

1 Preface

1.1 Aim of the specification

This E-ARK specification is part of a family of specifications that provide a common set of requirements for packaging digital information. These specifications are based on common, international standards for transmitting, describing and preserving digital data. They have been produced to help data creators, software developers and digital archives tackle the challenge of short-, medium- and long-term data management and reuse in a sustainable, authentic, cost-efficient, manageable and interoperable way.

The foundation for these specifications is the Reference Model for an Open Archival Information System (OAIS) which has Information Packages at its core. Familiarity with the core functional entities of OAIS is a prerequisite for understanding the specifications. A visualisation of the current specification network can be seen here:

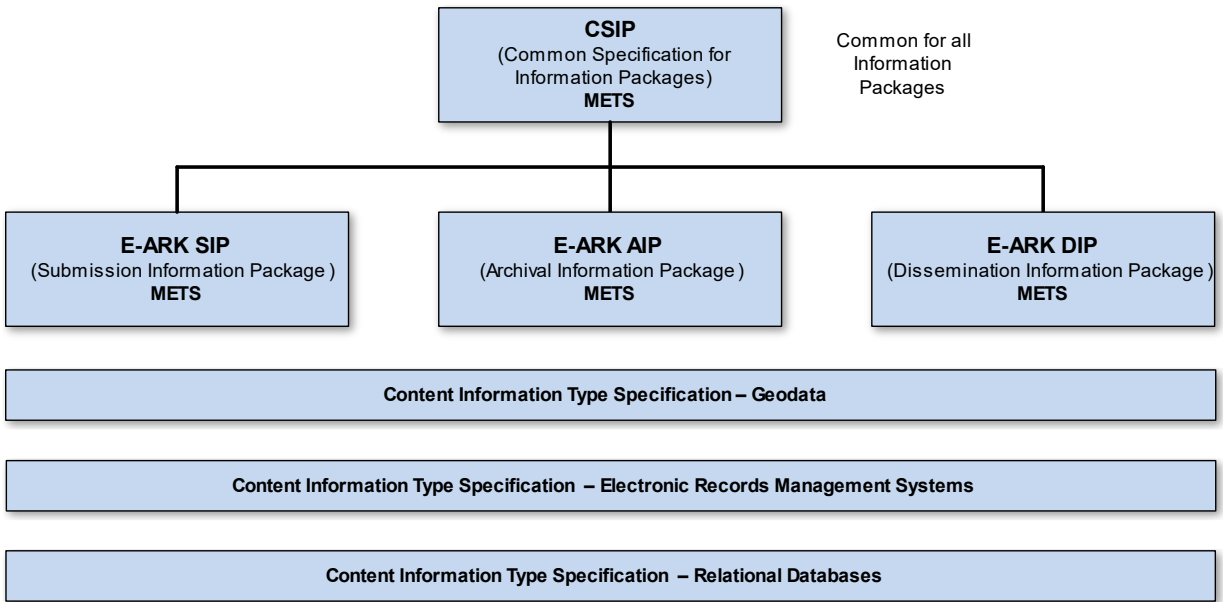


Figure 1: E-ARK specification dependency hierarchy

Table 1: Specification hierarchy aims and goals

Specification	Aim and Goals
Common Specification for Information Packages	<div>This document introduces the concept of a Common Specification for Information Packages (CSIP). Its three main purposes are to:</div> <ul style="list-style-type: none">Establish a common understanding of the requirements which need to be met in order to achieve interoperability of Information Packages.Establish a common base for the development of more specific Information Package definitions and tools within the digital

Specification	Aim and Goals
	<p>preservation community.</p> <ul style="list-style-type: none"> Propose the details of an XML-based implementation of the requirements using, to the largest possible extent, standards which are widely used in international digital preservation. <p>Ultimately the goal of the Common Specification is to reach a level of interoperability between all Information Packages so that tools implementing the Common Specification can be adopted by institutions without the need for further modifications or adaptations.</p>
E-ARK SIP	<p>The main aims of this specification are to:</p> <ul style="list-style-type: none"> Define a general structure for a Submission Information Package format suitable for a wide variety of archival scenarios, e.g. document and image collections, databases or geographical data. Enhance interoperability between Producers and Archives. Recommend best practices regarding metadata, content and structure of Submission Information Packages.
E-ARK AIP	<p>The main aims of this specification are to:</p> <ul style="list-style-type: none"> Define a generic structure of the AIP format suitable for a wide variety of data types, such as document and image collections, archival records, databases or geographical data. Recommend a set of metadata related to the structural and the preservation aspects of the AIP as implemented by the reference implementation eArchiving ToolBox (formerly earkweb). Ensure the format is suitable to store large quantities of data.
E-ARK DIP	<p>The main aims of this specification are to:</p> <ul style="list-style-type: none"> Define a generic structure of the DIP format suitable for a wide variety of archival records, such as document and image collections, databases or geographical data. Recommend a set of metadata related to the structural and access aspects of the DIP.
Content Information Type Specifications	<p>The main aim and goal of a Content Information Type Specification is to:</p> <ul style="list-style-type: none"> Define, in technical terms, how data and metadata must be formatted and placed within a CSIP Information Package in order to achieve interoperability in exchanging specific Content Information. <p>The number of possible Content Information Type Specifications is unlimited.</p>

1.2 Organisational support

This specification is maintained by the Digital Information LifeCycle Interoperability Standards Board (DILCIS Board). The DILCIS Board (<http://dilcis.eu/>) was created to enhance and

maintain the draft specifications developed in the European Archival Records and Knowledge Preservation Project (E-ARK project) which concluded in January 2017 (<http://eark-project.com/>). The Board consists of eight members, but there is no restriction on the number of participants in the work. All Board documents and specifications are stored in GitHub (<https://github.com/DILCISBoard>) while published versions are made available on the Board webpage. Since 2018 the DILCIS Board has been responsible for the core specifications in the Connecting Europe Facility eArchiving Building Block (<https://ec.europa.eu/cefdigital/wiki/display/CEFDIGITAL/eArchiving>).

1.3 Authors

A full list of contributors to this specification, as well as the revision history can be found in Appendix 1.

TABLE OF CONTENT

1	Preface	2
1.1	Aim of the specification.....	2
1.2	Organisational support.....	3
1.3	Authors	4
2	Introduction	7
2.1	Purpose.....	7
2.2	Scope	7
2.3	Layered Data Model	7
3	Specification	9
3.1	Requirements Structure	12
3.2	Principles	12
3.2.1	Principle – support for LOTAR conformance	12
3.2.2	Principle – use of PREMIS	12
3.3	Use cases for archiving of 3D product model data	14
3.4	LOTAR	15
3.4.1	Long-term archiving.....	15
3.4.2	Data Model	16
3.4.3	The Core Model	16
3.4.4	Metadata	16
3.4.5	Digital and Engineering Signatures.....	17
3.4.6	Validation.....	17
3.4.7	Verification	18
4	Implementation.....	19
4.1	Standards.....	19
4.1.1	LOTAR	19
4.1.2	PREMIS.....	19
4.1.3	METS	21
4.2	Information Packages.....	21
4.3	Package Structure.....	24
4.4	Descriptive Metadata	25
4.5	Preservation Metadata.....	25
5	Glossary.....	26
6	Postface.....	33

LIST OF TABLES

Table 1: Glossary 26

LIST OF FIGURES

Figure 1: E-ARK specification dependency hierarchy 2

Figure 2: Data Model Structure 8

Figure 3: CITS 3DPM Layered Data Model..... 9

Figure 4: Example Information Package Folder Structure 24

2 Introduction

2.1 Purpose

The purpose of this document is to accompany the CITS 3D Product Model Data Specification and to provide context and rationale for the principles and requirements of the specification. The specification is designed to be used for the transfer to archives as well as for records exchange between different 3D Product Information Model systems. The specification is supported by METS profiles for the Root and Representation METS files and this accompanying Guideline document.

