Standard course – Computer Science

Lesson SC3 –Preservation issues and solutions

Joan Masó – CREAF

Co-authors: Ester Prat, Lluís Pesquer (CREAF)
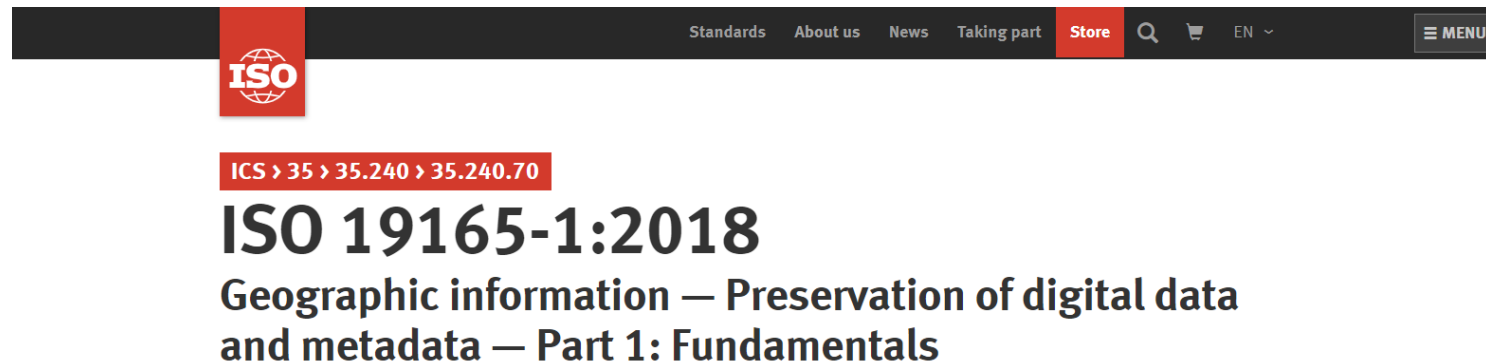
ISO 19165-1:2018 Geographic information — Preservation of digital data and metadata

ISO 19165-1:2018

- Preservation metadata extension of ISO 19115-1 :2014 - Geographic information — Metadata

- Defines the requirements for the **long-term preservation of digital geospatial data**.

- These data also include metadata, representation information, provenance, context, data formats and any other content items necessary to fully understand and reuse the archived data.

- It is inspired on the OAIS Reference Model (ISO 14721).

Geospatial Information Package (IP)

- It covers two main aspects: a **metadata model** to document preservation aspects and an **information package**.

- Geospatial data are preserved as a geospatial information package (IP).

- ISO 19165 defines the requirements of the geospatial **information archival package (IP)** as one of its central components. A geospatial archival IP is fully self-describing and allows a future reconstruction of the dataset without external documentation.

- The standard also details the **geospatial submission and the dissemination IPs**.

Considerations

- A geospatial dataset should always be linked to the corresponding metadata in a way that allows an unambiguous reference between metadata and data.

- In the geospatial domain 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.

- All data preservation-oriented tasks should allow future users to understand what they are working with (context information) and how the data was created (provenance information).

- 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).

Metadata Preservation Model

- The MD_Metadata of ISO 19115 has been extended into the GP_PreservationMetadata (that acts as the preservation description information PDI) by including some extra classes for the purpose of preservation.
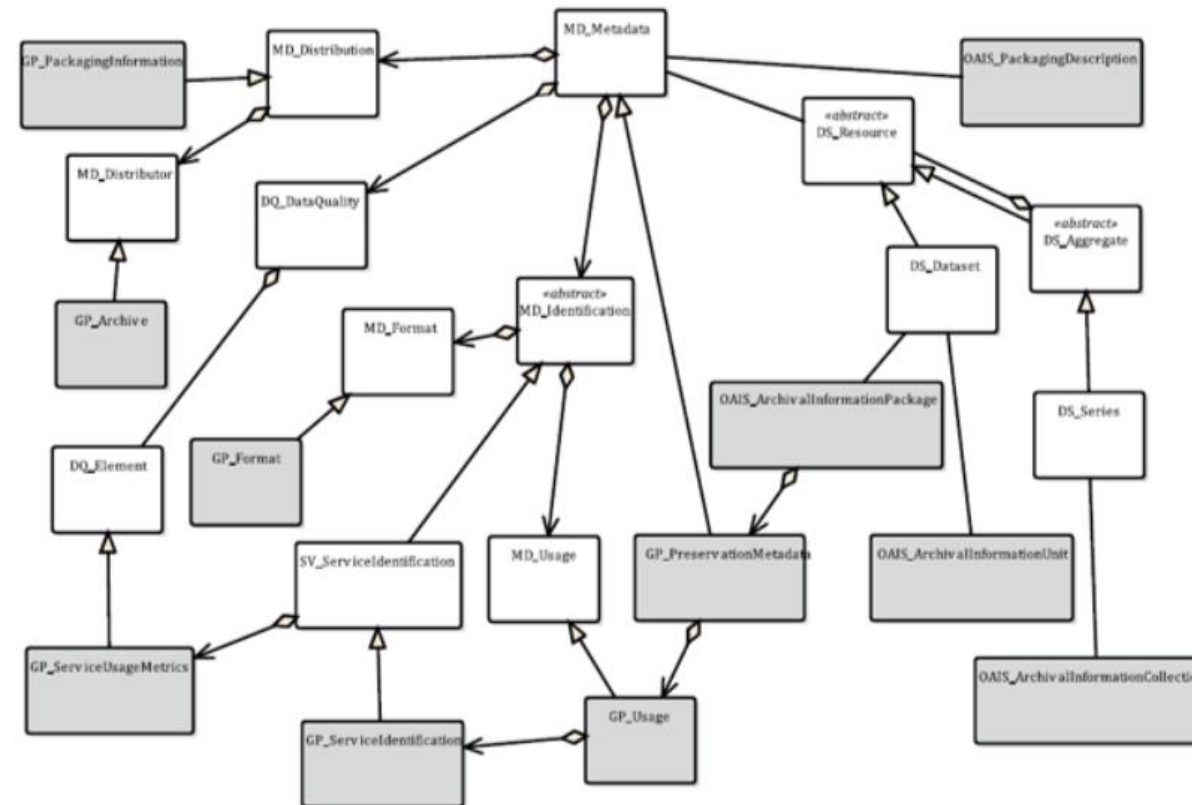
Key elements

