|  |  |
| --- | --- |
| © ISO 2014 – All rights reserved    **ISO/TC 211 N**  Date: 2016-01-~~03~~17  **ISO/WD 19165**  ISO/TC 211/WG 7  Secretariat : SN    **Geographic information – Preservation of digital data and metadata**  Information géographique – Archivage des données numériques et des métadonnées | |
| **Warning**  This document is not an ISO International Standard. It is distributed for review and comment. It is subject to change without notice and may not be referred to as an International Standard.  Recipients of this document are invited to submit, with their comments, notification of any relevant patent rights of which they are aware and to provide supporting documentation. |  |
|  |

|  |
| --- |
| **Copyright notice**  This ISO document is a working draft or committee draft and is copyright-protected by ISO. While the reproduction of working drafts or committee drafts in any form for use by participants in the ISO standards development process is permitted without prior permission from ISO, neither this document nor any extract from it may be reproduced, stored or transmitted in any form for any other purpose without prior written permission from ISO.  Requests for permission to reproduce this document for the purpose of selling it should be addressed as shown below or to ISO's member body in the country of the requester:  [Indicate the full address, telephone number, fax number, telex number, and electronic mail address, as appropriate, of the Copyright Manger of the ISO member body responsible for the secretariat of the TC or SC within the framework of which the working document has been prepared.]  Reproduction for sales purposes may be subject to royalty payments or a licensing agreement.  Violators may be prosecuted. |

Contents

Contents

Introduction

1. Scope

1. Conformance

1. Normative references

1. Terms and definitions

1. Symbols and abbreviated terms

1. Preservation
   1. Prioritization
   2. Structure
   3. Rights
   4. Time

1. Geospatial information package
   1. Overview
   2. Designated community
   3. Metadata
   4. Open Packaging Convention (OPC)

Annex A: Abstract test suite (normative)

Annex B: Summary of ISO 14721 (informative)

* 1. Scope
  2. OAIS model
  3. OAIS information
  4. OAIS responsibilities
  5. OAIS preservation perspectives
  6. Digital migration

Annex C: Requirements for data management regarding preservation (normative)

Annex D: Case-specific archival concept (informative)

Annex E: Functional requirements for a preservation archive (normative)

Bibliography

Foreword

ISO (the International Organization for Standardization) is a worldwide federation of national standards bodies (ISO member bodies). The work of preparing International Standards is normally carried out through ISO technical committees. Each member body interested in a subject for which, a technical committee has been established has the right to be represented on that committee. International organizations, governmental and nongovernmental, in liaison with ISO, also take part in the work. ISO collaborates closely with the International Electrotechnical Commission (IEC) on all matters of electrotechnical standardization.

International Standards are drafted in accordance with the rules given in the ISO/IEC Directives, Part 3.

The main task of technical committees is to prepare International Standards. Draft International Standards adopted by the technical committees are circulated to the member bodies for voting. Publication as an International Standard requires approval by at least 75 % of the member bodies casting a vote.

Attention is drawn to the possibility that some of the elements of this International Standard may be the subject of patent rights. ISO shall not be held responsible for identifying any or all such patent rights.

ISO 19165 was prepared by Technical Committee ISO/TC 211, *Geographic information / Geomatics*.

# Introduction

Today’s information is mostly stored on digital media. For several reasons the lifetime of those media is shorter than that of analogue media. Unless systematically archived the storage media will decay and thus the information lost. Digital information is also lost when due to other reasons such as loss of metadata describing the format of data. Unfortunately, this daunting scenario often exists. Consequently, the epoch in which we presently live is sometimes named the “Digital Dark Age”.

Traditional archives are understood as facilities or organizations which preserve records, originally generated by or for a government organization, institution, or corporation, for access by public or private communities. The archive accomplishes this task by taking ownership of the records, ensuring that they are understandable to the accessing community, and managing them so as to preserve their information content, data integrity and authenticity (ISO 16363/TDR). The major focus for preserving this information has been to ensure that they are on media with long term stability and that access to this media is carefully controlled (ISO 14721:2012).

Geospatial data possesses several structural peculiarities. They include:

* relation to a well-defined section of the Earth
* exchange by using theme-specific and sophisticated exchange formats
* link to thematic data (data bases)
* transformation between different coordinate reference systems
* visualization (map output)
* large data volumes
* existence of several levels-of-detail of the same dataset
* link between a geospatial dataset and rights

These structural peculiarities imply that geospatial data shall be preserved together with metadata content information that fully accounts for their structural peculiarities.

The ISO 14721:2012 “Space data and information transfer systems – Open archival information system (OAIS) – Reference model” defines a reference model for the archiving digital information. Despite being part of the “Space data and information transfer system” standard series ~~it has no focus on~~its application is not limited to space data and it is widely used by digital libraries. Regarding geospatial data in general however, the ISO 14721:2012 does not ~~satisfy~~ cover all the needs for digital data and metadata preservation. Therefore, the ISO 19165 addresses geospatial data, its data model structures, the multiplicity of data formats, and property rights. This standard is demanded by and developed for the geospatial community, and thus the OAIS preservation description information and descriptive information (described in the ISO 14721 section 4.2 Information model) is modelled as specialization of the ISO 19115-1 “Geographic information – Metadata – Part 1: Fundamentals”.

**Geographic information – Preservation of digital data and metadata**

## 1 Scope

This standard sets the rules for the long-term preservation of digital geospatial data. These data include metadata and other ancillary data that are necessary to fully understand and rebuild the archived digital environment.

Geospatial data are preserved as a geospatial archival information package. This standard defines its details. A geospatial archival information package shall be fully self-explanatory and shall allow a future reconstruction of the dataset without external documentation.

This standard refers to ~~existing~~ characteristics of data formats that ~~shall be used~~are useful for the purpose of archiving.

This standard has a close relation to the standards developed by the ISO/TC 211

“Geographic information / Geomatics” and to archival standards of ISO such as the ISO 14721 “Open archival information system (OAIS) – Reference model”.

**Komentarz [wk1]:**  Comment from Hampapuram Ramapriyan, 24.12.2015: Change to “content items that capture the knowledge (e.g., software, documents, workflow)”

**Komentarz [wk2]:**  Comment from Hampapuram Ramapriyan, 24.12.2015: This needs some clarification – what does it mean to rebuild the archived digital environment?” Isn’t the goal to be able to fully understand, reuse and/or regenerate the data in the future?

**Komentarz [HR3]:** Changed the language here since the standard does not call for any specific existing formats. It rather specifies that well documented formats be used.

**2 Conformance**

Details of the conformance classes are given in the Abstract test suite in Annex A.

## 3 Normative references

ISO 14721:2012 “Space data and information transfer systems – Open archival information system (OAIS) – Reference model”

ISO 19108:2002 “Geographic information – Temporal schema”

ISO 19115-1:2014 “Geographic information – Metadata – Part 1: Fundamentals”

ISO 19115-2:2009 “Geographic information – Metadata – Part 2: Extensions for imagery and gridded data”

ISO 19115-3 “Geographic information – Metadata – XML schema implementation of metadata fundamentals” (not yet published)

ISO 19157:2013 “Geographic information – Data quality”

ISO 19157-2 “Geographic information – Data quality – Part 2: XML schema implementation” (not yet published)

ISO/IEC 29500-2:2012 “Information technology – Document description and processing languages – Office Open XML File Formats – Part 2: Open Packaging Conventions

## 4 Terms and definitions

### 4.1 access rights information

The information that identifies the access restrictions pertaining to the content information, including the legal framework, licensing terms, and access control. It contains the access and distribution conditions stated within the submission agreement, related to both preservation (by the OAIS) and final usage (by the consumer). It also includes the specifications for the application of rights enforcement measures.

[SOURCE: ISO 14721:2012]

### 4.2 access software

A type of software that presents part of or all of the information content of an information object in forms understandable to humans or systems.

[SOURCE: ISO 14721:2012]

### 4.3 AIP edition

An AIP whose content information or preservation description information has been upgraded or improved with the intent not to preserve information, but to increase or improve it. An AIP edition is not considered to be the result of a digital migration.

[SOURCE: ISO 14721:2012]

### 4.4 AIP version

An AIP whose content information or preservation description information has undergone a transformation on a source AIP and is a candidate to replace the source AIP. An AIP version is considered to be the result of a digital migration.

[SOURCE: ISO 14721:2012]

### 4.5 archival information package (AIP)

An information package, consisting of the content information and the associated preservation description information (PDI), which is preserved within an OAIS.

[SOURCE: ISO 14721:2012]

### 4.5b content information

A set of information that is the original target of preservation or that includes part or all of that information.

Note 1 to entry: It is an information object composed of its content data object and its representation information.

### 4.6 data

A reinterpretable representation of information in a formalized manner suitable for communication, interpretation, or processing.

Note 1 to entry: Examples of data include a sequence of bits, a table of numbers, the characters on a page, the recording of sounds made by a person speaking, or a moon rock specimen.

[SOURCE: ISO 14721:2012, modified]

### 4.7 data dictionary

A formal repository of terms used to describe data.

[SOURCE: ISO 14721:2012]

### 4.8 data dissemination session

A delivery of media or a single telecommunications session that provides data to a consumer. The data dissemination session format/contents is based on a data model negotiated between the OAIS and the consumer in the request agreement. This data model identifies the logical constructs used by the OAIS and how they are represented on each media delivery or in the telecommunication session.

[SOURCE: ISO 14721:2012]

**Komentarz [HR4]:** Added “digital” here because “digital migration” is defined later in this section.

### 4.9 data submission session

A delivery of media or a single telecommunications session that provides data to an OAIS. The data submission session format/contents is based on a data model negotiated between the OAIS and the producer in the submission agreement. This data model identifies the logical constructs used by the producer and how they are represented on each media delivery or in the telecommunication session.

[SOURCE: ISO 14721:2012]

### 4.10 designated community

An identified group of potential consumers who should be able to understand a particular set of information. The designated community may be composed of multiple user communities. A designated community is defined by the archive and this definition may change over time.

[SOURCE: ISO 14721:2012]

### 4.11 digital migration

The transfer of digital information, while intending to preserve it, within the OAIS. It is distinguished from transfers in general by three attributes:

* a focus on the preservation of the full information content that needs preservation;

* a perspective that the new archival implementation of the information is a replacement for the old; and

* an understanding that full control and responsibility over all aspects of the transfer resides with the OAIS.

[SOURCE: ISO 14721:2012]

### 4.12 digital object

An object composed of a set of bit sequences.

[SOURCE: ISO 14721:2012]

### 4.13 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.

[SOURCE: ISO 14721:2012]

### 4.14 federated archives

A group of archives that has agreed to provide access to their holdings via one or more common finding aids.

[SOURCE: ISO 14721:2012]

### 4.15 geographic information system

information system dealing with information concerning phenomena associated with location relative to the Earth

[SOURCE: ISO 19101:2002, 4.18]

### 4.16 information package

A logical container composed of optional content information and optional associated preservation description information. Associated with this information package is packaging information used to delimit and identify the content information and package description Information used to facilitate searches for the content information.

[SOURCE: ISO 14721:2012]

### 4.17 knowledge base

A set of information, incorporated by a person or system, that allows that person or system to understand received information.

[SOURCE: ISO 14721:2012]

### 4.18 long term

A period of time long enough for there to be concern about the impacts of changing technologies, including support for new media and data formats, and of a changing designated community, on the information being held in an OAIS. This period extends into the indefinite future.

[SOURCE: ISO 14721:2012]

### 4.19 long term preservation

The act of maintaining information, independently understandable by a designated community, and with evidence supporting its authenticity, over the long term.

[SOURCE: ISO 14721:2012]

### 4.20 management

The role played by those who set overall OAIS policy as one component in a broader policy domain, for example as part of a larger organization.

[SOURCE: ISO 14721:2012]

### 4.21 metadata

information about a resource

[SOURCE: ISO 19115-1:2014, 4.10]

### 4.22 open archival information system (OAIS)

An archive, consisting of an organization, which may be part of a larger organization, 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'. The term 'open' in OAIS is used to imply that this recommendation and future related recommendations and standards are developed in open forums, and it does not imply that access to the archive is unrestricted.

[SOURCE: ISO 14721:2012]

### 4.23 package description

The information intended for use by access aids.

[SOURCE: ISO 14721:2012]

### 4.24 packaging information

The information that is used to bind and identify the components of an information package. For example, it may be the ISO 9660 volume and directory information used on a CD-ROM to provide the content of several files containing content information and preservation description information.

[SOURCE: ISO 14721:2012]

### 4.25 preservation description information (PDI)

The information which is necessary for adequate preservation of the content information and which can be categorized as provenance, reference, fixity, context, and access rights Information.