2.2 Scope

Use of 3D data is widespread across many domains, with a plethora of applications and data formats. This 3D Product Model content specification limits its scope to the area of 3D digital product data such as computer aided design (CAD) or product data model (PDM) data. There is an international standard for the long term archiving of this class of data in the LOTAR “Long Term Archiving and Retrieval of digital technical product information”¹, which is published as the EN and NAS 9300 series. However although LOTAR extensively references and extends ISO 14721 the “Open reference model for Archiving Information System”, (OAIS) it does not extend into areas detailed in the E-ARK common specification for information packages (CSIP). LOTAR also references and builds on ISO 10303, the Standard for the Exchange of Product model data (STEP) and so with this E-ARK 3DPM CITS we have the opportunity to add to a layered standards model as seen below in section 2.3.

2.3 Layered Data Model

This section introduces the role of the CITS 3DPM and its dependencies on the basic structures of the Information Package.

The specification is created based on the requirements of the Common Specification for Information Packages (CSIP), the specification for Submission Information Packages (E-ARK SIP) and the specification for Archival Information Packages (E-ARK AIP). To fully understand its requirements, we highly recommend that users review the requirements and the terminology of the source documents, before using the specification.

The data model structure is based on a layered approach for information package definitions (Figure 2). The Common Specification for Information Packages (CSIP) forms the outermost layer. The general SIP, AIP and DIP specifications add respectively, submission, archiving and dissemination information to the CSIP specification. The third layer of the model represents specific content information type specifications, such as this 3DPM specification. Additional layers for business-specific specifications and local variant implementations of any specification can be added to suit the needs of the organisation.

¹ <https://lotar-international.org/lotar-standard/>

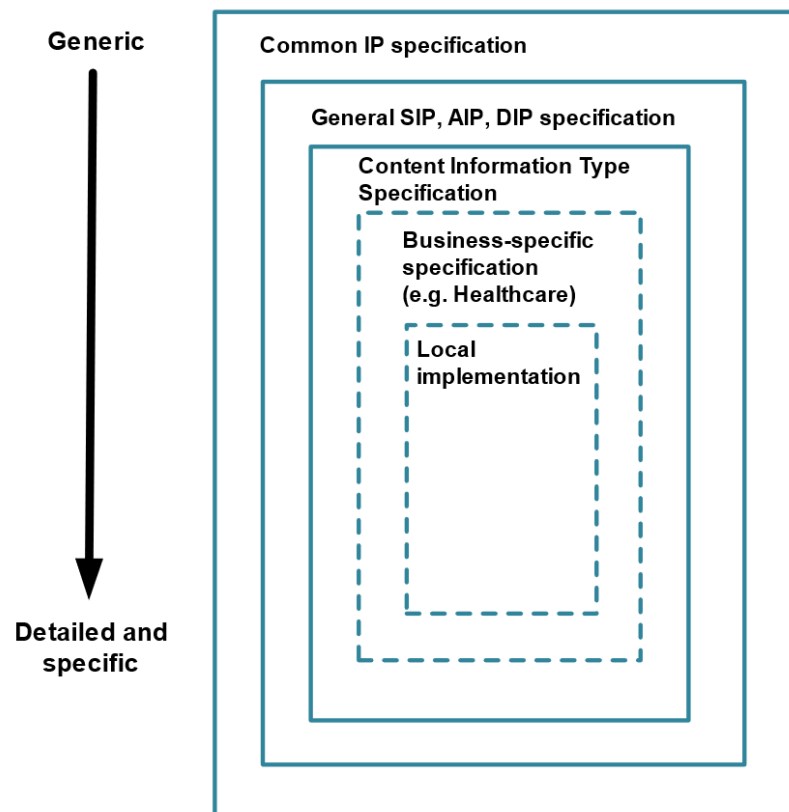


Figure 2: Data Model Structure

Every level in the data model structure inherits metadata entities and elements from the higher levels. In order to increase adoption, a flexible schema has been developed. This will allow for extension points where the schema in each layer can be extended to accommodate additional information on the next specific layer until, finally, the local implementation can add specific entities or metadata elements to satisfy specific local needs. Extension points can be implemented by:

- Embedding foreign extension schemas (in the same way as supported by METS [<http://www.loc.gov/standards/mets/>] and PREMIS [<http://www.loc.gov/standards/premis/>]). These both support increasing the granularity of existing metadata elements by using more detailed data structures as well as adding new types of metadata.
- Substituting metadata schemas for standards more appropriate for the local implementation.

The structure allows the addition of more detailed requirements for metadata entities, for example, by:

- Increasing the granularity of metadata elements by using more detailed data structures, or
- Adding local controlled vocabularies.

For consistency, design principles are reused between layers as much as possible.

The CITS 3DPM builds on the existing standard LOTAR standard for ‘long-term archiving of digital technical product information’ which in itself builds on the standard for an Open Archival Information System (OAIS, ISO 14721) and the Standard for the Exchange of Product Model Data (STEP, ISO 15926). So for CITS 3DPM in particular we have a layered data model as seen in Figure 3. Note however that compliance with LOTAR or STEP is not mandatory within 3DPM but is recommended. Individual organisational archiving strategies for 3D Product Model data may or may not include STEP Representations of model data and may include all or some of the elements of the LOTAR standard. The 3DPM CITS provides for accommodation of these standards but makes compliance an organisational choice.

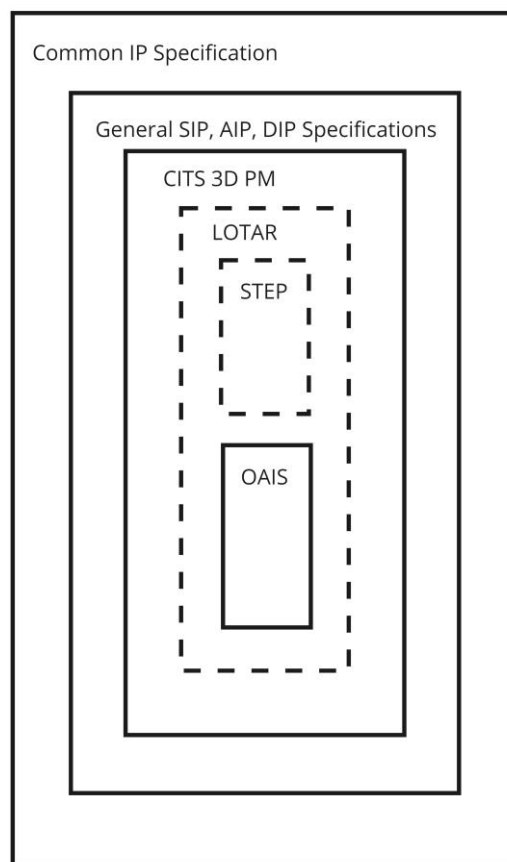


Figure 3: CITS 3DPM Layered Data Model

2.4 File Formats

3D product model file formats contain data representing the model and the necessary information for its display and editing plus contextual metadata to aid interpretation and usability. Most CAD software and hence file formats are proprietary, but open, neutral formats exist and can be produced from proprietary formats by direct translation or direct export from the authoring software.

