INTERNATIONAL STANDARD

ISO 19165-1

First edition
2018-05

Geographic information — Preservation of digital data and metadata —

Part 1: Fundamentals

Information géographique — Archivage des données numériques et des métadonnées —

Partie 1: Principes fondamentaux

Reference number
ISO 19165-1:2018(E)

ISO copyright office
CP 401 • Ch. de Blandonnet 8
CH-1214 Vernier, Geneva
Phone: +41 22 749 01 11
Fax: +41 22 749 09 47
Email: copyright@iso.org
Website: www.iso.org

Published in Switzerland

Contents

Page

Foreword ... **iv**
Introduction ... **v**
1 Scope ... **1**
2 Normative references ... **1**
3 Terms and definitions ... **1**
4 Abbreviated terms and conventions ... **8**
4.1 Abbreviated terms ... 8
4.2 Conventions ... 8
5 Conformance ... **9**
6 Preservation ... **9**
6.1 Prioritization ... 9
6.2 Structure ... 9
6.2.1 Data format ... 9
6.2.2 Data structures ... 10
6.2.3 Software and algorithms ... 10
6.2.4 Properties of geospatial data ... 10
6.2.5 Gold copy ... 11
6.3 Rights/licensing ... 11
6.4 Time ... 11
7 Geospatial information model ... **12**
7.1 Overview ... 12
7.2 Designated community ... 14
7.3 Metadata ... 15
7.3.1 Introduction to metadata ... 15
7.3.2 Preservation metadata classes ... 15
7.3.3 Data identifiers ... 16
7.3.4 Data, product and format specifications ... 17
7.3.5 Preservation actions ... 18
7.3.6 Association of resources ... 19
7.3.7 Value and strategy ... 20
7.3.8 Fixity ... 21
7.3.9 Information package description ... 22
7.3.10 Acquisition ... 23
7.3.11 Coordinate reference system ... 24
7.4 Open Packaging Convention (OPC) ... 25
Annex A (normative) **Abstract test suite** ... **27**
Annex B (normative) **Data dictionary** ... **31**
Annex C (informative) **Case-specific archival concept** ... **38**
Annex D (normative) **Functional requirements for a preservation archive** ... **39**
Bibliography ... **41**

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 non-governmental, 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.

The procedures used to develop this document and those intended for its further maintenance are described in the ISO/IEC Directives, Part 1. In particular the different approval criteria needed for the different types of ISO documents should be noted. This document was drafted in accordance with the editorial rules of the ISO/IEC Directives, Part 2 (see www.iso.org/directives).

Attention is drawn to the possibility that some of the elements of this document may be the subject of patent rights. ISO shall not be held responsible for identifying any or all such patent rights. Details of any patent rights identified during the development of the document will be in the Introduction and/or on the ISO list of patent declarations received (see www.iso.org/patents).

Any trade name used in this document is information given for the convenience of users and does not constitute an endorsement.

For an explanation on the voluntary nature of standards, the meaning of ISO specific terms and expressions related to conformity assessment, as well as information about ISO's adherence to the World Trade Organization (WTO) principles in the Technical Barriers to Trade (TBT) see the following URL: www.iso.org/iso/foreword.html.

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

Introduction

Today's information is mostly stored on digital media, which has a shorter lifetime than that of analogue media for a variety of reasons. Unless systematically archived, the storage media will decay and the information is lost. Missing or incorrect metadata describing the format of data can also result in lost digital information. Unfortunately, this daunting scenario occurs often. Consequently, the epoch in which we presently live is sometimes named the "Digital Dark Age".

Traditional archives are facilities or organizations that 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 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).

Geospatial data possess several distinguishing structural characteristics that may include:

— relations to a well-defined section of the Earth;

— exchange by using theme-specific and sophisticated exchange formats;

— links to thematic data (databases);

— transformation between different coordinate reference systems;

— visualization (map output);

— large data volumes;

— existence of several levels-of-detail of the same dataset;

— links between a geospatial dataset and rights.

These distinctive features suggest that geospatial data shall be preserved together with relevant metadata content that fully addresses these structural characteristics.

ISO 14721 defines a reference model for archiving digital information. The application of ISO 14721 is not limited to space data and it is widely used by digital libraries. However, ISO 14721 does not completely cover all the needs for digital data and metadata preservation for geospatial data in general. Therefore, the ISO 19165 series addresses geospatial data, its data model structures, the multiplicity of data formats, and intellectual property rights. ISO 19165 is required by and developed for the geospatial community. ISO/TC 211, Geographic information/Geomatics, has developed the ISO 19100 family of standards dedicated to geographic information. One of them is ISO 19115-1. ISO 19165 is modelled as a specialization of ISO 19115-1. This document is neither a profile nor an implementation of ISO 14721.

Apart from the ISO standards mentioned above, other standards for archival metadata exist. Examples are the provenance family of documents of the W3C[19] and PREMIS, the data dictionary for preservation metadata[17].

Geographic information — Preservation of digital data and metadata —

Part 1: Fundamentals

1 Scope

This document defines a preservation metadata extension of ISO 19115-1.

It defines the requirements for the long-term preservation of digital geospatial data. These data also include metadata, representation information, provenance, context and any other content items that capture the knowledge that are necessary to fully understand and reuse the archived data. This document also refers to characteristics of data formats that are useful for the purpose of archiving.

Geospatial data are preserved as a geospatial information package (IP). This document defines the requirements of the geospatial archival IP and details of the geospatial submission and the dissemination IPs. A geospatial archival IP is fully self-describing and allows a future reconstruction of the dataset without external documentation. The functional requirements for a preservation archive are defined in Annex D.

This document complements standards developed by ISO/TC 211 as well as other ISO standards such as ISO 14721.

2 Normative references

The following documents are referred to in the text in such a way that some or all of their content constitutes requirements of this document. For dated references, only the edition cited applies. For undated references, the latest edition of the referenced document (including any amendments) applies.

ISO 14721:2012, *Space data and information transfer systems — Open archival information system (OAIS) — Reference model*

ISO 19110, *Geographic information — Methodology for feature cataloguing*

ISO 19115-1:2014, *Geographic information — Metadata — Part 1: Fundamentals*

ISO/TS 19115-3, *Geographic information — Metadata — XML schema implementation of metadata fundamentals*

ISO 19157:2013, *Geographic information — Data quality*

3 Terms and definitions

For the purposes of this document, the following terms and definitions apply.

ISO and IEC maintain terminological databases for use in standardization at the following addresses:

— IEC Electropedia: available at http://www.electropedia.org/

— ISO Online browsing platform: available at http://www.iso.org/obp

3.1
access rights information
information that identifies the access restrictions pertaining to the content information, including the legal framework, licensing terms, and access control

Note 1 to entry: Access rights information 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, 1.7.2, modified — Supporting content has been relocated to Note 1 to entry.]

3.2
access software
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, 1.7.2, modified]

3.3
archival information package
AIP
information package (3.16), consisting of the *content information* (3.6) and the associated *preservation description information* (PDI) (3.25), which is preserved within an OAIS (3.22)

[SOURCE: ISO 14721:2012, 1.7.2, modified]

3.4
AIP edition
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

Note 1 to entry: This definition only refers to digital migration.

Note 2 to entry: An AIP edition is not considered to be the result of a migration.

[SOURCE: ISO 14721:2012, 1.7.2, modified — Supporting content has been relocated to Note 2 to entry.]

3.5
AIP version
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

Note 1 to entry: An AIP version is considered to be the result of a digital migration.

[SOURCE: ISO 14721:2012, 1.7.2, modified — Supporting content has been relocated to Note 1 to entry.]

3.6
content information
set of information that is the original target of preservation or that includes part or all of that information

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

[SOURCE: ISO 14721:2012, 1.7.2, modified — Supporting content has been relocated to Note 1 to entry.]

3.7
data dictionary
formal repository of terms used to describe data

[SOURCE: ISO 14721:2012, 1.7.2, modified]

3.8
data dissemination session
delivery of media or a single telecommunications session that provides data to a consumer

Note 1 to entry: 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, 1.7.2, modified — Supporting content has been relocated to Note 1 to entry.]

3.9
data submission session
delivery of media or a single telecommunications session that provides data to an OAIS

Note 1 to entry: 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, 1.7.2, modified — Supporting content has been relocated to Note 1 to entry.]

3.10
designated community
identified group of potential consumers who should be able to understand a particular set of information

Note 1 to entry: 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, 1.7.2, modified — Supporting content has been relocated to Note 1 to entry.]

3.11
digital migration
transfer of digital information, while intending to preserve it, within the OAIS

Note 1 to entry: Digital migration 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;

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

[SOURCE: ISO 14721:2012, 1.7.2, modified — Supporting content has been relocated to Note 1 to entry.]

3.12
digital object
object composed of a set of bit sequences

[SOURCE: ISO 14721:2012, 1.7.2, modified]

3.13
dissemination information package
DIP
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, 1.7.2, modified]

3.14
federated archives
group of archives that has agreed to provide access to their holdings via one or more common finding aids

[SOURCE: ISO 14721:2012, 1.7.2, modified]

3.15
geographic information system
information system dealing with information concerning phenomena associated with location relative to the Earth

[SOURCE: ISO 19101-1:2014, 4.1.20]

3.16
information package
logical container composed of optional content information and optional associated preservation description information

Note 1 to entry: 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, 1.7.2, modified — Supporting content has been relocated to Note 1 to entry.]

3.17
knowledge base
database of knowledge about a particular subject

Note 1 to entry: The database contains facts, inferences, and procedures needed for problem solution [Webster Computer].

Note 2 to entry: The set of information may be incorporated by, or understood by, a person or a system.

[SOURCE: ISO/TS 19101-2:2008, 4.18, modified]

3.18
long term
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

Note 1 to entry: This period extends into the indefinite future.

[SOURCE: ISO 14721:2012, 1.7.2, modified — Supporting content has been relocated to Note 1 to entry.]

3.19
long term preservation
act of maintaining information, independently understandable by a designated community, and with evidence supporting its authenticity, over the long term

[SOURCE: ISO 14721:2012, 1.7.2, modified]

3.20
management
<OAIS> 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, 1.7.2, modified]

3.21
metadata
information about a resource

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

3.22
open archival information system
OAIS
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

Note 1 to entry: An OAIS Archive meets a set of responsibilities that allows 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, 1.7.2, modified — Supporting content has been relocated to Note 1 to entry.]

3.23
package description
information intended for use by access aids

[SOURCE: ISO 14721:2012, 1.7.2, modified]

3.24
packaging information
information used to bind and identify the components of an information package

EXAMPLE The ISO 9660 volume and directory information is used on a CD-ROM to provide the content of several files containing content information and preservation description information.

[SOURCE: ISO 14721:2012, 1.7.2, modified — Supporting content has been relocated to an example.]

3.25
preservation description information
PDI
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, 1.7.2, modified]

3.26
profile
set of one or more base standards or subsets of base standards, and, where applicable, the identification of chosen clauses, classes, options and parameters of those base standards, that are necessary for accomplishing a particular function

Note 1 to entry: A profile is derived from base standards so that by definition, conformance to a profile is conformance to the base standards from which it is derived.

[SOURCE: ISO 19106:2004, 4.5]

3.27
producer
<OAIS> role played by those persons or client systems that provide the information to be preserved

Note 1 to entry: This can include other OAISes or internal OAIS persons or systems.

