White Paper for

**Long Term Archiving and Retrieval of Product Data within the Aerospace Industry (LOTAR)**

Technical Aspects of an approach for application of Project Group “LOTAR”:

Version: 1.0
Status: Released
Author: Project LOTAR
Date: 2002/08/30
Distribution: project distribution list
## History

<table>
<thead>
<tr>
<th>Version</th>
<th>Status</th>
<th>Authors</th>
<th>Changes</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.1</td>
<td>Draft</td>
<td>Project group</td>
<td>Creation</td>
</tr>
<tr>
<td>1.0</td>
<td>Released</td>
<td>Project group</td>
<td></td>
</tr>
</tbody>
</table>
Contents

Authority 5

Conventions 5

Forward 5

Abbreviations 6

Definition of terms 7

1 Background and motivation 9

2 Scope 10

3 General approach of project LOTAR 11

3.1 Approach on data level 11

3.2 Approach on process level 11

3.3 Approach on system architecture level 12

4 Relationships with other standards and efforts 13

4.1 OAIS 13

4.2 ISO 10303 (STEP) 14

5 Description methods 16

5.1 UML Use Case diagram for scope definitions 16

5.2 Simplified activity diagrams for business process analysis 17

5.3 Methods for data description 18

5.4 UML Package Diagram for system architecture description 19

6 Major requirements 21

6.1 Acceptance 21

6.2 Legal demand 21
6.3 Security

7 Scenarios

7.1 Start archiving process

7.1.1 Archiving digital documents

7.1.2 Archiving CAx Data

7.1.3 Archiving PDM data

7.1.4 Archiving PDM structure

7.2 Validation

7.3 Retrieval of archived data

7.4 Administration of archived data

8 General approaches for long-term archiving

8.1 Static aspects

8.1.1 Native format

8.1.2 Neutral format

8.1.3 Multi format

8.2 Dynamic aspects

8.2.1 Data

8.2.2 System architecture

9 Strategic recommendations for long-term archiving

9.1 General Strategy

9.2 Proposal of a technical solution

9.2.1 Recommendations for different roles

9.2.2 Processes

9.2.3 Data concept

9.2.4 System architecture

9.3 Administrative aspects

References
Authority

This document has been approved for publication by the project steering committee of project group “Aerospace LOTAR” and represents the consensus of technical agreements of the participating project partners. The document is a result of review and authorization as stated by the project partners.

Many thanks for the active participation of colleagues of following companies:

AIRBUS
MTU
EADS Military Aircraft
Fairchild Dornier
PROSTEP AG
ProSTEP iViP Association

Special thanks for the participation of colleagues of Fairchild Dornier. Despite the insolvency proceeding of Fairchild Dornier they remained highly motivated on the subject.

Participants of core team project group LOTAR are: Christine Grafe (Airbus, Project leader), Jean-Yves Delaunay (Airbus), Dr. Jörg Wirtz (EADS Military Aircraft), Heinz Knittel (MTU), Uwe Urra (MTU), Stefan Schulz (Fairchild Dornier), Jim Broughton (Fairchild Dornier), Andreas Trautheim (PROSTEP), Ludger Millarg (PROSTEP), Fahmi Bellalouna (ITM/ TUM)

Conventions

Within the context of this document the terms “shall” and “may” are used per Code of the US Federal Regulations (CFR) Title 14 Chapter 1, Part 1:

- “Shall” is used in an imperative sense.
- “May” is used in a permissive sense to state authority or permission to do the act prescribed, and the words "no person may …“ or "a person may not …“ mean that no person is required, authorized, or permitted to do the act prescribed.
- Includes means "includes but is not limited to".

Forward

This document describes the results of the project group based on:
Consolidation, common understanding, legal and business requirements

Development of a strategy and preferred approach

It has to be seen as a starting point to specify methods, scenarios, applications and process modules for the aircraft industry which support long-term archiving as well as re-use (readability first, furthermore ability for further processing) of product data including 3D geometry.

Since communication processes with suppliers, partners or customers manage archiving-relevant information, scenarios and resulting descriptions of processes are to be developed in a way to be applicable also for these communication processes (data exchange and data integration). The results of the specification shall be submitted first as AECMA standard.

**Abbreviations**

2D two Dimensional
3D three Dimensional
AP214 Application Protocol 214 – ISO 10303-214
AECMA European Association of Aerospace Industries
AIM Application Interpreted Model
AIP Archival Information Package
ARM Application Reference Model
CAD Computer Aided Design
CCx Conformance Class x; x stays for the number of the CC, e.g. 2 for geometric data or 6 for PDM data
CFR Code of the US Federal Regulations
DIP Dissemination Information Package
DMU Digital Mock-up
LOTAR Long Term Archiving and Retrieval
OAIS Reference Model for an Open Archival Information System
CUSTOMER Original Equipment Manufacturer
OMG Object Management Group
PDM Product Data Management
SIP Submission Information Package
Definition of terms

Archiving: Certified process of writing, storing and retrieval of selected information into an archive.

Assembly: A node in the part structure that is further decomposed into subassemblies or parts. The assembly structure is a non-configurable explicit part structure (i.e., without variations). The assembly structure is a hierarchical structure of arbitrary depth.

Asynchronous Data Exchange: Processes of offline data transferring from one site to another site via file transfer. The data is duplicated on the receiving site with no link to the original data on the sending site.

Business Process: A business process is a process related to an execution e.g. of a product data export and driven by an individual engineering step (e.g., communication of changes, establishing a collaboration).

CAD System: Application for creation and modification of geometric product model data

Digital Mock-up (DMU): Digital Mock-up is a virtual assembly of the complete product or components of the product in a computer system. The purpose of a Digital Mock-up is all kind of simulations concerning the geometric shape, kinematics or design studies.

Data Module: A data module is a set of data types (entities and its attributes) to support a specific functionality.

Design documents: Set of documents relevant for the definition of products.

Designated Community: An identified group of potential Consumers who should be able to understand a particular set of information. The Designated Community may be composed of multiple user communities.

Document: A document is a logical container for defined product data. A document may be associated to an item or another object and is either represented in physical or digital form.

File: A file is a container of data on a computer storage medium.