- Preservation metadata classes

- Data identifiers

- Data, product and format specifications

- Preservation actions

- Association of resources

- Value and strategy

- Fixity

- Information package description

- Acquisition

Preservation metadata classes

- Class **GP_PreservationMetadata** as a specialization of MD_Metadata

- It involves five subclasses:
  - GP_Usage
  - GP_ValueAndPreservationStrategy
  - GP_Acquisition
  - GP_AssociatedResource
  - GP_PackagingInformation)

- It also considers LI_Lineage as a type of preservation information (aggregation of GP_PresevationMetadata), and MD_Distribution as a specialization of GP_PackagingInformation.


class Fig. 3. Preservation Metadata
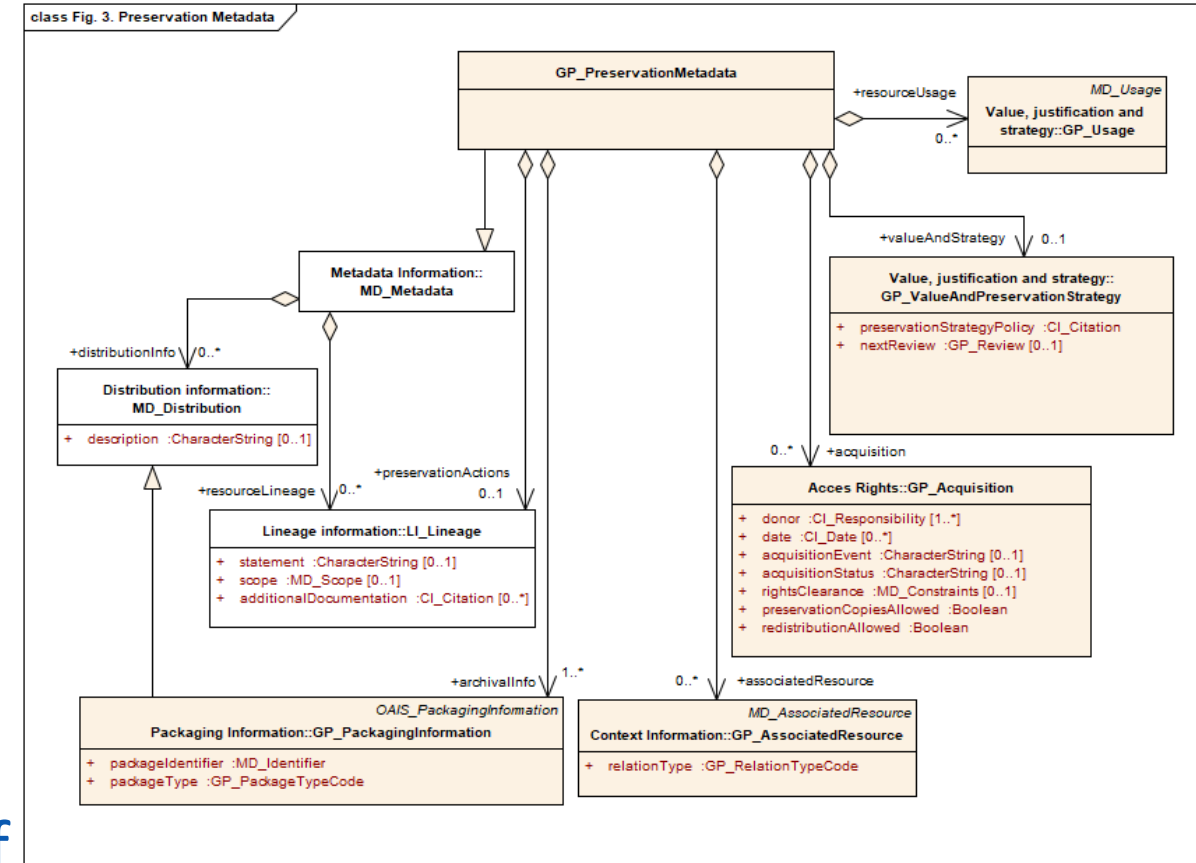
Data Identifiers