[SOURCE: ISO 14721:2012, 1.7.2, modified — Supporting content has been relocated to Note 1 to entry.]

3.28
provenance information
information that documents the history of the content information

Note 1 to entry: 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, 1.7.2, modified — Supporting content has been relocated to Note 1 to entry.]

3.29
reference information
information that is used as an identifier for the content information

Note 1 to entry: Reference information also includes identifiers that allow outside systems to refer unambiguously to particular content information. An example of reference information is an ISBN.

[SOURCE: ISO 14721:2012, 1.7.2, modified — Supporting content has been relocated to Note 1 to entry.]

3.30
reference model
framework for understanding significant relationships among the entities of some environment, and for the development of consistent standards or specifications supporting that environment

Note 1 to entry: 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, 1.7.2, modified — Supporting content has been relocated to Note 1 to entry.]

3.31
refreshment
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, 1.7.2]

3.32
repackaging
digital migration in which there is an alteration in the packaging information of the AIP

[SOURCE: ISO 14721:2012, 1.7.2]

3.33
replication
digital migration where there is no change to the packaging information, the content information, and the PDI

Note 1 to entry: The bits used to represent these information objects are preserved in the transfer to the same or new media instance.

[SOURCE: ISO 14721:2012, 1.7.2, modified — Supporting content has been relocated to Note 1 to entry.]

3.34
representation information
information that maps a data object into more meaningful concepts

EXAMPLE 1 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.

EXAMPLE 2 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, 1.7.2, modified — Supporting content has been relocated to examples 1 and 2.]

3.35
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 — addition of Notes 1, 2 and 3 to entry]

3.36
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 19152:2012, 4.1.21]

3.37
spatial source
source with the spatial representation of one (part of) or more spatial units

[SOURCE: ISO 19152:2012, 4.1.22, modified — The EXAMPLE was deleted.]

3.38
submission agreement
agreement reached, between an OAIS and the producer, that specifies a data model, and any other arrangements needed, for the data submission session

Note 1 to entry: 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, 1.7.2, modified — Supporting content has been relocated to Note 1 to entry.]

3.39
submission information package
SIP
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, 1.7.2, modified]

3.40
transformation
<OAIS> digital migration in which there is an alteration to the content information or PDI of an archival information package

EXAMPLE Changing ASCII codes to UNICODE in a text document being preserved is a transformation.

[SOURCE: ISO 14721:2012, 1.7.2, modified — Supporting content has been relocated to an example.]

4 Abbreviated terms and conventions

4.1 Abbreviated terms

AIP	archival information package
ASCII	American Standard Code for Information Interchange
CRS	coordinate reference system
DIP	dissemination information package
FITS	flexible image transport system
geo-AIP	geospatial archival information package
geo-DIP	geospatial dissemination information package
geo-SIP	geospatial submission information package
HTTP	hypertext transfer protocol
ISBN	international standard book number
JPEG	joint photographic expert group
LA	land administration
MIME	multipurpose internet mail extensions
OAIS	open archival information system
OPC	open packaging convention
PDI	preservation description information
SIP	submission information package
SOA	service oriented architecture
TAR	tape archiving (packaging format)
URI	uniform resource identifier
XML	extensible mark-up language
ZIP	file compression format

4.2 Conventions

Some of the classes and attributes are defined in other ISO geographic information standards. Those classes and attributes are identified by one of the following two-character codes.

CI = Citation (ISO 19115-1)

DS = Dataset (ISO 19115-1)

DQ = Data Quality (ISO 19157)

EX = Extent (ISO 19115-1)

GP = Geospatial Preservation (ISO 19165)

LI = Lineage (ISO 19115-1)

MD = Metadata (ISO 19115-1)

SV = Metadata for Services (ISO 19115-1)

TM = Temporal (ISO 19108)

5 Conformance

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

6 Preservation

6.1 Prioritization

The extremely rapid increase in the quantity of data prevents preservation of all information. Consequently, only a selected subset of data can move to a long term archive as determined by an appraisal strategy. The appraisal strategy shall primarily evaluate the relevance of the data to be archived. This appraisal should be done by the producer in cooperation with the archival institution, by the appropriate involvement of the consumer community. Preservation shall be included in the product life cycle and requires a decision on the archival procedure at the moment the data are created.

The value of geospatial datasets shall be appraised according to their content and categorized accordingly as having short-term or long term archival value with described time spans. The short-term value could, for example, be 1 year to 10 years. For these geospatial data with a short-term value, a disposal schedule may be created where the archival storage and possible disposal process are described. The preservation duration may be reassessed before the end of the term defined. Geospatial datasets that have been classified as having long term value could, for example, be 100 years or more.

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

The layers relevance may be distinguished by the time, the function and the relationship.

See Annex C for more details on topics to be discussed during the initial appraisal.

EXAMPLE

Function: A typical example is a future analysis of a land consolidation project. Though the administrative procedure will be fully complete after a few years, the information keeps its value as documentation of the change of landscape.

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

6.2 Structure

6.2.1 Data format

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

A geospatial dataset shall be archived together with the full documentation of its data format either by maintaining the full documentation on all geospatial formats held in the archive's collections or by linking via unique identifiers to well established file format registries. Archival in an undocumented format is not permitted. One of the key components of sustainable long term preservation is detailed

knowledge of the file format that houses data, because even file formats that are currently well understood may become incomprehensible over time.

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

In consideration of experience, cost, and resources, an implementation of this standard shall link to existing and acknowledged format-registries[18].

The use of a specified profile for any given file format is highly desirable knowledge to support long term preservation and 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.

A graphic representation of the file content (plotted map) shall be archived where feasible.

6.2.2 Data structures

Many geospatial data are structured and simple storage of the data is not sufficient to preserve the structure. In many cases, the storage requires software to access and interpret the data. Software is a form of representation information in the OAIS model. Consequently, the full content shall be transferred to the archive preserving the structure. This requires an archiving strategy that allows a persistent understanding of the technology for accessing this dataset.

6.2.3 Software and algorithms

When using algorithms or software to extract results out of the data, careful attention should be paid to preserve the replicability of the results. This can be facilitated by performing regular maintenance using tests and validation programs.

6.2.4 Properties of geospatial data

Geospatial data have a number of particular properties which require consideration before archiving. If the data are to be archived, an assumption about potential future uses of the data has to be made. Based on this assumption, decisions are made about which properties and their details will be archived and which others will be dropped. This is particularly important for data that have fine granularity, structured in objects or aggregated to larger units such as layers.

Some geospatial data are 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 exist between the layer units.

Often, geospatial data are stored redundantly, e.g. to serve different applications or to increase performance.

— If the data require different ways of access such as in full resolution or in reduced resolution in the form of an image pyramid, it shall be decided before preservation which of the resolution levels of the data shall be preserved.

— In some cases, only the raw data, i.e. the original data, and a description of the method of the creation of the derived product need to be archived. However, if it takes significant effort to generate derived products that are deemed valuable to the user community then those products will need to be archived as well.

— If identical data exist in redundant forms, it shall be decided before preservation which of the storage types shall be preserved.

The appraisal of geospatial data may use different criteria from 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.

For archiving, assumptions shall be made about typical geospatial data products, i.e. typical aggregations of the raw data used by the designated communities (see 7.2). 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 document 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 of the data archive shall be established in open, file based repositories, not databases, nor other complex environments. Often, this copy is called a gold copy.

EXAMPLE In the presence of multiple versions of time-series data specific temporal subsets from earlier versions are preserved as gold copies to facilitate verification of the ability to regenerate a given earlier version.

6.3 Rights/licensing

The problems of assuming sufficient control of mainly digital content information and preservation description information are addressed in three related categories, as follows (ISO 14721:2012, 3.2.2):

— 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, e.g. Creative Commons. The rights may be transferred to the archive together with the dataset and related items such as software and documentation. Where transferred, the rights shall be guaranteed by the archive.

With respect to mapping products, legal restrictions 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 requires that the datasets do not change within a certain time interval.

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

— 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. A system such as the ISO 19152 VersionedObject shall be in place to manage current and older data.

— 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.

Some organizations may choose to archive data by temporal sampling in order to save storage capacity, if it is appropriate to do so. In such cases, it is essential to ensure that the requirements of the designated community are met by any sampling constraints to ensure any data collected is a comfortable fit with the expressed future use cases. Any variations of collection intervals should be recorded in the archival information package (AIP) documentation, ensuring that current users are informed of the limits of the data being archived and future consumers of the data will be well aware of the contents of the archived dataset.

7 Geospatial information model

7.1 Overview

ISO 14721:2012, 2.2 and 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 what allows the designated community to understand the data. For geospatial information, the representation information is covered by ISO 19115-1, ISO 19157, ISO 19110, and ISO 19165 (this document). An example of representation information is the description of the dataset coordinate reference system (CRS).

The OAIS model also defines 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. Content information and preservation description information are both contained in information packages (IPs). See Annex B for details.

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

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

This document maps the PDI of ISO 14721 into an extension of the ISO 19115-1 metadata standard. The extension follows the rules specified in ISO 19115-1:2014, C.4 "Rules for creating an extension". This document has not been designed as a community profile and does not consider ISO 19115-1:2014, C.6 as it does not impose any new requirements on top of the classes defined in ISO 19115-1. For example, it does not make any optional elements in ISO 19115-1 mandatory. That is also the reason why this document does not reproduce the ISO 19115-1 classes, but only refers to them. The class diagram in Figure 1 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. For the sake of clarity, not all classes of the preservation metadata are shown. They are explained in detail below.

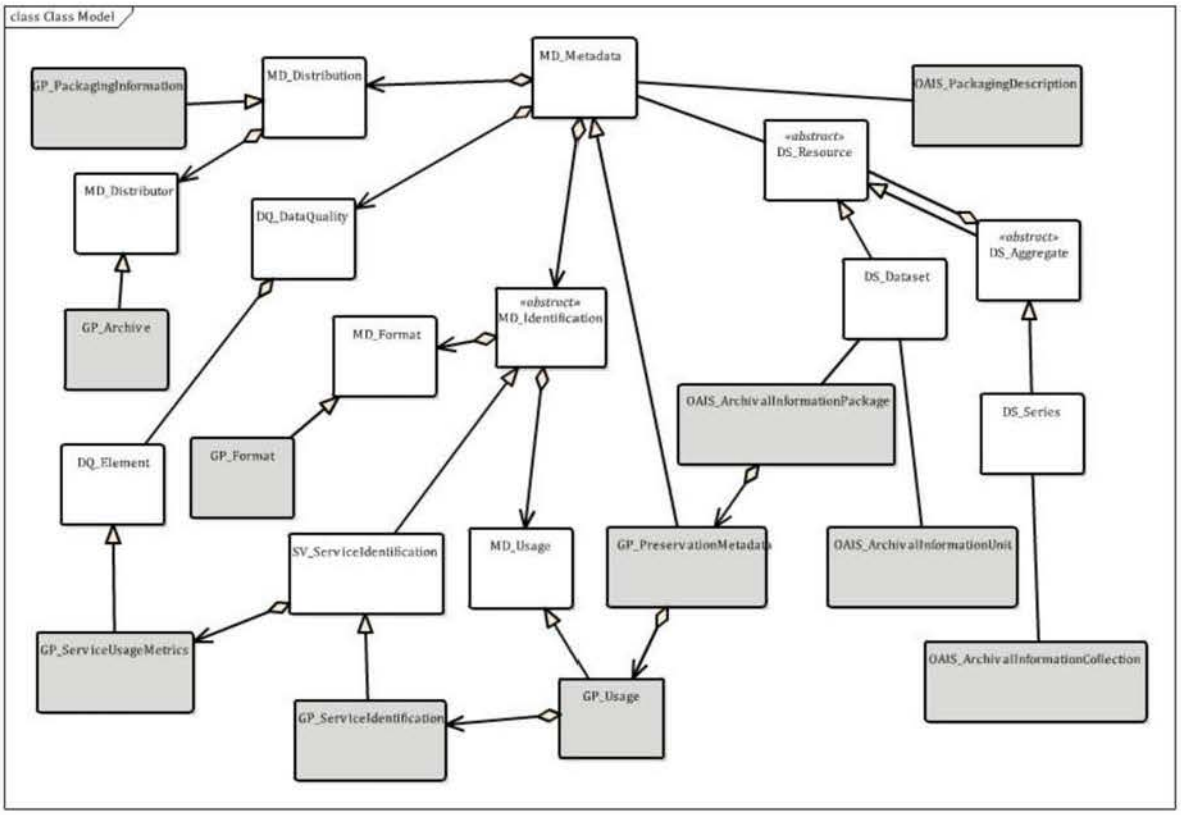

Figure 1 — Specialization of ISO 19115-1 for the preservation of geospatial data and metadata

The class diagram in Figure 2 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.