Geometric Model: A geometric model is a representation of a shape. The geometric data are typically generated by CAD systems in digital form.

Item: An item is an element of a product relevant for a bill of material. An item may be a single item, an assembly, raw material or non-geometrical components, e.g., lubricant.
Part 21 File: See STEP Physical File.

PDM Schema: Internationally harmonized intersectional subset of the STEP application protocols (AP) AP 203, AP212, AP 214 and AP 232.

PDM System: Application for creation and modification of meta data for management of product definition data and its life cycle

Retrieval: Certified process of providing archived data to designated community.

STEP: Is a synonym (see Abbreviations) for the entire set of partial standards (parts) of ISO10303.

STEP Physical File: A file that contains actual product model data encoded as specified in ISO 10303-21. This is the preferred format for the file based asynchronous exchange of product data between heterogeneous tools.

Version: A version is a formal status within the change history of an object, e.g., representing a specific stage of a part, document, geometric model or other objects. Terms like version, revision, iteration or issue are not distinguished within the current scope of document.
1 Background and motivation

Within the aircraft industry extremely high requirements exist to the processes for administration, archiving as well as re-use of product defining data. Archiving and retention of documents have to take place in such a way that they are usable mainly for the proof of:

- Legal constraints
- Certification constraints
- Constraints on product liability
- Contractual constraints
- Re-use of data belonging information after years
- Manufacturing processes
- Modifications on products and documents

The duration of retention differs for the different types of documents. For CAD and PDM data there are differences for Design and Type Design Data. But it is expected to archive this data up to 99 years or furthermore on duration, belonging to the California product liability regulation.

Long term archiving of CAD data meant so far archiving of 2D drawings as paper drawings, digital files of (usually) tiff format, micro fiche or aperture cards.

Today 3D models are used as master models and 2D drawings are generated from the 3D models. The result of the development within 3D master models is that the 3D models are no longer fully dimensioned, all necessary views are only within the 3D model available. Only the combination about product structure, 3D and 2D information together describe an object completely.

For these reasons the current situation is characterized by a huge number of 3D-CAD models, grown enormously, which are kept write-protected in databases in their native formats. They are quasi safe for about 10 years but not safe in sense of legal and product liability requirements. This data must therefore be checked and migrated regularly in order to conserve the required model quality.

Today a usable and legally accepted solution for long term archiving of 3D CAD and PDM data worldwide does not exist. The international aircraft industry has started to push the activities to develop practicable archiving solutions. The automotive industry has not yet started to specify solutions for archiving of 3D data because of low legal priorities.
2 Scope

This White Paper covers the main topics regarding long-term archiving within the Aerospace industry.

The document focuses furthermore on the results and experiences gained in the project and summarizes scenarios, concepts and technical solutions.

The approach of solution has to support:

- Internal processes
- Legal requirements
- Different scenarios (e.g. during single process steps or migration processes)
- Liability and quality of archived data (e.g. 3D and PDM data)

Out of scope is a more detailed analysis of legal requirements and requirements of other authorities. Nevertheless, these requirements are considered as far as possible within the technical approach. Further out of scope are the principles of archiving (e.g. redundant data storages, WORMs, refresh mechanisms) and hardware solutions for the archive.
3 General approach of project LOTAR

Existing archiving standards are still based on the assumption that the data is not modified when being archived and that the conformity of the original and the archived data can be directly validated by visualisation.

This was correct for the archiving of two-dimensional presentation with microfiche and aperture cards or even with tiff-files as a first digital archiving approach. However whilst the definition of the product more and more moves to three-dimensional data based on a digital and formal product data model, the visualization of the data is no more ensure by existing data archive storages themselves (e.g. paper, microfiches, TIFF files) and therefore needs to be independent of storage technology and format. Therefore the conformity of the original data and the archived data format can not be validated by simple visualization any more.

The original native formats of the CAD-systems change too frequently compared with the legal requirements for archiving periods due to changes of hardware platforms and operating systems. Therefore a transformation of the original product model data to a more stable format becomes necessary.

In the following the general approaches of the planned standardization of LOTAR project group on data- process- and system architecture level in order to fulfil the requirements of product model based archiving of three-dimensional and PDM data will be presented.

3.1 Approach on data level

Subsets of ISO10303 will be defined, containing the most rigid product data representation methods, aiming to achieve a maximum degree of unambiguity of the stored product models. Using an ISO-standardized data model for the archiving will be recommended to ensure the availability of the semantics of stored data as well as a maximum of system independency.

Recommendations will be given on the minimum subset of data types required to be archived depended on specific use cases.

Different ways of instantiating the archive data model shall be avoided to ensure the quality of processes and converted data and the correct presentation when retrieval and reusing the model after a long period of time

3.2 Approach on process level

The process to be defined within LOTAR must ensure the traceability of conversions of the original product model generated within the PDM and CAX -systems to the archive format and back from the archive format to the visualization or reuse of data.
The definition of the process must be on one hand defined rigid enough to satisfy legal requirements and must offer different options in order to ensure the adoptability and applicability of the archiving standard to a broad range of companies and application scenarios on the other hand. Therefore the usage of certified tools and interfaces (i.e. conversion tools), vaulting mechanisms and even the usage of digital signatures will be part of the proposal.

3.3 Approach on system architecture level

The LOTAR project group considers that existing system architectures within the various enterprises are very different. In order to make LOTAR applicable for a broad range of companies and scenarios, LOTAR shall apply to those different system architectural approaches. For example the planned standard will cover scenarios and requirements for a separate archiving database independent from operative databases. The standard will also address approaches where the management of archived data will be performed within operative Document Management or PDM systems.
4 Relationships with other standards and efforts

The LOTAR project has the objective of an efficient harmonization of requirements and approaches at the international aerospace community level, e.g., as already started with the LTA PDES Pilot for long term retention of type design data and if required with other branches like the automotive or ship industry.

The LOTAR project references mainly ISO 14721.2 Open Reference Model for archiving Information System (OAIS) and ISO 10303 (STEP).

4.1 OAIS

OAIS standardizes a reference model for a system architecture of archiving systems and processes. The OAIS model is already adapted by retention systems in several industries and is considered by the project group as a mature model.