[SOURCE: ISO 14721:2012]

### 4.26 producer

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

[SOURCE: ISO 14721:2012]

### 4.27 provenance information

The information that documents the history of the content information. This information tells the origin or source of the content information, any changes that may have taken place since it was originated, and who has had custody of it since it was originated. The archive is responsible for creating and preserving provenance Information from the point of Ingest; however, earlier provenance information should be provided by the producer. Provenance information adds to the evidence to support authenticity.

[SOURCE: ISO 14721:2012]

### 4.28 reference information

The information that is used as an identifier for the content information. It also includes identifiers that allow outside systems to refer unambiguously to a particular content information. An example of reference information is an ISBN.

[SOURCE: ISO 14721:2012]

### 4.29 reference model

A framework for understanding significant relationships among the entities of some environment, and for the development of consistent standards or specifications supporting that environment. A reference model is based on a small number of unifying concepts and may be used as a basis for education and explaining standards to a non-specialist.

[SOURCE: ISO 14721:2012]

### 4.30 refreshment

A digital migration where the effect is to replace a media instance with a copy that is sufficiently exact that all archival storage hardware and software continues to run as before.

[SOURCE: ISO 14721:2012]

### 4.31 repackaging

A digital migration in which there is an alteration in the packaging information of the AIP.

[SOURCE: ISO 14721:2012]

### 4.32 replication

A digital migration where there is no change to the packaging information, the content information, and the PDI. The bits used to represent these information objects are preserved in the transfer to the same or new media instance.

[SOURCE: ISO 14721:2012]

### 4.33 representation information

The information that maps a data object into more meaningful concepts. An example of representation information for a bit sequence which is a FITS file might consist of the FITS standard which defines the format plus a dictionary which defines the meaning in the file of keywords which are not part of the standard. Another example is JPEG software which is used to render a JPEG file; rendering the JPEG file as bits is not very meaningful to humans but the software, which embodies an understanding of the JPEG standard, maps the bits into pixels which can then be rendered as an image for human viewing.

[SOURCE: ISO 14721:2012]

### 4.34 resolution (of imagery)

smallest distance between two uniformly illuminated objects that can be separately resolved in an image

Note 1 to entry: This definition refers to the spatial resolution.

Note 2 to entry: In the general case, the resolution determines the possibility to distinguish between distinct neighbouring features (objects).

Note 3 to entry: Resolution can also refer to the spectral and the temporal resolution.

[SOURCE: ISO/TS 19130-2:2014, 4.61, modified]

### 4.34a source

document providing legal and/or administrative facts on which the land administration (LA) object [right, restriction, responsibility, basic administrative unit, party, or spatial unit] is based

[SOURCE: ISO 19152]

### 4.34b spatial source

source with the spatial representation of one (part of) or more spatial units

[SOURCE: ISO 19152]

### 4.35 submission agreement

The agreement reached between an OAIS and the producer that specifies a data model, and any other arrangements needed, for the data submission session. This data model identifies format/contents and the logical constructs used by the producer and how they are represented on each media delivery or in a telecommunication session.

[SOURCE: ISO 14721:2012]

### 4.36 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.

[SOURCE: ISO 14721:2012]

### 4.37 transformation

A digital migration in which there is an alteration to the content information or PDI of an archival information package. For example, changing ASCII codes to UNICODE in a text document being preserved is a transformation.

[SOURCE: ISO 14721:2012]

## 5 Symbols and abbreviated terms

### 5.1 Abbreviated terms

|  |  |
| --- | --- |
| AIP | archival information package |
| CRS | coordinate reference system |
| DIP | dissemination information package |
| OAIS | open archival information system |

## 6 Preservation

### 6.1 Prioritization

The exponential extremely rapid growth of the data volumes prevents a full archival of all data. Consequently, only a selected subset can go for a long term archive. The selection strategy shall primarily evaluate the relevance of the data to be archived. Preservation shall be included in the product life cycle and demands a decision on the archival procedure at the moment the data are created.

The temporal classification shall define an archival category of 1 year, 10 years, 100 years, or more than 100 years. Each geospatial dataset shall be classified to one of those categories and then archived accordingly. The classifications may be reassessed before the end of the term defined for the category in which a given dataset has been placed.

An appraisal of every layer of a geographic information system is required, because not all layers are equally relevant. However, often layers are interdependent. Archival shall guarantee consistency among independent layers.

The layer’s relevance may be distinguished by the time, the function and the relation.

Examples:

Time: The following data layers have a temporal meaning only, and thus may not be relevant for archival: network of meters for fine particulates.

Function: A typical example is a future analysis of a land consolidation project. Though the administrative procedure will be fully completed after some years, the documents keep their values as a documentation of the change of landscape.

|  |  |
| --- | --- |
| OPC | Open Packaging Convention |
| PDI | preservation description information |
| SIP | submission information package |

Relation: Often geospatial information is related to several topics and data sources at the same time. An example is drinking water which may have a network given as vectors, a map layer in raster, and written documentation.

### 6.2 Structure

**6.2.1 Data format**

Today, all geospatial data are stored in commonly accepted specialized data formats. Those formats have a specific structure and include metadata. Some of the formats are standardized by ISO and/or IEC, others are de-facto standards.

A geospatial dataset shall be archived together with a full documentation of its data format. One of the key components of sustainable long term preservation is detailed knowledge of the file format that houses data. Some file types are well understood by the technical user communities (e.g. TIFF v6) while other less common file format types suffer from significantly less technical comprehension.

**Komentarz [HR5]:** “exponential” has a specific mathematical meaning, and I am not sure it applies to data. Sometimes the growth is linear and sometimes has large spurts of growth as step functions of time, etc.

**Komentarz [HR6]:** This provides some flexibility – e.g., if we decide a dataset is to be preserved for 10 years and the community needs change requiring the data to be preserved longer.

The documentation of the data format shall include the format structure, its properties, the metadata, and eventually a means of interpreting the data to access and interpret the data.

In all cases it is desirable to ensure that the primary file format is recorded in a consistent and comprehensive way, ideally linking via unique identifiers to well established file format registries.

In consideration of experience, cost, and resources, an implementation of this standard shall link to existing and acknowledged format-registries of which an example is PRONOM [13].

The use of a specified profile for any given file format is highly desirable knowledge to long term preservation, and as such it should be meticulously recorded, along with pointers towards any conformance and/or validation tools and methods used to assert the quality of the profiled file.

An analogue graphic representation of file content shall be archived where feasible.

**6.2.2 Data base**

Most of the geospatial data are object-structured and stored in data bases. In order to preserve this structure a simple storage of the data is not sufficient. In ~~any~~ many of the cases, the archival requires the software to access and interpret the data. This is named representation information. Consequently, the full data base content shall be transferred to the archive, which demands an archiving strategy that allows a persistent understanding of the technology for accessing this dataset.

Unless this is possible, the content of objects is modified and thus a decision about the future use of the archived data required before archiving.

**6.2.3 Properties of geospatial data**

Geospatial data have a number of particular properties which require consideration before archiving. If the data is to be archived an assumption about potential future uses of the data has to be made. Based on this assumption it shall be decided which of the properties and which of their details will be archived and which ~~other~~ others will be dropped.

If the data are finely grained, structured in objects or aggregated to larger units such as layers, it shall be decided before archival, which of the properties shall be preserved. Some geospatial data is created and maintained taking into account complex geospatial and topological relationships between elements. For such complex products there are dependencies and topological relationships that can ~~extent~~ exist between the layer units.

If the data require different ways of access, for instance in full resolution or in reduced resolution in the form of an image pyramid, it shall be decided before archival, which of the resolution levels of the data shall be preserved.

If the data exist in a redundant ~~data storage~~forms, for instance in a raster and in a vector format it shall be decided before archival, which of the storage types shall be preserved.

**6.2.4 Level of aggregation**

The appraisal of geospatial data ~~follows~~ may use ~~other~~ criteria ~~than~~ different from those for other types of data. Often the same geospatial data exist at several servers with different levels of aggregation, from raw data to an aggregated data product [11].

Consequently, only the raw data, i.e. the original data, and a description of the method of the creation of the derived product shall be archived. Only in exceptional cases aggregated products and levels of less finely grained details shall also be archived.

The service-oriented architecture (SOA) is widely used in the Internet. SOA services are separated modules of data and functions that are combined and reused for any production of applications [10]. Geospatial products that are the outcome of SOA can hardly be archived because they only temporarily exist in a user’s environment and never in the environment of the data producer.

For archiving, assumptions shall be made about typical geospatial data products, i.e. typical aggregations of the raw data. Based on the assumption the raw data shall be archived in a way that allows a full derivation of the data product.

**6.2.5 Gold copy**

This standard sets several methods for securely archiving geospatial data. However, the totality of all methods can never guarantee a full recovery of the data after a very long period of time. In order to increase reliability, a separate copy version of the 100 year data archive shall be established in open, file based repositories, not databases, nor other complex environments. Often, this copy is called a gold copy.

### 6.3 Rights

The problems of assuming sufficient control of the content information and preservation description information, when they are largely digital, are addressed in three related categories, as follows (ISO 14721:2012):

 copyright implications, intellectual property and other legal restrictions on use;  authority to modify representation information;  agreements with external organizations.

Authoritative geospatial data often possess legal restrictions. The rights are transferred to the archive together with the dataset, and their rights shall be guaranteed by the archive.

Regarding mapping products, this may play a role in the copyright, fees for the use of a dataset, restricted access to personal data, or for other applications. Rights including those imposed on archived data are often a function of time.

### 6.4 Time

Many geospatial data are never obsolete or are continuously updated such as cadastral data. These kinds of geospatial datasets never become mature for archiving, if the criterion for archiving is that the datasets do not change with time.

In order to allow for archiving these data one of the following methods shall be applied.

~~In the case that~~If a system contains data of the same topic and the same area but of different epochs, then the older data shall not be overwritten by new data.

**Komentarz [HR7]:** This is quite peculiar when we try to apply it to remotely sensed data that we handle. Almost every case becomes an exceptional case there because the datasets at we produce from the “raw data” from remote sensing instruments require considerable attention and quality assurance. The “higher level” data products are not simply aggregated, but are computed using complex algorithms – calibration (radiometric and geometric), geophysical parameter extraction, interpolation to standard spatiotemporal grids, etc. See level definitions [- http://science.nasa.gov/earthscience/earth-science-data/dataprocessing-levels-for-eosdis-dataproducts/.](http://science.nasa.gov/earth-science/earth-science-data/data-processing-levels-for-eosdis-data-products/)

**Komentarz [wk8]:** This sentence was written having vector data in mind. Obviously this is not appropriate for most of the raster data. I propose to remove this sentence.

If overridden data are not kept in the system, a time slice shall be defined in periodic intervals, in order to allow moving the complete dataset to the archive.

In order to save storage capacity, large volume data can be archived in longer intervals. It is essential to ensure that the requirements of the designated community is met by any sampling constraints, ensuring that any data collected is a comfortable fit with the expressed future use cases. Any variations of collection intervals should be recorded in the AIP documentation, ensuring that current users are informed of the limits of the data being archived, and future consumers of the data ~~are~~ will be well ~~sighted on~~aware of the ~~shape~~ contents of the archived dataset.

## 7 Geospatial information model

### 7.1 Overview

The ISO 14721 section 2.2 and section 4.2 describe an information model for data preservation. The core concept is the content information that is composed by a data object and the representation information. In the geospatial case, the data object will be the geospatial digital dataset to be preserved and the representation information is ~~that~~ what allows the designated community to understand the data. ~~In the~~For geospatial information the representation information is covered by the standards ISO 19115-1, ~~the~~ ISO 19157, ~~the~~ ISO19110, and ~~the~~ ISO 19165 (this standard). An example of representation information is the description of the dataset coordinate reference system (CRS).

The OAIS model defines also the information package as a conceptual container of two types of information called content information and preservation description information (PDI). The latter is all the information necessary to preserve the content information, to identify it in the archive, and to understand the environment in which the content information was created. Both pieces of information are packaged in information packages (IPs). See Annex B for details.