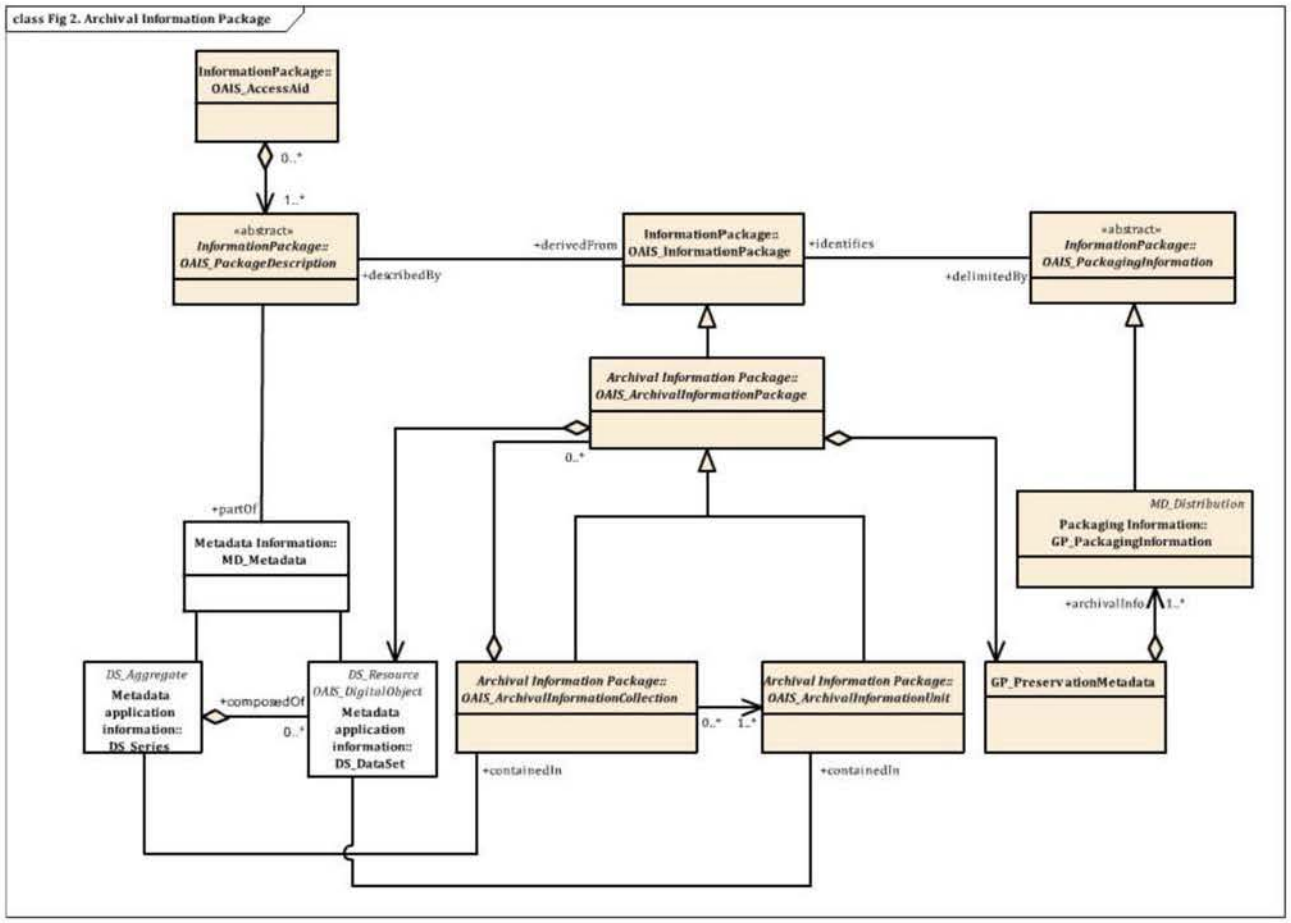

Figure 2 — Data sets and collections in geospatial data mapped to the OAIS model

7.2 Designated community

The designated community is a concept defined in the ISO 14721. This section describes its application to geospatial data.

A designated community is an identified group for whom the data are being preserved and who should be able to understand a particular set of information (4.10). A designated community is defined by the archive and includes information about expected users, designated purposes, typical software product for processing the data, companies and government organizations that use the data, other data typically linked to the preserved dataset, and the region of the usage. The definition of the designated community is useful because we can assume a minimum level of "common knowledge" that will not be required to be initially included in the information package. Typical current designated communities include land survey administrations, cadastral administration, and space imaging enterprises. The archive can assume that the communities have a minimum digital technologies background and knowledge about how to present and overlay geospatial information (a minimum set of geospatial information systems expertise).

The present day designated community should be defined 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. This is important since it is not possible to imagine what all future use cases of geospatial data may look like and therefore to articulate the specific requirements of a future designated community.

Any newly expressed use cases encountered may redefine the designated community or affect the knowledge background of the community in the future. Such changes shall be processed by the archive upon each periodic review of the dataset throughout its lifecycle, and any new elements of the metadata

and/or data added into the dataset AIP and SIP construction to allow the new designated community to be able to understand the data. For example, if the official CRS for the area changes and over time, the designated community could "forget" the existence of the old one and it may be necessary to review the archive and add a full CRS description for each archived package affected by the change.

7.3 Metadata

7.3.1 Introduction to metadata

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

A CRS is one of the most important metadata elements for geospatial data because a full recovery of an archived dataset requires the CRS and its properties. The archive shall include information about the underlying coordinate reference system.

Most of the geospatial metadata corresponds to the OAIS concept of representation information. However, preserving data for a future use requires extra effort on the metadata content that corresponds to the preservation description information. This document extends ISO 19115-1:2014 to allow preservation description information to be included.

The process of preservation is not just an act of saving everything. It starts by determining and selecting what data are going to be preserved and what data, or metadata are not worth being preserved. Examples of reasons for not preserving some data are: existence of a better revision of the same data, the existence of another format that is easier to use, or the possibility to regenerate the data from other sources. 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 preservation 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 provide 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 needs of the designated community and types of datasets (e.g. maps, remotely sensed data from satellites and airborne instruments, physical samples). Follow-up parts to this standard may be developed detailing content items appropriate to individual disciplines.

7.3.2 Preservation metadata classes

The class diagram in Figure 3 shows the class GP_PreservationMetadata as a specialization of MD_Metadata and the subclasses introduced in this standard.

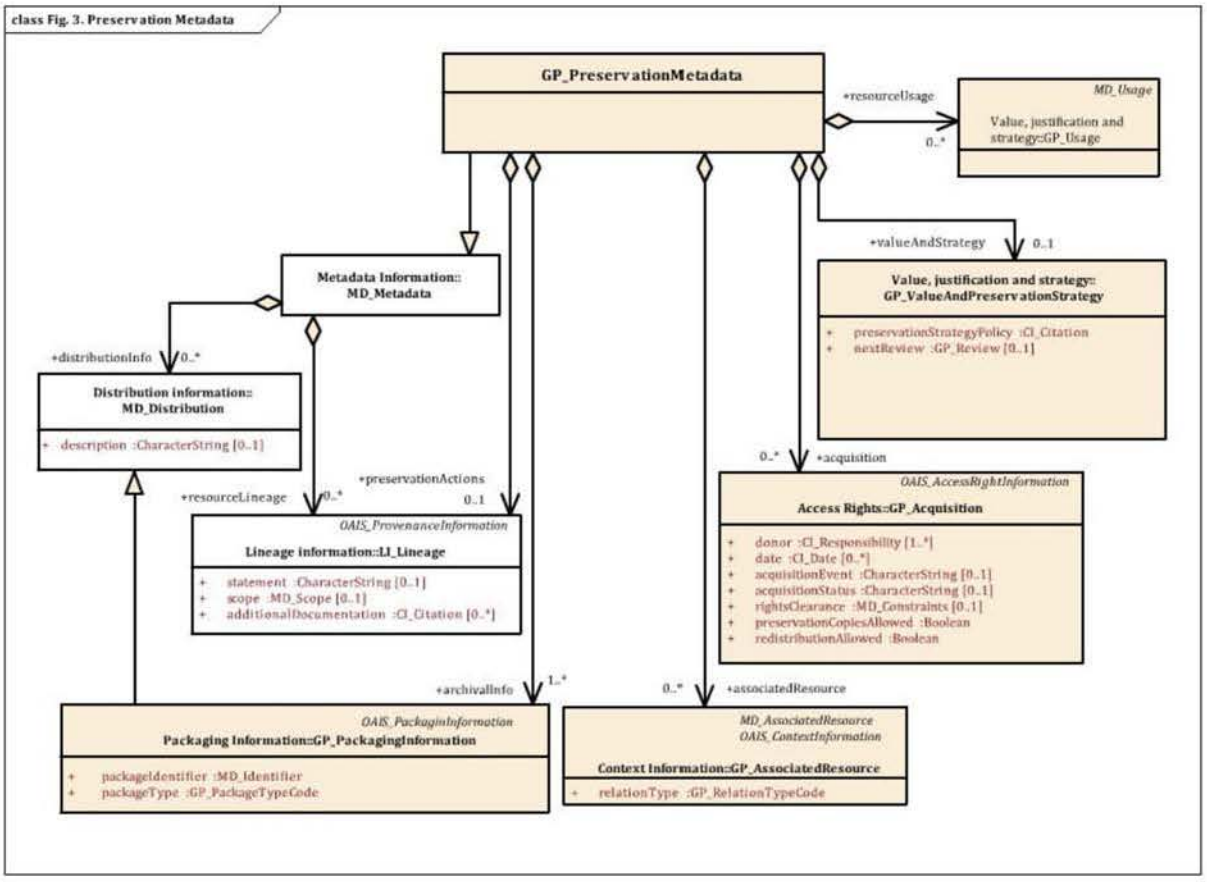

Figure 3 — Class GP_PreservationMetadata

7.3.3 Data identifiers

The class diagram in Figure 4 includes a specialization of MD_Identification and emphasizes the importance of including all resource identifiers that will help to identify this resource unique and avoid preserving it more than once. 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.