Use of file formats that offer the best long-term prospects in terms of usability, accessibility, and sustainability is recommended and is generally achieved with formats that have wide usage, have

open specifications, and are independent of specific software or developers. Neutral and open file formats are available, and there may be options to convert proprietary file formats to them; but there is risk which should be assessed for information loss in the transformation. Keeping original, proprietary formats is not ideal, but with the risk of loss of information in transforming file formats, the inclusion of original and format derivations in archival packages is recommended.

The following are neutral product model file formats, used primarily for data exchange²:

DXF (Drawing eXchange Format)

Developed by Autodesk in 1982 as their data interoperability solution between AutoCAD and other CAD systems. The DXF is primarily 2D-based and its format is a tagged data representation of all the information contained in an AutoCAD drawing file, which means that each data element in the file is preceded by an integer number that is called a group code indicating the type of following data element. As most commercial application software developers have chosen to support Autodesk's native DWG as the format for AutoCAD data interoperability, DXF has become less useful.

VDA-FS (Verband der Automobilindustrie – Flächenschnittstelle)

Created by the German Association of the Automotive Industry in 1982 as an interoperability method for free-form surfaces. This format differs from other formats in that it only supports the communication of free-form curve and surface data with associated comments, but no other geometric or non-geometric entities. Therefore, it is limited to representations by parametric polynomials, but this covers the great majority of free-form CAD systems. It includes Bézier, B-Spline and Coons tensor product types of surfaces and corresponding curves. The VDA-FS specification is released in the German Industrial Standard DIN 66301.

PDES (Product Data Exchange Specification)

Originated in 1988 under the Product Definition Data Interface (PDDI) study done by McDonnell Aircraft Corporation on behalf of the U.S. Air Force. PDES was designed to completely define a product for all applications over its expected life cycle, including geometry, topology, tolerances, relationships, attributes, and features necessary to completely define a part or assembly of parts. PDES can be viewed as an expansion of IGES where organizational and technological data have been added. In fact, the later PDES contained IGES. The development of PDES under the guidance of the IGES organization and in close collaboration with the International Organization for Standardization (ISO) led to the birth of STEP.

STEP (ISO 10303 – Standard for the Exchange of Product model data)

The work with the ISO 10303 standard was initiated in 1984 and initially published in 1994, with the objective to standardize the exchange of product data between Product Lifecycle Management (PLM) systems. It is a very comprehensive set of specifications covering many different product types and many life cycle phases. STEP uses the neutral ISO 10303-11 format, also known as an EXPRESS schema. EXPRESS defines not only the data types but also relations and rules applying to them. STEP supports data exchange, data sharing and data archiving. For data

² https://en.wikipedia.org/wiki/CAD_data_exchange

exchange, STEP defines the transitory form of the product data that is to be transferred between a pair of applications. It supports data sharing by providing access to and operation on a single copy of the same product data by more than one application, potentially simultaneously. STEP may also be used to support the development of the archive product data itself. STEP consists of several hundred documents called parts. Every year new parts are added or new revisions of older parts are released. This makes STEP the biggest standard within ISO. The 200-series parts STEP are called Application Protocols (AP), with the specific parts directly related to CAD systems:

- 203 (Configuration controlled 3D designs of mechanical parts and assemblies) – Mainly used for 3D design and product structure. A subset of AP214 but most widely used.
- 210 (Electronic assembly, interconnect and packaging design) – CAD systems for printed circuit board.
- 212 (Electrotechnical design and installation) – CAD systems for electrical installation and cable harness.
- 214 (Core data for automotive mechanical design processes) – How STEP is represented in a text file for interchange.
- 238 (STEP-NC Application interpreted model for computerized numerical controllers) – CAD, CAM, and CNC machining process information.
- 242 (Managed model based 3D engineering) – the merging of the two leading STEP application protocols, AP 203 and AP 214.

Parasolid XT

Part of the Parasolid geometric modeling kernel originally developed by Shape Data and currently owned by Siemens Digital Industries Software. Parasolid can represent wireframe, surface, solid, cellular and general non-manifold models. It stores topological and geometric information defining the shape of models in transmitting files. These files have a published format so that applications can have access to Parasolid models without necessarily using the Parasolid kernel. Parasolid is capable of accepting data from other modeler formats. Its unique tolerant modeling functionality can accommodate and compensate for less accurate data.

IGES (Initial Graphics Exchange Specification)

An outdated format originated in late 1979 and initially published by the American National Standards Institute (ANSI) in 1980 preceding the large-scale deployment of the CAD technology in the industry. This file format considers the product definition as a file of entities, with each entity being represented in an application-independent format. After the initial release of STEP (ISO 10303) in 1994, interest in further development of IGES declined, and Version 5.3 (1996) was the last published standard.

3 Specification

3.1 Requirements Structure

The Content Information Type Specification for 3D Product Model Data aims to define the necessary elements required to preserve the accessibility and authenticity of 3D Product Model Data over time and across changing technical environments. The specification builds on the international standard for long-term archiving of Product Model Data (LOTAR) and in order to achieve this, elevates the level (and adjusts the cardinality) of some of the requirements set out in the Common Specification (CSIP) and package specifications (namely SIP and AIP) and adds new requirements for the package structure, descriptive metadata, preservation metadata and accompanying METS files. It also introduces new requirements for authentication and Verification. The specification sets out general principles that underpin the specific requirements and further context for the requirements and justification for the principles follow in this accompanying guideline.

3.2 Principles

3.2.1 Principle – support for LOTAR conformance

The LOTAR “Long Term Archiving of digital technical product information” series is an international standard for the long term archiving of Product Model data (such as computer aided design CAD or product data model PDM data). LOTAR extensively references and extends ISO 14721 the “Open reference model for Archiving Information System”, (OAIS) but does not extend into areas detailed in the E-ARK common specification for information packages (CSIP). LOTAR also references and builds on ISO 10303, the Standard for the Exchange of Product Model data (STEP). This eArchiving 3D Product Model CITS creates a layered model for creating archival packages for Product Models that allows conformance to LOTAR and STEP whilst maintaining conformance with the CSIP and with the individual eArchiving package specifications (SIP, AIP and DIP). Specifically:

- no requirements in this specification should conflict with mandatory requirements of LOTAR;
- requirements of LOTAR with regard to essential data or metadata elements in Information Packages become optional should requirements in the 3DPM CITS;
- the scope of the E-ARK specifications are not altered to encompass areas covered by LOTAR but not covered by E-ARK, for example process requirements and management procedures.

