomposed into Transfer Objects (depicted as colored boxes in figure 2-1 above) holding one or more trees of Groups (usually denoting folders) organizing the Data Objects that are the subject of the transfer

(usually a single file or a set of files). A typical example of a Transfer Object could be an Earth Observation product composed of various metadata and data files (i.e., the Data Objects) organized in a tree of folders (i.e., the Groups). The PAIS standard supports the control of these objects through the description of their types, namely the Transfer Object Types, Group Types, and Data Object Types.

According to the PAIS standard, the definition of these SIP Content Types is given by a 'SIP Constraints' XML document that specifies the Transfer Object Types and their frequency of occurrence in each SIP Content Type. It can also specify the order in which SIP Content Types are transferred to the Archive. Referring again to figure 2-1, SIP Content Type A may have specified that only 'blue type' Transfer Objects may be present. SIP Content Type B may have specified that both 'blue and red type' Transfer Objects may be present. And finally the SIP Content Type C may require an 'orange type' Transfer Object and one or more 'green type' Transfer Objects. The SIP Constraints document may also have specified that SIPs of type A and type B need to be transferred prior to all type C SIPs. Alternatively, there may be no constraints on the order of transfer.

The PAIS standard provides a mechanism, called Transfer Object Descriptors, to formally define Transfer Object Types. These Descriptors are instantiated as XML documents. A Descriptor specifies a content tree composed of Group Types possibly holding other Group Types and/or Data Object Types. It also supports the specification of occurrences, sizes, and associations between types.

In addition, the PAIS standard specifies the attributes that need to be included in a SIP for the complete typing of all the objects it contains. These attributes link the objects (Transfer Objects, Group Objects, Data Objects) to their specifications within the PAIS Descriptors.

The PAIS standard also defines a default SIP format based on the CCSDS XFDU Recommended Standard (see reference [2]). According to the XFDU standard, the SIPs are containers of any type (e.g., usually a ZIP archive or a root folder), that hold XFDU Data Object files organized as an arbitrary number of nested folders called Content Units. This structure is accompanied by an XFDU manifest XML document that registers all the Data Objects and, when specialized as defined by the PAIS, uniquely identifies their types within a PAIS Producer-Archive Project (i.e., identifies the PAIS Data Object Types, Group Types, Transfer Object Types, SIP Content Type, etc.).

The list of methods for writing PAIS Descriptors is countless and it is likely that none is suitable for all project contexts. Nevertheless, the following workflow gives an overview of the major steps that are usually to be addressed during a Producer-Archive Project definition:

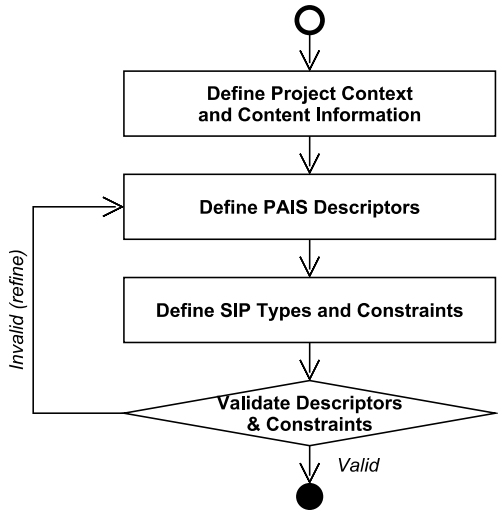

Figure 2-2: Typical Steps Driving a PAIS Producer-Archive Project Definition

In summary, Producer-Archive Projects can increase Archive ingest efficiency and quality control by adopting the PAIS standard. The Producer Archive Project will need to write a set of XML documents according to a formal XML schema to model the transfer. This facilitates validating these Descriptors against XML schema documents provided in the annex of the standard. It also enhances the understanding of the transfer project by both the Producer and the Archive.

If the Producer-Archive Project uses that model and adopts the PAIS standard for packaging the data into SIPs for transfer, it will gain the ingest efficiency and quality control benefits by using the tools for building, transferring, receiving and validating SIPs.

Adopting the PAIS standard involves some shifting of the amount and nature of the effort between the Producer and the Archive in order to gain the benefits. The following two scenarios attempt to typify this shifting and the resulting pros and cons.

Typical ad hoc Producer-Archive submission and ingest scenario:

The Producer and Archive discuss what information is to be submitted, including an estimate of how much data is involved and how it will be sent. A document indicating the format of the data is also exchanged. Typical Pros and Cons of this approach include the following:

Pros:

– There is a minimal effort on the part of the Producer as the data are shipped or electronically transferred in the most convenient form for the Producer, along with the Producer's existing documentation.

Cons:

- There is significant effort on the part of the Archive as it tries to understand unfamiliar documentation and as it prepares software to facilitate ingest of the data into the Archives storage system.
- The Archive attempts to use its understanding of the documentation to perform some level of validation, but this may be limited to expected sizes and manual examination of data structures to see if they match the documentation.
- Neither the Producer nor the Archive can be sure that the data, as ingested into the Archive's system, are correct and properly understood by the Archive so they can be subsequently parsed and delivered, with adequate documentation, to Customers of the Archive.
- Documentation and tracking of the ingest processing will need to be tailored to each ingest project.

Typical PAIS Producer-Archive submission and ingest scenario:

The Producer, using PAIS Descriptor generating software (see 7.2 for a prototype) provided by the Archive, models the information to be submitted, including documentation, in terms of well-defined Transfer Objects. This is electronically reviewed by the Archive and if there are ambiguities or insufficient details, the Archive suggests clarifications and possible additional details. Upon completion of this process, both Producer and Archive have a clear, formally modeled understanding of the information to be submitted. The Archive provides the Producer with an XML-based SIP Constraints template to be altered to indicate the types of SIPs (specifying allowed types of Transfer Objects) to be provided and to indicate any sequencing constraints on the order of types of SIPs. This is also electronically reviewed by the Archive and updates may be requested. Upon agreement, the Archive understands what levels of automated validation it can apply to the ingest activity and whether it wants to employ any manual validation.

The Archive provides the Producer with SIP-generating software (see 7.3 for a prototype) to automate the production of SIPs in accordance with the modeled agreements. This software makes use of the agreed Descriptors and SIP Constraints document to assist the Producer in generating the expected SIPs. The SIPs are transferred to the Archive using a previously agreed mechanism. Typical Pros and Cons of this approach include the following:

Pros:

- There is a clear, formally modeled understanding of the information to be submitted which allows the Archive to do automated validation and ingest into its Archive storage system, greatly reducing the time and effort otherwise involved.
- The Producer is aided in generating the correct information for transfer by the associated software.
- The Producer and Archive are assured that the information transferred is correct and complete.
- The Archive can be confident that its understanding can much more reliably service user requests for extracts of the information.

Cons:

- The Archive needs to prepare and maintain Descriptor and SIP generating software. However, this is reusable across multiple Producers and Producer-Archive Projects, and this effort is offset by the Archive's not having to generate specific ingest software and procedures for each ingest project.
- The Producer needs to take the time to understand the Descriptor and SIP generating software, and to complete the modeling process.

As the Producers will need to make additional effort, it is likely that they will need to be educated as to the benefits to the Producer-Archive Project that will result from adopting the PAIS standard.

3 MODELING TRANSFERS

3.1 INTRODUCTION

The PAIS standard in reference [1] specifies material for the modeling and control of the transfer of Digital Objects from a Producer to an Archive. This material consists of a set of XML language specifications and construction rules supporting the concrete implementation of the Producer-Archive Interface Methodology Abstract Standard (PAIMAS), itself deriving from the Reference Model for an Open Archival Information System (OAIS) (see references [3] and [4]). The XML schemas defined by the PAIS do not intend to implement all aspects of the broad PAIMAS abstract standard but are focused on the specification of the formal model and the associated SIPs during the Formal Definition Phase, the validation and follow-up of the SIPs during the Transfer Phase, and the Validation Phase as defined in PAIMAS.

NOTE – The figures in this section employ UML modeling to give an overview of the PAIS modeling using XML based Descriptors.

From the text in later sections it appears that the MOT is specifically the means for controlling the format and contents of the data that a PAIS compliant project can produce. See highlighted sentence in Sec 3.3. I think this may be clearer than what is stated here.

3.2 MODEL OF OBJECTS FOR TRANSFER

3.2.1 GENERAL

The MOT is a set of PAIS XML Descriptors controlling a Producer-Archive Project and is an instantiation of the PAIMAS formal model. The UML diagram in figure 3-1 shows that the MOT needs to be composed of at least one Collection Descriptor and one Transfer Object Type Descriptor.

The colors used in Fig 3-1, and in subsequen figures, seem to have a meaning, but this is nowhere described.

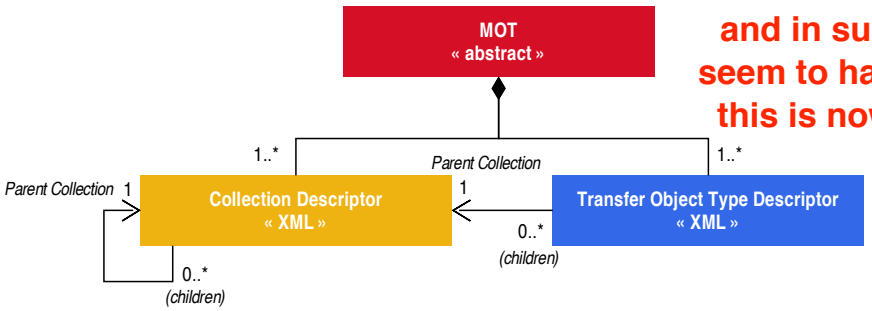

Figure 3-1: Model of Objects for Transfer

Each Transfer Object Type Descriptor models a unit of transfer broken down in a tree of Group Types and sub-Group Types of Data Objects. The Transfer Object Type Descriptor is further described in 3.2.2.

The Collection Descriptors organize the Transfer Object Types in a logical tree with unlimited levels of parent collections. The Collection Descriptor is further described in 3.2.3 below.

According to this definition, the minimal MOT is composed of two XML Descriptors and defines one Transfer Object Type as a part of one collection.

At least one …

3.2.2 TRANSFER OBJECT TYPE DESCRIPTOR

A Transfer Object Type Descriptor defines a Transfer Object Type. Objects conforming to the Descriptor can be transferred from the Producer to the Archive. Each Descriptor defines one and only one Transfer Object Type, and it has to be part of a collection.

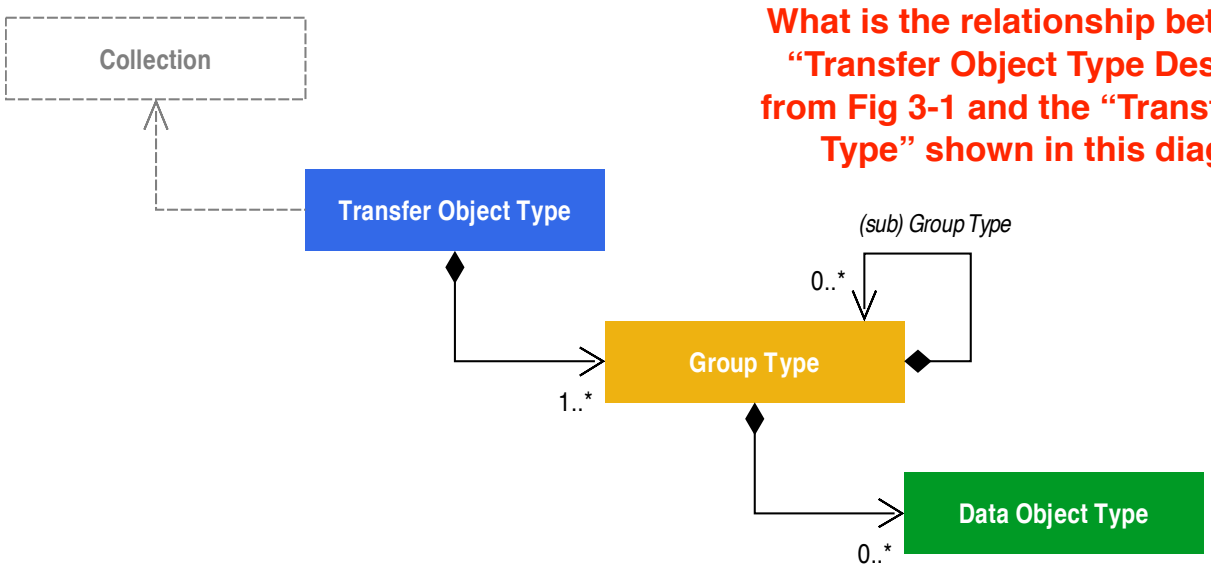

Figure 3-2: Transfer Object Descriptor

A Transfer Object Type is uniquely identified across the overall Producer-Archive Project and decomposes the object into one or more trees of Group Type nodes and Data Object Type leaves. A Transfer Object Type defines the minimal and maximal number of objects, or instances, of this type that could occur in the overall project. It may optionally define the minimal and maximal size of each individual object of this type. The complete specification of the Transfer Object Type is provided in subsection 3.2 of the PAIS standard (see reference [1]) and illustrated throughout section 4 of this report.

A Transfer Object Type needs to contain at least one Group Type, each accepting as many sub-Group Types as required up to the Data Object Type leaves. The Group Type definition specifies the minimal and maximum number of occurrences expected within its parent Group or its parent Transfer Object. The Group Types may correspond to directories or ZIP entries, or they may simply represent the concept of a set or sequence. It is important to note that when there are nested groups with multiple occurrences, the order in which the group instances should be instantiated is ambiguous. This should be addressed by specifying the order semantically in the text description that is a part of each group specification. It could also be addressed by incorporating one or more user defined attributes into the Transfer Object Descriptor. This issue is addressed further in 6.2.4 of the ISEE use case.

The Data Object Type is the lowest level of description in the MOT and usually corresponds to a single file type. It may also represent multiple files if this set of files can be considered as a single Data Object at the transfer level. This feature enables the avoidance of overly

detailed Descriptor specifications. For example, it may be convenient to consider an Earth Observation product as a single Data Object although it is composed of multiple files. The modeling of the header files, the image bands, and other auxiliary files composing this product may not be of interest if they are never disjoint and never referenced individually during the transfer. The Data Object Type specifies the minimal and maximal number of occurrences expected within its instance of parent Group Type. It is also possible to control the minimal and maximal number of files composing a Data Object.

Any of the Transfer Object Types, Group Types, and Data Object Types can be interrelated and/or associated to a collection. The semantics of an association are unconstrained by PAIS and are to be agreed between the Producer and the Archive during the Project definition. Typical semantics include the data/metadata relationship or the association with a Representation Information that may support the cataloguing and the validation activity on the Archive side. These relations are further discussed in 4.6 of this report.

3.2.3 COLLECTION DESCRIPTOR

Each Collection Descriptor defines a single collection. Each Collection Descriptor references a single parent collection. The parent collection attribute of the top-level collection has the value 'none'. Any collection can be referenced as a parent by zero or more other collections and any number of Transfer Object Types. Thus the set of collections defined in a MOT forms a tree.

What is the relationship between the "Collection Descriptor" from Fig 3-1 and the "Collection" shown in this diagram?

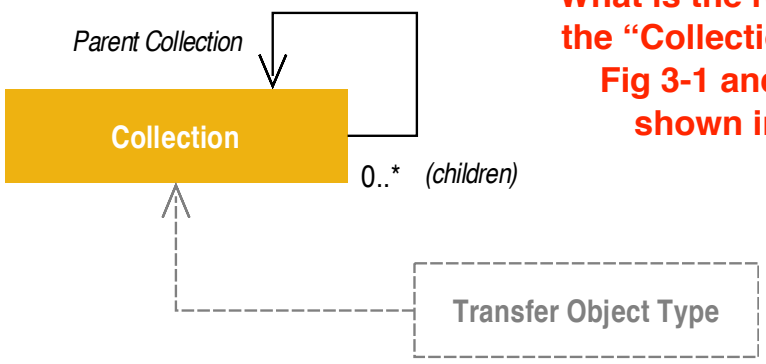

Figure 3-3: Collection Descriptor

A collection is a logical grouping concept whose semantics are to be determined by the Producer and the Archive. For example, the collection may support Archive behavior in terms of cataloguing, validation means, or storage. Typical collections could distinguish project documentation from data production and could group products according to their processing levels, the production phases, etc. The complete specification of collection is provided in subsection 3.3 of PAIS standard (see reference [1]) and is illustrated throughout section 4 of this report.

3.3 SUBMISSION INFORMATION PACKAGE

3.3.1 GENERAL

The SIP is the actual physical unit of transfer that can hold one or more Transfer Objects. The PAIS standard defines a SIP Constraints file and various rules for specifying and controlling the SIP contents based on the MOT elements introduced in the subsections above.

3.3.2 SIP CONSTRAINTS

The SIP Constraints file is an XML document defining all the SIP Content Types that may occur in a Producer-Archive Project. At least one SIP Content Type needs to be defined for the project. If more than one SIP Content Type is defined, then the sequencing of transfers to the Archive for those SIP Content Types can be specified.

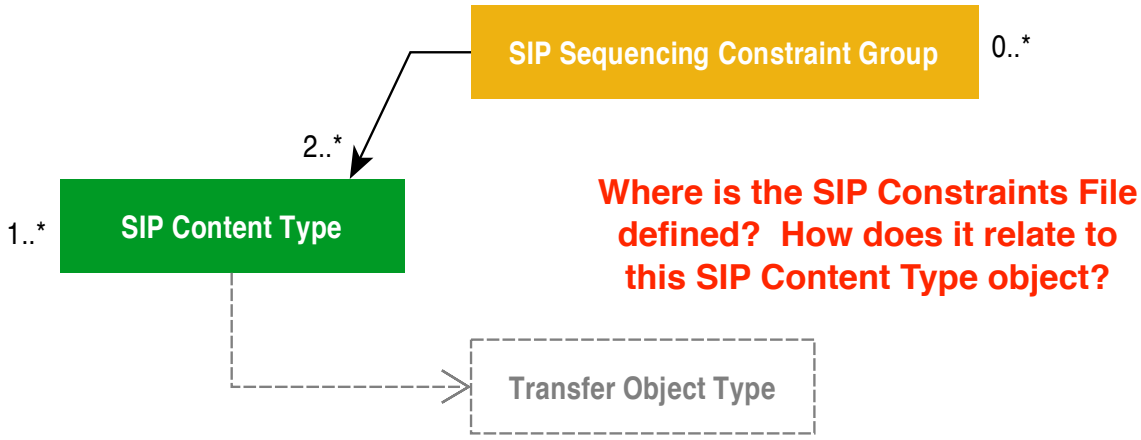

Figure 3-4: SIP Constraints

Each SIP Content Type specifies the Transfer Object Types authorized for this type of SIP and, optionally, the minimal and maximal numbers of occurrences allowed for each SIP instance.

The SIP Constraints file may also define one or more SIP Sequencing Constraint Groups that impose a sequencing order between two or more SIP Content Types. This feature may be convenient to secure, for example, the prior transfer of Representation Information required for the validation of the successive packages.

3.3.3 SIP MODEL

The PAIS standard defines an abstract SIP Model that has to be followed by any PAIS-compliant SIP. Unless specified otherwise, any reference to SIP in this report supposes compliance with the PAIS SIP Model.

TO's to Delete?

According to this SIP Model, a SIP is composed of one mandatory SIP Global Information section and optional Transfer Objects to delete and Transfer Object sections as outlined in the figure 3-5 below. At least one of the two sections is mandatory.

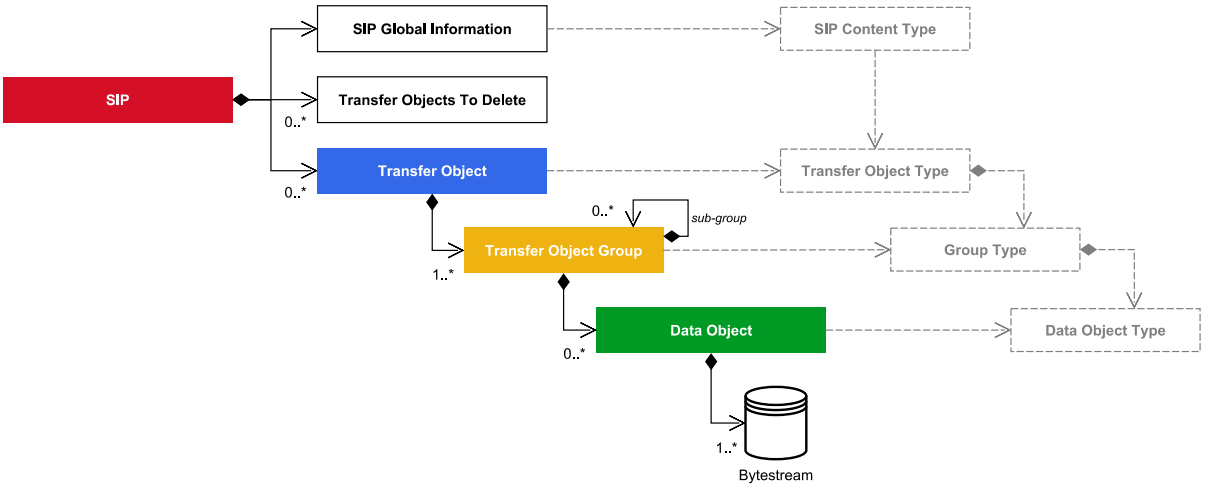

Figure 3-5: SIP Model

The SIP Global Information includes the identification of the SIP with respect to the Producer-Archive Project, the actual source that has generated this SIP and, in particular, a reference to the SIP Content Type it instantiates. The SIP Global Information may also include a sequence number necessary for disambiguation when the MOT Descriptors do not provide fixed occurrence values.

The Transfer Objects are instances of Transfer Object Types of the Producer-Archive Project's MOT. A Transfer Object in a SIP needs to be of a Transfer Object Type authorized by the SIP Content Type referenced by the SIP Global Information. Similarly, the number of Transfer Objects of an authorized type need to be within the range defined in the SIP Content Type.

Following the type hierarchy, the Transfer Objects are composed of Transfer Object Groups that are instances of Group Types, and Data Objects that are instances of Data Object Types. The actual instances need to explicitly reference the corresponding MOT types (i.e., Descriptor specifications) to allow the formal validation of the SIP structure at ingestion.

Finally, the Data Objects reference one or more byte streams typically instantiated as physical files.

3.3.4 RELATIONSHIPS BETWEEN SIP CONSTRAINTS AND MOT

The figure below is an overview of the MOT, SIP Constraints, and links between them defined during the Formal Definition Phase, as detailed in the 3.4.

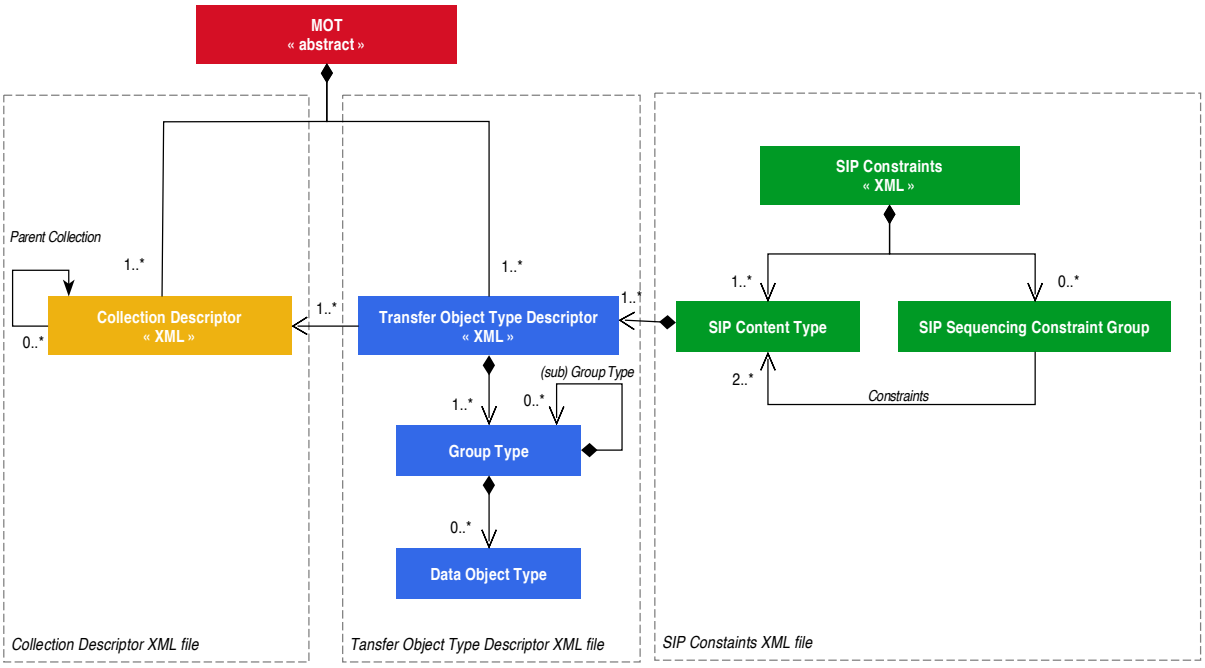

Figure 3-6: Relationships between SIP Constraints and MOT

3.4 A METHODOLOGY FOR MODELING A TRANSFER

As introduced in section 2 above, there are probably countless methods that could apply for defining the PAIS Descriptors, but it is likely that none is suitable for all project contexts. However, the workflow diagram shown in figure 3-7 introduces a typical methodology illustrating major steps that most implementers could follow.

The workflow steps can be summarized as follows:

– **Define Project Context**: according to the PAIMAS standard (see reference [3]), this step represents a possibly long preliminary phase during which the Producer and the Archive converge towards a common understanding of the transfer. At the PAIS level, the critical outputs are the Producer-Archive Project identifier and the potential Producer Source identifiers that will be necessary in the header of all SIPs.

– **Define Content Information**: at this step the Producer and the Archive agree on what kind of Digital Objects are to be transferred. From this step, the workflow forks into two branches involving the Producer and the Archive separately.

Are these colors meaningful in the context of prior figures?

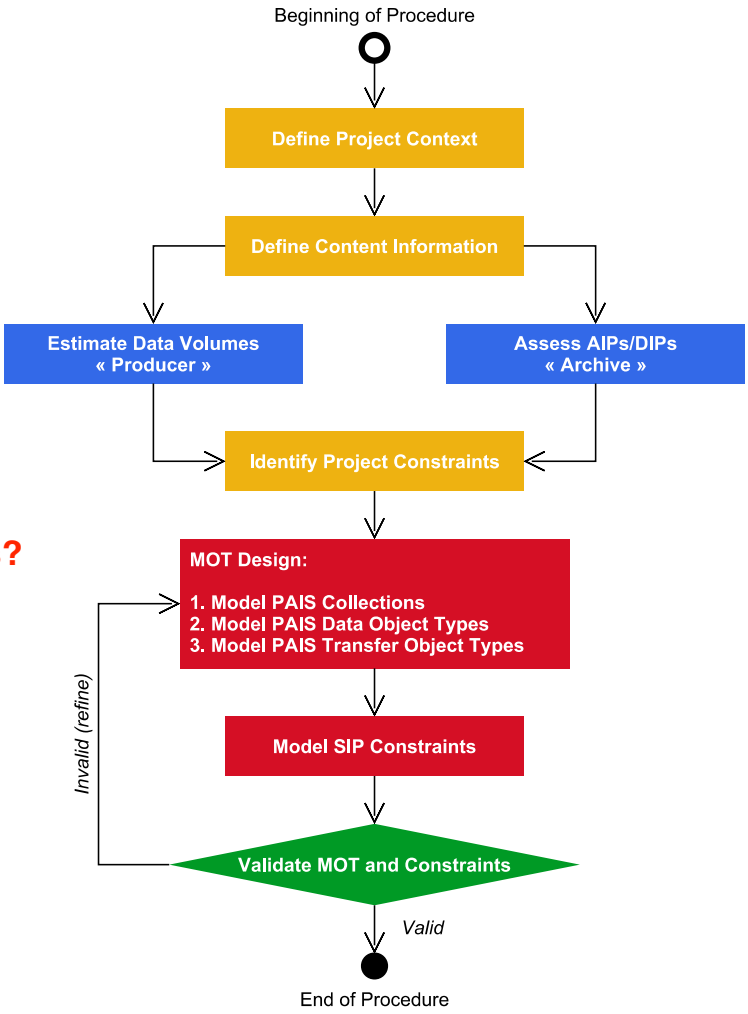

Figure 3-7: A Methodology for Modeling a Transfer

– **Estimate the Data Volumes**: at this step the Producer determines the volumes of each type of Digital Objects. This may depend on the variety of formats, of data sources, etc.

– **Assess AIPs/DIPs**: at this step the Archive has to consider how and where it will host/store the Digital Objects in term of Archival Information Packages (AIPs). A preliminary analysis of the Dissemination Information Packages (DIPs) is usually a good practice for the proper design of AIPs. This step may also identify the validation procedures that will be required at ingestion to populate the Archive and ensure adequate quality.

– **Identify Project Constraints**: this step consists in merging the information collected during the two previous parallel steps and ensuring that all parties can handle the planned transfer, e.g., network resources, hardware, manpower. The objective is to identify the project constraints that can influence the modeling of the objects to be transferred in terms of grouping, sizing, and sequencing.

- **MOT Design**: at this step, all the project information (e.g., Data Objects, formats, collections, etc.) and constraints are known, and it becomes possible to model the transfer and write the PAIS Descriptors for the transfer project, i.e., the Collection Descriptors and the Transfer Object Type Descriptors. It is recommended at this step to model only the elements that are actually required or justified for the transfer. For example, the PAIS does not require Descriptors that may serve as complete Representation Information outside the context of a transfer. Thus, modeling all files and directories because they are present may be more cumbersome than useful and should not be considered as a good practice.

- **Model SIP Constraints**: this step consists in writing the SIP Constraints XML document that defines the SIP Content Types allowed in the transfer and potentially the sequencing constraints; e.g., SIPs holding data specifications or auxiliary data may be required first to allow the validation or the reprocessing of the primary data.

- **Validate MOT and Constraints**: at this decision step the PAIS Descriptors and the SIP Constraints XML documents are validated. They have to be well-formed according to the XML recommendation and validated against the PAIS XML schemas provided in the annex of the standard. It is recommended to complete this automatic validation with reviews involving both the Producer and the Archive and to simulate the transfer of all potential SIP Content Types in an environment as close as possible to the target operational one. In case of failure, the workflow needs to be reiterated from the 'MOT Design' step.

The PAIS standard does not address the management of messages between the Producer and the Archive. Clearly there are many issues that could arise to require various communications. One of these is the possible need to manage updates to the Descriptors and/or to the SIP Constraints during the ongoing transfer of SIPs to the Archive. While this is outside the scope of the PAIS standard, it is recommended that any update to a working Descriptor should result in a new Descriptor ID and therefore a new version of the MOT. This allows the subsequent SIPs to properly reference the relevant Descriptors. An updated SIP Constraints document should be given a version identifier so that the Producer and Archive can communicate clearly as to which version is being used. In the most extreme case, a new Producer-Archive Project could be started. Another example is the exchange of messages for acknowledgment of deliveries and SIPs validation.

4 WRITING XML DESCRIPTORS AND SIP CONSTRAINTS

4.1 INTRODUCTION

This section describes the XML Descriptors composing the MOT (collections and Transfer Object Descriptors). A basic understanding of the MOT concepts introduced in 3.2 above is required. This section includes examples based on the POLDER Data Set, described in 4.2.1.

4.2 STRUCTURES AND CONSTRUCTION RULES

4.2.1 POLDER—A HIGH-LEVEL DESCRIPTION EXAMPLE

The POLDER data example is used to highlight typical points being described in the tutorial. This subsection presents a high-level description of the POLDER data without further details (as this amount of understanding should be sufficient for the readers' understanding of the examples).

POLDER (POLarization and Directionality of the Earth's Reflectances) was developed by the French space agency, CNES.

POLDER is a wide field of passive optical imaging radiometer and polarimeter that has provided the first global, systematic measurements of spectral, directional, and polarized characteristics of the solar radiation reflected by the Earth/atmosphere system. Its original observation capabilities have opened up new perspectives for discriminating the radiation scattered in the atmosphere from the radiation actually reflected by the surface.

The design of the MOT for the POLDER use case, as shown in figure 4-1, involves two Collection Descriptors, one for the raw data called 'L0' and one for the upper-level product called 'L1', and one Transfer Object Descriptor for the documentation.

The L0 collection contains a Transfer Object Descriptor for the raw data '`L0DATA`', the associated Representation Information '`L0REPINFO`', and the auxiliary data '`AUXDATA`' used to build the upper level products '`L1`'. For the example only three products are modeled.

The L1 collection contains three Transfer Object Descriptors, one for the L1 products '`LIG1DATA`', one for the associated browses '`L1GB1DATA`', and one for the L1 products Representation Information '`L1REPINFO`'.

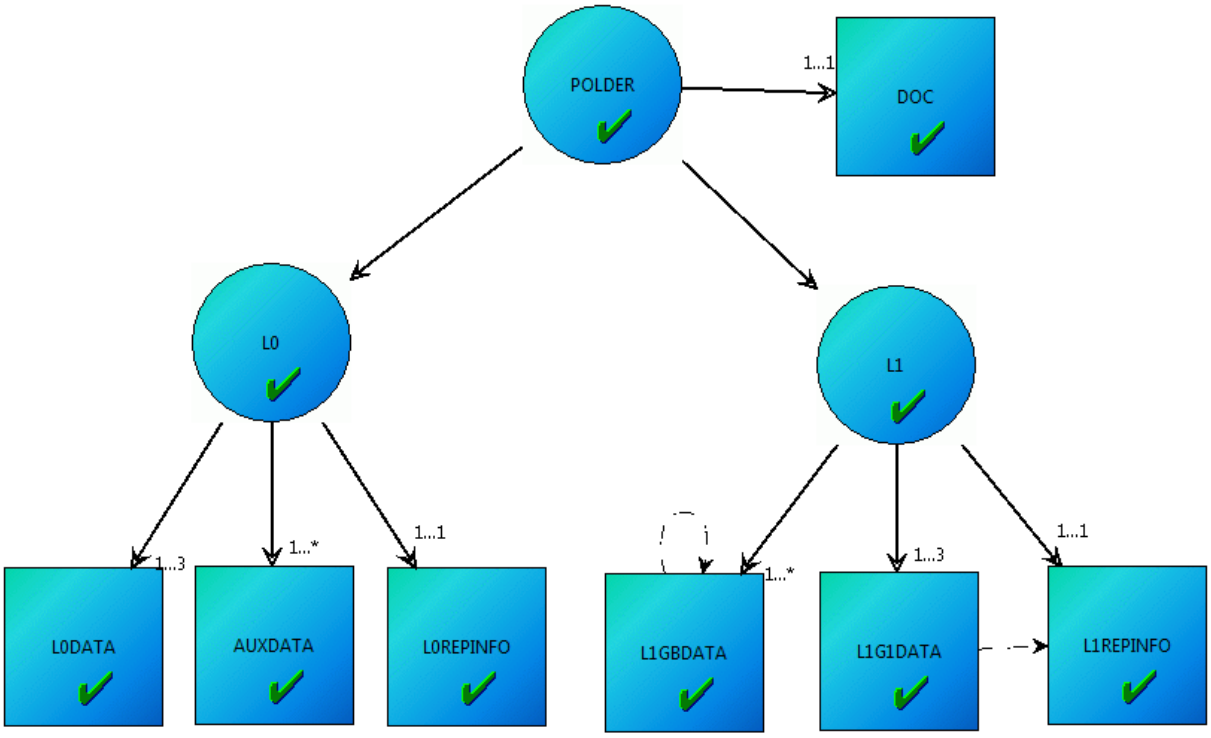

Figure 4-1: POLDER MOT

4.2.2 ORGANIZATION OF XML DOCUMENTS

The physical organization of XML documents is essentially unconstrained.

Figure 4-2 shows an example of a set of XML documents grouped in a directory. Files are named using the naming convention of

`<project-id>-pais-<type>[-<id>].xml`

Where:

`<type>` can either be `collection`, `transfer-object` or `sip-constraints`;

`[-<id>]` is either the `descriptorId` of the collection or Transfer Object Descriptors or does not exist for SIP Constraints files that are unique.

```
├── polder-pais-collection-l0.xml
├── polder-pais-collection-l1.xml
├── polder-pais-collection-polder.xml
├── polder-pais-sip-constraints.xml
├── polder-pais-transfer-object-auxdata.xml
├── polder-pais-transfer-object-l0data.xml
├── polder-pais-transfer-object-l0repinfo.xml
├── polder-pais-transfer-object-l1g1data.xml
├── polder-pais-transfer-object-l1gbdata.xml
├── polder-pais-transfer-object-l1repinfo.xml
└── polder-pais-transfer-object-doc.xml
```

Figure 4-2: Example of PAIS XML Documents in a Directory

4.2.3 XML NAMESPACE

4.2.3.1 General

PAIS Descriptors are standard XML 1.0 documents that should begin with a standard XML prolog:

```
<?xml version="1.0" encoding="UTF-8"?>
```

XML namespaces provide a method to avoid element name conflicts. A so-called namespace needs to be defined. The namespace is defined by the xmlns attribute in the start tag of an element. Any sub-element inherits the xmlns attribute—and so the namespace). The namespace declaration has the following syntax. xmlns[:prefix]="URI" (prefix is optional).

The current version of the PAIS produces elements in the '`urn:ccsds:schema:pais:1`' namespace:

```
Xmlns[:pais]="urn:ccsds:schema:pais:1"
```

This is defined in the CCSDS URN RFC, add as a reference

4.2.3.2 Default Namespace

The '`urn:ccsds:schema:pais:1`' namespace could be used as the default namespace, i.e., non-prefixed PAIS elements:

```
<?xml version="1.0" encoding="UTF-8"?>

<collectionDescriptor xmlns="urn:ccsds:schema:pais:1">
   <identification>
      <descriptorModelID>CCSD0015</descriptorModelID>
      […]
</collectionDescriptor>
```

In this example, the `collectionDescriptor` and all sub-elements *implicitly* belong to the `urn:ccsds:schema:pais:1 namespace`—even if not prefixed.

4.2.3.3 Namespace Prefix

The '`urn:ccsds:schema:pais:1`' namespace also could be used with explicit **namespace prefix**, i.e., preference to '`pais`' prefix, but any other token would work (this may be useful to distinguish PAIS elements from other from different namespaces—only valid for extended PAIS Descriptors):

```
<?xml version="1.0" encoding="UTF-8"?>

<pais:collectionDescriptor xmlns:pais="urn:ccsds:schema:pais:1">
   <pais:identification>
      <pais:descriptorModelID>CCSD0015</pais:descriptorModelID>
      […]
</pais:collectionDescriptor>
```

In this example, the `collectionDescriptor` explicitly belongs to the `urn:ccsds:schema:pais:1` namespace.

4.2.3.4 Default Namespace VS Namespace Prefix

As much as possible, it is recommended to use the **default namespace** (especially because this form is easier to read by humans).

The prolog and namespace declarations will not be repeated along the examples of this section. Tabular representation is preferred.

4.2.4 COLLECTIONS

One XML document per Collection Descriptor needs to be created.

Each collection needs to reference a parent collection, so each Collection Descriptor will include a single `parentCollection` element.

Table 4-1 shows a snippet of the contents of an example of a root Collection Descriptor: POLDER collection. One can identify that this is the root collection since the value of the `parentCollection` element is 'NONE' as is shown at ❷ in the table.

Table 4-1: Example of a Descriptor for a Root Collection

Element	Sample Value
collectionDescriptor ❶	
identification	
descriptorModelID	CCSD0015
descriptorModelVersion	V1.0
descriptorID	POLDER
description	
collectionTitle	POLDER 1 and 2 Products
collectionDescription	POLarization and Directionality of the Earth's Reflectance
Relation	
parentCollection ❷	NONE

Where are these values defined?

All elements in table 4-1 are mandatory. This example corresponds to the minimal set of elements required for a collection. Collection Descriptors accept a few more optional elements but this example would fit most use cases.

The descriptorModelID and descriptorModelVersion uniquely identify the model on which this Descriptor is based. It may be the Descriptor model as given in the PAIS standard, in which case the values are CCSD0015 and V1.0, respectively, as shown in table 4-1. However, the Descriptor model can be specialized, either by CCSDS within a revised standard in which case one or both of the elements will be given new values, or by the Archive in which case the values and specifications need to be managed by the Archive and agreed by the Producer.

This sounds like something that probably should be registered somewhere in the SANA. Is it?

The CCSDS will update only the descriptorModelVersion when the updated schema is backward compatible with the original schema (i.e., changes are simply restrictions of semantics or syntax). Otherwise CCSDS will update the descriptorModelID. In order to maintain unambiguous responsibility for model ownership, only CCSDS is allowed to update the desciptorModelVersion value in association with the CCSDS defined descriptorModelID values of the form CCSDxxxx.

Should an Archive decide to make any changes to the schema as given by the PAIS standard, it needs to use a new value for the descriptorModelID that is not of the form CCSDxxxx. It is recommended that the Archive use the same rule of backward compatibility in the management of its schema's model identifier and version values.

NOTE – All identifiers of the form CCSDxxxx, along with their specification and meaning, can be found in CCSDS standards.

The descriptorModelID and descriptorModelVersion are defined in the PAIS XML schema as unrestricted xsd:string allowing empty strings, tabs, etc. A good practice would be to restrict these elements to the values expected for the Producer-Archive project as described in 4.8 below.

The descriptorID uniquely identifies the collection within the Producer-Archive project.

The descriptorID is defined in the XML schema as an xsd:string which authorizes empty strings or whatever content of any length. As for descriptorModelID, a recommended practice is to restrict the string to a controlled naming space, for example through a pattern or an enumeration valid for the overall project. It should reject empty strings (see 4.8).

The collectionTitle and collectionDescription can greatly help the understanding of the purpose and scope of the collection and should be written with care. Empty strings or the duplication of the collection identifier is to be considered as a poor practice. Formal projects could allow restriction of the PAIS XML schemas to prevent those situations.

Table 4-2 presents an example of a child collection. Its descriptorID value is L0 and it can be referred to as the L0 collection. As shown at ❶, the relation/parentCollection element identifies the parent collection through its descriptorID, POLDER.

Table 4-2: Example of Child Collection

Element	Sample Value
collectionDescriptor	
Identification	
descriptorID	L0
… removed for brevity …	
Relation	
parentCollection ❶	**POLDER**

Only a single parentCollection can be defined. However, other associations can also be defined. A collection can be associated with other collections or other PAIS entities. Additional discussion of object relationships can be found in 4.6.

Collections are logical entities that by default do not have a physical representation in the actual transfer. They are simply referenced by Transfer Objects that are part of the collection in order for the receiving Archive to understand which piece of the transfer they are receiving.

4.2.5 TRANSFER OBJECTS

One XML document per Transfer Object Type Descriptor needs to be created.

Every Transfer Object needs to have a parent collection, so each Transfer Object Type Descriptor will include a single `parentCollection` element.

Unlike Collection Descriptors, Transfer Object Type Descriptors also serve as type definitions for the instantiation of one or more Transfer Objects that are to be incorporated into SIPs and sent to an Archive. Each such Transfer Object within a SIP will include a pointer whose value is the `descriptorID` of the corresponding Transfer Object Type Descriptor, thus linking each Transfer Object in the SIP to its type definition and allowing the recipient of the SIP to understand the Transfer Objects. This is discussed in more detail in 5.2.2.

Table 4-3 is an example of a Transfer Object Type Descriptor for POLDER Level 0 products consisting of one to three Transfer Objects.

Not clear why the source ID is optional

The `identification` section is the same as for the collection `identification` section described previously except for the additional optional element `producerSourceID`.

The `producerSourceID` uniquely identifies an entity of the Producer that has provided the Transfer Object. When used, an XML restriction to an enumeration may be defined to ensure that only valid values are provided. This identification section allows the Archive to uniquely identify the sending Producer entity, should the need for communication arise.

The `description section` is similar to the corresponding section in the collection for the `collectionTitle` and `collectionDescription`. However, because the Transfer Object Type Descriptor is a type definition, the additional `transferObjectTypeOccurrence` element controls the number of occurrences of the Transfer Objects of this type **within the overall Producer-Archive project**. (Refer to 4.4 for further explanations about occurrences.)

The optional `namePreservationRule` is provided to allow a statement as to how the Producer and the Archive should treat the names used for the Data Objects within the Transfer Objects derived from a given Transfer Object Descriptor, including how the names should be preserved by the Archive. In the example of table 4-3, the element's value is the statement 'Producer to use the Source Names'. The Producer understands that 'Source Names' refers to the names used in the Producer's environment. Since each Data Object will exist as one or more files in the Producer's environment, this rule instructs the Producer to use each file's existing name when instantiating each file of the Transfer Objects corresponding to the subject Transfer Object Descriptor. The existence of the rule means the receiving Archive needs to preserve each file name in the received Transfer Objects, although how it does this is at the discretion of the Archive. Most likely it would store the files with their received names. If this element were not included in the Descriptor, technically the Archive could preserve the files using some local naming convention unless instructed otherwise via an external agreement.

The RMP defines both a standard org ID that should be used in this context and also an org role type that should be used. Please see the RMP and align with this now that it exists. It is better than some arbitrary, unmanaged, ID and type.

Table 4-3: Example of Transfer Object Type for a Series of L0 Products

Element	Sample Value
transferObjectTypeDescriptor ❶	
identification	
descriptorModelID	CCSD0014
descriptorModelVersion	V1.0
producerSourceID	CNES
descriptorID	L0DATA
description	
transferObjectTypeTitle	POLDER Level 0 Transfer Object
transferObjectTypeDescription	A single POLDER Level 0 product (TAR)
transferObjectTypeOccurrence	
minOccurrence	1
maxOccurrence	3
namePreservationRule	*Producer to use the source names*
relation	
parentCollection	L0
groupType	
groupTypeID	L0GROUP
groupTypeStructureName	directory
groupTypeOccurrence	
minOccurrence	1
maxOccurrence	1
dataObjectType	
dataObjectTypeID	L0DATAOBJECT
dataObjectTypeOccurrence	
minOccurrence	1
maxOccurrence	1

Previous page says CCSDS0015

As another example of the namePreservationRule, its value may be the statement 'Archive is required to maintain the file name extensions as received'. This may be appropriate for multi-file Data Objects where the file extensions are used to distinguish the file types of the Data Object. This does not constrain the Producer, as in the previous example, but instead the Archive is constrained to preserve the file extensions but not necessarily the file names apart from their extensions. How it does this is at the discretion of the Archive.

As a third example, the Archive may be asked to alter the file names following an algorithm given in the statement of the rule. This may occur when the Producer and the Archive agree that the names used in the Producer's environment are not user-friendly for customers of the Archive. This could cause problems with documentation, so it would have to be carefully addressed.

The `namePreservationRule` is defined in the XML schema as `xsd:string`.

The `relation/parentCollection` element identifies the parent collection through its `descriptorID`. The example in table 4-3 states that there is a parent Collection Descriptor whose `descriptorID` is `L0`, which is the example child collection of table 4-2.

The `parentCollection` is defined in the XML schema as `xsd:string`.

The `relation` section can also contain other elements that allow this Transfer Object Type to be associated with other Transfer Object Types and with other collections. (Refer to 4.6 for further explanations about relations.)

As noted in 3.2.2, a Transfer Object Type is broken down into a tree of `groupTypes` and `dataObjectTypes`. It contains at least one `groupType`.

The `groupType` element and its children are described in 4.2.6.

The `dataType` element and its children are described in 4.2.7.

4.2.6 TRANSFER OBJECTS—GROUP

4.2.6.1 General

Each Transfer Object Type Descriptor needs to have at least one `groupType` element. Each `groupType` element may contain any number of additional `groupType` elements and an unlimited number of `dataObjectType` elements. These elements are convenient for modeling hierarchies such as directory structures.

The groupType is specified in the XML schema and is of type `transferObjectGroupType`.

Table 4-4 provides an example of `groupType` for the POLDER example.

Table 4-4: Example of Group Type

Element	Sample Value
transferObjectTypeDescriptor	
… removed for brevity …	
groupType ❶	
groupTypeID	L0GROUP
groupTypeDescription	*Level 0 Group Type*
groupTypeStructureName ❷	directory
groupTypeOccurrence	
minOccurrence	1
maxOccurrence	1
groupType ❸	
… removed for brevity …	
dataObjectType ❹	
… removed for brevity …	
dataObjectType ❺	
… removed for brevity …	

4.2.6.2 Identification and Description

The groupTypeID is required and uniquely identifies the group, within the scope of the Producer-Archive Project, for reference purposes and is given as L0GROUP in this example.

The groupTypeID is defined in the XML schema as xsd:string.

The optional groupTypeDescription provides the ability to have a meaningful description as to what the group means or how it is being used.

The groupTypeDescription is defined in the XML schema as xsd:string.

4.2.6.3 Structure

The groupTypeStructureName is required and provides a semantic identification as to how the structure of the group is organized.

Although groupTypeStructureName is defined in the XML schema as xsd:string, it is recommended that it be restricted to a limited enumeration. The standard has predefined the four values of 'directory', 'set', 'sequence', and 'undefined', and it is recommended that they be lowercase for interoperability purposes. Their meanings are as follows:

- **directory**: A `groupTypeStructureName` with the value '`directory`' is understood to be a directory structure with a name, corresponding to a physical file system folder or equivalent, as folders in a ZIP or Tar. All instances of `dataObjectTypes` (i.e., Data Objects) and other groups at the same level within this `groupType`, are understood to be within this directory. Table 4-4 shows that the POLDER Transfer Object is viewed as a directory structure.

- **set**: A `groupTypeStructureName` with a value of '`set`' is understood to state that the `groupType` holds an unordered grouping of all the instances of `dataObjectTypes` (i.e., Data Objects) and all the instances of groups at the same level, and nested immediately within this `groupType`. A '`set`' may or may not have a name. For example, the Data Objects could be several files related to observations of an event and there could be an included `groupType`, also specified as a '`set`', holding additional files related to calibration of the observing instrument.

- **sequence**: A `groupTypeStructureName` with a value of '`sequence`' is understood to state that the `groupType` holds an ordered grouping of either all the instances of `dataObjectTypes` (i.e., Data Objects) or all the instances of groups at the same level, and nested immediately within, this `groupType`. It is not permitted to mix groups and Data Objects under the same sequence and therefore, for example, a `groupType` with a `groupTypeStructureName` of '`set`' and containing a `dataObjectType` cannot also contain a `groupType`. As an explicit example, the Data Objects could be images of solar eruptions taken over 20 years and ordered by the size of the eruption from smallest to largest. It is recommended that the optional `groupTypeDescription` be used to specify how the ordering is to be recognized. As an example, for a sequence of Data Objects, they may be named alphabetically in the defined order.