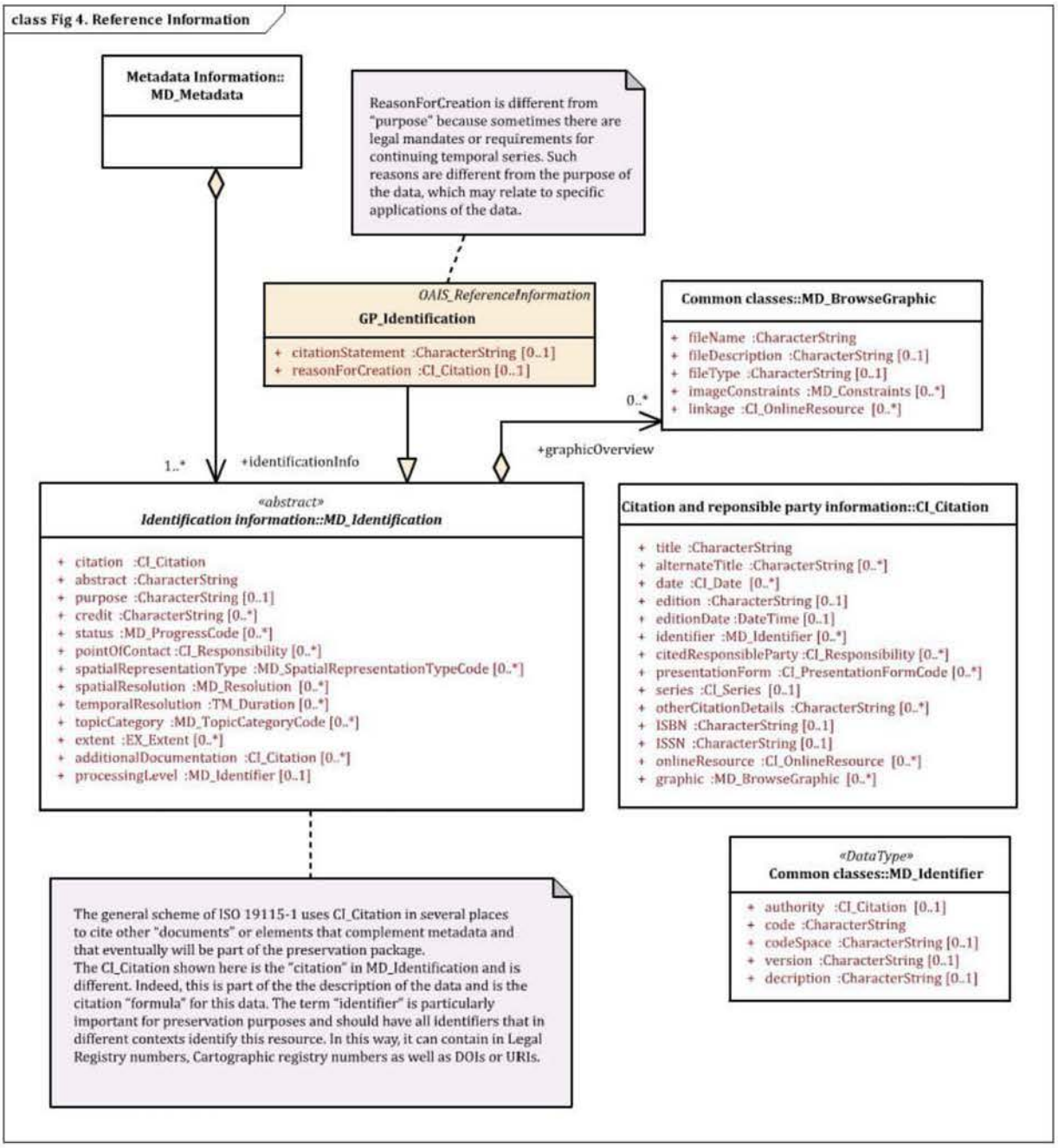

Figure 4 — Class GP_Identification

7.3.4 Data, product and format specifications

The class diagram in Figure 5 shows how the information regarding the product specifications, data dictionary references, and common format specifications can be linked to ISO 19115-1 and the need for a format specification profile specific to this product.

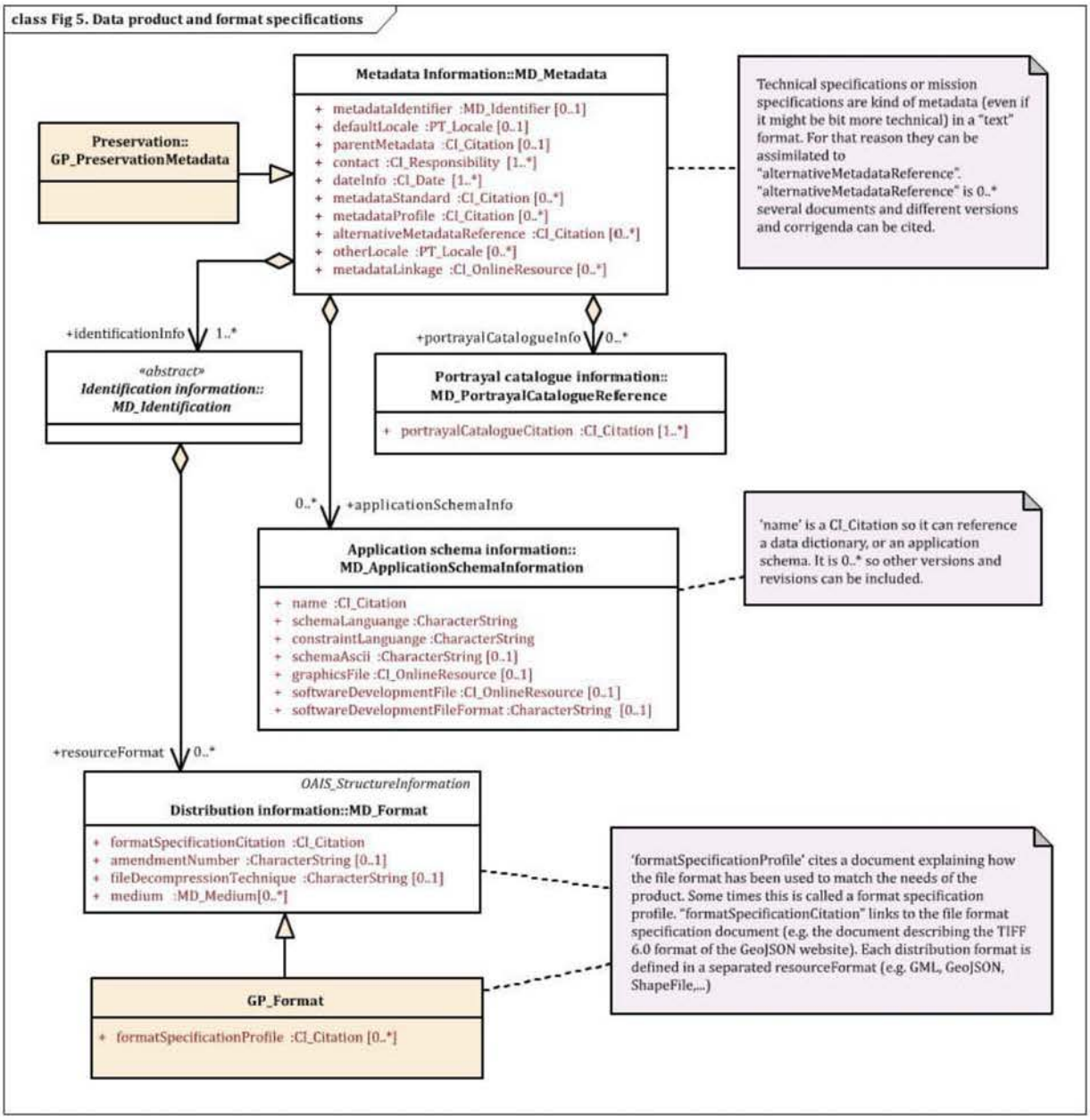

Figure 5 — Data, product and format specifications

7.3.5 Preservation actions

Preservation actions are performed as a part of the curation process when the creation and maintenance of the resource by the producer 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). Figure 6 shows how preservation can be encoded with the LI_Lineage class.

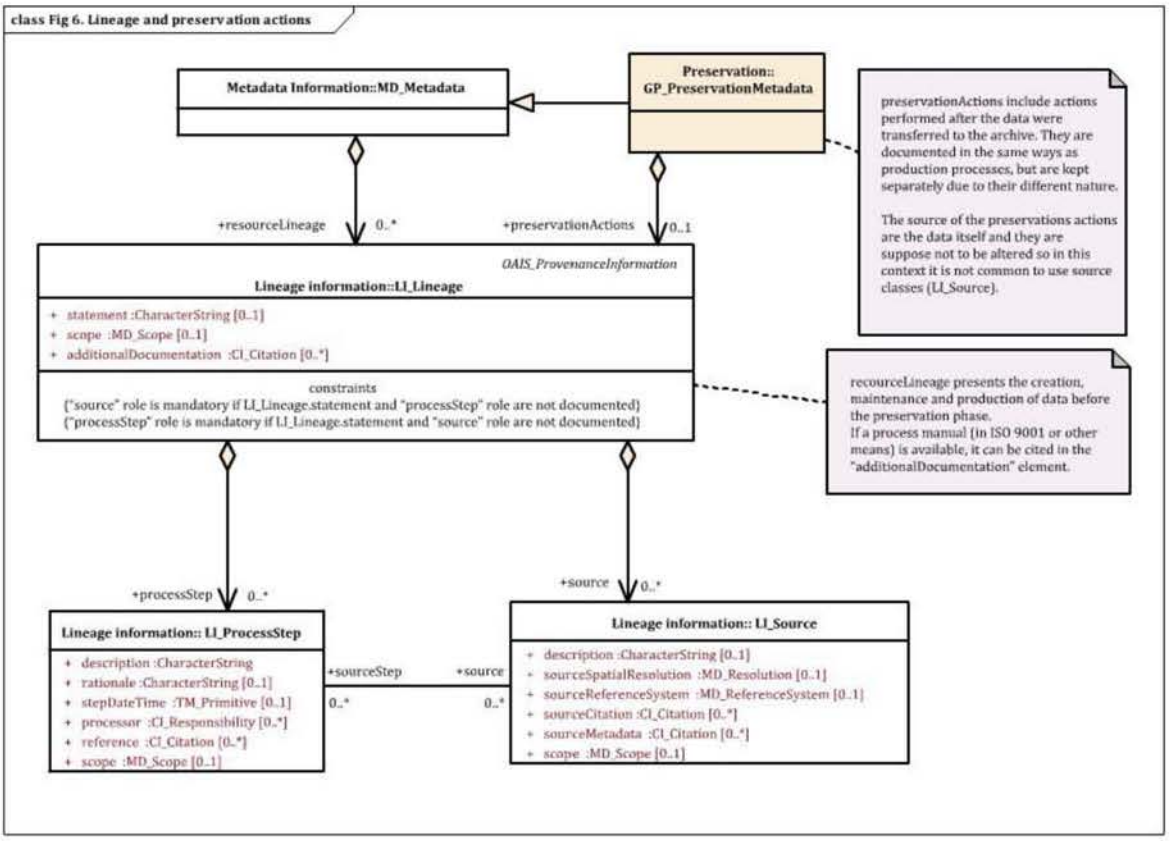

Figure 6 — Lineage and preservation actions information

7.3.6 Association of resources

The class diagram in Figure 7 shows the association of other resources and to express the context of the content information.