A conformant E-ARK 3DPM Information Package will not imply conformance or validation against LOTAR, but an archive will be able to use the E-ARK 3DPM CITS together with the other E-ARK package specifications to produce Information Packages that can support LOTAR compliance.

3.2.2 Principle – use of PREMIS

From the CSIP and PREMIS CITS:

- PREMIS should be used to record detailed technical metadata;
- Technical information should be included in PREMIS metadata by using extension schemas;

- Information about agents carrying out preservation actions must be recorded in the PREMIS metadata (this is because METS agents describe agents relevant for generic IP level events, such as the creation or submission of the package, not the preservation of the data);
- Event descriptions should be included in PREMIS metadata. Use of the official PREMIS event vocabulary (<https://id.loc.gov/vocabulary/preservation/eventType.html>) is recommended;
- Detailed rights information should be included in PREMIS. High-level rights information in METS indicates restrictions. Detailed, object-specific rights information will be included in the PREMIS metadata;
- File format information for all files should be included as Persistent Unique Identifier (PUID) values in the appropriate PREMIS semantic units.

Technical and preservation metadata in the context of the 3DPM CITS can include:

- Creating agent;
- Reference to the content information standard (e.g. STEP);
- Reference to the LOTAR standard part for the content information type;
- Information about the generating system.

Event descriptions include:

- Creation events;
- Conversion or change events;
- Electronic signature events;
- Verification events and results;
- Validation events and results.

Detailed technical metadata in the context of the 3DPM CITS include:

- File format, characterisation, checksums;
- Detailed part number, version, product model and issue information;
- Relationships for the digital object (is part of, contains parts).

Rights information includes:

- Access rights;
- Export controls;
- License restrictions;
- Copyright owner;
- Security classification;

- Personal identifiable information restrictions;
- Company specific classifications.

Use of PREMIS must conform to the requirements of the Preservation Metadata CITS.

3.3 Use cases for archiving of 3D product model data

According to the Common Specification: “Regardless of the formats and systems in question, it is necessary to consider whether the information needs to be retained and managed for longer periods of time.” The reasons for this might be, for example:

- to meet legal and regulatory obligations;
- to provide for efficient reuse;
- to satisfy historical, cultural, scientific and business interest.

In LOTAR the objectives for keeping data for long-term are clearly distinguished into two major categories³:

- Legal requirements/certification requirements, such as for proof of technical documentation for actions in law;
- Business requirements, such as keeping knowledge.

Within the two categories LOTAR offers four characteristics which describe the objectives in more detail:

- to preserve the *original* data (generated by a source system) so that it can be used as evidence of what the data was at a particular date. This characteristic fits with the subcategory ‘legal requirement’;
- to keep *data available* to new users over the period for which it is kept. This characteristic fits with the subcategories ‘legal requirement’ and ‘business requirement’;
- to be able to *preserve* the *source* of the kept data. This characteristic fits with the subcategory ‘business requirement’;
- to be able to *reuse* the data, for example, by modifying design data to meet new requirements. This characteristic fits with the subcategory ‘business requirement’.

The two major categories for keeping data stated by LOTAR imply a private (as opposed to public) archive which is under the control of the organisation which hopes to benefit from the value of the content in the future (business requirements) or to manage future risk with respect to the data (legal/certification requirements). Requirements for interoperability in this situation will be

³ EN 9300 Part 003, 4.1.3

intra-organisational rather than inter-organisational, where there may over time be exchange or movement of data between locations and as a result of changes in organisation structure.

The use cases considered within the specification are therefore:

1. To enable the submission of 3D Product Model data from engineering departments in an organisation to a centralised or distributed archive, in a common format;
2. To store archival 3D Product Model data in a manner that will allow consolidation of archives intra-organisationally or with sources added through mergers or acquisitions;
3. To allow dissemination of Preoduct Model archival data within the organisation or to external regulatory bodies preserving both the integrity of the data objects and the information packages.

3.4 LOTAR

3.4.1 Long-term archiving

Within LOTAR, long-term archiving has a specific, defined meaning as follows⁴:

- Storage of a copy of data in an appropriate way for record, certification and legal purposes;
- Data will be preserved and kept available for use within the archive and possibly for further use;
- With certified conversion processes, the Native Data Representation generated by the source system can be converted into a Representation which is appropriate for long term archiving. To fulfil legal and certification requirements, the stored form can be an accurate or approximate Representation of the source;
- Integrity must be ensured by a Digital Signature;
- The data is retained over the long term;
- Invariance is mandatory;
- Business, legal and certification requirements are covered.

The E-ARK Common Specification for Information Packages states as its aim:

“...specifications which aim to provide a common set of usage descriptions of international standards for packaging digital information for archiving purposes. These specifications are based on common, international standards for transmitting, describing and preserving digital data. They also utilise the Reference Model for an Open Archival Information System (OAIS), which has Information Packages as its foundation.”

E-ARK CSIP does not attempt to make a definition of archiving, but as both CSIP and LOTAR use OAIS as a foundation it is useful to cite the OAIS Reference Model definition of an Open Archive Information System (OAIS), as follows:

⁴ EN 9300 Part 003, 4.2.6

“An Archive system consisting of hardware, software, information, and policy-based processes and procedures put in place and operated by an organization and its staff. The organization has accepted the responsibility to preserve information and make it available for a Designated Community. The organization may be part of a larger organization. The system meets a set of mandatory responsibilities that allows an OAIS Archive to be distinguished from other uses of the term ‘archive’. The term ‘Open’ in OAIS is used to imply that this Recommended Practice and future related Recommended Practices and standards are developed in open forums, and it does not imply that access to the Archive is unrestricted.”

3.4.2 Data Model

The following is the LOTAR description of its data model. The contents of data packages are also described in LOTAR as detailed below. Broadly, the LOTAR data model considers the data comprising a Representation of the product model as follows⁵:

- The core data model
- The required metadata
- The engineering approvals data (as a Digital Signature)
- The Validation information

Note that LOTAR defines two different data authentication processes, namely Validation and Verification which are considered as distinct events in the 3DPM CITS. In some parts of LOTAR however, the term Validation is used to cover all data authentication processes.

3.4.3 The Core Model

The core model identifies the essential minimum of data which is required to preserve the design intent for a given purpose. The domain specific parts of LOTAR identify a purpose or set of purposes through appropriate use cases, and therefore the core model which is required to support the business cases. The core model is defined as a system of data elements together with their representation information, interpretation information and data quality criteria they must meet.

3.4.4 Metadata

Metadata considered within the LOTAR data model is limited to data used to retrieve the package and data used to Validate the data and its provenance.