- Identify resources uniquely and avoid preserving it more than once is very important.

- The specialization of MD_Metadata, **GP_Identification**, is the used class to identify uniquely any geospatial resource.
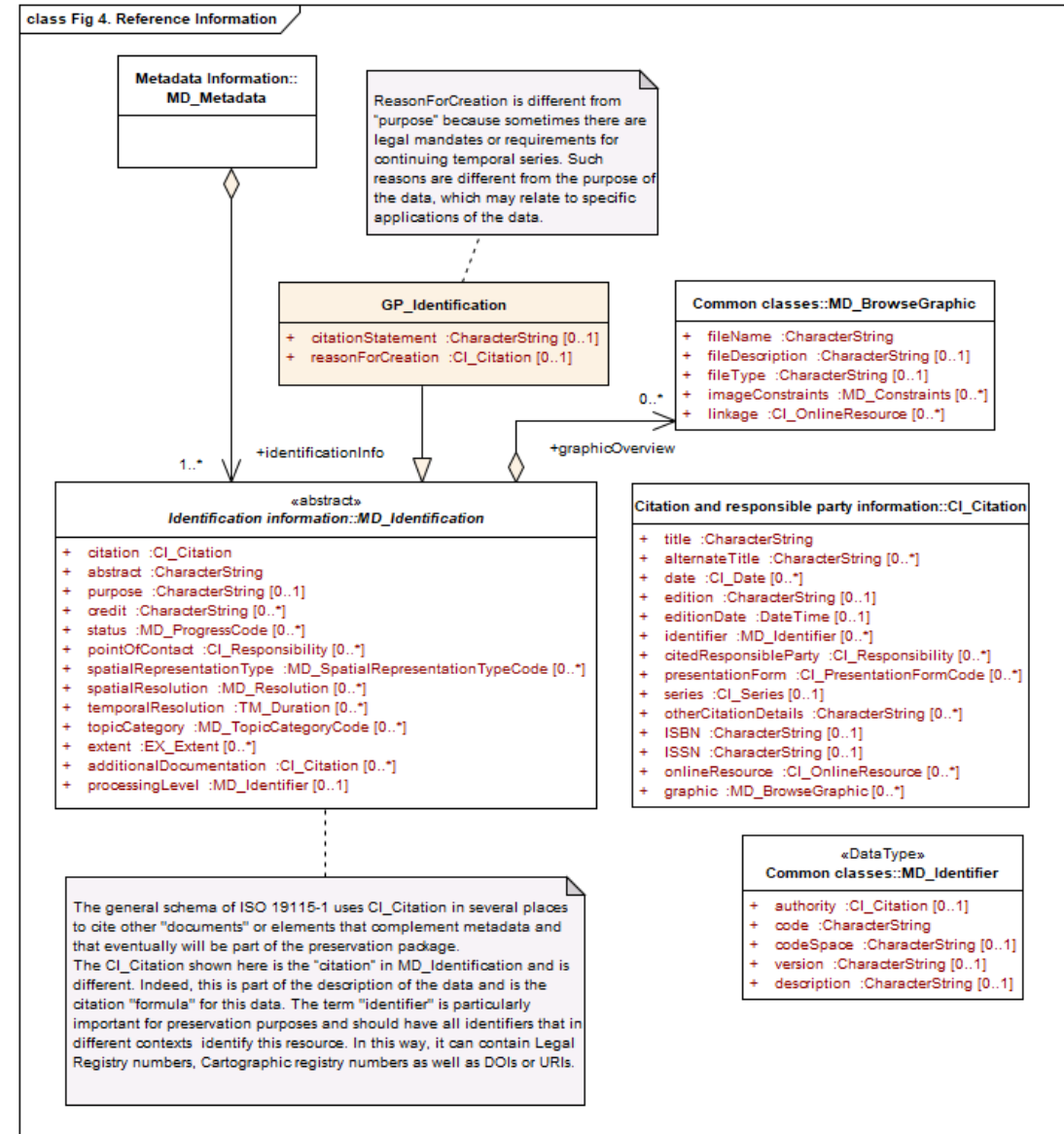
Data, product and format specifications

- Information regarding the product specifications, data dictionary references, common format specifications and the need for a product-specific format specification profile can be linked to ISO 19115-1, concretely the subclasses of MD_Metadata, MD_PortrayalCatalogueReference and MD_AplicationSchemaInformation.