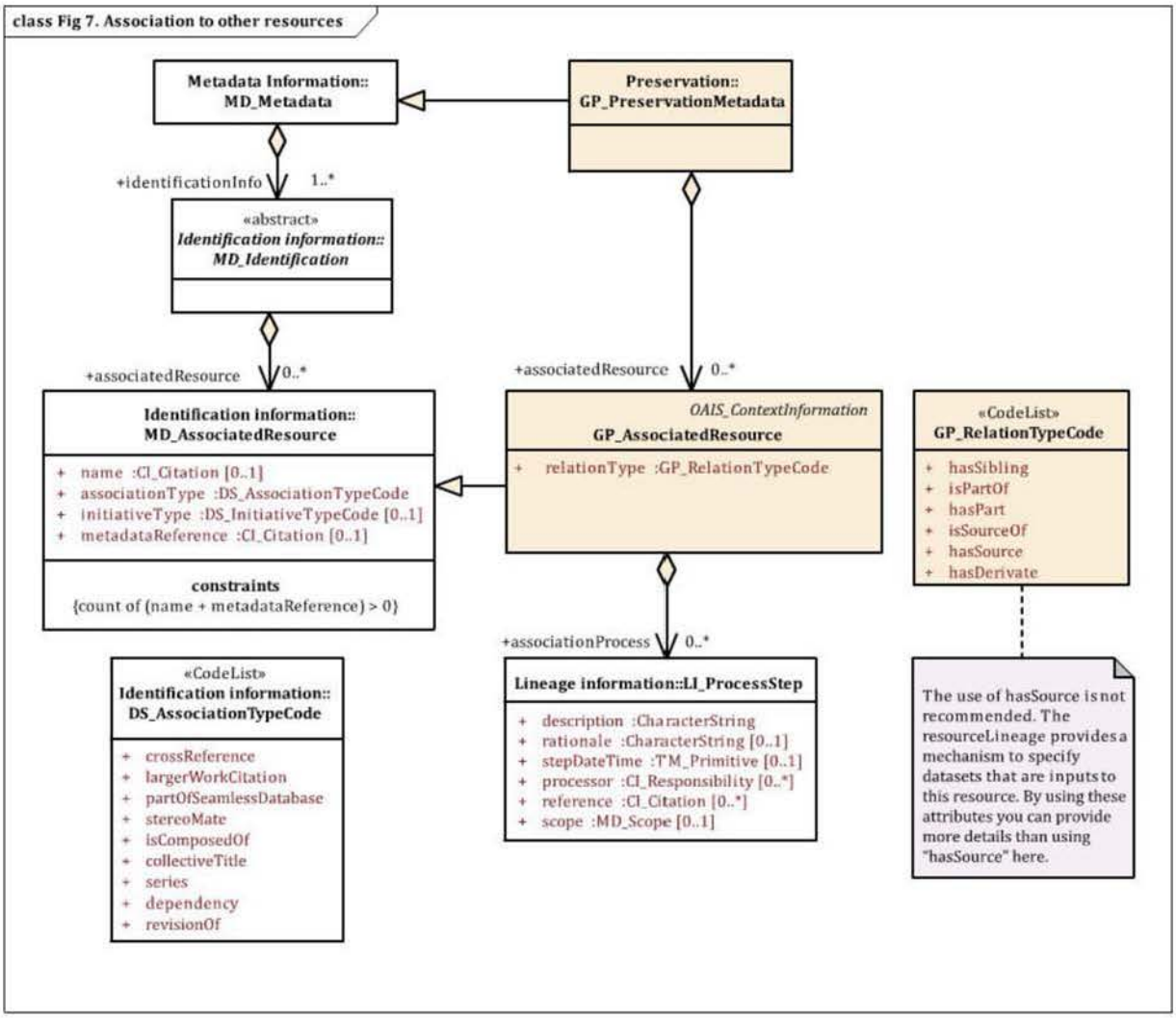

Figure 7 — Association of other resource

7.3.7 Value and strategy

The value, use and justification components address 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 Figure 8 shows the classes involved.

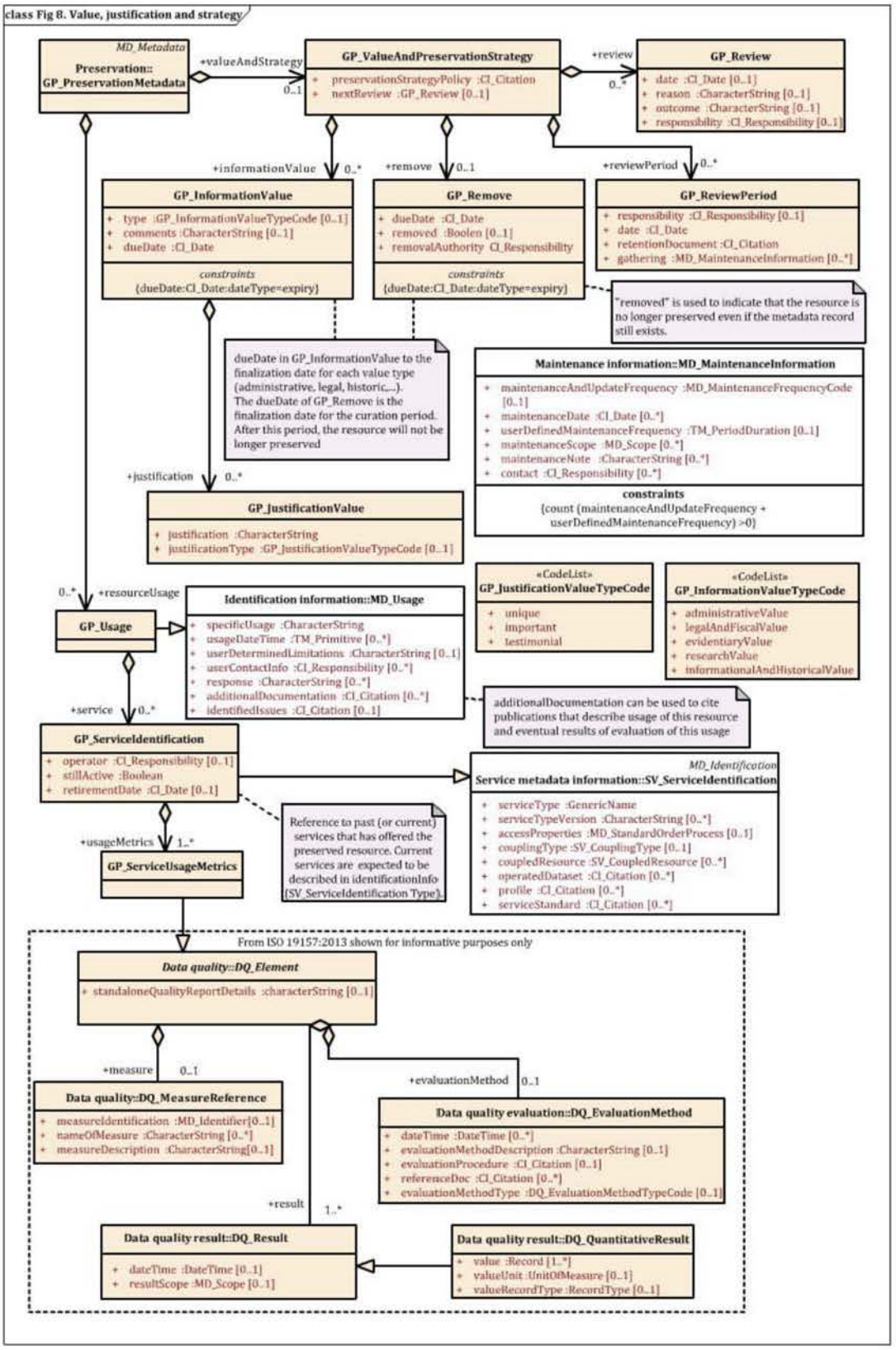

Figure 8 — Value and strategy classes diagram

7.3.8 Fixity

The questions related to verification of integrity, signature and certification of data are not discussed in this document because many digital repositories have their own systems. The defined class allows

citing the methods used and, eventually, also allows for providing integrity and signature information if there is the need to include them in the metadata. The class diagram in Figure 9 shows the class involved. Fixity information can be captured at the package level or at the individual part level as shown in Figure 10.

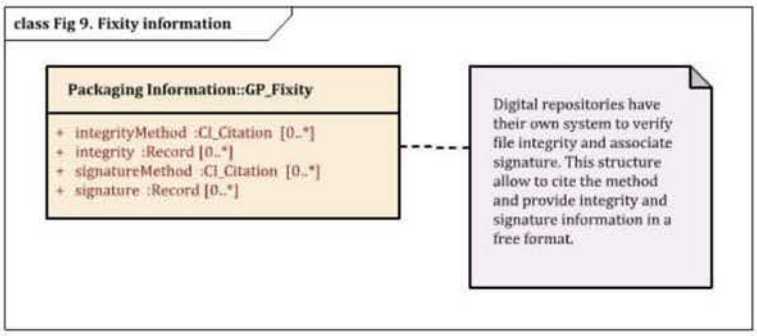

Figure 9 — Fixity class diagram

7.3.9 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 Figure 10 shows all classes related to the information package description.