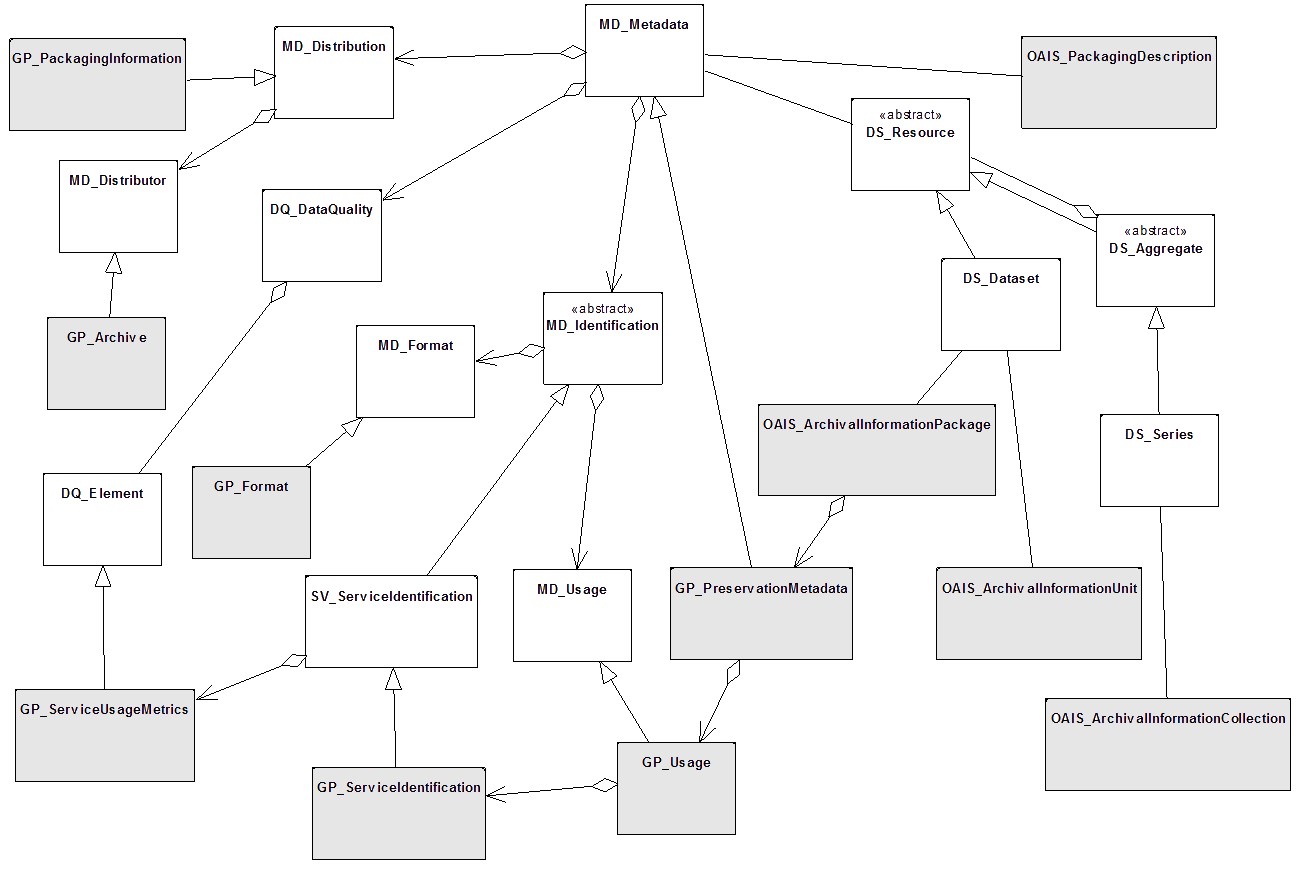
The ISO 19165 defines as one of its central components the elements of the archival information package (AIP). This package should be ready to be shared with other organizations, including those outside the geospatial community.

The ISO 19165 specializes versions of the IPs named Geo-SIP, Geo-AIP, and Geo-DIP (geospatial submission information package, geospatial archival information package, geospatial dissemination information package). Their special properties include, lossless compression, cartographic series support, i.e. a manageable regional size, and a container for information regarding geometry (vector and raster), attributes, topology, metadata, quicklooks, and recommendation on how to symbolize the data.

This standard maps the PDI of ISO 14721 into a specialization of the ISO 19115-1 metadata standard. The class diagram in Fig. 2 shows how MD\_Metadata has been extended into the GP\_PreservationMetadata (that acts as the PDI) by including some extra classes for the purpose of preservation. Not all classes of the preservation metadata are shown.

**Komentarz [HR9]:** Does this mean can be sampled for archiving? So, if a dataset was generated by making some measurements every day, does this imply that we may archive the measurements from every week or every month? If so, this does not apply universally to all the types of data addressed by this standard. A better way to state this might be to open the paragraph with: “Some organizations may choose to archive data by sampling them temporally in order to save storage capacity, if it is appropriate to do so. In such cases, it is essential to ensure …”

**Komentarz [wk10]:** In my eyes the proposed change of words would better express the idea of the sentence.



**Fig. 2:** Specialization of the ISO 19115-1 for the preservation of geospatial data and metadata.

The class diagram in Fig. 3 shows how DS\_Dataset and DS\_Series can be mapped to the OAIS model. In particular DS\_Dataset can be mapped directly to an OAIS archival information unit while the DS\_Series can be mapped in the OAIS archival information collection

**class Archival Information Package**

*MD\_Distribution*

**Archival Packaging Information::**

**GP\_PackagingInformation**

***Archival Information Package::***

***OAIS\_ArchivalInformationCollection***

***Archival Information Package::***

***OAIS\_ArchivalInformationUnit***

**GP\_PreservationMetadata**

***Archival Information Package::***

***OAIS\_ArchivalInformationPackage***

**Metadata Information::**

**MD\_Metadata**

«abstract»

***Archival Information Package::***

***OAIS\_PackageDescription***

«abstract»

***Archival Information Package::***

***OAIS\_PackagingInformation***

**Archival**

**Information**

**Package::**

**OAIS\_AccessAid**

*DS\_Resource*

**Metadata**

**application**

**information::**

**DS\_DataSet**

*DS\_Aggregate*

**Metadata**

**application**

**information::**

**DS\_Series**

1..\*

0..\*

containedIn

+

packagingInformation

+

1..\*

delimitedBy

+

+

identifies

+

containedIn

composedOf

+

0..\*

derivedFrom

+

describedBy

+

partOf

+

1..\*

0..\*

**Fig. 3:** Datasets and collections in geospatial data mapped to the OAIS model.

### 7.2 Designated community

A designated community is an identified group of potential customers (users) who should be able to understand a particular set of information (section 4). A designated community is defined by the archive. This definition shall include information about expected users, designated purposes, typical software product for processing the data, companies and government organisations that use the data, other data typically linked to the preserved dataset, and the region of the usage.

Further to the definition of a designated community it should be noted that it is not possible to imagine what a future use case of geospatial data looks like, and therefore to articulate the specific requirements of a future designated community. Accordingly, the present day designated community should be described in such a way to unambiguously describe the expected use and/or requirements of the dataset, to allow future consumers of any preserved dataset to comprehensively understand the limits, usage assumptions and constraints found in any given dataset.

Any newly expressed use cases encountered in the future shall be processed by the archive upon each periodic review of the dataset throughout its lifecycle, and any new metadata and/or data requirements folded back into the dataset and SIP construction.

### 7.3 Metadata

A geospatial dataset is always linked to a dataset of metadata. The metadata shall be archived in a way that allows an ~~undoubtable~~ unambiguous reference between both datasets. The definition of the metadata elements has to follow ISO 19115-1, ISO 19115-3, ISO 19157 and ISO 19110.

One of the most important metadata of geospatial data is a coordinate reference system

(CRS), because a full recovery of an archived dataset requires the CRS and its properties. The ~~archival~~ archive shall include information about the underlying coordinate reference system.

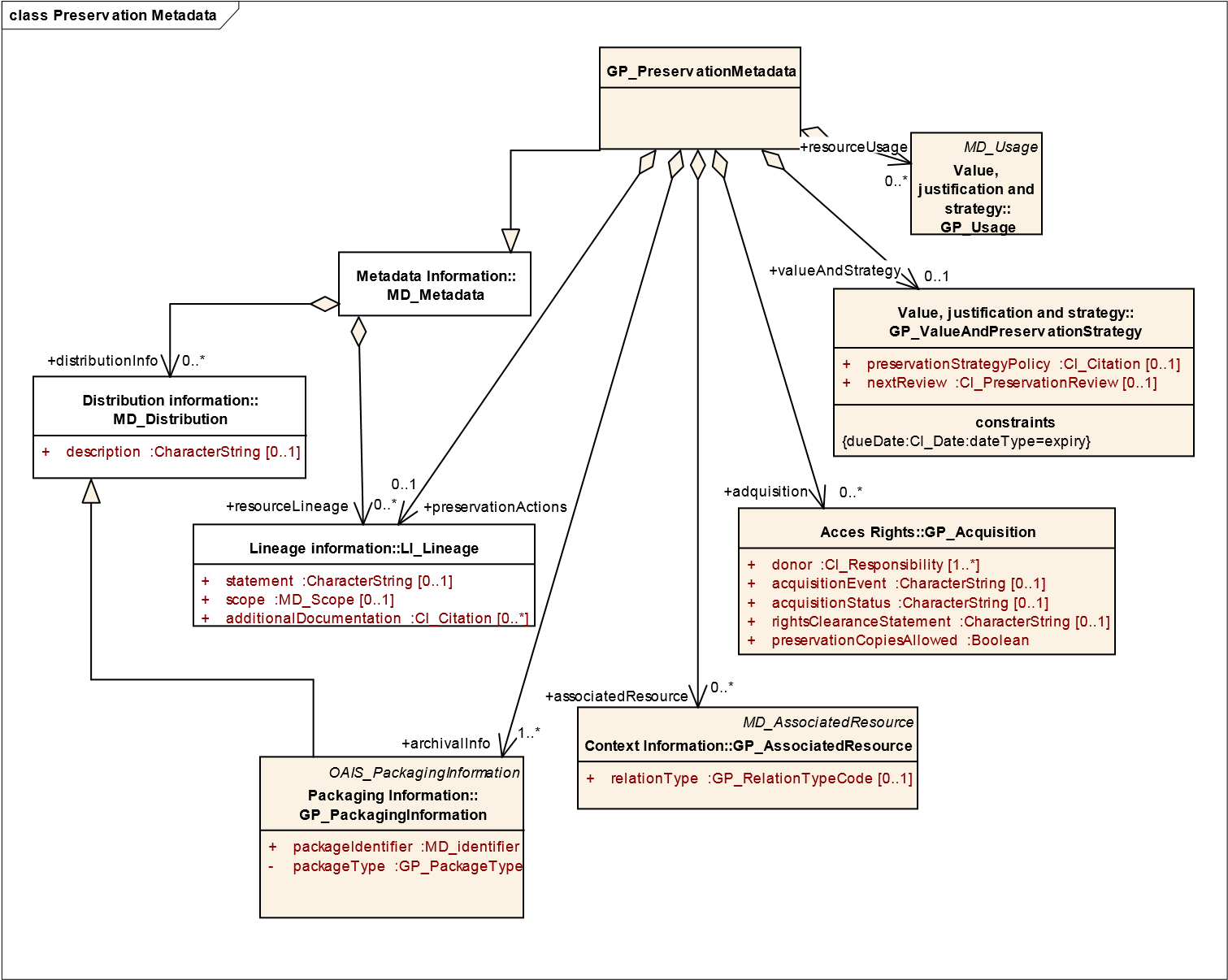
Most of the geospatial metadata corresponds to the OAIS concept of representation information. But preserving data for a future use requires extra effort on the metadata that corresponds to the preservation description information. This standard extends ISO 191151:2014 to allow preservation description information to be included.

Archival often requires the selection of certain data and leaving other data not be archived. In order to let a future user of the data understand the decisions made, the metadata shall include information about the assumptions made for the future use of the dataset when archival took place.

In preserving data, future users need to understand what they are working with (context information) and how the data were created (provenance information). Because most Earth science data involve complex physics and mathematics, the metadata shall include sufficient documentation (or pointers thereto) that provides the derivation of the algorithms used to generate the dataset. Likewise, the metadata shall include pointers to calibration data and ancillary data that were needed to produce the dataset. The specific content items needed to preserve the full provenance and context of the data and associated metadata depend on the individual disciplines and types of datasets (e.g., maps, remotely sensed data from satellites and airborne instruments, physical samples). Follow-up parts to this standard (ISO 19165) shall be developed detailing the content items appropriate to individual disciplines.

**7.3.1 Preservation Metadata classes**

The class diagram in Fig. 4 shows the class GP\_PreservationMetadata as a specialization of MD\_Metadata and the subclasses introduced in this standard.



**Fig. 4:** Class GP\_PreservationMetadata.

**Data identifiers**

The class diagram in Fig. 5 includes a specialization of MD\_Identification and emphasizes ~~on the importance of~~ the importance of including all resource identifiers that will help to consider this resource unique and avoid preserving it more than once accidentally. It also shows the current ISO 19115-1 capability of citing an image that can be used to link to a quicklook of the product or the legend giving clues to future users to symbolize the resource correctly.