According to LOTAR⁶: “The scope of the metadata depends on the particular use cases it applies to, and is detailed in the domain specific parts. These also detail any data quality criteria applicable to the metadata. Additional metadata may be applied, and should be agreed in the ingest (submission) agreement. Data used in the internal management of the archive is outside of the scope of LOTAR. In addition the detailed intellectual property rights information may be explicitly identified, described, documented in the Archival Information package during ingest and after the retrieval process.”

⁵ EN 9300 Part 003, 7.3.2.1

⁶ EN 9300 Part 003, 7.3.2.3

It is implicit in LOTAR that metadata accompanies data in the submission to the archive and is used within the digital repository or OAIS to aid the retrieval of data and to provide proof of provenance. It is not explicitly required that metadata is included in the package, but E-ARK CSIP sets minimum standards for this such that information packages are completely self describing and interoperable outside the digital repository or OAIS.

3.4.5 Digital and Engineering Signatures

A Digital Signature is a sort of seal of digital data. It is produced by using mathematical algorithm with the help of a private cryptographic key. With the help of the related public key the signature can be checked at any time to identify the signature-key-owner and to proof the integrity of the data. The Digital Signature may follow the rules given by, for example: Directive in 1999/93/EC of the European parliament and the council from the 13th of December, 1999 about collective basic conditions of Digital Signature.

A Digital Signature has to be renewed every approximately 5-6 years. For LOTAR, therefore, the Digital Signature is only used to safeguard the data integrity for the short period between the producers' release and the transfer to the archive system. Time stamps (Digital Signatures authenticated by 3rd party services such as to identify producers and time of creation are used within LOTAR alongside stored data (AIPs). As this information is held outside of the archival package it is out of scope of the E-ARK specification. The encoding of Digital Signatures within PREMIS in Submission Information Packages (SIPs), should be adequate for Verification on ingest into the archive.

According to LOTAR⁷: “An Engineering Signature supports a business release process. With it the data producer asserts that the prepared data fits with the process and quality requirements from the engineering point of view. It may also identify the approver. Similarly to metadata the Engineering Signature is domain dependent and the use of it should be agreed between the producer and the archive.”

Engineering Signatures are described in LOTAR as being Digital Signatures as they are accompanying digital product models. It is important to note that in LOTAR the Engineering Signature is attesting the quality of the product data as submitted, not the long term authenticity or quality of the archived data.

3.4.6 Validation

LOTAR defines sets of Validation Properties. The Validation Properties are used within automatic or manual processes to check the consistency of data content during transformations from one Representation and/or format into another Representation and/or format. LOTAR defines domain specific Validation Properties, for example, for CAD and PDM but further extensions of Validation properties are possible by user agreement. The use of Validation properties is mandatory within LOTAR, ensuring data integrity and process security. Recommended subsets of STEP Validation properties are described within LOTAR-1xx series and LOTAR-2xx series.

According to LOTAR⁸: “A Validation Property should be something that is relatively simple to compute, compact to store, but which detects the most important errors. It is recommended to

⁷ EN 9300 Part 003, 7.3.2.4

⁸ EN 9300 Part 003, 7.3.5.2

use at a minimum Validation properties for the key and the global characteristics of the object the Validation Properties are assigned to. For example, geometric models may include Validation criteria such as surface area, volume, or centre of gravity, which are highly sensitive to errors in the model. Engineering correctness is out of scope for Validation Properties.”

Validation Properties in LOTAR analyse only the correctness of the Representation of the core model. However, it should be noted that an invalid Representation is often an indication of an engineering error.

The Producer calculates Validation Properties and produces a results report either manually or automatically during data preparation according to Validation Property Rules. The Validation Properties are integrated into Preservation Description Information (PDI) during data preparation and the results report included in the SIP.

LOTAR requires that data selected for archiving undergoes data quality or Validation checks involving the generation of Validation properties from the data and checks that each Representation of the data meets with the recommended data quality criteria (Validation rules data). Good archival practice dictates that we should not only include the results of those Validation checks but that Validation rules data and Validation properties should be included in the package. The specification also recommends that Validation events and results are recorded within preservation metadata (PREMIS).

3.4.7 Verification

The intention of Verification is to ensure that data is correctly represented. Verification rules ensure that a data Representation meets the quality requirements within defined tolerances. A Verification is successful if no Verification rule is violated. The Verification rules are domain specific and are defined for example within the LOTAR-1xx series and LOTAR-2xx series. Only Verification rules belonging to the core models are part of LOTAR.

Data Verification of a Representation of data (e.g. a STEP Representation) for a SIP will result in a Verification Results report. LOTAR requires that a reference to the Verification Results report should be included in the SIP Descriptive Information (metadata) and a copy of the report may be also included.

LOTAR requires that each Representation of the Product Model be verified using data quality rules. The data Verification process is supported by the following documents:

- Data quality rules (Verification)
- Data Verification results report

A reference to the Verification results report should be included in Descriptive Information (LOTAR). Good archival practice states that we should not only include the reference to Verification results in descriptive metadata but should include the Data Quality Rules and the Verification reports within each Representation. The CITS also suggests that Verification events and results are recorded with preservation metadata (PREMIS).

4 Implementation

4.1 Standards

4.1.1 LOTAR

LOTAR is a standard for the long term archiving and retrieval of digital technical product documentation such as 3D, Computer Aided Design (CAD) and Product Digital Model (PDM) data.

In common with E-ARK, LOTAR references and extends OAIS, but there is a different but complementary focus within the standard and the specifications. Firstly, E-ARK is concerned with the Information Packages described in OAIS (SIP, AIP, DIP), refining their definition such as to:

1. Establish a common understanding of the requirements which need to be met in order to achieve interoperability of Information Packages;
2. Establish a common base for the development of more specific Information Package definitions and tools within the digital preservation community;
3. Propose the details of an XML-based implementation of the requirements using, to the largest possible extent, standards which are widely used in international digital preservation.

E-ARK is not prescriptive on process and encompasses multiple use cases for submission of content to an archive, which can involve multiple different entities or organisations. For example E-ARK can envisage situations where SIPs are created at a Producer or at an Archive.

Secondly LOTAR is not concerned with package structure but with the data content of the packages and with the main phases of archiving as described by OAIS: data preparation, ingest, archival storage, access and retrieval. It does not consider interoperability, as the Archives considered will typically be owned and managed by the data owners.

LOTAR extends OAIS with detailed process views, which provide normative descriptions of key aspects of the process, together with extended descriptions of roles and responsibilities. The process views identify any data format conversions required, the preparation of Validation information and the execution of Validation and Verification tests. These process views have the objective of establishing a certifiable process for archiving.

4.1.2 PREMIS

From the EARK PREMIS Content Information Type Specification (CITS):

“When using preservation metadata together with the Common Specification for Information Packages (CSIP) (<http://earkcsip.dilcis.eu>), it is recommended that these are included in the information package in PREMIS format. Although this is not mandatory, all tools claiming to be able to validate CSIP compliant Information Packages must also be able to validate PREMIS metadata once it exists within the package”. The two high-level requirements for the use of PREMIS in Common Specification IPs are that:

- All preservation metadata is created according to official PREMIS guidelines;
- All PREMIS metadata is referenced from the amdSec/digiprovMD element of the appropriate METS file.