- **undescribed**: A `groupTypeStructureName` with a value of '`undescribed`' is understood to state that the detailed modeling of the `groupType` and all data structures instantiated under this `groupType` in the SIP have intentionally been left undescribed. Such a `groupType` is not allowed to contain any `dataObjectTypes` or other `groupTypes`, although all other elements of `groupType` may be present. This allows the modeling of Transfer Objects of complex structure to be terminated when the PAIS formal modeling effort becomes impractical or other modeling specifications already exist. For example, if the Transfer Object is a complex directory structure, it may be sufficient to model only the first directory level and to treat nested `groupTypes` as '`undescribed`' because specifications for the directory structure exist elsewhere. This could give an Archive sufficient information to recognize and partially validate the receipt of an expected Transfer Object without requiring an onerous modeling effort by the Producer. The Producer needs to select, when creating the Transfer Object for instantiation into a SIP, the data that is to be included in the '`undescribed`' `groupType`. Upon receipt of the SIP and encountering a Transfer Object with an '`undescribed`' `groupType`, the Archive is expected to preserve all data found to be associated with this `groupType`, and to maintain its relationship with the rest of the Transfer Object.

The PAIS puts only two constraints on the mixing of `groupTypeStructureName` values that may be put into a hierarchy of `groupTypes` within a Transfer Object. One exception is the use of '`sequence`' for Data Objects, which use stops any further hierarchy under this `groupType` and therefore any further `groupTypeStructureName` values. The other exception is the use of '`undescribed`' in a `groupType`, which use stops any further modeling of possible hierarchy under this `groupType` and therefore any further `groupTypeStructureName` values.

This modeling flexibility is available to describe the Transfer Object as it will appear in a SIP. It should be noted that it may or may not correspond to the organization of the data in the Producer's environment and it does not require that the described representation be maintained as the actual organization in the Archive environment. In other words, the structures described are the structures to be used for the SIP transfers and need not be the same as the physical implementations at either the Producer or the Archive site. For example, the tree of groups may be interleaved with groups not present in the Producer environment's physical structure. The tree of groups may also skip some levels of the Producer environment's physical structure. This was discussed briefly in 3.3.4 and is present in the use case example of 6.2. However, for convenience, the structures used for transfer may, if desired, correspond to the physical structure at the Producer's site, or may, if desired, match the expected organization at the Archive.

As noted in 3.2, there is ambiguity in the order of instantiation when a tree of groups includes multiple instances of the groups. This can be resolved by incorporating the needed semantics within the `groupTypeDescription`. (See also 6.2 for a use case example.)

4.2.6.4 Occurrences

The `groupTypeOccurrence` is required and specifies the number of instances of this `groupType` that should be present in the Transfer Object. This may be expressed as one, or more, or as a range of values. This number may not be known at the time of specification. In the case of a unique value, the minimal occurrence and the maximal occurrence values will be the same.

The `groupTypeOccurrence` is defined in the XML schema as type `occurrenceType` and is a complexType with the three elements `minOccurrence` of type `xsd:nonNegativeInteger`, `maxOccurrence` of type `xsd:nonNegativeInteger`, and `maxUnknown` of type `anySimpleType`.

Table 4-4 shows an occurrence of the `groupType` of exactly one. A more complete discussion is presented in 4.4.

4.2.6.5 Other elements

Table 4-4 does not show all the PAIS-defined elements that may be present in a `groupType`. The Transfer Object data associated with the `groupType` may be encoded, such as via a compression algorithm. It may be used iteratively to describe possible nested encodings, in which case they are described in the order of application. The result of encoding is a single Data Object instantiated as a single file in the Transfer Object. The possible presence of a Data Object (`dataObjectType`) specification and any additional `groupType` specifications under this `groupType` specification is understood to comprise a detailed model of this group after the encoding has been reversed. With the use of this encoding element, validation of the encoded content may not be possible because the resulting data structure will not carry any of the identifiers associated with the detailed modeling of this group. A `groupType` that is specified to be encoded may also be specified to be 'undescribed'. In this case there is no modeling of the result of reversing the encoding. Encoding of groups is supported using the complex element `groupTypeEncoded` of type `pais:encodingType` as defined in the XML schema. Details are provided in 4.5.

A `groupType` may also establish a relation with other entities within the Producer-Archive project. For example, it could be of interest to establish the relation 'described by' between the top level POLDER directory and a document describing the directory structure, provided in a different POLDER collection (not actually true for the POLDER use case). This is supported using the complex element `groupTypeAssociation` of type `associationType` as defined in the XML schema. Details are provided in 4.6.

A `groupType` may include any number of other `groupTypes`. Table 4-4 shows that the POLDER Transfer Object has nested `groupTypes` consistent with a directory structure.

A `groupType` may also have any number of `dataObjectTypes`. A `dataObjectType` is used to describe a single file, or a few files that are typically used together such as in a standard format. Table 4-4 shows that each `groupType` includes at least one `dataObjectType`. The `dataObjectType` is described in detail in 4.2.7.

4.2.7 TRANSFER OBJECTS—DATA OBJECTS

A Data Object is the lowest level of modeling in the Transfer Object Descriptor and is typically instantiated as a file or groups of files.

Data Objects may be physically transferred, or may be referenced by URLs.

Each Data Object needs to have a parent `groupType` ❷.

The Data Object is specified in the XML schema and is of type `dataObjectType` ❶.

Table 4-5 provides an example of `dataObjectType` for the POLDER example.

Table 4-5: Example of Data Object Type

Element	Content
transferObjectTypeDescriptor	
… removed for brevity …	
groupType ❷	
… removed for brevity …	
dataObjectType ❶	
dataObjectTypeID	L0DATAOBJECT
dataObjectDescription	*Level 0 Data Object*
dataObjectTypeOccurrence	
minOccurrence	1
maxOccurrence	1
dataObjectTypeFileOccurrence ❹	
minOccurrence	2
maxOccurrence	2
dataObjectTypeFormat ❸	
mimeType	application/binary

The dataObjectTypeID and the dataObjectTypeOccurrence are the two required elements for the Data Object Type description:

– The dataObjectTypeID uniquely identifies the file, or group of files, within the scope of the Producer-Archive Project, for reference purposes, and is given as L0DATAOBJECT in this example. It is defined in the XML schema as xsd:string.

– The dataObjectTypeOccurrence is required and specifies the number of instances of this DataObjectType that should be present in the Group. This may be expressed as one, or more, or as a range of values. This number may not be known at the time of specification. In the case of a unique value, the minimal occurrence and the maximal occurrence values will be the same.

The optional dataObjectTypeDescription provides the ability to have a meaningful description as to what the Data Object means. It is defined in the XML schema as xsd:string.

Table 4-5 does not show all the PAIS-defined elements that may be present in a dataObjectType.

The Data Objects may be encoded (the description is the same as for the group encoding—see 4.2.5), and may be related to other elements in the MOT (see 4.6.3 for Association Type description).

If the Data Object is composed of different types of files (for example, a header file along with a set of measurement files), then the `dataObjectTypeFileOccurrence` ❹ should be used to give the number of files expected. The example table 4-5 expresses that each L0 Data Object is made up of two related and different files, a header file and a data file. If the Transfer Object Descriptor includes the namePreservationRule element, then the names of each file belonging to a Data Object of a Transfer Object corresponding to the Descriptor needs to be preserved by the Archive.

The optional `dataObjectTypeFormat` ❸ may also be present (the example table 4-5 expressed that the header and data file are binary files). This information could be used on the Archive side to perform validation, for example.

When a Data Object is composed of different types of files, it is recommended to use the `dataObjectTypeFormat` as follows:

- If the different files have the same format, then the `dataObjectTypeFormat` describes this common value.
- If not, then the format information should be included in a companion related Transfer Object.

NOTE – It is also possible to model the group of files as a Group with a Set of Data Objects, each having its own format information.

4.3 MANAGEMENT OF MOT IDENTIFIERS

MOT identifiers are all elements whose name ends with 'ID' (e.g., `descriptorModelID`, `descriptorID`, `dataObjectTypeID`…). The only exception is `producerSourceID` which is not a MOT identifier—but a reference to an external Producer identifier.

All MOT identifiers have to be unique across the overall Producer-Archive project and therefore across all XML Descriptors composing a MOT.

This is a mandatory constraint because the IDs are required to be referenced **from outside the MOT** and especially from the SIPs. The ID references from outside the MOT are required to uniquely identify the target resource independently from its semantics.

MOT IDs are typed as `xs:string`. There is no recommendation about ID names. Users are free to create any ID, as long as they keep uniqueness of IDs in the MOT. Users should use some methodology for verifying the uniqueness of these ID in the MOT. All platforms and free solutions should be preferred.

4.4 OBJECTS OCCURRENCES AND SIZES

4.4.1 INTRODUCTION

This subsection describes how to control the occurrences of the Transfer Objects, the Transfer Object Groups, the Data Objects, and the Files of Data Objects. It also covers the control of the Transfer Objects size.

4.4.2 OCCURRENCE TYPE

4.4.2.1 General

The occurrence type is a common type that is used by Transfer Objects, Transfer Object Groups, Data Objects, and the Files of Data Objects. Its use is the same regardless of the object type. It allows the user to describe the expected number of objects allowed within the immediate parent grouping. It does NOT define the total number of objects allowed within the SIP. The `transferObjectOccurrence`, `transferObjectGroupOccurrence`, and `dataObjectOccurrence` are required elements, and `dataObjectFileOccurrence` is an optional element.

If an occurrence element is supplied for an object, it will include

– a minOccurrence and a maxOccurrence; or

– a minOccurrence and a maxUnknown.

Table 4-6: Definition of Occurrence Type

Element	Type
`occurrenceType`	
`minOccurrence`	**`xsd:nonNegativeInteger`** ❶
❸ { **`maxOccurrence`**	**`xsd:nonNegativeInteger`**
`maxUnknow`	*`- empty when used -`* ❷

A `minOccurrence` is required to have a value of zero or a positive integer defined by `xsd:nonNegativeInteger` (as shown at ❶ in table 4-6 above). The value represents the minimum number of occurrences allowed for the objects. A zero value for `minOccurrence` indicates that this is an optional object.

> **Implementation Note:** Within the users' computing environment it may be desirable to restrict xsd:`nonNegativeInteger` to a maximum suitable for computer. (See 4.8.4 for more information about how to restrict types.)

The maxOccurrence is required to have a value of zero or positive integer. However, that value needs to be greater than or equal to the minOccurrence value.

A zero value may be confusing and could correspond to a situation where the objects of this type are unexpected, disabled, or denied. This may help during the development of the PAIS Descriptors or during some ad hoc situations.

> **Implementation Note:** Within the users computing environment it may be desirable to restrict xsd:nonNegativeInteger to a maximum suitable for computer. (See 4.8.4 for more information about how to restrict types.) The local community may also decide to restrict the use of zero as a maxOccurrence value.

If it is used, maxUnknown is an empty element as shown at ❷ in the table above. It expresses that the number of occurrences is variable and that there is no upper limit on the number of objects or that the limit is unknown.

The maxOccurrence and maxUnknown are mutually exclusive (as indicated at ❸ in the table above) so only one should be specified in the same parent element.

4.4.2.2 Occurrence Control Use Cases

A number of occurrence examples are presented in the following tables.

Table 4-7: A Bounded Number of Transfer Objects

Element	Value
transferObjectTypeOccurrence	
minOccurrence	26 ❶
maxOccurrence	53 ❷

For the example in table 4-7, a variable number of objects between 26 (as shown at ❶) and 53 (as shown at ❷) inclusive are required to appear in the parent group. If a measurement is taken every week or two for a year and it is known that a measurement will be taken at least every two weeks, between 26 and 53 sets of measurements will be expected over the year.

Table 4-8: An Unlimited Number of Groups

Element	Value
groupTypeOccurrence	
minOccurrence	444 ❶
maxUnkown	*- none -* ❷

For the example in table 4-8, a minimum of 444 (as shown at ❶) objects inclusive are required to appear in the parent group and the maximum number of object is unknown (as shown at ❷).

Table 4-9: A Fixed Number of Data Objects

Element	Value
dataObjectTypeFileOccurrence	
minOccurrence	2 ❶
maxOccurrence	2 ❷

Table 4-9 provides an example where a fixed number of objects appear in the parent group. This is indicated by a minimum value (as shown at ❶) being equal to the maximum value (as shown at ❷). A fixed number of Data Objects might be used if readings were being received from two spacecraft. For this particular mission and the particular science needs, readings from both spacecraft are needed, and readings from a single spacecraft would not be useful. If information is received from a single spacecraft, it should be discarded. So it should be required that two Data Objects be received.

Table 4-10: An Optional Transfer Object

Element	Value
transferObjectTypeOccurrence	
minOccurrence	0 ❶
maxOccurrence	1 ❷

The example in table 4-10 represents a single optional Transfer Object appearing in the parent group. This is defined by a minimum value of 0 (as shown at ❶) and a maximum value of 1 (as shown at ❷).

Table 4-11: A Denied Data Object

Element	Value
dataObjectTypeOccurrence	
minOccurrence	0 ❶
maxOccurrence	0 ❷

The example in table 4-11 is a special case. Although a denied object is allowed by the standard, it is expected that it will rarely be used. This case will likely be used primarily to document that an object exists and to document that it is not intended for transfer to the Archive. A denied object is identified by having a zero value for both the minimum value (as shown at ❶) and the maximum value (as shown at ❷). A possible use for this construct would be to model data on the Producer's site where a file exists that records user access to the set of data and where the Producer wants to ensure that the file is not transferred in order to protect the privacy of the users.

Table 4-12: Invalid Occurrence—Attempted Negative Bound

Element	Value
dataObjectTypeFileOccurrence	
minOccurrence	-1 ❶
maxOccurrence	2

The example in table 4-12 shows a case that should never exist. Providing a negative value (as shown at ❶) for either the minimum value or the maximum value is an error.

Table 4-13: Invalid Occurrence—Attempted Maximum Value Less Than Minimum Value

Element	Value
dataObjectTypeFileOccurrence	
minOccurrence	5 ❶
maxOccurrence	2 ❷

The example in table 4-13 shows another case that should never exist. The maximum value should never be less than the minimum value. In this case an attempt is made to define the minimum value as 5 (as shown at ❶) while defining the maximum value as 2 (as shown at ❷) which is less than the minimum value and therefore is an error.

4.4.3 TRANSFER OBJECT SIZES

The Transfer Object Type Size is a type that provides an optional element that gives an indication of the total size range for Transfer Objects. The size range may not be known at the time of the Descriptor creation and therefore the corresponding element may be omitted.

Table 4-14: Definition of Transfer Object Type Size

Element	Type
`transferObjectTypeSize`	
`minSize`	**`xsd:float`** ❶
`maxSize`	**`xsd:float`** ❷
`unitsType`	**`xsd:String`** ❸ xsd:enumeration value="**KB**" xsd:enumeration value="**MB**" xsd:enumeration value="**GB**" xsd:enumeration value="**TB**" xsd:enumeration value="**PB**"

A `unitsType` is required to have a value that, as shown at ❸ in table 4-14, is one of the strings 'KB', 'MB', 'GB', 'TB', or 'PB' which indicates that the size values provided are in kilobytes, megabytes, gigabytes, terabytes, or petabytes respectively.

> **Implementation Note:** The PAIS Standard does not indicate whether the prefixes represent multiples of 1000 bytes or 1024 bytes. The Producer and Archive should come to an agreement on the meaning of this for the specific Producer-Archive Project.

A `minSize` is required to have a value defined by `xsd:float` (as shown at ❶ in table 4-14 above). A `maxSize` is required to have a value defined by `xsd:float` (as shown at ❷). However, the `minSize` value needs to be less than or equal to the `maxSize` value. If the `minSize` value equals the `maxSize` value, then the size of the Transfer Object needs to be the given size. In both cases the value represents the size in units defined by `unitsType`.

> **Implementation Note:** The `xsd:float` Type allows both positive and negative numbers. Since the PAIS standard does not define the meaning of negative sizes, it may be desirable to restrict `xsd:float` to a non-negative values. (See 4.8.4 for more information about how to restrict types.)

4.5 OBJECTS ENCODINGS

4.5.1 INTRODUCTION

This subsection addresses how to describe the encoding of Transfer Object Groups and Data Objects.

When applied to a Transfer Object Group, using the element `groupTypeEncoded` of type `encodedType`, it is applied to all the data associated with all contained Data Objects and all contained Transfer Object Groups. It results in a single file in the Transfer Object.

When applied to a Data Object, using the element `dataObjectTypeEncoded` of type `encodedType`, it is applied to all the files associated with the Data Object and the result is a single file in the Transfer Object for this Data Object.

4.5.2 ENCODED TYPE

4.5.2.1 General

The `encodedType` is specified in the XML schema. It is a complex type consisting of two elements, `encodingName` of type `xsd:string` and `encodingDescription` of type `xsd:string`, as shown in table 4-15.

Table 4-15: Definition of Encoded Type

Element	Type
`encodedType`	
`encodingName`	**`xsd:string`**
`encodingDescription`	**`xsd:string`**

4.5.2.2 Encoded Type with Groups

As an example, the top level `groupType` of table 4-4 has been altered to include two `groupTypeEncoded` elements. When there are multiple encodings such as in this case, they are to be performed in the order given. Therefore the directory structure associated with the various `groupTypes` will be maintained by the tar function, and the data files associated with the various `dataObjectTypes` will be incorporated accordingly. Once the tar file has been generated, it will be gzip encoded into another file for inclusion in the Transfer Object in a SIP. While it is not a requirement, for this example the descriptive names for the encoding are included along with the mime type expressions.

Table 4-16: Example of Group Type with Encoding

Element	Sample Value
transferObjectTypeDescriptor	
identification	
descriptorID	L0DATA
… removed for brevity …	
groupType ❶	
groupTypeID	L0GROUP
groupTypeDescription	*Level 0 Group Type*
groupTypeStructureName ❷	directory
groupTypeOccurrence	
minOccurrence	1
maxOccurrence	1
groupTypeEncoded	
encodingName	tar
encodingDescription	application/x-tar
groupTypeEncoded	
encodingName	gzip
encodingDescription	application/x-gzip
groupType ❸	
… removed for brevity …	
dataObjectType ❹	
… removed for brevity …	
dataObjectType ❺	
… removed for brevity …	

Upon receipt by an Archive, the associated Transfer Object will include an identifier pointing to L0DATA and therefore the Archive can associate this Descriptor with the encoded data file. However, upon unpacking the encoded data file, which is to be performed in reverse order to that given in the groupType, the resulting directory structure will not include any identifiers of the nested groupTypes or dataObjectTypes that have been modeled. Therefore it may not be possible to make much use of the modeling that was done for groupTypes under the groupType having the encoding elements. Section 5 discusses the instantiation of Transfer Objects within a SIP in more detail.

4.5.2.3 Encoded Type with Data Objects

As an example, the `transferObjectTypeDescriptor` of table 4-4 has been altered to include a `dataObjectTypeEncoded` element under the `dataObjectType` element of the first nested `groupType`.

Table 4-17: Example of Data Object Type with Encoding

Element	Sample Value
transferObjectTypeDescriptor	
identification	
descriptorID	L0DATA
… removed for brevity …	
groupType ❶	
groupTypeID	L0GROUP
groupTypeDescription	*Level 0 Group Type*
groupTypeStructureName ❷	directory
groupTypeOccurrence	
minOccurrence	1
maxOccurrence	1
groupType ❸	
groupTypeStructureName	directory
… removed for brevity …	
dataObjectType ❹	
dataObjectTypeID	L0DATAOBJECT
dataObjectDescription	*Level 0 Data Object*
dataObjectTypeOccurrence	
minOccurrence	1
maxOccurrence	1
dataObjectTypeEncoded	
encodingName	compress
encodingDescription	LZW compression
dataObjectTypeFormat	
mimeType	PDF
registrationInformation	Application/pdf
dataObjectType ❺	
… removed for brevity …	

It states that the data of this dataObjectType, shown to be a single PDF file, is to be encoded using the LZW compression algorithm before being instantiated into the Transfer Object within a SIP. The instantiated Transfer Object will reflect the directory structure of this transferObjectTypeDescriptor, having a top level directory with a single dataObjectType (details not provided for brevity) and a subdirectory with a compressed PDF data file. Section 5 discusses the instantiation of Transfer Objects within a SIP in more detail.

4.6 OBJECTS RELATIONS

4.6.1 GENERAL

The PAIS standard supports directed binary relationships expressed between entities in the MOT using the complex element relation, and it sub-elements, as defined in the XML schema. This element is mandatory in both Transfer Object Descriptors and Collection Descriptors. It contains the mandatory element parentCollection and the optional elements association and any as shown in table 4-18. The any element allows users of the PAIS to define new relation elements. Details on the use of any are discussed in 4.8.

Table 4-18: Definition of Relation

Element	Type
relation	
parentCollection	**xsd:string**
association	**pais:associationType**
any	**pais:extensionType**

4.6.2 PARENT COLLECTION RELATION

A Descriptor's mandatory parentCollection element supports an aggregate view of the collection to which this Descriptor belongs. It is of type xsd:string as defined in the XML schema.

The value of parentCollection needs to be the identifier of the single collection to which the Descriptor (be it Transfer Object Descriptor or Collection Descriptor) belongs. In other words, the value needs to be the descriptorID value of the parent collection. The top level collection, of which there is only one, needs to have a parentCollection whose value is none.

Figure 4-3 provides a schematic view of a MOT, consisting of stylized XML Descriptors, with the parentCollection relations identified.

Figure 4-3: Snippet of MOT with Parent Collection Relations Identified

4.6.3 ASSOCIATION TYPE RELATION

The `pais:associationType` is a complex common type as defined in the XML schema and shown in table 4-19. It is the type assigned to three elements: `association`, `groupTypeAssociation`, and `dataObjectTypeAssociation`. While each is optional, each may be used as often as needed to establish all the relationships that are to be modeled.

Table 4-19: Definition of associationType Relation

Element	Type
associationType	
targetID	**xsd:string**
relationDescription	
relationType	**xsd:string**
relationTextualDescription	**xsd:string**

The element `association` establishes a relationship from the Descriptor (i.e., Transfer Object Descriptor [Transfer Object Type] or Collection Descriptor) in which it is found **toward** a target Descriptor or target entity within a Descriptor. The target entity within a Descriptor is either a Transfer Object Group Type or a Data Object Type.

The element `groupTypeAssociation` establishes a relationship from the Transfer Object Descriptor's Group Type in which it is found toward a target Descriptor or target entity within a Descriptor. The target entity within a Descriptor is either a Transfer Object Group Type or a Data Object Type.

The element `dataObjectTypeAssociation` establishes a relationship from the Transfer Object Descriptor's Data Object Type in which it is found toward a target Descriptor or target entity within a Descriptor. The target entity within a Descriptor is either a Transfer Object Group Type or a Data Object Type.

As the Transfer Object Descriptor is also a type definition, the relationships also apply to the instantiated Transfer Objects, Transfer Object Groups, and Data Objects as applicable.

The `targetID` element has the type `xsd:string` as defined in the XML schema. The value of `targetID` needs to be a valid identifier of one of the allowed targets, either `descriptorID`, `groupTypeID`, or `dataObjectTypeID`.

The complex `relationDescription` element is defined in the XML schema. It is composed of the mandatory `relationType` element and the optional `relationTextualDescription` element.

Table 4-20 shows a partial expansion of the Transfer Object Descriptor `TOD1` from the MOT of figure 4-1 and it highlights the use of the association type of relationships for the three elements `association`, `groupTypeAssociation`, and `dataObjectTypeAssociation`.

Under the `relation` element, the complex `association` element shows a `targetID` value of `COL2`. This states that the relationship is being established from this Transfer Object Descriptor to the Collection Descriptor whose `collectionID` value is `COL2`. This Collection Descriptor is shown in figure 4-1. The relationship direction is always from the entity holding the described relationship toward the targeted entity. The `relationType` is stated to be `Data described by` and the optional `relationTextualDescription` is given as `Data described by targeted collection of documentation`. Although not shown in figure 4-1, table 4-20 assumes that the collection `COL2` has three Transfer Object Descriptors as children and that they are describing different types of documentation related to the data described by Transfer Object Descriptor `TOD1`.

Table 4-20: Example of Transfer Object Type with Associations

Element	Sample Value
transferObjectTypeDescriptor	
identification	
descriptorID	**TOD1**
… removed for brevity …	
relation	
parentCollection	COL1
association	
targetID	**COL2**
relationDescription	
relationType	**Data described by**
relationTextualDescription	**Data described by targeted collection of documentation**
… removed for brevity …	
groupType	
groupTypeID	L0GROUP
groupTypeAssociation	
targetID	**COL2TO-1**
relationDescription	
relationType	**Data related to**
relationTextualDescription	**Data from instrument has targeted instrument description**
groupTypeAssociation	
targetID	**COL2TO-2**
relationDescription	
relationType	**Data related to**
relationTextualDescription	**Data from mission as targeted mission description**
… removed for brevity …	
dataObjectType	
dataObjectTypeID	L0DATAOBJECT
dataObjectTypeAssociation	
targetID	**COL2TO-3**
relationDescription	
relationType	**Data formatted as**
relationTextualDescription	**Data whose format is targeted format description**
… removed for brevity …	

Under groupType, there are two different groupTypeAssociation elements. The first states that this groupType is related via the targetID to an entity whose ID is COL2TO-1. Although not shown, this is the descriptorID of a Transfer Object Descriptor under the COL2 collection. This Descriptor describes text documentation about the instrument used to generate the data associated with the TOD1 Descriptor. The relationType is given as Data related to and the optional relationTextualDescription is given as Data from instrument has targeted instrument description. More precisely it is the modeling of the data that is being related to the modeling of the instrument documentation by the use of the Descriptor, but the relationship carries through to their respective Transfer Objects.

The second groupTypeAssociation element is much like the first; however, its targetID refers to a different Transfer Object Descriptor whose descriptorID is COL2TO-2. This also is not shown in figure 4-1 because of space limitations. This Descriptor describes text documentation about the mission under which the instrument was used to generate the data associated with Transfer Object Descriptor TOD1.

Under dataObjectType, there is one complex dataObjectTypeAssociation element. Using the targetID element, it establishes a relationship from the Descriptor's dataObjectType specification to an entity with the identifier COL2TO-3. This is the descriptorID value of a Transfer Object Descriptor (not shown) that is also a child under the COL2 collection. This Descriptor describes a text document specifying the format of the data for the Data Object Type shown in table 4-20. The relationType is given as Data formatted as and the optional relationTextualDescription is given as Data whose format is targeted format description.

Figure 4-4 is a partial expansion of the MOT snippet of figure 4-1. It shows some of the parentCollection relationships (solid arrows), an association relationship (dotted arrow), and a groupTypeAssociation relationship (dotted arrow).

In summary, the association type relation can be used between any two pairs of the following four Descriptor entities: Transfer Object Descriptor, Collection Descriptor, Group Type, and Data Object Type. Pairings involving Group Type and Data Object Type may even be within the same Transfer Object Descriptor. When the relation involves a Transfer Object Descriptor or one of its sub-entities, the relation extends to the Descriptor's instantiated Transfer Objects.

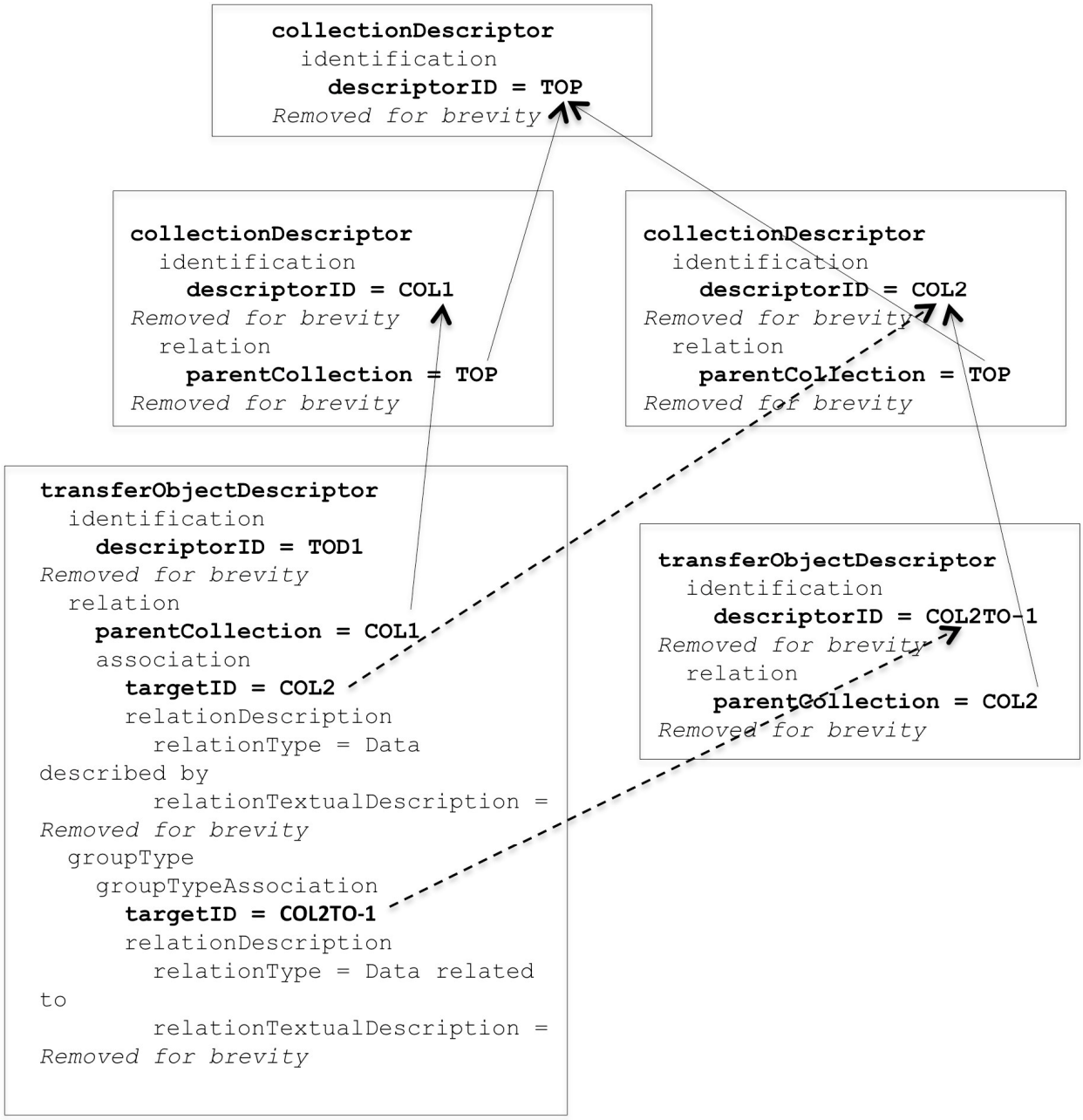

Figure 4-4: Snippet of MOT Showing Parent and Association Relationships

4.7 SIP CONSTRAINTS

There are two types of constraints that apply to the actual SIPs:

a) defines SIP types and defines what data, i.e., Transfer Object Types, appears in each type of SIP:

b) defines the order in which SIPs should arrive at the Archive.

An example of when these constraints would be used is a case where the Producer and Archive have agreed that the Producer will send descriptions of the content in SIPs before the Producer sends the primary content in following SIPs.

Table 4-21: SIP Constraints

Element	Type
sipConstraints	
producerArchiveProjectID	**xsd:string**
sipContentType	
sipContentTypeID	**xsd:string**
authorizedDescriptor	
descriptorID	**xsd:string**
occurrence	**pais:occurrenceType**
minOccurrence	**xsd:integer**
maxOccurrence	**xsd:integer**
maxUnknown	**xsd:string**
sipSequencingConstraintGroup	
groupName	**xsd:string**
constraintItem	
sipContentTypeID	**xsd:string**
constraintSerialNumber	**xsd:integer**

An example of a minimal SIP Constraints file, for which there is only one Content Type and no sequencing constrains, is as follows:

```
<sipConstraints xmlns="urn:ccsds:schema:pais:1">
   <producerArchiveProjectID>MyProject</producerArchiveProjectID>
   <sipContentType>
      <sipContentTypeID>Content Type A❶</sipContentTypeID>
      <authorizedDescriptor>
         <descriptorID>Blue Descriptor ID❷</descriptorID>
         <occurrence>❸
            <minOccurrence>2</minOccurrence>
            <maxOccurrence>2</maxOccurrence>
         </occurrence>
      </authorizedDescriptor>
   </sipContentType>
</sipConstraints>
```

Table 4-22: Example of SIP Constraints Content

Node	Content
sipConstraints	
@xmlns	urn:ccsds:schema:pais:1
producerArchiveProjectID	MyProject
sipContentType	
sipContentTypeID	Content Type A❶
authorizedDescriptor	
descriptorID	Blue Descriptor ID❷
occurrence❸	
minOccurrence	2
maxOccurrence	2

The Content Type A accepts only one Transfer Object Type identified as 'Blue Descriptor ID' ❷. The example also defines that two and only two objects of this type are expected per SIP of this Content Type ❸.

As a second example of a SIP Constraints file, there is the case with two different SIPS: one containing Representation Information and the other one containing the primary data. Representation Information is sent to the Archive prior to sending the primary data. The example file is as follows:

```
<sipConstraints xmlns="urn:ccsds:schema:pais:1">
   <producerArchiveProjectID>MyProject2</producerArchiveProjectID>
   <sipContentType>
      <sipContentTypeID>RepInfo Content Type❶</sipContentTypeID>
      <authorizedDescriptor>
         <descriptorID>IDRepInfo❷</descriptorID>
         <occurrence>❸
            <minOccurrence>1</minOccurrence>
            <maxOccurrence>1</maxOccurrence>
         </occurrence>
      </authorizedDescriptor>
   </sipContentType>
   <sipContentType>
      <sipContentTypeID>Raw Data ContentType❶
       </sipContentTypeID>
      <authorizedDescriptor>
         <descriptorID>IDRawData❷</descriptorID>
         <occurrence>❸
            <minOccurrence>12</minOccurrence>
            <maxOccurrence>366</maxOccurrence>
         </occurrence>
      </authorizedDescriptor>
   </sipContentType>
   <sipSequencingConstraintGroup>
      <groupName>My Single Restraint - RepInfo Before Data
      </groupName>
      <constraintItem>
         <sipContentTypeID>IDRepInfo</sipContentTypeID>
```

```
         <constraintSerialNumber>1</constraintSerialNumber>
      </constraintItem>
      <constraintItem>
         <sipContentTypeID>IDRawData</sipContentTypeID>
         <constraintSerialNumber>2</constraintSerialNumber>
      </constraintItem>
   </sipSequencingConstraintGroup
</sipConstraints>
```

Table 4-23: Example of SIP Constraints Content

Node	Content
sipConstraints	
@xmlns	urn:ccsds:schema:pais:1
producerArchiveProjectID	MyProject2
sipContentType	
sipContentTypeID	RepInfo Content Type❶
authorizedDescriptor	
descriptorID	IDRepInfo❷
occurrence❸	
minOccurrence	1
maxOccurrence	1
sipContentType	
sipContentTypeID	Raw Data Content Type❶
authorizedDescriptor	
descriptorID	IDRawData❷
occurrence❸	
minOccurrence	12
maxOccurrence	366
sipSequencingConstraintGroup	
groupName	My Single Constraint— RepInfo Before Raw Data
constraintItem	
sipContentID	IDRepInfo
sequenceNumber	1
constraintItem	
sipContentID	IDRawData
sequenceNumber	2

The first SIP identified as the RepInfo Content Type ❶ accepts only one Transfer Object Type identified as 'IDRepInfo' ❷. The example defines that exactly one object of this type is expected in this type of SIP ❸.

The second SIP identified as the Raw Data Content Type ❶ also accepts only one Transfer Object Type identified as '`IDRawData`' ❷, since the Raw Data for the example is a year's worth of data and it may be collected daily, but it is required that at least one measurement be made every month. The example defines between twelve and three hundred sixty-six objects of this type are expected per SIP of this Content Type ❸.

The single sequencing constraints group specifies that SIPs identified by the '`IDRepInfo`' ID need to be delivered before any SIPs identified by the '`IDRawData`' ID.

4.8 CUSTOMIZATION—EXTENSIONS AND SPECIALIZATIONS

4.8.1 GENERAL

XML schema is a formal language to describe XML files.

PAIS XML schemas constitute the generic and formal definition of any Digital Objects to be transferred and they assist in creating SIPs.

PAIS XML schemas define the minimum required information that needs to be shared by the Producer and the Archive, to

- produce and package Digital Objects for transfer;
- receive and interpret in an unambiguous manner the transferred Digital Objects.

PAIS XML schemas may be used 'as is' by projects. Projects may also specialize the PAIS XML schemas in order to provide a more precise specification of the model. There are many ways of specializing the PAIS XML schemas (see following subsections), each of them following a single simple rule: the model defined by the specialized schemas needs to be fully compatible with the generic XML schema defined by the PAIS standard; in other words XML files produced using the specialized model (i.e., valid against the specialized schemas) need to be interpretable using the generic model (i.e., valid against the generic schemas). When a PAIS XML schema is specialized, it is required also to be renamed (e.g., <project>-pais-descriptor-collection.xsd is a specialization of ccsds-pais-descriptor-collection.xsd).

The targetNamespace of the specialized schema remains identical to the targetNamespace of the generic PAIS XML schema (for compatibility).

For example, when specializing the Collection Descriptor schema, the `descriptorModelID` is no longer `CCSD0015` since the schema used is a specialized schema, and the specialized schema fixes the value for that element (e.g., `CNES0023`).

Using this example, it is recommended an annotation be added in the specialized schema in the form of:

```
<xsd:annotation>
  <xsd:documentation>
    CollectionDescriptor for <project>, based on ccsds-pais-descriptor-collection schema.
     (descriptorModelID, descriptorModelVersion) = (CNES0023 , 2.1) based on (CCSD0015 , 1.0)
  </xsd:documentation>
</xsd:annotation>
```

4.8.2 PAIS XML SCHEMAS

PAIS XML schemas for Descriptors are composed of three XML schemas:

- ccsds-pais-descriptor-collection.xsd;
- ccsds-pais-descriptor-transfer-object.xsd;
- ccsds-pais-common-types.xsd.

Of these, ccsds-pais-descriptor-collection.xsd specifies a collection, and ccsds-pais-descriptor-transfer-object.xsd specifies a Transfer Object. Both schemas use some common definitions, gathered in ccsds-pais-common-types.xsd.

In addition, there are schemas for the SIP constraints for the implementation of SIPs using the XFDU standard. The PAIS standard (reference [1]) provides the complete view and description of these schemas.

PAIS XML schemas can be found on the official CCSDS SANA XML schemas repository: http://sanaregistry.org/r/daixml

4.8.3 EXTENSIONS

PAIS XML schemas contain special elements, named '`any`'; these elements are extension points, which means that any additional element or attribute may be added at these points, the XML file remaining valid against the generic schema.

Extension points are the opportunity for a project to specify additional information for the transfer of Digital Objects.

The additional part (or fragment) has its own XML namespace, since it has not been defined by the CCSDS. The XML namespace and associated prefix are free; nevertheless, it is recommended a similar rule be used to build the namespace of a project-specific part.

The following text presents a concrete example of an extended schema.

For example, the project POLDER may need to order and date the transferred collections. For that purpose, the following XML namespace is defined: **`urn:cnes:schema:polder:1`**

The additional XML schema part, **`polder-extensions.xsd`** is defined: this schema contains the definition of one (or more) element(s), designed to complete the generic schema. In this example, the element '`collectionID`' is defined to complete the identification part: it is a sequence of a number and a date.

```
<?xml version="1.0" encoding="UTF-8"?>
<xs:schema xmlns:xs="http://www.w3.org/2001/XMLSchema" xmlns:polder="urn:cnes:schema:polder:1" targetNamespace="
urn:cnes:schema:polder:1" elementFormDefault="qualified" attributeFormDefault="unqualified" version="1.0">
   <xs:element name="collectionID">
      <xs:complexType>
         <xs:sequence>
            <xs:element name="order">
               <xs:annotation>
                  <xs:documentation>Collection are transferred in a specific order. First collection has number 1.</xs:documentation>
               </xs:annotation>
               <xs:simpleType>
                  <xs:restriction base="xs:int">
                     <xs:minInclusive value="1"/>
                  </xs:restriction>
               </xs:simpleType>
            </xs:element>
            <xs:element name="date" type="xs:date">
               <xs:annotation>
                  <xs:documentation>Date of production</xs:documentation>
               </xs:annotation>
            </xs:element>
         </xs:sequence>
      </xs:complexType>
   </xs:element>
</xs:schema>
```

The PAIS XML schema is renamed to polder-pais-descriptor-collection.xsd. The schema contains an import link to polder-extensions.xsd.

```
<?xml version="1.0" encoding="UTF-8"?>
<xsd:schema xmlns:xsd="http://www.w3.org/2001/XMLSchema" xmlns:pais="urn:ccsds:schema:pais:1" xmlns:polder="
urn:cnes:schema:polder:1" targetNamespace="urn:ccsds:schema:pais:1" elementFormDefault="qualified" attributeFormDefault="
unqualified">
   <xsd:include schemaLocation="ccsds-pais-common-types.xsd"/>
   <xsd:import namespace="urn:cnes:schema:polder:1" schemaLocation="polder-extensions.xsd"/>
   <xsd:element name="collectionDescriptor">
      <xsd:annotation>
         <xsd:documentation>CollectionDescriptor for POLDER, based on ccsds-pais-descriptor-collection schema
         (descriptorModelID, descriptorModelVersion) = (CNES0023 , 2.1) based on (CCSD0015 , 1.0)</xsd:documentation>
      </xsd:annotation>
```

The `targetNamespace` remains identical to the `targetNamespace` of the PAIS schema.

The any element of the identification structure is redefined. It contains now a reference to the element '`collectionID`'.

```
<xsd:element name="any" minOccurs="0">
  <xsd:complexType>
    <xsd:sequence>
      <xsd:element ref="polder:collectionID"/>
    </xsd:sequence>
  </xsd:complexType>
</xsd:element>
```

The following table presents an example of an extended collection.

Table 4-24: Example of Extended Collection

Element	Content
collectionDescriptor	
identification	
descriptorModelID❹	CNES0023
descriptorModelVersion	2.1
descriptorID	POLDER
any ❶	
polder:collectionID ❷	
@xmlns ❸	urn:cnes:schema:polder:1
polder:order	1
polder:date	2015-06-18

NOTES

1 ❶ allowed only once and accepts only one child ❷ with a namespace ❸ different from the one of the PAIS.

2 ❹ descriptorModelID is no longer CCSD0015 since the schema used is a specialized schema.

4.8.4 RESTRICTIONS

In XML schema implementation, restriction is the standard mechanism for subtyping, or controlling built-in type constraints, e.g., pattern, ranges, etc. Restrictions are very common for XML schemas that are not only based on direct reference to the built-in types.

Another way of specializing the generic PAIS XML schema is therefore to restrict definitions:

– restrict the number of allowed values in an enumeration;

– restrict the range of allowed values for an integer or for a real;

– limit the size of a character string;

– specialize a character string using a pattern (for an unlimited number of allowed values);

– specialize a character string using an enumeration type (for a limited number of allowed values);

– specify the maximum occurrence number of some elements (e.g., `pais:association` may occur 0, 1, or more times, up to ∞ (often set to 'unbounded' in the generic schema);

– delete optional elements;

– specify that an element is required while it is optional in the generic schema;

– change a choice between multiple elements into a sequence of a single element;

– or any other restriction that makes the new schema compatible with the generic schema.

The following table suggests restrictions for some elements of the PAIS XML schema ccsds-pais-descriptor-collection:

Table 4-25: How to Specialize a Collection Descriptor

Element	Restrictions
Collection Descriptor	
`descriptorModelID`	`fixed`❶
`descriptorModelVersion`	`fixed`❶
`descriptorID`	`pattern recommended, length > 1`
`collectionTitle`	`length > 5`
`collectionDescription`	`length > 5`
`minSize`	`Should be positive or null`
`maxSize`	`May be restricted to the actual/total Archive capabilities or any intermediate system that could not handle the size, e.g., ZIP.`
`unitsType`	`restrict enum to minimal set, e.g., MB or GB`
`parentCollection`	
`targetID`	`pattern recommended, length > 1`❷
`relationType`	`restrict to actually used, e.g., representationInformation, dependency, use, etc.` ❸
`relationTextualDescription`	

NOTE – ❶ fixed value may be defined using the '`fixed`' attribute or using a single-value enumeration type.

The following table illustrates ❷:

Table 4-26: Example of Restricted Type—Simple Type—String Patterns

Element	Content
xs:element	
@name	targetID
xs:simpleType	
xs:restriction	
@base	xs:string
xs:pattern	
@value	POLDER_.*
xs:minLength	
@value	10
xs:whiteSpace	
@value	preserve

The following table illustrates ❸:

Table 4-27: Example of Restricted Type—Simple Type—Enumeration Type

Element	Content
xs:element	
@name	relationType
xs:simpleType	
xs:restriction	
@base	xs:string
xs:enumeration	
@value	representationInformation
@value	dependency

The following table suggests restrictions for some elements of the PAIS XML schema ccsds-pais-descriptor-transfer-object:

Table 4-28: How to Specialize a Transfer Object Descriptor

Element	Restrictions
Transfer Object Type Descriptor	
descriptorModelID	fixed❶
descriptorModelVersion	fixed❶
descriptorID	pattern recommended, length > 1
producerSourceID	pattern recommended, length > 1
transferObjectTypeTitle	length > 5
transferObjectTypeDescription	length > 5

Element	Restrictions
Transfer Object Type Descriptor	
minOccurrence	restrict xs:nonNegativeInteger to a type applicable for the project and for the target hardware, e.g., xs:short, xs:int, etc. May constrain to an actual minimum of minimum.
maxOccurrence	restrict xs:nonNegativeInteger to a type applicable for the project and for the target hardware, e.g., xs:short, xs:int, etc. May constrain to an actual maximum of maximum
maxUnknown	deny if not applicable
minSize	Should be positive or null
maxSize	May be restricted to the actual/total Archive capabilities or any intermediate system that could not handle the size, e.g., ZIP.
unitsType	restrict to an agreed enumeration (MB, GB…)
namePreservationRule	deny if not used or restrict to an enumeration
parentCollection	pattern recommended, length > 1
targetID	pattern recommended, length > 1
relationType	restrict to actually used, e.g., representationInformation, dependency, use, etc.❷
relationTextualDescription	free text (no restriction)
groupTypeID	pattern recommended, length > 1
groupTypeDescription	free text (no restriction)
groupTypeStructureName	restrict to an agreed enumeration (directory, set, sequence…)
encodingName	restrict to an agreed enumeration (zip, tar…)
encodingDescription	free text (no restriction)
dataObjectTypeID	pattern recommended, length > 1
mimeType	restrict to an agreed enumeration (text/xml, image/jpeg…).
registrationAuthority	deny if not accepted or restrict to an agreed enumeration
registeredID	deny if not accepted or restrict to an agreed enumeration or at a minimum, a pattern (not recommended)

Each project should adopt a common policy for restricting all ID values, e.g., at minimum one character, printable characters, pattern.

5 BUILDING AND MANIPULATING SIPS

5.1 INTRODUCTION

The generation of the MOT, consisting of all the XML Descriptors for a given Archive Project, together with the XML constraints document, establishes the basis for the generation of the SIPs. These XML documents are used by the Producer, in conjunction with the actual data files in the Producer's environment, as input to the SIP building process.

The SIP building process needs to take the XML modeling information and create a conforming mapping from the Producer's data to actual SIPs. This mapping is instantiated, conceptually, by a combination of SIP building software and the information supplied by the Producer's agent who is using the software. This mapping is not explicitly addressed by the PAIS. However, the PAIS provides an abstract view of resulting SIPs that incorporate the required Transfer Objects along with other required information, such as a set of global SIP attributes and whether any previously sent Transfer Objects are to be deleted.

The following subsections address the abstract SIP (5.2), the PAIS specified SIP instantiated using the XFDU packaging mechanism (5.3), and the abstract SIP instantiated by a non-PAIS specified packaging mechanism (5.4).

5.2 UNDERSTANDING THE PAIS ABSTRACT SIP

The PAIS abstract SIP is a specification that has been registered with the CCSDS and has been given the identifier 'CCSD0017'. The fact that the SIP specification is registered with CCSDS simply means that a CCSDS document exists that defines this specification. Definition of an abstract SIP is significant because the PAIS allows for the instantiation of SIPs using different physical SIP packaging mechanisms. While the PAIS Recommended Standard specifies the XFDU packaging mechanism and details how this should be adapted, a user-defined specification using some other packaging mechanism is also possible. The user-defined specification should reference the PAIS abstract SIP specification CCSD0017 as its foundation.

As the PAIS standard notes,

> SIPs carry the data, or pointers to the data, being transferred to an Archive. The abstract SIP, or SIP Model, is an abstraction that puts constraints on all possible SIPs. It conceptually conveys one or more complete Transfer Objects. It also conceptually conveys a number of attributes about the SIP. The framework for this SIP model is based on the concept of containers. The SIP Model is a container that holds any number of internal containers which themselves may have containers, and so on, thus supporting multiple hierarchies of containers. A container may also hold attributes about itself.

The PAIS provides several diagrams with text detailing the SIP container and the allowed containers within a SIP. These containers are identified as follows:

- A SIP container holds:
  - one SIP Global Information container;
  - any number of Transfer Object containers; and
  - any number of Transfer Object to Delete containers.
- A Transfer Object container holds:
  - one Transfer Object Identification and Status container; and
  - one or more Transfer Object Group containers.
- A Transfer Object Group container holds:
  - Transfer Object Group Identification container;
  - any number of other Transfer Object Group containers; and
  - any number of Data Object containers.
- A Data Object container holds:
  - Data Object Identification container; and
  - One or more Byte Stream containers.

A summary diagram from PAIS, presented in figure 5-1 below, shows the containers and their relationships. The containers and their attributes are discussed below.