**class Reference Information**

**Citation and responsible party information::CI\_Citation**

+

title :CharacterString

+

alternateTitle :CharacterString [0..\*]

+

date :CI\_Date [0..\*]

+

edition :CharacterString [0..1]

+

editionDate :DateTime [0..1]

+

identifier :MD\_Identifier [0..\*]

+

citedResponsibleParty :CI\_Responsibility [0..\*]

+

presentationForm :CI\_PresentationFormCode [0..\*]

+

series :CI\_Series [0..1]

+

otherCitationDetails :CharacterString [0..\*]

+

ISBN :CharacterString [0..1]

+

ISSN :CharacterString [0..1]

+

onlineResource :CI\_OnlineResource [0..\*]

+

graphic :MD\_BrowseGraphic [0..\*]

«DataType»

**Common classes::MD\_Identifier**

+

authority :CI\_Citation [0..1]

+

code :CharacterString

+

codeSpace :CharacterString [0..1]

+

version :CharacterString [0..1]

+

description :CharacterString [0..1]

**Metadata Information::**

**MD\_Metadata**

«abstract»

***Identification information::MD\_Identification***

+

citation :CI\_Citation

+

abstract :CharacterString

+

purpose :CharacterString [0..1]

+

credit :CharacterString [0..\*]

+

status :MD\_ProgressCode [0..\*]

+

pointOfContact :CI\_Responsibility [0..\*]

+

spatialRepresentationType :MD\_SpatialRepresentationTypeCode [0..\*]

+

spatialResolution :MD\_Resolution [0..\*]

+

temporalResolution :TM\_Duration [0..\*]

+

topicCategory :MD\_TopicCategoryCode [0..\*]

+

extent :EX\_Extent [0..\*]

+

additionalDocumentation :CI\_Citation [0..\*]

+

processingLevel :MD\_Identifier [0..1]

The general schema of ISO 19115-1 uses CI\_Citation in several places

to cite other "documents" or elements that complement metadata and

that eventually will be part of the preservation package.

This CI\_Citation that is shown here is the "citation" in

MD\_Indentification, and is deferent. Indeed, this is part of the description

of the data and is the citation "formula" for this data. The term

"identifier" is particularly important for preservation purposes and should

have all identifiers that in different contexts identify this resource. In this

way, it can contain Legal Registry numbers, Cartographic registry

numbers as well as DOIs or URIs.

**Common classes::MD\_BrowseGraphic**

+

fileName :CharacterString

+

fileDescription :CharacterString [0..1]

+

fileType :CharacterString [0..1]

+

imageConstraints :MD\_Constraints [0..\*]

+

linkage :CI\_OnlineResource [0..\*]

**GP\_Identification**

+

citationStatement :CharacterString [0..1]

+

reasonForCreation :CharacterString [0..1]

reasonForCreation is different form the

"purpose" due to some times there are

legal mandates reasons or data

"continuity of the temporal series"

reasons that are different from the

purpose of the data

+

graphicOverview

0..\*

identificationInfo

+

1..\*

**Fig.**

**5**

**:**

Class GP\_Identification

.

**7.3.3**

**Data, produc**

**t and format specifications**

**Komentarz [HR11]:**

It is not possible

to edit words in the figure directly, so

here are some corrections:

1.

ReasonForCreation is different from

“purpose” be

cause sometimes there

are legal mandates or requirements for

continuing temporal series. Such

reasons are different from the purpose

of the data, which may relate to specific

applications of the data.

2.

(

First sentence is OK). The

CI\_Citation shown here is

the “citation”

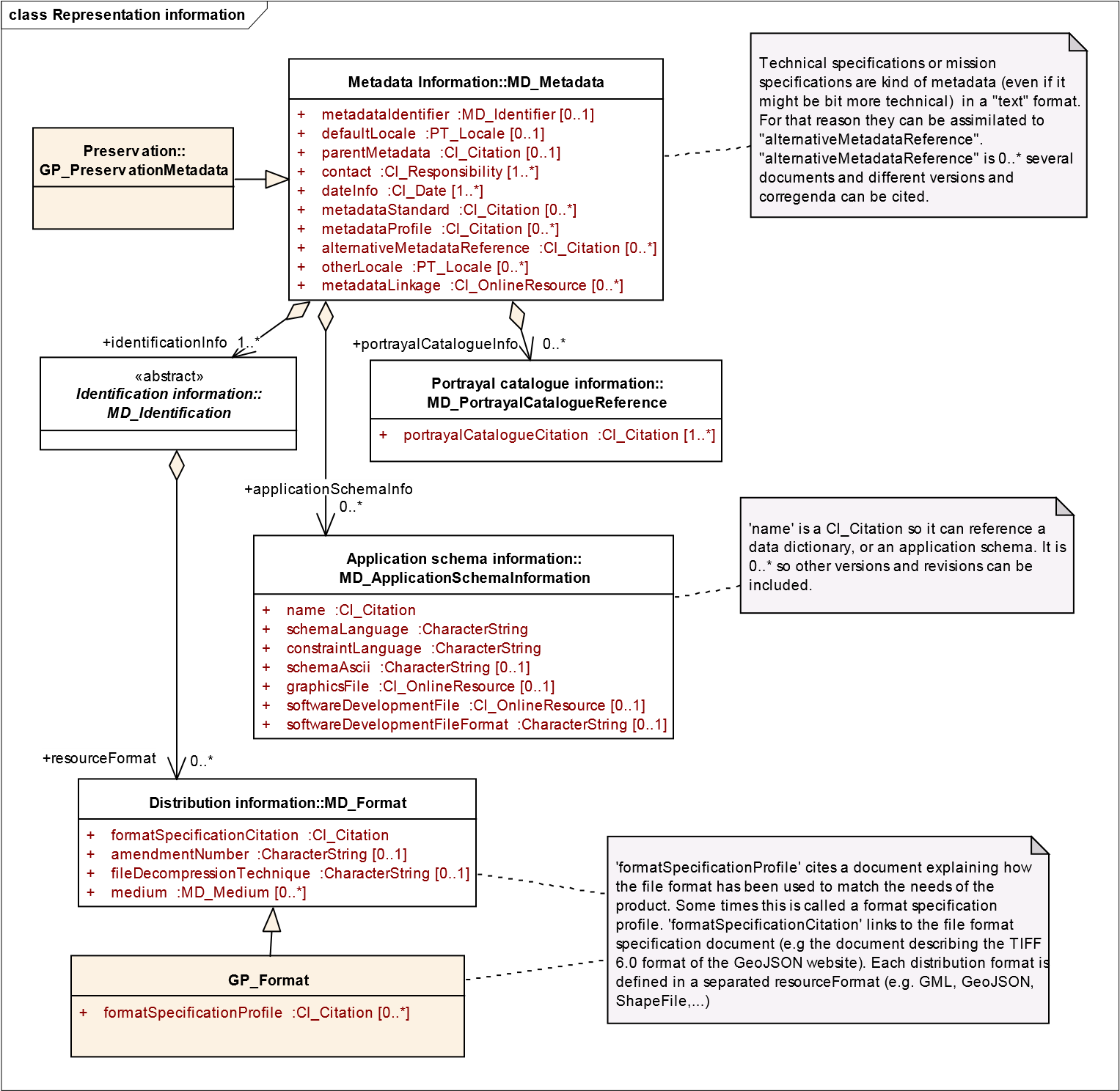
in MD\_Identification and is different.

(

Rest of the paragraph is OK

).

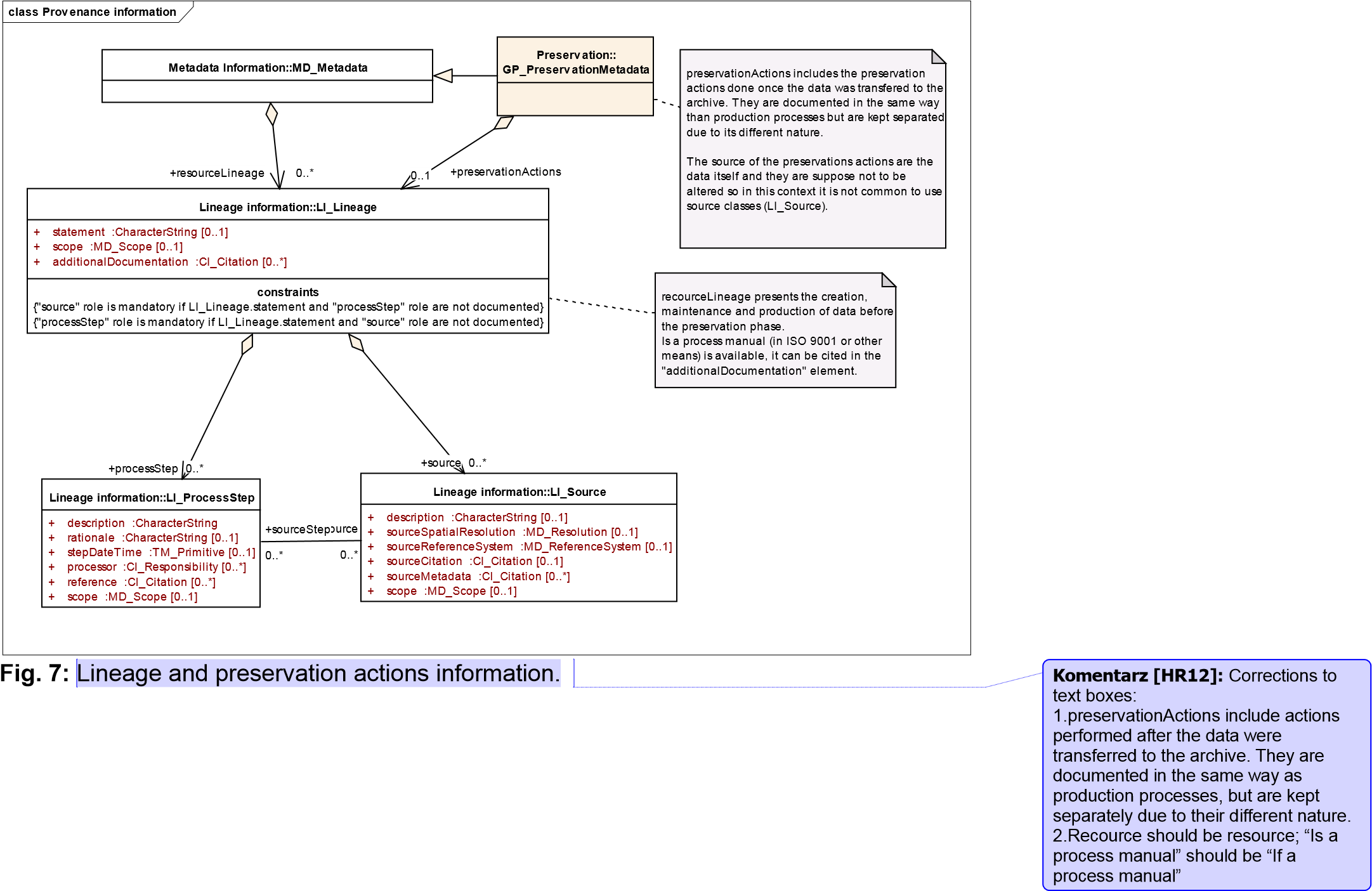
The class diagram in Fig. 6 shows how the information regarding the product specifications, data dictionary references, and common format specifications can be linked to the ISO 19115-1 and the need for a format specification profile specific to this product.