The OAIS functional model is shown in Figure 1

![Figure 1 Functional Model of OAIS](image)

The project group focuses first on the following five main process modules Preservation Planning, Data Management, Access, Ingest and Administration. The process module Archival Storage is assumed to be already covered by basic functions of existing archival systems (Data base, Storage, Media, etc.).
The project group expects that the OAIS definition of Archival Storage will be also applicable for 3D and PDM data without modifications or extensions. Therefore this process module is not considered within this project.

**4.2 ISO 10303 (STEP)**

The STEP Standard (ISO 10303) will be the basis for the logical data models, semantics and format needed to ensure the accessibility and possibility of interpretation by the designated community of the data for the retention period.

![Figure 2 Overview about Structure of ISO 10303](image)

The key approach of project LOTAR is to apply STEP methods to transform real (product) data descriptions into a neutral format for long term retention without loss of semantic.
Figure 3 Application example, formal schema definition and mapping onto Part21 file format
5 Description methods

Project LOTAR uses accepted standardized methods to support an equal level of understanding. Within the White Paper and further specifications within the planned standardization the methods are dedicated to four main categories:

- scope
- processes
- data
- system architecture

For the description of the scope the modeling methods are based on UML Use Case Diagrams. The description of processes will be done with Simplified Activity Diagrams. Within a following specification the relevant data modules will be described with Express G Diagrams. To support the development of a system architecture the modeling method of UML Package Diagrams are used.

5.1 UML Use Case diagram for scope definitions

The Use Case Diagrams, based on Unified Modeling Language (UML) focuses to apply the permutation of the requirements into specific scenarios.

UML is an industry-standard language for specifying, visualizing, constructing, and documenting the artifacts of software systems. It simplifies the complex process of software design, making a "blueprint" for construction. The diagrams are realized with the specification of UML version 1.4.

The UML Use Case Diagram describes the dependencies which can occur between identified use cases and involved participants (actors) within the environment of a specific system or domain.

The dependencies between the use cases are described by different formats of arrows. Dashed arrows describe the relationships between the use cases (include or extend). Solid line Arrows describe the heredity between the use cases, solid lines describe the interaction between actors and use cases. Figure 4 gives an example.
Figure 4 Example UML Use case diagram

The “start archiving process” will be triggered by the actor “producer”. The use case includes the sub case “initialization of archiving process”, which inherits all functionalities of the sub case “immediate archiving at release”. “Archive digital documents” is marked with a thick line i.e. the thick line visualizes a reference to another detail description.

5.2 Simplified activity diagrams for business process analysis

The detailed description and analysis of scenarios and resulting processes are figured out by simplified activity diagrams based on the UML and IDEF 0/1.

Figure 5: Example of a simplified activity diagram
The simplified activity diagram identifies the roles participating within the process. The roles perform several activities. The data flow represents the interaction between the single activities within the process chain.

The representation differs between activities (oval form), decisions (hexagons) and resources (rectangles) like user interfaces, system functionalities or automatic processes. The shadow behind activities or processes indicate that a further detailed description for a level 2 activity or process is available.

5.3 Methods for data description

The description of the relevant data will be done by graphical representation via Express G Diagrams and Data Dictionary in tabular form as a union.

The Express G (ISO 10303-11) Diagrams describe formally the data elements and their constrains within the defined archiving processes. The specification will use the Express-G modeling method as part of ISO 10303. It will visualize the logical context between the identified information objects.

Elements of the Express G Diagrams are displayed Figure 7.

<table>
<thead>
<tr>
<th>Schema</th>
<th>Attribute / relationship</th>
<th>Inter-Schema-References</th>
<th>Data type</th>
<th>Page references</th>
</tr>
</thead>
<tbody>
<tr>
<td>process</td>
<td>SCHEMA</td>
<td></td>
<td>process</td>
<td>page#, ref#</td>
</tr>
<tr>
<td></td>
<td>process</td>
<td></td>
<td>ENTITY</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>things</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>choice</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>what_it_is</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>REAL</td>
<td></td>
</tr>
</tbody>
</table>

Figure 6: Express G Syntax

Figure 7 describes relationships between the entities person and organization. The entity person, person in organization and organization are further described by their relevant attributes.
The data dictionary describes the semantic of entities and attributes represented in the data model and gives further information over value allocations, rules and constrains. It extends the graphical representation in Express-G.

5.4 **UML Package Diagram for system architecture description**

To specify the first approach of the system architecture UML package diagrams are used. The UML package diagram gives an overview of functional models. Figure 8 shows an example.
Figure 8: Example for an UML package diagram

The diagram should be read from the left to the right side, i.e., the module “Session-handling” has a dependency from the module “GUI” and/or “Administration”, thereby implicitly also process dependencies are illustrated.
6 Major requirements

This chapter identifies the key requirements for LOTAR. The LOTAR project group analyzed other recommendations (see references) which also describe requirements for archiving. The result of the analysis is a consolidation of the necessary requirements according the categories as follows:

- Acceptance
- Legal demand
- Security

6.1 Acceptance

The first requirement category belongs to the users acceptance addressing the manageability of archived documents and the minimisation of costs of archiving within the long term archiving system:

<table>
<thead>
<tr>
<th>Nr.</th>
<th>Requirement</th>
</tr>
</thead>
<tbody>
<tr>
<td>AC1</td>
<td>Each document has to be retrievable by suitable retrieval-techniques.</td>
</tr>
<tr>
<td>AC2</td>
<td>Each document has to be found promptly.</td>
</tr>
<tr>
<td>AC3</td>
<td>Minimize workflow expenses for archiving</td>
</tr>
<tr>
<td>AC4</td>
<td>Make data available for further use</td>
</tr>
<tr>
<td>AC5</td>
<td>Provide a planned end for archived documents/data</td>
</tr>
</tbody>
</table>

Table 1: Requirements of category “Acceptance”

6.2 Legal demand

The second requirement category belongs to the compliance of legal demands. Because of the fact that legal demands are not fully described by laws the following list of requirements satisfies the known and expected demands.