- In addition, **MD_Format** (subclass of MD_Identification) is used to include information about the format of the geospatial resource to preserve.
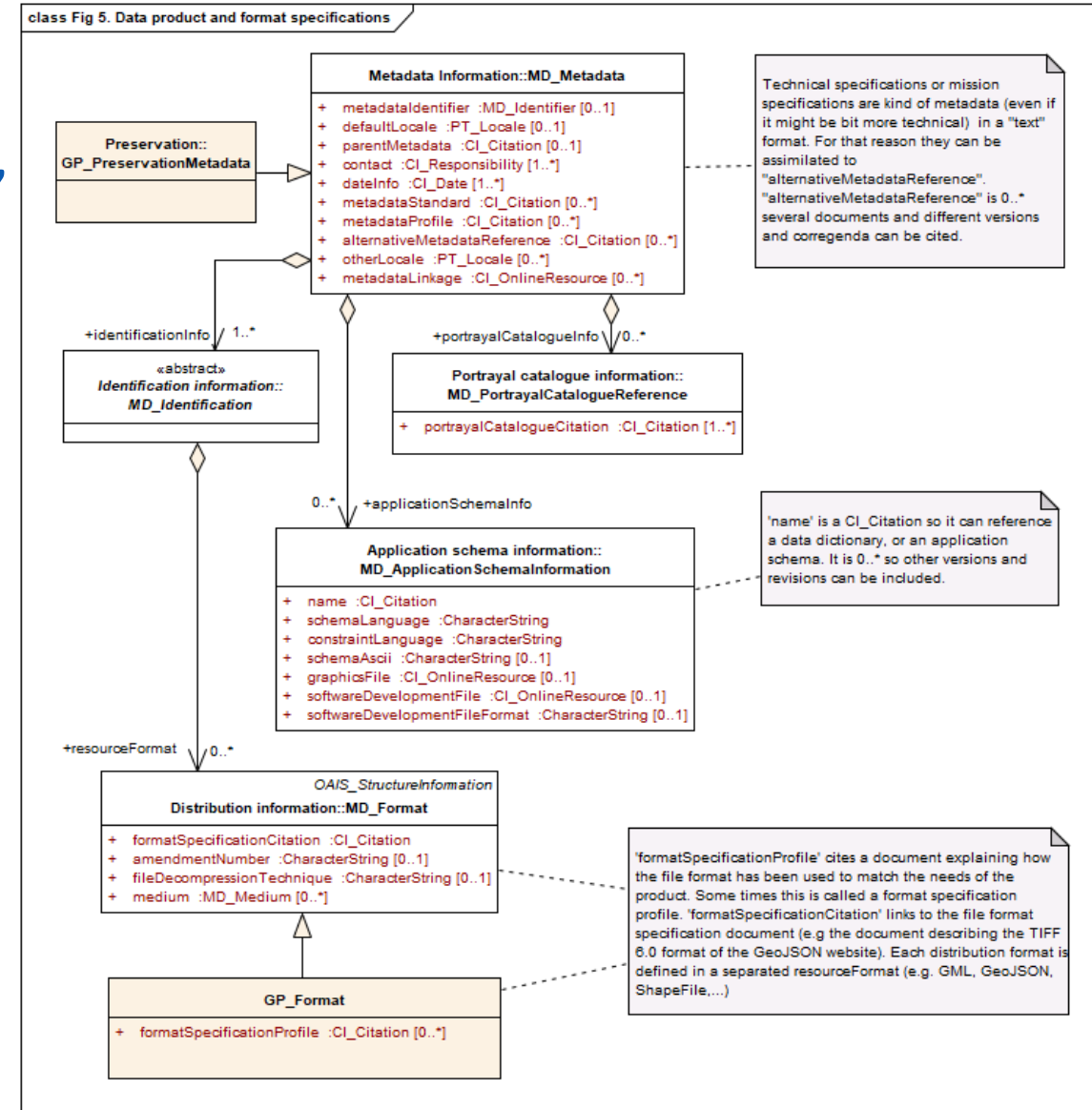
Preservation actions

- Preservation actions are the activities performed as part of the curation process once the creation and maintenance of a resource has ended.

- The main objective is to ensure the preservation of data. All preservation actions done in order to preserve data should be recorded in the LI_Lineage model.