Further, to enhance the interoperability of the CSIP and to strengthen the management of information packages (IPs) in an archive, this specification imposes additional requirements regarding the use of PREMIS for describing IPs. The principles adopted in the CSIP for deciding which additional PREMIS semantic units are required are:

- PREMIS should be used to record detailed technical metadata;
- Technical information should be included in PREMIS metadata by using extension schemas;
- Information about agents carrying out preservation actions must be recorded in the PREMIS metadata (this is because the METS agents describe agents relevant for generic IP level events, such as the creation or submission of the package, not the preservation of the data);
- Event descriptions should be included in PREMIS metadata. Use of the official PREMIS event vocabulary (<https://id.loc.gov/vocabulary/preservation/eventType.html>) is recommended (note that more elaborate descriptions can be made than are made in this CITS);
- Detailed rights information should be included in PREMIS. High-level rights information in METS indicates restrictions. Detailed, object-specific rights information will be included in the PREMIS metadata;
- File format information for all files should be included as Persistent Unique Identifier (PUID) values in the appropriate PREMIS semantic units.

PREMIS is one way to record Preservation Description Information for an information package (the other being METS) and specifically for LOTAR, the following preservation actions need to be recorded:

1. Engineering (Digital) Signatures;
2. Engineering (Digital) Signature events;
3. Validation events;
4. Verification events.

The use of PREMIS within the 3DPM CITS requires that information packages also conform to the CITS PREMIS (use of PREMIS is not mandatory within CSIP but if used there are a number of MUST and SHOULD requirements. So, the following PREMIS semantic units SHOULD be included:

1. Checksums;
2. File format;
3. Relationship (if IP is part of another IP);
4. Event identifier;
5. Link to agent/object;
6. Migration event type;
7. PREMIS agent;

8. PREMIS rights.

4.1.3 METS

The main requirement for METS files in a CSIP Information Package is that these need to follow the official METS Schema version 1.12 <http://www.loc.gov/standards/mets/mets-schemadocs.html> (used by CSIP version 2.1) and the extension schema developed for the CSIP and published by the DILCIS Board. As new versions of METS Schema become available the DILCIS Board will evaluate these and, if necessary, update the CSIP.

The METS specification requires a METS profile document which describes the use of METS and the METS elements. All the requirements described in this specification are also expressed with METS profiles for the CITS 3DPM root and representation METS which can be found at: [CSIP https://github.com/DILCISBoard/E-ARK-CSIP/tree/master/profile](https://github.com/DILCISBoard/E-ARK-CSIP/tree/master/profile).

CSIP specifies that METS files be located at the root of the package folder structure (Root METS) and optionally in each of the Representations within its respective root folder (representation METS). CITS 3DPM has a requirement to contain at least one Representation and so will contain at minimum a root METS and a representation METS file.

4.2 Information Packages

The E-ARK SIP specification states that: “According to the Open Archival Information System Reference Model (OAIS) every submission of information to an archive occurs as one or more discrete transmissions of Submission Information Packages (SIPs).”

The OAIS Reference Model does not specify the format, structure or contents of these information packages. The EU funded E-ARK project (2014-2017) first acknowledged this problem and started to develop a solution in the form of an information package specification. This specification is now part of a set of specifications managed by an independent body named the Digital Information LifeCycle Interoperability Standards Board (DILCIS Board).

According to LOTAR⁹: “The formats of the packages are defined in a Submission Agreement between the producer and the archive.” In the case of building E-ARK and LOTAR compliant information packages, the submission agreement will reference the E-ARK Common Specification for Information Packages, the specification for Submission Information Packages (SIP), the specification for Archival Information Packages (AIP), the specification for Dissemination Information Packages (DIP) and the Content Information Type Specification for 3D Product Data (CITS 3DPM).

LOTAR defines the minimum content for information packages and the processes for creating them, definitions for the LOTAR terms can be found in the glossary and descriptions of the involved data below.

SIP Contents

- Content Information (CI)
- Preservation Description Information (PDI)
- Description Information (DI)

⁹ EN 9300 Part 011, 6.9

- Packaging Information (PI)

AIP Contents

- Content Information (CI)
- Preservation Description Information (PDI)
- Packaging Information (PI)
- Digital Signature Information

DIP Contents

- Content Information (CI) – which can be from more than one AIP
- Packaging Information (PI)

The following are descriptions taken from LOTAR regarding the involved entities in Information Packages. 3DPM CITS describes how these entities can be represented in E-ARK Information Packages.

Content Information (CI): Content Information includes the set of information that is the original target of preservation (and includes different Representations of that data).

Description Information (DI): The set of information consisting primarily of Package Descriptions (Descriptive Metadata) which is provided to a data management function in the repository (and in the case of CSIP included in Information Packages) to support the finding, ordering and retrieving by Consumers of archived information holdings. Descriptive Information can be extended with storage information and with reference information for the Product Model. (storage information is not included in E-ARK packages but as this is an extension of Descriptive Information that is passed to the repository, this is not an issue).

Descriptive Information includes archive metadata meeting the standards for the specific archive (according to ISO 14721:2003 – OAIS). Descriptive Information also includes a reference to the data Verification results report. For E-ARK this should be more correctly included in preservation metadata (METS or PREMIS) but can also be included in Descriptive Information for compliance purposes.

Preservation Description Information (PDI): PDI is information which is necessary to preserve the interpretation of the Content Information and which can be categorized as Provenance, Reference, Fixity, and Context information.

According to OAIS, the four categories can briefly be described as follows:

- Provenance describes the source of the Content Information, who has had custody of it since its origin, and its history (including processing history);
- Context describes how the Content Information relates to other information outside the Information Package. For example, it would describe why the Content Information was produced, and it may include a description of how it relates to another Content Information object that is available in the archive;
- Reference provides one or more identifiers, or systems of identifiers, by which the Content Information may be uniquely identified;

- Fixity provides a wrapper, or protective shield, that protects the Content Information from undocumented alteration. For example, it may involve a checksum over the Content Information of a digital Information Package.

Within LOTAR PDI also includes Validation and Verification information including: Validation Properties/reports and Verification reports.

Packaging Information (PI): Administrative data needed for archive data management. Packaging information is that information used to bind and identify the components of an Information Package. - information which, either actually or logically, binds, identifies and relates the CI and PDI.

Engineering (Digital) Signature: In the context of the LOTAR, a Digital Signature shall be¹⁰:

- uniquely linked to the signatory;
- capable of identifying the signatory;
- created using means that the signatory can maintain:
- under their sole control.
- linked to the data to which it relates in such a manner that:
- any subsequent change of the data is detectable.

An Engineering Signature is a special type of Digital Signature that expresses and fixes a volition of the signatory. It gives evidence of¹¹:

- the process of testifying quality of data against process / quality requirements by linking the signature owner to the data;
- the identity of the signatory by usage of appropriate methods of authentication;
- the integrity of the data by using appropriate methods protecting the signed object against unauthorized changes.