<table>
<thead>
<tr>
<th>Nr.</th>
<th>Requirement</th>
</tr>
</thead>
<tbody>
<tr>
<td>LD1</td>
<td>Ability to verify the conformity of a part with its documentation</td>
</tr>
<tr>
<td>LD2</td>
<td>The system has to enable the user to assure the provisions of a law regarding data security and protection of data privacy over the life cycle of the archives.</td>
</tr>
</tbody>
</table>
Table 2: Requirements of category “Legal Demand”

Goal of requirement LD1 is to ensure the verification of conformity of a part with the associated documentation. Requirement LD 2 ensures the ability of the long term archiving system to provide information about aspects of IT security. Throughout the IT Security the long term archiving system ensures the unchangeability of archived data during retention time and certified archiving and retrieval processes.

6.3 Security

The third requirement category describes the requirements belonging the aspects of security.

<table>
<thead>
<tr>
<th>Nr.</th>
<th>Requirement</th>
</tr>
</thead>
<tbody>
<tr>
<td>S1</td>
<td>There must be no possibility to destroy a document during it's scheduled lifetime.</td>
</tr>
<tr>
<td>S2</td>
<td>Exactly the document must be retrieved that was searched for.</td>
</tr>
<tr>
<td>S3</td>
<td>All actions that cause changes in organization and structure within the archives must be recorded so that the retrieval of the original status is possible.</td>
</tr>
<tr>
<td>S4</td>
<td>Preserve the ability of the documentation (PDM Structure + documents, physical documents, and geometric models) to be retrievable for &gt;50 years</td>
</tr>
<tr>
<td>S5</td>
<td>Provide evidence that the design was state of the art for &gt;50 years</td>
</tr>
<tr>
<td>S6</td>
<td>Guarantee the integrity of data for &gt;50 years</td>
</tr>
<tr>
<td>S7</td>
<td>Minimize the risk/ expenses/costs during change of system /release</td>
</tr>
<tr>
<td>S8</td>
<td>Electronic archives must be designed for migration to new platforms, media, software versions and components without loss of information.</td>
</tr>
<tr>
<td>S9</td>
<td>Each document has to be archived unchangeable.</td>
</tr>
<tr>
<td>S10</td>
<td>The information content of the document shall not be changed during the archiving, retrieval and migration processes</td>
</tr>
<tr>
<td>S11</td>
<td>Each document may be displayed and printed exactly in the form as recorded, this will fulfill the aspect of unchangeability of archived data</td>
</tr>
<tr>
<td>S12</td>
<td>No document may get lost on the way to the archives or into the archives themselves, this shall be ensured with appliance of certified processes</td>
</tr>
<tr>
<td>S13</td>
<td>The integration of the electronic signatures shall be possible. Evidence of security by secured process is also possible</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td>S14</td>
<td>Ensure the information is understandable by the Designated Community without the assistance of the information producers</td>
</tr>
<tr>
<td>S15</td>
<td>Logbook of the archiving process i.e. this offers a possibility to provide information about the secure doing of relevant processes in case of needed evidence</td>
</tr>
<tr>
<td>S16</td>
<td>Note information about collection management i.e. the information about archiving processes which includes more than one e.g. model, document or assemblies. Containment of the note are mostly information about references</td>
</tr>
<tr>
<td>S17</td>
<td>Note information about the management of rights.</td>
</tr>
</tbody>
</table>

Table 3: Requirements of category “Security”
7 Scenarios

The scenarios are developed on a further analysis of the requirements. They summarize requirements of similar content. The scenarios are defined with the goal of a modular structure for further process analysis to support a high degree of flexibility of the approach against project and company specific system architectures and business processes.

The scenarios are categorized as:

- Start archiving
  - Archiving digital documents
  - Archiving Cax data
  - Archiving PDM data
  - Archiving PDM Structure
- Validation
- Retrieval of archived data
- Administration of archived data

According to business process analysis the scenarios were founded on the OAIS system architecture.

7.1 Start archiving process

The main process is triggered by the role producer. Important for the scenario is the main distinction of archiving digital and analogue data. Archiving analogue data is out of scope for this document. The approach shall support an immediate archiving during the release process as well as archiving of data which were released times before.
Scenario 1: Start archiving process

7.1.1 Archiving digital documents

This scenario describes the processes needed to archive digital data. In scope of archiving process are digital documents like PDM or CAx documents. To ensure the archiving process the documents shall be validated (e.g. validation of the process of conversion from native into STEP format or other data migration), a digital signature shall ensure the status of unchangeability, to reduce the amount of data the documents shall be compressed after validation and signature. With aspects of archiving process quality the entire scenario shall be logged.
Scenario 2: Archiving digital documents

7.1.2 Archiving CAx Data

This scenario defines the scope of CAx data for the LOTAR archive. In scope are digital 2D and 3D CAD models. The 3D CAD data include information about electric harness, 3D Part and 3D assembly models. Out of scope are for the LOTAR archive CAe data, which may be described in any future specifications.

Scenario 3: Archiving CAx Documents
7.1.3 Archiving PDM data

This scenario defines the scope of PDM data. Necessary for a correct archiving process are information about PDM structure, PDM part and PDM baseline. PDM baseline means a grouping and archiving of PDM data (e.g. documents or objects) in a single package for a special situation or in a specific context (e.g. all data, which were part of a specific DMU milestone or type design certification). An advantage to archive PDM baseline is that the producer is able to recover the environment information in case of product liability. To archive PDM data means in detail to archive the part data as well as the part structure.

Scenario 4: Archiving PDM data

7.1.4 Archiving PDM structure

This scenario supports the archiving of various relationships of PDM parts.
Scenario 5: Archiving PDM structure

7.2 Validation

The scenario validation describes the processes needed to ensure the consistent data quality of the content information to be archived. The validation shall take place within the ingest and retrieval process after each conversion of data from native format into STEP format and vice versa. The favoured validation process shall be an automatic validation. The process will be triggered by the archiving system and shall include a selection of validation method. The validation process and results shall be logged. In case of an erroneous validation a manual validation shall take place. The manual validation will be triggered by the role producer and proceed with e.g. a viewing tool.
Scenario 6: Validation

7.3 Retrieval of archived data

The scenario describes the processes needed to ensure the quality for requested data. The process will be triggered by the role consumer. The consumer will be supported by the archive through functionalities like search and navigation. The consumer is able to preselect the requested data. The retrieval process includes the decompression of archived data and the validation of following reconversion process (STEP format into current or future CAD / PDM format). The scenario shall also include a data authenticity check and data integrity assurance. This functionalities ensure the status of unchangeability of data.
Scenario 7: Retrieval of archived data

### 7.4 Administration of archived data

The scenario administration is currently out of scope of this document. However, the administration processes are already described in detail within OAIS. Therefore it is planned to describe scenario administration more in detail within the next LOTAR specification.
8 General approaches for long-term archiving

In principle the archiving and retrieval processes can be associated to the mechanisms of data exchange between a sender and a receiver. The data creating application sends these data to an archive. The archive plays the role of a receiver which has to support the interpretation of this data without loss and in semantically correct way. During the retrieval process the archive plays the role of the sender. In this case the process has to be flexible against changing and unknown future receiving systems, functionalities and environments.