- LI_Lineage acts as a specialization of GP_PreservationMetadata
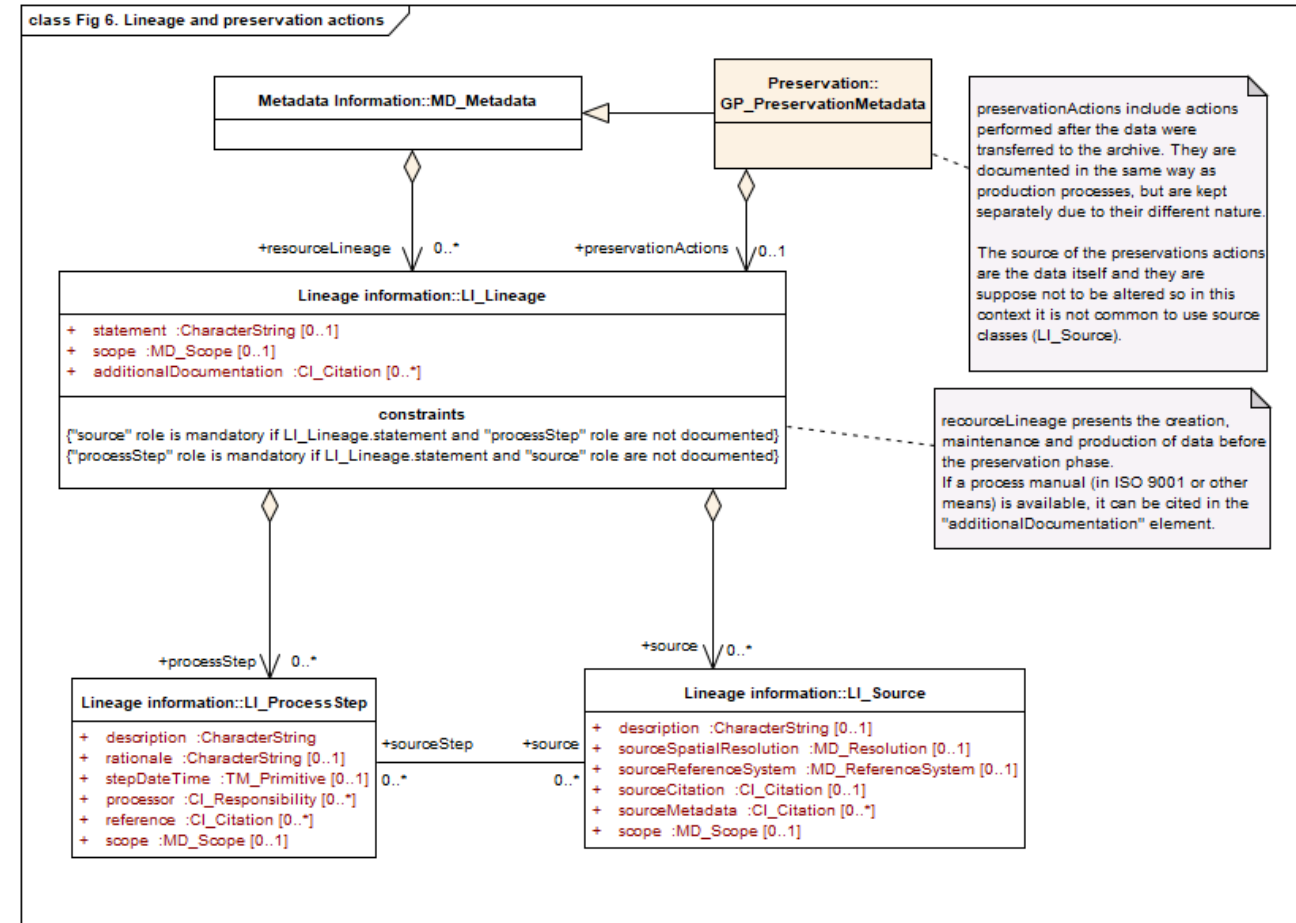
Association of resources

- Providing context for the content information by means of describing associations with other resources is also part of the preservation metadata.