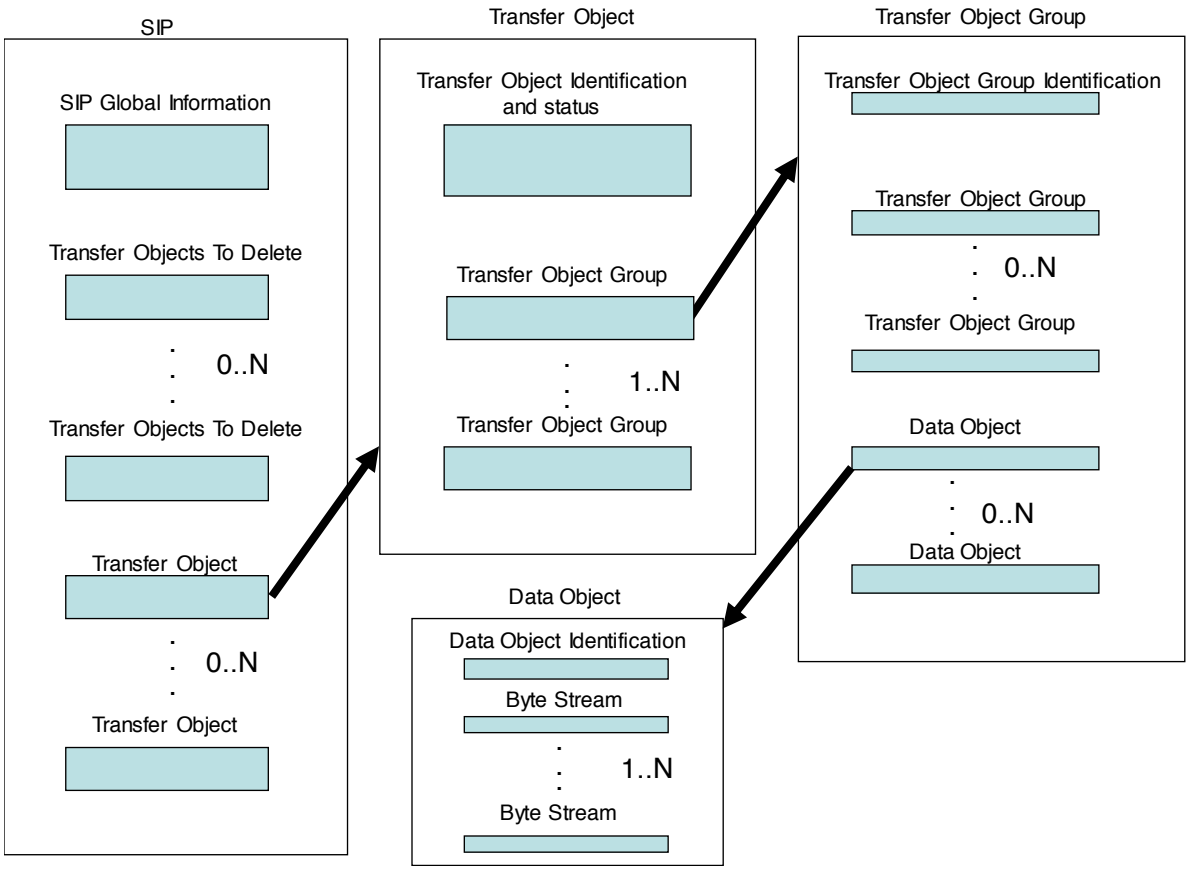

Figure 5-1: Abstract View of SIP, Transfer Object, Transfer Object Group, and Data Object

5.2.1 SIP CONTAINER

5.2.1.1 General

The SIP container holds any number of Transfer Object containers, any number of Transfer Object to Delete containers, and one SIP Global Information container.

5.2.1.2 SIP Global Information Container

The SIP Global Information container holds a number of attributes as follows:

SIP ID

The SIP ID is a mandatory attribute that needs to have a unique value across all SIPs within a given Producer-Archive Project. If there is a single entity within the Producer's environment that is creating and sending SIPs, then this identifier could be simply a sequence number. However, if there are multiple entities within the Producer's environment creating and sending SIPs, then this identifier could be made unique by prepending a unique identifier of

each sending entity to a sequence number generated by each sending entity. Each sending entity is referred to as a 'Producer Source'. Both the Producer and Archive need to agree on the form of this identifier. A unique SIP ID allows checking for duplicate submissions and provides a common identifier for the Producer and Archive if they need to communicate about this submission.

Producer Source ID

The Producer Source ID is a mandatory attribute that uniquely identifies the sending entity within a given Producer-Archive Project. Should the Archive need to contact the sending entity of a given SIP, for example to resolve a problem, this identifier makes clear which of several possible sending entities originated that SIP. Both the Producer and the Archive need to agree on the form of this identifier.

Producer-Archive Project ID

The Producer-Archive Project ID is a mandatory attribute assigned by the Archive to ensure uniqueness across all such Producer-Archive Projects involving the Archive. Its presence within a SIP enables the Archive to uniquely identify the project to which a SIP belongs and thus to identify the MOT and constraints document applicable to that SIP.

SIP Content Type ID

The SIP Content Type ID is a mandatory attribute that enables the identification, within the constraints document, of the specification as to the types of Transfer Objects allowed in the SIP and their frequency of occurrence. Its value should be checked by the Archive upon receipt of the SIP. It also allows the Archive to check as to whether receipt of the SIP has violated any constraints on the order in which the SIPs are to be received.

SIP Sequence Number

The SIP Sequence Number is an optional attribute that indicates the order in which the SIPs have been sent. It needs to be unique within the context of a given Producer Source. When a given Producer Source is sending Transfer Objects whose Descriptor does not specify the number of Transfer Objects to be sent, this attribute becomes mandatory for all its SIPs to enable the Archive to check that it has not missed any Transfer Objects.

Any other attributes

The abstract SIP allows the user to define additional attributes that may be included as Global Information. There are many possible reasons to define additional attributes. Some examples might be for a Producer to track the individual who generated this SIP or to identify an agreement or contract between the Producer and the Archive under which this SIP falls.

5.2.1.3 Transfer Object to Delete container

The Transfer Object to Delete optional container holds one or more attributes as follows:

Transfer Object to Delete ID

The Transfer Object to Delete ID is a mandatory attribute giving the value of the Transfer Object ID for a Transfer Object that should be deleted by the Archive. Any number of such attributes may be present.

Any other attributes

The abstract SIP allows the user to define additional attributes that may be included in the Transfer Object to Delete container. This may include attributes such as the reason for deletion and authorization details regarding the deletion.

5.2.2 TRANSFER OBJECT CONTAINER

5.2.2.1 General

The Transfer Object container holds one Transfer Object Identification and Status container and one or more of Transfer Object Group containers.

5.2.2.2 Transfer Object Identification and Status Container

The Transfer Object Identification and Status container holds a number of attributes supporting the unique identification of this Transfer Object, identification of the associated Descriptor, identification of whether this Transfer Object is to replace a previously sent Transfer Object, and identification of whether this is the last Transfer Object of this type that the Producer Source will be transferring, as follows:

Descriptor ID

The Descriptor ID is a mandatory attribute that identifies the Transfer Object Type Descriptor that describes this Transfer Object. It is obtained from the MOT. Its presence allows the Archive to compare the received Transfer Object with its specification (type definition) as agreed between the Producer and the Archive.

Transfer Object ID

The Transfer Object ID is a mandatory attribute that uniquely identifies this Transfer Object within the Producer-Archive Project. The form of the identifier needs to be agreed between the Producer and the Archive. For example, it could be constructed by prepending the unique SIP ID to a sequence number generated by each Producer Source entity. It could also be constructed by concatenating the Producer Source ID, the Descriptor ID, and a sequence number generated by each Producer Source entity.

Last Transfer Object Flag

The Last Transfer Object Flag is an optional attribute indicating that this is the last Transfer Object of this type (i.e., of the associated Descriptor) that this Producer Source entity expects to send. It is particularly useful when the total number of Transfer Objects of a given type, to be sent, has not been specified in the associated Descriptor. In this case there needs to be a SIP Sequence Number, and, if there is a single Producer Source entity sending Transfer Objects, this identifier allows the Archive to determine when all of the Transfer Objects of this type have been received. However, if there are multiple Producer Source entities sending SIPs, it only indicates that this Producer Source sending entity does not expect to send any additional Transfer Objects of this type. There may be another Producer Source entity that may, or may not, be sending additional Transfer Objects of this type, and thus the Archive may not know when it has received all such Transfer Objects. In this case additional Producer-Archive communications or agreements will be needed.

Replacement Transfer Object ID

The Replacement Transfer Object ID is an optional attribute stating that the value of this attribute is the Transfer Object ID of a previously sent Transfer Object that the Archive is to replace with this Transfer Object.

Any other attributes

The abstract SIP allows the user to define additional attributes that may be included in the Transfer Object Identification and Status container. Additional attributes that could be added would be things like who approved the Transfer Object as being ready for archiving or information on the date the Transfer Object was last updated.

5.2.3 TRANSFER OBJECT GROUP CONTAINER

5.2.3.1 General

The Transfer Object Group container holds any number of additional Transfer Object Group containers, any number of Data Object containers, and one Transfer Object Group Identification container.

For each Transfer Object Group Type described by a Transfer Object Group Type specification within a Descriptor, there needs to be one or more Transfer Object Group Type containers in the SIP. Generally there will be multiple containers when there are multiple instances of the group type. However, there are two special cases that are exceptions, as follows:

Case 1:

When a Transfer Object Group Type specification includes the Transfer Object Group Type Encoded attribute, and regardless of any other attributes, the abstract SIP will have a Data Object container instead of a Transfer Object Group Type container. The instantiation of the Data Object container in the actual SIP will be a single file.

Case 2:

When a Transfer Object Group Type specification does NOT include the Transfer Object Group Type Encoded attribute, but its Transfer Object Group Type Structure Name has the value 'undescribed', then this Transfer Object Group Type specification results in the abstract SIP's having some combination of Transfer Object Group containers and Data Object containers that was not described when the Transfer Object Group was modelled. A data structure in the Producer's environment that has been modeled in the Descriptor as a group type that is 'undescribed' with no encoding is to be fully instantiated in the SIP. The Transfer Object Group containers are used for directories or for holding groupings of files, and the Data Object containers are used for files. The Producer needs to ensure that the correct data are included in the SIP for each 'undescribed' group. The Archive cannot use the associated Descriptor to completely verify that data as it has not been fully modeled.

5.2.3.2 Transfer Object Group Identification Container

The Transfer Object Group Identification container holds a number of attributes identifying the type of the group and optionally naming the group instance.

Associated Descriptor Group Type ID

The Associated Descriptor Group Type ID is a mandatory attribute that identifies the associated group type description within the associated Descriptor.

- When the group type description is not specifying that it is '`undescribed`', then this identifier's value is the Transfer Object Group Type ID of the subject Descriptor's group type specification.
- When the group type description is specifying that it is '`undescribed`', then this same identifier value is used for the Associated Descriptor Group Type ID not only for this group container, but for all nested group containers and nested Data Object containers. An example of this situation is shown in annex A titled 'Associated Descriptor Data Identifiers'.

There is a choice of which of the next two attributes—Transfer Object Group Instance Name or Transfer Object Group Preservation Name—is used. The primary difference between the two is that if Transfer Object Group Preservation Name is used, then the Archive should maintain that name with the group.

Transfer Object Group Instance Name

The Transfer Object Group Instance Name is an optional attribute that may be inserted by the Producer to name the group, such as to provide a directory name or to name a set of Data Objects or other groups. If this attribute is used, the Transfer Object Group Preservation Name may not be used.

Transfer Object Group Preservation Name

The Transfer Object Group Preservation Name is an optional attribute that may be inserted by the Producer to name the group and to indicate to the Archive that this name needs to be preserved with the group. If this attribute is used, the Transfer Object Group Instance Name may not be used.

Any other attributes

The abstract SIP allows the user to define additional attributes that may be included in the Transfer Object Group Identification container.

5.2.4 DATA OBJECT CONTAINER

5.2.4.1 General

The Data Object container holds one or more Byte Stream containers and a single Data Object Identification container.

5.2.4.2 Data Object Identification Container

The Data Object Identification container holds one or more attributes identifying the type of Data Object and optionally supplies a name that is to be preserved in association with the byte stream or streams.

Associated Descriptor Data ID

The Associated Descriptor Data ID is a mandatory attribute that identifies a part of a Descriptor that is to be associated with this Data Object. The part of a Descriptor that is to be used depends on several factors as follows:

- If this Data Object is an instance of a Data Object Type defined in a Descriptor, then the Associated Descriptor Data ID should be the value of the Data Object Type ID of that Data Object Type specification.
- If this Data Object is an instance of a Transfer Object Group Type defined in the Descriptor to be encoded and thus the group becomes a single Data Object, then the Associated Descriptor Data ID should be the value of the Transfer Object Group Type ID of that Transfer Object Group Type specification.
- If this Data Object is an instance that is transferred within the context of a Descriptor-defined Transfer Object Group Type whose Transfer Object Group Type Structure Name has the value ‘undescribed’, then the Associated Descriptor Data ID should be the value of the Transfer Object Group Type ID of that Transfer Object Group Type specification.

An example is given in annex A, titled ‘Associated Descriptor Data Identifiers’.

Data Object Preservation Name

The Data Object Preservation Name is an optional attribute that may be inserted into the SIP by the Producer to tell the Archive that this name is to be preserved in association with this Data Object. When the Data Object is a single file, this name is in addition to the name of the file in the Transfer Object. When the Data Object is composed of multiple byte streams, or files, the name is associated with all of them. The use of Data Object Preservation Name provides a second name apart from the names of the files in the Transfer Object. Therefore there is no conflict if the associated Transfer Object Descriptor has included the optional namePreservationRule element giving a rule for how the Data Object names (i.e., file names) are to be constructed by the Producer or possibly altered and subsequently preserved by the Archive.

An example is a case in which each Data Object consists of a single file composed of a day's observations of magnetic field values taken at 60-second-averaged intervals and has been given a file name of the form spacecraft_mag_julianday.asc. The Producer may decide that it is desirable to associate a more user friendly name, such as spacecraft_mag_year_day.asc, with each Data Object by inserting the Data Object Preservation Name attribute, with this value, into the SIP. The Archive would be required to preserve this name in association with the Data Object.

Another example is a case in which each Data Object consists of two files with different file extensions taking the forms spacecraft_mag_julianday.asc and spacecraft_mag_julianday.bin. These files contain data averaged over 60 seconds during each day. However, because of gaps in the data, the mid-point of a day does not always conform to the mid-point of the observations. The Producer may decide that it is desirable not only to have a single preserved name for the pair of files but also to indicate a weighted mid-point of the observations within each day by constructing a name of the form spacecraft_mag_year_day_hour. The Producer would insert this value, for the Data Object Preservation Name attribute, into the SIP. Again, the Archive would be required to preserve this name in association with multi-file Data Object. There is no requirement in the PAIS standard that the Data Object Preservation Name value be unique, but it is recommended that it be unique at least within the context of each Transfer Object.

Any other attributes

The abstract SIP allows the user to define additional attributes that may be included in the Data Object Identification container.

5.2.4.3 Byte Stream Container

The Byte Stream container specifies a number of attributes in order to provide a single byte stream within the SIP, or to provide a pointer to a byte stream outside the SIP, or both. If both mechanisms are used, the Producer and Archive need to agree on the relationship between the two.

Byte Stream

The optional Byte Stream attribute is the byte stream that will be instantiated as a single file within the SIP.

Byte Stream Checksum

The optional Byte Stream Checksum attribute provides a checksum value covering the stream of bytes. Its instantiation within the SIP will also require identification of the checksum algorithm, such as MD5.

Pointer to Byte Stream

The optional Pointer to Byte Stream attribute provides a mechanism to point to a byte stream outside of the SIP. For example, it may be instantiated in the SIP as a URL.

Any Other Attributes

The abstract SIP allows the user to define additional attributes that may be included in the Byte Stream container.

5.3 XFDU SIPS

5.3.1 GENERAL

The previous subsections described an abstract view of the SIP. Any SIP implementation should include an implementation of all the attributes and requirements discussed in that section. The PAIS standard proposes one implementation of the recommended SIP Model based on the CCSDS XML Formatted Data Units (XFDU) packaging standard (see reference [2]). The XFDU implementation adopts the use of standard XFDU mechanisms to convey some of the attributes and uses XFDU's extensibility to include the remaining attributes.

As a brief introduction to the most relevant features of this standard, an XFDU package is a container, usually a ZIP archive file, composed of one manifest XML file referencing all the other files of the package. The manifest file consists of a series of sections among which are a Package Header containing general information, an Information Package Map providing a logical break-down of the package content in a tree of nested Content Units, and an XFDU Data Object Section referencing all the packaged files.

The XFDU PAIS SIP definition makes use of the XFDU extendibility mechanisms to implement the SIP Model introduced in the previous section. This implementation is depicted in figure 5-2 and is completely defined in the PAIS standard (see section 6 of reference [1]). The XFDU standard (see reference [2]) defines the base XFDU structures. The PAIS standard provides a number of rules for how and where the abstract SIP items appear in an XFDU instance (see section 6 in reference [6]) and it defines several XML schema snippets (see annex E in reference [6]) that fully define the additional XML elements that allow for the inclusion of PAIS information within the XFDU document.

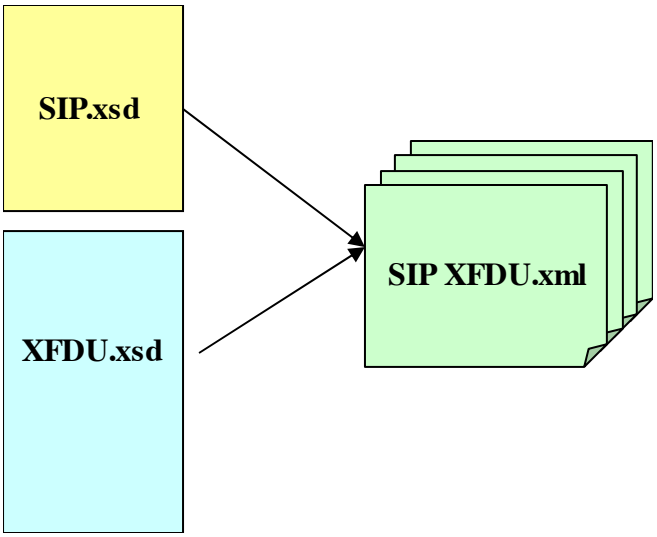

Figure 5-2: Combination of XFDU Schema and PAIS Schema to Form XFDU SIP Schema

This resulting extended XFDU implementation is depicted in figure 5-3 below.

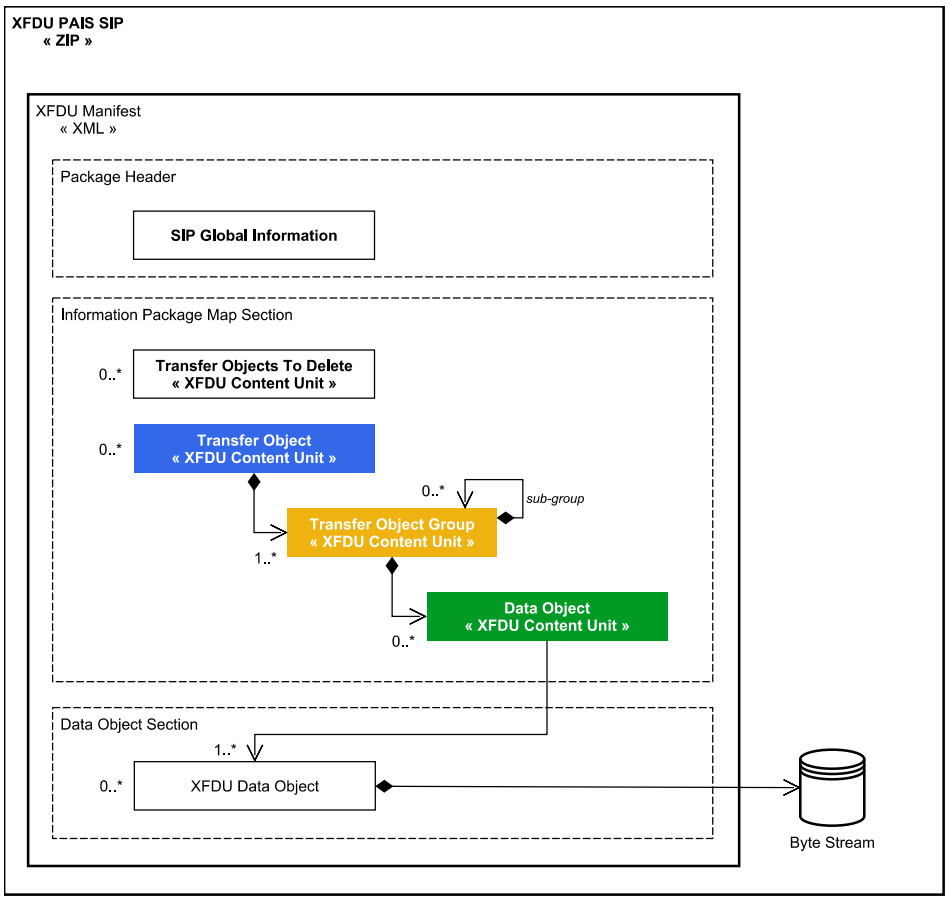

Figure 5-3: XFDU PAIS SIP

The SIP Global Information is implemented as an extension of the Package Header section of the XFDU manifest.

The Transfer Objects to Delete are implemented as an extension of a Content Unit of the Information Package Map Section of the XFDU manifest.

The Transfer Objects and their Transfer Object Groups and Data Objects children are also implemented as extensions of Content Units of the Information Package Map Section. However, because the XFDU does not authorize the direct references to the packaged files, a Data Object Content Unit makes use of XFDU defined Data Object Pointers to reference entries into the Data Object Section of the XFDU manifest. The byte streams are then referenced from this latter Data Object Section. When a PAIS SIP Data Object is composed of multiple byte streams, the XFDU Content Unit implementing that Data Object will have multiple XFDU Data Object Pointers referencing multiple XFDU Data Objects in the XFDU Data Object Section. Thus there is not always a one-to-one mapping between PAIS-defined Data Objects and XFDU-defined Data Objects. The term 'Data Object' normally refers to PAIS-defined Data Objects unless the context makes clear that the XFDU Data Object is intended.

A SIP that did not transfer any information would not be very useful, so the PAIS standard requires that at least one Data Object is required within a SIP.

When the Producer creates the SIP instance, it includes a number of items that allow the Archive to verify the SIP constructs against the MOT, i.e., the agreement between the Producer and the Archive about what is being transferred. What allows these checks to be made is the inclusion in the SIP of IDs that point back to definitions of those constructs in the MOT. For example the SIP `TranferObject` implementation includes the `descriptorID` that links back to the definition of that Transfer Object Type. Other examples are Transfer Object Groups containing the `associatedDescriptorGroupTypeID` element and Data Objects containing the `associatedDescriptorDataID` element.

As a reminder, the Producer is able to specify instance names for Transfer Object Groups (directories) (the `transferObjectGroupInstanceName` or the `transferObjectGroupPreservationName`) and Data Objects (files) (the `dataObjectInstanceName`) as part of the SIP. The names used may be the actual names of the directories and files in the Producer's data environment but other names may be provided if desired.

The XFDU standard includes many features that have not been adopted for the SIP/XFDU implementation. The primary role for the XFDU is an implementation of the abstract SIP to allow the various PAIS-defined constructs to be found and compared with their definitions in the MOT. The PAIS puts few explicit restrictions on use of the XFDU features. One of these restrictions is the limit of the XFDU `byteStream` element to a single `fileLocation` element.

The Producer and Archive may make a separate agreement to make use of other XFDU features to include information not required by the implementation of the abstract SIP.

While this is primarily outside the scope of this Green Book, one feature of note is the ability of the XFDU to include a mime type value in association with an XFDU Data Object and/or in association with an XFDU `byteStream` element. Such a mime type value needs to be not in conflict with the associated Descriptor-defined format information or mime type value.

5.3.2 LINKAGE BETWEEN DESCRIPTOR IDS AND SIP

Subsection 5.2 describes the abstract SIP, or SIP Model, and it includes extensive discussion regarding the various identifiers that need to appear in the SIP to relate the SIP structures to their descriptions in the MOT Descriptors. These identifiers are:

- `descriptorID`—relating a Transfer Object to its Descriptor;
- `associatedDescriptorGroupTypeID`—relating a Transfer Object Group to its description in the Descriptor;
- `associatedDescriptorDataID`—relating a Data Object to its description in the Descriptor.

Annex A expands this discussion for the `associatedDescriptorDataID`.

Each identifier is implemented in the XFDU SIP manifest file within the appropriate XFDU Content Unit, either the Transfer Object Content Unit, the Group Content Unit, or the Data Object Content Unit. Examples of these Content Units, extracted from a complete XFDU SIP manifest file given in annex subsection B5, which is related to the ISEE use case of 6.2, are as follows:

```
      <xfdu:contentUnit>
         <extension>
            <sip:sipTransferObject>
               <sip:descriptorID>ISEE_Mag_Data_TC2</sip:descriptorID>
               <sip:transferObjectID>ISEE_Mag_Data_TC2-
0001</sip:transferObjectID>
            </sip:sipTransferObject>
         </extension>
         <xfdu:contentUnit>
            <extension>
               <sip:sipTransferObjectGroup>

<sip:associatedDescriptorGroupTypeID>Satellite_Group</sip:associatedDescrip
torGroupTypeID>

<sip:transferObjectGroupInstanceName>isee1</sip:transferObjectGroupInstance
Name>
               </sip:sipTransferObjectGroup>
            </extension>
               <xfdu:contentUnit>
```

```
                    <extension>
                       <sip:sipDataObject>

<sip:associatedDescriptorDataID>ISEE_Mag_Data_File</sip:associatedDescripto
rDataID>
                       </sip:sipDataObject>
                    </extension>
                    <dataObjectPointer   dataObjectID="DO-ISEE_Mag_Data_File-
0001"/>
                </xfdu:contentUnit>
```

The Content Units use the XFDU extension capability to include the SIP elements. The Transfer Object Content Unit includes the `descriptorID` with a value of **`ISEE_Mag_Data_TC2`**. Nested within this Content Unit is a Transfer Object Group Content Unit that includes the `associatedDescriptorGroupTypeID` with a value of **`Satellite_Group`**. Nested within this Content Unit is a Data Object Content Unit that includes the `associatedDescriptorDataID` with a value of **`ISEE_Mag_Data_File`**. Again, the Archive needs to ensure that such identifiers are unique within the Archive Project during the development of the MOT.

Complete and practical examples of XFDU PAIS SIP implementations are provided and discussed in section 6 below.

5.4 NON-XFDU SIPS

Within the PAIS standard, two levels of PAIS conformance are identified. One level of conformance makes use of the PAIS XFDU implementation defined in that standard. Another level of conformance, the abstract level of PAIS conformance, allows for development of non-XFDU PAIS implementations that still fulfill all the requirements for information sharing and for automation of the Producer-to-Archive information transfer.

In the previous subsections of this document the XFDU implementation has been discussed. While the XFDU implementation is useful in many domains, there are also many other domains and communities that primarily make use of general packaging mechanisms other than XFDU. Within those domains and communities it may be worthwhile considering developing or making use of a PAIS implementation in the communities' dominant packaging mechanism that conforms to the abstract PAIS.

For communities with an existing non-XFDU dominant packaging mechanism, it may be better to use that mechanism for a PAIS implementation to gain greater PAIS adoption. It is also possible that this non-XFDU implementation of PAIS could be standardized within that community. The CCSDS participants who developed the PAIS standard may be able to support those efforts, or they may be able to review what that community develops.

If a decision is reached to develop a non-XFDU implementation of PAIS, it is recommended that the community try to make use of the native features of their chosen packaging mechanism to represent the required PAIS information, rather than just tacking on a separate PAIS module.

Another point is that once a non-XFDU implementation of PAIS is developed, it may be possible to develop associated tools that would support conversions between the XFDU PAIS implementation and the non-XFDU implementation. This is particularly true if the non-XFDU PAIS implementation is an XML-based implementation. In that case, it is likely that an XSL transformation between the two implementations could quickly be developed.

An example of a non-XFDU implementation is provided in section 6 where a Metadata and Encoding Transmission Standard (METS) implementation of PAIS is presented. The METS implementation was developed by the French National library (BnF) in consultation with the CCSDS PAIS development team. METS is a widely used standard for transmission of metadata and packaging information within the cultural heritage and library communities.

The BnF developed this METS implementation to study the applicability of PAIS to their environment. BnF is not currently using PAIS in its Scalable Preservation and Archiving Repository (SPAR). Nonetheless, PAIS could be profitably implemented in the case of complex and predictable Transfer Objects in the future.

6 USE CASES

6.1 INTRODUCTION

This section registers a series of use cases that were generally elaborated during the development and validation of the PAIS standard. These examples do not claim to provide turnkey solutions for operations. They may, however, improve the understanding of the PAIS standard through concrete cases built from various aspects. They may also help implementers start up their projects with patterns and snippets they can arrange at their discretion.

NOTE – XML Descriptors, SIP constraints, and extracts of SIP manifests associated to these use cases are provided in annexes B to E.

6.2 ISEE—A TYPICAL USE CASE

6.2.1 CONTEXT AND BENEFITS

This ISEE use case is based on data acquired by the NASA ISEE 1 and ISEE 2 'tandem' spacecraft launched in 1977. The use case is designed to exercise the major features of the PAIS standard. It covers the following:

– consideration of the data and its organization in the Producer's environment;

– how the Producer plans to describe and organize subsets of it into individual Transfer Objects; and

– how the Transfer Objects should be put into for transmission to an Archive.

It also briefly covers the role of the Archive in reviewing and approving the planned descriptions of the data and the organization of the data into SIPs. This allows both the Producer and Archive to have a common understanding of the data and its organization to ensure it meets the objective of both parties. This gives the Archive the ability to apply some automation in reviewing the received SIPs so they can be checked for conformance to the agreed plans, and this helps to reduce errors. It also covers the role of the Producer in using software to create the SIPs according to the agreed plans.

6.2.2 OBJECTS TO BE TRANSFERRED

The data chosen were resident on a NSSDCA (NASA Space Science Data Coordinated Archive) server; however, they were truncated for convenience prior to building the SIPs and therefore at that point no longer bore any real resemblance to the actual ISEE data.

The organization of the data in the Producer's (NSSDCA) environment is as shown in figure 6-1.

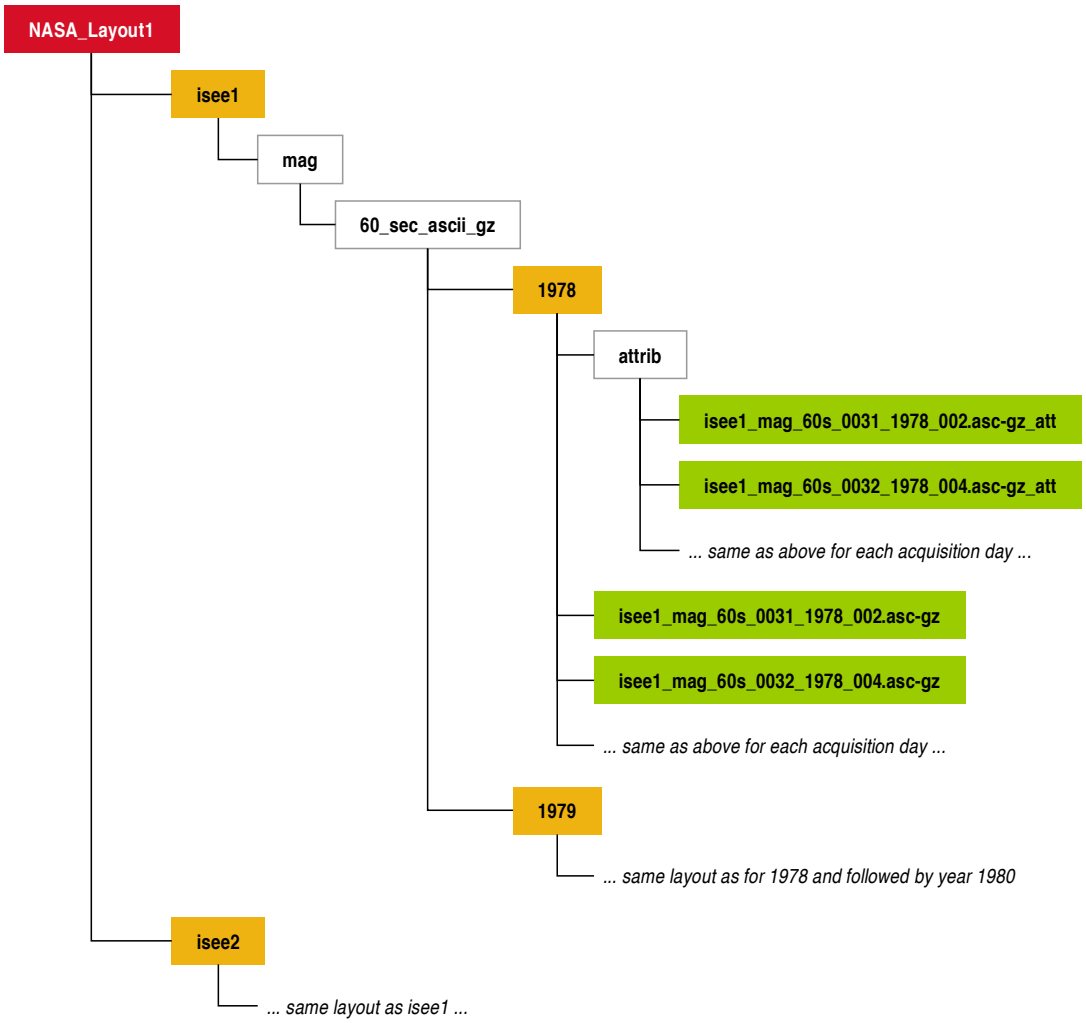

Figure 6-1: ISEE 1/ISEE 2 Data Repository Layout

The hierarchical levels correspond to different directory levels. There are many more directories present than are shown in figure 6-1. Only the directories in 'yellow' are to be included in the transmission to the Archive. The data used for the testing are from the ISEE 1 and ISEE 2 spacecraft selected from the 60 second ASCII magnetometer data under the '`mag`' directories. These are shown in 'green' and include the actual observations with file extension '`.asc-gz`' and corresponding attribute files with file extension '`.asc-gz-att`'. The 'asc-gz' file is a zip file containing many 60 second ASCII files for a given day. The 'asc-gz-att' file is also a daily zip file containing many attribute files, each corresponding to a single 60 second file of the same day.

NOTE – The file extensions are those inherited from the original repository. They do not correspond to any practice recommended by the present report. Probably, the use of extensions such as '`.txt.gz`' or '`-att.txt.gz`' would have improved the usability of these files on present computer environments.

The Producer needs to decide:

- what data is to be transferred;
- how it should be divided into Transfer Objects;
- what Transfer Objects should go into what SIPs; and
- whether some SIPs should be sent to the Archive prior to sending others.

Since a Transfer Object cannot be split across SIPs, determining what is to constitute a Transfer Object is a key consideration. Generally it will be most convenient for the Producer, when constructing one or more Transfer Objects, to maintain the same hierarchical and sequential relationships among the data files as exist in the Producer's environment. However, in some cases the Producer may want to reorganize the data as they are instantiated as Transfer Objects and put into SIPs in order to better support agreements with the Archive as to how the data will be made available to future users. As noted earlier, for this use case some directories are to be skipped and thus are not to be included in the Transfer Objects.

Additionally, the Producer needs to consider the mechanism, or transfer protocol, that will be used to transfer the SIPs as this may put a limit on the size of the SIPs and thus on the size or number of Transfer Objects in a given SIP. A size limitation may also be imposed by the Archive based on its data ingest handling capabilities.

The sequence in which the SIPs are received may be important to the Archive for validation purposes, or for management of the ingest process generally. For example, the Producer may be sending a formal description of formats which the Archive needs to have to support validation requirements, and thus the Archive wants the format information prior to receiving and processing the related data files.

In this test case, the Producer is proposing to send a small sample of 60 second magnetometer data from each spacecraft. It was decided that there should be two types of Transfer Objects, with multiple instances of each type, and there should be two distinct SIPs. One type of Transfer Object, referred to as 'data', should have both ISEE 1 and ISEE 2 files, with extension .asc-gz, taken from **one** of the years in the range 1978–1980. The result will be three Transfer Objects as there are three years in the range 1978–1980. To limit the size of these Transfer Objects, the daily files are restricted to days in the range 001–007. These decisions will result in Transfer Objects whose data organization will not be a direct extraction of the organization in the Producer's environment shown in figure 6-1. Rather there will be the equivalent of an ISEE1 directory containing a year 1978 directory containing several data files with days in the range 001–007. This will be followed by an ISEE2 directory containing a year 1978 directory containing several data files with days in the range 001–007. These restrictions can be met by examining the directory and file names in the Producer's environment and instructing the SIP creation software accordingly. This is a capability supported by the software (see 7.3) that will be used to automate the creation of these SIPs.

The second type of Transfer Object, referred to as 'metadata', should have the same restrictions, except its files will have the extension `.asc-gz_att`. These metadata files

provide a number of attributes about the primary data files. One type of SIP will be used to send the 'data' Transfer Objects, and the second type of SIP will be used to send the 'metadata' Transfer Objects. The 'metadata' SIP will be sent prior to sending the 'data' SIP.

To achieve the above objectives, the Transfer Object types, types of SIPs, and the SIP sequencing constraints need to be formally defined. This is discussed in the next subsection.

6.2.3 MODEL OF OBJECTS FOR TRANSFER AND SIP CONSTRAINTS

6.2.3.1 MOT

The Producer and the Archive jointly define the MOT, taking into consideration the level of detail needed for understanding and the level of validation desired. Considerations include what objects are to be transferred, their frequency of occurrence, what relationships exist among the objects, and under what format they will be provided to the Archive. The Producer and Archive also specify the naming rules for the different identifiers of collections, Transfer Object Types, Producer Sources, etc. The practical generation of most MOTs will require supporting software. The MOT for this use case was generated with such software, and it presents a graphical user interface.

The schematic of the MOT for the ISEE use case, as shown in figure 6-2, involves one Collection Descriptor as the parent of two Transfer Object Descriptors. The Collection Descriptor semantics calls for ISEE 1 and ISEE 2 magnetic field data and metadata grouped by spacecraft. The term 'metadata' is used here to refer to the attribute files with extension '.asc-gz-att' as shown in figure 6-1. The collection size is specified to lie between 9 and 22 MB. Two association relationships, which are optional in the standard, of 'contains' are specified corresponding to the data and metadata Transfer Object Descriptors. The XML content of the Collection Descriptor is provided in annex subsection B1.

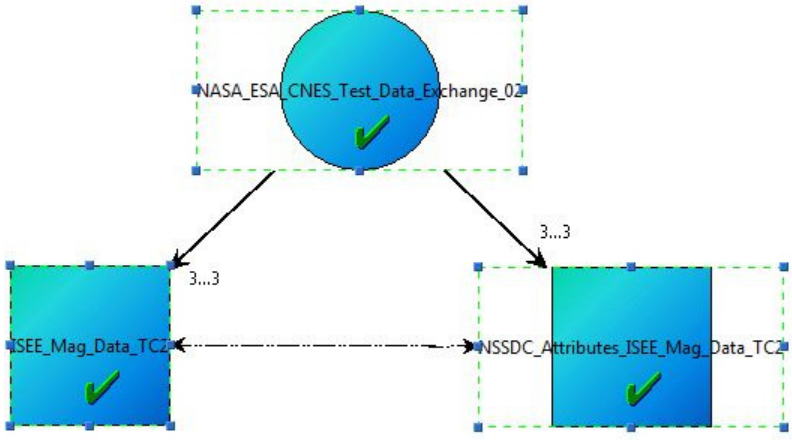

Figure 6-2: ISEE 1/ISEE 2 MOT

The **data** Transfer Object Descriptor (whose parent is `NASA_ESA_CNES_Test_Data_Exchange_02` as shown in figure 6-2) is an XML

object that is fully specified in annex subsection B2. Its semantics calls for each Transfer Object to contain two satellite groups (ISEE 1 and ISEE 2), with each group containing a single yearly directory group taken from the range 1978 through 1980. The yearly directory group will hold data for days 001–007 inclusive. The relevant excerpt is as follows:

```
            <transferObjectTypeTitle>Annual Directory of ISEE 1,2
Magnetic_Field Data

            </transferObjectTypeTitle>
            <transferObjectTypeDescription>Annual Directory of ISEE 1,2
      magnetic field data (no metadata) grouped by Spacecraft (ISEE 1 and
      ISEE 2) and then for a Yearly Directory in range 1978 through 1980
      for days 001-007 inclusive.

            </transferObjectTypeDescription>
```

The Descriptor also specifies that there will be three Transfer Objects, corresponding to the three years 1978–1980, and each will have a size between 3 and 7 MB. The relevant excerpt is as follows:

```
            <transferObjectTypeOccurrence>
                  <minOccurrence>3</minOccurrence>
                  <maxOccurrence>3</maxOccurrence>
            </transferObjectTypeOccurrence>
            <transferObjectTypeSize>
                  <minSize>3</minSize>
                  <maxSize>7</maxSize>
                  <unitsType>MB</unitsType>
            </transferObjectTypeSize>
```

Inside each yearly group there need to be between two and four Data Objects. The Data Objects are specified to be plain text that has been gzip encoded. The relevant excerpt is as follows:

```
<groupType>
            <groupTypeID>Yearly_Group</groupTypeID>
            <groupTypeDescription>Each group will contain 1 year taken from
range 1978-1980</groupTypeDescription>

      <groupTypeStructureName>directory</groupTypeStructureName>
                  <groupTypeOccurrence>
                        <minOccurrence>1</minOccurrence>
                        <maxOccurrence>1</maxOccurrence>
                  </groupTypeOccurrence>
                  <dataObjectType>

      <dataObjectTypeID>ISEE_Mag_Data_File</dataObjectTypeID>
                        <dataObjectTypeDescription>ISEE magnetometer data
file whose file name contains a day identifier in the range 001-
007.</dataObjectTypeDescription>
                        <dataObjectTypeOccurrence>
```

```
                              <minOccurrence>2</minOccurrence>
                              <maxOccurrence>4</maxOccurrence>
                        </dataObjectTypeOccurrence>
                        <dataObjectTypeFormat>
                              <mimeType>text/plain</mimeType>
                        </dataObjectTypeFormat>
                        <dataObjectTypeEncoded>
                              <encodingName>gzip</encodingName>
                              <encodingDescription>application/x-
gzip</encodingDescription>
                        </dataObjectTypeEncoded>
```

It also specifies an association relationship of 'data' with respect to a referenced group of metadata (attribute) files. The relevant excerpt is as follows:

```
      <relation>

      <parentCollection>NASA_ESA_CNES_Test_Data_Exchange_02</parentCollecti
on>
            <association>
                  <targetID>NSSDC_Attributes_ISEE_Mag_Data_TC2</targetID>
                  <relationDescription>
                        <relationType>Data</relationType>
                        <relationTextualDescription>group of data files
corresponding to the Target Id's group of metadata
files</relationTextualDescription>
                  </relationDescription>
            </association>
      </relation>
```

The **metadata** Transfer Object Descriptor (whose parent is also NASA_ESA_CNES_Test_Data_Exchange_02 as shown in figure 6-2) is an XML object that is fully specified in annex subsection B3. Its semantics calls for each Transfer Object to contain two satellite groups (ISEE 1 and ISEE 2) with each group containing a single yearly directory group taken from the range 1978 through 1980. The yearly directory group will hold metadata for days 001–007 inclusive. The relevant excerpt is as follows:

```
      <description>
            <transferObjectTypeTitle>Annual Directory of NSSDC Attributes
for ISEE 1,2 Magnetic_Field Data</transferObjectTypeTitle>
            <transferObjectTypeDescription>Annual Directory of NSSDC
Attributes for ISEE 1,2 magnetic field data grouped by Spacecraft (ISEE 1
and ISEE 2) and then for a Yearly Directory (in range 1978 through 1980)
for days 001-007 inclusive.  </transferObjectTypeDescription>
```

The Descriptor also specifies that there will be three Transfer Objects, corresponding to the three years 1978–1980, and each will have a size between 8 and 24 KB. The relevant excerpt is as follows:

```
            <transferObjectTypeOccurrence>
                  <minOccurrence>3</minOccurrence>
                  <maxOccurrence>3</maxOccurrence>
            </transferObjectTypeOccurrence>
```

```
            <transferObjectTypeSize>
                  <minSize>8</minSize>
                  <maxSize>24</maxSize>
                  <unitsType>KB</unitsType>
            </transferObjectTypeSize>
```

Inside each yearly group there need to be between two and four Data Objects (referred to as metadata objects). The metadata objects are specified to be plain text. The relevant excerpt is as follows:

```
      <groupType>
            <groupTypeID>Satellite_Group</groupTypeID>
            <groupTypeDescription>There are 2 satellite groups, ISEE1 and
ISEE2</groupTypeDescription>
            <groupTypeStructureName>directory</groupTypeStructureName>
            <groupTypeOccurrence>
                  <minOccurrence>2</minOccurrence>
                  <maxOccurrence>2</maxOccurrence>
            </groupTypeOccurrence>
            <groupType>
                  <groupTypeID>Yearly_Group</groupTypeID>
                  <groupTypeDescription>Each group will contain 1 year taken
from range 1978 - 1980</groupTypeDescription>
                  <groupTypeStructureName>directory</groupTypeStructureName>
                  <groupTypeOccurrence>
                        <minOccurrence>1</minOccurrence>
                        <maxOccurrence>1</maxOccurrence>
                  </groupTypeOccurrence>
                  <dataObjectType>

      <dataObjectTypeID>NSSDC_Attributes_ISEE_Mag_Data_File</dataObjectTypeID>
                        <dataObjectTypeDescription>NSSDC generated metadata
file, corresponding to ISEE magnetometer data file, whose file name contains a
day identifier in the range 001-007.</dataObjectTypeDescription>
                        <dataObjectTypeOccurrence>
                              <minOccurrence>2</minOccurrence>
                              <maxOccurrence>4</maxOccurrence>
                        </dataObjectTypeOccurrence>
                        <dataObjectTypeFormat>
                              <mimeType>text/plain</mimeType>
                        </dataObjectTypeFormat>
```

It also specifies an association relationship of ‘metadata’ with respect to a referenced group of data files. The relevant excerpt is as follows:

```
      <relation>
            <parentCollection>NASA_ESA_CNES_Test_Data_Exchange_02</parentCollection>
            <association>
                  <targetID>ISEE_Mag_Data_TC2</targetID>
                  <relationDescription>
                        <relationType>Metadata</relationType>
                        <relationTextualDescription>group of metadata files
corresponding to the Target Id’s group of data files</relationTextualDescription>
                  </relationDescription>
            </association>
      </relation>
```

6.2.3.2 ISEE SIP Constraints

The ISEE use case also includes a SIP constraints file that identifies two types of SIPs. `SIP_01` is allowed to contain from one to three data Transfer Objects while `SIP_02` is allowed to contain from one to three metadata Transfer Objects. In addition, it specifies that the `SIP_02` metadata Transfer Objects are to be transferred to the Archive prior to sending the `SIP_01` data Transfer Objects. The XML content of the SIP Constraints is provided in annex subsection B4.

6.2.4 SIPS

The PAIS specifies a standard packaging mechanism for the implementation of PAIS SIPs. It is based on use of the XFDU packaging standard. When this is followed, and the semantics of PAIS section 5 are followed, the resulting implementation is said to be 'XFDU PAIS SIP Conformant'. However, it is acceptable to use other packaging mechanisms. In this case the resulting SIP implementation can be said to be 'Abstract PAIS SIP Conformant' provided it also adheres to the semantics of PAIS section 5.

These ISEE SIPs are based on the XFDU standard as extended by the PAIS schema (see section 6 of the PAIS standard) and therefore consist of

– an **xfdumanifest.xml** file providing the following information:

  • the packageHeader containing in particular the PAIS sipGlobalInformation: sipID, producerSourceID, producerArchiveProjectID, sipContentTypeID, sipSequenceNumber;

  • the informationPackageMap describing the contained PAIS sipTransferObject corresponding to the Transfer Object Types authorized within the SIP;

  • The dataObjectSection listing the different transferred objects: a unique identifier, the size in bytes of the file, its location within the SIP, its MD5 checksum for integrity validation;

– the **different transferred objects** as files or directories/subdirectories containing files.

As noted in 6.2.3.2, two types of SIPs are generated for this use case. The first type, whose SIP Content Type ID is '`SIP_01`', can contain from one to three data Transfer Objects. As noted in 6.2.3.1 and also in the data Transfer Object Descriptor in annex subsection B2, these data Transfer Objects range in size from 3 to 7 MB. Therefore the decision was made to put all three data Transfer Objects into a single SIP. Only the `SIP_01` type will be described as the `SIP_02` type has a parallel construction.

The practical generation of most SIPs will require the use of supporting software. Such software, called the ESA SIP Builder (see 7.3), was used to generate the manifest for the data SIP (`SIP_01`) as given in annex subsection B5. A tabular view is given in figure 6-3. Items in bold correspond to terms from the standards (PAIS and XFDU) while non-bold items are values explicit to this use case.

The first section of the manifest contains the SIP Global Information. It contains a number of identifiers. Annex D of the PAIS standard provides an informative summary of the various PAIS identifiers including references to relevant sections of the standard.

The SIP ID is provided by the Producer, and it is checked by the Archive to ensure it is unique within this particular Producer-Archive Project. In this case, the SIP ID is given as '`NASA_ESA_CNES_Test_Data_Exchange_02-SIP-0002`', where the Producer has decided to append the term SIP followed by a sequence number to ensure uniqueness.

It is possible that there may be multiple actual Producers submitting SIPs for a given Producer-Archive Project. The Archive needs to understand who has submitted each SIP in case there are issues that need to be resolved. This is handled by providing a Producer-Source ID which is agreed to jointly by the Producer and Archive. In this case there is only one source and the ID agreed is 'NASA_ESA_Source1'.

When the Archive receives a SIP, it needs to understand which Producer-Archive Project it is associated with. The Archive establishes the Producer-Archive Project ID to ensure uniqueness within its environment. In this case it chose '`NASA_ESA_CNES_Test_Data_Exchange_02`'. The Producer then used this ID as the basis for its generation of SIP IDs. There is no requirement to do this but it is more informative than simply using something like a sequence number for the SIP ID.

When the Archive receives a SIP, it also needs to understand what type of SIP it conforms to because each type of SIP has different constraints. This is handled by including a SIP Content Type ID that is jointly decided between the Producer and Archive when there is agreement on the SIP Constraints file, which is an XML object (see annex subsection B4). In this case the allowed IDs are `SIP_01` and `SIP_02`. `SIP_01` identifies the SIP as containing ISEE data (but not metadata).

The PAIS also specifies an optional 'SIP Sequence Number' that can be useful in identifying missing SIP deliveries. This is particularly useful when the exact number of Transfer Objects of each type has not been specified in the Descriptors. Although the exact number of Transfer Objects has been specified in this ISEE use case, the Producer decided to include the SIP Sequence Number. It is given as '2' because this data SIP is sent after the metadata SIP.