**Fig. 6:** Data, product and format specifications.

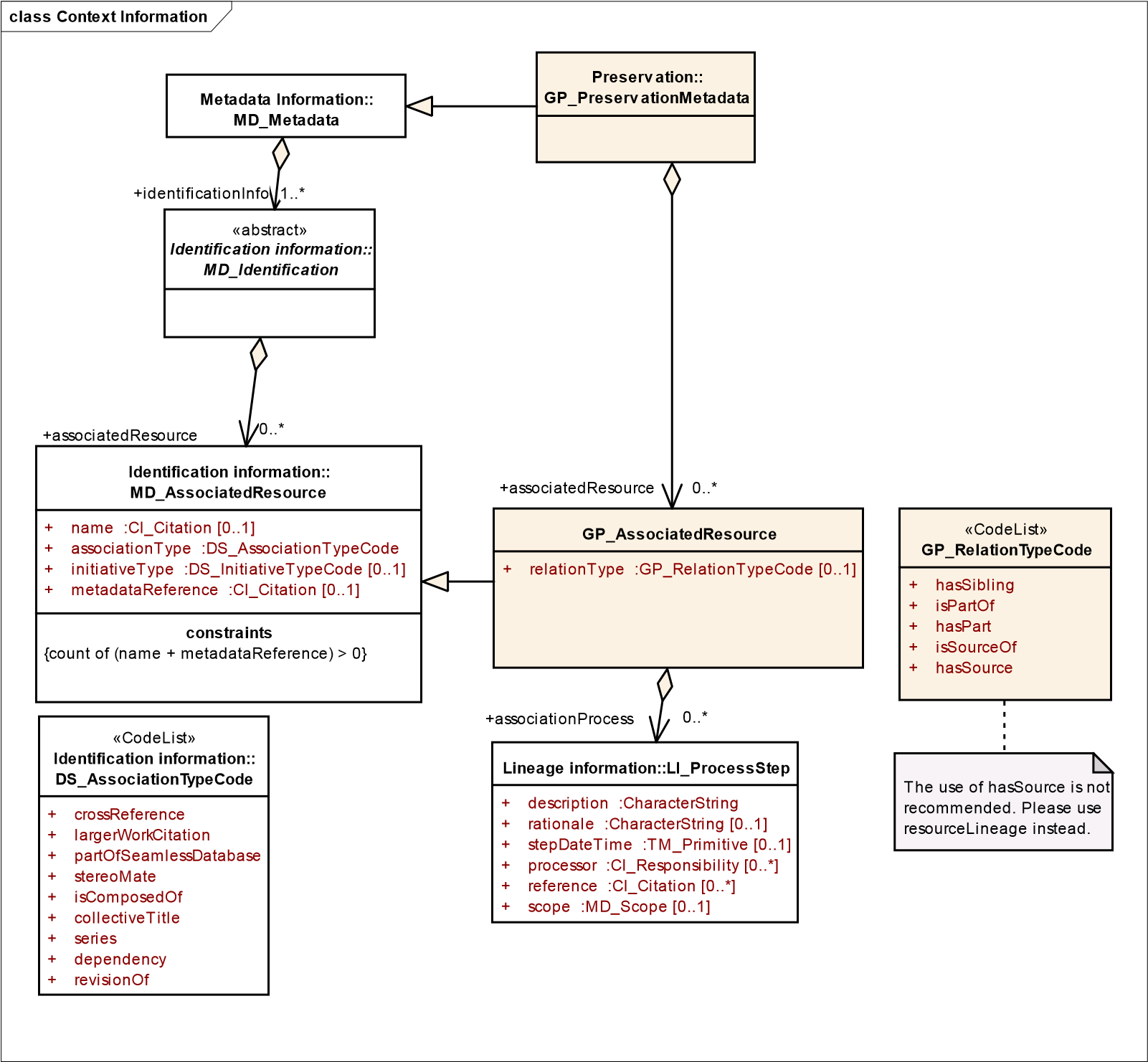
**Preservation actions**

Preservation actions ~~are processes that were done~~ are performed as a part of the curation process when the creation and maintenance of the resource has ended. They are often done by a responsible party different from the one that created or maintained the resource and have the objective of preserving the data (e.g. documenting a media migration). Fig 7 shows how can be encoded with the LI\_Lineage class.



**Association of resources**

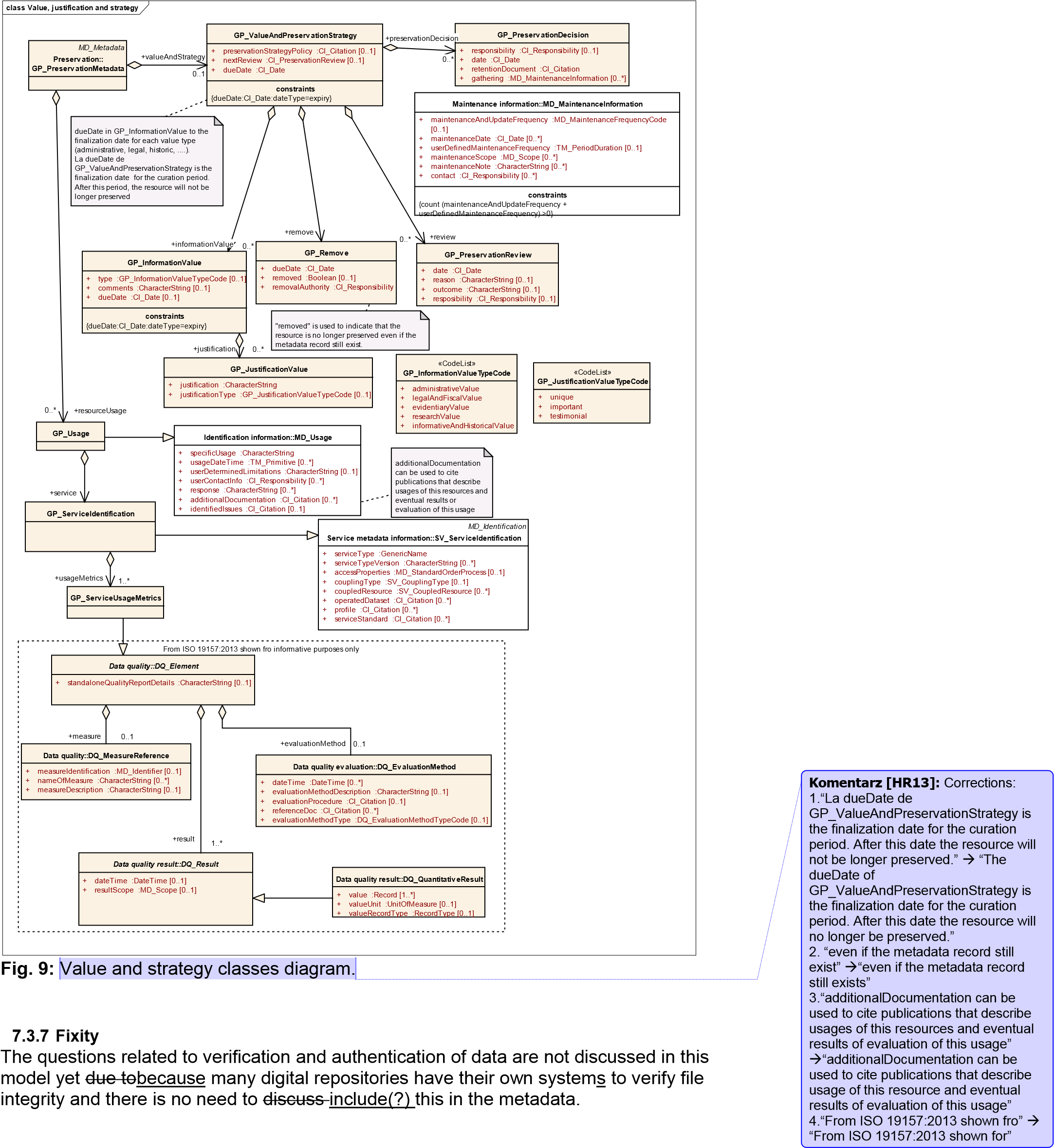
The class diagram in Fig. 8 shows the association of other resources and to express the context of the content information.



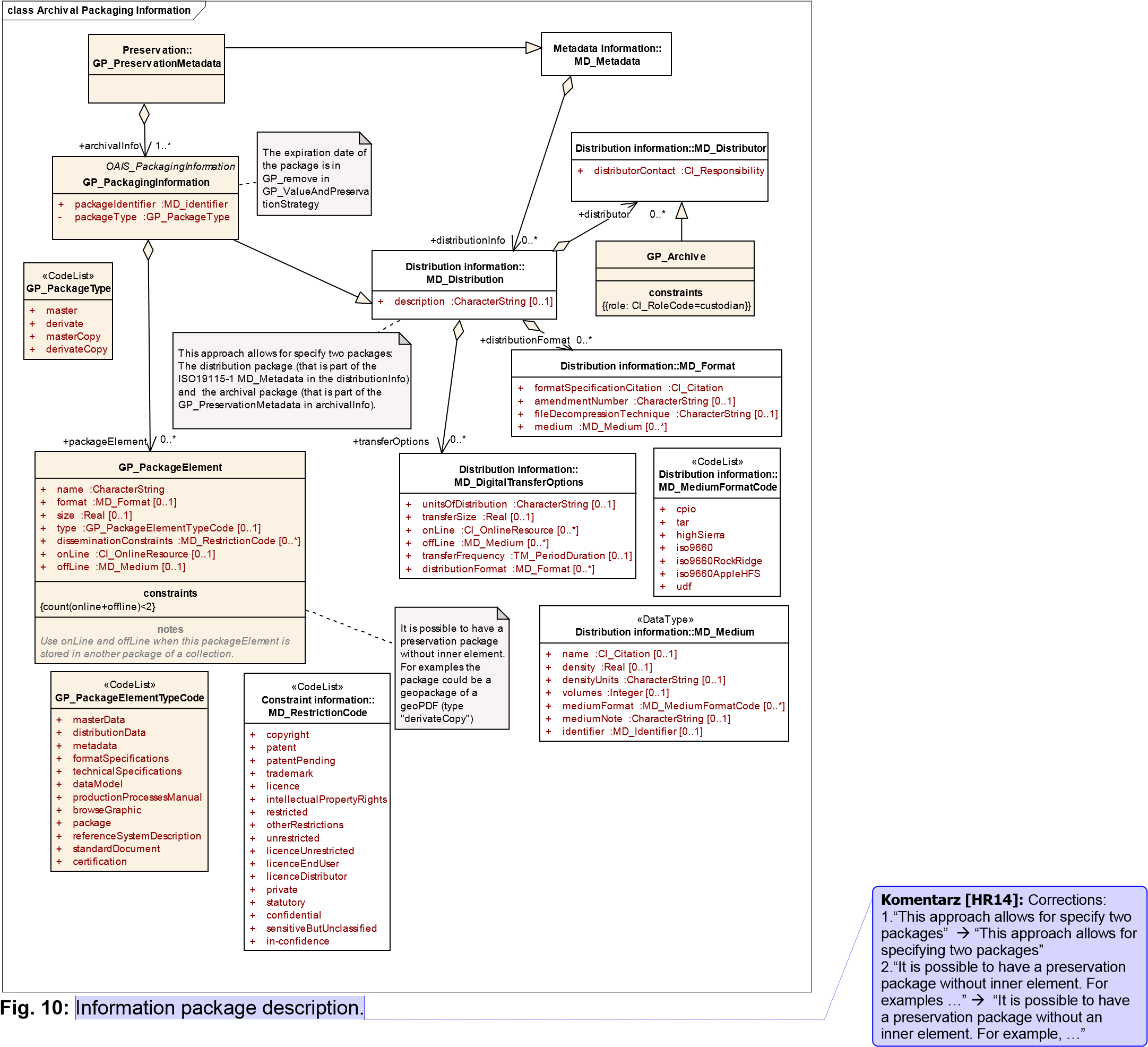
**Fig. 8:** Association of other resources.

**Value and strategy**

Value, use and justification deals with documenting the administrative, legal, evidentiary, research or historical recognized value of the resource and the justification for preserving it (e.g. documenting the legal mandate of preserving the dataset for 10 years); it also includes geospatial services usage statistics (e.g. documenting the number of times the dataset was visualized in a web map service) using the resource as another means of justifying its importance. Preservation strategy and review dates are also considered, including the eventual decision of discontinuing the preservation of the resource. The class diagram in Fig. 9 shows the classes involved.