- To this purpose the model includes the **GP_AssociatedResource**, a specialization of GP_PreservationMetadata, which is a subclasse of MD_AssociatedResource.
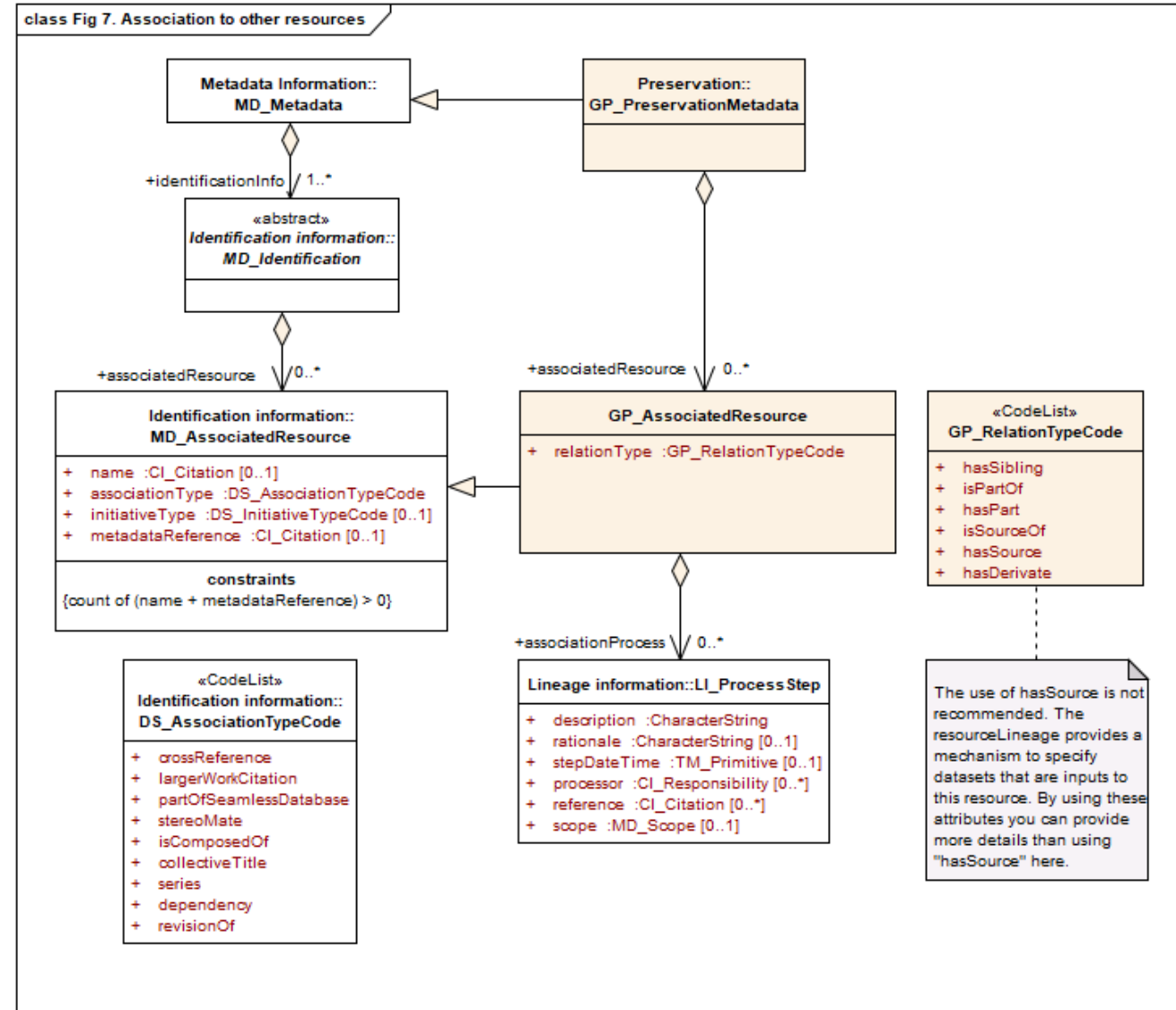
Value and strategy

- 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 **GP_ValueAndPreservationStrategy** includes three different specialization classes: GP_InformationValue, GP_Remove and GP_ReviewPeriod.