An Engineering Signature is created individually for each document and is used within LOTAR for the authentication of the data ready for archiving and storage. An Engineering Signature includes¹²:

- owner name;
- public key from owner and certification company;
- signature owner name and key;
- information about algorithm for use of the public key;
- start and end time of the certificate;

¹⁰ EN 9300 Part 005, 3.4

¹¹ EN 9300 Part 005, 3.4.1

¹² EN 9300 Part 995, 7.1

- identification number of the certificate;

The Digital Signature is used to ensure that:

- the document file is an exact copy of the stored original;
- the document is approved by the originating company;
- with digital engineering signatures; the document is approved by the given approver;

This process provides evidential weight for legal proceedings.

4.3 Package Structure

The CITS 3D Product Model information structure inherits its package structure from the E-ARK Common Specification for Information Packages and is shown in Figure 4.

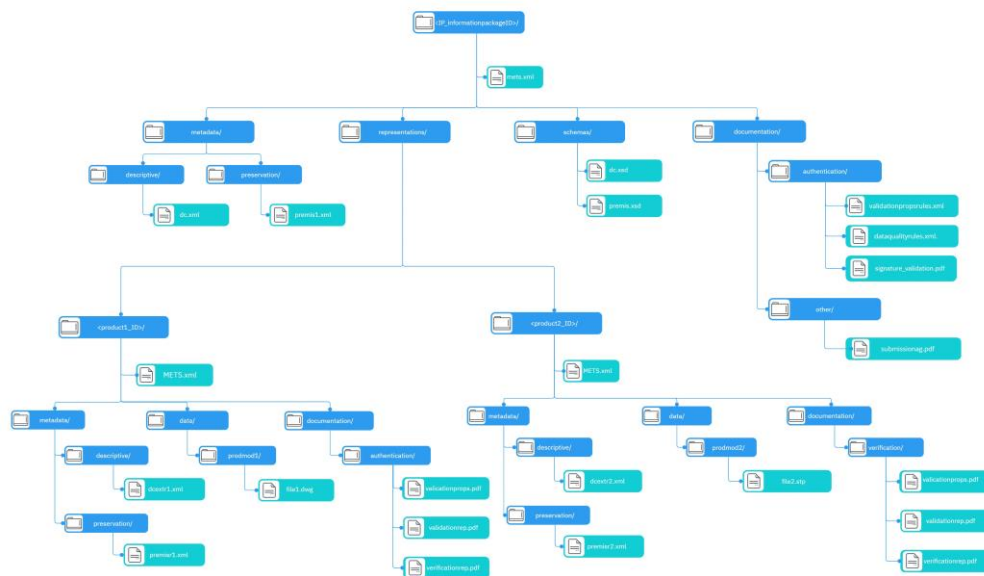


Figure 4: Example Information Package Folder Structure

A 3D Product Model Information Package package can consist of one to many Representations of the same Product Model which comprise Product Information Models of the same product. It is likely for example that a Product Model archival package will at minimum contain a Representation in Native Format and one in Standardised Open Format (e.g. STEP). Long-term usability of proprietary Native Format data is a risk as detailed in LOTAR, conversely open derivations such as STEP are at risk of losing some properties from the original.

LOTAR requires the inclusion of files or documents to support Validation and Verification of the data and Representations, including:

- Validation properties rules data

- Validation properties
- Validation reports
- Data Quality Rules (Verification)
- Verification reports

Validation and Verification rules information relate to the package and their outputs (properties, reports) relate to each individual representation. Within CITS 3DPM, a sub-folder named Authentication is included at root and representation levels within the Documentation folders to store the respective information related to Validation and Verification.

4.4 Descriptive Metadata

According to LOTAR¹³: “The producer creates a set of Descriptive Information (DI), which includes archive metadata meeting the archives requirements (according to ISO 14721:2003 – OAIS).”

Neither CSIP or LOTAR prescribe specific schemas for Descriptive Information and so this is left to be determined by the user organisation. CSIP states that this should be according to a standardised schema and EAD is generally recommended.

LOTAR makes integration of data Verification results reports into Descriptive Information as optional. and as the data Verification process is conducted for each Representation, according to CSIP rules this descriptive metadata should be held at Representation level. In addition, although LOTAR asks only for a reference to the Verification report, it is good archival practice to also include the reports, together with the criteria for data quality (Data Quality Rules).

If a standardised metadata schema is used at Representation level then a suitable location for the reference will need to be found in the chosen schema (for example EAD), or provided in a schema extension and this location specified in the Submission Agreement. Extensions to standardised metadata schemas should also be included in the package Schemas folder.

4.5 Preservation Metadata

LOTAR requires that Information Packages include Preservation Description Information (PDI) which in turn comprises information related to:

- Provenance
- Reference
- Fixity
- Context

According to the CITS PREMIS: “When using preservation metadata together with the Common Specification for Information Packages (CSIP) (<http://eakcsip.dilcis.eu>), it is recommended that these are included in the Information Package in PREMIS format. Although this is not mandatory, all tools claiming to be able to validate CSIP compliant Information Packages must also be able to

¹³ EN 9300 Part 0011, 6.8

validate PREMIS metadata once it exists within the package. The two high-level requirements for the use of PREMIS in Common Specification IPs are that:

- All preservation metadata is created according to official PREMIS guidelines;
- All PREMIS metadata is referenced from the amdSec/digiprovMD element of the appropriate METS file.

It is recommended that users review the CITS PREMIS specification.

In addition, LOTAR states that¹⁴ “the producer shall integrate Validation properties into the PDI to ensure that the existence of Validation Properties for later automatic Validation processes. As the Content Objects and Validation Properties related to Content Objects are held at Representation and as CITS 3DPM also requires the inclusion of Digital Signatures with each Representation then CITS 3DPM recommends the use of PREMIS at Representation level.

5 Glossary

Table 2: Glossary

Term	Description
Archival Creator	Organisation unit or individual that creates records and/or manages records during their active use.
Archival Information Package (AIP)	An information package, consisting of the Content Information and the associated Preservation Description Information (PDI), which is preserved within an Open Archival Information System (OAIS).
Asymmetric Keys	Asymmetric keys are pairs of keys, created in one step; they can be used in both directions. Encryption with the public key can only be decrypted with the private key; if the encryption is done with the private key, the decryption can only be done with the public key; such a key pair can be used for encryption and for signing
Authentication	Term needs verifying by DILCIS
Cardinality	The term describes the possible number of occurrences for elements in a set. The numbers have the following meanings: (1..1) – in each set, there is exactly 1 such element present