As an example contracted data exchange guidelines define the requirements for data type, format and quality of product data to be enabled by a supplier. The same “contracts” are defined within the OAIS as submission and dissemination agreements between the archive and the designated community (Producer and Consumer).

With this assumption experiences of exchanging 3D geometry and PDM data between customers and suppliers can be adopted for the approach.

8.1 Static aspects

Depending on the data format which will be archived, different concepts for the realization are the result. The possible data formats are:

- native format
- neutral format
- multi format

8.1.1 Native format

The native format is used for data exchange between equal systems without a conversion process.

Advantages may be for example:

- no conversion into another data format is necessary
- usually there are no difficulties during import and export processes
- it is possible to use specific functionalities of CAD- and PDM – Systems, i.e. functionalities which are only available in these systems

Disadvantages can be for example
- the data exchange is restricted to the same type of systems and usually with the same type of customization i.e. a data exchange between different systems or different customizations is usually not successful
- the change of hard- or software has to be realized by both exchange partners
- actually it is not possible to support a converting chain, which enables a certified process of retention after many change cycles of native data formats and systems
- with the actual experiences the re-use of native data in later releases of the involved applications is very problematic and requires a re-issuing / release of all archived data (e.g. migration from Catia V4 to V5)

8.1.2 Neutral format

The neutral format is used for data exchange between different systems and customizations with the conversion of data from native format of the sender into the neutral format and possibly from the neutral to native format of the receiver.

Advantages are:

- enables the definition of common core standard for archiving systems and data
- the neutral format concept decrease the dependency from the involved sending and receiving systems
- increases the data longevity enormously and is therefore applicable for long term archiving
- supports the re-use of definitions, tools and components of the archiving architecture
- changes at the e.g. sending system usually does not mean that equivalent changes at the receiving system has to be implemented
- changes of data at sending site does not lead to changes of existing data at receiver site (e.g., it is the only possibility to save a digital signature for existing data)
- therefore efforts and costs in changing functionality of the exchange process will be restricted in most cases to one site, sender or receiver

Disadvantage:
for each system, which participates at the data exchange process a specific converter (processor) is necessary.

The recommended neutral standard used within the LOTAR will be STEP. For the realization of an archiving system it is necessary to extract the relevant scope of needed data. The preferred candidate for the data backbone of LOTAR is the Application Protocol ISO 10303-214 (AP214). It supports the substantial data related to the scenarios as addressed in chapter 7.

8.1.3 Multi format

The multi format is means a combination of different formats. Therefore both aspects as described in chapters 8.1.1 and 8.1.2 are relevant. It enables the possibility of combining the advantages of neutral format of long term retention requirements and the higher functionality of native formats for short term aspects.

8.2 Dynamic aspects

8.2.1 Data

The major dynamic aspect arises from the frequent changes of supported functionality and formats the involved applications enable. Therefore the LOTAR archiving architecture will be based on the neutral data format STEP. Even though the standard itself is not a 100% static concept, the LOTAR project group considers that a mature and stable kernel of the standard exists, which is applicable for a long time of retention.

To ensure the readability of archived data it is necessary that future applications, processors and interfaces, which will read and write archiving data shall be downward compatible. This includes the functionality to convert data from / to STEP semantics and format.

8.2.2 System architecture

The planned system architecture will be also based on a dynamic concept. No matter how well LOTAR maintains its current configuration, it may be necessary from time to time to migrate definitions and data to different media and/or to different hardware or software environments.

However, it is usually not recommended to migrate the archived data itself, otherwise a new release of the derived data representation including a new digital signature is mandatory. The LOTAR project group recommends not to plan such a data migration due to economic (costs and effort) and product liability aspects.
Digital migration defined was within the OAIS to be the transfer of digital information, while intending to preserve it. Major motivators for a digital migration may be:

- **Improved Cost-Effectiveness:** The rapid pace of hardware (e.g., disk/tape drives) and software evolution provides greatly increasing storage capacities and transfer bandwidths at reducing costs. It also drives the obsolescence of some media types well before they have time to decay. In addition, improved data packaging designs may be less dependent on underlying media and supporting systems, and therefore simplified migration efforts may be recognized. To remain cost-effective, LOTAR must take advantage of these technologies. Depending on the particular technologies involved, the information may have to be moved to new media types not previously supported and it may have to revise its implementations to take advantage of the new technologies.

- **New Consumer-Service Requirements:** The Consumers of LOTAR also experience the benefits of new technologies and consequently raise their expectations of the types and levels of service they expect from LOTAR. These increased services may require new forms of DIPs (see 9.2.3) to service particular Designated Communities, which in turn may drive LOTAR to hold new forms of AIPs to reduce output conversions. Additionally, AIPs typically go through popularity swings and LOTAR may need to provide different levels of access performance to meet Consumer demands over time. This is likely to be satisfied by moving some AIPs to different media that provide increased or decreased levels of access performance.

- **Media Decay:** Digital media, over time, become increasingly unreliable as secure preservers of bits. Even those that are used with some level of error correction eventually need to be replaced. The net result of media decay is that AIP information must be moved to newer media.

Finally, the Designated Community for a provided data may be broadened, resulting in the need to revise requested data in forms so as to be understandable and usable by this broader community. This issue shall be solved by improved and certified converters from AIP to DIP instead of a data migration within the archive as described above.