Value and strategy
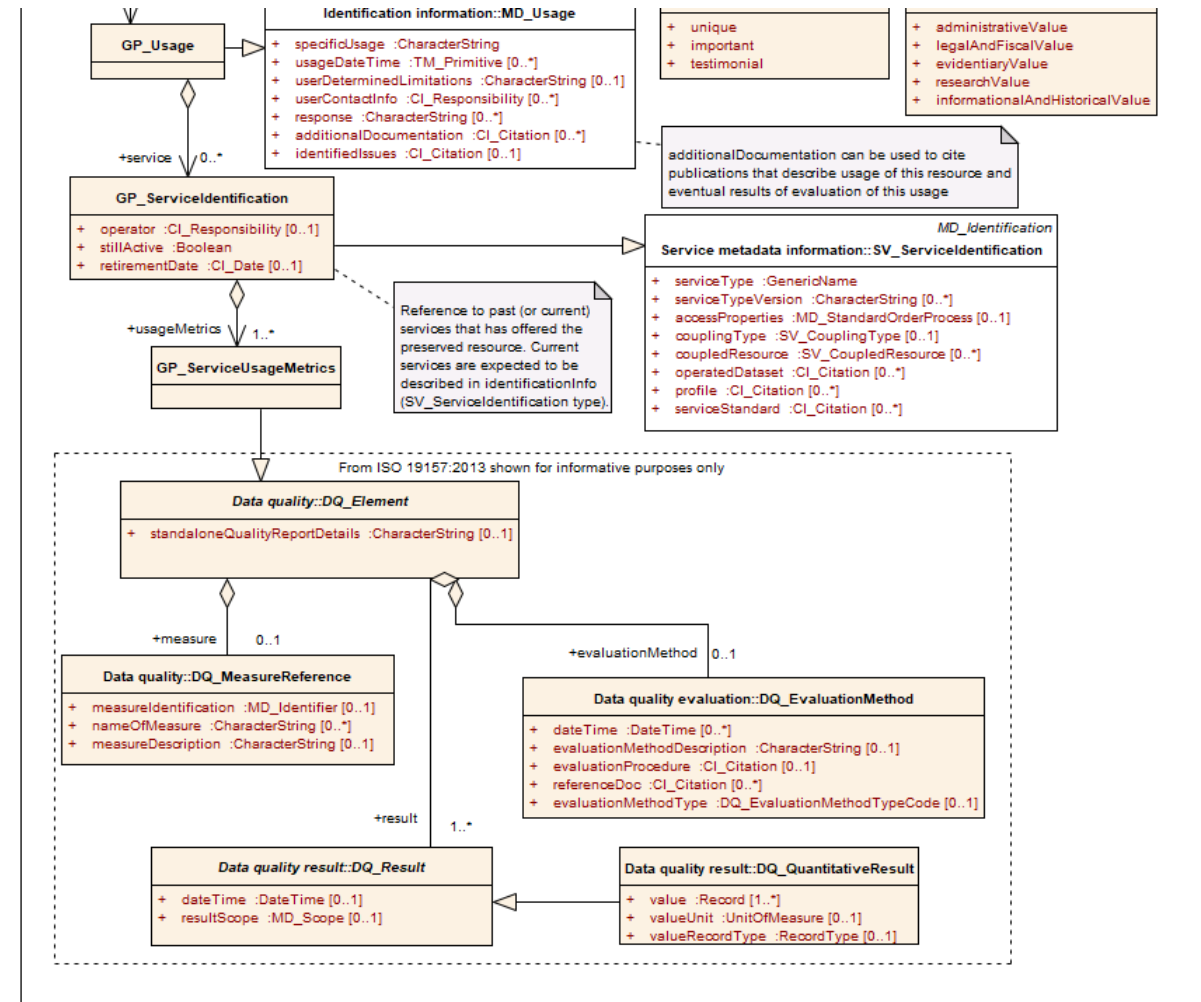
Fixity

- The Fixity information cares about the data integrity checks for the content element in order to guarantee that it has not been altered in an undocumented manner.

- It can be captured at the package level or at the individual part level.

- The integrity methods are described by using the **CI_Citation** methods.

Information package description

- The GP_PreservationMetadata uses the **GP_PackagingInformation** specification, which is based on OAIS_PackagingInformation.
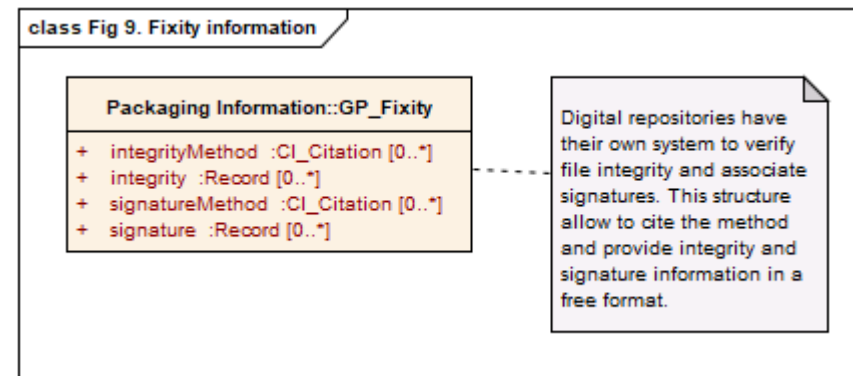
Acquisition