PAIS elements/items	Contents
SIPGlobalInformation	
	sipID : NASA_ESA_CNES_Test_Data_Exchange_02-SIP-0002 **producerSourceID :** NASA_ESA_Source1 **producerArchiveProjectID:** NASA_ESA_CNES_Test_Data_Exchange_02 **sipContentTypeID:** SIP_01 **sipSequenceNumber:** 2
sipTransferObject	
	descriptorID: ISEE_Mag_Data_TC2 **transferObjectID:** ISEE_Mag_Data_TC2-0001 **sipTransferObjectGroup**
	associatedDescriptorGroupTypeID: Satellite_Group
	transferObjectGroupInstanceName: isee1
	sipTransferObjectGroup
	associatedDescriptorGroupTypeID: Yearly_Group **transferObjectGroupInstanceName:**1978
	sipDataObject
	associatedDescriptorDataID:ISEE_Mag_Data_File
	dataObjectPointer:dataObjectID= "DO-ISEE_Mag_Data_File-0001"
	sipDataObject
	associatedDescriptorDataID: ISEE_Mag_Data_File
	dataObjectPointer:dataObjectID= "DO-ISEE_Mag_Data_File-0002"
	sipDataObject
	associatedDescriptorDataID: ISEE_Mag_Data_File
	dataObjectPointer:dataObjectID= "DO-ISEE_Mag_Data_File-0003"
	sipTransferObjectGroup
	associatedDescriptorGroupTypeID: Satellite_Group
	transferObjectGroupInstanceName: isee2
	sipTransferObjectGroup
	associatedDescriptorGroupTypeID: Yearly_Group **transferObjectGroupInstanceName:**1978
	sipDataObject
	associatedDescriptorDataID:ISEE_Mag_Data_File
	dataObjectPointer:dataObjectID= "DO-ISEE_Mag_Data_File-0004"
	sipDataObject
	associatedDescriptorDataID: ISEE_Mag_Data_File
	dataObjectPointer:dataObjectID= "DO-ISEE_Mag_Data_File-0005"
	sipDataObject
	associatedDescriptorDataID: ISEE_Mag_Data_File
	dataObjectPointer:dataObjectID= "DO-ISEE_Mag_Data_File-0006"

Figure 6-3: Tabular View of an ISEE Data SIP Manifest File

PAIS elements/items	Contents
sipTransferObject	
	descriptorID: ISEE_Mag_Data_TC2 **transferObjectID:** ISEE_Mag_Data_TC2-0002 **sipTransferObjectGroup**
	associatedDescriptorGroupTypeID: Satellite_Group
	transferObjectGroupInstanceName: isee1
	associatedDescriptorGroupTypeID: Yearly_Group **transferObjectGroupInstanceName:**1979
The subsequent organization for 1979 and 1980 follows that above for 1978, but is not shown here for brevity. Following this is the dataObject section which is shown below only for 1978.	
dataObject	
	ID="DO-ISEE_Mag_Data_File-0001" **size=**"128"
	byteStream
	size="128" **fileLocation locatorType="URL" href="** **file:**isee1/1978/isee1_mag_60s_0031_1978_002.asc-gz " **checksum checksumName=**"MD5" 7cc53dd29fb89105352e5f50f9af06b5
dataObject	
	ID="DO-ISEE_Mag_Data_File-0002" **size=**"128"
	byteStream
	size="128" **fileLocation locatorType="URL" href="file:** isee1/1978/isee1_mag_60s_0032_1978_004.asc-gz " **checksum checksumName=**"MD5" 8b1e38c7109f4b39ae5f0ec456ba1569
dataObject	
	ID="DO-ISEE_Mag_Data_File-0003" **size=**"128"
	byteStream
	size="128" **fileLocation locatorType**="**URL**" **href="file:** isee1/1978/isee1_mag_60s_0033_1978_007.asc-gz" **checksum checksumName=**"MD5" 92f7668852b9006f4091becd3b3e7ab7
dataObject	
	ID="DO-ISEE_Mag_Data_File-0004" **size=**"128"
	byteStream
	size="128" **fileLocation locatorType="URL" href="** **file:**isee2/1978/isee2_mag_60s_0031_1978_002.asc-gz " **checksum checksumName=**"MD5" 15e56b31c9c576f7ca50785f31bc8528

Figure 6-3: Tabular View of an ISEE Data SIP Manifest File (continued)

PAIS elements/items	Contents	
dataObject		
	ID="DO-ISEE_Mag_Data_File-0005" **size=**"128"	
	byteStream	
		size="128" **fileLocation locatorType="URL" href="** **file:**isee2/1978/isee2_mag_60s_0032_1978_002.asc-gz " **checksum checksumName=**"MD5" 0aa1312e75d78a68c98cf0063c6115a6
dataObject		
	ID="DO-ISEE_Mag_Data_File-0006" **size=**"128"	
	byteStream	
		size="128" **fileLocation locatorType="URL" href="** **file:**isee2/1978/isee2_mag_60s_0033_1978_002.asc-gz " **checksum checksumName=**"MD5" 06b5550d1907056737992c60530045be

Figure 6-3: Tabular View of an ISEE Data SIP Manifest File (continued)

The next section starts with the first Transfer Object. Following the specifications and semantics of the Transfer Object Descriptor, it contains two-top level Transfer Object Groups corresponding to ISEE1 data and ISEE2 data respectively. Each of these groups contains another group corresponding to data for the first year, or 1978. Within each of these secondary groups are three SIP Data Objects. Each corresponds to a daily data file taken from the day range one to seven. Each Data Object uses the XFDU `dataObjectPointer` to give an identifier of a set of information that can be found later in the manifest under the XFDU `dataObject` section. This section, shown in the tabular view, has a number of XFDU `dataObject` elements. Each dataObject has an identifier that matches one of those given by the `dataObjectPointer`. It also gives information about the Data Object such as size, file location, and checksum. It could also refer to multiple files but in this use case each Data Object is a single file. Because the actual files used to create the SIPs were very truncated versions of the original data, the file sizes of 128 bytes each result in Transfer Objects very much smaller than the minimum size of 3 MB as specified in the corresponding Transfer Object Descriptor. As a result, the Archive would find that these Transfer Objects fail a validation check on Transfer Object sizes. The Archive would be expected to contact the Producer (`NASA_ESA_Source1`) to resolve the issue.

The second Transfer Object is much like the first, but it corresponds to data for the year 1979. For brevity, only the beginning elements are shown. For the same reason, the third Transfer Object, for 1980, is not shown.

The complete SIP for this use case consists of a zip file containing the manifest file and the directories and files consistent with those given by the file location information in the XFDU `dataObject` elements. As noted above, building the Transfer Objects within a SIP requires examining the associated Descriptor and mapping its specifications to the organization of data

within the Producer's environment so that proper Transfer Objects are generated. When the initial SIPs were generated for this use case, the semantics within the Description elements of the Descriptor's groups were not closely followed. In particular, when groups contain other groups, and they have multiple occurrences, it can be ambiguous as to the order in which the groups and subgroups are instantiated. This leads to the initial SIPs' having Transfer Objects that have various mixtures of ISEE1 and ISEE2, instead of each Transfer Object's having both ISEE1 and ISEE2 satellite groups. This is shown schematically in figure 6-4, where TO stands for Transfer Object and view (a) is the correct view as described above, while (b) is the initial view that was inconsistent with the Descriptor semantics.

When there are multiple occurrences, it is important to be sure the needed semantics regarding instantiation of the groups is clear. This could also involve the use of a user-defined element to more formally define the order of group instantiations, leading to greater automation in correctly generating Transfer Objects and SIPs.

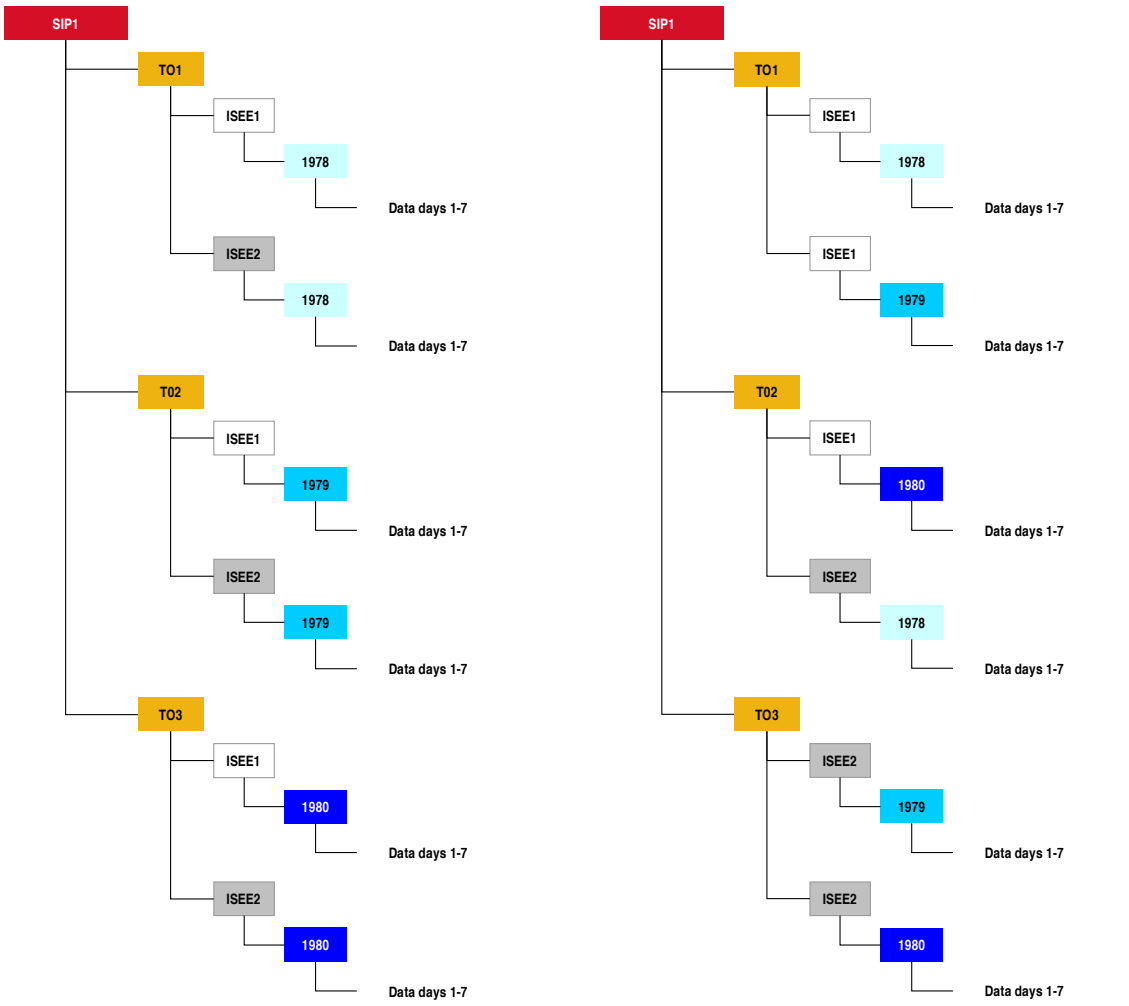

Figure 6-4: Views of Correct and Incorrect Ordering of ISEE Satellite Group Instantiations

6.3 ESA-SAFE—TRANSFER OF SAFE PRODUCTS

6.3.1 CONTEXT AND BENEFITS

Standard Archive Format for Europe (SAFE) (see reference [7]) is an Earth Observation data archiving format standardized through the efforts of several European national, institutional, and industrial space stakeholders. It provides a specification for the organization and content of an OAIS-compatible AIP.

The ESA-SAFE use case is based on data and documentation provided by ESA (European Space Agency). These data are samples and subset from the European Remote Sensing Satellite (ERS) Synthetic Aperture Radar (SAR), and tailored for the scope of this test case.

In this use case, SAFE 2.0 Packages are wrapped inside SIPs for submission to an Archive. This tutorial illustrates the possibility of:

– detailing the contents of the SAFE packages or not (packages seen as black boxes);

– defining relationships (provides context for, provides representation information for, provides documentation of, …) among the packages or among the files contained in the packages, which might be used, for example, to identify the Representation Information for data format validation at the Archives side;

– defining and controlling sequencing constraints among the delivered SIPs;

– validating the integrity of the SIPs containing SAFE packages.

Two ways of modeling the transfer between the Producer and the Archive are proposed in this section so as to underline that such modeling depends on the level of details needed and the level of validation desired by the Producer and required by the Archive. These two ways of modeling are referred to further in the section as:

– ‘simple case’: when the SAFE packages are transferred as black boxes;

– ‘detailed case’: when the SAFE packages are described in detail.

NOTES

1 The SAFE 2.0 standard is still undergoing development and certain aspects may change without notice. The files and packages provided as examples are not actual SAFE Packages guaranteed to be valid against the SAFE 2.0 standard.

2 For practical reasons, the ‘MEASUREMENT.DAT’ file included in the example SAFE 2.0 EO Product Package is a truncated version of an actual ERS SAR Level-0 strip line (which is typically a few gigabytes large).

6.3.2 OBJECTS TO BE TRANSFERRED

The objects selected for transfer consist of four examples of an ERS SAR SAFE 2.0 EO (Earth Observation) product package with their corresponding Representation Information (three other SAFE packages) and associated documentation. For convenience, the former packages are called SAFE Data Products.

Here is a short description of all packages pertaining to this use case:

a) Four SAFE Data Products transferred one by one. These products are:

   – ER2_OPER_SAR_IM__0P_19970923T212658_19970923T213316_43E3.SAFE which corresponds to a set of files as defined by the SAFE 2.0 format;

   – three copies of it (with XXXX, YYYY, and ZZZZ instead of the string '43E3') to simulate sequential delivery of additional products.

b) The associated Representation Information, in turn made up of three SAFE packages:

   – The first package is the so-called Base Schemas Package, which is a SAFE 2.0 Representation Information Package. This package has to be present in any SAFE 2.0 compliant Archive. Its name in this case is ESA_101213T122045_1A3F_BAS.SAFE.

   – The two other packages are also SAFE 2.0 Representation Information Packages but provide ERS SAR L0 Representation Information of the metadata (XSD schemas) and data (Data Format Description Language [DFDL] schemas) components present in all ERS SAR L0 SAFE Data Products. They are named, respectively,

      • ERS-AMI-SAR-LEVEL-0_101213T122045_35AF_MTD.SAFE; and

      • ERS-AMI-SAR-LEVEL-0_101213T122045_1F25_DAT.SAFE.

c) A set of documents related to ERS missions (in PDF format) providing context information.

The objects to transfer to the Archive are gathered under three different directories:

– a directory named 'SAFE_REPINFO' for the three SAFE 2.0 Representation Information Packages;

– a directory named 'EO_PRODUCTS' for the four SAFE Data Products;

– a directory named 'ERS-DOC' for the PDF documents.

The organization of the 'simple case' is outlined in figure 6-5, where all packages are seen as black boxes (zip archives).

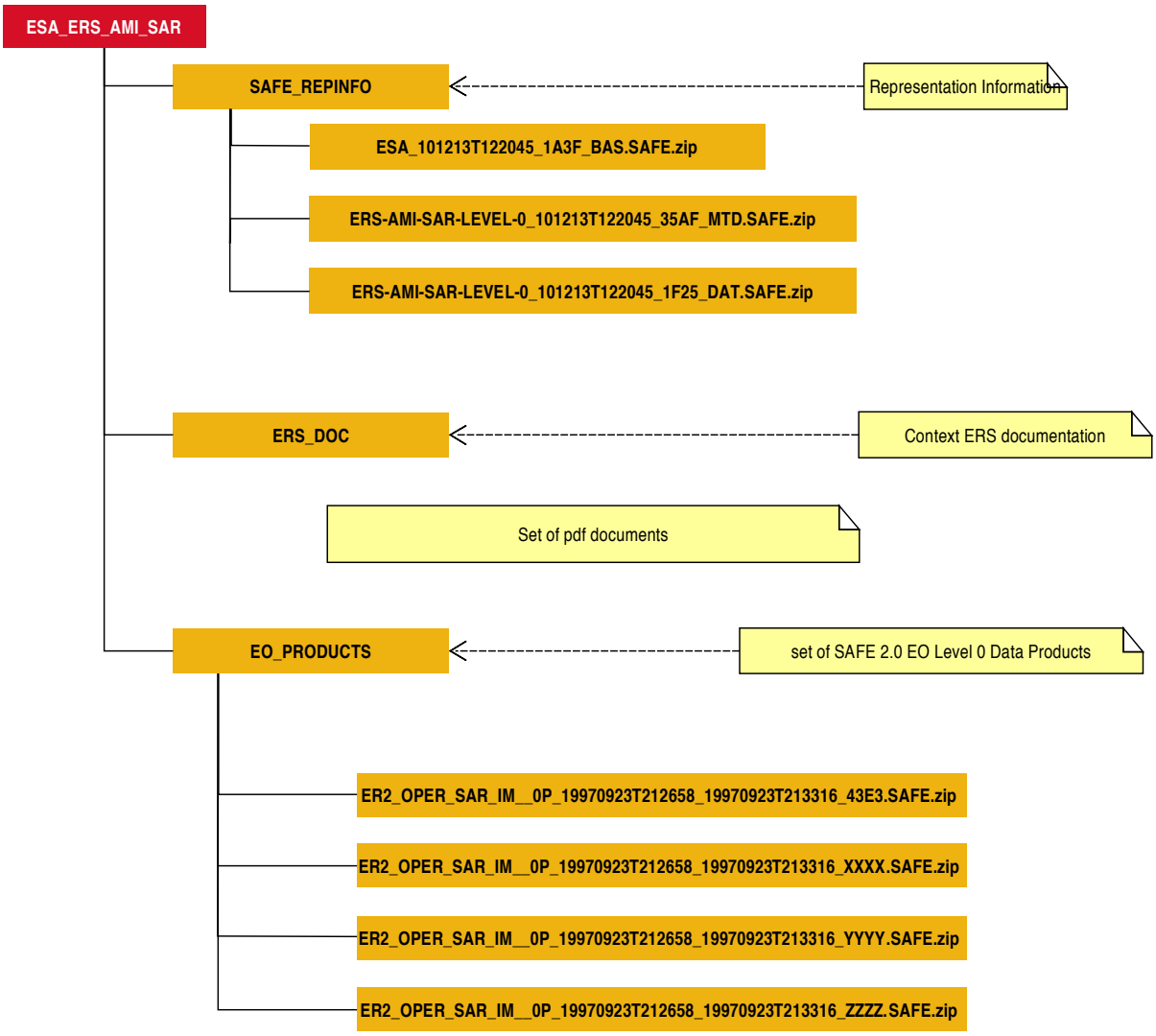

Figure 6-5: ESA SAFE Repository Layout in 'Simple Case'

The organization of the 'detailed case' is outlined in figure 6-6, where the contents of the packages are detailed at file level (not all files are shown here).

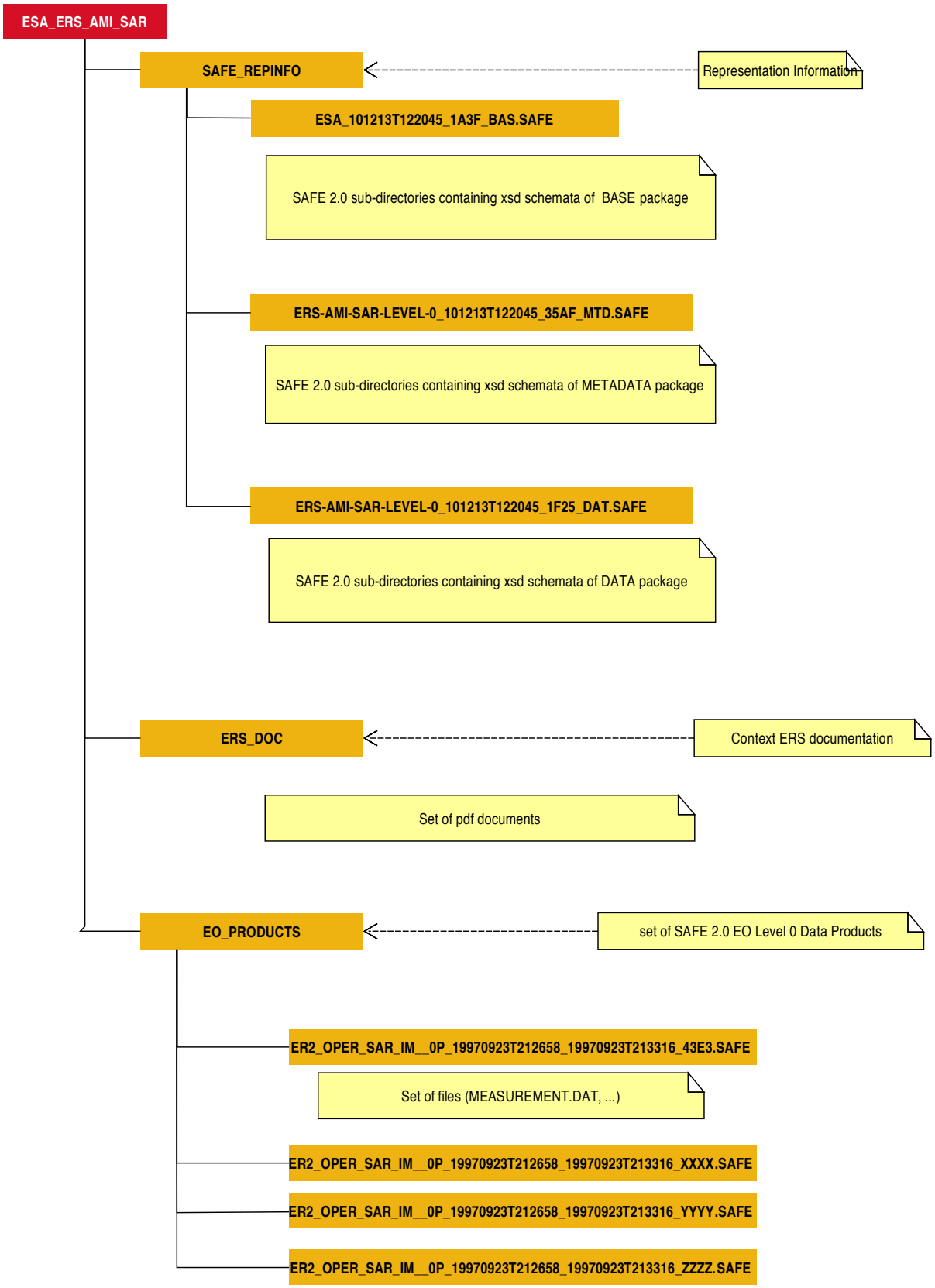

Figure 6-6: ESA SAFE Repository Layout in 'Detailed Case'

6.3.3 MODEL OF OBJECTS FOR TRANSFER AND SIP CONSTRAINTS

6.3.3.1 MOT

The Producer and the Archive jointly define the MOT, taking into consideration the level of detail needed for understanding and the level of validation desired. Considerations include what objects are to be transferred, their frequency of occurrence, what relationships exist among the objects, and under what format they will be provided to the Archive. The Producer and Archive also specify the naming rules for the different identifiers of collections, Transfer Object Types, Producer Sources, etc.

Table 6-1 presents a comparison between the two modeling choices:

Table 6-1: Comparison between Simple and Detailed MOT

	Simple Case	Detailed Case
Data as described (Producer)	Zip packages (black boxes)	Unpacked data
Model of Transfer	Minimal description	Up to the individual files inside packages
Links	High level between zip files	Between the individual files inside packages
Integrity	Checksum of SAFE package	Checksum of each Data Object (including SAFE manifest) inside packages
Validations	Checksum (SAFE package), sequencing constraints, SIPs, and conformity to the Model	Checksum of each individual Data Object, sequencing constraints, SIPs, and conformity to the Model

As three kinds of objects have been identified (data, Representation Information, documentation), the ESA-SAFE MOT chose to specify three collections under the project root collection ESA_ERS_AMI_SAR:

– COL_ERS-AMI-SAR-EO-DATA (Level 0 data);

– COL_ERS-AMI-SAR-Rep-Info (Representation Information); and

– COL_ERS-AMI-SAR-DOC (documentation).

The COL_ERS-AMI-SAR-EO-DATA collection logically gathers the description of a SAFE Data Product (with at least one occurrence and an unknown number of delivered data products): TRF_ER2_OPER_SAR_IM_PRODUCT.

The COL_ERS-AMI-SAR-Rep-Info collection logically gathers the Representation Information conveyed by three Transfer Object Types (with only one occurrence for each) corresponding to each kind of SAFE 2.0 Representation Information package:

– TRF_BASE_PACKAGE_REP_INFO describes the Base Schemas Package;

– TRF_MTD_PACKAGE_REP_INFO describes the Metadata Schemas Package;

– TRF_DAT_PACKAGE_REP_INFO describes the Data Schemas Package.

The COL_ERS-AMI-SAR-DOC collection logically gathers the description of the associated multi-mission documents (with only one occurrence): TRF_ERS_AMI_SAR_DOC.

Associations can be defined in both the simple and the detailed cases as follows:

– representation association between the SAFE Data Product (TRF_ER2_OPER_SAR_IM_PRODUCT) and its Representation Information:

  • with TRF_MTD_PACKAGE_REP_INFO for the Representation Information of the metadata components of the SAFE Data Product;

  • with TRF_DAT_PACKAGE_REP_INFO for the Representation Information of the data components of the SAFE Data Product;

– documentation association between the SAFE Data Product (TRF_ER2_OPER_SAR_IM_PRODUCT) and the general documentation:

  • with TRF_ERS_AMI_SAR_DOC for the context documentation for the data package;

– links (dependencies) among the Representation Information packages:

  • between metadata Representation Information package (TRF_MTD_PACKAGE_REP_INFO) and the base package (TRF_BASE_PACKAGE_REP_INFO);

  • between the data Representation Information package (TRF_DAT_PACKAGE_REP_INFO) and the base package (TRF_BASE_PACKAGE_REP_INFO).

Figure 6-7 is a screenshot taken via the CNES prototype showing the MOT (collections, Transfer Object Types, and relationships) for the ESA-SAFE use case. The visualization is the same for both cases (simple and detailed).

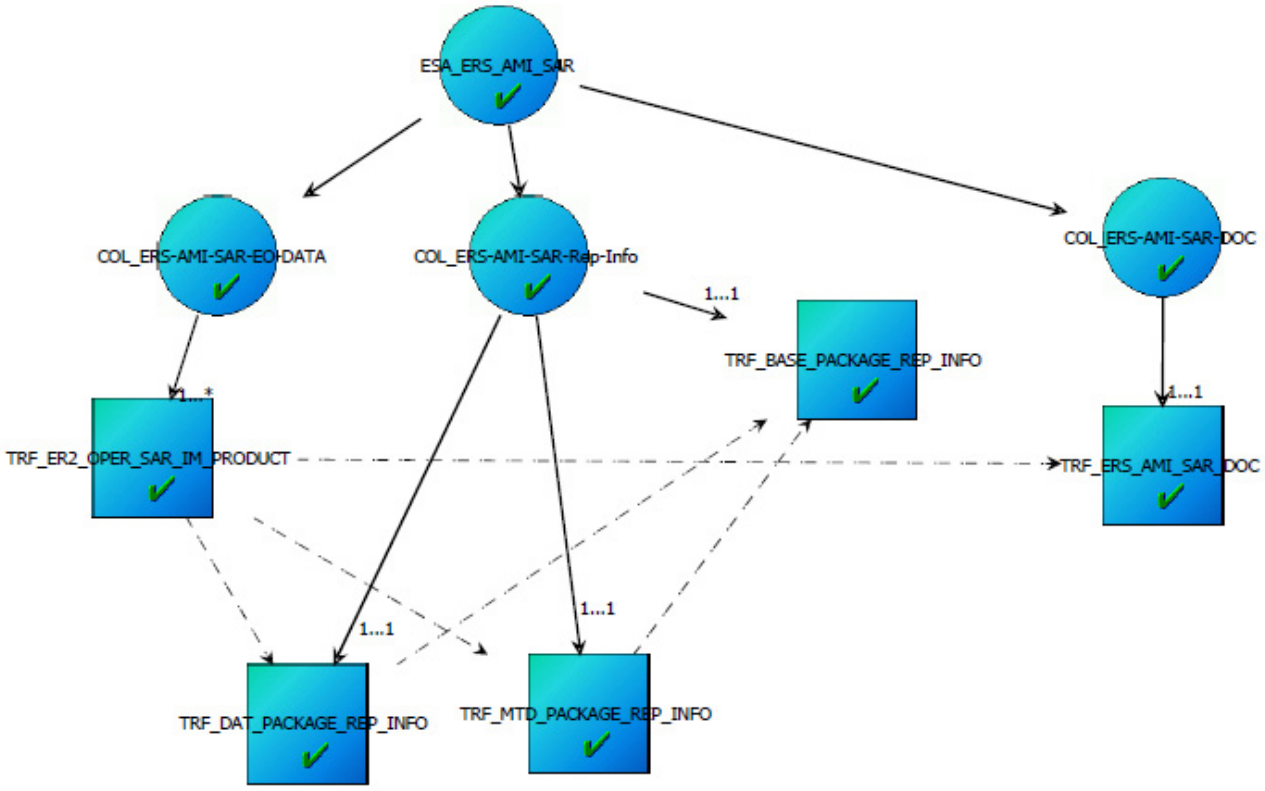

Figure 6-7: ESA SAFE MOT

In the ‘detailed case’ the description of a SAFE Level 0 Data Product corresponds to the transfer of four files (MEASUREMENT.SIX, MEASUREMENT.DAT, and two XML files).

If a Validation Phase has to be performed on the content of the product with regards to its DFDL description, each concerned file has to be clearly identified within its corresponding transfer description in order to add the required links at object level, i.e., between the object description corresponding to the MEASUREMENT.DAT file within TRF_ER2_OPER_SAR_IM_PRODUCT transfer description and the object description corresponding to the DFDL file within TRF_DAT_PACKAGE_REP_INFO transfer description.

The ‘detailed case’ does not implement this level of detail just explained (see annex C for further information).

6.3.3.2 SIP constraints

Once the MOT has been specified, the next steps are:

- the identification of the Producer Archive Project Identifier;
- the identification of the SIPs;
- the sequence order of the SIPs.

The known delivery order constraints are:

a) SAFE Data Products are transferred one by one, one after the other for practical reasons since the products are a few Gigabytes large and for demonstration reasons (simulating a mission lifecycle when products are regularly provided).

b) SAFE Data Products have to be ingested in the Archive after all Representation Information Packages because SAFE Data products point to SAFE Representation Information packages.

NOTE – This is also the case if the Archive wants to apply data file validation against the DFDL and XSD schemas; that is not the reason here.

c) Among the Representation Information Packages, the following constraints exist:

- The Base Representation Information Schemas Package has to be ingested before any other package and is to be provided only once.
- The two other Representation Information packages have to be ingested before any SAFE Product Package and are to be provided once.

d) The documents can be ingested all at the same time and at any moment.

These constraints lead to the definition of the following sequence order: base Representation Information package first, followed by metadata Representation Information schemas package, followed by data Representation Information schemas package, then by documentation package, finally by 1 up to *n* (=4) ESA SAFE Data Products. The sequence order is the same in both cases (simple and detailed). Consequently, five types of SIPs are defined to match this delivery sequence order, each containing a single Transfer Object type. (Alternatively, only two types of SIPs could have been defined to accomplish the task: one SIP could contain the three Representation Information schema packages—base, metadata, and data—as well as the documentation package since they could all arrive together, and a second type of SIP would contain the ESA SAFE Data Product).

Figure 6-8 presents the five types of SIPs, their contents, and their associated sequencing constraints.

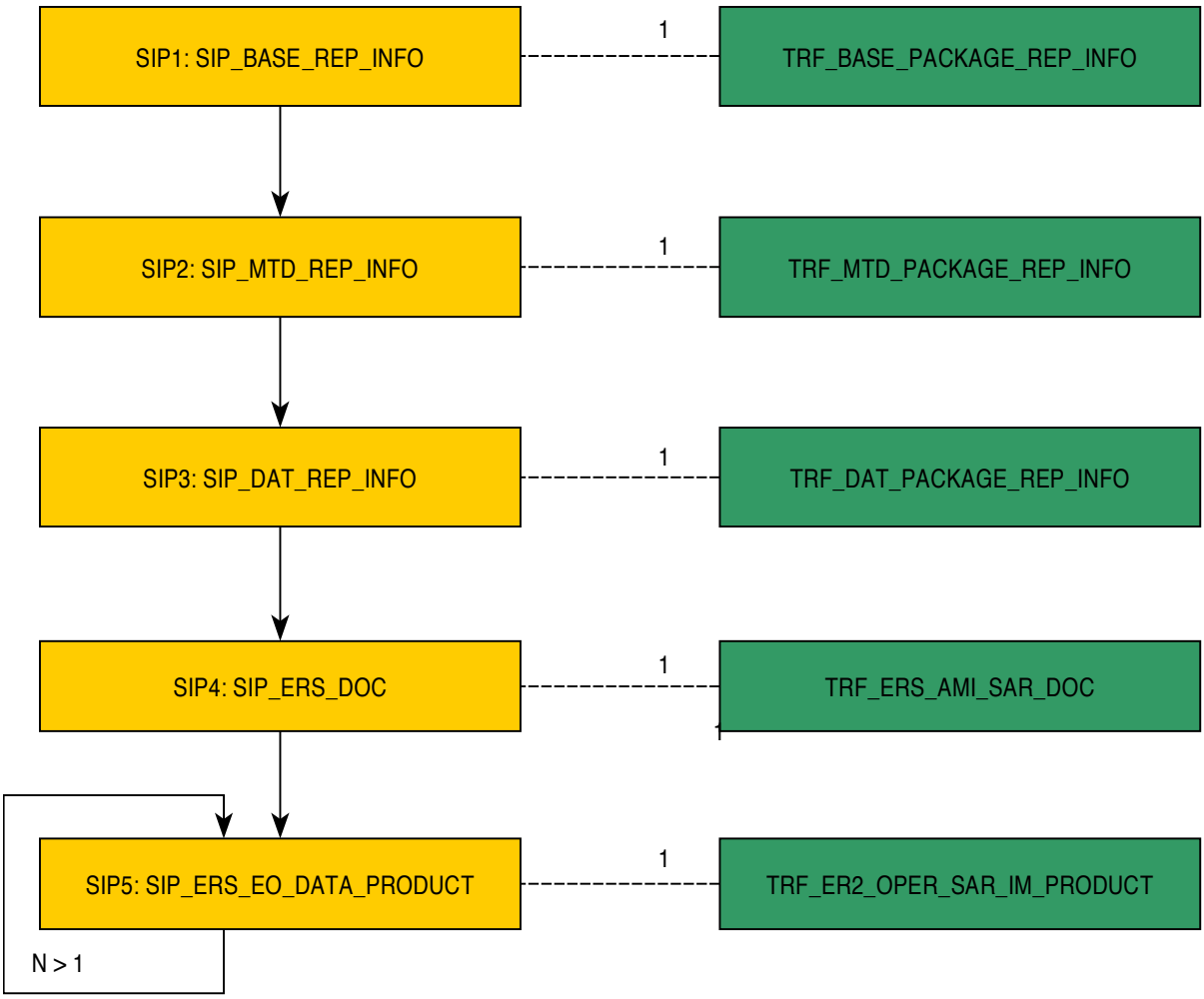

Figure 6-8: ESA SAFE SIP and Sequencing Constraints SIPS

Up to eight SIPs are generated as zip files and transferred to the Archive: the three Representation Information SIPs (SIP1 for BASE, SIP2 for MTD, and SIP3 for DAT), then the documentation SIP4 followed by n (=1 to 4) SIPs (SIP5 up to SIP8), each containing one SAFE Data Product.

A SIP is based on the XFDU standard extended by the PAIS schema and therefore consists of

– An **xfdumanifest.xml** file providing the following information:

  • the packageHeader containing in particular the PAIS sipGlobalInformation: sipID, producerSourceID, producerArchiveProjectID, sipContentTypeID, sipSequenceNumber;

  • the informationPackageMap describing the contained PAIS sipTransferObject corresponding to the Transfer Object Types authorized within the SIP;

- the dataObjectSection listing the different transferred objects: a unique identifier, the size in bytes of the file, its location within the SIP, its MD5 checksum for integrity validation.

– The **different transferred objects** as files or directories/subdirectories containing files.

The following manifest is the one contained in SIP5 corresponding to a SIP transferring a SAFE Data Product in the 'simple case' where the SAFE Data Products are handled as black boxes. Its sequence number is 5; there is only one object corresponding to a zip file containing the Data Product **ER2_OPER_SAR_IM__0P_19970923T212658_19970923T213316_43E3.SAFE.**

Table 6-2: Manifest File of a SIP of Type 'SIP_ERS_EO_DATA_PRODUCT' for 'Simple Case'

PAIS elements/items	Contents	
SIPGlobalInformation		
	sipID : ESA_ERS_AMI_SAR-SIP-0005 **producerSourceID :** ESA_ERS_PROJECT **producerArchiveProjectID:** ESA_ERS_AMI_SAR **sipContentTypeID:** SIP_ERS_EO_DATA_PRODUCT **sipSequenceNumber:** 5	
sipTransferObject		
	descriptorID: TRF_ER2_OPER_SAR_IM_PRODUCT **transferObjectID:** TRF_ER2_OPER_SAR_IM_PRODUCT-0001 **lastTransferObjectFlag:** FALSE	
sipTransferObjectGroup		
	associatedDescriptorGroupTypeID: GroupEODataProductPackagePart	
sipDataObject		
	associatedDescriptorDataID: GroupEODataProductPackagePart_DO **dataObjectPointer:dataObjectID**= "DO-GroupEODataProductPackagePart_DO-0001"	
dataObject		
	ID="DO-GroupEODataProductPackagePart_DO-0001" **size=**"8921"	
	byteStream	
		size="8921" fileLocation locatorType="**URL**" href="file:**ER2_OPER_SAR_IM__0P_19970923T212658_19970923T213316_43E3.SAFE.zip**" checksum checksumName="MD5" **e6c99ab18c96015a267874c371accfc0**

The following manifest is an extract of the one contained in SIP5 corresponding to a SIP transferring a SAFE Data Product in the 'detailed case' where the contents of the SAFE Data Products are detailed. Its sequence number is 5; there are four objects corresponding to the

different files contained in the SAFE Data Product**:** MEASUREMENT.DAT, SAFE-SAR-ERS-AMI-SAR-LEVEL0.XML, MEASUREMENT.SIX and the SAFE manifest file.

Table 6-3: Manifest File of a SIP of Type 'SIP_ERS_EO_DATA_PRODUCT' for 'Detailed Case'

PAIS elements/items	Contents	
SIPGlobalInformation		
	sipID : ESA_ERS_AMI_SAR_det-SIP-0005 **producerSourceID :** ESA_ERS_PROJECT **producerArchiveProjectID:** ESA_ERS_AMI_SAR_det **sipContentTypeID:** SIP_ERS_EO_DATA_PRODUCT_det **sipSequenceNumber:** 5	
sipTransferObject		
	descriptorID: TRF_ER2_OPER_SAR_IM_PRODUCT_det **transferObjectID:** TRF_ER2_OPER_SAR_IM_PRODUCT_det-0001 **lastTransferObjectFlag:** FALSE	
sipTransferObjectGroup		
	associatedDescriptorGroupTypeID: GroupEODataProductPackagePart_det	
	sipTransferObjectGroup	
		associatedDescriptorGroupTypeID: GroupEODataProduct_det **transferObjectGroupInstanceName:** ER2_OPER_SAR_IM__0P_19970923T212658_19970923T213316_43E3.SAFE
sipDataObject		
	associatedDescriptorDataID: GroupEODataProduct_Meas_det **dataObjectPointer:dataObjectID**= " DO-GroupEODataProduct_Meas_det-0001"	
sipDataObject		
	associatedDescriptorDataID: GroupEODataProduct_XML_det **dataObjectPointer:dataObjectID**= " DO-GroupEODataProduct_XML_det-0001"	
sipDataObject		
	associatedDescriptorDataID: GroupEODataProduct_XML_det **dataObjectPointer:dataObjectID**= " DO-GroupEODataProduct_XML_det-0002"	
sipDataObject		
	associatedDescriptorDataID: GroupEODataProduct_SIX_det **dataObjectPointer:dataObjectID**= " DO-GroupEODataProduct_SIX_det-0001"	
dataObject		
	ID=" DO-GroupEODataProduct_Meas_det-0001" **size=**" 2097152"	
	byteStream	
		size=" 2097152" fileLocation locatorType="**URL**" href="file:

PAIS elements/items	Contents	
		ER2_OPER_SAR_IM__0P_19970923T212658_19970923T213316_43E3.SAFE/MEASUREMENT.DAT" checksum checksumName="MD5" **b2d1236c286a3c0704224fe4105eca49**
dataObject		
	ID=" DO-GroupEODataProduct_XML_det-0001" **size=**" 11239"	
	byteStream	
		size=" 11239" fileLocation locatorType="**URL**" href="file: **ER2_OPER_SAR_IM__0P_19970923T212658_19970923T213316_43E3.SAFE/MANIFEST.XML**" checksum checksumName="MD5" **10349ee55dc45733fbb2f674e706a7a8**
dataObject		
	ID=" DO-GroupEODataProduct_XML_det-0002" **size=**" 5458"	
	byteStream	
		size=" 5458" fileLocation locatorType="**URL**" href="file: **ER2_OPER_SAR_IM__0P_19970923T212658_19970923T213316_43E3.SAFE/SAFE-SAR-ERS-AMI-SAR-LEVEL0.XML**" checksum checksumName="MD5" **d81bc5df27f32d9562beeec587b87775**
dataObject		
	ID=" DO-GroupEODataProduct_SIX_det-0001" **size=**" 2772"	
	byteStream	
		size=" 2772" fileLocation locatorType="**URL**" href="file: **ER2_OPER_SAR_IM__0P_19970923T212658_19970923T213316_43E3.SAFE/MEASUREMENT.SIX**" checksum checksumName="MD5" **a7f8208e7e86929979ab0fdd407b605b**

6.4 COROT—END OF MISSION BULK TRANSFER

6.4.1 CONTEXT AND BENEFITS

The CoRoT—End of Mission Bulk Transfer use case deals with the transfer of a full set of science and auxiliary data acquired by the CoRoT space mission. The science data is in an unprocessed form that is referred to as ‘Level 0’, or sometimes as ‘raw’ data. The auxiliary data includes data about the spacecraft and instrument status, such as currents and temperatures, which is referred to as ‘housekeeping’ data. It also includes data from instrument calibration runs. The setup of this use case was contemporary to the actual transfer of CoRoT data from the acquisition centers to the long-term archiving center at CNES after the end of the mission. It served as a PAIS demonstration of capabilities for the bulk transfer of medium-large data sets.

CoRoT is a space astronomy mission devoted to the study of the variability with time of a star's brightness, with an extremely high accuracy (100 times better than from the ground), on very long durations (up to 150 days) and a very high duty cycle (more than 90 percent). The original scientific objectives were focused on the study of stellar pulsations (asteroseismology) and the detection of small exoplanets. However, the data collected are now feeding many domains of stellar physics. The mission was led by CNES in association with four French laboratories, and seven participating countries and agencies (Austria, Belgium, Brazil, Germany, Spain, and the ESA Science Programme). It was launched on December 27, 2006 by a Soyuz Rocket, from Baikonur. The mission lasted almost six years (the nominal three years duration and a three years extension) and has observed more than 160,000 stars. It stopped sending data suddenly on November 2, 2012.

This use case provides an example of PAIS configurations for bulk transfers and highlights the control of sequencing as the housekeeping data needs to be transferred before the science data.

This use case does not cover the transfer of documents or representational information about the content of the transferred files. That information was to be transferred through a separate means. The transfer of metadata, documentation, and the relationships with the actual data is covered by the ISEE use case (see 6.2).

6.4.2 OBJECTS TO BE TRANSFERRED

On the Producer side, the CoRoT repository of Level 0 data and accompanying auxiliary data is composed of about 460,000 files representing a total size of about 3.2 Tb. The file size spans from 1 KB to 1.2 GB with an average value of 1.5 MB.

The data chosen were resident on a portable external hard drive provided by the CoRoT project at CNES; however, they have been truncated to facilitate testing with this use case and no longer bear any real resemblance to the actual CoRoT data.

The logical layout of the repository is depicted in figure 6-8 below. The logical layout, also known as static layout, represents the hierarchy of directory and file 'types' with their cardinalities. It does not represent the full list of directory and file instances.

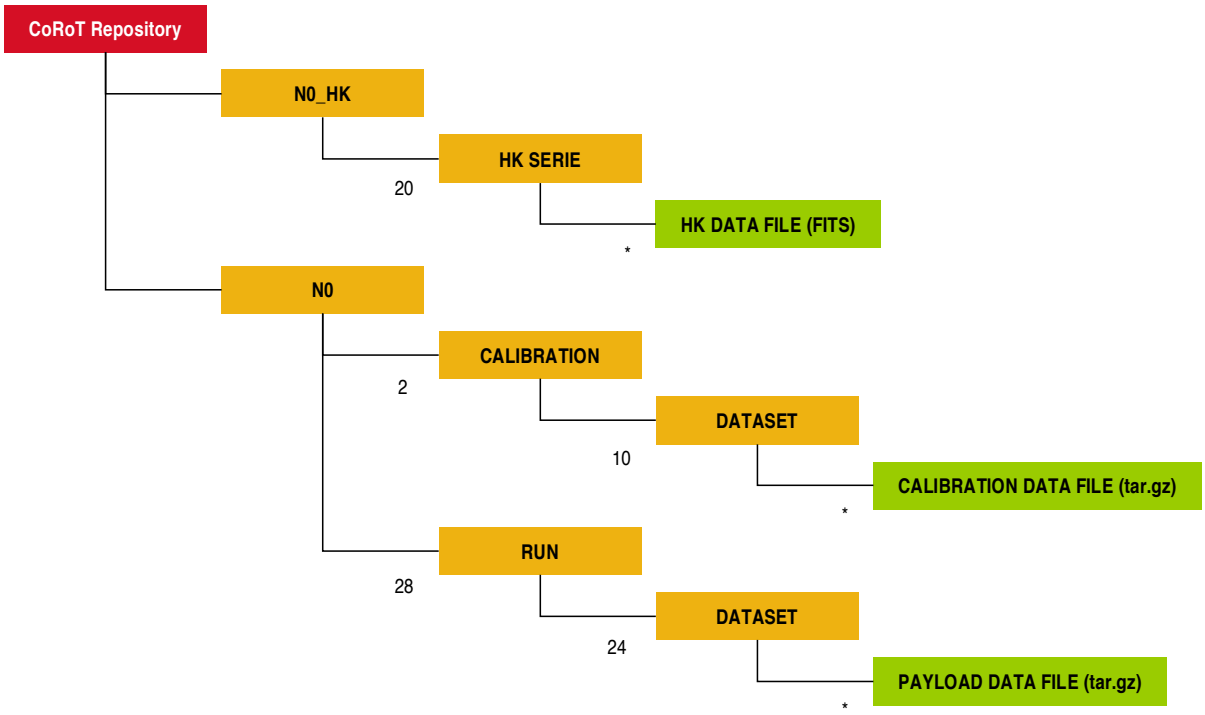

Figure 6-9: CoRoT Repository—Logical Layout

As mentioned before, the CoRoT data to be transferred consist of auxiliary (housekeeping and calibration) and Level 0 data.

The housekeeping data are those dedicated to the Level 0 data only and are filed under an `N0_HK` folder of the repository. The `N0` is an abbreviation of 'Niveau 0' meaning 'Level 0' in French, and the `HK` stands for housekeeping. The housekeeping data is then distributed in 20 distinct series of parameters as currents, temperatures, etc. The folders of '`HK SERIE`' type depicted in the above figure can take the following names: `FRACTIOPPS1`, `FRACTIOPPS2`, `LATCHEDOBT`, etc. Those folders contain a variable number of files formatted according to the Flexible Image Transport System (FITS) specifications (a common space science format).

The Level 0 data is filed in an `N0` folder containing 28 sub-folders corresponding to the CoRoT observation 'runs' that represent continuous observations of 20 to 150 days. The run folder names follow the pattern `RUN{NN}_{CODE}` where `{NN}` is a counter and the `{CODE}` is a non-null string whose definition will not be detailed here, e.g., `RUN12_LRC03`. The counter starts with '03'. A separate run named `CALIBRATION` was used for calibration and validation of the CoRoT telescope. It has the same structure and content types as the other runs.

Each run folder is further subdivided into subfolders called datasets corresponding to different types of data or different processing levels. A dataset is a set of files in FITS format that are archived and compressed in tar/gz format. A run may not contain all datasets but most of them include a majority of the datasets defined in table 6-4.

Table 6-4: CoRoT Level-0 Datasets

Dataset	Total Size	Max. File Size	File Number	Title
AN0_BKGROUND	16.8 GB	392 MB	190	Asteroseismology Background
AN0_ECARTO_AFPS	6.8 GB	404 MB	43	Ecartometry Fine Pointing mode
AN0_ECARTO_ARPS	66.5 Mo	14 MB	38	Ecartometry Rough Pointing mode
AN0_ECARTO_Undefined	1.1 MB	0.2 MB	60	Ecartometry mode non-detected
AN0_FULLIMAGE	549 MB	29.6 MB	76	Asteroseismology Full Image
AN0_FULLWINDOW	317.3 MB	35.8 MB	36	Asteroseismology Full Window
AN0_IMAGETTE❶	**82.80 GB**	**1.3 GB**	**194**	**Asteroseismology imagette**
AN0_MASK	164 KB	2 KB	141	Asteroseismology templates
AN0_OFFSET	6.62 GB	268 MB	58	Asteroseismology Offset
AN0_STARWIND	15.85 GB	185 MB	216	Asteroseismology channel
AN0_THRESHOLDIMAGE	65.7 GB	11 MB	16	Asteroseismology Threshold Image
EN0_BKGROUND_MONOCHROM	1.37 GB	1.2 MB	5544	Monochr. Exoplanet Bkgd.
EN0_BKGROUND_SAMPLEM	3.04 GB	7.6 MB	1623	Monochr. Oversampled Exoplanet Bkgd.
EN0_BRIGHT_PIX_32	1.64 GB	5.3 MB	1372	Exoplanet sky bkg. impacted pixels 32
EN0_BRIGHT_PIX_512	710 MB	768 KB	4116	Exoplanet sky bkg. impacted pixels 512
EN0_FULLIMAGE	1.71 GB	111 MB	74	Exoplanet Full Image
EN0_FULLWINDOW	282.2 MB	663 KB	6898	Exoplanet Full Window
EN0_IMAGETTE	72.64 GB	168 MB	1015	Exoplanet imagette
EN0_OFFSET_SAMPLEM	1.08 GB	32 MB	74	Offset Exoplanet Oversampled Mono.
EN0_STARWIND_CHROM	17.34 GB	969 MB	45987	Chromatic Exoplanet Observation
EN0_STARWIND_MONOCHROM	28.3 GB	567 MB	120143	Mono. Exoplanet Observation
EN0_STARWIND_SAMPLEC❷	**101.7 GB**	**6.5 MB**	**41917**	**Chromatic Overspld. Exoplanet Obs.**
EN0_STARWIND_SAMPLEM	9.48 GB	3.7 MB	7777	Mono. Oversampled Exoplanet Obs.
EN0_TEMPLATE	6.08 MB	1 KB	6658	Exoplanet Templates

The scientific or engineering nature of the datasets will not be further described in this use case. Their understanding could have helped to establish a more detailed model of transfer, for example, by describing relationships between the datasets. However, the above table provides the size and count properties that are significant for the design of the transfer model. For example, the total size of the EN0_STARWIND_SAMPLEC ❷ exceeds 100 GB and shows that it is probably not a good idea to define CoRoT datasets as the smallest unit for the transfer model. A modeling down to the file level would allow the transfer of packets of

more reasonable sizes. At the file level, the table shows that the maximum size of the individual files is 1.3 GB for the `AN0_IMAGETTE` ❶ which is the smallest maximum Transfer Object size that the model can impose for this dataset since the PAIS does not allow a file to span over multiple Transfer Objects or SIPs. So the maximum file size is an interesting property used to define the transfer. If this value is not reasonable, it would have been necessary to consider a change, for example, by slicing some of the big files into chunks of acceptable sizes.

Figure 6-10 shows a partial view of the physical layout of the CoRoT repository. Because of the large number of files, it is not possible to depict all of them, but the figure provides actual examples of folder and file names for CoRoT runs, datasets, or housekeeping series.