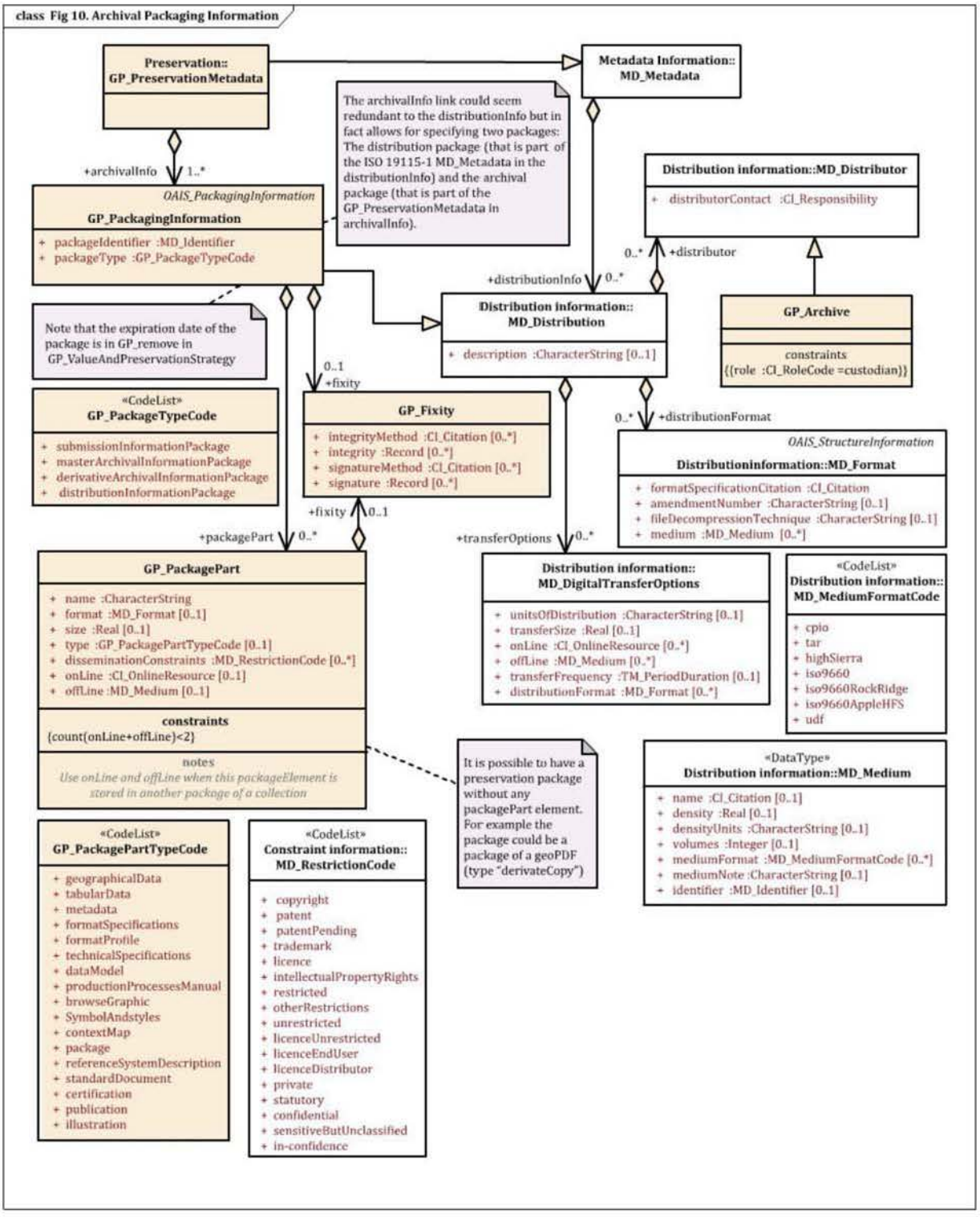

Figure 10 — Information package description

7.3.10 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 Figure 11 shows the class related to the acquisition.

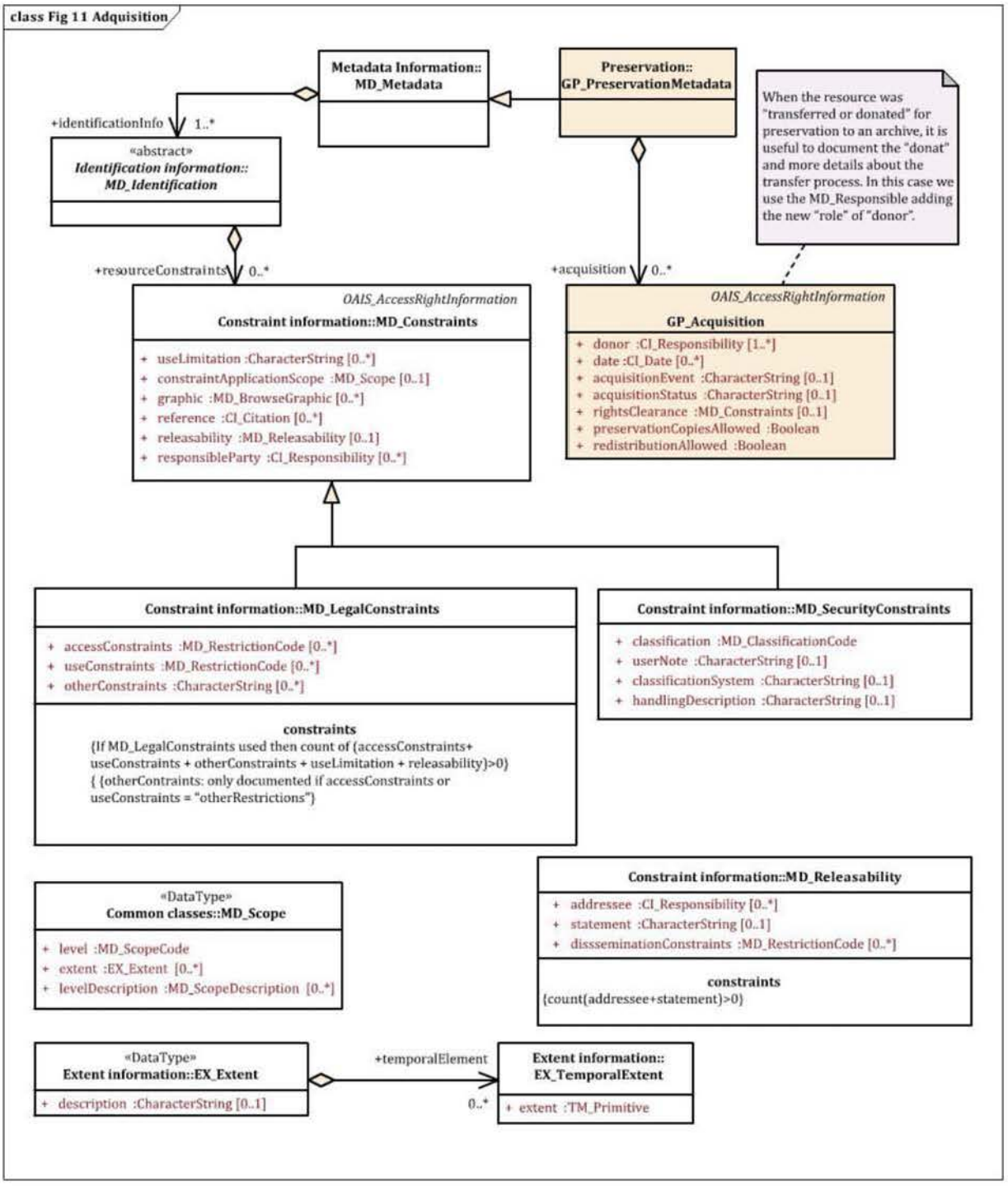

Figure 11 — Acquisition

7.3.11 Coordinate reference system

The coordinate reference system (CRS) designates the resource as geospatial. This document sets three methods for defining a CRS:

1) Method defined in the ISO 19115-1;

2) Full description of the CRS together with the archived data;

3) Reference to a format-registry.

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

ISO 19115-1 only explicitly describes referencing CRSs and suggests using ISO 19111[1] and ISO 19111-2[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 and for a SIP, for preservation purposes the archive should consider 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 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 CRS-registries based on ISO standards.

7.4 Open Packaging Convention (OPC)

The OAIS defines the idea of the information package as a container that includes the content information, the preservation description information (PDI), and descriptive information. This is particularly relevant for the geospatial information. The way geospatial data are disseminated on the Internet has forced a separation between data and metadata for practical reasons. It is common practice that geospatial data are hosted and disseminated in the producer's services while metadata are 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. The OAIS information package offers us a mechanism to use a "container" where data and metadata are kept together.

In fact, OAIS defines three different information packages depending on their function. To submit information to an archive, a SIP is proposed. To archive the data, an AIP is proposed. To disseminate the archived data to users, a 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 as an example 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.

Currently, a possible medium for package exchange is the World Wide Web with 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 hypertext transfer protocol (HTTP) interaction. In the web, the file can be associated to a multipurpose internet mail extension (MIME) type that web clients can easily deal with by delegating its management to a client helper application. In the future, new technologies may replace the mentioned ones. However, the need for addressing the package as a single address will most certainly remain.

This document proposes to use Open Packaging Convention (OPC) format that is described by ISO/IEC 29500-2 (and ECMA-376[10]) 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. This guarantees 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 information 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 is 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 ISO 19115-1 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 a 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/dateType
creator	MD_Identification/pointOfContact[CI_Responsibility/role = resourceProvider]/CI_Responsability/party/CI_Party/name
description	MD_Identification/abstract
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_Responsibility/party/CI_Party/name|MD_DataIdentification/environmentDescription
lastPrinted	MD_Identification/citation/CI_Citation.date[CI_Date/dateType = publication]/CI_Date/dateType
modified	MD_Identification/citation/CI_Citation.date[CI_Date/dateType = revision]/CI_Date/dateType
revision	*no direct match*[a]
subject	MD_Identification/purpose
title	MD_Identification/citation/CI_Citation/title
version	MD_Identification/citation/CI_Citation/edition

[a] Note: Revision: No direct match can be found in the ISO 19115-1 metadata model. In case the producer has a revision number he/she most probably will save it in the ISO metadata in MD_Identification/citation/CI_Citation/otherCitationDetails.

Annex A
(normative)

Abstract test suite

A.1 Semantics

Conformance to this document consists of either service conformance or data conformance.

The abstract test suite has seven conformance classes:

a) Prioritization;

b) Structure;

c) Rights;

d) Time;

e) Geospatial information package;

f) Metadata;

g) Open Packaging Convention.

A.2 Prioritization

a) Test Purpose: to verify the use of the appropriate temporal classification into the categories set by this document.

b) Test Method: Inspect the archival information package.

c) Reference: 6.1.

A.3 Structure

A.3.1 Data format

a) Test Purpose: to verify the fulfillment of all demands set by this document for preserving the knowledge about the data format.

b) Test Method: Inspect the archival information package.

c) Reference: 6.2.1.

A.3.2 Database

a) Test Purpose: to verify the fulfillment of all requirements set by this document for preserving the knowledge about the database.

b) Test Method: Inspect the archival information package.

c) Reference: 6.2.2.

A.3.3 Properties of geospatial data

a) Test Purpose: to verify that the properties of geospatial data have been considered.

b) Test Method: Inspect the archival information package.

c) Reference: 6.2.3.

A.3.4 Level of aggregation

a) Test Purpose: to verify that the level of aggregation has been considered.

b) Test Method: Inspect the archival information package.

c) Reference: 6.2.4.

A.3.5 Gold copy

a) Test Purpose: to verify the existence of a gold copy.

b) Test Method: Inspect the archival information package.

c) Reference: 6.2.5.

A.4 Rights

a) Test Purpose: to verify that all rights associated with the data have been considered.

b) Test Method: Inspect the archival information package.

c) Reference: 6.3.

A.5 Time

a) Test Purpose: to verify that all temporal aspects, e.g. preservation intervals, that are defined by this document, have been considered.

b) Test Method: Inspect the archival information package.

c) Reference: 6.4.

A.6 Geospatial information package

a) Test Purpose: to verify the completeness of the geospatial IP to meet the designated community requirements.

b) Test Method: Inspect the information package.

c) Reference: 7.1.

A.7 Metadata

A.7.1 Preservation metadata

a) Test Purpose: to verify the use of the appropriate metadata elements for the preservation metadata.

b) Test Method: Inspect the archival information package.

c) Reference: 7.3.2.

A.7.2 Data identifier

a) Test Purpose: to verify the use of the appropriate metadata elements for data identifier.

b) Test Method: Inspect the archival information package.

c) Reference: 7.3.3.

A.7.3 Data, product, and format specification

a) Test Purpose: to verify the use of the appropriate metadata elements for the data, product and format specification.

b) Test Method: Inspect the archival information package.

c) Reference: 7.3.4.

A.7.4 Preservation actions

a) Test Purpose: to verify the use of the appropriate metadata elements for preservation actions.

b) Test Method: Inspect the archival information package.

c) Reference: 7.3.5.