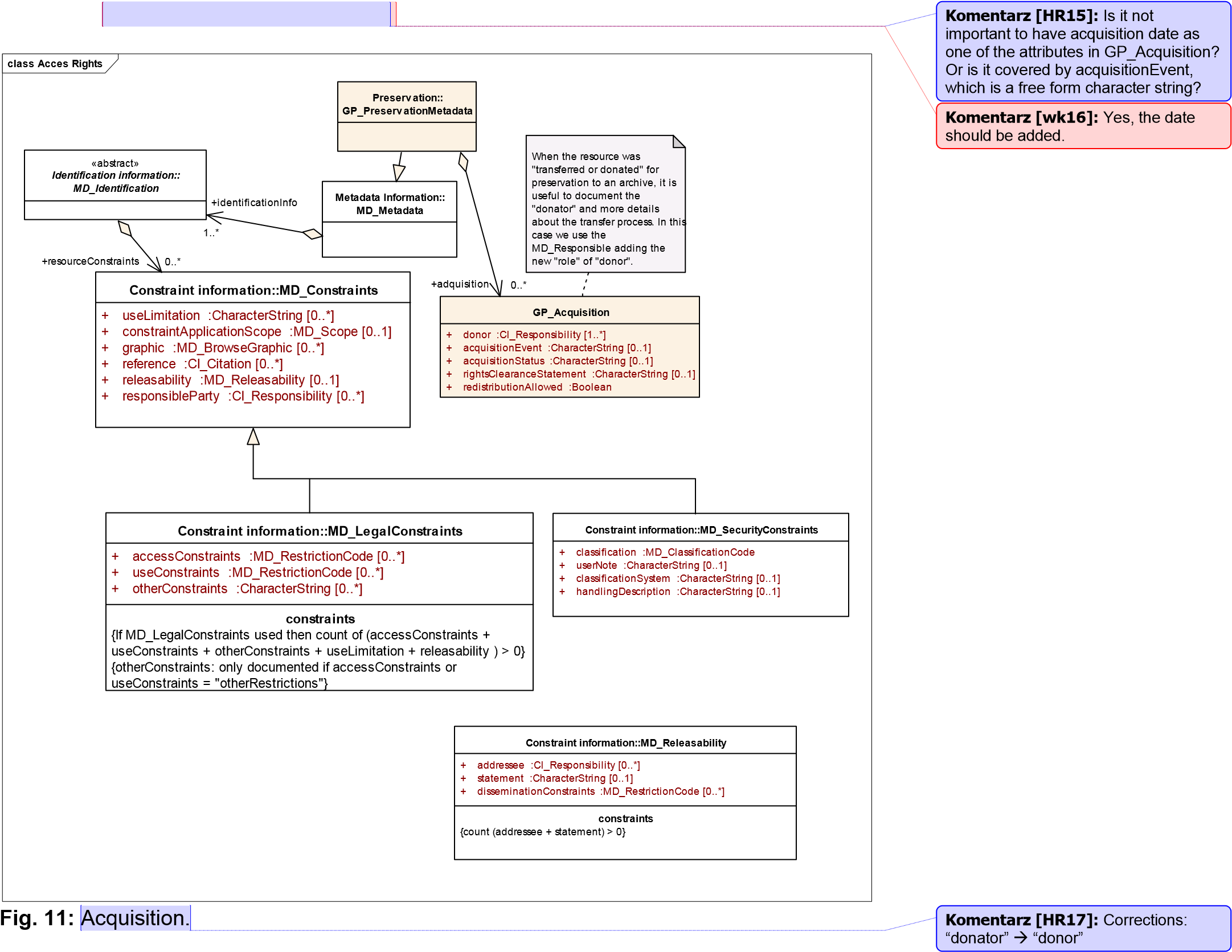


**7.3.8 Information package description**

Preservation package description lists all the parts (files and documents) that together form the AIP (e.g. enumerating all files covering the data values, additional data tables, the metadata, the product specifications, the format specifications, the symbols used in the visualization, a quicklook in PDF format, etc.). The class diagram in Fig. 10 shows all classes related to the information package description.

**7.3.9 Acquisition**

The acquisition class documents the transfer process from the production and maintenance to the curation process including licensing and rights transfer. The class diagram in Fig. 11 shows the class related to the acquisition.



**7.3.10 Coordinate reference system**

The coordinate reference system (CRS) is what makes the resource a geospatial one. This standard sets four methods for defining a CRS:

1.) Method defined in the ISO 19115-1,

2.) EPSG-codes,

3.) Full description of the CRS together with the archived data, and

4.) Reference to a format-registry

The ISO 19115-1 includes two ways to describe the reference system: by reference to a CRS identifier ~~of~~ or by a detailed description of it using MD\_CRS.

The use of EPSG codes as references to CRSs is a common practice in the geospatial web services and EPSG codes are considered stable~~;~~ , even if some modifications ~~has~~ have been made with time and they have become more dependent of the CRS database version.

ISO19115-1 only explicitly describes referencing CRSs and suggests ~~to use~~using ISO19111 and ISO19111-2 to fully describe the CRS. While a reference to a CRS identifier can be perfectly acceptable for a dataset that is currently in use ~~dataset~~ and for a SIP, for preservation purposes the archive should consider ~~to expand~~expanding the definition of the CRS code into a full description of datum, projection and ellipsoid and their parameters. As an alternative the archive could consider ~~to save~~saving a copy of the external CRS databases into the catalogue for preservation purposes.

A way to ensure that the CRS is recorded in a consistent and comprehensive way~~,~~ is linking via unique identifiers to well established CRS registries. In consideration of experience, cost, and resources, an implementation of this standard shall link to existing, acknowledged, and ISO-recommended CRS-registries.

### 7.4 Open Packaging Convention (OPC)

The OAIS defines the idea of the information package as a container that includes the content information and the PDI (preservation description information) and it is described by a descriptive information (see Annex B for a full a description of the OAIS Information model). This is particularly relevant for the geospatial information. The way geospatial data is disseminated on the Internet has forced a separation between data and metadata for practical reasons. It is common practice that geospatial data is hosted and disseminated in the producer’s services while metadata is harvested and replicated in centralised metadata catalogues. Both services are independently managed and in many cases, once a metadata record is discovered and identified as describing a useful dataset by a user, it is not obvious how to get to the data service. This separation is not convenient for preservation purposes.

~~so that~~The OAIS information package offers us a mechanism to use a “container” where data and metadata are kept together.

In fact, OAIS defines 3 different information packages depending on their function. To submit information to an archive a submission information package (SIP) is proposed. To archive the data an archival information package is proposed. To disseminate the archived data to users a dissemination information package (DIP) is proposed. OAIS states that the package can be logical or physical and specific physical formats are not proposed. However, a TAR format is quoted in the document. Nevertheless, it is expected that the SIP and the DIP are physical deliveries of data from the producer to the archive and from the archive to the consumer respectively. A possible ~~media~~ medium for package exchange is the world wide web and, due to the way the web is functioning, it is very useful to have a format that contains all the elements of the information package inside and can be exposed as a single link that can be transferred in a single HTTP (hypertext transfer protocol) interaction, e.g. by clicking on it in a web browser. In the web, the file can be associated to a MIME (multipurpose internet mail extensions) type that web clients can easily deal with by delegating its management to a client helper application.

This standard proposes to use Open Packaging Convention (OPC) format that is described by the ISO 29500-2 (and ECMA-376) to build a geospatial information package. By integrating the needs of the modern Internet environment, the OPC standard can be considered a modern version of the TAR format: it combines a ZIP compression of the parts composing the package (respecting a directory structure) with XML documents that describe the package content (that can be used to store the OAIS descriptive information), the web media types present in the package, the relations between the parts composing the package and a quicklook of the data inside. It also integrates data encryption system to guarantee both data integrity and respect the predefined access rights. An OPC package can contain several files (the “parts”). Names and paths of each part have to follow the URI restrictions and conventions. All these extra capabilities allow for the interoperability of some basic data maintenance, such as the extraction of some parts of the package, thus guaranteeing that all the related resources can be extracted without the need to understand the actual encoding of the individual parts included in the package.

In addition, the OPC standard also introduces the possibility of relating files outside the package (external relations). A practical application of this is to exclude from the package some elements that are considered too remote or too big but keep the relation to them (keep the context information). The files filtered out will be left as a remote URI for further download by the OPC enabled client application.

OPC can be used without any modification for encapsulating geospatial data and metadata by mapping the OAIS concepts and the ones defined by this standard into the OPC concepts. Indeed, OAIS data object is the dataset that is accompanied by OAIS representation information (commonly expressed in ISO 19115-1). OAIS data object (the geospatial dataset) is mapped to one or more “parts” of the OPC package. The OAIS representation rnformation is mapped to ISO 19115-1 metadata, symbolization information, data specification description documents, etc.; all for them are also “parts” of the OPC package. Then PDI is added, using the extensions of ISO 19115-1 proposed in previous sections of this document, as “parts” of the OPC. OPC includes its own additional parts such as a quicklook of the data. The OAIS packaging information in mapped to the internal ZIP headers. Finally, the OAIS packaging description Information is mapped to OPC core metadata stored in the core.xml “part” (see table 1) that can be extracted from discovery elements of the ISO19115-1 metadata. All these “parts” composing the OAIS information package can be included in an OPC package, allowing for both an easy submission of data to the archive for preservation (SIP) and clean way to disseminate the preserved information from the archive to the users (DIP).

Table 1: Mapping between OPC core.xml metadata and ISO 19115-1 metadata elements. This constitutes the OAIS package description information.

|  |  |
| --- | --- |
| **OPC core metadata** | **ISO 19115-1 Metadata – Part 1: Fundamentals** |
| category | MD\_DataIdentification/topicCategory |
| contentStatus | MD\_Identification/status |
| created | MD\_Identification/citation/CI\_Citation.date[CI\_Date/dateType=creation]/CI\_Date/d ateType |
| creator | MD\_Identification/pointOfContact[CI\_Responsibility/role=resourceProvider]/CI\_Re sponsability/party/CI\_Party/name |
| description | MD\_DataIdentification/supplementalInformation |
| identifier | MD\_Identification/citation/CI\_Citation/identifier/MD\_identifier/code |
| keywords | MD\_Identification/descriptiveKeywords |
| language | MD\_DataIdentification/language |
| lastModifiedBy | MD\_Identification/pointOfContact[CI\_Responsibility/role=processor]/CI\_Responsib ility/party/CI\_Party/name|MD\_DataIdentification/environmentDescription |
| lastPrinted | MD\_Identification/citation/CI\_Citation.date[CI\_Date/dateType=publication]/CI\_Dat e/dateType |
| modified | MD\_Identification/citation/CI\_Citation.date[CI\_Date/dateType=revision]/CI\_Date/d |
| **OPC core metadata** | **ISO 19115-1 Metadata – Part 1: Fundamentals** |
|  | ateType |
| revision |  |
| subject | MD\_Identification/abstract |
| title | MD\_Identification/citation/CI\_Citation/title |
| version | MD\_Identification/citation/CI\_Citation/identifier/MD\_identifier/version |

**Annex A**

# (normative) Abstract test suite

## A.1 Semantics

Conformance to this standard, ISO 19165, consists of either service conformance or data conformance.

The abstract test suite has eight conformance classes.

1. Prioritization
2. Structure
3. Rights
4. Time
5. Geospatial information model
6. Designated community
7. Metadata
8. Open Packaging Convention

## A.2 Prioritization

1. Test Purpose: to verify the use of the appropriate temporal classification into the categories set by this standard.
2. Test Method: Inspect the archival information package.
3. Reference: 6.1

## A.3 Structure

### A.3.1 Data format

1. Test Purpose: to verify the fulfillment of all demands set by this standard for preserving the knowledge about the data format
2. Test Method: Inspect the archival information package.
3. Reference: 6.2.1

### A.3.2 Data base

1. Test Purpose: to verify the fulfillment of all demands set by this standard for preserving the knowledge about the data base
2. Test Method: Inspect the archival information package.
3. Reference: 6.2.2

### A.3.3 Properties of geospatial data

1. Test Purpose: to verify ~~the regard of~~that the properties of geospatial data have been considered
2. Test Method: Inspect the archival information package.
3. Reference: 6.2.3

### A.3.4 Level of aggregation

1. Test Purpose: to verify ~~the regard of~~that the level of aggregation has been considered
2. Test Method: Inspect the archival information package.
3. Reference: 6.2.4

### A.3.5 Gold copy

1. Test Purpose: to verify the existence of a gold copy
2. Test Method: Inspect the archival information package.
3. Reference: 6.2.5

## A.4 Rights

1. Test Purpose: to verify ~~the regard of~~that all rights ~~that rest on~~associated with the data have been considered
2. Test Method: Inspect the archival information package.
3. Reference: 6.3

## A.5 Time

1. Test Purpose: to verify ~~the regard of~~that all temporal aspects, e.g. preservation intervals, that are defined by this standard, have been considered
2. Test Method: Inspect the archival information package.
3. Reference: 6.4

## A.6 Geospatial information package

1. Test Purpose: to verify the correct construction of the archival information package (Geo-AIP) and related structures such as Geo-SIP and Geo-DIP defined by this standard
2. Test Method: Inspect the archival information package.
3. Reference: 7.1

## A.7 Designated community

1. Test Purpose: to verify the definition of the designated community as demanded by this standard
2. Test Method: Inspect the archival information package.
3. Reference: 7.2

## A.8 Metadata

### A.8.1 Preservation metadata

1. Test Purpose: to verify the use of the appropriate metadata elements for the preservation metadata
2. Test Method: Inspect the archival information package.
3. Reference: 7.3.1

### A.8.2 Data identifier

1. Test Purpose: to verify the use of the appropriate metadata elements for data identifier
2. Test Method: Inspect the archival information package.
3. Reference: 7.3.2

### A.8.3 Data, product, and format specification

1. Test Purpose: to verify the use of the appropriate metadata elements for the data, product and format specification
2. Test Method: Inspect the archival information package.
3. Reference: 7.3.3