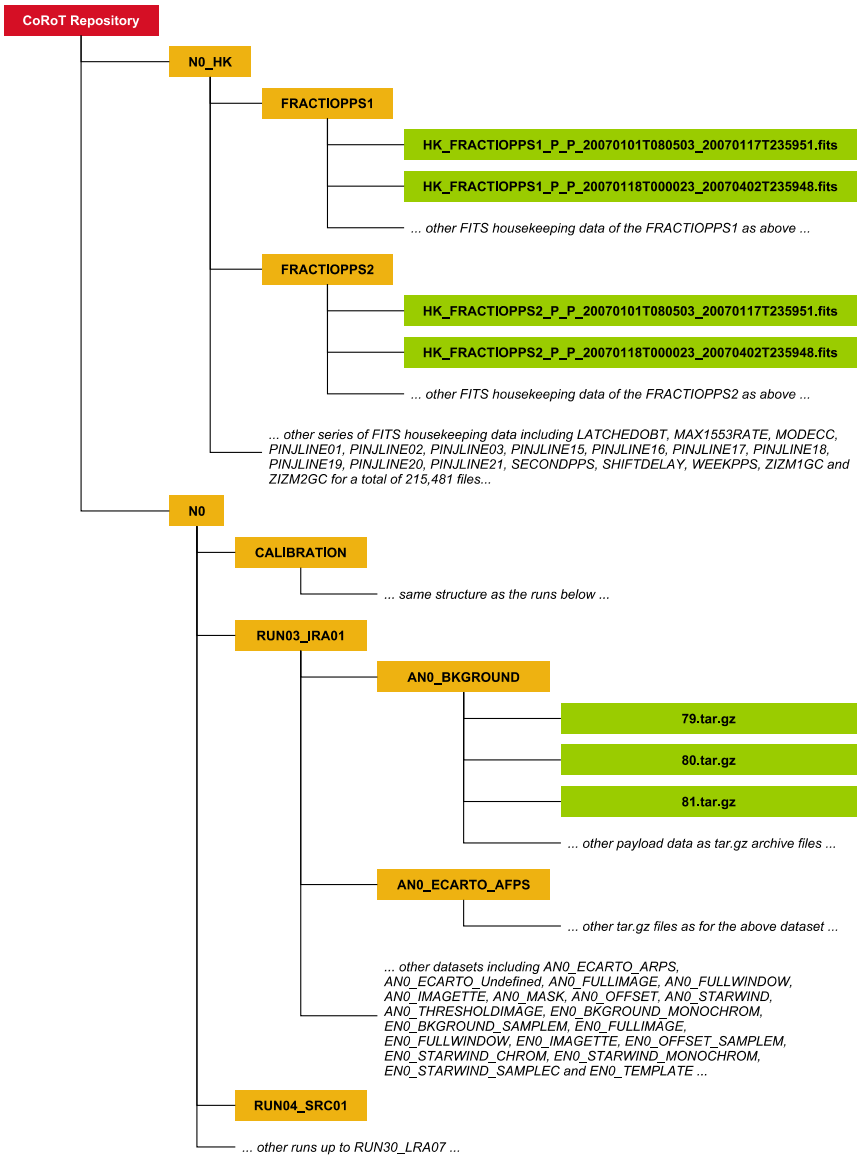

Figure 6-10: CoRoT Repository Physical Layout

6.4.3 MODEL OF OBJECTS FOR TRANSFER AND SIP CONSTRAINTS

6.4.3.1 General

The transfer of a full run as a single SIP is not practical because some may exceed 100 GB. The transfer model built splits the runs into subparts of less than 4 GB, but with the assurance that each transferred SIP deals with only one run and only one type of dataset, e.g., RUN03 and AN0_BACKGROUND. However, a single run may require multiple SIPs to be fully transferred, depending on the total size of the dataset. (See figure 6-11.)

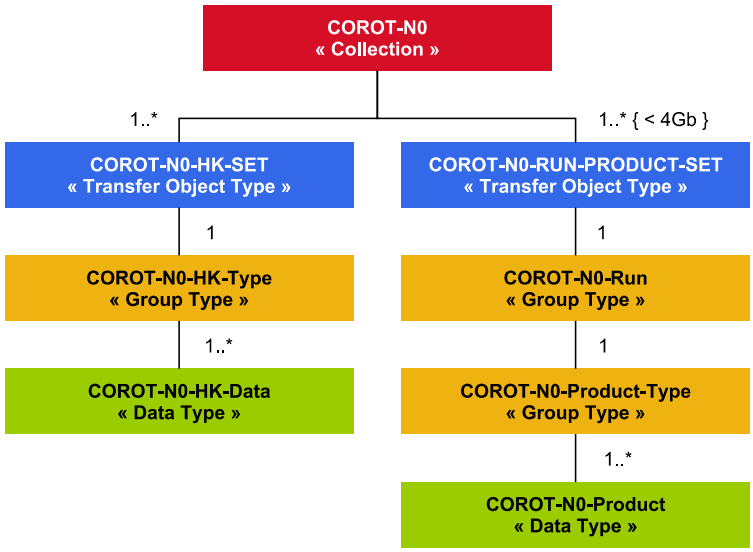

Figure 6-11: CoRoT Model of Objects for Transfer

6.4.3.2 MOT

The XML Descriptors are provided in annex D.

The Model of Objects for Transfer, which is a hierarchical tree, can be summarized as follows:

– One root collection ‘CoRoT-N0’.

– A first Transfer Object Type ‘CoRoT-N0-RUN-PRODUCT-SET’ represents the payload stream of CoRoT N0 products. There can be an unlimited number of objects of this type in the transfer project but each has a limited size of 4 GB. It is made up of:

  • A ‘CoRoT-N0-Run’ Group Type—assures that a ‘CoRoT-N0-RUN-PRODUCT-SET’ object contains data dealing with one and only one RUN.

- A ‘CoRoT-N0-Product-Type’ Group Type—assures that a ‘CoRoT-N0-RUN-PRODUCT-SET’ object contains data of the same set, e.g., AN0_BACKGROUND.

– A second Transfer Object Type ‘CoRoT-N0-HK-SET’ represents the payload stream of CoRoT HK auxiliary data. There can be an unlimited number of objects of this type in the transfer project with no constraint on size. It contains:

- A ‘CoRoT-N0-HK-Type’ Group Type—assures that a ‘CoRoT-N0-HK-SET’ object contains data dealing with one and only one series, e.g., FRACTIOPPS1.

Figure 6-12 below is a snapshot of the MOT viewed through the CNES prototype. It shows the collections and the Transfer Objects levels as described in this section.

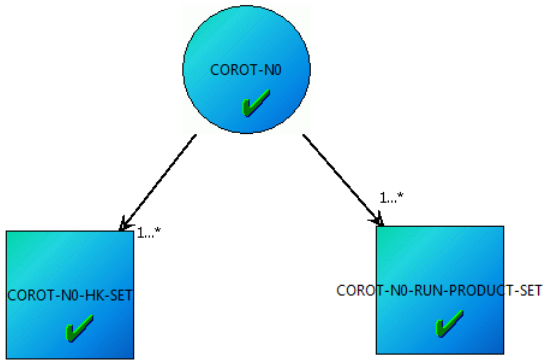

Figure 6-12: CoRoT Model of Objects for Transfer

6.4.3.3 SIP Constraints

The SIP Constraints XML document is provided in annex subsection D5.

This test case defines two SIP Content Types, one for each of the categories of data:

– **SIP-CoRoT-N0-PRODUCT-SET**—authorizing only one CoRoT-N0-RUN-PRODUCT-SET Transfer Object per SIP of this type. It corresponds to one homogeneous set of dataset of the same type accumulated up to a maximum of 4 GB.

– **SIP-CoRoT-N0-HK-SET**—authorizing only one CoRoT-N0-HK-SET Transfer Object per SIP of this type. It corresponds to one homogeneous set of FITS files accumulated without limit of count or size.

This test case also defines a sequence order imposing the housekeeping data to be transferred first or at least before any N0 product.

Currently SIP Constraints are created manually via an XML editing tool or a text editing tool.

6.4.4 SIPS

6.4.4.1 SIPs Generation

In this case, the SIPs are generated from the CoRot repository via the SIP Builder software (see 7.3) with a configuration file provided in annex subsection D8. The generated SIPs are 'XFDU PAIS SIP Conformant' as defined in the PAIS standard.

173 SIPs have been generated with the first 20 SIPs conveying housekeeping auxiliary data followed by 153 packages of N0 products. More SIPs should have been generated from the CoRoT repository but for the demonstration the process has been voluntarily limited to the first nine runs, i.e., RUN01 to RUN09.

6.4.4.2 SIPs Contents

The first series of SIPs are, as expected, of SIP-CoRoT-N0-HK type illustrated by the abstract tree below:

```
├── CoRoT-N0-SIP-0001.zip          <—— First HK SIP
│   ├── N0_HK
│   │   └── FRACTIOPPS1
│   │       ├── HK_FRACTIOPPS1_P_P_20070101T080503_20070117T235951.fits
│   │       ├── HK_FRACTIOPPS1_P_P_20070118T000023_20070402T235948.fits
│   │       ├── ...
│   │       ├── HK_FRACTIOPPS1_P_P_20120705T000009_20121001T235932.fits
│   │       └── HK_FRACTIOPPS1_P_P_20121001T000004_20121103T235941.fits
│   └── xfdumanifest.xml
├── CoRoT-N0-SIP-0002.zip
│   ├── N0_HK
│   │   └── FRACTIOPPS2
│   │       ├── HK_FRACTIOPPS2_P_P_20070101T080503_20070117T235951.fits
│   │       ├── HK_FRACTIOPPS2_P_P_20070118T000023_20070402T235948.fits
│   │       ├── ...
│   │       ├── HK_FRACTIOPPS2_P_P_20120705T000009_20121001T235932.fits
│   │       └── HK_FRACTIOPPS2_P_P_20121001T000004_20121103T235941.fits
│   └── xfdumanifest.xml
├── CoRoT-N0-SIP-0003.zip
│   ├── N0_HK
│   │   └── LATCHEDOBT
│   │       ├── HK_LATCHEDOBT_P_P_20070101T080431_20070117T235951.fits
│   │       ├── HK_LATCHEDOBT_P_P_20070118T000023_20070402T235948.fits
│   │       ├── ...
│   │       ├── HK_LATCHEDOBT_P_P_20120705T000009_20121001T235932.fits
│   │       └── HK_LATCHEDOBT_P_P_20121001T000004_20121103T235941.fits
│   └── xfdumanifest.xml
├── ...
└── CoRoT-N0-SIP-0020.zip          <—— Last HK SIP
    ├── N0_HK
    │   └── ZIZM2GC
    │       ├── HK_ZIZM2GC_P_P_20070226T103801_20070402T235956.fits
    │       ├── HK_ZIZM2GC_P_P_20070402T000004_20070509T235954.fits
    │       ├── ...
    │       ├── HK_ZIZM2GC_P_P_20120705T000001_20121001T235956.fits
    │       └── HK_ZIZM2GC_P_P_20121001T000004_20121103T235957.fits
    └── xfdumanifest.xml
```

Tables 6-5–6-7 show snippets of the manifest `xfdumanifest.xml` file of the first SIP containing the first housekeeping series. The extracts are focused on the PAIS extension elements and are generally stripped of XFDU elements. These latter are reported only when they bring value to the example and are explicitly prefixed by `xfdu` to avoid confusion, although that may not be the case in the actual manifest file. The complete manifest document is provided in subsection D6 of annex D.

Table 6-5: SIP-CoRoT-N0-HK SIP Manifest—Header

Element	Content
sipGlobalInformation ❶	
sipID	**CoRoT-N0-SIP-0001**
producerSourceID	CNES
producerArchiveProjectID	CoRoT-N0
sipContentTypeID	**SIP-CoRoT-N0-HK**
sipSequenceNumber	1

The header contains general information associated to the whole package: identifier of SIP (created during SIP construction, uniqueness to be checked within the Producer-Archive Project), type of SIP (checked against the SIP constraints where it has been defined). The Producer Source ID and the Producer-Archive Project ID make links with the MOT where they have been defined (the Producer-Archive Project ID is the ID of the root node in the MOT).

Table 6-6: SIP-CoRoT-N0-HK SIP Manifest—Information Package Map

sipTransferObject ❷		
	descriptorID	**CoRoT-N0-HK**
	transferObjectID	CoRoT-N0-HK-0001
	lastTransferObjectFlag	FALSE
	sipTransferObjectGroup ❸	
	associatedDescriptorGroupTypeID	CoRoT-N0-HK-GROUP
	transferObjectGroupInstanceName	**FRACTIOPPS1**
❹	**sipDataObject**	
	associatedDescriptorDataID	**CoRoT-N0-HK-DATA**
	xfdu:dataObjectPointer	
	@dataObjectID	**DO-CoRoT-N0-HK-DATA-0001**
	sipDataObject	
	associatedDescriptorDataID	CoRoT-N0-HK-DATA
	xfdu:dataObjectPointer	
	@dataObjectID	**DO-CoRoT-N0-HK-DATA-0029**

The Information Package Map describes the hierarchical content of the package by making links with the MOT through the Descriptor, Descriptor Group Types, and Descriptor Data IDs. These IDs are checked against the MOT for conformity with the expected Data Objects. The Transfer Object ID is inserted during SIP building and identifies the Transfer Object. This ID should be kept in a log, for potential update or deletion (this is the lowest delivery granule). Transfer Object Group Instance Name is the instantiated name of the Transfer Object Group Type Structure specified in the MOT (without path information).

The Information Package Map also points towards the physical Data Objects in the Data Object section through the Data Object Pointers.

Table 6-7: SIP-CoRoT-N0-HK SIP Manifest—Data Object Section

xfdu:dataObject	
@ID	**DO-CoRoT-N0-HK-DATA-0001**
byteStream	
fileLocation	
@locatorType	URL
@href	**N0_HK/FRACTIOPPS1/HK_FRACTIOPPS1_P_P_20070101T080503_20070117T235951.fits**
Checksum	d41d8cd98f00b204e9800998ecf8427e
@checksumName	MD5
xfdu:dataObject	
@ID	**DO-CoRoT-N0-HK-DATA-0029**
byteStream	
fileLocation	
@locatorType	URL
@href	**N0_HK/FRACTIOPPS1/HK_FRACTIOPPS1_P_P_20121001T000004_20121103T235941.fits**
Checksum	d41d8cd98f00b204e9800998ecf8427e
@checksumName	MD5

The Data Object Section contains the physical location of the Data Objects as described in the Information Package Map. This is also the place to indicate checksums or file sizes.

The second series of SIPs is, as expected, of SIP-CoRoT-N0-RUN type, illustrated by the abstract tree below:

```
├── CoRoT-N0-SIP-0021.zip          <——— First N0 for 'RUN' data SIP
│   ├── N0
│   │   └── RUN03_IRA01
│   │       └── AN0_BKGROUND
│   │           ├── 79.tar.gz
│   │           ├── 80.tar.gz
│   │           ├── 81.tar.gz
│   │           ├── 82.tar.gz
│   │           └── 83.tar.gz
│   └── xfdumanifest.xml
├── CoRoT-N0-SIP-0022.zip
│   ├── N0
│   │   └── RUN03_IRA01
│   │       └── AN0_ECARTO_AFPS
│   │           ├── 0000000116.tar.gz
│   │           └── 0000000223.tar.gz
│   └── xfdumanifest.xml
├── CoRoT-N0-SIP-0023.zip
│   ├── N0
│   │   └── RUN03_IRA01
│   │       └── AN0_ECARTO_ARPS
│   │           ├── 0000000116.tar.gz
│   │           └── 0000000223.tar.gz
│   └── xfdumanifest.xml
├── ...
└── CoRoT-N0-SIP-0173.zip          <——— Last N0 SIP
    ├── N0
    │   └── RUN09_SRC02
    │       └── EN0_TEMPLATE
    │           ├── 0.tar.gz
    │           ├── 100.tar.gz
    │           ├── ...
    │           ├── 98.tar.gz
    │           └── 99.tar.gz
    └── xfdumanifest.xml
```

SIP-021 to SIP-0173 contains the scientific Level 0 data. The tabulated snippets represent the three nested repositories containing the data grouped in the form of tar limited in size (as specified in the MOT).

6.4.4.3 SIPs Ingestion

In this case, the SIPs are submitted and ingested by the CNES Prototype (see 7.2).

The CNES Prototype main validation and ingestion steps are shown in figure 6-13.

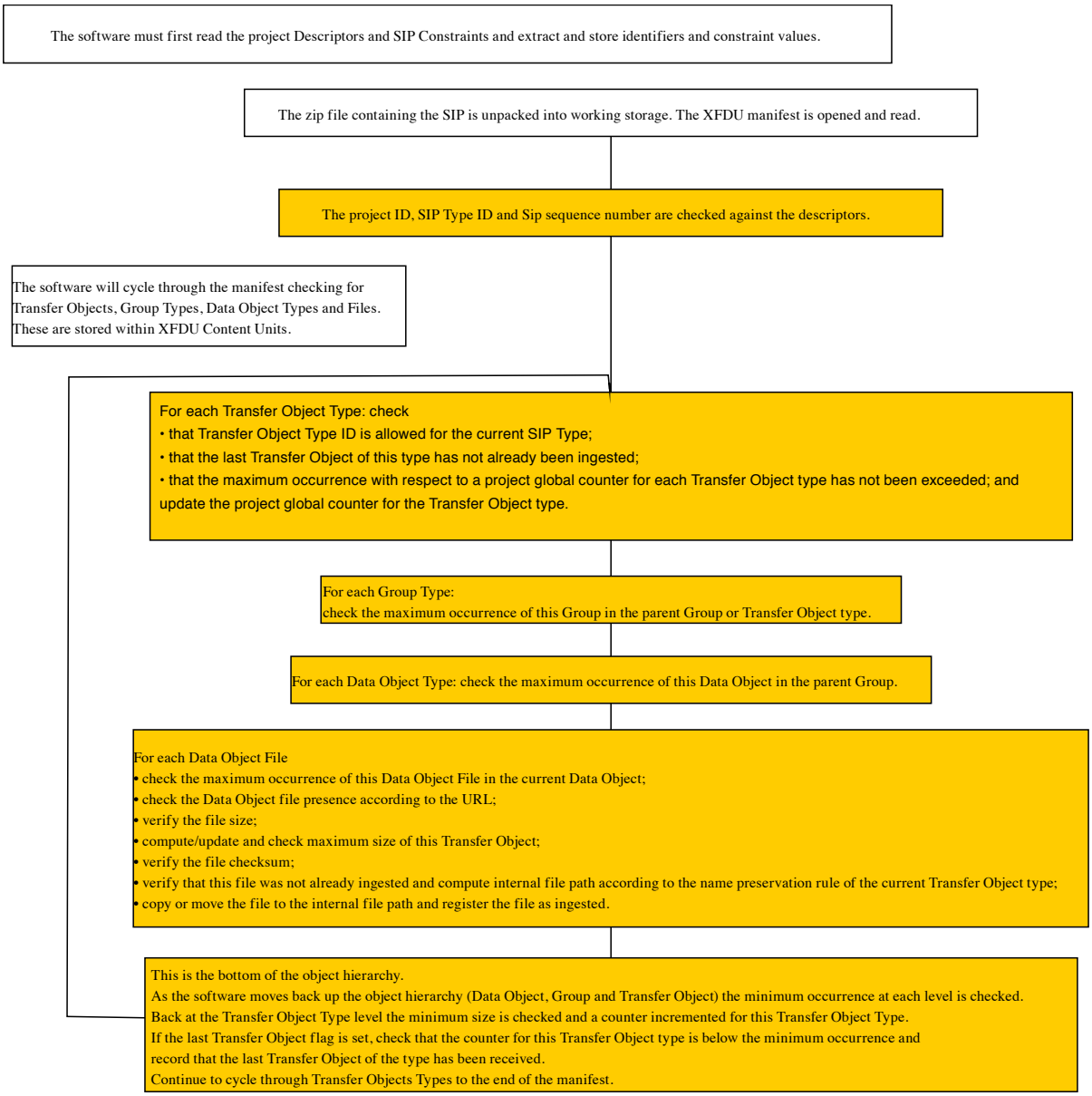

Figure 6-13: SIP Main Validation and Ingestion Steps

In this case the Archive internal repository is a reconstruction of the original CoRoT repository on the Producer side. The three examples below show the progressive construction of the repository on the Archive side after ingestion.

Example of Archive internal repository after ingestion of the first SIP of CoRoT housekeeping data:

```
└── N0_HK
    └── FRACTIOPPS1    <——— From the first HK SIP
        ├── HK_FRACTIOPPS1_P_P_20070101T080503_20070117T235951.fits
        ├── HK_FRACTIOPPS1_P_P_20070118T000023_20070402T235948.fits
        ├── ...
        ├── HK_FRACTIOPPS1_P_P_20120705T000009_20121001T235932.fits
        └── HK_FRACTIOPPS1_P_P_20121001T000004_20121103T235941.fits
```

Example of Archive internal repository after ingestion of second SIP of CoRoT housekeeping data:

```
└── N0_HK
    ├── FRACTIOPPS1
    │   ├── HK_FRACTIOPPS1_P_P_20070101T080503_20070117T235951.fits
    │   ├── HK_FRACTIOPPS1_P_P_20070118T000023_20070402T235948.fits
    │   ├── ...
    │   ├── HK_FRACTIOPPS1_P_P_20120705T000009_20121001T235932.fits
    │   └── HK_FRACTIOPPS1_P_P_20121001T000004_20121103T235941.fits
    └── FRACTIOPPS2    <——— From the second HK SIP
        ├── HK_FRACTIOPPS2_P_P_20070101T080503_20070117T235951.fits
        ├── HK_FRACTIOPPS2_P_P_20070118T000023_20070402T235948.fits
        ├── ...
        ├── HK_FRACTIOPPS2_P_P_20120705T000009_20121001T235932.fits
        └── HK_FRACTIOPPS2_P_P_20121001T000004_20121103T235941.fits
```

Example of Archive internal repository after ingestion of first SIP of CoRoT Level 0 data:

```
├── N0_HK
│   ├── FRACTIOPPS1
│   │   ├── HK_FRACTIOPPS1_P_P_20070101T080503_20070117T235951.fits
│   │   ├── HK_FRACTIOPPS1_P_P_20070118T000023_20070402T235948.fits
│   │   ├── ...
│   │   ├── HK_FRACTIOPPS1_P_P_20120705T000009_20121001T235932.fits
│   │   └── HK_FRACTIOPPS1_P_P_20121001T000004_20121103T235941.fits
│   ├── FRACTIOPPS2
│   │   ├── HK_FRACTIOPPS2_P_P_20070101T080503_20070117T235951.fits
│   │   ├── HK_FRACTIOPPS2_P_P_20070118T000023_20070402T235948.fits
│   │   ├── ...
│   │   ├── HK_FRACTIOPPS2_P_P_20120705T000009_20121001T235932.fits
│   │   └── HK_FRACTIOPPS2_P_P_20121001T000004_20121103T235941.fits
│   │
│   └── … <——— Other HK folders not represented here for brevity
└── N0
    └── RUN03_IRA01
        └── AN0_BKGROUND     <——— From the first N0 SIP
            ├── 79.tar.gz
            ├── 80.tar.gz
            ├── 81.tar.gz
            ├── 82.tar.gz
            └── 83.tar.gz
```

6.5 BNF AND METS—A NON-XFDU SIP IMPLEMENTATION

6.5.1 GENERAL

The BnF is not currently using PAIS in its SPAR. Nonetheless, PAIS could be profitably implemented in the case of complex and predictable Transfer Objects. The METS is a widely used standard for transmission of metadata and packaging information within the cultural heritage and library communities. That is the main reason why the BnF chose METS as the format for packaging information after a preliminary study. METS and XFDU provide similar high level sections to describe Data objects and structure of the Information packages. This test case describes one (among others) possible implementation of PAIS elements in METS.

In order to analyze the benefits of an implementation of PAIS in SPAR, a simple example has been taken from the library's collections: a periodical issue of the newspaper *Le Temps* published on April 29th of 1861. A Model of Transfer is created for submission of any number of issues, but only the transfer of that single issue is detailed below.

6.5.2 TRANSFERRING PERIODICAL ISSUES OF A DIGITIZED NEWSPAPER

6.5.2.1 Context

BnF is digitizing French printed heritage for two decades. The automatic preservation of digitized material in SPAR began in 2010. Although born-digital documents are collected now and stored in SPAR, the reproduction of analog documents is still the major workflow.

The example now presented is the transfer of a digitized issue of an ancient newspaper. This kind of digital material, preserved into SPAR *via* a 'track' (a specific workflow) named 'fil_num_cons', shares the following properties:

- it is a digital version of an analog document;
- it is a specified type of a visual document (manuscript, newspaper, book, photograph, engraving, etc.);
- its content is copyright free (but its use may be restricted);
- it is to be long-term preserved.

The FIL_NUM_CONS track is subdivided into several channels defined by their technical peculiarities. In each track, other channels can be defined if needed as new specific types of documents appear to be preserved.

- FIL_NUM_CONS_A: 2D documents belonging to the BnF, digitized by its own service providers;
- FIL_NUM_CONS_B: 2D documents belonging to the BnF digitized by partners following unusual technical specifications;

– FIL_NUM_CONS_C: 2D documents belonging to partner institutions;

– FIL_NUM_CONS_D: 3D documents belonging to the BnF.

As the first objective of SPAR was to preserve the digitization of printed documents, the channel FIL_NUM_CONS_A was the first to be implemented in SPAR. Indeed, BnF considers that differences between original document types (manuscripts, engravings, etc.) do not imply fundamental differences between their digitized copies. In this example, FIL_NUM_CONS_A is considered as the PAIS 'project'.

Requirements on the SIP structure and on the transferred objects are expressed by Service Level Agreements (SLAs) and by the METS profile. METS manifests conformance with the submission agreement between the Producer and the Archive are checked against the SLAs (SIP maximum and minimum size, service availability, file formats accepted, etc.). METS profiles encapsulate other requirements expressed using schematron (checking the structure of the manifest, the required metadata) and are applied to the SIP manifest.

The main reason why SPAR has not implemented strict rules about the Transfer Objects structure is because it relies on previous controls implemented in the digitization delivery chain. Considered as the Producer by SPAR, the digitization delivery chain carries out many tests about the Transfer Objects conformance to BnF requirements (e.g., Transfer Object groups expected, image resolution and compression, color depth, etc.). When the SIP is ingested by SPAR, controls focus on the validity of the transferred files (specified format characteristics).

The other reason is because the digital documents preserved in SPAR do not have a complex or predictable structure (number of pages, number of issues are extremely variable). However, as new documents preserved in SPAR are more complex (3D digitized objects) and/or born-digital material (Web archive), PAIS mechanisms could prove to be useful to SPAR.

6.5.2.2 Objects to Be Transferred

The Transfer Objects are periodical issues of a newspaper. Each issue is transferred by a single SIP (to date, SIPs are always carrying a single Transfer Object). An issue is composed of one or two file groups: a 'master' group (image files produced by the digitization of the issue) and an 'ocr' group (text files produced by the optical character recognition of the images). This latter group is not mandatory.

Digitized old newspapers have no predictable structure: the publication might be irregular, the library's collection might be incomplete and the number of pages of each issue is unknown. Moreover, at all times, the formerly transferred objects can be updated by a new Transfer Object group (e.g., image master files completed by XML-ALTO OCR files and/or by an EPUB file).

Issues are linked to a collection that represents all the maintained issues of the periodical. These collections are preserved in SPAR as AICs with no Data Object associated. However, the 'digitized_newspapers_1' collection does not exist in SPAR (the information about the document's being a periodical is only mentioned in the descriptive metadata). Nonetheless, it could be implemented, considering that digitized newspapers have some characteristics in common.

Collection - digitized_newspapers_1

Collection - ark:/12148/cb34431794k (ID for the newspaper *Le Temps* (1861–1942))

Transfer Object - issue (1..n)

Group Object - master (1..1)

Data Object - image (1..n)

Group Object - ocr (0..1)

Data Object - text (1..n)

NOTE – As the requirements for Transfer Objects have changed since 2010, the element descriptorModelVersion is used to record the current version of the Descriptor.

6.5.2.3 Model of Objects for Transfer and SIP Constraints

6.5.2.3.1 MOT

The MOT is extremely simple, as there is only one Transfer Object Type (i.e., issue) allowed for the collection ark:/12148/cb34431794k ('*Le Temps*').

Figure 6-14: Model of Objects for Transfer

As a Transfer Object corresponds to a single issue, there will be as many such Transfer Objects as there are issues in the 'Le Temps' collection. Each Transfer Object will have a single master group containing one or more Data Objects where each corresponds to a single page. Each Transfer Object may also have an OCR group containing one or more Data Objects where each corresponds to a single page. The collection and Transfer Object Descriptors are given in annex subsections E1, E2, and E3.

6.5.2.3.2 SIP Constraints

For an 'issue_ SIP', the only Transfer Object that can be transferred is of 'issue' type. Only one "issue" is accepted in this SIP.

Unlike other projects (e.g., Web harvesting), there are no different SIP content types and no sequencing constraints. If a complementary set of files needs to be added to an AIP, the Producer needs to create a new complete SIP, subjected to the same constraints as the first delivery SIP. However, another choice could have been to create a second SIP content type "OCR_addition" which would transfer a Transfer Object with a single Transfer Object Group of "ocr" type. The SIP Constraints specification is given in annex subsection E4.

6.5.2.4 PAIS SIP Implementation in METS

6.5.2.4.1 Principles

In METS, extensibility is permitted in the elements <dmdSec>, descriptive metadata section, <amdSec>, administrative metadata section, and its child elements, i.e., <techMD>, technical metadata, <rightsMD>, rights metadata, <digiprovMD>, digital provenance metadata, and <sourceMD>, source metadata. BnF suggests integrating:

- the SIP Global Information metadata into a digiprovMD element (information about the ingest process);
- the Transfer Object, the Transfer Object Group, the Data object metadata into a techMD element.

METS metadata sections need to be associated to corresponding sections in <fileSec> and <structMap> elements using an ID/IDREF mechanism. The BnF's choices regarding elements, considered as Transfer Object, Transfer Object Group, and Data Object description levels, to which PAIS metadata are related, are described below.

6.5.2.4.2 BnF Test Case

Every Transfer Object stored in SPAR has a physical structure expressed in the METS element <structMap>. The table below presents the generic nested architecture and the corresponding description levels of a periodical.

Generic structure	Periodical description level
<structMap TYPE="physical"> <fptr>	 Periodical Issue Page File

This structure partly corresponds to the Transfer Objects description levels, except on one point: the Transfer Object Groups are based on a technical representation of the Transfer Object (master, OCR, epub).

The byteStream level can only be associated with the <fileSec><fileGrp><file><stream> level, as in METS the <stream> element is only authorized in a <file> element.

 (associated with SIP Global Information)

 (associated with Transfer Object)

<fileSec><fileGrp> (associated with Transfer Object group)

<fileSec><fileGrp><file> (associated with Data object)

<fileSec><fileGrp><file><stream> (associated with byteStream)

The Transfer Object is the newspaper issue (= the 'group' level object described in the manifest). The Transfer Object metadata will be associated to the 'group' level in the <structMap>. Transfer Object groups PAIS metadata are associated with the <fileGrp>s in the <fileSec>.

The Data Object level is the file level. It can be associated to the <structMap> <fptr>.

NOTES

1 As BnF uses PREMIS to describe events, PAIS SIP Global Information metadata could be integrated into the PREMIS extension of an 'ingest completion' PREMIS event. Once ingested, some packages happen to be disseminated, enriched, and ingested again; the original SIP Global Information could this way be preserved and dated.

2 As stated above, there are many ways to implement PAIS in METS. One alternative would have been to create a new <structMap> with an attribute TYPE="transfer" with nested corresponding to PAIS levels, as follows.

```
<structMap TYPE="transfer">
   <div TYPE="SIP" ID="DIV.1" ADMID="AMD.22">
     <div TYPE="transferObject" ID="DIV.2" ADMID="AMD.30">
       <div TYPE="transferObjectGroup" LABEL="master" ID="DIV.3" ADMID="AMD.31">
         <div TYPE="dataObject" ORDER="1" ID="DIV.4" ADMID="AMD.101">
          <fptr FILEID="master.1"/>
         </div>
         <div TYPE="dataObject" ORDER="2" ID="DIV.5" ADMID="AMD.101">
          <fptr FILEID="master.2"/>
         </div>
         <!-- etc. -->
       </div>
       <div TYPE="transferObjectGroup" LABEL="ocr" ID="DIV.6" ADMID="AMD.32">
         <div TYPE="dataObject" ORDER="1" ID="DIV.7" ADMID="AMD.102">
          <fptr FILEID="ocr.1"/>
         </div>
         <div TYPE="dataObject" ORDER="2" ID="DIV.8" ADMID="AMD.102">
           <fptr FILEID="ocr.2"/>
         </div>
         <!-- etc. -->
       </div>
     </div>
   </div>
 </structMap>
```

For comparison, a mapping between the PAIS abstract SIP and both the XFDU and METS packaging is given in annex subsection E6. Again, other mappings to METS are possible.

6.5.2.4.3 SIPs

SIPs are created by a special pre-ingest module as soon as the digitization delivery chain has performed the tests and validated the Data Objects. They are zipped folders containing two sets of files (TIFF master files and XML-ALTO OCR files) and a METS XML manifest.

Table 6-8: PAIS SIP Elements and Values

Element	Content
sipGlobalInformation	
sipID	**NUM_220908_20140201155415**
producerSourceID	NUM
producerArchiveProjectID	info:bnf/spar/context/fil_num_cons_a
sipContentTypeID	**issue_SIP**
sipTransferObject	
descriptorID	**Issue**
transferObjectID	b85f1350-8b50-11e3-909e-00144f80ca6b
sipTransferObjectGroup	
associatedDescriptorGroupTypeID	Master
transferObjectGroupInstanceName	**GRP.1**
sipDataObject	
associatedDescriptorDataID	Image
sipTransferObjectGroup	
associatedDescriptorGroupTypeID	Ocr
transferObjectGroupInstanceName	**GRP.2**
sipDataObject	
associatedDescriptorDataID	Text

a) The Producer source ID is always 'NUM' within the 'fil_num_cons_a' channel. It represents the digitization delivery chain.

b) The SIP ID is made of the Producer identifier, the production identifier, and a date (in order to distinguish updates or editions of the same object).

c) The Producer-Archive project ID is the info: URI of the 'fil_num_cons_a' channel.

d) The Transfer Object ID is the UUID of the ingest completion event.

The complete METS SIP manifest is given in annex subsection E5.

7 SOFTWARE TOOLS

7.1 INTRODUCTION

This section describes the two prototypes, developed as part of the CCSDS PAIS standard (reference [1]), that are being made available. They are not to be considered as operational tools, and in each case the software is delivered 'as is', without maintenance.

The Descriptors, data, SIPs, SIP constraints, are included with the CNES prototype.

7.2 CNES PROTOTYPE

The CNES prototype is available at the following address: https://logiciels.cnes.fr. It is a Java implementation that conforms to the PAIS Recommended Standard based on the PAIS XML schemas, and it uses other components, such as a Derby database for the management of the Data Objects received by the Archive. This software is governed by the CeCILL-C license under French law and the rules of distribution of free software. It uses the dataChecker library under the term of EPL license, and the jgraph library under the term of LGPL license.

The CNES implementation was developed to cover the following functions during the following main phases of a Producer-Archive Project:

- During the Formal Definition Phase:
  - MOT creation and visualization:
    - MOT structure design;
    - Descriptors instantiation and validation with XSD Models, or direct import of Descriptors received from the Producer;
    - MOT visualization (see figure 4-1, and MOT figures provided with each practical case in section 6), using an easy GUI;
  - SIP specification: SIP content definition, and sequencing constraints between SIPs;
- During the Transfer and Validation Phases (on the Archive side):
  - State of the delivered Transfer Objects in the MOT and detailed follow-up (the number of delivered objects can be seen on the same window, instead of the occurrence);
  - Visualization of the transfer using the same graphical interface;
  - Validations performed on the delivered SIPs (integrity, Data Objects occurrences, SIP contents, SIP sequencing constraints); they are applied on the XFDU manifest (not on the attached data).

It is supposed to be installed on the Archive side.

It is implemented in the Java programming language. It includes the open-source Xample (XML form) and JGraph (graphical design of the MOT).

7.3 ESA SIP BUILDER

The SIP Builder is available at the following address: https://github.com/gael-systems/ccsds-sip-builder. It is under LGPL v3 license, described in the README.txt file, available along with the tool.

The ESA prototype, also called 'ESA SIP Builder', is a Java command line software dedicated to the generation of SIPs according to the PAIS XFDU SIP Model specified in section 6 of the PAIS Recommended Standard (see reference [1]). The following diagram identifies the main PAIS elements that are critical for the ESA SIP Builder software.

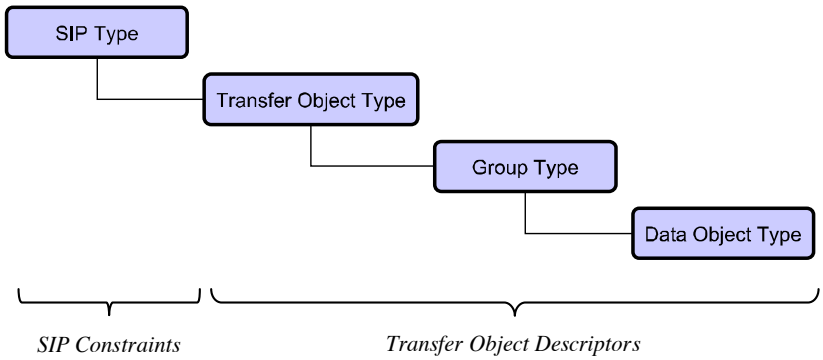

Figure 7-1: Identification of PAIS Elements Used by the ESA SIP Builder

On top of these standard PAIS definitions, the ESA SIP Builder introduces the concept of so-called Collectors responsible for the supply of the actual data files and folders to be conveyed by the output SIPs.

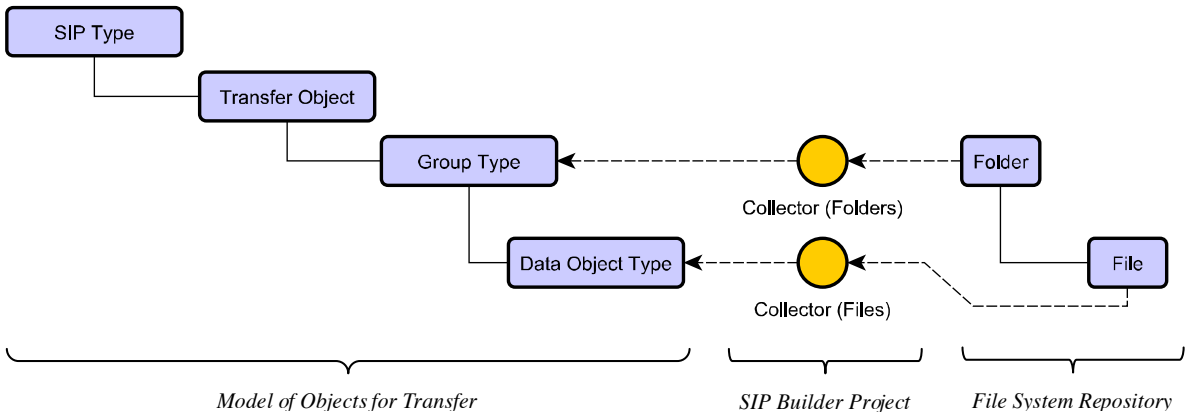

Figure 7-2: File and Folder Collectors

The ESA SIP Builder Collectors are attached to a target Group Type or a Data Object Type and hold one or more inclusion or exclusion patterns (file naming rules) that select files and folders in a given file system repository.

The ESA SIP Builder processes the input Descriptors and collectors in an internal in-memory tree of objects before any conversion from/to the disk as XFDU packages. For disambiguation with PAIS and XFDU elements or objects, the internal in-memory objects are further prefixed with the term 'Logical', e.g., Logical SIP composed of Logical Groups themselves composed of Logical Data Object.

The main steps of the ESA SIP Builder processing are the following:

- **Read and validate** the Collection Descriptors and Transfer Object Descriptors
- **Attach Collectors** to the Descriptors Group Types and Data Objects Types
- **Select the SIP Types** to be produced from the SIP Constraints
- **Sort the SIP Types** from the SIP Constraints, if any
- For each SIP Type:
  - For each authorized Descriptor of the SIP Type:
    - **For each Group Type** and sub-Group Type of the Descriptor:
      - Run the attached Collector, if any
      - For each collected folder or once if none:
        - For each Data Object Type:
          - **Run the attached Collector** considering the current Group Type instance name as context location
          - For each collected file:
            - **Create a Logical Data Object** of the current Data Object Type
      - **Create as many Logical Groups as necessary** to hold the created Logical Data Objects or intermediate Logical Groups
    - Create as many Logical Transfer Objects as necessary to hold the created Logical Groups
  - Create as many Logical SIPs as necessary to hold the created Logical Transfer Objects
- **Validate the Logical SIPs** against Descriptor constraints and limits, e.g., occurrences, sizes, etc.
- **Convert the Logical SIPs** to definitive **XFDU** packages

Special attention has to be paid to the three '**foreach**' used in the above processing tree. The general strategy is a top-down analysis of the PAIS elements from the SIP Types to the Data Object Type leaves. However, the ESA SIP Builder Collectors may select more folders or files than allowed for a given Data Object Type or a Group Type; cf. `maxOccurrence` usage in PAIS Descriptors and constraints file. The ESA SIP Builder considers that extra instances of the parent Group Types, Transfer Object Types, or even SIP Types need to be created. The use of `maxOccurrences` is even the only means for building multiple instances of Group Types or Data Object Types in the current implementation. It is not said that this is the best strategy, but it seemed reasonable so far, limiting the complexity of configuration to simple patterns. Future improvements could be considered by implementing less 'passive' Collectors capable of, for example, expressing grouping rules or more complex mapping.

The following practical example is probably more efficient than any further theory. It derives from a test case elaborated from data acquired by the ISEE first two spacecraft.

The input data set of about 2 GB has the following structure with two main directories isee1 and isee2 that share the same layout. The data are broken down in individual directories matching the year of acquisition. Each data file (`.asc-gz`) corresponds to a day of acquisition and has an associated metadata file (`.asc-gz_att`) in an `attrib` sub directory.

```
.
├── isee1
│   └── mag
│       └── 60_sec_ascii_gz
│           ├── 1977
│           │   ├── attrib
│           │   │   ├── isee1_mag_60s_0001_1977_295.asc-gz_att
│           │   │   ├── isee1_mag_60s_0002_1977_298.asc-gz_att
│           │   │   ├── isee1_mag_60s_0003_1977_300.asc-gz_att
│           │   │   ├── isee1_mag_60s_0004_1977_302.asc-gz_att
│           │   │   └── ...
│           │   ├── isee1_mag_60s_0001_1977_295.asc-gz
│           │   ├── isee1_mag_60s_0002_1977_298.asc-gz
│           │   ├── isee1_mag_60s_0003_1977_300.asc-gz
│           │   ├── isee1_mag_60s_0004_1977_302.asc-gz
│           │   └── ...
│           ├── 1978
│           │   └── ...
│           ├── ...
│           └── 1987
│               └── ...
└── isee2
    └── mag
        └── 60_sec_ascii_gz
            ├── 1979
            │   ├── attrib
            │   │   ├── isee2_mag_60s_0001_1977_295.asc-gz_att
            │   │   └── ...
            │   ├── isee2_mag_60s_0001_1977_295.asc-gz
            │   └── ...
            ├── ...
            └── 1987
                └── ...
```

In order to simulate an archiving project of this data set, a series of PAIS Descriptors and SIP constraints files were generated. The detail about these files is out of the scope of the present document, but the main elements relevant for the present illustration are summarized in figure 7-3.

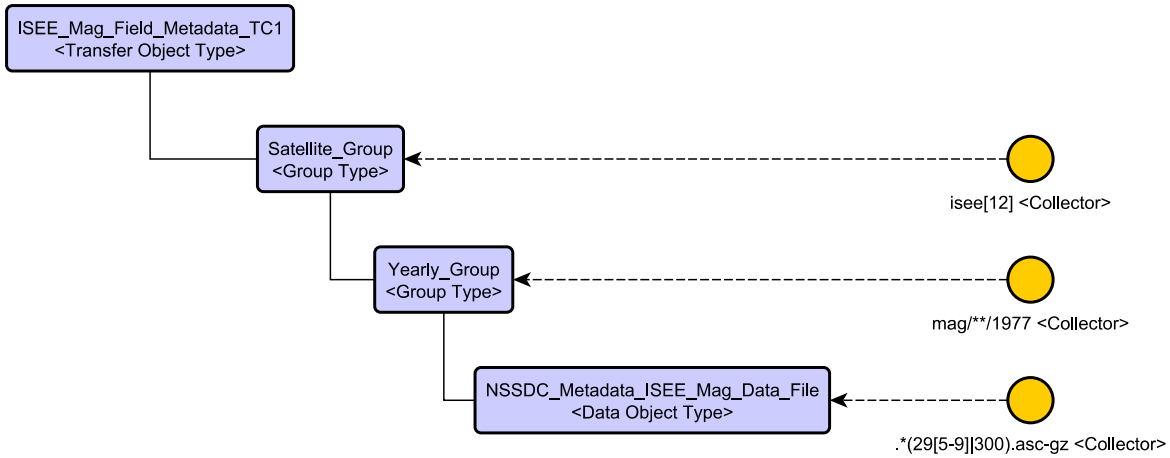

Figure 7-3: Example of Collectors Supplying TC1 Groups and Data Objects

The Descriptors define, among others, a Transfer Object Type identified by `ISEE_..._Metadata_TC1`. This latter is described as composed of a `Satellite_Group` Group Type, itself composed of a `Yearly_Group` Group Type that accepts a `NSSDC_Metadata_ISEE_Mag_Data_File` Data Object Type.

This tree of types matches quite well the layout of the data set described earlier where the `Satelite_Group` would correspond to the `isee1` and `isee2` directories, the `Yearly_Group` to the `1977`, `1978`, ... and `1987` directories, and finally the `NSSDC_..._Mag_Data_File` would represent the metadata files in the `attrib` subdirectories. Three collectors are necessary to express this mapping to the ESA SIP Builder. They are represented in yellow in the above diagram.

Different patterns could have produced the same results. For example, the pattern '`isee[12]`' could have been replaced by '`isee`' (any trailing character in regular expression) because only `isee1` and `isee2` are present in the input data set. Similarly, the '`mag`' prefix for the `Yearly_Group` filters nothing in that there is no other directory than the `mag` one for any spacecraft.

More specifically, some collecting patterns include some constraints that limit the file selection to a subset of what is actually available in the input data set. For example, the collector attached to the `Yearly_Group` type selects only the files under the `1977` directory. Similarly, the "`.*(29[5-9]|300).asc-gz_att`" of the Data Object Type collector filters only those files acquired between the 295th and the 300th days of the year, inclusive. These constraints are derived from the Descriptors documentation, e.g., a `transferObjectTypeDescription` element contained "Selected ISEE 1,2 `Magnetic_Field` Metadata grouped by Spacecraft and then by Yearly Directories (1977 only) for days 295–300 inclusive" and thus required specific collector patterns for building SIPs with the right content.

The ESA SIP Builder interpretation of the input Descriptors and collection patterns for this example can be summarized by figure 7-4, where a Transfer Object, on the left, will be generated with two sub-groups, one for each satellite, each containing a single year group containing three files. The tree on the right corresponds to the input data set.

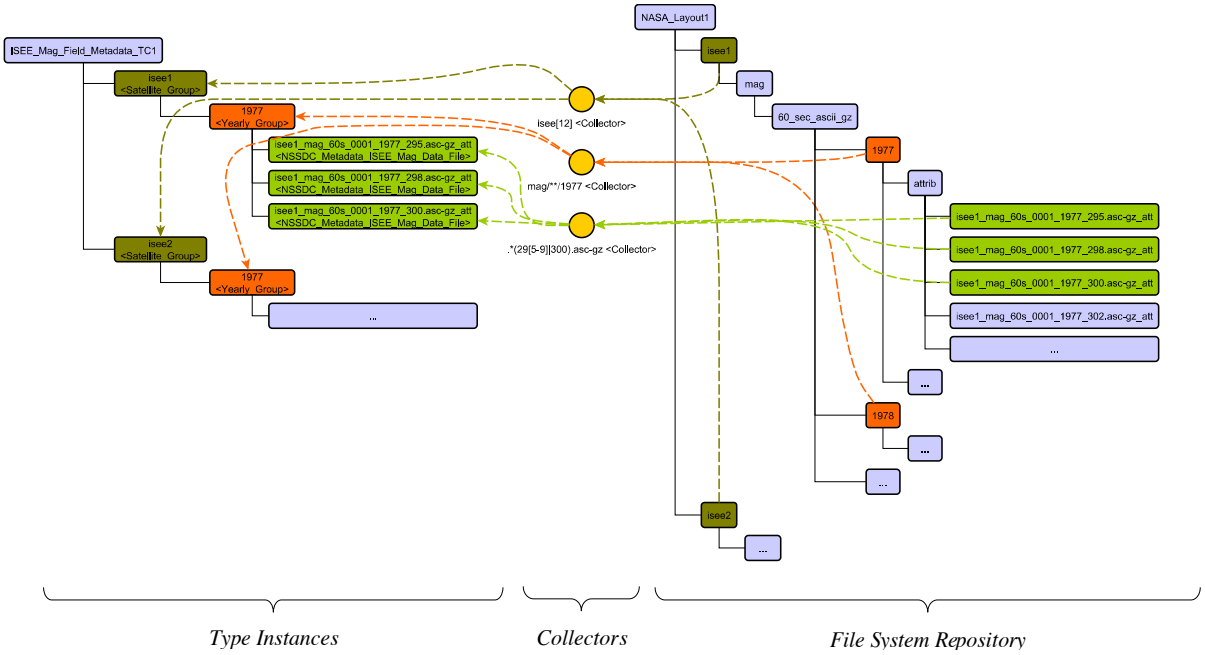

Figure 7-4: Collectors Mapping from Test Data to SIP Groups and Data Objects

The configuration of the ESA SIP Builder is provided through a so-called project file, which is an XML document.

ANNEX A

ASSOCIATED DESCRIPTOR DATA IDENTIFIERS

A1 INTRODUCTION

Under 5.2.4.2 (Data Object Identification container), one of the required attributes is Associated Descriptor Data ID. This attribute is used to identify that part of a Descriptor that is to be associated with a given Data Object. It takes on three different types of identification values depending on several factors as follows:

- Type A: If this Data Object is an instance of a Data Object Type defined in a Descriptor, then its Associated Descriptor Data ID should be the value of the Data Object Type ID of that Data Object Type specification.
- Type B: If this Data Object is an instance of a Transfer Object Group Type defined in the Descriptor to be encoded, and thus the group becomes a single Data Object, then its Associated Descriptor Data ID should be the value of the Transfer Object Group Type ID of that Transfer Object Group Type specification. This will be the case even when the Group Type is also defined to be undescribed.
- Type C: If this Data Object is an instance that is transferred within the context of a Descriptor-defined Transfer Object Group Type whose Transfer Object Group Type Structure Name has the value `undescribed`, then its Associated Descriptor Data ID should be the value of the Transfer Object Group Type ID of that Transfer Object Group Type specification.