Based on the description of the different concepts the LOTAR project group avowed that a dynamic concept shall be the grounding for LOTAR.
9 Strategic recommendations for long-term archiving

9.1 General Strategy

During analysis of requirements for the archive and the possible system architecture the LOTAR project group recognized that it is necessary to develop an archiving system which considers aspects of an open framework concept. The architecture shall be based on a stable kernel. This means that a realization of LOTAR shall be independent as long as possible from structures within companies, projects and environments.

The main goal for the project group is to define the relevant processes which are needed so that archiving of data accomplish the substantial requirements of acceptance, legal demands and IT-security. The physical implementation of LOTAR and its relevant processes belongs to the respective companies / administrators responsibilities.

Due to the independence of LOTAR to existing structures within e.g. companies it is important that the management has to decide about how to realize a possible archive. The project group acquired three potential concepts to realize an archive:

- as a part function within a PDM backbone system
- a stand alone archiving system
- a mixed system environment with a distribution of archival and retrieval functions into both, may include a leading system

A further decision during a realization is necessary about organisational criteria. The management shall define if the long term archiving process responsibility is defined as a central attendance (including the responsibility for correct proceeding, date and format of archiving) or if e.g. each design domain is responsible for date, procedure and format of data to be archived.

9.2 Proposal of a technical solution

9.2.1 Recommendations for different roles

Several roles are defined. The definition of the roles are based on the OAIS. These roles are fundamental for further detailed definition of processes and specifications for implementation.

9.2.1.1 Producer

A producer is the role played by those persons (designer or engineer…), which provide the information to be preserved. The producer is responsible for the integrity, completeness and consistency of the data to be archived.
9.2.1.2 Consumer

Consumer is the role played by those persons, that interacts with services of the archive to find and acquire preserved information of interest. The consumer starts a request and is not allowed to change any archived data.

9.2.1.3 Administrator

The role administration performs the administrative functions like, user management, preservation planning, consistency, availability, performance of the archive, removal of planned end data and procedure of tool and hardware migrations.

9.2.1.4 Management

The role management is the role played by those who set overall archiving policy as one component in a broader policy domain.

9.2.2 Processes

The LOTAR project group strongly recommends an implementation of the processes belonging to the subject areas ingest (and archiving), dissemination and removal according to OAIS. The processes shall represent the first level description for audits aiming for data security, quality assurance.

Furthermore, reporting mechanisms are required which give the former defined roles the feedback about pass or failure of ingest (and archiving), dissemination and removal.

Chapter 9.2.3 gives a description of data packages (AIP, SIP and DIP) according to OAIS.

9.2.2.1 Ingest and Archiving Process

Figure 9 gives an overview about ingest of data into the archive and the associated quality procedures as part of an approval process.
The start of process is triggered by the roles producer or administrator. The approval request will be submitted to the archiving front end. The front end may be implemented within a CAD / PDM System, may be web client or be integrated within the archive itself.

After submitting the request a collection of relevant data, the Submission Information Package (SIP), will be transferred to the archive. For the ingest process it is important that the SIP shall conform to the criteria of data quality, checked by the quality assurance module. In a failure case the Ingest process will be interrupted, the producer and/or administrator has to enhance the quality of application data in a following bug fix process.

After the successful quality check the content information of the Archival Information Package (AIP) will be generated by converting the SIP data into the neutral archiving format. A validation of the generated AIP will take place to ensure safeguard the transformation into AIP format. This validation can be automatically or manually performed, depending on the allegations. The administrator usually will perform an error handling in case of validation problems.

After successful validation the required meta data (descriptive information) will be added to the AIP and the resulting package is ready to store at the archiving media. A confirmation follows to document the success of the entire archiving process. Afterwards the setting of the approval status can be accepted.
9.2.2.2 Dissemination

Figure 10. defines the dissemination process to retrieve data by a consumer of the designated community.

The start of the process will be triggered by the consumer or administrator. An access control check prevents an unauthorized data retrieval. In addition it has to be checked if the consumer has permission rights to the requested data content. Due to security reasons reports about failure request should not be displayed. A single request may lead to numerous data, which will be collected automatically and offer them to the consumer for further selection. Afterwards the selection of a copy of AIP will be prepared for the dissemination. The resulting information will be submitted to the archive where the data collection will be verified within the “confirmation of selection” process.

In case of possible failures an error report will be sent to the participating role which has to perform an error handling. Thereafter the Dissemination Information Package (DIP) will be generated out of the AIP including meta data. The AIP will remain unchanged within the archive. The generation of DIP include also the conversion into the requested native data format.
In case of errors the archive will send an error report to the administrator who has to perform bug fix. If the validation is successful the archive will provide a report about the entire dissemination process.

9.2.2.3 Remove

The LOTAR project group recommends a possibility to remove archived data. Be a constrained process. The process is displayed within Figure 11.

![Figure 11: LOTAR Remove process](image)

The automatic remove request is based on the predefined planned end information within the archive. The interactive remove request of the manager (e.g. if requirements for retention periods will change in general) needs an access control and submission of the request.

Equal to the dissemination process an access filter shall be supported. The archive shall identify related data and collect them into the set of AIPs to be removed. But the process has to safeguard the consistency of the archive, i.e., data can not be submitted for removal if it is related to from other ones. Before deletion of the AIP content information the archive offers a selection list for final confirmation by the manager or administrator.

The LOTAR project group recommends to hold the descriptive information for documentation including the archive logging information. In the case of success a final confirmation will take place.
9.2.3 Data concept

The conceptual structure for supporting the long term archiving system is the information package according to OAIS. An information package contains two types of information: Containing information and preservation descriptive information.

There are three types of data packages defined within long term archiving system:

- Submission Information Packages (SIP).
  The SIP is the package that is sent to the long term archiving system. Its content will be defined by the producer, who initiates the forwarding of the package from the creation application to the long term archiving system. The SIP format will be defined by the long term archiving policies.