### A.8.4 Preservation actions

1. Test Purpose: to verify the use of the appropriate metadata elements for preservation actions
2. Test Method: Inspect the archival information package.
3. Reference: 7.3.4

### A.8.5 Association of resources

1. Test Purpose: to verify the use of the appropriate metadata elements for the association of resources
2. Test Method: Inspect the archival information package.
3. Reference: 7.3.5

### A.8.6 Value and strategy

1. Test Purpose: to verify the use of the appropriate metadata elements for the value and strategy metadata
2. Test Method: Inspect the archival information package.
3. Reference: 7.3.6

### A.8.7 Fixity

1. Test Purpose: to verify the appropriate addressing of fixity
2. Test Method: Inspect the archival information package.
3. Reference: 7.3.7

### A.8.8 Information package description

1. Test Purpose: to verify the use of the appropriate metadata elements for the information package description
2. Test Method: Inspect the archival information package.
3. Reference: 7.3.8

### A.8.9 Acquisition

1. Test Purpose: to verify the use of the appropriate metadata elements for the description of acquisition
2. Test Method: Inspect the archival information package.
3. Reference: 7.3.9

### A.8.10 Coordinate reference system

1. Test Purpose: to verify ~~the correct way~~correctness of describing the associated coordinate reference system
2. Test Method: Inspect the archival information package.
3. Reference: 7.3.10

## A.9 Open Packaging Convention

1. Test Purpose: to verify ~~the correct~~correctness of mapping to the Open Packaging Convention model
2. Test Method: Inspect the archival information package.
3. Reference: 7.4

**Annex B**

# (informative) Summary of ISO 14721

## B.1 Scope

The scope of the ISO 14721:2012 is defined as follows:

The ISO 14721 defines a reference model for an open archival information system (OAIS). An OAIS is an archive, consisting of an organization, which may be part of a larger organization, of people and systems that has accepted the responsibility to preserve information and make it available for a designated community.

The term “open” in OAIS is used to emphasize that the standard is an open standard maintained by the ISO community (in opposition to a de-facto standard maintained by a single company), and it does not imply that access to the archive is unrestricted.

## B.2 OAIS model

The ISO 14721 defines a model that consists of the OAIS as well as the producers, consumers and management.

Producer is the role played by those persons, or client systems, which provide the information to be preserved.

Management is the role played by those who set overall OAIS policy as one component in a broader policy domain, for example as part of a larger organization. In the sense of this standard this is not involved in the day-to-day archive operations.

Consumer is the role played by those persons, or client systems, that interact with OAIS services to find and acquire preserved information of interest. A special class of consumers is the designated community. The designated community is the set of consumers who should be able to understand the preserved information.

## B.3 OAIS information

A clear definition of information is central to the ability of an OAIS to preserve it.

A person, or system, can be said to have a knowledge base, which allows that person or system to understand received information. Information is defined as any type of knowledge that can be exchanged, and this information is always expressed, i.e. represented, by some type of data in an exchange.

The archive has the mission to preserve data objects that could be physical objects or digital objects composed by bits. The digital objects are only interpretable if the necessary representation information is also provided. The representation information is composed by structure information about the data format used to encode the bits, some semantic information to give meaning to the data and other information that can help in the data interpretation. The representation information designed for a designated community. The designated community is formed by people and/or systems that have the same level of associated knowledge base and for whom the information is being preserved by the archive. The representation information can be considered the first level of metadata. The data object and the representation information together form the content information.

The archive has to be aware that the designated community can change over time and thus the knowledge base will change too. The archive will have to extend the representation information with the information that represents the knowledge that the designed community no longer retains.

The content information can be described by the preservation description information (PDI). The PDI must include information that is necessary to adequately preserve the particular content information with which it is associated. It is specifically focused on describing the past and present states of the content information, ensuring it is uniquely identifiable, and ensuring it has not been unknowingly altered. PDI can be considered a second level of metadata. The PDI types are:

* Reference information, which identifies, and if necessary describes, one or more mechanisms used to provide assigned identifiers for the content information. It also provides those identifiers that allow outside systems to refer, unambiguously, to this particular content information.
* Context information, which documents the relationships of the content information to its environment. This includes why content information was created and how it relates to other content information objects existing elsewhere.
* Provenance information, which documents the history of the content information. This tells the origin or source of the content information, any changes that may have taken place since it was originated, and who has had custody of it since it was originated, providing an audit trail for the content information.
* Fixity information, which provides the data integrity checks or validation/verification keys used to ensure that the particular content information object has not been altered in an undocumented manner.
* Access rights information, which identifies the access restrictions pertaining to the content information, including the legal framework, licensing terms, and access control.

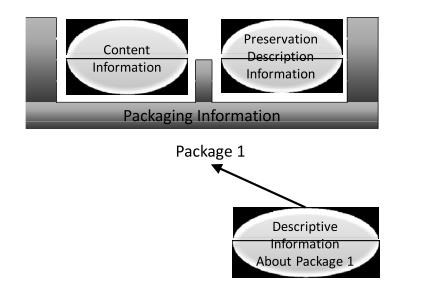


Fig. B1: information package concept.

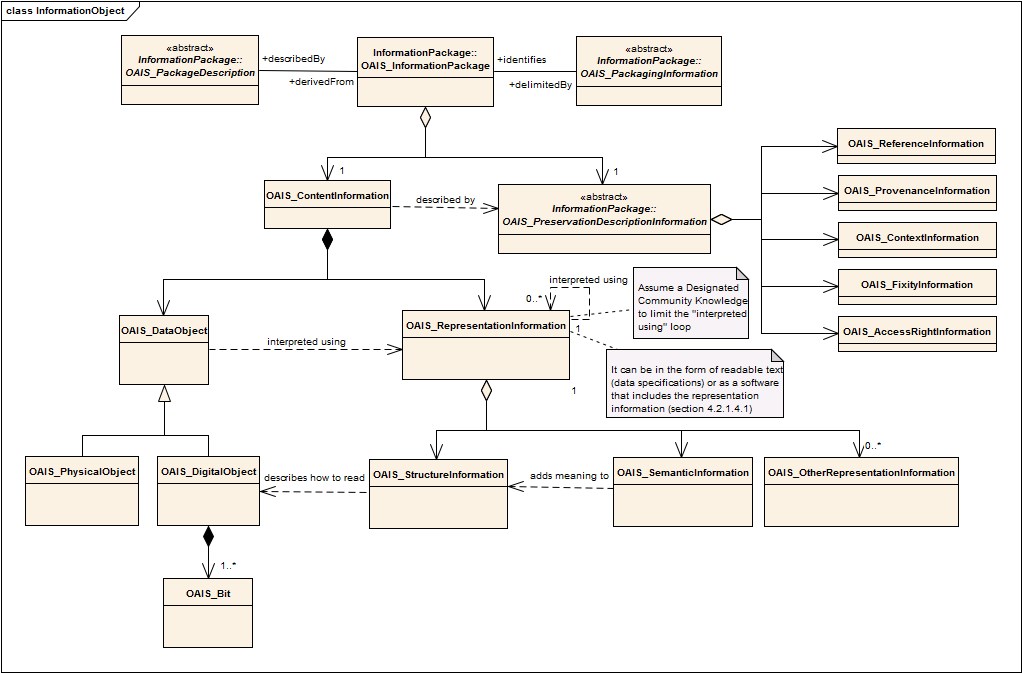


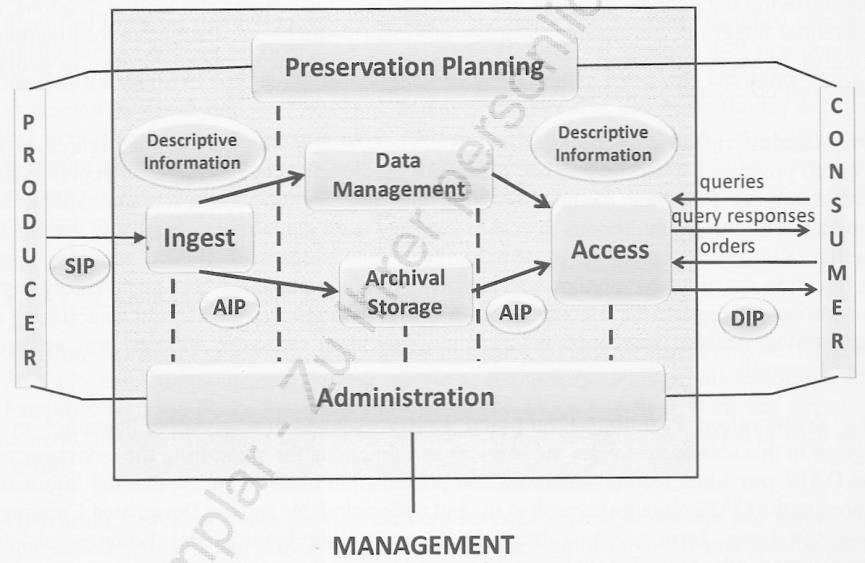
Fig. B2: Different components of the information that is related to the information package. A digital object is described using for different levels of metadata and encapsulated in the information package.

Every submission of information to an OAIS by a producer, and every dissemination of information to a consumer, occurs as one or more discrete transmissions. Therefore, it is convenient to define the concept of an information package.

An information package is a container of two types of information called content information and preservation description information (PDI). The content information and PDI are viewed as being encapsulated and identifiable by the packaging Information that is considered the third level of metadata. The resulting package is viewed as being discoverable by virtue of the descriptive information that is a fourth level of metadata. See Fig. B1 for a graphical representation of the information package composition. A complete perspective of all these levels of metadata and information components can be seen in Fig. B2.

For the operation of an OAIS it is necessary to distinguish between three variants of information packages. These variants are referred to as the submission information package (SIP), the archival information package (AIP), and the dissemination information package (DIP). Submitted information packages often have insufficient representation information or PDI. The OAIS may organize this information very ~~different~~ differently for preservation. In addition, the OAIS may provide to customers information that does not include all representation information or PDI.

Fig. B3 illustrates the management model of the administration of preservation data including the workflow from the owner of the data (producer) to the management of the archive and further to a user of the archived data (consumer). The administration of an archive includes preservation planning, entrance of data (ingest), archival storage, and response to queries (access).



**Fig. B3:** OAIS functional entities (SIP = submission information package, AIP = archival information package, DIP = dissemination information package).

## B.4 OAIS responsibilities

The ISO 14721 defines the mandatory responsibilities that an organization must discharge in order to operate an OAIS archive.

The OAIS shall:

* Negotiate for and accept appropriate information from information producers under consideration of the ISO 19131:2007 “Geographic information – Data product specifications”.

* Obtain sufficient control of the information provided to the level needed to ensure long term preservation.

The problems of assuming sufficient control of the content information and preservation description information, when they are largely digital, are addressed in three related categories, as follows:

* + copyright implications, intellectual property and other legal restrictions on use;
  + authority to modify representation information; o agreements with external organizations.

* Determine, either by itself ~~of~~ or in conjunction with other parties, which communities should become the designated community and, therefore, should be able to understand the information provided, thereby defining its knowledge base.

* Ensure that the information to be preserved is independently understandable to the designated community. In particular, the designated community should be able to understand the information without needing special resources such as the assistance of the experts who produced the information.

* Follow documented policies and procedures which ensure that the information is preserved against all reasonable contingencies, including the demise of the archive, ensuring that it is never deleted unless allowed as part of an approved documented strategy. There should be no ad-hoc deletions.

* Make the preserved information available to the designated community and enable the information to be disseminated as copies of, or as traceable to, the original submitted data object with evidence supporting its authenticity.

## B.5 OAIS preservation perspectives

The fast-changing nature of the computer industry and the ephemeral nature of electronic data storage media are at odds with the key purpose of an OAIS: to preserve information over a long period of time. No matter how well an OAIS maintains its current holdings, it will eventually need to migrate much of its holding to different media.

Digital migration of an archival information package (AIP) can include

* copying content data object or representation information bits to new media
* altering or adding to content data object or representation information bits
* altering or adding to preservation description information bits
* altering or adding to operational software whose role is essential for content information preservation, i.e. it is part of representation information
* altering or adding to the bits that make up the AIP’s packaging information

## B.6 Digital migration

Digital migration is defined to be the transfer of digital information, while intending to preserve it, within the OAIS. It is distinguished from transfers in general by three attributes:  a focus on the preservation of the full information content intended for preservation;

* a perspective that the new archival implementation of the information is a replacement for the old; and
* full control and responsibility over all aspects of the transfer resides with the OAIS.