This annex provides examples of these three types of Associated Descriptor Data ID values. The first example is for Type A defined above.

A2 TYPE A

The data structure for the Type A example is taken from the ISEE use case of 6.2, and its data layout is reproduced here as figure A-1.

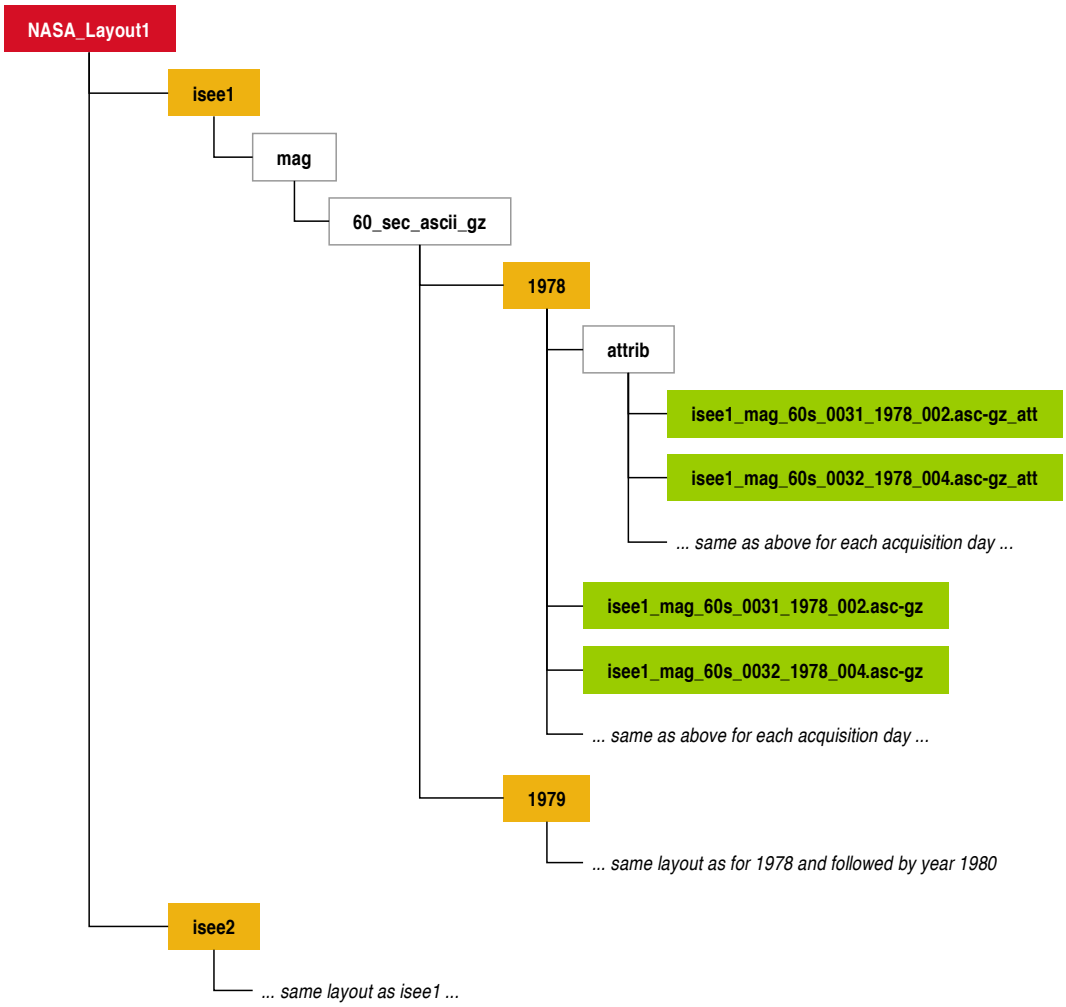

Figure A-1: ISEE1/ISEE 2 Data Repository Layout

For the purpose of highlighting the use of the Type A Associated Descriptor Data ID in the abstract SIP and in a resulting XFDU SIP, a simple Transfer Object Descriptor is sufficient. It is assumed that the Producer will be submitting two ISEE 1 files from 1978 to an Archive. A possible resulting Transfer Object Descriptor, in table form, is specified as follows:

Table A-1: Example of a Transfer Object Type Descriptor for Single File

Element	Sample Value
transferObjectTypeDescriptor ❶	
identification	
descriptorModelID	CCSD0014
descriptorModelVersion	V1.0
descriptorID	ISEE_1978_mag_data
description	
transferObjectTypeTitle	ISEE 1978 Demo Example 1
transferObjectTypeDescription	Demonstrate Associated Descriptor Data ID for Type A situation
transferObjectTypeOccurrence	
minOccurrence	1
maxOccurrence	2
relation	
parentCollection	Demo 1 Parent Collection
groupType	
groupTypeID	Yearly_Group
groupTypeStructureName	directory
groupTypeOccurrence	
minOccurrence	1
maxOccurrence	1
dataObjectType	
dataObjectTypeID	ISEE_1978_Mag_60s
dataObjectTypeOccurrence	
minOccurrence	1
maxOccurrence	1

This specifies that there can be at most two Transfer Objects, and each consists of a directory with a single file (see `dataObjectTypeOccurrence` above).

Assuming that the SIP Constraints specify that only one Transfer Object of the above type may be included in a SIP, a possible resulting abstract SIP is as follows:

Table A-2: Example of an Abstract SIP for Table A-1 Descriptor

Containers and Attributes	Example Attribute Values
SIP container❶	
SIP Global Information container	
SIP ID	SIP-Demo-1_01
Producer Source ID	Demo_1_Producer
Producer-Archive Project ID	Demo_1_Project
SIP Content Type ID	SIP_Demo_1_Single
Transfer Object Container	
Transfer Object Identification and Status container	
Descriptor ID	**ISEE_1978_mag_data**
Transfer Object ID	Demo_1_TO_01
Transfer Object Group container	
Transfer Object Group Identification container	
Associated Descriptor Group Type ID	**Yearly_Group**
Transfer Object Group Instance Name	1978
Data Object container	
Data Object Identification container	
Associated Descriptor Data ID	**ISEE_1978_Mag_60s**
Byte Stream container	
Byte Stream	PGTesv^&895…etc.

The attribute values in a bold font are those that reference values from the associated Descriptor. The SIP begins with a container for the SIP Global Information and follows with containers for Transfer Objects of which there is only one, because in this scenario it has been assumed that the SIP constraints require only one Transfer Object per SIP. It references the associated Descriptor through the Descriptor ID with a value of **ISEE_1978_mag_data**. There needs to be at least one Transfer Object Group container, and the Descriptor specifies it to be a directory. The abstract SIP references this Group container through the Associated Group Type ID value of Yearly_Group. The Transfer Group Instance Name is given as 1978, and this is one of the directory names shown in figure A-1. There is one Data Object container as there is only one file being transferred. It references the relevant Descriptor information by the attribute Associated Descriptor Data ID with a value of **ISEE_1978_Mag_60s**.

This abstract SIP can be implemented using the PAIS specified XFDU packaging mechanism. Most of the containers are implemented as XFDU content units using its

`extension` capability. The exception is the Byte Stream container because the XFDU has its own Byte Stream mechanism. The resulting SIP will not be further detailed as it is not relevant to the objective of this annex. However, 6.2 and its relevant annex contain a more complete example of ISEE Descriptors, SIP Constraints, and an XFDU SIP.

A3 TYPE B

The Type B case occurs when a Descriptor has a group type that is encoded. The example Descriptor is given in table A-3 below. It is identical to the Descriptor in table A-1 except many elements have been removed for brevity, and the element `groupTypeEncoded` has been inserted. It states that the group will be encoded into a single tar file. Thus the file will appear as a Data Object in the SIP.

Table A-3: Example of a Transfer Object Type Descriptor for Encoded Group

Element	Sample Value
`transferObjectTypeDescriptor` ❶	
`…removed for brevity…`	
`groupType`	
`groupTypeID`	`Yearly_Group`
`groupTypeStructureName`	`directory`
`groupTypeEncoded`	
`encodingName`	`tar`
`encodingDescription`	`application/x-tar`
`groupTypeOccurrence`	
`minOccurrence`	`1`
`maxOccurrence`	`1`
`dataObjectType`	
`dataObjectTypeID`	`ISEE_1978_Mag_60s`
`dataObjectTypeOccurrence`	
`minOccurrence`	`1`
`maxOccurrence`	`1`

The resulting abstract SIP would look as follows:

Table A-4: Example of an Abstract SIP for Table A-3 Descriptor

Containers and Attributes	Example Attribute Values
SIP container❶	
SIP Global Information container	
…removed for brevity	
Transfer Object Container	
Transfer Object Identification and Status container	
Descriptor ID	**ISEE_1978_mag_data**
Transfer Object ID	Demo_1_TO_01
Data Object container	
Data Object Identification container	
Associated Descriptor Data ID	**Yearly_Group**
Byte Stream container	
Byte Stream	Rtvexig*345… etc.

The abstract SIP's Transfer Object container no longer has a group container as this has been replaced by a Data Object container. The section of the Descriptor that is referenced by its Associated Descriptor Data ID is `Yearly_Group`, which the specification for the Transfer Object Group Type. This links the tar file with its specification in the Descriptor. The associated byte stream of the Data Object container will be the tar file.

A4 TYPE C

The Type C case occurs when a Descriptor has a group type that is `undescribed`. The example Descriptor is given in table A-5 below.

Table A-5: Example of a Transfer Object Type Descriptor for `undescribed` Group

Element	Sample Value
transferObjectTypeDescriptor ❶	
…removed for brevity…	
groupType	
groupTypeID	Yearly_Group
groupTypeStructureName	undescribed

It is identical to the Descriptor in table A-1 up to the Transfer Object Group Type, except many elements have been removed for brevity. The `groupTypeStructureName` is given as `undescribed`. This means that the Group Type and all data under that Group Type,

regardless of complexity, have not been modeled by this Descriptor. Other elements, except for `dataObjectType` and `groupType`, may be in the undescribed `groupType`. Nevertheless, the Producer still needs to package directories as groups and data files as Data Objects. For this example, the Producer needs to package the directory '1978' as a Transfer Object Group container and a single file as a Data Object container. The resulting abstract SIP would look as follows:

Table A-6: Example of an Abstract SIP for Table A-5 Descriptor

Containers and Attributes	Example Attribute Values
SIP container❶	
SIP Global Information container	
… removed for brevity…	
Transfer Object Container	
Transfer Object Identification and Status container	
Descriptor ID	**ISEE_1978_mag_data**
Transfer Object ID	Demo_1_TO_01
Transfer Object Group container	
Transfer Object Group Identification container	
Associated Descriptor Group Type ID	**Yearly_Group**
Transfer Object Group Instance Name	1978
Data Object container	
Data Object Identification container	
Associated Descriptor Data ID	**Yearly_Group**
Byte Stream container	
Byte Stream	PGTesv^&895…etc.

This abstract SIP looks like the abstract SIP of table A-2 with the exception that the Associated Descriptor Data ID points to the same section of the Descriptor as does the Associated Descriptor Group Type ID. All that is known, from this modeling information, is that the Data Object container is associated with the `Yearly_Group` part of the Descriptor. When a Descriptor employs a `groupTypeStructureName` of `undescribed`, it is good practice for the details of the `undescribed` data to be provided to the Archive in other documentation.

ANNEX B

ISEE USE CASE

B1 ISEE COLLECTION DESCRIPTOR

```
<?xml version="1.0" encoding="UTF-8"?>
<collectionDescriptor xmlns:x0="http://www.w3.org/2001/XMLSchema">
      <identification>
            <descriptorModelID>CCSD0015</descriptorModelID>
            <descriptorModelVersion>V1.0</descriptorModelVersion>

      <descriptorID>NASA_ESA_CNES_Test_Data_Exchange_02</descriptorID>
      </identification>
      <description>
            <collectionTitle>Selection 2 of ISEE 1,2 Magnetic_Field Data
and Metadata</collectionTitle>
            <collectionDescription>Selected ISEE 1,2 Magnetic_Field Data
and Metadata grouped by Spacecraft and then by Yearly Directories (1978
through 1980 only) for days 001-007 inclusive.   </collectionDescription>
            <collectionSize>
                  <minSize>9</minSize>
                  <maxSize>22</maxSize>
                  <unitsType>MB</unitsType>
            </collectionSize>
      </description>
      <relation>
            <parentCollection>none</parentCollection>
            <association>
                  <targetID>ISEE_Mag_Data_TC2</targetID>
                  <relationDescription>
                        <relationType>Contains</relationType>
                        <relationTextualDescription>Collection contains
data Transfer Objects of this type</relationTextualDescription>
                  </relationDescription>
            </association>
            <association>

            <targetID>NSSDC_Attributes_ISEE_Mag_Data_TC2</targetID>
                        <relationDescription>
                              <relationType>Contains</relationType>
                              <relationTextualDescription>Collection
      contains metadata Transfer Objects of this
      type</relationTextualDescription>
                        </relationDescription>
                  </association>
            </relation>
            </collectionDescriptor>
```

B2 ISEE DATA DESCRIPTOR

```
<?xml version="1.0" encoding="UTF-8"?>
<transferObjectTypeDescriptor xmlns:x0="http://www.w3.org/2001/XMLSchema">
      <identification>
            <descriptorModelID>CCSD0014</descriptorModelID>
            <descriptorModelVersion>V1.0</descriptorModelVersion>
            <descriptorID>ISEE_Mag_Data_TC2</descriptorID>
            <producerSourceID>NASA_ESA_Source1</producerSourceID>
      </identification>
      <description>
            <transferObjectTypeTitle>Annual Directory of ISEE 1,2
Magnetic_Field Data</transferObjectTypeTitle>
            <transferObjectTypeDescription>Annual Directory of ISEE 1,2
magnetic field data (no metadata) grouped by Spacecraft (ISEE 1 and ISEE 2)
and then for a Yearly Directory in range 1978 through 1980 for days 001-007
inclusive.  </transferObjectTypeDescription>
            <transferObjectTypeOccurrence>
                  <minOccurrence>3</minOccurrence>
                  <maxOccurrence>3</maxOccurrence>
            </transferObjectTypeOccurrence>
            <transferObjectTypeSize>
                  <minSize>3</minSize>
                  <maxSize>7</maxSize>
                  <unitsType>MB</unitsType>
            </transferObjectTypeSize>
            <namePreservationRule>Use the Source
names</namePreservationRule>
      </description>
      <relation>

      <parentCollection>NASA_ESA_CNES_Test_Data_Exchange_02</parentCollecti
on>
            <association>
                  <targetID>NSSDC_Attributes_ISEE_Mag_Data_TC2</targetID>
                  <relationDescription>
                        <relationType>Data</relationType>
                        <relationTextualDescription>group of data files
corresponding to the Target Id's group of metadata
files</relationTextualDescription>
                  </relationDescription>
            </association>
      </relation>
      <groupType>
            <groupTypeID>Satellite_Group</groupTypeID>
            <groupTypeDescription>There are 2 satellite groups, ISEE1 and
ISEE 2</groupTypeDescription>
```

```
            <groupTypeStructureName>directory</groupTypeStructureName>
            <groupTypeOccurrence>
                  <minOccurrence>2</minOccurrence>
                  <maxOccurrence>2</maxOccurrence>
            </groupTypeOccurrence>
            <groupType>
                  <groupTypeID>Yearly_Group</groupTypeID>
                  <groupTypeDescription>Each group will contain 1 year
taken from range 1978-1980</groupTypeDescription>

      <groupTypeStructureName>directory</groupTypeStructureName>
                  <groupTypeOccurrence>
                        <minOccurrence>1</minOccurrence>
                        <maxOccurrence>1</maxOccurrence>
                  </groupTypeOccurrence>
                  <dataObjectType>

      <dataObjectTypeID>ISEE_Mag_Data_File</dataObjectTypeID>
                        <dataObjectTypeDescription>ISEE magnetometer data
file whose file name contains a day identifier in the range 001-
007.</dataObjectTypeDescription>
                        <dataObjectTypeOccurrence>
                              <minOccurrence>2</minOccurrence>
                              <maxOccurrence>4</maxOccurrence>
                        </dataObjectTypeOccurrence>
                        <dataObjectTypeFormat>
                              <mimeType>text/plain</mimeType>
                        </dataObjectTypeFormat>
                        <dataObjectTypeEncoded>
                              <encodingName>gzip</encodingName>
                              <encodingDescription>application/x-
gzip</encodingDescription>
                        </dataObjectTypeEncoded>
                        <dataObjectTypeAssociation>

      <targetID>NSSDC_Attributes_ISEE_Mag_Data_File</targetID>
                              <relationDescription>
                                    <relationType>Data</relationType>
                                    <relationTextualDescription>Data file
corresponding to the Target ID's metadata file</relationTextualDescription>
                              </relationDescription>
                        </dataObjectTypeAssociation>
                  </dataObjectType>
            </groupType>
      </groupType>
</transferObjectTypeDescriptor>
```

B3 ISEE METADATA DESCRIPTOR

```
<?xml version="1.0" encoding="UTF-8"?>
<transferObjectTypeDescriptor xmlns:x0="http://www.w3.org/2001/XMLSchema">
      <identification>
            <descriptorModelID>CCSD0014</descriptorModelID>
            <descriptorModelVersion>V1.0</descriptorModelVersion>
            <descriptorID>NSSDC_Attributes_ISEE_Mag_Data_TC2</descriptorID>
            <producerSourceID>NASA_ESA_Source1</producerSourceID>
      </identification>
      <description>
            <transferObjectTypeTitle>Annual Directory of NSSDC Attributes
for ISEE 1,2 Magnetic_Field Data</transferObjectTypeTitle>
            <transferObjectTypeDescription>Annual Directory of NSSDC
Attributes for ISEE 1,2 magnetic field data grouped by Spacecraft (ISEE 1
and ISEE 2) and then for a Yearly Directory (in range 1978 through 1980)
for days 001-007 inclusive.  </transferObjectTypeDescription>
            <transferObjectTypeOccurrence>
                  <minOccurrence>3</minOccurrence>
                  <maxOccurrence>3</maxOccurrence>
            </transferObjectTypeOccurrence>
            <transferObjectTypeSize>
                  <minSize>8</minSize>
                  <maxSize>24</maxSize>
                  <unitsType>KB</unitsType>
            </transferObjectTypeSize>
            <namePreservationRule>Use the Source
names</namePreservationRule>
      </description>
      <relation>

      <parentCollection>NASA_ESA_CNES_Test_Data_Exchange_02</parentCollecti
on>
            <association>
                  <targetID>ISEE_Mag_Data_TC2</targetID>
                  <relationDescription>
                        <relationType>Metadata</relationType>
                        <relationTextualDescription>group of metadata files
corresponding to the Target Id's group of data
files</relationTextualDescription>
                  </relationDescription>
            </association>
      </relation>
      <groupType>
            <groupTypeID>Satellite_Group</groupTypeID>
            <groupTypeDescription>There are 2 satellite groups, ISEE1 and
ISEE 2</groupTypeDescription>
```

```
            <groupTypeStructureName>directory</groupTypeStructureName>
            <groupTypeOccurrence>
                  <minOccurrence>2</minOccurrence>
                  <maxOccurrence>2</maxOccurrence>
            </groupTypeOccurrence>
            <groupType>
                  <groupTypeID>Yearly_Group</groupTypeID>
                  <groupTypeDescription>Each group will contain 1 year
taken from range 1978 - 1980</groupTypeDescription>

      <groupTypeStructureName>directory</groupTypeStructureName>
                  <groupTypeOccurrence>
                        <minOccurrence>1</minOccurrence>
                        <maxOccurrence>1</maxOccurrence>
                  </groupTypeOccurrence>
                  <dataObjectType>

      <dataObjectTypeID>NSSDC_Attributes_ISEE_Mag_Data_File</dataObjectType
ID>
                        <dataObjectTypeDescription>NSSDC generated metadata
file, corresponding to ISEE magnetometer data file, whose file name
contains a day identifier in the range 001-007.</dataObjectTypeDescription>
                        <dataObjectTypeOccurrence>
                              <minOccurrence>2</minOccurrence>
                              <maxOccurrence>4</maxOccurrence>
                        </dataObjectTypeOccurrence>
                        <dataObjectTypeFormat>
                              <mimeType>text/plain</mimeType>
                        </dataObjectTypeFormat>
                        <dataObjectTypeAssociation>
                              <targetID>ISEE_Mag_Data_File</targetID>
                              <relationDescription>
                                    <relationType>Metadata</relationType>
                                    <relationTextualDescription>Metadata
files corresponding to the Target ID's data
files</relationTextualDescription>
                              </relationDescription>
                        </dataObjectTypeAssociation>
                  </dataObjectType>
            </groupType>
      </groupType>
</transferObjectTypeDescriptor>
```

B4 ISEE SIP CONSTRAINTS

```
<?xml version="1.0" encoding="UTF-8"?>
<sipConstraints xmlns:x0="http://www.w3.org/2001/XMLSchema"
xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
xsi:noNamespaceSchemaLocation="D:\PUBLIC\PAIS_DATA\projectsV2.7\NASA1\conf\
sip_constraints_RB1.xsd">
      <producerArchiveProjectID>NASA_ESA_CNES_Test_Data_Exchange_02</produc
erArchiveProjectID>
      <sipContentTypes>
            <sipContentTypeID>SIP_01</sipContentTypeID>
            <authorizedDescriptor>
                  <descriptorID>ISEE_Mag_Data_TC2</descriptorID>
                  <occurrence>
                        <minOccurrence>1</minOccurrence>
                        <maxOccurrence>3</maxOccurrence>
                  </occurrence>
            </authorizedDescriptor>
            <sipContentTypeID>SIP_02</sipContentTypeID>
            <authorizedDescriptor>

      <descriptorID>NSSDC_Attributes_ISEE_Mag_Data_TC2</descriptorID>
                  <occurrence>
                        <minOccurrence>1</minOccurrence>
                        <maxOccurrence>3</maxOccurrence>
                  </occurrence>
            </authorizedDescriptor>
      </sipContentTypes>
      <sipSequencingConstraintGroup>
            <groupName>Normal Group-A</groupName>
            <constraintItem>
                  <sipContentTypeID>SIP_02</sipContentTypeID>
                  <constraintSerialNumber>1</constraintSerialNumber>
            </constraintItem>
            <constraintItem>
                  <sipContentTypeID>SIP_01</sipContentTypeID>
                  <constraintSerialNumber>2</constraintSerialNumber>
            </constraintItem>
      </sipSequencingConstraintGroup>
</sipConstraints>
```

B5 MANIFEST OF THE ISEE DATA SIP

```
?xml version="1.0" encoding="UTF-8"?>

<xfdu:XFDU xmlns:xfdu="urn:ccsds:schema:xfdu:1"
xmlns:sip="urn:ccsds:schema:pais:1">
   <packageHeader ID="NASA_ESA_CNES_Test_Data_Exchange_02-SIP-0002">
      <volumeInfo>
         <specificationVersion>1.0</specificationVersion>
      </volumeInfo>
      <environmentInfo>
         <extension>
            <sip:sipGlobalInformation>
               <sip:sipID>NASA_ESA_CNES_Test_Data_Exchange_02-SIP-
0002</sip:sipID>

<sip:producerSourceID>NASA_ESA_Source1</sip:producerSourceID>

<sip:producerArchiveProjectID>NASA_ESA_CNES_Test_Data_Exchange_02</sip:prod
ucerArchiveProjectID>
               <sip:sipContentTypeID>SIP_01</sip:sipContentTypeID>
               <sip:sipSequenceNumber>2</sip:sipSequenceNumber>
            </sip:sipGlobalInformation>
         </extension>
      </environmentInfo>
   </packageHeader>
   <informationPackageMap>
      <xfdu:contentUnit>
         <extension>
            <sip:sipTransferObject>
               <sip:descriptorID>ISEE_Mag_Data_TC2</sip:descriptorID>
               <sip:transferObjectID>ISEE_Mag_Data_TC2-
0001</sip:transferObjectID>
            </sip:sipTransferObject>
         </extension>
         <xfdu:contentUnit>
            <extension>
               <sip:sipTransferObjectGroup>

<sip:associatedDescriptorGroupTypeID>Satellite_Group</sip:associatedDescrip
torGroupTypeID>

<sip:transferObjectGroupInstanceName>isee1</sip:transferObjectGroupInstance
Name>
               </sip:sipTransferObjectGroup>
            </extension>
            <xfdu:contentUnit>
               <extension>
                  <sip:sipTransferObjectGroup>

<sip:associatedDescriptorGroupTypeID>Yearly_Group</sip:associatedDescriptor
GroupTypeID>
```

```
<sip:transferObjectGroupInstanceName>1978</sip:transferObjectGroupInstanceN
ame>
                  </sip:sipTransferObjectGroup>
               </extension>
               <xfdu:contentUnit>
                  <extension>
                     <sip:sipDataObject>

<sip:associatedDescriptorDataID>ISEE_Mag_Data_File</sip:associatedDescripto
rDataID>
                     </sip:sipDataObject>
                  </extension>
                  <dataObjectPointer dataObjectID="DO-ISEE_Mag_Data_File-
0001"/>
               </xfdu:contentUnit>
               <xfdu:contentUnit>
                  <extension>
                     <sip:sipDataObject>

<sip:associatedDescriptorDataID>ISEE_Mag_Data_File</sip:associatedDescripto
rDataID>
                     </sip:sipDataObject>
                  </extension>
                  <dataObjectPointer dataObjectID="DO-ISEE_Mag_Data_File-
0002"/>
               </xfdu:contentUnit>
               <xfdu:contentUnit>
                  <extension>
                     <sip:sipDataObject>

<sip:associatedDescriptorDataID>ISEE_Mag_Data_File</sip:associatedDescripto
rDataID>
                     </sip:sipDataObject>
                  </extension>
                  <dataObjectPointer dataObjectID="DO-ISEE_Mag_Data_File-
0003"/>
               </xfdu:contentUnit>
            </xfdu:contentUnit>
         </xfdu:contentUnit>
         <xfdu:contentUnit>
            <extension>
               <sip:sipTransferObjectGroup>

<sip:associatedDescriptorGroupTypeID>Satellite_Group</sip:associatedDescrip
torGroupTypeID>

<sip:transferObjectGroupInstanceName>isee2</sip:transferObjectGroupInstance
Name>
               </sip:sipTransferObjectGroup>
            </extension>
            <xfdu:contentUnit>
               <extension>
```

```
                  <sip:sipTransferObjectGroup>

<sip:associatedDescriptorGroupTypeID>Yearly_Group</sip:associatedDescriptor
GroupTypeID>

<sip:transferObjectGroupInstanceName>1978</sip:transferObjectGroupInstanceN
ame>
                  </sip:sipTransferObjectGroup>
               </extension>
               <xfdu:contentUnit>
                  <extension>
                     <sip:sipDataObject>

<sip:associatedDescriptorDataID>ISEE_Mag_Data_File</sip:associatedDescripto
rDataID>
                     </sip:sipDataObject>
                  </extension>
                  <dataObjectPointer dataObjectID="DO-ISEE_Mag_Data_File-
0004"/>
               </xfdu:contentUnit>
               <xfdu:contentUnit>
                  <extension>
                     <sip:sipDataObject>

<sip:associatedDescriptorDataID>ISEE_Mag_Data_File</sip:associatedDescripto
rDataID>
                     </sip:sipDataObject>
                  </extension>
                  <dataObjectPointer dataObjectID="DO-ISEE_Mag_Data_File-
0005"/>
               </xfdu:contentUnit>
               <xfdu:contentUnit>
                  <extension>
                     <sip:sipDataObject>

<sip:associatedDescriptorDataID>ISEE_Mag_Data_File</sip:associatedDescripto
rDataID>
                     </sip:sipDataObject>
                  </extension>
                  <dataObjectPointer dataObjectID="DO-ISEE_Mag_Data_File-
0006"/>
               </xfdu:contentUnit>
            </xfdu:contentUnit>
         </xfdu:contentUnit>
      </xfdu:contentUnit>
      <xfdu:contentUnit>
         <extension>
            <sip:sipTransferObject>
               <sip:descriptorID>ISEE_Mag_Data_TC2</sip:descriptorID>
               <sip:transferObjectID>ISEE_Mag_Data_TC2-
0002</sip:transferObjectID>
            </sip:sipTransferObject>
         </extension>
```

```
         <xfdu:contentUnit>
            <extension>
               <sip:sipTransferObjectGroup>

<sip:associatedDescriptorGroupTypeID>Satellite_Group</sip:associatedDescrip
torGroupTypeID>

<sip:transferObjectGroupInstanceName>isee1</sip:transferObjectGroupInstance
Name>
               </sip:sipTransferObjectGroup>
            </extension>
            <xfdu:contentUnit>
               <extension>
                  <sip:sipTransferObjectGroup>

<sip:associatedDescriptorGroupTypeID>Yearly_Group</sip:associatedDescriptor
GroupTypeID>

<sip:transferObjectGroupInstanceName>1979</sip:transferObjectGroupInstanceN
ame>
                  </sip:sipTransferObjectGroup>
               </extension>
               <xfdu:contentUnit>
                  <extension>
                     <sip:sipDataObject>

<sip:associatedDescriptorDataID>ISEE_Mag_Data_File</sip:associatedDescripto
rDataID>
                     </sip:sipDataObject>
                  </extension>
                  <dataObjectPointer dataObjectID="DO-ISEE_Mag_Data_File-
0007"/>
               </xfdu:contentUnit>
               <xfdu:contentUnit>
                  <extension>
                     <sip:sipDataObject>

<sip:associatedDescriptorDataID>ISEE_Mag_Data_File</sip:associatedDescripto
rDataID>
                     </sip:sipDataObject>
                  </extension>
                  <dataObjectPointer dataObjectID="DO-ISEE_Mag_Data_File-
0008"/>
               </xfdu:contentUnit>
               <xfdu:contentUnit>
                  <extension>
                     <sip:sipDataObject>

<sip:associatedDescriptorDataID>ISEE_Mag_Data_File</sip:associatedDescripto
rDataID>
                     </sip:sipDataObject>
                  </extension>
```

```
                  <dataObjectPointer dataObjectID="DO-ISEE_Mag_Data_File-
0009"/>
               </xfdu:contentUnit>
            </xfdu:contentUnit>
         </xfdu:contentUnit>
         <xfdu:contentUnit>
            <extension>
               <sip:sipTransferObjectGroup>

<sip:associatedDescriptorGroupTypeID>Satellite_Group</sip:associatedDescrip
torGroupTypeID>

<sip:transferObjectGroupInstanceName>isee2</sip:transferObjectGroupInstance
Name>
               </sip:sipTransferObjectGroup>
            </extension>
            <xfdu:contentUnit>
               <extension>
                  <sip:sipTransferObjectGroup>

<sip:associatedDescriptorGroupTypeID>Yearly_Group</sip:associatedDescriptor
GroupTypeID>

<sip:transferObjectGroupInstanceName>1979</sip:transferObjectGroupInstanceN
ame>
                  </sip:sipTransferObjectGroup>
               </extension>
               <xfdu:contentUnit>
                  <extension>
                     <sip:sipDataObject>

<sip:associatedDescriptorDataID>ISEE_Mag_Data_File</sip:associatedDescripto
rDataID>
                     </sip:sipDataObject>
                  </extension>
                  <dataObjectPointer dataObjectID="DO-ISEE_Mag_Data_File-
0010"/>
               </xfdu:contentUnit>
               <xfdu:contentUnit>
                  <extension>
                     <sip:sipDataObject>

<sip:associatedDescriptorDataID>ISEE_Mag_Data_File</sip:associatedDescripto
rDataID>
                     </sip:sipDataObject>
                  </extension>
                  <dataObjectPointer dataObjectID="DO-ISEE_Mag_Data_File-
0011"/>
               </xfdu:contentUnit>
               <xfdu:contentUnit>
                  <extension>
                     <sip:sipDataObject>
```

```
<sip:associatedDescriptorDataID>ISEE_Mag_Data_File</sip:associatedDescriptorDataID>
                    </sip:sipDataObject>
                 </extension>
                 <dataObjectPointer dataObjectID="DO-ISEE_Mag_Data_File-0012"/>
              </xfdu:contentUnit>
           </xfdu:contentUnit>
        </xfdu:contentUnit>
     </xfdu:contentUnit>
     <xfdu:contentUnit>
        <extension>
           <sip:sipTransferObject>
              <sip:descriptorID>ISEE_Mag_Data_TC2</sip:descriptorID>
              <sip:transferObjectID>ISEE_Mag_Data_TC2-0003</sip:transferObjectID>
           </sip:sipTransferObject>
        </extension>
        <xfdu:contentUnit>
           <extension>
              <sip:sipTransferObjectGroup>

<sip:associatedDescriptorGroupTypeID>Satellite_Group</sip:associatedDescriptorGroupTypeID>

<sip:transferObjectGroupInstanceName>isee1</sip:transferObjectGroupInstanceName>
              </sip:sipTransferObjectGroup>
           </extension>
           <xfdu:contentUnit>
              <extension>
                 <sip:sipTransferObjectGroup>

<sip:associatedDescriptorGroupTypeID>Yearly_Group</sip:associatedDescriptorGroupTypeID>

<sip:transferObjectGroupInstanceName>1980</sip:transferObjectGroupInstanceName>
                 </sip:sipTransferObjectGroup>
              </extension>
              <xfdu:contentUnit>
                 <extension>
                    <sip:sipDataObject>

<sip:associatedDescriptorDataID>ISEE_Mag_Data_File</sip:associatedDescriptorDataID>
                    </sip:sipDataObject>
                 </extension>
                 <dataObjectPointer dataObjectID="DO-ISEE_Mag_Data_File-0013"/>
              </xfdu:contentUnit>
              <xfdu:contentUnit>
```

```
                  <extension>
                     <sip:sipDataObject>

<sip:associatedDescriptorDataID>ISEE_Mag_Data_File</sip:associatedDescripto
rDataID>
                     </sip:sipDataObject>
                  </extension>
                  <dataObjectPointer dataObjectID="DO-ISEE_Mag_Data_File-
0014"/>
               </xfdu:contentUnit>
               <xfdu:contentUnit>
                  <extension>
                     <sip:sipDataObject>

<sip:associatedDescriptorDataID>ISEE_Mag_Data_File</sip:associatedDescriptor
rDataID>
                     </sip:sipDataObject>
                  </extension>
                  <dataObjectPointer dataObjectID="DO-ISEE_Mag_Data_File-
0015"/>
               </xfdu:contentUnit>
            </xfdu:contentUnit>
         </xfdu:contentUnit>
         <xfdu:contentUnit>
            <extension>
               <sip:sipTransferObjectGroup>

<sip:associatedDescriptorGroupTypeID>Satellite_Group</sip:associatedDescrip
torGroupTypeID>

<sip:transferObjectGroupInstanceName>isee2</sip:transferObjectGroupInstance
Name>
               </sip:sipTransferObjectGroup>
            </extension>
            <xfdu:contentUnit>
               <extension>
                  <sip:sipTransferObjectGroup>

<sip:associatedDescriptorGroupTypeID>Yearly_Group</sip:associatedDescriptor
GroupTypeID>

<sip:transferObjectGroupInstanceName>1980</sip:transferObjectGroupInstanceN
ame>
                  </sip:sipTransferObjectGroup>
               </extension>
               <xfdu:contentUnit>
                  <extension>
                     <sip:sipDataObject>

<sip:associatedDescriptorDataID>ISEE_Mag_Data_File</sip:associatedDescripto
rDataID>
                     </sip:sipDataObject>
                  </extension>
```

```
                  <dataObjectPointer dataObjectID="DO-ISEE_Mag_Data_File-
0016"/>
               </xfdu:contentUnit>
               <xfdu:contentUnit>
                  <extension>
                     <sip:sipDataObject>

<sip:associatedDescriptorDataID>ISEE_Mag_Data_File</sip:associatedDescripto
rDataID>
                     </sip:sipDataObject>
                  </extension>
                  <dataObjectPointer dataObjectID="DO-ISEE_Mag_Data_File-
0017"/>
               </xfdu:contentUnit>
               <xfdu:contentUnit>
                  <extension>
                     <sip:sipDataObject>

<sip:associatedDescriptorDataID>ISEE_Mag_Data_File</sip:associatedDescripto
rDataID>
                     </sip:sipDataObject>
                  </extension>
                  <dataObjectPointer dataObjectID="DO-ISEE_Mag_Data_File-
0018"/>
               </xfdu:contentUnit>
            </xfdu:contentUnit>
         </xfdu:contentUnit>
      </xfdu:contentUnit>
   </informationPackageMap>
   <dataObjectSection>
      <dataObject ID="DO-ISEE_Mag_Data_File-0001" size="128">
         <byteStream size="128">
            <fileLocation locatorType="URL"
href="file:isee1/1978/isee1_mag_60s_0031_1978_002.asc-gz"/>
            <checksum
checksumName="MD5">7cc53dd29fb89105352e5f50f9af06b5</checksum>
         </byteStream>
      </dataObject>
      <dataObject ID="DO-ISEE_Mag_Data_File-0002" size="128">
         <byteStream size="128">
            <fileLocation locatorType="URL"
href="file:isee1/1978/isee1_mag_60s_0032_1978_004.asc-gz"/>
            <checksum
checksumName="MD5">8b1e38c7109f4b39ae5f0ec456ba1569</checksum>
         </byteStream>
      </dataObject>
      <dataObject ID="DO-ISEE_Mag_Data_File-0003" size="128">
         <byteStream size="128">
            <fileLocation locatorType="URL"
href="file:isee1/1978/isee1_mag_60s_0033_1978_007.asc-gz"/>
            <checksum
checksumName="MD5">92f7668852b9006f4091becd3b3e7ab7</checksum>
         </byteStream>
```

```
      </dataObject>
      <dataObject ID="DO-ISEE_Mag_Data_File-0004" size="128">
         <byteStream size="128">
            <fileLocation locatorType="URL"
href="file:isee2/1978/isee2_mag_60s_0031_1978_002.asc-gz"/>
            <checksum
checksumName="MD5">15e56b31c9c576f7ca50785f31bc8528</checksum>
         </byteStream>
      </dataObject>
      <dataObject ID="DO-ISEE_Mag_Data_File-0005" size="128">
         <byteStream size="128">
            <fileLocation locatorType="URL"
href="file:isee2/1978/isee2_mag_60s_0032_1978_004.asc-gz"/>
            <checksum
checksumName="MD5">0aa1312e75d78a68c98cf0063c6115a6</checksum>
         </byteStream>
      </dataObject>
      <dataObject ID="DO-ISEE_Mag_Data_File-0006" size="128">
         <byteStream size="128">
            <fileLocation locatorType="URL"
href="file:isee2/1978/isee2_mag_60s_0033_1978_007.asc-gz"/>
            <checksum
checksumName="MD5">06b5550d1907056737992c60530045be</checksum>
         </byteStream>
      </dataObject>
      <dataObject ID="DO-ISEE_Mag_Data_File-0007" size="128">
<byteStream size="128">
            <fileLocation locatorType="URL"
href="file:isee1/1979/isee1_mag_60s_0184_1979_002.asc-gz"/>
            <checksum
checksumName="MD5">cd4ba939abbab267def1888133a57a0f</checksum>
         </byteStream>
      </dataObject>
      <dataObject ID="DO-ISEE_Mag_Data_File-0008" size="128">
         <byteStream size="128">
            <fileLocation locatorType="URL"
href="file:isee1/1979/isee1_mag_60s_0185_1979_005.asc-gz"/>
            <checksum
checksumName="MD5">a4c03823a04e77e01f091e94e851b506</checksum>
         </byteStream>
      </dataObject>
      <dataObject ID="DO-ISEE_Mag_Data_File-0009" size="128">
         <byteStream size="128">
            <fileLocation locatorType="URL"
href="file:isee1/1979/isee1_mag_60s_0186_1979_007.asc-gz"/>
            <checksum
checksumName="MD5">92e12fc7928b0c43f71b3cdef70fff49</checksum>
         </byteStream>
      </dataObject>
      <dataObject ID="DO-ISEE_Mag_Data_File-0010" size="128">
         <byteStream size="128">
            <fileLocation locatorType="URL"
href="file:isee2/1979/isee2_mag_60s_0184_1979_002.asc-gz"/>
```

```
            <checksum
checksumName="MD5">d9eb55a41bb39f06b8def40a5bb4361c</checksum>
         </byteStream>
      </dataObject>
      <dataObject ID="DO-ISEE_Mag_Data_File-0011" size="128">
         <byteStream size="128">
            <fileLocation locatorType="URL"
href="file:isee2/1979/isee2_mag_60s_0185_1979_005.asc-gz"/>
            <checksum
checksumName="MD5">8000efd93d427604b6552df0dbc658b7</checksum>
         </byteStream>
      </dataObject>
      <dataObject ID="DO-ISEE_Mag_Data_File-0012" size="128">
         <byteStream size="128">
            <fileLocation locatorType="URL"
href="file:isee2/1979/isee2_mag_60s_0186_1979_007.asc-gz"/>
            <checksum
checksumName="MD5">d51f95e20c288d3ada95fb3017679d26</checksum>
         </byteStream>
      </dataObject>
      <dataObject ID="DO-ISEE_Mag_Data_File-0013" size="128">
         <byteStream size="128">
            <fileLocation locatorType="URL"
href="file:isee1/1980/isee1_mag_60s_0336_1980_001.asc-gz"/>
            <checksum
checksumName="MD5">b125c3e15a9299aadcd6b2ce2c1592d4</checksum>
         </byteStream>
      </dataObject>
      <dataObject ID="DO-ISEE_Mag_Data_File-0014" size="128">
         <byteStream size="128">
            <fileLocation locatorType="URL"
href="file:isee1/1980/isee1_mag_60s_0337_1980_003.asc-gz"/>
            <checksum
checksumName="MD5">bc19c66de561583eed74165a0cab0ca4</checksum>
         </byteStream>
      </dataObject>
      <dataObject ID="DO-ISEE_Mag_Data_File-0015" size="128">
         <byteStream size="128">
            <fileLocation locatorType="URL"
href="file:isee1/1980/isee1_mag_60s_0338_1980_006.asc-gz"/>
            <checksum
checksumName="MD5">333f8d57f730909697bfbd8b61f07dae</checksum>
         </byteStream>
      </dataObject>
      <dataObject ID="DO-ISEE_Mag_Data_File-0016" size="128">
         <byteStream size="128">
            <fileLocation locatorType="URL"
href="file:isee2/1980/isee2_mag_60s_0336_1980_001.asc-gz"/>
            <checksum
checksumName="MD5">0fb14b0bf6586b12f6306bad6d1cb52b</checksum>
         </byteStream>
      </dataObject>
      <dataObject ID="DO-ISEE_Mag_Data_File-0017" size="128">
```

```
         <byteStream size="128">
            <fileLocation locatorType="URL"
href="file:isee2/1980/isee2_mag_60s_0337_1980_003.asc-gz"/>
            <checksum
checksumName="MD5">43e388d77d6ad8d5d0ff87b1f4910ac6</checksum>
         </byteStream>
      </dataObject>
      <dataObject ID="DO-ISEE_Mag_Data_File-0018" size="128">
         <byteStream size="128">
            <fileLocation locatorType="URL"
href="file:isee2/1980/isee2_mag_60s_0338_1980_006.asc-gz"/>
            <checksum
checksumName="MD5">2d9a2b38ace885affb587ba8ab2c0dbe</checksum>
         </byteStream>
      </dataObject>
   </dataObjectSection>
</xfdu:XFDU>
```

B6 EXAMPLE OF SIP BUILDER SOFTWARE CONFIGURATION FILE

```
<?xml version="1.0" encoding="UTF-8"?>

<project xmlns="urn:fr:gael:schema:ccsds:pais:sip-builder:1">

   <descriptors>
      <descriptor file="ISEE Descriptor- data TC2.xml"/>
      <descriptor file="ISEE Descriptor- attributes TC2.xml"></descriptor>
      <descriptor file="ISEE Collection Descriptor TC2.xml"/>
      <descriptor file="ISEE SIP Constraints TC2.xml" ></descriptor>
   </descriptors>

   <collectors
      baseDirectory="../../test-data/nasa-isee12-tds-20130705-shrunken-
128b">
       <collector typeId="Satellite_Group">
           <include>isee[12]</include>
       </collector>
       <collector typeId="Yearly_Group">
           <include>mag/**/19(78|79|80)</include>
       </collector>
       <collector typeId="ISEE_Mag_Data_File">
           <include>.*00[1-7].asc-gz</include>
       </collector>
       <collector typeId="NSSDC_Attributes_ISEE_Mag_Data_File">
           <include>**/.*00[1-7].asc-gz_att</include>
       </collector>
   </collectors>

</project>
```

ANNEX C

ESA-SAFE USE CASE

C1 INTRODUCTION

This annex contains the PAIS XML Descriptors of the ESA-SAFE detailed and simple use cases.