A.7.5 Association of resources

a) Test Purpose: to verify the use of the appropriate metadata elements for the association of resources.

b) Test Method: Inspect the archival information package.

c) Reference: 7.3.6.

A.7.6 Value and strategy

a) Test Purpose: to verify the use of the appropriate metadata elements for the value and strategy metadata.

b) Test Method: Inspect the archival information package.

c) Reference: 7.3.7.

A.7.7 Fixity

a) Test Purpose: to verify the appropriate addressing of fixity.

b) Test Method: Inspect the archival information package.

c) Reference: 7.3.8.

A.7.8 Information package description

a) Test Purpose: to verify the use of the appropriate metadata elements for the information package description.

b) Test Method: Inspect the archival information package.

c) Reference: 7.3.9.

A.7.9 Acquisition

a) Test Purpose: to verify the use of the appropriate metadata elements for the description of acquisition.

b) Test Method: Inspect the archival information package.

c) Reference: 7.3.10.

A.7.10 Coordinate reference system

a) Test Purpose: to verify correctness of describing the associated coordinate reference system.

b) Test Method: Inspect the archival information package.

c) Reference: 7.3.11.

A.8 Open Packaging Convention

a) Test Purpose: to verify correctness of mapping to the Open Packaging Convention model.

b) Test Method: Inspect the archival information package.

c) Reference: 7.4.

Annex B
(normative)

Data dictionary

B.1 Data dictionary

The following subclause provides a detailed description of each of the classes and each class attribute in the models presented in this document in the form of a tabular data dictionary.

B.2 Presentation metadata

	Name/Role Name	Definition	Obligation/ Condition	Max Occurrence	Data Type/ Class	Domain
1.	GP_Preservation-Metadata 7.3.2	root entity which defines preservation metadata about a resource or resources	M	Use maximum occurrence from referencing object if referenced from DS_Resource	Specified Class (MD_ Metadata)	lines 2–7
2.	acquisition	details about the archival acquisition	O	N	Class	GP_Acquisition
3.	associatedResource	associated resource information	O	N	Class	GP_AssociatedResource
4.	archivalInfo	description of the information package	M	N	Class	GP_Packaging information
5.	resourceUsage	basic information about specific application(s) for which the resource has been or is being used by different users	O	N	Class	GP_Usage
6.	valueAndStrategy	requirements for assessing the value of the resource and its future preservation strategy	O	1	Class	GP_ValueAndPreservationStrategy
7.	preservationActions	information about the provenance, source(s), and/or the production process(es) applied to the resource by the archive.	O	1	Class	LI_Lineage (ISO 19115-1)
8.	GP_Usage 7.3.2	brief description of ways in which the resource(s) is/are currently or has been used	Use obligation/ condition from referencing object	Use maximum occurrence from referencing object	Specified Class (MD_Usage)	line 9
9.	service	services that currently or has been using this resource(s)	M	N	Class	GP_ServiceIdentification
10.	GP_ValueAndPreservationStrategy 7.3.2	reason, justification and strategy for archiving and curating the resource(s)	Use obligation/ condition from referencing object	Use maximum occurrence from referencing object	Specified Class (GP_ValueAndPreservationStrategy)	lines 11–16

	Name/Role Name	Definition	Obligation/ Condition	Max Occurrence	Data Type/ Class	Domain
11.	preservationStrategyPolicy 7.3.2	document reference with the archive preservation strategy	M	1	Class	CI_Citation (ISO 19115-1)
12.	nextReview 7.3.2	information about the next foreseen review of the archived package including date and intention/reason	O	1	Class	GP_Review
13.	informationValue	reason and justification of the value of the resource	O	N	Class	GP_InformationValue
14.	remove	foreseen date for the finalization of the preservation	O	1	Class	GP_Remove
15.	review	dates of the preservation revisions	O	N	Class	GP_Review
16.	reviewPeriod	periodicity of the foreseen reviews	O	N	Class	GP_ReviewPeriod
17.	GP_Acquisition 7.3.2, 7.3.10	information about the acquisition of the resource(s)	Use obligation/ condition from referencing object	Use maximum occurrence from referencing object	Specified Class (GP_Acquisition)	lines 18–24
18.	donor 7.3.2, 7.3.10	party responsible that delivered the resource(s) to the archive	M	N	Class	CI_Responsibility (ISO 19115-1)
19.	date	date of the delivery of the resource(s) to the archive	O	N	Class	CI_Date (ISO 19115-1)
20.	acquisitionEvent 7.3.2, 7.3.10	description of the acquisition process	O	1	CharacterString	Free text
21.	acquisitionStatus 7.3.2, 7.3.10	status of the acquisition process	O	1	CharacterString	Free text
22.	rightsClearance 7.3.2, 7.3.10	statement describing how and with whom the resource can be shared	O	1	Class	MD_Constraint (ISO 19115-1)
23.	preservationCopiesAllowed 7.3.2	indication of whether or not the resource(s) has can be duplicated by the archive	O	1	Boolean	0 = no 1 = yes
24.	redistributionAllowed 7.3.10	indication of whether or not the resource(s) can be redistributed	O	1	Boolean	0 = no 1 = yes
25.	GP_AssociateResource 7.3.2, 7.3.6	definition	Use obligation/ condition from referencing object	Use maximum occurrence from referencing object	Specified Class (MD_AssociateResource)	lines 26–27
26.	relationType 7.3.2	relation dependency	M	1	Class	GP_RelationTypeCode <<Codelist>>
27.	associationProcess	process that made the relation between this resource and the cited resource	O	N	Class	LI_ProcessStep (ISO 19115-1)
28.	GP_Identification 7.3.3	identification of the archived resource(s)	Use obligation/ condition from referencing object	Use maximum occurrence from referencing object	Specified Class (MD_Identification)	lines 29–30

	Name/Role Name	Definition	Obligation/ Condition	Max Occurrence	Data Type/ Class	Domain
29.	citationStatement 7.3.3	indication on how you have to cite or knowledge the attribution of the resource(s)	O	1	Character-String	Free text
30.	reasonForCreation 7.3.3	reason for creating the resource. In many cases this is different from the "purpose". In the citations title one can use "legal mandate" or "data continuity of the temporal series". If there is a legal mandate, one can be more precise by fully citing a legal document or a contract containing the reason or the mandate.	O	1	Class	CI_Citation (ISO 19115-1)
31.	GP_Format 7.3.4	format stored in the archive for this resource(s).	Use obligation/ condition from referencing object	Use maximum occurrence from referencing object	Specified Class (MD_Format)	line 32
32.	formatSpecification-Profile 7.3.4	cites a document explaining how the file format has been used to match the needs of the product. Sometimes this is called a "format specification profile" ('formatSpecificationCitation' links to the file format specification document (e.g the document describing the TIFF 6.0 format).	O	N	Class	CI_Citation (ISO 19115-1)
33.	GP_ReviewPeriod 7.3.7	information about the review policy and period of the archiving resource	Use obligation/ condition from referencing object	Use maximum occurrence from referencing object	Specified Class (GP_ReviewPeriod)	lines
34.	responsibility 7.3.7	responsible party for the revision of this resource of the archive	O	1	Class	CI_Responsibility (ISO 19115-1)
35.	date 7.3.7	date of the next foreseen review	M	1	Class	CI_Date ISO 19115-1
36.	retentionDocument 7.3.7	reference to a document that describes the revision procedure	M	1	Class	CI_Citation ISO 19115-1
37.	gathering 7.3.7	information about the periodicity of the revision	O	N	Class	MD_Maintenan-ceInformation ISO 19115-1
38.	GP_InformationValue 7.3.7	reason and justification for archiving the resource	Use obligation/ condition from referencing object	Use maximum occurrence from referencing object	Specified Class (GP_Informa-tionValue)	lines 39–42
39.	type 7.3.7	the reason for archiving the resource and gives it value	O	1	Class	GP_InformationVal-ueTypeCode <<Codelist>>

	Name/Role Name	Definition	Obligation/ Condition	Max Occurrence	Data Type/ Class	Domain
40.	comments 7.3.7	additional explanation about the reason for archiving the resource	O	1	Character-String	Free text
41.	dueDate 7.3.7	expiration date of the reason for archiving the resource	O	1	Class	CI_Date (ISO 19115-1)
42.	justification	justification of importance of the resource in context of the general reason	O	N	Class	GP_Justification-Value
43.	GP_Remove 7.3.7	information about the intent of removing the resource in the future or evidence that the resource has been removed.	Use obligation/ condition from referencing object	Use maximum occurrence from referencing object	Specified Class (GP_Remove)	lines 44–46
44.	dueDate 7.3.7	finalization date for the archival period. After this period, the resource will not be longer preserved	M	1	Class	CI_Date (ISO 19115-1)
45.	removed 7.3.7	indication of whether or not the resource(s) has been removed and is no longer archived	O	1	Boolean	0 = no 1 = yes
46.	removeAuthority 7.3.7	responsible party that has the decision on removing the resource form the archive	M	1	Class	CI_Responsibility (ISO 19115-1)
47.	GP_Review 7.3.7	information about a review (foreseen or done)	Use obligation/ condition from referencing object	Use maximum occurrence from referencing object	Specified Class (GP_Review)	lines 48–51
48.	date 7.3.7	date of the review	M	1	Class	CI_Date (ISO 19115-1)
49.	reason 7.3.7	reason for the review	O	1	Character-String	Free text
50.	outcome 7.3.7	outcome of the review	O	1	Character-String	Free text
51.	responsibility 7.3.7	responsible party for the review	O	1	Class	CI_Responsibility (ISO 19115-1)
52.	GP_JustificationValue 7.3.7	Explicit justification why this resource is important in the general context	Use obligation/ condition from referencing object	Use maximum occurrence from referencing object	Specified Class (GP_JustificationValue)	lines 53–54
53.	justification 7.3.7	statement on why this resource is important	M	1	Character-String	Free text
54.	justificationType 7.3.7	justification why this resource is important	O	1	Class	GP_Justification-TypeCode
55.	GP_ServiceIdentification 7.3.7	estimation of the usage of the resource by registering a service and service metrics that host or hosted the resource	Use obligation/ condition from referencing object	Use maximum occurrence from referencing object	Specified Class (GP_ServiceIdentification)	line 56–59
56.	operator	Who was (or is) the responsible party maintaining the service	O	1	Class	CI_Responsibility (ISO 19115-1)