Three major motivators are seen to drive digital migration of AIPs within an OAIS. These are

* Improved cost-effectiveness: The rapid pace of hardware 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 and it drives the obsolescence of software employed as part of representation information.
* New consumer-service requirements
* Media decay

# (normative) Requirements for data management regarding preservation

## C.1 Archiving package for geospatial data registries

### C.1.1 Geo submission information package (Geo-SIP)

The ISO 14721:2012 defines the general concepts SIP, AIP, and DIP (submission information package, archival information package, and dissemination information package).

This standard defines a specialized geo-type of these packages, i.e. Geo-SIP, Geo-AIP, and Geo-DIP. The purpose of these is the inclusion of geo-specific information in the package and a standardized internal structure.

A Geo-SIP shall have the following properties:

* The Geo-SIP-files shall be limited in size to let them be processed by state-of-the-art systems. A Geo-SIP shall have a manageable regional size, e.g. tiles of 1 km x 1 km or 10 km x 10 km. This size applies to the complete full size of the Geo-SIP and not to a part such as a tile.
* The Geo-SIP-files shall have an internal directory structure.
* The names of the files to be archived shall be preserved in the Geo-SIP.
* Geometry, vector and raster, is organized in entities which may include a coordinate dataset and a definition of a line, an area or a solid. Consequently, the packaging for archiving purposes shall not split such entities.
* Topology defines neighbourhood-relations between geospatial entities. A complete or an almost complete regional coverage causes a sequence of links that practically binds all entities together. Consequently, such a “borderless” dataset shall not be split for packaging unless a loss of topological information is ~~accepted~~acceptable.
* Metadata models and catalogues shall be archived together with geospatial dataset. The content of the metadata models and catalogues shall be reduced to the elements that are links to elements, e.g. objects, of the geospatial dataset (homogenization).

The Geo-AIP and the Geo-DIP shall have the same structure. The applied character set shall be defined according ISO 19115-1 as a metadata element.

### C.1.2 Access

Geographical data shall be preserved in a way that non geo-specialists can handle.

Current experience with users of archived geospatial data shows that the demand is for the data to be usable in a GIS. Therefore, the access to archived data shall be arranged in the same way as to non-archived data.

**Komentarz [wk18]:** Yes, delete the second sentence of this bullet.

**Komentarz [HR19]:** This is still confusing. Does a Geo-SIP have many files that are called Geo-SIP files? I really don’t think this bullet is necessary. If we need it, I would stop after the first sentence. We want the files sizes to be limited to what state-ofthe-art systems can handle. The details vary a lot from system to system and organization to organization.

# (informative) Case-specific archival concept

The development of an archiving concept requires a close cooperation between the archival experts and the geospatial data domain experts, e.g., in land survey administration and spatial data infrastructure offices.

A questionnaire may help to address and clarify the most relevant topics:

* Institution or agency
* Contact
* Name of application
* Theme of application
* Name and version of GIS-software
* Developer and vender of GIS- software
* Geo-reference of thematic data
* Regional coverage
* Update cycle
* Storage of data history
* Data format
* Availability of metadata
* Usage of geospatial base-data of other institutions
* Other institutions, if yes

# (normative) Functional requirements for a preservation archive

## E.1 Introduction

The purpose of this annex is to document a set of high level functional requirements that should be satisfied by an archive that preserves geospatial data and metadata. In the following sections, the term “preservation content” is used to refer to data, metadata and all the related documentation needed to be preserved to ensure understandability and reusability of the data.

## E.2 Data ingest

**E.2.1** The archive ingests the preservation content and prepares it for storage and preservation.

**E.2.2** The archive ingests the preservation content from each provider in accordance with appropriate documented interface specifications.

**E.2.3** The archive ensures ingest data integrity through the use of appropriate technologies. **E.2.4** The archive verifies the quality of the preservation content upon ingest (e.g. dataset size, dataset name) for each unique dataset.

## E.3 Data storage and preservation

### E.3.1 Overview

The archive preserves designated data products and distributes them on request to users. Some products may be created for distribution by an on-demand processing and should be subject to the same delivery requirements as products that are stored in the archive.

### E.3.2 Data storage

**E.3.2.1** The archive stores all designated data products or creates products on demand. The archive ensures that products generated on demand (also known as virtual products) are identical to the corresponding products that would otherwise be archived, having undergone operational and scientific quality assessment. Designated products are established as a part of the archive’s planning process.

**E.3.2.2** The archive provides preservation planning to ensure the preservation content remains accessible to and understandable by the user community throughout the data life cycle.

**E.3.2.3** The archive is capable of archiving multiple versions of selected archive data. **E.3.2.4** The archive maintains metadata on all products that are available from the archive. The metadata content shall be compliant with the ISO standards quoted in section 3, Normative References

**E.3.2.5** The archive maintains long-term archiving, distribution and user services functions for designated data products.

**E.3.2.6** The archive defines and adheres to retention requirements for all data and information identified for preservation.

**E.3.2.7** The archive maintains an off-site backup copy of all data ~~which~~ that would otherwise be impossible or difficult to recover in case of loss.

**E.3.2.8** The archive maintains the capability to restore its archive to avoid permanent loss of archived data.

**E.3.2.9** The archive allows for new technology integration for archival data.

**E.3.2.10** The archive allows old versions of data to be removed from the archive.

**E.3.2.11** The archive manages (populates, maintains and accesses) the information identifying and documenting the stored data and information and produce reports on the stored data and information.

### E.3.3 Preservation management

**E.3.3.1** The archive maintains an electronic inventory of all archived data.

**E.3.3.2** The archive provides the capability of retrieving any data granule stored in the archive.

**E.3.3.3** The archive interoperates with other archives of data relevant to the user community.

## E.4 Data distribution

**E.4.1** The archive enables users ~~in determining~~to determine the existence, description and availability of stored data and information and ~~allowing~~ allows consumers to request and receive the data and information.

**E.4.2** The archive distributes to users, upon request, data products, metadata, ancillary/auxiliary data, calibration data, science software, and documentation.

**E.4.3** The archive distributes data in standard formats commonly accepted within the user community for the types of data for which the archive is responsible. **E.4.4** The archive distributes data to users via electronic networks.

**E.4.5** The archive provides data to processing systems to support product generation, reprocessing and quality assurance in a timely manner.

**E.4.6** The archive provides subsetted data and/or subsetting, reprojection and format conversion tools appropriate to the archive’s data holdings to ensure efficient distribution of data to users.

**E.4.7** The archive ensures data integrity through the use of appropriate technologies such as checksums on distribution in order to satisfy interface requirements with external systems.

**E.4.8** The archive makes available for distribution standard metadata using appropriate tagging (e.g., extensible mark-up language [XML]).

**E.4.9** The archive implements a user feedback process to improve the data, the information, its usefulness and the available services.

# Bibliography

1. BAR, 2013: Projekt Ellipse, Konzeption der Archivierung von Geobasisdaten des **Sformatowano:** Niemiecki (Niemcy) Bundesrechts, Bundesamt für Landestopografie swisstopo and Schweizerisches Bundesarchiv BAR. –

http://www.swisstopo.admin.ch/internet/swisstopo/de/home/topics/geodata/geoarchive .parsysrelated1.29968.downloadList.70421.DownloadFile.tmp/konzeptberichtellipsev1 .3publikationde.pdf (18.5.2014).

1. CEOS WGISS-DSIG, 2015, Long Term Preservation of Earth Observation Space Data: Preservation Guidelines, Version 1.0, http://ceos.org/document\_management/Working\_Groups/WGISS/Interest\_Groups/Dat a\_Stewardship/Recommendations/EO%20Data%20Preservation%20Guidelines\_v1.1. pdf (1.9.2015).
2. CEOS WGISS-DSIG, 2015, Long Term Preservation of Earth Observation Space Data: Preservation Workflow, Version 1.0, http://ceos.org/document\_management/Working\_Groups/WGISS/Interest\_Groups/Dat a\_Stewardship/Best\_Practices/Preservation%20Workflow\_v1.0.pdf (1.3.2015).
3. CEOS WGISS-DSIG, 2015, Long Term Preservation of Earth Observation Space Data Earth Observation: Preserved Data Set Content, Version 1.0,

http://ceos.org/document\_management/Working\_Groups/WGISS/Interest\_Groups/Dat a\_Stewardship/Recommendations/EO%20Preserved%20Data%20Set%20Content\_v

1.0.pdf (1.9.2015).

1. CEOS WGISS-DSIG, 2015, Long Term Preservation of Earth Observation Space Data: Generic Earth Observation Data Set Consolidation Process, Version 1.0, http://ceos.org/document\_management/Working\_Groups/WGISS/Interest\_Groups/Dat a\_Stewardship/Best\_Practices/GenericEarthObservationDataSetConsolidationProces s\_v1.0.pdf (1.3.2015).
2. Dorfey, B., Graf, S., Grau, B., Homberg, J., Huth, K., Keitel, C., Sandner, P. & **Sformatowano:** Niemiecki (Niemcy) Wettmann, A., 2009: Handreichung zur Archivierung elektronisch vorliegender

Geodaten. – http://www.bundesarchiv.de/imperia/md/content/bundesarchiv\_de/fachinformation/ark/ handreichung\_geodaten\_20090928.pdf (18.5.2014).

1. Engel, H., Engelmeier, S. Grams, G., Gültlinger, M., Häußermann, M., Hermann, C.,

Hoss, H., Keitel, C., Koch, E., Lang, R., Naumann, K., Schurer, H., Spohrer, M.,

Strietter, O., Witke, T., Wöhrle, M., Wütherich, T., 2013: Archivierung von

Geobasisdaten im Kontext der Gesamtüberlieferung des Vermessungswesens. – Abschlussbericht der AG „Archivierung von Geobasisdaten des LGL“, Stuttgart, Germany.

1. ESA, 2012: Long Term Data Preservation, Earth Observation Preserved Data Set Content. – http://earth.esa.int/gscb/ltdp/LTDP\_PDSC\_4.0.pdf (18.5.2014).
2. Higgins, S., 2008: The DCC Curation Lifecycle Model (Digital Curation Centre). – The International Journal of Digital Curation 1 (3). – http://www.ijdc.net/index.php/ijdc/article/download/69/48 (18.5.2014).
3. Jobst, M. & Gartner, G., 2011: Structural Aspects for the Digital Cartographic Heritage. – Jobst, Markus (ed.): Preservation in Digital Cartography – Archiving Aspects, Springer.
4. Morris, S., 2010: Appraisal and Selection of Geospatial Data, Prepared for Library of Congress. – National Digital Information Infrastructure and Preservation Program

(NDIIPP). –

http://www.digitalpreservation.gov/meetings/documents/othermeetings/AppraisalSelec tion\_whitepaper\_final.pdf (18.5.2014).

1. NASA, 2013: NASA Eartth Science Data Preservation Content Specification, 423- **Sformatowano:** Czcionka: Nie

SPEC-01, [https://earthdata.nasa.gov/files/423-SPEC-](https://earthdata.nasa.gov/files/423-SPEC-001_NASA%20ESD_Preservation_Spec_OriginalCh01_0.pdf) Pogrubienie

[001\_NASA%20ESD\_Preservation\_Spec\_OriginalCh01\_0.pdf.](https://earthdata.nasa.gov/files/423-SPEC-001_NASA%20ESD_Preservation_Spec_OriginalCh01_0.pdf)

~~[12]~~[13] PREMIS, 2012: Data Dictionary for Preservation Metadata, version 2.2, July 2012 (premis-2-2.pdf). – http://www.loc.gov/standards/premis (18.5.2014).

~~[13]~~[14] PRONOM, 2015: http://apps.nationalarchives.gov.uk/PRONOM/Default.aspx **Sformatowano:** Polski (31.12.2015).

~~[14]~~[15] Rönsdorf, C., Mason, P., Gerber, U., Bos, M., Shaon, A., Naumann, K., Kirstein, M., Samuelsson, G., Rantala, M., Kvarteig, S. & Stößel, W., 2012: GI+100: Long term preservation of digital Geographic Information – 16 fundamental principles agreed by

National Mapping Agencies and Archives. – http://www.eurosdrireland.net/archiving/GI+100%20-

%2016%20EuroSDR%20Archiving%20Principles\_V3%201.pdf (18.5.2014).

~~[15]~~[16] Sandner, P., 2012: Langzeitarchivierung von Geodaten. – **Sformatowano:** Niemiecki (Niemcy)

http://www.swisstopo.admin.ch/internet/swisstopo/de/home/docu/Kolloquien/120323.p arsys.6785.downloadList.22057.DownloadFile.tmp/

20120323sandnerlangzeitarchivierunggeodatenpub.pdf (18.5.2014).

~~[16]~~[17] ISO 19131:2007 “Geographic information – Data product specifications”. ISO, 2007.