- The Acquisition class documents the transfer process from the production and maintenance to the curation process including licensing and rights transfer.
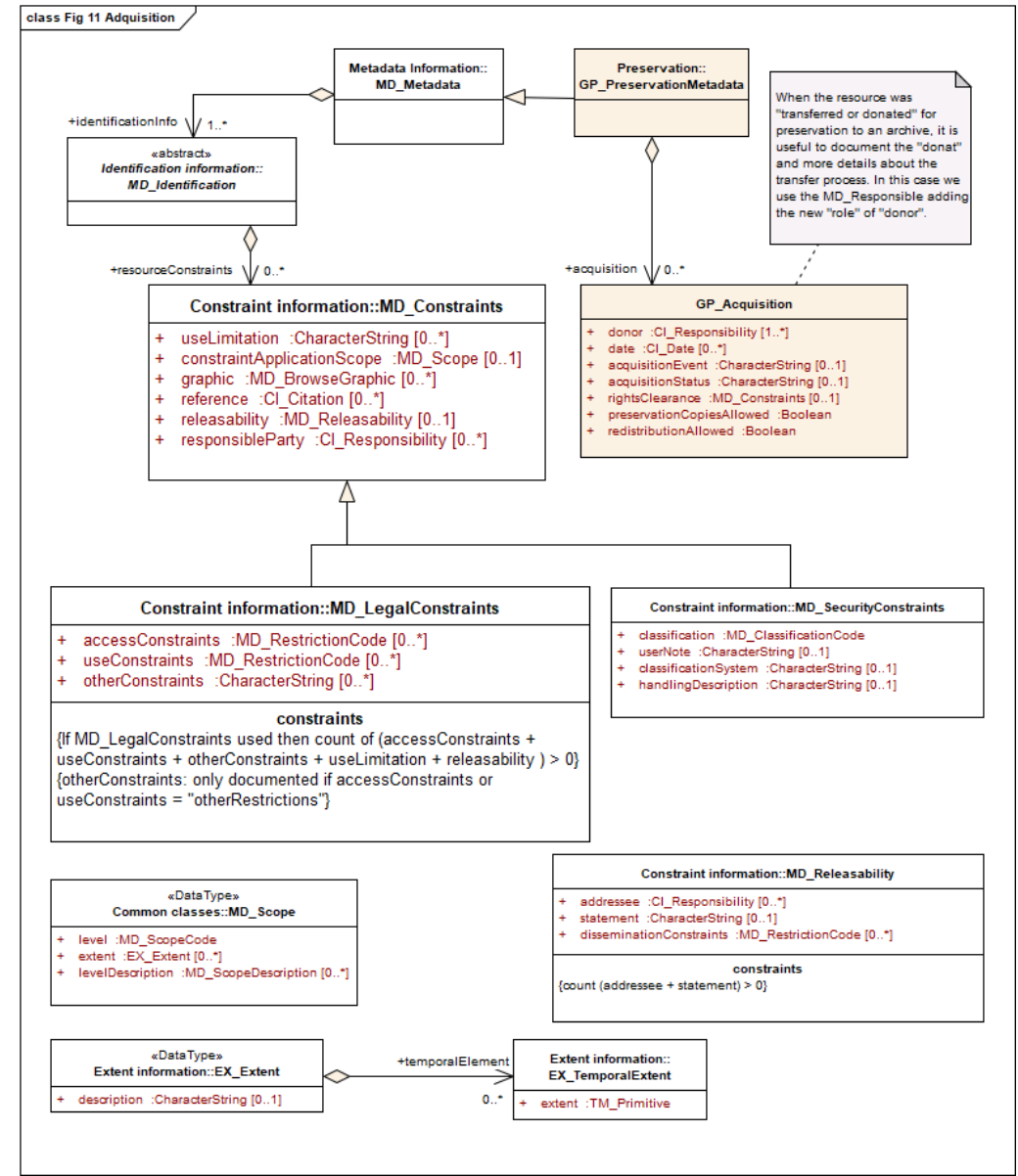
Schemas and examples

- Link to schemas:

  https://standards.iso.org/iso/19165/gpm/1.0/

- Examples of XML encoding:

  https://standards.iso.org/iso/19165/gpm/1.0/examples/

- Link to schemas GitHub:

  https://github.com/ISO-TC211/XML/tree/master/standards.iso.org/iso/19165/gpm/1.0

MMZx preservation package

Geospatial Open Packaging Conventions (OPC)

- 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) along 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 was initially designed as a package for the **MicroSoft office packaged documents** (docx, xlsx, etc) but is now used by other software companies to support their multipart formats (see a list of formats in the wikipedia page) including two GIS brands such as **Autodesk AutoCAD** (for a Design Web Format (DWFx) file format) and **MiraMon** (for its open compressed map (MMZx) file format).

Information technology - Document description and processing languages - Office Open XML File Formats - Part 2: Open Packaging Conventions

MMZ and MMZX

- MMZx is the MiraMon software implementation of the Geospatial Open Packaging Convention (OPC) standard. **MiraMon** is a Geographic Information System (GIS) and Remote Sensing software.

- The **MMZ** is a binary file able to compress different files and links to Internet resources in a single file.

- It was initially intended for the MiraMon maps compression (MMM) containing raster, images, vector, tables, symbolization, etc.

- The separated parts are compressed with a **gzip algorithm** and stored together in a multipart file with a specific header format including metadata about the original files (Read the MMZ format specification).

MMZX format

- The new **MMZX** format keeps the idea and the spirit of the original MMZ, but it improves the former:
  - It takes a format based on an ISO specification (specifically the ISO 19165:2018 of Preservation of Geographic Information) to convert it not only into an open format, but also to an standardized one
  - It includes a number of additional properties, such as the support to different entry points of the "maps" (map project) for the different formats (proprietary or free) it may contain, or the inclusion of thumbnails of the map itself.

- It takes a renewed strength for the added repesenting contribution to the preservation of data following the needs marked by the preservation standard **Open Archival Information System** (OAIS).

- It can be used, with its great simplicity of creation and use, to save "time frozen versions" of layers or maps and also with the sophistication of supporting different data models, metadata, symbolization, etc.