- Archival Information Package (AIP)
  The AIP is the result of a transformation process within the archive. The incoming SIP will be transferred into an AIP. The AIP is defined to provide a set of information that has all qualities needed for the permanent preservation. The LOTAR project group addresses only a 1:n relationship between an SIP and AIP in contrast to the OAIS system architecture (an AIP shall only contain a single SIP; one SIP may be split up into several AIPs)

- Dissemination Information Package (DIP)
  The dissemination package will be provided by long term archiving system as an answer to a search request by the role consumer or administrator. The package may include collections of AIPs.
Within the following specification the information packages will be described in detail. The specification will give a subset of recommendations about a consistent data model which includes at least the mandatory information to produce a successful data exchange between participating systems. The data model will be based on existing standards like ISO 10303 (STEP) AP214 and will reference further standards like OAIS. The detailed information, like use and constraints of mandatory attributes, will be displayed within a data dictionary.

9.2.4 System architecture

First rough definitions of the main functional modules are based on the results of process analysis and the determination of data information packages. The modules provide elementary services required for process chains mature for certification.

The functional modules are divided into the subject areas Ingest and Archiving, Archival Storage, Data Management, Access and Dissemination. Sub functions are defined additionally and titled in *Italic*. Detailed definitions are part of later specification.

The approach is further based on the idea to use standard tools of data exchange like STEP Processors, Viewing Tools, ENGDAT ENGPART- Package Information.

9.2.4.1 Ingest

This module provides the services and functions to accept SIP’s from Producers (or from internal processes under Administration control) and prepare the content for storage and management within the archive. Ingest and archiving functions include the functions as briefly described as follows:

**Provide data for retention**

This functionality supports a collection of relevant data which shall be part of the approval set of PDM data and related documents. It is mainly dependent on the business requirements and shall include at least the minimum of data which will be addressed later on with the specification.

**Quality assurance**

Quality assurance will take place before converting data into STEP format. The function will perform a check if the data within an SIP satisfies the demands of data quality as contracted between the archive and its customers (producer). With this check it is assured that incoming data correspond with the given guidelines. Furthermore, it aims for avoiding failure cases in the following archiving processes, especially conversion into the AIP format. If the SIP does not pass the check the producer has to change the data.

**Generate AIP function**
The AIP generation may result in a combination of one or more STEP files which implies content information from the PDM and CAD data of SIP.

Content information are represented by the data section of the Part 21 STEP file. Furthermore the STEP File may contain preservation and representation information within the header section. By identifying the applied application protocol of ISO 10303 the reference to the knowledge base is given.

The generation of the AIP will proceed in several process steps starting after quality assurance and ending with validation of the AIP.

**Generate descriptive information**

Within the AIP archive metadata are also required, i.e., descriptive information, packaging information and preservation information.

Descriptive information are required for retrieval of content within the archive database, e.g., part number and title, model number and title, project ID, system, status. The descriptive information may partly be derived from the content information of the SIP of the PDM system.

Preservation information are required for reuse of content information, e.g., file format (application protocol and conformance class of STEP file), identification of preprocessor (e.g. Theorem V1.2), and former format of CAD and PDM data (e.g. CATIA V4.2.2; Metaphase 3.2). In addition validation criteria are also included for reconversion within the dissemination quality assurance of the data files (e.g. Mass; centre of gravity in case of geometry).

Packaging information are required to pack and unpack the archived content data, e.g. file names, method of packaging (Zip- or TAR format) and compression rate.

**Validation**

The validation function is divided into two possible sub functions: automatic validation. and manual validation

An automatic validation (e.g. Mass; centre of gravity in case of geometry) will take place after conversion of data into the archive format. It will be feasible if a processor includes validation algorithms which are certified. An automatic validation may be combined with a manual once. However, if an automatic validation is not applicable the manual once including a separate setting of a digital signature is the only way to ensure the authenticity of the archived data. For the validation some error detection methods, such as checksum or cylindrical redundancy check, shall be used. The results shall be associated with the data.
9.2.4.2 Archival Storage

This subject area provides the services for the storage, maintenance and retrieval of AIP’s. Archival Storage functions include receiving storage requests of AIP’s from Ingest and adding them to permanent storage, managing the storage hierarchy, refreshing the media on which archive holdings are stored, performing routine and special error checking, providing disaster recovery capabilities, and providing AIP’s to Access to fulfill orders.

Within the Archival Storage function the following functionalities should be implemented.

**Coordinate update function**

The function is responsible for transferring the AIP’s to Archival Storage and the Descriptive Information to Data Management. A transfer of the AIP includes a storage request and may represent an electronic or a physical transfer. After the transfer is completed and verified, from Archival Storage function a storage confirmation must be generated and sent to Ingest function. The Coordinates Updates functions also merges the storage identification information with the Descriptive Information of the AIP’s. The merged information will be managed by the Data Management function.

Data Management updates may take place without a corresponding Archival Storage transfer when the SIP contains Descriptive Information for an AIP already in Archival Storage.

**Receive data function**

The Receive Data function manages the incoming storage request from Ingest and is responsible for the transfer of AIP’s to the archiving media. After selection of archiving devices the functions performs the physical transfer of the AIP’s. Following the successful transfer the receive data functions shall send a successful storage confirmation to Ingest.

**Manage Storage Hierarchy function**

The Manage Storage Hierarchy function positions, via commands, the contents of the AIP’s on the appropriate media based on storage management policies, operational statistics, or directions from Ingest via the storage request. It will also conform to any special levels of service required for the AIP, or any special security measures that are required, and ensures the appropriate level of protection for the AIP. These include on-line, off-line or near-line storage, required throughput rate, maximum allowed bit error rate, or special handling or backup procedures.

**Set Signature function**
The set signature function will mark the AIP’s with a digital signature to ensure the unchangeability of the data. The unchangeability of data is a strong requirement in case of product liability.

**Compression of data function**
The function compression of data will perform a data compression with standard methods like Zip, RAR. The compression will take place before the data will be transferred to archiving media by the Receive Data function.

**Replace Media function**
The Replace Media function provides the reproduction of archived data within stipulated periods or within a change of archiving technique. The function may perform several different strategies, like Refreshment, Repackaging and Replication.

**Error Checking**
This function provides statistically acceptable assurance that no components of the AIP are corrupted during any internal Archival Storage data transfer. The error checking function requires that all hardware and software within the archive provide notification of potential errors and that these errors are routed to standard error logs that are checked by the Archival Storage staff.

**Disaster Recovery**
This function provides a mechanism for duplicating the contents of LOTAR archive in a physically separate facility. The duplication of the data will be done by copying the archive contents to some form of removable storage media (e.g., digital linear tape, compact disc) or via hardware transport or network data transfers.