¹⁴ EN 9300 Part 011, 6.7

	<p>(0..1) – the set can contain from 0 to 1 of such elements</p> <p>(1..n) – the set contains at least one element</p> <p>(0..n) – the set can contain up to n of such elements, but it is not mandatory</p> <p>(0..0) – the element is prohibited to use</p>
Content Data Object	The Data Object, that together with associated Representation Information comprises the Content Information (OAIS – ISO 14721:2012).
Content Information	A set of information that is the original target of preservation or includes part or all of that information. It is an Information Object composed of its Content Data Object and its Representation Information (OAIS – ISO 14721:2012).
Data File	A component which contains data and has an associated MIME file type. A Data File can encapsulate multiple bit streams and metadata according to a standard such as a DICOM but must have a recognised MIME file type. A Data File may comprise one or more subsidiary Byte Streams; for example, an MP4 file might contain separate audio and video streams, each of which has its own associated metadata.
Data Quality Rules	Data Quality or Verification rules ensure that a representation of a product model meets quality requirements within defined tolerances, i.e. that a specific representation represents the Product Model with sufficient accuracy.
Derived Representation	A transformation of the native data, which may be based on a Native Format or a Standardised Format, e.g. an html version may be derived from a text document as an alternative representation.
Descriptive Information	Descriptive metadata described by LOTAR as including archive metadata meeting the archive's requirements (according to ISO 14721:2003 – OAIS).
Digital Signature	An Digital Signature is a defined method to sign an object in electronic environments; it provides means to authenticate the signatory and the signed object in an unambiguous and safe way by attaching to or logically associating data in

	electronic form to other electronic objects
Dissemination Information Package (DIP)	An Information Package, derived from one or more AIPs and sent by Archives to the Consumer in response to a request to the OAIS.
Document	A single or group of related Data Files with common metadata. For example, a Document may consist of a PDF file together with associated attachments or a word file with a separate image signature sheet. A document can be considered to be an entity that is approved/signed as a whole by a practitioner.
Information Package	A logical container composed of optional Content Information and optional associated Preservation Description Information used to delimit and identify the Content Information and Package Description information used to facilitate searches for the Content Information.
Internal Archival Long Term Preservation guidelines	This type of guideline can have different names depending on the creator. Generally, archives specify technical guidelines and/or regulations for formats, specifying what they will accept and maintain for the long term/ Depending on the archive and available technical resources, the criteria for the selected formats can differ from archive to archive.
Level	<p>The level of requirements of the element following RFC 2119 http://www.ietf.org/rfc/rfc2119.txt</p> <p>MUST – this means that the definition is an absolute requirement</p> <p>SHOULD – this means that in particular circumstances, valid reasons may exist to ignore the requirement, but the full implications must be understood and carefully weighed before choosing a different course. http://www.ietf.org/rfc/rfc2119.txt</p> <p>MUST NOT – this means that the prohibition described in the requirement is an absolute prohibition of the use of the element.</p> <p>SHOULD NOT – this means that in particular circumstances, violating the prohibition described in the requirement is acceptable or even useful, but the full implications should be understood and the case carefully weighed before doing</p>

	<p>so. The requirement text should clarify such circumstances.</p> <p>MAY – means that a requirement is entirely optional.</p>
Open Archival Information System (OAIS)	An Archive consisting of an organisation, which may be part of a larger organisation, of people and systems, that has accepted the responsibility to preserve information and make it available for a Designated Community. It meets a set of responsibilities that allows an OAIS Archive to be distinguished from other uses of the term 'Archive'.
Original Product Model or Native Representation	Used specially to keep the design intent for long term archiving in the context of certification and legal requirements for proof. It can be stored in native or standardised formats. (LOTAR)
Preservation Description Information (PDI)	The information which is necessary for adequate preservation of the Content Information and which can be categorised as Provenance, Reference, Fixity and Access Rights Information. (LOTAR).
Product Information Model	A Product Information Model represents an information model which provides an abstract description of facts, concepts and instructions about a product e.g. STEP (ISO 10303-1:1994) Application reference model or STEP Application interpreted model.
Product Model	A Product Model represents an occurrence of a product information model for a particular product, e.g. the geometric model of a part 123. Companies will create Product Models of different types, depending on the life cycle stages or disciplines, e.g. there are Product Models of type 'space allocation mock up'. Product Models are independent from their presentation.
Public Key	A public key is the part of the asymmetric key pair that is known to everyone
Private Key	A private key is the part of the asymmetric key pair that is only known by the owner of the asymmetric key pair
RDBMS	Relational Database Management System
Representation	A Representation within an Information Package contains archival data. If an Information Package contains the same data in two or more different formats (i.e. an original and a long term preservation format) or in different types of organisations (arrangements), they are placed within two or more separate Representations within the Representations folder of the Information Package of the Information

	Package.
Representation Information	The Representation Information must enable or allow the re-creation of the significant properties of the original data object.
Standardised Machine-readable Documentation	A standardised machine-readable document is a document whose content can be readily processed by computers and is based on a commonly accepted standard. Such documents are distinguished from machine-readable data by virtue of having sufficient structure to provide the necessary context to support the business processes for which they are created.
Standardised Open Format	A format of data in a syntax which is derived by as broad community, such as ISO and which is independent of a specific system or interface. "Open" means completely and precisely documented in syntax and semantics and is applicable for free. In addition, standardisation processes regulate the change process for the standard. (LOTAR)
Submission Agreement	The agreement reached between an archive and the submission producer that specifies a submission format (e.g. 3DPM CITS) and any other arrangements needed, for the data submission session. Any special conditions on patient confidentiality could be specified in the submission agreement.
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.
Submitting Organisation	Name of the organisation submitting the package to the archive.
Time Stamp/Signature	A time signature is created automatically as part of a certified process and requires certified hardware; it provides a legal guarantee for time and owner of the data. (LOTAR).
Validation	Validation is applied to guarantee the integrity of the content of a Document throughout the entire process of long-term archiving. In the context of LOTAR the validation will be done by calculation and comparison of Validation Properties. A set of Validation Properties is like a fingerprint for the content of the document. Each change of the content changes one or more attributes of the Validation Properties. Validation properties should be independent from the system and representation within given deviations (LOTAR).

Validation Properties	Validation properties are measurable characteristics of a given Product Model that can demonstrate the veracity of a representation of the model. E.g. weight, center of gravity. Validation properties are calculated during the process of Validation for each representation of the model, e.g. STEP (LOTAR).
Validation Properties Rules Data	Validation Properties Rules Data is the original Validation Properties derived from the source Product Model (LOTAR).
Verification	A process to ensure that data is correctly represented (e.g. in a package representation). Verification rules ensure that a data representation meet the quality requirements within defined tolerances. Verification rules are domain specific (CAD, PDM, Electronic Assembly, Fluid Dynamics) and are defined within LOTAR (LOTAR).

6 Postface

AUTHOR(S)	
Name(s)	Organisation(s)
Stephen Mackey	Penwern Limited

REVIEWER(S)	
Name(s)	Organisation(s)
[Name]	[Affiliation]
[Name]	[Affiliation]
[Name]	[Affiliation]

Project co-funded by the European Commission within the ICT Policy Support Programme		
Dissemination Level		
P	Public	x
C	Confidential, only for members of the Consortium and the Commission Services	

REVISION HISTORY AND STATEMENT OF ORIGINALITY

Submitted Revisions History

Revision No.	Date	Authors(s)	Organisation	Description
[Version]	[Date]	[Who]	[Affiliation]	[What]

Statement of originality:

This deliverable contains original unpublished work except where clearly indicated otherwise. Acknowledgement of previously published material and of the work of others has been made through appropriate citation, quotation or both.