C2 SIMPLE CASE

C2.1 ESA-SAFE—ESA_ERS_AMI_SAR COLLECTION DESCRIPTOR (ROOT)

```
<?xml version="1.0" encoding="UTF-8"?>
<collectionDescriptor xmlns="urn:ccsds:schema:pais:1">
  <identification>
    <descriptorModelID>CCSD0015</descriptorModelID>
    <descriptorModelVersion>1.0</descriptorModelVersion>
    <descriptorID>ESA_ERS_AMI_SAR</descriptorID>
  </identification>
  <description>
    <collectionTitle>Level 0 of ERS AMI SAR experiments</collectionTitle>
    <collectionDescription>LO data corresponding to Earth's oceans and land
images obtained with a suite of instruments, one of which is a SAR on the
AMI, onboard a ERS satellite</collectionDescription>
  </description>
  <relation>
    <parentCollection>NONE</parentCollection>
  </relation>
</collectionDescriptor>
```

C2.2 ESA-SAFE—COL_ERS_AMI_SAR-REP-INFO COLLECTION DESCRIPTOR

```
<?xml version="1.0" encoding="UTF-8"?>
<collectionDescriptor xmlns="urn:ccsds:schema:pais:1">
  <identification>
    <descriptorModelID>CCSD0015</descriptorModelID>
    <descriptorModelVersion>1.0</descriptorModelVersion>
    <descriptorID>COL_ERS-AMI-SAR-Rep-Info</descriptorID>
  </identification>
  <description>
    <collectionTitle>ERS-AMI-SAR Representation Information for L0 data
   </collectionTitle>
    <collectionDescription>A set of SAFE v2.0 packages providing all the
Representation Information necessary for L0 data products
</collectionDescription>
  </description>
  <relation>
    <parentCollection>ESA_ERS_AMI_SAR</parentCollection>
  </relation>
</collectionDescriptor>
```

C2.3 ESA-SAFE—COL_ERS_AMI_SAR-DOC COLLECTION DESCRIPTOR

```
<?xml version="1.0" encoding="UTF-8"?>
<collectionDescriptor xmlns="urn:ccsds:schema:pais:1">
  <identification>
    <descriptorModelID>CCSD0015</descriptorModelID>
    <descriptorModelVersion>1.0</descriptorModelVersion>
    <descriptorID>COL_ERS-AMI-SAR-DOC</descriptorID>
  </identification>
  <description>
    <collectionTitle>ERS-AMI-SAR associated documentation</collectionTitle>
    <collectionDescription>All documents necessary for the understanding of
the mission, instruments, products, and sips</collectionDescription>
  </description>
  <relation>
    <parentCollection>ESA_ERS_AMI_SAR</parentCollection>
  </relation>
</collectionDescriptor>
```

C2.4 ESA-SAFE—COL_ERS-SAR-EO-DATA COLLECTION DESCRIPTOR

```
<?xml version="1.0" encoding="UTF-8"?>
<collectionDescriptor xmlns="urn:ccsds:schema:pais:1">
  <identification>
    <descriptorModelID>CCSD0015</descriptorModelID>
    <descriptorModelVersion>1.0</descriptorModelVersion>
    <descriptorID>COL_ERS-AMI-SAR-EO-DATA</descriptorID>
  </identification>
  <description>
    <collectionTitle>EO Product of level 0 for ERS-AMI-SAR
   </collectionTitle>
    <collectionDescription>COL_ERS-AMI-SAR-LEVEL-0</collectionDescription>
  </description>
  <relation>
    <parentCollection>ESA_ERS_AMI_SAR</parentCollection>
  </relation>
</collectionDescriptor>
```

C2.5 ESA-SAFE—TRF_BASE_PACKAGE_REP_INFO TRANSFER DESCRIPTOR

```
<?xml version="1.0" encoding="UTF-8"?>
<transferObjectTypeDescriptor xmlns="urn:ccsds:schema:pais:1">
  <identification>
    <descriptorModelID>CCSD0014</descriptorModelID>
    <descriptorModelVersion>1.0</descriptorModelVersion>
    <descriptorID>TRF_BASE_PACKAGE_REP_INFO</descriptorID>
  </identification>
  <description>
    <transferObjectTypeTitle>Base Package Representation Information
</transferObjectTypeTitle>
<transferObjectTypeDescription>&lt;transferObjectTypeDescription&gt;Represe
```

```
ntation information to be present in any SAFE 2.0 compliant archive. It
should be ingested before any other packages.
</transferObjectTypeDescription>
    <transferObjectTypeOccurrence>
      <minOccurrence>1</minOccurrence>
      <maxOccurrence>1</maxOccurrence>
    </transferObjectTypeOccurrence>
  </description>
  <relation>
    <parentCollection>COL_ERS-AMI-SAR-Rep-Info</parentCollection>
  </relation>
  <groupType>
    <groupTypeID>GroupBaseRepInfoPackagePart</groupTypeID>
    <groupTypeStructureName>set</groupTypeStructureName>
    <groupTypeOccurrence>
      <minOccurrence>1</minOccurrence>
      <maxOccurrence>1</maxOccurrence>
    </groupTypeOccurrence>
    <dataObjectType>
      <dataObjectTypeID>GroupBaseRepInfoPackagePart_DO</dataObjectTypeID>
      <dataObjectTypeOccurrence>
        <minOccurrence>1</minOccurrence>
        <maxOccurrence>1</maxOccurrence>
      </dataObjectTypeOccurrence>
      <dataObjectTypeFormat>
        <mimeType>binary</mimeType>
      </dataObjectTypeFormat>
      <dataObjectTypeEncoded>
        <encodingName>application/zip</encodingName>
        <encodingDescription />
      </dataObjectTypeEncoded>
    </dataObjectType>
  </groupType>
</transferObjectTypeDescriptor>
```

C2.6 ESA-SAFE—TRF_MTD_PACKAGE_REP_INFO TRANSFER DESCRIPTOR

```
<?xml version="1.0" encoding="UTF-8"?>
<transferObjectTypeDescriptor xmlns="urn:ccsds:schema:pais:1">
  <identification>
    <descriptorModelID>CCSD0014</descriptorModelID>
    <descriptorModelVersion>1.0</descriptorModelVersion>
    <descriptorID>TRF_MTD_PACKAGE_REP_INFO</descriptorID>
  </identification>
  <description>
    <transferObjectTypeTitle>Meta Data Package Representation Information
    </transferObjectTypeTitle>
    <transferObjectTypeDescription>General Documentation associated to ERS
EO
SIP formats: Strip line and Strip scene
    </transferObjectTypeDescription>
    <transferObjectTypeOccurrence>
      <minOccurrence>1</minOccurrence>
      <maxOccurrence>1</maxOccurrence>
    </transferObjectTypeOccurrence>
  </description>
  <relation>
    <parentCollection>COL_ERS-AMI-SAR-Rep-Info</parentCollection>
```

```
  </relation>
  <groupType>
    <groupTypeID>GroupMTDRepInfoPackagePart</groupTypeID>
    <groupTypeStructureName>Set</groupTypeStructureName>
    <groupTypeOccurrence>
      <minOccurrence>1</minOccurrence>
      <maxOccurrence>1</maxOccurrence>
    </groupTypeOccurrence>
    <dataObjectType>
      <dataObjectTypeID>GroupMTDRepInfoPackagePart_DO</dataObjectTypeID>
      <dataObjectTypeOccurrence>
        <minOccurrence>1</minOccurrence>
        <maxOccurrence>1</maxOccurrence>
      </dataObjectTypeOccurrence>
      <dataObjectTypeFormat>
        <mimeType>binary</mimeType>
      </dataObjectTypeFormat>
      <dataObjectTypeEncoded>
        <encodingName>application/zip</encodingName>
        <encodingDescription>Zip Archive file</encodingDescription>
      </dataObjectTypeEncoded>
        <dataObjectTypeAssociation>
       <targetID>TRF_BASE_PACKAGE_REP_INFO</targetID>
      <relationDescription>
        <relationType>linked Representation Information</relationType>
        <relationTextualDescription>meaningful with base Package
Representation Information
        </relationTextualDescription>
        </relationDescription>
      </dataObjectTypeAssociation>
    </dataObjectType>
  </groupType>
</transferObjectTypeDescriptor>
```

C2.7 ESA-SAFE—TRF_DAT_PACKAGE_REP_INFO TRANSFER DESCRIPTOR

```
<?xml version="1.0" encoding="UTF-8"?>
<transferObjectTypeDescriptor xmlns="urn:ccsds:schema:pais:1">
  <identification>
    <descriptorModelID>CCSD0014</descriptorModelID>
    <descriptorModelVersion>1.0</descriptorModelVersion>
    <descriptorID>TRF_DAT_PACKAGE_REP_INFO</descriptorID>
  </identification>
  <description>
    <transferObjectTypeTitle>Data Package Representation Information
    </transferObjectTypeTitle>
    <transferObjectTypeDescription>ERS SAR L0 Representation Information of
             the data components (DFDL schemas)
    </transferObjectTypeDescription>
    <transferObjectTypeOccurrence>
      <minOccurrence>1</minOccurrence>
      <maxOccurrence>1</maxOccurrence>
    </transferObjectTypeOccurrence>
  </description>
 <relation>
    <parentCollection>COL_ERS-AMI-SAR-Rep-Info</parentCollection>
  </relation>
  <groupType>
```

```
    <groupTypeID>GroupDatRepInfoPackagePart</groupTypeID>
    <groupTypeStructureName>Set</groupTypeStructureName>
    <groupTypeOccurrence>
      <minOccurrence>1</minOccurrence>
      <maxOccurrence>1</maxOccurrence>
    </groupTypeOccurrence>
    <dataObjectType>
      <dataObjectTypeID>GroupDatRepInfoPackagePart_DO</dataObjectTypeID>
      <dataObjectTypeOccurrence>
        <minOccurrence>1</minOccurrence>
        <maxOccurrence>1</maxOccurrence>
      </dataObjectTypeOccurrence>
      <dataObjectTypeFormat>
        <mimeType>binary</mimeType>
      </dataObjectTypeFormat>
      <dataObjectTypeEncoded>
        <encodingName>application/zip</encodingName>
        <encodingDescription>Zip Archive file</encodingDescription>
      </dataObjectTypeEncoded>
        <dataObjectTypeAssociation>
       <targetID>TRF_BASE_PACKAGE_REP_INFO</targetID>
      <relationDescription>
        <relationType>linked Representation Information</relationType>
        <relationTextualDescription>meaningful with base Package
                  Representation Information
        </relationTextualDescription>
        </relationDescription>
      </dataObjectTypeAssociation>
    </dataObjectType>
  </groupType>
</transferObjectTypeDescriptor>
```

C2.8 ESA-SAFE—TRF_ERS_AMI_SAR_DOC TRANSFER DESCRIPTOR

```
<?xml version="1.0" encoding="UTF-8"?>
<transferObjectTypeDescriptor xmlns="urn:ccsds:schema:pais:1">
  <identification>
    <descriptorModelID>CCSD0014</descriptorModelID>
    <descriptorModelVersion>1.0</descriptorModelVersion>
    <descriptorID>TRF_ERS_AMI_SAR_DOC</descriptorID>
  </identification>
  <description>
    <transferObjectTypeTitle>Documentation associated with the data
products
    </transferObjectTypeTitle>
    <transferObjectTypeDescription>Various documents related to ERS L0 data
         products
    </transferObjectTypeDescription>
    <transferObjectTypeOccurrence>
      <minOccurrence>1</minOccurrence>
      <maxOccurrence>1</maxOccurrence>
    </transferObjectTypeOccurrence>
  </description>
  <relation>
    <parentCollection>COL_ERS-AMI-SAR-DOC</parentCollection>
  </relation>
  <groupType>
    <groupTypeID>GroupEODoc</groupTypeID>
```

```
    <groupTypeStructureName>Directory</groupTypeStructureName>
    <groupTypeOccurrence>
      <minOccurrence>1</minOccurrence>
      <maxOccurrence>1</maxOccurrence>
    </groupTypeOccurrence>
    <dataObjectType>
      <dataObjectTypeID>GroupEODoc_DO</dataObjectTypeID>
      <dataObjectTypeOccurrence>
        <minOccurrence>1</minOccurrence>
        <maxOccurrence>6</maxOccurrence>
      </dataObjectTypeOccurrence>
      <dataObjectTypeFormat>
        <mimeType>application/pdf</mimeType>
      </dataObjectTypeFormat>
    </dataObjectType>
  </groupType>
</transferObjectTypeDescriptor>
```

C2.9 ESA-SAFE—TRF_ER2_OPER_SAR_IM_PRODUCT TRANSFER DESCRIPTOR

```
<?xml version="1.0" encoding="UTF-8"?>
<transferObjectTypeDescriptor xmlns="urn:ccsds:schema:pais:1">
  <identification>
    <descriptorModelID>CCSD0014</descriptorModelID>
    <descriptorModelVersion>1.0</descriptorModelVersion>
    <descriptorID>TRF_ER2_OPER_SAR_IM_PRODUCT</descriptorID>
  </identification>
  <description>
    <transferObjectTypeTitle>EO Data Product</transferObjectTypeTitle>
    <transferObjectTypeDescription>ERS SAR L0 data product
    </transferObjectTypeDescription>
    <transferObjectTypeOccurrence>
      <minOccurrence>1</minOccurrence>
      <maxUnknown />
    </transferObjectTypeOccurrence>
  </description>
  <relation>
    <parentCollection>COL_ERS-AMI-SAR-EO-DATA</parentCollection>
  </relation>
  <groupType>
    <groupTypeID>GroupEODataProductPackagePart</groupTypeID>
    <groupTypeStructureName>Set</groupTypeStructureName>
    <groupTypeOccurrence>
      <minOccurrence>1</minOccurrence>
      <maxOccurrence>1</maxOccurrence>
    </groupTypeOccurrence>
    <dataObjectType>
      <dataObjectTypeID>GroupEODataProductPackagePart_DO</dataObjectTypeID>
      <dataObjectTypeOccurrence>
        <minOccurrence>1</minOccurrence>
        <maxOccurrence>1</maxOccurrence>
      </dataObjectTypeOccurrence>
      <dataObjectTypeFormat>
        <mimeType>binary</mimeType>
      </dataObjectTypeFormat>
      <dataObjectTypeEncoded>
        <encodingName>application/zip</encodingName>
```

```
        <encodingDescription>Zip Archive file</encodingDescription>
      </dataObjectTypeEncoded>
      <dataObjectTypeAssociation>
        <targetID>TRF_DAT_PACKAGE_REP_INFO</targetID>
        <relationDescription>
          <relationType>linked Representation Information</relationType>
          <relationTextualDescription>DATA Package Representation
Information
          </relationTextualDescription>
        </relationDescription>
      </dataObjectTypeAssociation>
      <dataObjectTypeAssociation>
        <targetID>TRF_MTD_PACKAGE_REP_INFO</targetID>
        <relationDescription>
          <relationType>linked Representation Information</relationType>
          <relationTextualDescription>METADATA Package Representation
                        Information
        </relationTextualDescription>
        </relationDescription>
      </dataObjectTypeAssociation>
        <dataObjectTypeAssociation>
        <targetID>TRF_ERS_AMI_SAR_DOC</targetID>
        <relationDescription>
          <relationType>linked context documentation</relationType>
          <relationTextualDescription>Associated documents
          </relationTextualDescription>
        </relationDescription>
      </dataObjectTypeAssociation>
    </dataObjectType>
  </groupType>
</transferObjectTypeDescriptor>
```

C2.10 ESA-SAFE—SIP CONSTRAINTS

```
<?xml version="1.0" encoding="UTF-8"?>
<sipConstraints xmlns="urn:ccsds:schema:pais:1">
  <producerArchiveProjectID>ESA_ERS_AMI_SAR</producerArchiveProjectID>
  <sipContentType>
    <sipContentTypeID>SIP_BASE_REP_INFO</sipContentTypeID>
    <authorizedDescriptor>
      <descriptorID>TRF_BASE_PACKAGE_REP_INFO</descriptorID>
      <occurrence>
        <minOccurrence>1</minOccurrence>
        <maxOccurrence>1</maxOccurrence>
      </occurrence>
    </authorizedDescriptor>
  </sipContentType>
  <sipContentType>
    <sipContentTypeID>SIP_MTD_REP_INFO</sipContentTypeID>
    <authorizedDescriptor>
      <descriptorID>TRF_MTD_PACKAGE_REP_INFO</descriptorID>
      <occurrence>
        <minOccurrence>1</minOccurrence>
        <maxOccurrence>1</maxOccurrence>
      </occurrence>
    </authorizedDescriptor>
  </sipContentType>
  <sipContentType>
```

```
    <sipContentTypeID>SIP_DAT_REP_INFO</sipContentTypeID>
    <authorizedDescriptor>
      <descriptorID>TRF_DAT_PACKAGE_REP_INFO</descriptorID>
      <occurrence>
        <minOccurrence>1</minOccurrence>
        <maxOccurrence>1</maxOccurrence>
      </occurrence>
    </authorizedDescriptor>
  </sipContentType>
  <sipContentType>
    <sipContentTypeID>SIP_ERS_DOC</sipContentTypeID>
    <authorizedDescriptor>
      <descriptorID>TRF_ERS_AMI_SAR_DOC</descriptorID>
      <occurrence>
        <minOccurrence>1</minOccurrence>
        <maxOccurrence>1</maxOccurrence>
      </occurrence>
    </authorizedDescriptor>
  </sipContentType>
  <sipContentType>
    <sipContentTypeID>SIP_ERS_EO_DATA_PRODUCT</sipContentTypeID>
    <authorizedDescriptor>
      <descriptorID>TRF_ER2_OPER_SAR_IM_PRODUCT</descriptorID>
      <occurrence>
        <minOccurrence>1</minOccurrence>
        <maxOccurrence>1</maxOccurrence>
      </occurrence>
    </authorizedDescriptor>
  </sipContentType>
  <sipSequencingConstraintGroup>
    <groupName>ERS Group</groupName>
    <constraintItem>
      <sipContentTypeID>SIP_BASE_REP_INFO</sipContentTypeID>
      <constraintSerialNumber>1</constraintSerialNumber>
    </constraintItem>
    <constraintItem>
      <sipContentTypeID>SIP_MTD_REP_INFO</sipContentTypeID>
      <constraintSerialNumber>2</constraintSerialNumber>
    </constraintItem>
    <constraintItem>
      <sipContentTypeID>SIP_DAT_REP_INFO</sipContentTypeID>
      <constraintSerialNumber>3</constraintSerialNumber>
    </constraintItem>
    <constraintItem>
      <sipContentTypeID>SIP_ERS_DOC</sipContentTypeID>
      <constraintSerialNumber>4</constraintSerialNumber>
    </constraintItem>
    <constraintItem>
      <sipContentTypeID>SIP_ERS_EO_DATA_PRODUCT</sipContentTypeID>
      <constraintSerialNumber>5</constraintSerialNumber>
    </constraintItem>
  </sipSequencingConstraintGroup>
</sipConstraints>
```

C2.11 EXAMPLE OF SIP BUILDER SOFTWARE CONFIGURATION FILE

```
<?xml version="1.0" encoding="UTF-8"?>

<project xmlns="urn:fr:gael:schema:ccsds:pais:sip-builder:1"
         outputDirectory="C:\PAIS_DATA\datafiles\TC_ESA\TDS2\SIPs"
         outputPackaging="zip">
   <descriptors
baseDirectory="C:\PAIS_DATA\datafiles\TC_ESA\TDS2\pais_descriptors">
      <descriptor file="TRF_BASE_PACKAGE_REP_INFO.xml"/>
      <descriptor file="TRF_DAT_PACKAGE_REP_INFO.xml"/>
      <descriptor file="TRF_MTD_PACKAGE_REP_INFO.xml"/>
      <descriptor file="TRF_EO-SIP-DOC.xml"/>
      <descriptor file="TRF_ERS-DOC.xml"/>
      <descriptor file="TRF_ER2_OPER_SAR_IM_PRODUCT.xml"/>
      <descriptor file="sip-constraints.xml"/>
   </descriptors>
   <collectors
baseDirectory="C:\PAIS_DATA\datafiles\TC_ESA\TDS2\test_data">
      <!-- gathering files for SIP BASE.SAFE -->
      <collector typeId="GroupBaseRepInfoPackagePart_DO"
baseDirectory="RepInfo">
         <include>BASE.*PACKAGE.*.SAFE.zip</include>
      </collector>
       <!-- gathering files for SIP DAT.SAFE -->
       <collector typeId="GroupDatRepInfoPackagePart_DO"
baseDirectory="RepInfo">
          <include>ERS-AMI-SAR-LEVEL-0_.*DAT.SAFE.zip</include>
       </collector>
       <!-- gathering files for SIP MTD.SAFE -->
       <collector typeId="GroupMTDRepInfoPackagePart_DO"
baseDirectory="RepInfo">
          <include>ERS-AMI-SAR-LEVEL-0_.*MTD.SAFE.zip</include>
       </collector>
       <!-- gathering files for SIP ERS_DOC -->
       <collector typeId="GroupERSDoc" baseDirectory="ERS_DOC">
          <include>.*.zip</include>
       </collector>
       <!-- gathering files for SIP EO_SIP -->
       <collector typeId="GroupEOSIPDoc" baseDirectory="EO_SIP">
          <include>.*.pdf</include>
       </collector>
       <!-- gathering SAFE ERS AMI SAR L0 products -->
       <collector typeId="GroupEODataProductPackagePart_DO"
baseDirectory="EO_PRODUCT">
          <include>ER2_OPER_SAR_IM__0P_.*.SAFE.zip</include>
       </collector>
    </collectors>
</project>
```

C3 DETAILED CASE

C3.1 ESA-SAFE—ESA_ERS_AMI_SAR_DET COLLECTION DESCRIPTOR (ROOT)

```
<?xml version="1.0" encoding="UTF-8"?>
<collectionDescriptor xmlns="urn:ccsds:schema:pais:1">
  <identification>
    <descriptorModelID>CCSD0015</descriptorModelID>
    <descriptorModelVersion>1.0</descriptorModelVersion>
    <descriptorID>ESA_ERS_AMI_SAR_det</descriptorID>
  </identification>
  <description>
    <collectionTitle>Level 0 of ERS AMI SAR experiments</collectionTitle>
    <collectionDescription>LO data corresponding to Earth's oceans and land
images obtained with a suite of instruments, one of which is a SAR on the
AMI, onboard a ERS satellite</collectionDescription>
  </description>
  <relation>
    <parentCollection>NONE</parentCollection>
  </relation>
</collectionDescriptor>
```

C3.2 ESA-SAFE—COL_ERS-AMI-SAR-REP-INFO_DET COLLECTION DESCRIPTOR

```
<?xml version="1.0" encoding="UTF-8"?>
<collectionDescriptor xmlns="urn:ccsds:schema:pais:1">
  <identification>
    <descriptorModelID>CCSD0015</descriptorModelID>
    <descriptorModelVersion>1.0</descriptorModelVersion>
    <descriptorID>COL_ERS-AMI-SAR-Rep-Info_det</descriptorID>
  </identification>
  <description>
    <collectionTitle>ERS-AMI-SAR Representation Information for L0
data</collectionTitle>
    <collectionDescription>A set of SAFE v2.0 packages providing all the
Representation Information necessary for L0 data
products</collectionDescription>
  </description>
  <relation>
    <parentCollection>ESA_ERS_AMI_SAR_det</parentCollection>
  </relation>
</collectionDescriptor>
```

C3.3 ESA-SAFE—COL_ERS-AMI-SAR-DOC_DET COLLECTION DESCRIPTOR

```
<?xml version="1.0" encoding="UTF-8"?>
<collectionDescriptor xmlns="urn:ccsds:schema:pais:1">
  <identification>
    <descriptorModelID>CCSD0015</descriptorModelID>
    <descriptorModelVersion>1.0</descriptorModelVersion>
    <descriptorID>COL_ERS-AMI-SAR-DOC_det</descriptorID>
  </identification>
  <description>
    <collectionTitle>ERS-AMI-SAR associated documentation</collectionTitle>
    <collectionDescription>All documents necessary for the comprehension of
the mission, instruments, products, and sips</collectionDescription>
  </description>
  <relation>
    <parentCollection>ESA_ERS_AMI_SAR_det</parentCollection>
  </relation>
</collectionDescriptor>
```

C3.4 ESA-SAFE—COL_ERS-AMI-SAR-DOC_DET COLLECTION DESCRIPTOR

```
<?xml version="1.0" encoding="UTF-8"?>
<collectionDescriptor xmlns="urn:ccsds:schema:pais:1">
  <identification>
    <descriptorModelID>CCSD0015</descriptorModelID>
    <descriptorModelVersion>1.0</descriptorModelVersion>
    <descriptorID>COL_ERS-AMI-SAR-EO-DATA_det</descriptorID>
  </identification>
  <description>
    <collectionTitle>EO Product of Level 0 for ERS-AMI-
SAR</collectionTitle>
    <collectionDescription>COL_ERS-AMI-SAR-LEVEL-0</collectionDescription>
  </description>
  <relation>
    <parentCollection>ESA_ERS_AMI_SAR_det</parentCollection>
  </relation>
</collectionDescriptor>
```

C3.5 ESA-SAFE—TRF_BASE_PACKAGE_REP_INFO_DET TRANSFER DESCRIPTOR

```
<?xml version="1.0" encoding="UTF-8"?>
<transferObjectTypeDescriptor xmlns="urn:ccsds:schema:pais:1">
  <identification>
    <descriptorModelID>CCSD0014</descriptorModelID>
    <descriptorModelVersion>1.0</descriptorModelVersion>
    <descriptorID>TRF_BASE_PACKAGE_REP_INFO_det</descriptorID>
  </identification>
  <description>
    <transferObjectTypeTitle>Base Package Representation Information
    </transferObjectTypeTitle>
    <transferObjectTypeDescription>Representation information to be present
in any SAFE 2.0 compliant archive. It should be ingested before any other
packages.
   </transferObjectTypeDescription>
    <transferObjectTypeOccurrence>
```

```
      <minOccurrence>1</minOccurrence>
      <maxOccurrence>1</maxOccurrence>
    </transferObjectTypeOccurrence>
  </description>
  <relation>
    <parentCollection>COL_ERS-AMI-SAR-Rep-Info</parentCollection>
    <association>
      <targetID>TRF_ERS-DOC_det</targetID>
      <relationDescription>
        <relationType>context information</relationType>
        <relationTextualDescription>associated documentation
        </relationTextualDescription>
      </relationDescription>
    </association>
  </relation>
  <groupType>
    <groupTypeID>GroupBasePackage_det</groupTypeID>
    <groupTypeStructureName>Directory</groupTypeStructureName>
    <groupTypeOccurrence>
      <minOccurrence>1</minOccurrence>
      <maxOccurrence>1</maxOccurrence>
    </groupTypeOccurrence>
    <groupType>
      <groupTypeID>GroupConformance_det</groupTypeID>
      <groupTypeDescription>There will be one Conformance directory
      </groupTypeDescription>
      <groupTypeStructureName>directory</groupTypeStructureName>
      <groupTypeOccurrence>
        <minOccurrence>1</minOccurrence>
        <maxOccurrence>1</maxOccurrence>
      </groupTypeOccurrence>
      <dataObjectType>
        <dataObjectTypeID>GroupConformance_DO_det</dataObjectTypeID>
        <dataObjectTypeDescription>contents of Conformance directory
        </dataObjectTypeDescription>
        <dataObjectTypeOccurrence>
          <minOccurrence>1</minOccurrence>
          <maxUnknown />
        </dataObjectTypeOccurrence>
        <dataObjectTypeFormat>
          <mimeType>text/plain</mimeType>
        </dataObjectTypeFormat>
      </dataObjectType>
    </groupType>
    <groupType>
      <groupTypeID>GroupDocumentation_det</groupTypeID>
      <groupTypeDescription>There will be one Documentation directory
      </groupTypeDescription>
      <groupTypeStructureName>directory</groupTypeStructureName>
      <groupTypeOccurrence>
        <minOccurrence>1</minOccurrence>
        <maxOccurrence>1</maxOccurrence>
      </groupTypeOccurrence>
      <dataObjectType>
        <dataObjectTypeID>GroupDocumentation_DO_det</dataObjectTypeID>
        <dataObjectTypeDescription>contents of documentation directory
        </dataObjectTypeDescription>
        <dataObjectTypeOccurrence>
          <minOccurrence>1</minOccurrence>
          <maxUnknown />
```

```
      </dataObjectTypeOccurrence>
      <dataObjectTypeFormat>
        <mimeType>text/plain</mimeType>
      </dataObjectTypeFormat>
    </dataObjectType>
  </groupType>
  <groupType>
    <groupTypeID>GroupIndex_det</groupTypeID>
    <groupTypeDescription>There will be one Index directory
    </groupTypeDescription>
    <groupTypeStructureName>Directory</groupTypeStructureName>
    <groupTypeOccurrence>
      <minOccurrence>1</minOccurrence>
      <maxOccurrence>1</maxOccurrence>
    </groupTypeOccurrence>
    <dataObjectType>
      <dataObjectTypeID>GroupIndex_DO_det</dataObjectTypeID>
      <dataObjectTypeDescription />
      <dataObjectTypeOccurrence>
        <minOccurrence>1</minOccurrence>
        <maxUnknown />
      </dataObjectTypeOccurrence>
      <dataObjectTypeFormat>
        <mimeType>text/plain</mimeType>
      </dataObjectTypeFormat>
    </dataObjectType>
  </groupType>
  <groupType>
    <groupTypeID>GroupMetaData_det</groupTypeID>
    <groupTypeDescription>There will be one metadata directory
    </groupTypeDescription>
    <groupTypeStructureName>Directory</groupTypeStructureName>
    <groupTypeOccurrence>
      <minOccurrence>1</minOccurrence>
      <maxOccurrence>1</maxOccurrence>
    </groupTypeOccurrence>
    <dataObjectType>
      <dataObjectTypeID>GroupMetaData_DO_det</dataObjectTypeID>
      <dataObjectTypeDescription />
      <dataObjectTypeOccurrence>
        <minOccurrence>1</minOccurrence>
        <maxUnknown />
      </dataObjectTypeOccurrence>
      <dataObjectTypeFormat>
        <mimeType>text/plain</mimeType>
      </dataObjectTypeFormat>
    </dataObjectType>
  </groupType>
  <groupType>
    <groupTypeID>GroupXFDU_det</groupTypeID>
    <groupTypeDescription>There will be one xfdu directory
    </groupTypeDescription>
    <groupTypeStructureName>Directory</groupTypeStructureName>
    <groupTypeOccurrence>
      <minOccurrence>1</minOccurrence>
      <maxOccurrence>1</maxOccurrence>
    </groupTypeOccurrence>
    <dataObjectType>
      <dataObjectTypeID>GroupXFDU_DO_det</dataObjectTypeID>
      <dataObjectTypeDescription>contents of xfdu directory
```

```
        </dataObjectTypeDescription>
        <dataObjectTypeOccurrence>
          <minOccurrence>1</minOccurrence>
          <maxUnknown />
        </dataObjectTypeOccurrence>
        <dataObjectTypeFormat>
          <mimeType>text/plain</mimeType>
        </dataObjectTypeFormat>
      </dataObjectType>
    </groupType>
    <groupType>
      <groupTypeID>GroupOtherFiles_det</groupTypeID>
      <groupTypeDescription>Set</groupTypeDescription>
      <groupTypeStructureName>Directory</groupTypeStructureName>
      <dataObjectType>
        <dataObjectTypeID>GroupOtherFiles_DO_det</dataObjectTypeID>
        <dataObjectTypeDescription />
        <dataObjectTypeOccurrence>
          <minOccurrence>1</minOccurrence>
          <maxOccurrence>2</maxOccurrence>
        </dataObjectTypeOccurrence>
      </dataObjectType>
    </groupType>
  </groupType>
</transferObjectTypeDescriptor>
```

C3.6 ESA-SAFE—TRF_MTD_PACKAGE_REP_INFO_DET TRANSFER DESCRIPTOR

```
<?xml version="1.0" encoding="UTF-8"?>
<transferObjectTypeDescriptor xmlns="urn:ccsds:schema:pais:1">
  <identification>
    <descriptorModelID>CCSD0014</descriptorModelID>
    <descriptorModelVersion>1.0</descriptorModelVersion>
    <descriptorID>TRF_MTD_PACKAGE_REP_INFO_det</descriptorID>
  </identification>
  <description>
    <transferObjectTypeTitle>Meta Data Package Representation
Information</transferObjectTypeTitle>
    <transferObjectTypeDescription>General Documentation associated to ERS
EO SIP formats: Strip line and Strip scene</transferObjectTypeDescription>
    <transferObjectTypeOccurrence>
      <minOccurrence>1</minOccurrence>
      <maxOccurrence>1</maxOccurrence>
    </transferObjectTypeOccurrence>
  </description>
  <relation>
    <parentCollection>COL_ERS-AMI-SAR-Rep-Info</parentCollection>
    <association>
      <targetID>TRF_BASE_PACKAGE_REP_INFO_det</targetID>
      <relationDescription>
        <relationType>linked Representation Information</relationType>
        <relationTextualDescription>meaningful with base Package
         Representation Information
      </relationTextualDescription>
      </relationDescription>
    </association>
  </relation>
```

```
<groupType>
  <groupTypeID>GroupMTDPackage_det</groupTypeID>
  <groupTypeStructureName>Directory</groupTypeStructureName>
  <groupTypeOccurrence>
    <minOccurrence>1</minOccurrence>
    <maxOccurrence>1</maxOccurrence>
  </groupTypeOccurrence>
<groupType>
  <groupTypeID>GroupMTDMetaData_det</groupTypeID>
  <groupTypeDescription>There will be one metadata directory
  </groupTypeDescription>
  <groupTypeStructureName>Directory</groupTypeStructureName>
  <groupTypeOccurrence>
    <minOccurrence>1</minOccurrence>
    <maxOccurrence>1</maxOccurrence>
  </groupTypeOccurrence>
  <dataObjectType>
    <dataObjectTypeID>GroupMTDMetaData_DO_det</dataObjectTypeID>
    <dataObjectTypeDescription>contents of metadata directory
    </dataObjectTypeDescription>
    <dataObjectTypeOccurrence>
      <minOccurrence>1</minOccurrence>
      <maxUnknown />
    </dataObjectTypeOccurrence>
    <dataObjectTypeFormat>
      <mimeType>text/plain</mimeType>
    </dataObjectTypeFormat>
  </dataObjectType>
</groupType>
<groupType>
  <groupTypeID>GroupMTDXFDU_det</groupTypeID>
  <groupTypeDescription>There will be one xfdu directory
  </groupTypeDescription>
  <groupTypeStructureName>Directory</groupTypeStructureName>
  <groupTypeOccurrence>
    <minOccurrence>1</minOccurrence>
    <maxOccurrence>1</maxOccurrence>
  </groupTypeOccurrence>
  <dataObjectType>
    <dataObjectTypeID>GroupMTDXFDU_DO_det</dataObjectTypeID>
    <dataObjectTypeDescription>contents of xfdu directory
    </dataObjectTypeDescription>
    <dataObjectTypeOccurrence>
      <minOccurrence>1</minOccurrence>
      <maxUnknown />
    </dataObjectTypeOccurrence>
    <dataObjectTypeFormat>
      <mimeType>text/plain</mimeType>
    </dataObjectTypeFormat>
  </dataObjectType>
</groupType>
<groupType>
  <groupTypeID>GroupMTDOtherFiles_det</groupTypeID>
  <groupTypeStructureName>Set</groupTypeStructureName>
  <groupTypeOccurrence>
    <minOccurrence>1</minOccurrence>
    <maxOccurrence>1</maxOccurrence>
  </groupTypeOccurrence>
  <dataObjectType>
    <dataObjectTypeID>GroupMTDOtherFiles_DO_det</dataObjectTypeID>
```

```
      <dataObjectTypeDescription>stand-alone files : manifest + schema
      </dataObjectTypeDescription>
      <dataObjectTypeOccurrence>
        <minOccurrence>1</minOccurrence>
        <maxOccurrence>2</maxOccurrence>
      </dataObjectTypeOccurrence>
      <dataObjectTypeFormat>
        <mimeType>text/plain</mimeType>
      </dataObjectTypeFormat>
    </dataObjectType>
  </groupType>
   </groupType>
</transferObjectTypeDescriptor>
```

C3.7 ESA-SAFE—TRF_DAT_PACKAGE_REP_INFO_DET TRANSFER DESCRIPTOR

```
<?xml version="1.0" encoding="UTF-8"?>
<transferObjectTypeDescriptor xmlns="urn:ccsds:schema:pais:1">
  <identification>
    <descriptorModelID>CCSD0014</descriptorModelID>
    <descriptorModelVersion>1.0</descriptorModelVersion>
    <descriptorID>TRF_DAT_PACKAGE_REP_INFO_det</descriptorID>
  </identification>
  <description>
    <transferObjectTypeTitle>Data Package Representation
Information</transferObjectTypeTitle>
    <transferObjectTypeDescription>ERS SAR L0 Representation Information of
the data components (DFDL schemas)</transferObjectTypeDescription>
    <transferObjectTypeOccurrence>
      <minOccurrence>1</minOccurrence>
      <maxOccurrence>1</maxOccurrence>
    </transferObjectTypeOccurrence>
  </description>
  <relation>
    <parentCollection>COL_ERS-AMI-SAR-Rep-Info</parentCollection>
    <association>
      <targetID>TRF_BASE_PACKAGE_REP_INFO_det</targetID>
      <relationDescription>
        <relationType>linked Representation Information</relationType>
        <relationTextualDescription>meaningful with base Package
              Representation Information
      </relationTextualDescription>
      </relationDescription>
    </association>
  </relation>
  <groupType>
    <groupTypeID>GroupDATPackage_det</groupTypeID>
    <groupTypeStructureName>Directory</groupTypeStructureName>
    <groupTypeOccurrence>
      <minOccurrence>1</minOccurrence>
      <maxOccurrence>1</maxOccurrence>
    </groupTypeOccurrence>
    <groupType>
      <groupTypeID>GroupDATMeasurementData_det</groupTypeID>
      <groupTypeDescription>There will be one measurement directory
      </groupTypeDescription>
```

```
      <groupTypeStructureName>Directory</groupTypeStructureName>
      <groupTypeOccurrence>
        <minOccurrence>1</minOccurrence>
        <maxOccurrence>1</maxOccurrence>
      </groupTypeOccurrence>
      <dataObjectType>
        <dataObjectTypeID>GroupDATMeasurementData_DO_det</dataObjectTypeID>
        <dataObjectTypeDescription>contents of measurement directory
        </dataObjectTypeDescription>
        <dataObjectTypeOccurrence>
          <minOccurrence>1</minOccurrence>
          <maxUnknown />
        </dataObjectTypeOccurrence>
        <dataObjectTypeFormat>
          <mimeType>text/plain</mimeType>
        </dataObjectTypeFormat>
      </dataObjectType>
    </groupType>
    <groupType>
      <groupTypeID>GroupDATOtherFiles_det</groupTypeID>
      <groupTypeStructureName>Set</groupTypeStructureName>
      <groupTypeOccurrence>
        <minOccurrence>1</minOccurrence>
        <maxOccurrence>1</maxOccurrence>
      </groupTypeOccurrence>
      <dataObjectType>
        <dataObjectTypeID>GroupDATOtherFiles_DO_det</dataObjectTypeID>
        <dataObjectTypeDescription>stand-alone files : manifest + schema
        </dataObjectTypeDescription>
        <dataObjectTypeOccurrence>
          <minOccurrence>1</minOccurrence>
          <maxOccurrence>2</maxOccurrence>
        </dataObjectTypeOccurrence>
        <dataObjectTypeFormat>
          <mimeType>text/plain</mimeType>
        </dataObjectTypeFormat>
      </dataObjectType>
    </groupType>
  </groupType>
</transferObjectTypeDescriptor>
```

C3.8 ESA-SAFE—TRF_ERS_AMI_SAR_DOC_DET TRANSFER DESCRIPTOR

```
<?xml version="1.0" encoding="UTF-8"?>
<transferObjectTypeDescriptor xmlns="urn:ccsds:schema:pais:1">
  <identification>
    <descriptorModelID>CCSD0014</descriptorModelID>
    <descriptorModelVersion>1.0</descriptorModelVersion>
    <descriptorID>TRF_ERS_AMI_SAR_DOC_det</descriptorID>
  </identification>
  <description>
    <transferObjectTypeTitle>Documentation associated with the data
products </transferObjectTypeTitle>
    <transferObjectTypeDescription>Various documents related to ERS L0 data
products </transferObjectTypeDescription>
    <transferObjectTypeOccurrence>
      <minOccurrence>1</minOccurrence>
      <maxUnknown>1</maxUnknown>
```

```
    </transferObjectTypeOccurrence>
  </description>
  <relation>
    <parentCollection>COL_ERS-AMI-SAR-DOC_det</parentCollection>
  </relation>
  <groupType>
    <groupTypeID>GroupEODoc_det</groupTypeID>
    <groupTypeStructureName>Directory</groupTypeStructureName>
    <groupTypeOccurrence>
      <minOccurrence>1</minOccurrence>
      <maxOccurrence>1</maxOccurrence>
    </groupTypeOccurrence>
    <dataObjectType>
      <dataObjectTypeID>GroupEODoc_DO_det</dataObjectTypeID>
        <dataObjectTypeDescription>A document in pdf format relative to ERS
          L0 EO products
      </dataObjectTypeDescription>
      <dataObjectTypeOccurrence>
        <minOccurrence>1</minOccurrence>
        <maxOccurrence>6</maxOccurrence>
      </dataObjectTypeOccurrence>
      <dataObjectTypeFormat>
        <mimeType>application/pdf</mimeType>
      </dataObjectTypeFormat>
    </dataObjectType>
  </groupType>
</transferObjectTypeDescriptor>
```

C3.9 ESA-SAFE—TRF_ERS_AMI_SAR_IM_DET TRANSFER DESCRIPTOR

```
<?xml version="1.0" encoding="UTF-8"?>
<transferObjectTypeDescriptor xmlns="urn:ccsds:schema:pais:1">
  <identification>
    <descriptorModelID>CCSD0014</descriptorModelID>
    <descriptorModelVersion>1.0</descriptorModelVersion>
    <descriptorID>TRF_ER2_OPER_SAR_IM_PRODUCT_det</descriptorID>
  </identification>
  <description>
    <transferObjectTypeTitle>EO Data Product</transferObjectTypeTitle>
    <transferObjectTypeDescription>ERS SAR L0 data product
    </transferObjectTypeDescription>
    <transferObjectTypeOccurrence>
      <minOccurrence>1</minOccurrence>
      <maxUnknown />
    </transferObjectTypeOccurrence>
  </description>
  <relation>
    <parentCollection>COL_ERS-AMI-SAR-EO-DATA</parentCollection>
  </relation>
  <groupType>
    <groupTypeID>GroupEODataProductPackagePart_det</groupTypeID>
    <groupTypeStructureName>Directory</groupTypeStructureName>
    <groupTypeOccurrence>
      <minOccurrence>1</minOccurrence>
      <maxOccurrence>1</maxOccurrence>
    </groupTypeOccurrence>
    <groupType>
      <groupTypeID>GroupEODataProduct_det</groupTypeID>
```

```
      <groupTypeDescription>There will be one SAFE directory per product
      </groupTypeDescription>
      <groupTypeStructureName>Directory</groupTypeStructureName>
      <groupTypeOccurrence>
        <minOccurrence>1</minOccurrence>
        <maxOccurrence>1</maxOccurrence>
      </groupTypeOccurrence>
      <dataObjectType>
        <dataObjectTypeID>GroupEODataProduct_Meas_det</dataObjectTypeID>
            <dataObjectTypeDescription>the MEASUREMENT.DAT file
        </dataObjectTypeDescription>
        <dataObjectTypeOccurrence>
          <minOccurrence>1</minOccurrence>
          <maxOccurrence>1</maxOccurrence>
        </dataObjectTypeOccurrence>
        <dataObjectTypeFormat>
          <mimeType>application/binary</mimeType>
        </dataObjectTypeFormat>
      </dataObjectType>
      <dataObjectType>
        <dataObjectTypeID>GroupEODataProduct_XML_det</dataObjectTypeID>
            <dataObjectTypeDescription>the MANIFEST.XML file and the SAFE-
                SAR-ERS-AMI-SAR-LEVEL0.XML file
        </dataObjectTypeDescription>
        <dataObjectTypeOccurrence>
          <minOccurrence>1</minOccurrence>
          <maxOccurrence>2</maxOccurrence>
        </dataObjectTypeOccurrence>
        <dataObjectTypeFormat>
          <mimeType>application/xhtml+xml</mimeType>
        </dataObjectTypeFormat>
      </dataObjectType>
      <dataObjectType>
        <dataObjectTypeID>GroupEODataProduct_SIX_det</dataObjectTypeID>
        <dataObjectTypeDescription>the MEASUREMENT.SIX file
        </dataObjectTypeDescription>
        <dataObjectTypeOccurrence>
          <minOccurrence>1</minOccurrence>
          <maxOccurrence>1</maxOccurrence>
        </dataObjectTypeOccurrence>
        <dataObjectTypeFormat>
          <mimeType>application/octet-stream</mimeType>
        </dataObjectTypeFormat>
      </dataObjectType>
    </groupType>
  </groupType>
</transferObjectTypeDescriptor>
```

C3.10 .10 SIP CONSTRAINTS

```
<?xml version="1.0" encoding="UTF-8"?>
<sipConstraints xmlns="urn:ccsds:schema:pais:1">
<producerArchiveProjectID>ESA_ERS_AMI_SAR_det</producerArchiveProjectID>
   <sipContentType>
      <sipContentTypeID>SIP_BASE_REP_INFO_det</sipContentTypeID>
      <authorizedDescriptor>
            <descriptorID>TRF_BASE_PACKAGE_REP_INFO_det</descriptorID>
            <occurrence>
                  <minOccurrence>1</minOccurrence>
                  <maxOccurrence>1</maxOccurrence>
            </occurrence>
      </authorizedDescriptor>
      </sipContentType>
      <sipContentType>
            <sipContentTypeID>SIP_MTD_REP_INFO_det</sipContentTypeID>
            <authorizedDescriptor>
                  <descriptorID>TRF_MTD_PACKAGE_REP_INFO_det</descriptorID>
                  <occurrence>
                        <minOccurrence>1</minOccurrence>
                        <maxOccurrence>1</maxOccurrence>
                  </occurrence>
            </authorizedDescriptor>
      </sipContentType>
      <sipContentType>
            <sipContentTypeID>SIP_DAT_REP_INFO_det</sipContentTypeID>
            <authorizedDescriptor>
                  <descriptorID>TRF_DAT_PACKAGE_REP_INFO_det</descriptorID>
                  <occurrence>
                        <minOccurrence>1</minOccurrence>
                        <maxOccurrence>1</maxOccurrence>
                  </occurrence>
            </authorizedDescriptor>
      </sipContentType>
      <sipContentType>
        <sipContentTypeID>SIP_ERS_DOC_det</sipContentTypeID>
            <authorizedDescriptor>
                  <descriptorID>TRF_ERS_AMI_SAR_DOC</descriptorID>
                  <occurrence>
                        <minOccurrence>1</minOccurrence>
                        <maxOccurrence>1</maxOccurrence>
                  </occurrence>
            </authorizedDescriptor>
      </sipContentType>
      <sipContentType>
        <sipContentTypeID>SIP_ERS_EO_DATA_PRODUCT_det</sipContentTypeID>
            <authorizedDescriptor>
            <descriptorID>TRF_ER2_OPER_SAR_IM_PRODUCT_det</descriptorID>
                  <occurrence>
                        <minOccurrence>1</minOccurrence>
                        <maxOccurrence>1</maxOccurrence>
                  </occurrence>
            </authorizedDescriptor>
      </sipContentType>
      <sipSequencingConstraintGroup>
            <groupName>Group Safe2</groupName>
            <constraintItem>
```

```
      <sipContentTypeID>SIP_BASE_REP_INFO_det</sipContentTypeID>
                  <constraintSerialNumber>1</constraintSerialNumber>
            </constraintItem>
            <constraintItem>
                  <sipContentTypeID>SIP_MTD_REP_INFO_det</sipContentTypeID>
                  <constraintSerialNumber>2</constraintSerialNumber>
            </constraintItem>
            <constraintItem>
                  <sipContentTypeID>SIP_DAT_REP_INFO_det</sipContentTypeID>
                  <constraintSerialNumber>3</constraintSerialNumber>
            </constraintItem>
            <constraintItem>
                  <sipContentTypeID>SIP_ERS_DOC_det</sipContentTypeID>
                  <constraintSerialNumber>4</constraintSerialNumber>
            </constraintItem>
            <constraintItem>
      <sipContentTypeID>SIP_ERS_EO_DATA_PRODUCT_det</sipContentTypeID>
                  <constraintSerialNumber>5</constraintSerialNumber>
            </constraintItem>
    </sipSequencingConstraintGroup>
</sipConstraints>
```

C3.11 MANIFEST OF THE EO DATA PRODUCT

```
<?xml version="1.0" encoding="UTF-8"?>
<xfdu:XFDU xmlns:xfdu="urn:ccsds:schema:xfdu:1"
xmlns:pais="urn:ccsds:schema:pais:1">
   <packageHeader ID="ESA_ERS_AMI_SAR_det-SIP-0005">
      <volumeInfo>
         <specificationVersion>1.0</specificationVersion>
      </volumeInfo>
      <environmentInfo>
         <extension>
            <pais:sipGlobalInformation>
               <pais:sipID>ESA_ERS_AMI_SAR_det-SIP-0005</pais:sipID>
                <pais:producerSourceID> ESA_ERS_PROJECT
                </pais:producerSourceID>
               <pais:producerArchiveProjectID>ESA_ERS_AMI_SAR_det
</pais:producerArchiveProjectID>
      <pais:sipContentTypeID>SIP_ERS_EO_DATA_PRODUCT_det</pais:sipContentTypeID>
                  <pais:sipSequenceNumber>5</pais:sipSequenceNumber>
            </pais:sipGlobalInformation>
         </extension>
      </environmentInfo>
   </packageHeader>
   <informationPackageMap>
      <xfdu:contentUnit>
         <extension>
            <pais:sipTransferObject>
      <pais:descriptorID>TRF_ER2_OPER_SAR_IM_PRODUCT_det</pais:descriptorID>
               <pais:transferObjectID>TRF_ER2_OPER_SAR_IM_PRODUCT_det-0001
              </pais:transferObjectID>
      <pais:lastTransferObjectFlag>FALSE</pais:lastTransferObjectFlag>
            </pais:sipTransferObject>
         </extension>
         <xfdu:contentUnit>
            <extension>
```

```
                <pais:sipTransferObjectGroup>
      <pais:associatedDescriptorGroupTypeID>GroupEODataProductPackagePart_det
      </pais:associatedDescriptorGroupTypeID>
                   <pais:transferObjectGroupInstanceName/>
                </pais:sipTransferObjectGroup>
             </extension>
             <xfdu:contentUnit>
                <extension>
                   <pais:sipTransferObjectGroup>

<pais:associatedDescriptorGroupTypeID>GroupEODataProduct_det
                </pais:associatedDescriptorGroupTypeID>
             <pais:transferObjectGroupInstanceName>
             ER2_OPER_SAR_IM__0P_19970923T212658_19970923T213316_43E3.SAFE
          </pais:transferObjectGroupInstanceName>
                   </pais:sipTransferObjectGroup>
                </extension>
                <xfdu:contentUnit>
                   <extension>
             <pais:sipDataObject>
      <pais:associatedDescriptorDataID>GroupEODataProduct_Meas_det
                        </pais:associatedDescriptorDataID>
                      </pais:sipDataObject>
                   </extension>
                   <dataObjectPointer dataObjectID="DO-
GroupEODataProduct_Meas_det-0001"/>
                </xfdu:contentUnit>
                <xfdu:contentUnit>
                   <extension>
                      <pais:sipDataObject>
             <pais:associatedDescriptorDataID>GroupEODataProduct_XML_det
                        </pais:associatedDescriptorDataID>
                      </pais:sipDataObject>
                   </extension>
                   <dataObjectPointer dataObjectID="DO-
GroupEODataProduct_XML_det-0001"/>
                </xfdu:contentUnit>
                <xfdu:contentUnit>
                   <extension>
                      <pais:sipDataObject>
             <pais:associatedDescriptorDataID>GroupEODataProduct_XML_det
             </pais:associatedDescriptorDataID>
                      </pais:sipDataObject>
                   </extension>
                   <dataObjectPointer dataObjectID="DO-
GroupEODataProduct_XML_det-0002"/>
                </xfdu:contentUnit>
                <xfdu:contentUnit>
                   <extension>
                      <pais:sipDataObject>
             <pais:associatedDescriptorDataID> GroupEODataProduct_SIX_det
             </pais:associatedDescriptorDataID>
                      </pais:sipDataObject>
                   </extension>
                   <dataObjectPointer dataObjectID="DO-
GroupEODataProduct_SIX_det-0001"/>
                </xfdu:contentUnit>
             </xfdu:contentUnit>
          </xfdu:contentUnit>
      </xfdu:contentUnit>
```

```
   </informationPackageMap>
   <dataObjectSection>
      <dataObject ID="DO-GroupEODataProduct_Meas_det-0001" size="2097152">
         <byteStream size="2097152">
            <fileLocation locatorType="URL"
href="file:ER2_OPER_SAR_IM__0P_19970923T212658_19970923T213316_43E3.SAFE/ME
ASUREMENT.DAT"/>
            <checksum
checksumName="MD5">b2d1236c286a3c0704224fe4105eca49</checksum>
         </byteStream>
      </dataObject>
      <dataObject ID="DO-GroupEODataProduct_XML_det-0001" size="11239">
         <byteStream size="11239">
            <fileLocation locatorType="URL"
href="file:ER2_OPER_SAR_IM__0P_19970923T212658_19970923T213316_43E3.SAFE/MA
NIFEST.XML"/>
            <checksum
checksumName="MD5">10349ee55dc45733fbb2f674e706a7a8</checksum>
         </byteStream>
      </dataObject>
      <dataObject ID="DO-GroupEODataProduct_XML_det-0002" size="5458">
         <byteStream size="5458">
            <fileLocation locatorType="URL"
href="file:ER2_OPER_SAR_IM__0P_19970923T212658_19970923T213316_43E3.SAFE/SA
FE-SAR-ERS-AMI-SAR-LEVEL0.XML"/>
            <checksum
checksumName="MD5">d81bc5df27f32d9562beeec587b87775</checksum>
         </byteStream>
      </dataObject>
      <dataObject ID="DO-GroupEODataProduct_SIX_det-0001" size="2772">
         <byteStream size="2772">
            <fileLocation locatorType="URL"
href="file:ER2_OPER_SAR_IM__0P_19970923T212658_19970923T213316_43E3.SAFE/ME
ASUREMENT.SIX"/>
            <checksum
checksumName="MD5">a7f8208e7e86929979ab0fdd407b605b</checksum>
         </byteStream>
      </dataObject>
   </dataObjectSection>
</xfdu:XFDU>
```

C3.12 EXAMPLE OF SIP BUILDER SOFTWARE CONFIGURATION FILE

```
<?xml version="1.0" encoding="UTF-8"?>
<project xmlns="urn:fr:gael:schema:ccsds:pais:sip-builder:1"
         outputDirectory="C:\PAIS_DATA\datafiles\TC_ESA\TDS1\SIPs"
         outputPackaging="zip">
   <descriptors
baseDirectory="C:\PAIS_DATA\datafiles\TC_ESA\TDS1\pais_descriptors">
      <descriptor file="TRF_BASE_PACKAGE_REP_INFO_det.xml"/>
      <descriptor file="TRF_DAT_PACKAGE_REP_INFO_det.xml"/>
      <descriptor file="TRF_MTD_PACKAGE_REP_INFO_det.xml"/>
      <descriptor file="TRF_EO-SIP-DOC_det.xml"/>
      <descriptor file="TRF_ERS-DOC_det.xml"/>
      <descriptor file="TRF_ER2_OPER_SAR_IM_PRODUCT_det.xml"/>
      <descriptor file="sip-constraints.xml"/>
   </descriptors>
```

```
   <collectors
baseDirectory="C:\PAIS_DATA\datafiles\TC_ESA\TDS1\test_data">
      <!-- gathering files for SIP BASE.SAFE -->
      <collector typeId="GroupBasePackage_det" baseDirectory="RepInfo">
         <include>BASE.*PACKAGE.*[.]SAFE</include>
      </collector>
      <collector typeId="GroupConformance_det">
         <include>conformance</include>
      </collector>
      <collector typeId="GroupConformance_DO_det">
         <include>**/.*[.]xsd</include>
      </collector>
      <collector typeId="GroupDocumentation_det">
         <include>documentation</include>
      </collector>
      <collector typeId="GroupDocumentation_DO_det">
         <include>**/.*[.]xsd</include>
      </collector>
      <collector typeId="GroupIndex_det">
         <include>index</include>
      </collector>
      <collector typeId="GroupIndex_DO_det">
         <include>**/.*[.]xsd</include>
      </collector>
      <collector typeId="GroupMetaData_det">
         <include>metadata</include>
      </collector>
      <collector typeId="GroupMetaData_DO_det">
         <include>**/.*[.]xsd</include>
      </collector>
      <collector typeId="GroupXFDU_det">
         <include>xfdu-safe</include>
      </collector>
      <collector typeId="GroupXFDU_DO_det">
         <include>**/.*[.]xsd</include>
      </collector>
      <collector typeId="GroupOtherFiles_det">
         <include>BASE.*PACKAGE.*[.]SAFE</include>
      </collector>
      <collector typeId="GroupOtherFiles_DO_det">
         <include>.*[.](xsd|XML)</include>
      </collector>
      <!-- gathering files for SIP DAT.SAFE -->
      <collector typeId="GroupDATPackage_det" baseDirectory="RepInfo">
         <include>ERS-AMI-SAR-LEVEL-0_.*DAT[.]SAFE</include>
      </collector>
      <collector typeId="GroupDATDocumentation_det">
         <include>documentation</include>
      </collector>
      <collector typeId="GroupDATDocumentation_DO_det">
         <include>**/.*[.]xsd</include>
      </collector>
      <collector typeId="GroupDATIndex_det">
         <include>index</include>
      </collector>
      <collector typeId="GroupDATIndex_DO_det">
         <include>**/.*[.]xsd</include>
      </collector>
      <collector typeId="GroupDATMeasurementData_det">
         <include>measurement</include>
```

```
</collector>
<collector typeId="GroupDATMeasurementData_DO_det">
   <include>**/.*[.]xsd</include>
</collector>
<collector typeId="GroupDATOtherFiles_det">
   <include>ERS-AMI-SAR-LEVEL-0_.*DAT[.]SAFE</include>
</collector>
<collector typeId="GroupDATOtherFiles_DO_det">
   <include>.*[.](xsd|XML)</include>
</collector>
<!-- gathering files for SIP MTD.SAFE -->
<collector typeId="GroupMTDPackage_det" baseDirectory="RepInfo">
   <include>ERS-AMI-SAR-LEVEL-0_.*MTD[.]SAFE</include>
</collector>
<collector typeId="GroupMTDConformance_det">
   <include>conformance</include>
</collector>
<collector typeId="GroupMTDConformance_DO_det">
   <include>**/.*[.]xsd</include>
</collector>
<collector typeId="GroupMTDMetaData_det">
   <include>metadata</include>
</collector>
<collector typeId="GroupMTDMetaData_DO_det">
   <include>**/.*[.]xsd</include>
</collector>
<collector typeId="GroupMTDXFDU_det">
   <include>xfdu-safe</include>
</collector>
<collector typeId="GroupMTDXFDU_DO_det">
   <include>**/.*[.]xsd</include>
</collector>
<collector typeId="GroupMTDOtherFiles_det">
   <include>ERS-AMI-SAR-LEVEL-0_.*MTD[.]SAFE</include>
</collector>
<collector typeId="GroupMTDOtherFiles_DO_det">
   <include>.*[.](xsd|XML)</include>
</collector>
<!-- gathering files for SIP ERS_DOC -->
<collector typeId="GroupERSDoc_det" baseDirectory="ERS_DOC">
   <include>.*[.]zip</include>
</collector>
<!-- gathering files for SIP EO_SIP -->
<collector typeId="GroupEOSIPDoc_det" baseDirectory="EO_SIP">
   <include>.*[.]pdf</include>
</collector>
<collector typeId="GroupEODataProductPackagePart_DO_det">
   <include>.*ER2_OPER_SAR_IM__0P_.*[.]SAFE[.]zip</include>
</collector>
<!-- gathering files for SIP MTD.SAFE -->

<collector typeId="GroupMTDPackage_det" baseDirectory="RepInfo">
   <include>ERS-AMI-SAR-LEVEL-0_.*MTD[.]SAFE</include>
</collector>
<collector typeId="GroupMTDConformance_det">
   <include>conformance</include>
</collector>
<collector typeId="GroupMTDConformance_DO_det">
   <include>**/.*[.]xsd</include>
</collector>
```

```
      <collector typeId="GroupMTDMetaData_det">
         <include>metadata</include>
      </collector>
      <collector typeId="GroupMTDMetaData_DO_det">
         <include>**/.*[.]xsd</include>
      </collector>
      <collector typeId="GroupMTDXFDU_det">
         <include>xfdu-safe</include>
      </collector>
      <collector typeId="GroupMTDXFDU_DO_det">
         <include>**/.*[.]xsd</include>
       </collector>
      <collector typeId="GroupMTDOtherFiles_det">
         <include>ERS-AMI-SAR-LEVEL-0_.*MTD[.]SAFE</include>
      </collector>
      <collector typeId="GroupMTDOtherFiles_DO_det">
         <include>.*[.](xsd|XML)</include>
      </collector>
      <!-- gathering files for SIP EO.SAFE -->
      <collector typeId="GroupEODataProductPackagePart_det"
baseDirectory="EO_PRODUCT">
         <include>ER2_OPER_SAR_IM__0P_.*[.]SAFE</include>
      </collector>
      <collector typeId="GroupEODPIndex_det">
         <include>index</include>
      </collector>
      <collector typeId="GroupEODPIndex_DO_det">
         <include>.*[.]SIX</include>
      </collector>
      <collector typeId="GroupEODPMetadata_det">
         <include>metadata</include>
      </collector>
      <collector typeId="GroupEODPMetadata_DO_det">
         <include>.*[.]XML</include>
      </collector>
      <collector typeId="GroupEODPMeasurement_det">
         <include>measurement</include>
      </collector>
      <collector typeId="GroupEODPMeasurement_DO_det">
         <include>MEASUREMENT.DAT</include>
      </collector>
      <collector typeId="GroupEODPOtherFiles_DO_det">
         <include>MANIFEST.XML</include>
      </collector>

   </collectors>
</project>
```

ANNEX D

COROT USE CASE

D1 INTRODUCTION

This annex contains the PAIS XML files from the the CoRoT use case (see 6.4.3): Descriptors and SIP constraints (D2–D5), examples of SIP XFDU (D6–D7), and an example of SIP Builder software configuration file for the generation of XFDU PAIS Conformant SIPs (D8).