**Provide Data function**
The function provides the a copy of archived data after a consumer request to access. The request must be validated if the consumer has the necessary access right. After validation of the request and the transfer of the data the function shall send a confirmation about the successful provision.

9.2.4.3 Data Management
This subject area provides the services and functions for populating, maintaining, and accessing both Descriptive Information which identifies and documents archive holdings and administrative data used to manage the archive. Data Management function packages include administering the archive database functions (maintaining schema and view definitions, and referential integrity), performing database updates (loading new descriptive
information or archive administrative data), performing queries on the data management data to generate result sets, and producing reports from these result sets.

Within the Administer Database function packages following functionalities should be implemented.

**Administer Database function**

This function is responsible for maintaining the integrity of the Data management database, which could contain the Descriptive Information and system information. The system information is used to support archive operations. The function will provide the capability to create, maintain an access customized views of the data within LOTAR. The further functionalities shall be carried out in accordance with established policies (defined within the specification) of LOTAR.

**Perform Queries function**

The perform queries function receives a query request from Access and executes the query to generate a result set that is transmitted to the requester.

**Generate Report function**

The generate report function receives a report request from Ingest, Access or Administration and executes any queries or other processes necessary to generate the report that it supplies to the requester. Typical reports might include summaries of archive holdings by category, or usage statistics for accesses to archive holdings. It may also receive a report request from Access and provides descriptive information for a specific AIP.

9.2.4.4 Access and Dissemination

This subject area provides the services that support Consumers in determining the existence, description, location and availability of information stored in the LOTAR, and allowing Consumers to request and receive information. Access function packages include communicating with Consumers to receive requests, applying controls to limit access to specially protected information, coordinating the execution of requests to successful completion, generating responses (Dissemination Information Packages, result sets, reports) and delivering the responses to Consumers.

Within the Access and Dissemination functional package the following functions should be implemented.

**Coordinate Access Activities**

The coordinate activities function provides a single user interface to the information holdings of LOTAR. The interface can be a part of the PDM System or a special computer network.
Access rights control check
The access rights control check function supports a profile based control over actions of registered consumers, who sends requests from ingest to access. Within the profile the access rights of each consumer will be memorized. With a dissemination request under way the consumers access rights will be compared with the profile and in case the request will be denied.

Generate DIP
This function provides the conversion service which includes the data transformation from the archiving format into the future format. It might also contains a decompression of data.

Provide data for dissemination
This function provides a selection of requested data in case of multiple result during search activities by the Consumer / Administrator.

It may be useful to support the dissemination process by specific STEP tools (e.g. STEP PDM and 3D geometry viewer, browser), as part of the archiving architecture to enhance the readability of the associated representation information so that it is again readily understandable to the Designated Community. These tools may help the Designated Community to select and retrieve information for additional processing or reuse of the data.

For example such tools can support a representation of the assembly structure view of a complex product like an A380 section under the control of an independent (from the end user applications) archiving architecture, without retrieving and converting the full associated geometry. Based on the Provide Data for Dissemination function the consumer can confirm or not the dissemination or may precise the selection.

Copy AIP
This function retrieves the AIP from Archival Storage, and moves a copy of the data to a staging area for further processing.

9.2.4.5 Preservation Planning
This entity provides the services and functions for monitoring the environment of the LOTAR archive and providing recommendations to ensure that the information stored in the LOTAR archive remains accessible for consumers over the long term. The accessibility shall be ensured even, if the former hardware becomes obsolete. The functionalities contains also the migration of new archive standards and policies for LOTAR.
9.2.4.6 Administration

The administration of LOTAR processes describe the functionality of the archive itself. Because of the fact that the system architecture of LOTAR is based on the OAIS reference model the description of processes is out of scope within the White Paper and will be described in detail within the specification.

9.3 Administrative aspects

This White Paper gives an overview about the motivation, needs, problems and possible solutions for a realization of a long term archiving system. The White Paper can only contain a first approach of a description of basic processes and procedures needed for an archiving system and shall not describe all information in detail. A necessary following specification shall specify the system architecture, mapping recommendations for the data conversion from native data format into STEP, the processes within the archive and recommendations for the realization concerning hardware and archive infrastructure.

The White Paper cannot give information about the cost of a realization because of different existing archiving policies and hard-/software infrastructure within the companies. The realization of a new archiving system is an expensive process which will follow large scale changes within the companies, e.g. procedures of workflow will change, employees have to instruct with the new recommendations, new concepts for the administration of the archive will be realized and harmonization procedures between CUSTOMER and supplier depending on access rights concept shall take place.

An identified cost driver will be the dynamic concept of the LOTAR recommendations. The slow but steady migration of new hardware and the depending converting programs are necessary to ensure the readability of the archived data.

The risks of realization of an archive system (data loss, unreadability of data…) will bear the companies. So they have to decide which kind of strategy they will realize for the archive system.
References

Regulations

**Unified Modeling Language, v1.4:** A specification defining a graphical language for visualizing, specifying, constructing, and documenting the artifacts of distributed object systems.

**MOSLA - Maturity of Standard for Long term CAD Data Archiving:** MOSLA is a standard for long term archiving of 2-D and 3-D CAD data, based on AP203, AP202 and Part 21 of STEP standard

**JAR21.139:** Part 21 - Certification and airworthiness requirements for aircrafts and parts. 139 – Subpart G Production Certification / Quality system

**OCLC/RLG Preservating data project group:** tasked with examining current practice in the use of preservation metadata, and developing a comprehensive preservation metadata framework applicable to a broad range of digital preservation activities

**BoingMDB:** – Data format for multi database

**VOI - AIT/7704:** EU project for advanced information technology in design and manufacturing

**US Federal Regulations (CFR)**

Normative


**ISO 10303-21:** 1994, Industrial automation systems and integration - Product data representation and exchange - Part 21: Implementation methods: Clear text encoding of the exchange structure


**ISO 14721.2 (DIS): OAIS - Open Archiving Information Model – Reference Model**
Informative

XML 1.0: EXtensible Markup Language; W3C-Recommendation 10-February-1998

XMDT: EXchange and Management of Technical Data