	Name/Role Name	Definition	Obligation/ Condition	Max Occurrence	Data Type/ Class	Domain
57.	stillActive	indication of whether or not the service has been retired and is no longer active	M	1	Boolean	0 = no 1 = yes
58.	retirementDate	retirement date of the service (in the past or in the future).	O	1	Class	CI_Date (ISO 19115-1)
59.	usageMetrics	measure of the usage of the service	M	N	Class	GP_ServiceUsageMetrics
60.	GP_ServiceUsageMetrics 7.3.7	a measure value and method to determine the service usage	Use obligation/ condition from referencing object	Use maximum occurrence from referencing object	Specified Class (DQ_Element)	ISO 19157
61.	GP_Fixity	citation of the integrity and signature methods used in the packaging information followed by the actual integrity and signature information	Use obligation/ condition from referencing object	Use maximum occurrence from referencing object	Specified Class (GP_Fixity)	liness 62–65
62.	integrityMethod	citation of the integrity method specification	O	N	Class	CI_Citation (ISO 19115-1)
63.	integrity	integrity data	O	N	Class	Record (ISO 19115-1)
64.	signatureMethod	citation of the signature method specification	O	N	Class	CI_Citation (ISO 19115-1)
65.	signature	signature data	O	N	Class	Record (ISO 19115-1)
66.	GP_PackagingInformation 7.3.2, 7.3.9	information about a package	Use obligation/ condition from referencing object	Use maximum occurrence from referencing object	Specified Class (GP_PackagingInformation)	lines 67–70
67.	packageIdentifier 7.3.2	identifier of the package	M	1	Class	MD_Identifier (ISO 19115-1)
68.	packageType 7.3.2	type of the package	M	1	Class	GP_PackageType <<Codelist>>
69.	fixity	fixity information for the package	O	1	Class	GP_Fixity
70.	packagePart	part in a package	C	N	Class	GP_PackagePart
71.	GP_PackagePart 7.3.9	information about a package part	Use obligation/ condition from referencing object	Use maximum occurrence from referencing object	Specified Class (GP_PackageElement)	lines 72–79
72.	name 7.3.9	part name	M	1	CharacterString	Free text
73.	format 7.3.9	format of the part	O	1	Class	MD_Format (ISO 19115-1)
74.	size 7.3.9	size of the part in bytes	O	1	Real	
75.	type 7.3.9	type of part	O	1	Class	GP_PackagePartCode <<Codelist>>

	Name/Role Name	Definition	Obligation/ Condition	Max Occurrence	Data Type/ Class	Domain
76.	disseminationConstraints 7.3.9	dissemination constraints that apply to this part only	O	N	Class	MD_Restriction-Code (ISO 19115-1)
77.	onLine 7.3.9	link to an online resource to get the part	C	1	Class	CI_OnlineResource (ISO 19115-1)
78.	offLine 7.3.9	description of the medium for the part in the package	C	1	Class	MD_Medium (ISO 19115-1)
79.	fixity	fixity information for the part in the package	O	1	Class	GP_Fixity
80.	GP_Archive 7.3.10	Description of the archive as a distributor.	Use obligation/ condition from referencing object	Use maximum occurrence from referencing object	Specified Class (MD_ Distributor)	

B.3 CodeLists

B.3.1 General

The stereotype class <<CodeList>> as used in this document, does not contain any "other" values and are extendable. Consult ISO 19115-1:2014, Annex C for information about how to extend <<CodeList>>. The concept name is the name of the item (this document is in English and should be translated into the language of the nation or entity developing a profile). The code is a language neutral identifier.

B.3.2 GP_RelationTypeCode < <CodeList> >

	Name	Domain Code	Definition
1.	GP_RelationTypeCode 7.3.6	reTyCd	type of relation
2.	hasSibling	001	the resource(s) has a sibling associated resource
3.	isPartOf	002	the resource(s) is part of a bigger associated resource
4.	hasPart	003	the resource(s) is composed by smaller associated resources
5.	isSourceOf	004	the resource(s) is the source of an associated resource
6.	hasSource	005	the resource(s) has an associated source resource
7.	hasNoRelation	006	the resource(s) has no relation to other resources

B.3.3 GP_InformationValueTypeCode < <CodeList> >

	Name	Domain Code	Definition
7.	GP_InformationValueTypeCode	InVTCd	type of reason that makes this resource valuable
8.	administrativeValue	001	administrative value
9.	legalAndFiscalValue	002	legal and/or fiscal value
10.	evidentiaryValue	003	evidentiary value
11.	researchValue	004	research value
12.	informationalAndHistorical-Value	005	informational and historical value

B.3.4 GP_JustificationValueTypeCode < <CodeList> >

	Name	Domain Code	Definition
13.	GP_JustificationValueTypeCode	JuVTCd	type
14.	unique	001	the resource is unique in its kind
15.	important	002	the resource is considered important in its kind
16.	testimonial	003	the resource is a testimonial of some important event

B.3.5 GP_PackageTypeCode < <CodeList> >

	Name	Domain Code	Definition
17.	GP_PackageTypeCode	PaTyCd	type of package
18.	submissionInformationPackage	001	submission information package (SIP)
19.	masterArchivalInformation-Package	002	master archival information package (AIP)
20.	derivativeArchivalInforma-tionPackage	003	derivative archival information package (AIP)
21.	distributionInformationPack-age	004	distribution information package (DIP)

B.3.6 GP_PackagePartTypeCode < <CodeList> >

	Name	Domain Code	Definition
22.	GP_PackagePartTypeCode	PaETCd	type of part of package
23.	geographicalData	001	geographical data
24.	tabularData	002	tabular data
25.	metadata	003	metadata
26.	formatSpecification	004	format specification
27.	formatProfile	005	format profile
28.	technicalSpecification	006	technical specification
29.	dataModel	007	data model
30.	productionProcessManual	008	production process manual
31.	browseGraphic	009	browse graphic or quick look
32.	symbolsAndStyles	010	symbols and styles
33.	contextMap	011	context or map linking a list of data resources
34.	package	012	package (in a package)
35.	referenceSystemDescription	013	reference system description
36.	standardDocument	014	standards document
37.	certification	015	certification
38.	publication	016	publication
39.	illustration	017	illustration

Annex C
(informative)

Case-specific archival concept

This annex shall support the initial steps of the preservation strategy development.

The purpose of the questionnaire is to identify the resources of the organization in charge of producing and curating the geospatial data as well as clarify the most relevant topics:

— Institution or agency;

— Contact;

— Name of the resource;

— Theme of the resource;

— Name and version of the software used in production;

— Developer and vender of software;

— Geo-reference data;

— Regional coverage;

— Update cycle and versioning;

— Storage of data history;

— Data formats;

— Computer hardware and operating system;

— Approximate volume of data;

— Availability of metadata and context information;

— Usage of geospatial base-data of other institutions;

— Other institutions involved in the production.

This process can be completed in close cooperation between the archival experts and the geospatial data domain experts, e.g. in land survey administration and spatial data infrastructure offices. The completion of this preliminary work supports the development of an archive according to this document and other related standards.

Annex D
(normative)

Functional requirements for a preservation archive

D.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 understanding and reusability of the data.

D.2 Data ingest

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

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

D.2.3 The archive ensures ingest data integrity through the use of appropriate technologies.

D.2.4 The archive verifies the quality of the preservation content upon ingest (e.g. data set size, data set name) for each unique data set.

D.3 Data storage and preservation

D.3.1 Overview

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

D.3.2 Data storage

D.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.

D.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.

D.3.2.3 The archive is capable of archiving multiple versions of selected archive data.

D.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 Clause 2.

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

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

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

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

D.3.2.9 The archive allows for new technology integration for archival data.

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

D.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.

D.3.3 Preservation management

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

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

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

D.4 Data distribution

D.4.1 The archive enables users to determine the existence, description and availability of stored data and information and allows consumers to request and receive the data and information.

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

D.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.

D.4.4 The archive distributes data to users via electronic networks.

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

D.4.6 The archive provides a subset of the data and/or subset, reprojection, and format conversion tools appropriate to the archive's data holdings to ensure efficient distribution of data to users.

D.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.

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

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

Bibliography

[1] ISO 19111:2007, *Geographic information — Spatial referencing by coordinates*

[2] ISO 19111-2:2009, *Geographic information — Spatial referencing by coordinates — Part 2: Extension for parametric values*

[3] ISO 19131:2007, *Geographic information — Data product specifications*

[4] BAR. 2013: Projekt Ellipse, Konzeption der Archivierung von Geobasisdaten des Bundesrechts, Bundesamt für Landestopografie swisstopo and Schweizerisches Bundesarchiv BAR — https://www.swisstopo.admin.ch/de/wissen-fakten/geoinformation/landschaftsgedaechtnis/langzeitaufbewahrung/archivierung-geodaten.html (22.1.2018)

[5] CEOS WGISS-DSIG. 2015, Long Term Preservation of Earth Observation Space Data: Preservation Guidelines, Version 1.0, https://earth.esa.int/documents/1656065/2265358/EO-Data-Preservation-Guidelines (22.1.2018)

[6] 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/Data_Stewardship/Best_Practices/Preservation%20Workflow_v1.0.pdf (1.3.2015)

[7] CEOSWGISS-DSIG. 2015, LongTermPreservationofEarthObservationSpaceDataEarthObservation: Preserved Data Set Content, Version 1.0, http://ceos.org/document_management/Working_Groups/WGISS/Interest_Groups/Data_Stewardship/Recommendations/EO%20Preserved%20Data%20Set%20Content_v1.0.pdf (1.9.2015)

[8] 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/Data_Stewardship/Best_Practices/GenericEarthObservationDataSetConsolidationProcess_v1.0.pdf (1.3.2015)

[9] Dorfey B., Graf S., Grau B., Homberg J., Huth K., Keitel C. 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)

[10] ECMA 376, 2016, http://www.ecma-international.org/publications/standards/Ecma-376.htm (February 19, 2016)

[11] Engel H., Engelmeier S., Grams G., Gültlinger M., Häussermann M., Hermann C. 2013: Archivierung von Geobasisdaten im Kontext der Gesamtüberlieferung des Vermessungswesens. Abschlussbericht der AG "Archivierung von Geobasisdaten des LGL", Stuttgart, Germany

[12] 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)

[13] 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)

[14] Jobst M., & Gartner G. Structural Aspects for the Digital Cartographic Heritage. In: *Preservation in Digital Cartography — Archiving Aspects*, (Jobst M. ed.). Springer, 2011

[15] 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/AppraisalSelection_whitepaper_final.pdf (18.5.2014)

[16] NASA. 2013: NASA Earth Science Data Preservation Content Specification, 423-SPEC-01, https://earthdata.nasa.gov/files/423-SPEC-001_NASA%20ESD_Preservation_Spec_OriginalCh01_0.pdf

[17] 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)

[18] PRONOM. 2015: http://apps.nationalarchives.gov.uk/PRONOM/Default.aspx (31.12.2015)

[19] PROV. 2017: https://www.w3.org/TR/prov-overview/ (24.9.2017)

[20] RÖNSDORF C., MASON P., GERBER U., BOS M., SHAON A., NAUMANN K. 2012: GI+100: Long term preservation of digital Geographic Information — 16 fundamental principles agreed by National Mapping Agencies and Archives — http://www.eurosdr-ireland.net/archiving/GI+100%20-%2016%20EuroSDR%20Archiving%20Principles_V3%201.pdf (18.5.2014)

[21] SANDNER P. 2012: Langzeitarchivierung von Geodaten — https://www.bundesarchiv.de/imperia/md/content/bundesarchiv_de/fachinformation/ark/handreichung_geodaten_20090928.pdf (22.1.2018)

[22] SHAON A., & WOOLF A. 2011: Long-term Preservation for Spatial Data Infrastructures: A Metadata Framework and Geo-portal Implementation. D-Lib. 2011 September/October, 17 (9/10)

[23] ISO 9660, *Information processing — Volume and file structure of CD-ROM for information interchange*

[24] ISO 16363, *Space data and information transfer systems — Audit and certification of trustworthy digital repositories*

[25] ISO 19108, *Geographic information — Temporal schema*

[26] ISO 19152, *Geographic information — Land Administration Domain Model (LADM)*

[27] ISO/IEC 29500-2:2012, *Information technology — Document description and processing languages — Office Open XML File Formats — Part 2: Open Packaging Conventions*

ICS 35.240.70

Price based on 42 pages