D2 COROT N0—COLLECTION DESCRIPTOR

```
<?xml version="1.0" encoding="UTF-8"?>

<collectionDescriptor xmlns="urn:ccsds:schema:pais:1">

   <identification>
      <descriptorModelID>CCSD0015</descriptorModelID>
      <descriptorModelVersion>1.0</descriptorModelVersion>
      <descriptorID>CoRoT-N0</descriptorID>
   </identification>

   <description>
      <collectionTitle>CoRoT N0 Collection</collectionTitle>
      <collectionDescription>

Collection of CoRoT N0 data
      </collectionDescription>
   </description>

   <relation>
      <parentCollection>NONE</parentCollection>
   </relation>

</collectionDescriptor>
```

D3 COROT N0 PRODUCTS—TRANSFER OBJECT DESCRIPTOR

```
<?xml version="1.0" encoding="UTF-8"?>

<transferObjectTypeDescriptor xmlns="urn:ccsds:schema:pais:1">

   <identification>
      <descriptorModelID>CCSD0014</descriptorModelID>
      <descriptorModelVersion>V1.0</descriptorModelVersion>
      <descriptorID>CoRoT-N0-RUN </descriptorID>
      <producerSourceID>CNES</producerSourceID>
   </identification>

   <description>
      <transferObjectTypeTitle>
```

```
            CoRoT N0 RUN
         </transferObjectTypeTitle>
         <transferObjectTypeDescription>
            A set of CoRoT N0 Dataset of the same type and belonging
            to a single Run.
         </transferObjectTypeDescription>
         <transferObjectTypeOccurrence>
            <minOccurrence>1</minOccurrence>
            <maxUnknown/>
         </transferObjectTypeOccurrence>
         <transferObjectTypeSize>
            <maxSize>4</maxSize>
            <unitsType>GB</unitsType>
         </transferObjectTypeSize>
      </description>

      <relation>
         <parentCollection>CoRoT-N0</parentCollection>
      </relation>

      <groupType>

         <groupTypeID>CoRoT-N0-RUN</groupTypeID>
         <groupTypeDescription>
            A group denoting a single Run.
         </groupTypeDescription>
         <groupTypeStructureName>directory</groupTypeStructureName>
         <groupTypeOccurrence>
            <minOccurrence>1</minOccurrence>
            <maxOccurrence>1</maxOccurrence>
         </groupTypeOccurrence>

         <groupType>

            <groupTypeID>CoRoT-N0-DATASET-GROUP</groupTypeID>
            <groupTypeDescription>
               A group of CoRoT N0 Dataset of the same type.
            </groupTypeDescription>
            <groupTypeStructureName>directory</groupTypeStructureName>
            <groupTypeOccurrence>
               <minOccurrence>1</minOccurrence>
               <maxOccurrence>1</maxOccurrence>
            </groupTypeOccurrence>

            <dataObjectType>
               <dataObjectTypeID>CoRoT-N0-DATASET</dataObjectTypeID>
               <dataObjectTypeDescription>
                  A CoRoT N0 Dataset
               </dataObjectTypeDescription>
               <dataObjectTypeOccurrence>
                  <minOccurrence>1</minOccurrence>
                  <maxUnknown/>
               </dataObjectTypeOccurrence>
            </dataObjectType>

         </groupType>
      </groupType>
   </transferObjectTypeDescriptor>
```

D4 COROT HOUSKEEPING DATA (HK)—TRANSFER OBJECT DESCRIPTOR

```
<?xml version="1.0" encoding="UTF-8"?>

<transferObjectTypeDescriptor xmlns="urn:ccsds:schema:pais:1">

   <identification>
      <descriptorModelID>CCSD0014</descriptorModelID>
      <descriptorModelVersion>V1.0</descriptorModelVersion>
      <descriptorID>CoRoT-N0-HK </descriptorID>
      <producerSourceID>CNES</producerSourceID>
   </identification>

   <description>
      <transferObjectTypeTitle>CoRoT N0 - HK </transferObjectTypeTitle>
      <transferObjectTypeDescription>
         A set of CoRoT N0 housekeeping data
      </transferObjectTypeDescription>
      <transferObjectTypeOccurrence>
         <minOccurrence>1</minOccurrence>
         <maxUnknown/>
      </transferObjectTypeOccurrence>
   </description>

   <relation>
      <parentCollection>CoRoT-N0</parentCollection>
   </relation>

   <groupType>
      <groupTypeID>CoRoT-N0-HK-GROUP</groupTypeID>
      <groupTypeDescription>
         A group type for CoRoT N0 housekeeping data
      </groupTypeDescription>
      <groupTypeStructureName>directory</groupTypeStructureName>
      <groupTypeOccurrence>
         <minOccurrence>1</minOccurrence>
         <maxOccurrence>1</maxOccurrence>
      </groupTypeOccurrence>

      <dataObjectType>
         <dataObjectTypeID>CoRoT-N0-HK-DATA</dataObjectTypeID>
         <dataObjectTypeDescription>
            A CoRoT N0 housekeeping data
         </dataObjectTypeDescription>
         <dataObjectTypeOccurrence>
            <minOccurrence>1</minOccurrence>
            <maxUnknown/>
         </dataObjectTypeOccurrence>
      </dataObjectType>

   </groupType>

</transferObjectTypeDescriptor>
```

D5 COROT—SIP CONSTRAINTS

```
<?xml version="1.0" encoding="UTF-8"?>

<sipConstraints xmlns="urn:ccsds:schema:pais:1">

   <producerArchiveProjectID>CoRoT-N0</producerArchiveProjectID>

   <!-- SIPs of CoRoT N0 RUNs -->

   <sipContentType>
      <sipContentTypeID>SIP-CoRoT-N0-RUN</sipContentTypeID>
      <authorizedDescriptor>
         <descriptorID>CoRoT-N0-RUN </descriptorID>
         <occurrence>
            <minOccurrence>1</minOccurrence>
            <maxOccurrence>1</maxOccurrence>
         </occurrence>
      </authorizedDescriptor>
   </sipContentType>

   <!-- SIPs of CoRoT N0 housekeeping (HK) data -->

   <sipContentType>
      <sipContentTypeID>SIP-CoRoT-N0-HK </sipContentTypeID>
      <authorizedDescriptor>
         <descriptorID>CoRoT-N0-HK </descriptorID>
         <occurrence>
            <minOccurrence>1</minOccurrence>
            <maxOccurrence>1</maxOccurrence>
         </occurrence>
      </authorizedDescriptor>
   </sipContentType>

   <!-- Constraints: force HK before RUNs -->

   <sipSequencingConstraintGroup>
      <groupName>CoRoT N0</groupName>
      <constraintItem>
         <sipContentTypeID>SIP-CoRoT-N0-HK </sipContentTypeID>
         <constraintSerialNumber>1</constraintSerialNumber>
      </constraintItem>
      <constraintItem>
         <sipContentTypeID>SIP-CoRoT-N0-RUN</sipContentTypeID>
         <constraintSerialNumber>2</constraintSerialNumber>
      </constraintItem>
   </sipSequencingConstraintGroup>

</sipConstraints>
```

D6 FIRST SIP OF HOUSEKEEPING SERIES

```
<?xml version="1.0" encoding="UTF-8"?>

<xfdu:XFDU xmlns:pais="urn:ccsds:schema:pais:1"
           xmlns:xfdu="urn:ccsds:schema:xfdu:1">

   <packageHeader ID="CoRoT-N0-SIP-0001">
      <volumeInfo>
         <specificationVersion>1.0</specificationVersion>
      </volumeInfo>
      <environmentInfo>
         <extension>
            <pais:sipGlobalInformation>
               <pais:sipID>CoRoT-N0-SIP-0001</pais:sipID>
               <pais:producerSourceID>CNES</pais:producerSourceID>
               <pais:producerArchiveProjectID>CoRoT-N0↵
                  </pais:producerArchiveProjectID>
               <pais:sipContentTypeID>SIP-CoRoT-N0-HK↵
                  </pais:sipContentTypeID>
               <pais:sipSequenceNumber>1</pais:sipSequenceNumber>
            </pais:sipGlobalInformation>
         </extension>
      </environmentInfo>
   </packageHeader>

   <informationPackageMap>

      <xfdu:contentUnit>
         <extension>
            <pais:sipTransferObject>
               <pais:descriptorID>CoRoT-N0-HK</pais:descriptorID>
               <pais:transferObjectID>CoRoT-N0-HK-0001↵
                  </pais:transferObjectID>
               <pais:lastTransferObjectFlag>↵
                  FALSE</pais:lastTransferObjectFlag>
            </pais:sipTransferObject>
         </extension>

         <xfdu:contentUnit>
            <extension>
               <pais:sipTransferObjectGroup>
                  <pais:associatedDescriptorGroupTypeID>
                     CoRoT-N0-HK-
GROUP↵</pais:associatedDescriptorGroupTypeID>
                  <pais:transferObjectGroupInstanceName>↵
                     FRACTIOPPS1</pais:transferObjectGroupInstanceName>
               </pais:sipTransferObjectGroup>
            </extension>
            <xfdu:contentUnit>
               <extension>
                  <pais:sipDataObject>
                     <pais:associatedDescriptorDataID>↵
                        CoRoT-N0-HK-DATA</pais:associatedDescriptorDataID>
                  </pais:sipDataObject>
               </extension>
               <dataObjectPointer dataObjectID="DO-CoRoT-N0-HK-DATA-0001"/>
```

```
            </xfdu:contentUnit>


            <xfdu:contentUnit>
               <extension>
                  <pais:sipDataObject>
                     <pais:associatedDescriptorDataID>↵
                        CoRoT-N0-HK-Data</pais:associatedDescriptorDataID>
                  </pais:sipDataObject>
               </extension>
               <dataObjectPointer dataObjectID="DO-CoRoT-N0-HK-DATA-0029"/>
            </xfdu:contentUnit>

         </xfdu:contentUnit>
      </xfdu:contentUnit>
   </informationPackageMap>

   <dataObjectSection>

      <dataObject ID="DO-CoRoT-N0-HK-DATA-0001">
         <byteStream size="0">
            <fileLocation locatorType="URL"
               href="file:N0_HK/FRACTIOPPS1/↵

HK_FRACTIOPPS1_P_P_20070101T080503_20070117T235951.fits"/>
            <checksum checksumName="MD5">↵
               d41d8cd98f00b204e9800998ecf8427e</checksum>
         </byteStream>
      </dataObject>


      <dataObject ID="DO-CoRoT-N0-HK-DATA-0029">
         <byteStream size="0">
            <fileLocation locatorType="URL"
               href="file:N0_HK/FRACTIOPPS1/↵

HK_FRACTIOPPS1_P_P_20121001T000004_20121103T235941.fits"/>
            <checksum checksumName="MD5">↵
               d41d8cd98f00b204e9800998ecf8427e</checksum>
         </byteStream>
      </dataObject>

   </dataObjectSection>

</xfdu:XFDU>
```

D7 FIRST SIP OF LEVEL 0 DATASETS

```
<?xml version="1.0" encoding="UTF-8"?>

<xfdu:XFDU xmlns:pais="urn:ccsds:schema:pais:1"
           xmlns:xfdu="urn:ccsds:schema:xfdu:1">

   <packageHeader ID="CoRoT-N0-SIP-0021">
      <volumeInfo>
         <specificationVersion>1.0</specificationVersion>
      </volumeInfo>
      <environmentInfo>
         <extension>
            <pais:sipGlobalInformation>
               <pais:sipID>CoRoT-N0-SIP-0021</pais:sipID>
               <pais:producerSourceID>CNES</pais:producerSourceID>
               <pais:producerArchiveProjectID>CoRoT-N0
                  </pais:producerArchiveProjectID>
               <pais:sipContentTypeID>SIP-CoRoT-N0-RUN
                  </pais:sipContentTypeID>
               <pais:sipSequenceNumber>21</pais:sipSequenceNumber>
            </pais:sipGlobalInformation>
         </extension>
      </environmentInfo>
   </packageHeader>

   <informationPackageMap>

      <xfdu:contentUnit>
         <extension>
            <pais:sipTransferObject>
               <pais:descriptorID>CoRoT-N0-RUN
                  </pais:descriptorID>
               <pais:transferObjectID>CoRoT-N0-RUN-0001
                  </pais:transferObjectID>
               <pais:lastTransferObjectFlag>FALSE
                  </pais:lastTransferObjectFlag>
            </pais:sipTransferObject>
         </extension>

         <xfdu:contentUnit>
            <extension>
               <pais:sipTransferObjectGroup>
                  <pais:associatedDescriptorGroupTypeID>CoRoT-N0-RUN
                     </pais:associatedDescriptorGroupTypeID>
                  <pais:transferObjectGroupInstanceName>RUN03_IRA01
                     </pais:transferObjectGroupInstanceName>
               </pais:sipTransferObjectGroup>
            </extension>

            <xfdu:contentUnit>
               <extension>
                  <pais:sipTransferObjectGroup>
                     <pais:associatedDescriptorGroupTypeID>
                        CoRoT-N0-DATASET-GROUP
                     </pais:associatedDescriptorGroupTypeID>
                     <pais:transferObjectGroupInstanceName>AN0_BKGROUND
                        </pais:transferObjectGroupInstanceName>
```

```
            </pais:sipTransferObjectGroup>
         </extension>

         <xfdu:contentUnit>
            <extension>
               <pais:sipDataObject>
                  <pais:associatedDescriptorDataID>CoRoT-N0-DATASET
                     </pais:associatedDescriptorDataID>
               </pais:sipDataObject>
            </extension>
            <dataObjectPointer
               dataObjectID="DO-CoRoT-N0-DATASET-0001"/>
         </xfdu:contentUnit>

         <xfdu:contentUnit>
            <extension>
               <pais:sipDataObject>
                  <pais:associatedDescriptorDataID>CoRoT-N0-DATASET
                     </pais:associatedDescriptorDataID>
               </pais:sipDataObject>
            </extension>
            <dataObjectPointer
               dataObjectID="DO-CoRoT-N0-DATASET-0002"/>
         </xfdu:contentUnit>

         <xfdu:contentUnit>
            <extension>
               <pais:sipDataObject>
                  <pais:associatedDescriptorDataID>CoRoT-N0-DATASET
                     </pais:associatedDescriptorDataID>
               </pais:sipDataObject>
            </extension>
            <dataObjectPointer
               dataObjectID="DO-CoRoT-N0-DATASET-0003"/>
         </xfdu:contentUnit>

         <xfdu:contentUnit>
            <extension>
               <pais:sipDataObject>
                  <pais:associatedDescriptorDataID>CoRoT-N0-DATASET
                     </pais:associatedDescriptorDataID>
               </pais:sipDataObject>
            </extension>
            <dataObjectPointer
               dataObjectID="DO-CoRoT-N0-DATASET-0004"/>
         </xfdu:contentUnit>

         <xfdu:contentUnit>
            <extension>
               <pais:sipDataObject>
                  <pais:associatedDescriptorDataID>CoRoT-N0-DATASET
                     </pais:associatedDescriptorDataID>
               </pais:sipDataObject>
            </extension>
            <dataObjectPointer
               dataObjectID="DO-CoRoT-N0-DATASET-0005"/>
         </xfdu:contentUnit>
      </xfdu:contentUnit>
```

```
         </xfdu:contentUnit>
      </xfdu:contentUnit>
   </informationPackageMap>

   <dataObjectSection>

      <dataObject ID="DO-CoRoT-N0-DATASET-0001">
         <byteStream size="0">
            <fileLocation locatorType="URL"
               href="file:N0/RUN03_IRA01/AN0_BKGROUND/79.tar.gz"/>
            <checksum

checksumName="MD5">d41d8cd98f00b204e9800998ecf8427e</checksum>
         </byteStream>
      </dataObject>

      <dataObject ID="DO-CoRoT-N0-DATASET-0002">
         <byteStream size="0">
            <fileLocation locatorType="URL"
               href="file:N0/RUN03_IRA01/AN0_BKGROUND/80.tar.gz"/>
            <checksum

checksumName="MD5">d41d8cd98f00b204e9800998ecf8427e</checksum>
         </byteStream>
      </dataObject>

      <dataObject ID="DO-CoRoT-N0-DATASET-0003">
         <byteStream size="0">
            <fileLocation locatorType="URL"
               href="file:N0/RUN03_IRA01/AN0_BKGROUND/81.tar.gz"/>
            <checksum

checksumName="MD5">d41d8cd98f00b204e9800998ecf8427e</checksum>
         </byteStream>
      </dataObject>

      <dataObject ID="DO-CoRoT-N0-DATASET-0004">
         <byteStream size="0">
            <fileLocation locatorType="URL"
               href="file:N0/RUN03_IRA01/AN0_BKGROUND/82.tar.gz"/>
            <checksum

checksumName="MD5">d41d8cd98f00b204e9800998ecf8427e</checksum>
         </byteStream>
      </dataObject>

      <dataObject ID="DO-CoRoT-N0-DATASET-0005">
         <byteStream size="0">
            <fileLocation locatorType="URL"
               href="file:N0/RUN03_IRA01/AN0_BKGROUND/83.tar.gz"/>
            <checksum

checksumName="MD5">d41d8cd98f00b204e9800998ecf8427e</checksum>
         </byteStream>
      </dataObject>

   </dataObjectSection>
</xfdu:XFDU>
```

D8 EXAMPLE OF SIP BUILDER SOFTWARE CONFIGURATION FILE

```
<?xml version="1.0" encoding="UTF-8"?>

<project xmlns="urn:fr:gael:schema:ccsds:pais:sip-builder:1">

   <descriptors>
      <descriptor file="corot-pais-transfer-object-run.xml"/>
      <descriptor file="corot-pais-transfer-object-hk.xml"/>
      <descriptor file="corot-pais-sip-constraints.xml" />
   </descriptors>


   <collectors baseDirectory="../../../test-data/cnes-corot-tds-20140506">

       <!-- N0 Products -->
       <collector typeId="CoRoT-N0-RUN">
           <include>N0/RUN0.*</include>
       </collector>
       <collector typeId="CoRoT-N0-DATASET-GROUP">
           <include>(A|E)N0.*</include>
       </collector>
       <collector typeId="CoRoT-N0-DATASET">
           <include>.*tar.gz</include>
       </collector>

       <!-- N0 HK Data -->
       <collector typeId="CoRoT-N0-HK-GROUP">
           <include>N0_HK/.*</include>
       </collector>
       <collector typeId="CoRoT-N0-HK-DATA">
           <include> .*fits</include>
       </collector>

   </collectors>

</project>
```

ANNEX E

BNF USE CASE

E1 COLLECTION DESCRIPTOR: DIGITIZED NEWSPAPERS

```
<?xml version="1.0" encoding="UTF-8"?>
<pais:collectionDescriptor xmlns:pais="urn:ccsds:schema:pais:1"
xmlns:xsi="http://www.w3.org/2001/XMLSchema">
  <pais:identification>
    <pais:descriptorModelID>CCSD0015</pais:descriptorModelID>
    <pais:descriptorModelVersion>2.0</pais:descriptorModelVersion>
    <pais:descriptorID>digitized_newspapers_1</pais:descriptorID>
  </pais:identification>
  <pais:description>
    <pais:collectionTitle>Digitized newspapers</pais:collectionTitle>
    <pais:collectionDescription>A collection of all the digitized
newspapers</pais:collectionDescription>
  </pais:description>
  <pais:relation>
    <pais:parentCollection>NONE</pais:parentCollection>
  </pais:relation>
</pais:collectionDescriptor>
```

E2 COLLECTION DESCRIPTOR: LE TEMPS (1861–1942)

```
<?xml version="1.0" encoding="UTF-8"?>
<pais:collectionDescriptor xmlns:pais="urn:ccsds:schema:pais:1"
xmlns:xsi="http://www.w3.org/2001/XMLSchema">
  <pais:identification>
    <pais:descriptorModelID>CCSD0015</pais:descriptorModelID>
    <pais:descriptorModelVersion>1.0</pais:descriptorModelVersion>
    <pais:descriptorID>ark:/12148/cb34431794k</pais:descriptorID>
  </pais:identification>
  <pais:description>
    <pais:collectionTitle>Digitized newspaper Le Temps (1861-
1942)</pais:collectionTitle>
    <pais:collectionDescription>A collection of all the digitized issues of
Le Temps</pais:collectionDescription>
  </pais:description>
  <pais:relation>
    <pais:parentCollection>digitized_newspapers_1</pais:parentCollection>
  </pais:relation>
</pais:collectionDescriptor>
```

E3 TRANSFER OBJECT TYPE DESCRIPTOR: PERIODICAL ISSUE

```
<?xml version="1.0" encoding="UTF-8"?>
<pais:transferObjectTypeDescriptor xmlns:pais="urn:ccsds:schema:pais:1"
xmlns:x0="http://www.w3.org/2001/XMLSchema">
  <pais:identification>
    <pais:descriptorModelID>CCSD0014</pais:descriptorModelID>
    <pais:descriptorModelVersion>2.0</pais:descriptorModelVersion>
    <pais:descriptorID>issue</pais:descriptorID>
  </pais:identification>
  <pais:description>
    <pais:transferObjectTypeTitle>issue</pais:transferObjectTypeTitle>
    <pais:transferObjectTypeDescription>a single issue of a
newspaper</pais:transferObjectTypeDescription>
    <pais:transferObjectTypeOccurrence>
      <pais:minOccurrence>1</pais:minOccurrence>
      <pais:maxUnknown /> <!-- newspapers can have any number of issues -->
    </pais:transferObjectTypeOccurrence>
  </pais:description>
  <pais:relation>
    <pais:parentCollection>digitized_newspapers_1</pais:parentCollection>
  </pais:relation>
  <pais:groupType>
    <pais:groupTypeID>master</pais:groupTypeID>
    <pais:groupTypeStructureName>sequence</pais:groupTypeStructureName>
    <pais:groupTypeOccurrence>
      <pais:minOccurrence>1</pais:minOccurrence>
      <pais:maxOccurrence>1</pais:maxOccurrence>
    </pais:groupTypeOccurrence>
    <pais:dataObjectType>
      <pais:dataObjectTypeID>image</pais:dataObjectTypeID>
      <pais:dataObjectTypeDescription>master image of a single page of the
newspaper</pais:dataObjectTypeDescription>
      <pais:dataObjectTypeOccurrence>
        <pais:minOccurrence>1</pais:minOccurrence>
        <pais:maxUnknown />   <!-- no limit on number of master files -->
      </pais:dataObjectTypeOccurrence>
      <pais:dataObjectTypeFormat>
        <pais:mimeType>image/tiff</pais:mimeType>
      </pais:dataObjectTypeFormat>
    </pais:dataObjectType>
  </pais:groupType>
  <pais:groupType>
```

```
    <pais:groupTypeID>ocr</pais:groupTypeID>
    <pais:groupTypeStructureName>sequence</pais:groupTypeStructureName>
    <pais:groupTypeOccurrence>
      <pais:minOccurrence>0</pais:minOccurrence> <!-- ocr is not mandatory -->
      <pais:maxOccurrence>1</pais:maxOccurrence>
    </pais:groupTypeOccurrence>
    <pais:groupTypeAssociation>
        <pais:targetID>master</pais:targetID>
        <pais:relationDescription>
          <pais:relationType>ocerized version of the
issue</pais:relationType>
        </pais:relationDescription>
      </pais:groupTypeAssociation>
    <pais:dataObjectType>
      <pais:dataObjectTypeID>text</pais:dataObjectTypeID>
      <pais:dataObjectTypeDescription>Text originally generated from an OCR
of the master image of a page</pais:dataObjectTypeDescription>
      <pais:dataObjectTypeOccurrence>
        <pais:minOccurrence>1</pais:minOccurrence>
        <pais:maxUnknown />   <!-- no limit on number of OCR files -->
      </pais:dataObjectTypeOccurrence>
      <pais:dataObjectTypeFormat>
        <pais:mimeType>text/xml</pais:mimeType>
      </pais:dataObjectTypeFormat>
    </pais:dataObjectType>
  </pais:groupType>
</pais:transferObjectTypeDescriptor>
```

E4 SIP CONSTRAINTS

```
<?xml version="1.0" encoding="UTF-8"?>
<sipConstraints xmlns="urn:ccsds:schema:pais:1">
      <producerArchiveProjectID>info:bnf/spar/context/fil_num_cons_a</produ
cerArchiveProjectID> <!-- channel identifier -->
      <sipContentType>
            <sipContentTypeID>issue_SIP</sipContentTypeID>
            <authorizedDescriptor>
                  <descriptorID>issue</descriptorID>
                  <occurrence>
                        <minOccurrence>1</minOccurrence>
                        <maxOccurrence>1</maxOccurrence>
                  </occurrence>
            </authorizedDescriptor>
   </sipContentType>
</sipConstraints>
```

E5 SIP MANIFEST

```
<?xml version="1.0" encoding="UTF-8"?>
<mets xmlns="http://www.loc.gov/METS/"
xmlns:premis="info:lc/xmlns/premis-v2"
xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
xmlns:pais="urn:ccsds:schema:pais:1"
xmlns:dc="http://purl.org/dc/elements/1.1/"
xmlns:dcterms="http://purl.org/dc/terms/"
xmlns:spar_dc="http://bibnum.bnf.fr/ns/spar_dc"
xmlns:xlink="http://www.w3.org/1999/xlink"
xmlns:detailsOperation="http://bibnum.bnf.fr/ns/detailsOperation"
xsi:schemaLocation="http://www.loc.gov/METS/
http://www.loc.gov/standards/mets/mets.xsd http://bibnum.bnf.fr/ns/spar_dc
http://bibnum.bnf.fr/ns/spar_dc.xsd">
  <dmdSec ID="DMD.1">
    <mdWrap MIMETYPE="text/xml" MDTYPE="DC">
      <xmlData>
        <spar_dc:spar_dc>
          <dc:title>Le Temps (Paris. 1861)</dc:title>
          <dc:title>Le Temps</dc:title>
          <dc:identifier xsi:type="spar_dc:issn">11501073</dc:identifier>
          <dc:publisher>[s.n.] (Paris)</dc:publisher>
          <dc:date>1861-1942</dc:date>
          <dc:language xsi:type="dcterms:ISO639-3">fre</dc:language>
          <dc:relation
xsi:type="spar_dc:ark">ark:/12148/cb34431794k</dc:relation>
        </spar_dc:spar_dc>
      </xmlData>
    </mdWrap>
  </dmdSec>
  <dmdSec ID="DMD.2">
    <mdWrap MIMETYPE="text/xml" MDTYPE="DC">
      <xmlData>
        <spar_dc:spar_dc>
          <dc:title>1861/04/29 (Numéro 4)</dc:title>
          <dc:description xsi:type="spar_dc:sequentialDesignation1">Numéro
4</dc:description>
          <dc:publisher>[s.n.] (Paris)</dc:publisher>
          <dc:date>1861/04/29</dc:date>
          <dc:type>periodical</dc:type>
          <dcterms:provenance>BnF</dcterms:provenance>
          <dc:language xsi:type="dcterms:ISO639-3">fre</dc:language>
          <dc:relation
xsi:type="spar_dc:ark">ark:/12148/cb34431794k</dc:relation>
        </spar_dc:spar_dc>
```

```
      </xmlData>
    </mdWrap>
  </dmdSec>
  <amdSec>
    <techMD ID="AMD.30">
      <mdWrap MIMETYPE="text/xml" MDTYPE="OTHER" OTHERMDTYPE="PAIS">
        <xmlData>
          <pais:sipTransferObject>
            <pais:descriptorID>issue</pais:descriptorID>
            <pais:transferObjectID>b85f1350-8b50-11e3-909e-
00144f80ca6b</pais:transferObjectID>
          </pais:sipTransferObject>
        </xmlData>
      </mdWrap>
    </techMD>
    <techMD ID="AMD.31">
      <mdWrap MIMETYPE="text/xml" MDTYPE="OTHER" OTHERMDTYPE="PAIS">
        <xmlData>
          <pais:sipTransferObjectGroup>

<pais:associatedDescriptorGroupTypeID>master</pais:associatedDescriptorGrou
pTypeID>

<pais:transferObjectGroupInstanceName>GRP.1</pais:transferObjectGroupInstan
ceName>
          </pais:sipTransferObjectGroup>
        </xmlData>
      </mdWrap>
    </techMD>
    <techMD ID="AMD.32">
      <mdWrap MIMETYPE="text/xml" MDTYPE="OTHER" OTHERMDTYPE="PAIS">
        <xmlData>
          <pais:sipTransferObjectGroup>

<pais:associatedDescriptorGroupTypeID>ocr</pais:associatedDescriptorGroupTy
peID>

<pais:transferObjectGroupInstanceName>GRP.2</pais:transferObjectGroupInstan
ceName>
          </pais:sipTransferObjectGroup>
        </xmlData>
      </mdWrap>
    </techMD>
    <techMD ID="AMD.101">
```

```
      <mdWrap MIMETYPE="text/xml" MDTYPE="OTHER" OTHERMDTYPE="PAIS">
        <xmlData>
        <pais:sipDataObject>

<pais:associatedDescriptorDataID>image</pais:associatedDescriptorDataID>
        </pais:sipDataObject>
        </xmlData>
      </mdWrap>
    </techMD>
    <techMD ID="AMD.102">
      <mdWrap MIMETYPE="text/xml" MDTYPE="OTHER" OTHERMDTYPE="PAIS">
        <xmlData>
        <pais:sipDataObject>

<pais:associatedDescriptorDataID>text</pais:associatedDescriptorDataID>
        </pais:sipDataObject>
        </xmlData>
      </mdWrap>
    </techMD>
    <sourceMD ID="AMD.1">
      <mdWrap MIMETYPE="text/xml" MDTYPE="DC">
        <xmlData>
          <spar_dc:spar_dc>
            <dc:description xml:lang="fr">Papier massicoté en demi-
teinte</dc:description>
          </spar_dc:spar_dc>
        </xmlData>
      </mdWrap>
    </sourceMD>
    <digiprovMD ID="AMD.22">
      <mdWrap MIMETYPE="text/xml" MDTYPE="OTHER" OTHERMDTYPE="PAIS">
        <xmlData>
          <pais:sipGlobalInformation>
            <pais:sipID>NUM_220908_20140201155415</pais:sipID>
            <pais:producerSourceID>NUM</pais:producerSourceID>

<pais:producerArchiveProjectID>info:bnf/spar/context/fil_num_cons_a</pais:p
roducerArchiveProjectID>
            <pais:sipContentTypeID>issue_SIP</pais:sipContentTypeID>
          </pais:sipGlobalInformation>
        </xmlData>
      </mdWrap>
    </digiprovMD>
```

```
<!-- The <digiprovMD MDTYPE="premis:event"> describing the digitization and
ocerization events have been removed -->
    <digiprovMD ID="AMD.2">
      <mdWrap MIMETYPE="text/xml" MDTYPE="PREMIS:EVENT">
        <xmlData>
          <premis:event>
            <premis:eventIdentifier>
              <premis:eventIdentifierType>UUID</premis:eventIdentifierType>
              <premis:eventIdentifierValue>b85f1350-8b50-11e3-909e-
00144f80ca6b</premis:eventIdentifierValue>
            </premis:eventIdentifier>
            <premis:eventType>packageCreation</premis:eventType>
            <premis:eventDateTime>2014-02-
01T15:54:15.814+01:00</premis:eventDateTime>
            <premis:eventDetail>Création d'un paquet compatible avec
SPAR</premis:eventDetail>
            <premis:linkingAgentIdentifier>

<premis:linkingAgentIdentifierType>BnFApplication</premis:linkingAgentIdent
ifierType>

<premis:linkingAgentIdentifierValue>info:bnf/spar/agent/preingest_fil_num_c
ons_a_act_52</premis:linkingAgentIdentifierValue>
              <premis:linkingAgentRole>performer</premis:linkingAgentRole>
            </premis:linkingAgentIdentifier>
            <premis:linkingAgentIdentifier>

<premis:linkingAgentIdentifierType>producerIdentifier</premis:linkingAgentI
dentifierType>

<premis:linkingAgentIdentifierValue>NUM</premis:linkingAgentIdentifierValue
>
              <premis:linkingAgentRole>issuer</premis:linkingAgentRole>
            </premis:linkingAgentIdentifier>
            <premis:linkingAgentIdentifier>

<premis:linkingAgentIdentifierType>channelIdentifier</premis:linkingAgentId
entifierType>

<premis:linkingAgentIdentifierValue>info:bnf/spar/context/fil_num_cons_a</p
remis:linkingAgentIdentifierValue>
              <premis:linkingAgentRole>authorizer</premis:linkingAgentRole>
            </premis:linkingAgentIdentifier>
            <premis:linkingObjectIdentifier>

<premis:linkingObjectIdentifierType>productionIdentifier</premis:linkingObj
ectIdentifierType>
```

```
<premis:linkingObjectIdentifierValue>220908</premis:linkingObjectIdentifier
Value>
            </premis:linkingObjectIdentifier>
          </premis:event>
        </xmlData>
      </mdWrap>
    </digiprovMD>
  </amdSec>
  <fileSec>
    <fileGrp USE="master" ID="GRP.1" ADMID="AMD.31">
      <file CHECKSUMTYPE="MD5" CHECKSUM="e46b1e07dd8994be8766494814ed2a9a"
ADMID="AMD.1 AMD.3 AMD.101" ID="master.1">
        <FLocat xlink:type="simple" LOCTYPE="URL"
xlink:href="master/T0000001.tif"/>
      </file>
      <file CHECKSUMTYPE="MD5" CHECKSUM="5345dfe7240e882396f3b3bd01e5cca2"
ADMID="AMD.1 AMD.3 AMD.101" ID="master.2">
        <FLocat xlink:type="simple" LOCTYPE="URL"
xlink:href="master/T0000002.tif"/>
      </file>
      <file CHECKSUMTYPE="MD5" CHECKSUM="8135b55cf1f598a5bdfa44055c55cc4a"
ADMID="AMD.1 AMD.3 AMD.101" ID="master.3">
        <FLocat xlink:type="simple" LOCTYPE="URL"
xlink:href="master/T0000003.tif"/>
      </file>
      <file CHECKSUMTYPE="MD5" CHECKSUM="c7d8f4cb6152834c66ffddb35679000e"
ADMID="AMD.1 AMD.3 AMD.101" ID="master.4">
        <FLocat xlink:type="simple" LOCTYPE="URL"
xlink:href="master/T0000004.tif"/>
      </file>
    </fileGrp>
    <fileGrp USE="ocr" ID="GRP.2" ADMID="AMD.32">
      <file CHECKSUMTYPE="MD5" CHECKSUM="c6d2997b8d46c17bf87e8a5e7c082342"
ADMID="AMD.4 AMD.102" ID="ocr.1">
        <FLocat xlink:type="simple" LOCTYPE="URL"
xlink:href="ocr/X0000001.XML"/>
      </file>
      <file CHECKSUMTYPE="MD5" CHECKSUM="917ccf7eeaacd42c648453f12d842c38"
ADMID="AMD.4 AMD.102" ID="ocr.2">
        <FLocat xlink:type="simple" LOCTYPE="URL"
xlink:href="ocr/X0000002.XML"/>
      </file>
      <file CHECKSUMTYPE="MD5" CHECKSUM="61af865fdb34a5cc2bea93369b5f859e"
ADMID="AMD.4 AMD.102" ID="ocr.3">
```

```
        <FLocat xlink:type="simple" LOCTYPE="URL"
xlink:href="ocr/X0000003.XML"/>
      </file>
      <file CHECKSUMTYPE="MD5" CHECKSUM="8392d60f3592348fca3826ada32ab855"
ADMID="AMD.4 AMD.102" ID="ocr.4">
        <FLocat xlink:type="simple" LOCTYPE="URL"
xlink:href="ocr/X0000004.XML"/>
      </file>
    </fileGrp>
  </fileSec>
  <structMap TYPE="physical">
    <div TYPE="set" ID="DIV.1" DMDID="DMD.1">
      <div TYPE="group" ID="DIV.2" DMDID="DMD.2" ADMID="AMD.2 AMD.22 AMD.30
AMD.5">
        <div TYPE="object" ORDERLABEL="NP" ORDER="1" ID="DIV.3">
          <fptr FILEID="master.1"/>
          <fptr FILEID="ocr.1"/>
        </div>
        <div TYPE="object" ORDERLABEL="NP" ORDER="2" ID="DIV.4">
          <fptr FILEID="master.2"/>
          <fptr FILEID="ocr.2"/>
        </div>
        <div TYPE="object" ORDERLABEL="NP" ORDER="3" ID="DIV.5">
          <fptr FILEID="master.3"/>
          <fptr FILEID="ocr.3"/>
        </div>
        <div TYPE="object" ORDERLABEL="NP" ORDER="4" ID="DIV.6">
          <fptr FILEID="master.4"/>
          <fptr FILEID="ocr.4"/>
        </div>
      </div>
    </div>
  </structMap>
</mets>
```

E6 SIP MAPPING TO XFDU AND METS

NOTES

1 This proposal mixes generic METS elements with BnF-specific structure vocabulary. A view of this structure is provided in 6.5.2.4.2.

2 The PAIS schema for METS is the same as the PAIS schema for XFDU in every instance where METS uses a PAIS element.

PAIS Abstract SIP Concepts	**Abstract SIP Concept Definition**	**XFDU Mechanism to Instantiate Concept**	**Possible SPAR mechanism to instantiate concept**
SIP	Transferable package able to hold any type of information	xml manifest, typically within a ZIP file, incorporating PAIS specified schema defined elements for XFDU	XML manifest, incorporating METS and PAIS specified schema defined elements
Container	Able to hold attributes and other containers; has a mandatory or optional status	XFDU Content Unit	METS <amdSec> (<digiprovMD> or <techMD>) linked to the corresponding
Attribute	Able to hold a value; has a mandatory or optional status	XML elements defined in PAIS and incorporated into XFDU using XFDU extension elements; also some native XFDU elements	XML elements defined in PAIS and incorporated into METS using METS extension <mdWrap><xmlData>; also some native METS elements
PAIS Abstract SIP Structure	**Structure Definitions**	**XFDU Mechanism to Instantiate Structure**	**Possible SPAR mechanism to instantiate structure**
SIP Global Information container	SIP structure holding a set of attributes supporting the unique identification of each SIP within the Producer-Archive Project and the ability to optionally track the sequencing of SIPs	XFDU manifest element <environmentInfo> using XFDU extension element	METS manifest element <digiprovMD> linked to the BnF-specific "group" level in the <structMap> element using <mdWrap><xmlData> element.
- SIP ID (1,,1)	Identifier of the SIP within the Archive Project	PAIS schema specified element for XFDU	PAIS schema specified element for XFDU

PAIS Abstract SIP Concepts	Abstract SIP Concept Definition	XFDU Mechanism to Instantiate Concept	Possible SPAR mechanism to instantiate concept
- Producer-Archive Project ID (1,,1)	identifier of the Producer-Archive Project within the Archive	PAIS schema specified element for XFDU	PREMIS element <premis:linkingAgentIdentifier>
- Producer Source ID (1,,1)	identifier of the Producer Source for contact purposes	PAIS schema specified element for XFDU	PREMIS element <premis:linkingAgentIdentifier>
- SIP Content Type ID (1,,1)	Identifier of the specification as to which Transfer Object Types are allowed in this type of SIP	PAIS schema specified element for XFDU	PAIS schema specified element for METS
- SIP Sequence Number (0,,1)	Number indicating the order in which this SIP has been sent	PAIS schema specified element for XFDU	PAIS schema specified element for METS
- ANY (0,,N)	mechanism that allows a SIP to have any additional attributes within this container	PAIS schema specified element for XFDU	PAIS schema specified element for METS
Transfer Object to Delete container (0,,N)	SIP structure holding one or more attributes giving the identification of previously sent Transfer Objects that needs to be deleted by the Archive.	Optional first XFDU Content Unit, with extension for PAIS/XFDU specified elements	premis:event of "deletionRequest" type
- Transfer Object to Delete (1,,N)	Identifier of a previously sent Transfer Object that needs to be deleted	PAIS schema specified element for XFDU	/
- ANY (0,,1)	mechanism that allows a SIP to have any additional attributes within this container	PAIS schema specified element for XFDU	/

PAIS Abstract SIP Concepts	**Abstract SIP Concept Definition**	**XFDU Mechanism to Instantiate Concept**	**Possible SPAR mechanism to instantiate concept**
Transfer Object container (0,,N)	SIP Structure that holds two types of containers: one Transfer Object Identification and Status container and one or more Transfer Object Group containers.	First, or second, XFDU Content Unit with extension for PAIS/XFDU specified elements	METS element <div> of BnF-specific "group" type
- Transfer Object Identification and Status container (1,,1)	SIP structure that holds a set of attributes supporting unique identification and replacement status information about this Transfer Object	XFDU Content Unit nested within the Content Unit for the Transfer Object, with extension for PAIS/XFDU specified elements	METS manifest element <techMD> linked to the BnF-specific "group" level <div> in the <structMap> element using <mdWrap><xmlData> element.
- Descriptor ID (1,,1)	Identifier of the Transfer Object Type Descriptor that describes this type of Transfer Object	PAIS schema specified element for XFDU	PAIS schema specified element for METS
- Transfer Object ID (1,,1)	Identifier of each delivered Transfer Object within the Producer-Archive Project.	PAIS schema specified element for XFDU	PAIS schema specified element for METS
- Last Transfer Object Flag (0,,1)	Indicator specifying that this Transfer Object is the last one of its type (within the Descriptor) being delivered by the Producer Source.	PAIS schema specified element for XFDU	PAIS schema specified element for METS
- Replacement Transfer Object ID (0,,1)	Identifier of a previously sent Transfer Object that is to be replaced by this Transfer Object	PAIS schema specified element for XFDU	Replacement request
- ANY (0,,N)	mechanism that allows a SIP to have any additional attributes within this container	PAIS schema specified element for XFDU	PAIS schema specified element for METS

CESG APPROVAL COPY - NOT FOR DISTRIBUTION

PAIS Abstract SIP Concepts	Abstract SIP Concept Definition	XFDU Mechanism to Instantiate Concept	Possible SPAR mechanism to instantiate concept
- Transfer Object Group container (1,,N)	SIP structure that holds three types of containers;	XFDU Content Unit nested within the Content Unit for the Transfer Object, with extension for PAIS/XFDU specified elements	Nested s of BnF-specific "object" type
- Transfer Object Group Identification container (1,,1)		XFDU Content Unit nested within the XFDU Content Unit for the Transfer Object Group, with extension for PAIS/XFDU specified elements	METS manifest elements <techMD> linked to the METS element <flieGrp>
- Associated Descriptor Group Type ID (1,,1)	Identifier of the associated group description within the associated Descriptor	PAIS schema specified element for XFDU	PAIS schema specified element for METS
- Choice of one of the following two attributes (0,,1):		PAIS schema specified element for XFDU	PAIS schema specified element for METS
- Transfer Object Group Instance Name (1,,1)	Name given to the group such as a directory name	PAIS schema specified element for XFDU	PAIS schema specified element for METS
- Transfer Object Group Preservation Name (1,,1)	Name given to the group, such as a directory name, that is to be preserved by the Archive	PAIS schema specified element for XFDU	PAIS schema specified element for METS
- ANY (0,,N)	mechanism that allows a SIP to have any additional attributes within this container	PAIS schema specified element for XFDU	PAIS schema specified element for METS
- Transfer Object Group container (0,,N)	SIP structure that holds any number of additional Transfer Object Group containers	XFDU Content Unit nested within the Content Unit for the Transfer Object Group, with extension for PAIS/XFDU specified attributes	Nested s within the s corresponding to the Transfer Object group containers
- Data Object container (0,,N)	SIP structure that conceptually holds two or more containers: Data Object Identification container and one or more Byte Stream containers	XFDU Content Unit nested within the Content Unit for the Transfer Object Group, with extension for PAIS/XFDU specified attributes	METS element <file>s

PAIS Abstract SIP Concepts	**Abstract SIP Concept Definition**	**XFDU Mechanism to Instantiate Concept**	**Possible SPAR mechanism to instantiate concept**
- Data Object Identification container (1,,1)	SIP structure that holds a set of attributes identifying the type of Data Object	XFDU Content Unit nested within the Content Unit for the Data Object, with extension for PAIS/XFDU specified attributes	METS element <techMD> linked to the <file>s
- Associated Descriptor Data ID (1,,1)	Identifier of the associated data description within the associated Descriptor	PAIS schema specified element for XFDU	PAIS schema specified element for METS
- Data Object Preservation Name (0,,1)	Name to be preserved in association with the Data Object instance	PAIS schema specified element for XFDU	PAIS schema specified element for METS
- ANY (0,,N)	mechanism that allows a SIP to have any additional attributes within this container	PAIS schema specified element for XFDU	PAIS schema specified element for METS
- Byte Stream container (1,,N)	SIP structure that holds a set of attributes that provide a byte steam and/or a pointer to a byte stream outside the SIP	XFDU Data Object Pointer Element nested within the XFDU Content Unit for the Data Object	METS element <stream> nested within a <file> element corresponding to a Data Object
- Byte Stream (0,,1)	Stream of bytes	Uses XFDU <byteStream> elements with file location and/or file content elements	Uses METS < stream> elements with file location and linked technical metadata
- Byte Stream Checksum (0,,1)	Checksum covering the stream of bytes	Uses XFDU <byteStream> elements with file checksum element	Uses METS <stream> elements with <file> checksum
- Pointer to Byte Stream (0,,1)	A pointer to a byte stream outside the SIP	Uses XFDU <byteStream> elements with file location element	Uses METS <stream> elements with <file> location
- ANY (0,,N)	mechanism that allows a SIP to have any additional attributes within this container	PAIS schema specified element for XFDU	PAIS schema specified element for METS

CESG APPROVAL COPY - NOT FOR DISTRIBUTION