After populating INSPIRE with datasets, how to ensure geospatial digital preservation?

Joan Masó, Ester Prat, Carme Montaner, and Josep Lluis Colomer

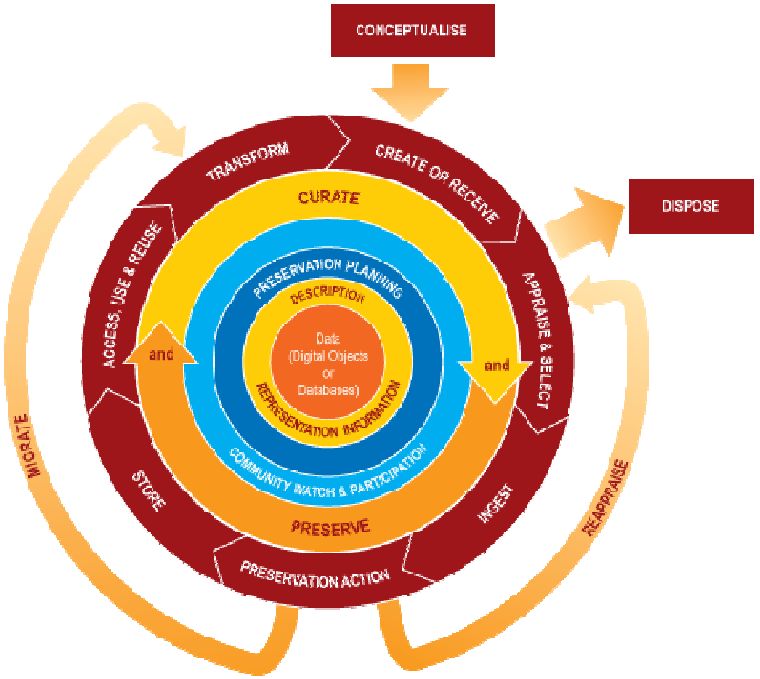
Introduction

- After a period, fighting for having good digital data, he have started to detect an increase of interest in data preservation.
- The reason: Datasets will be **reviewed and maintained** in a **continuous** way and there is a risk of **losing the temporal series** that could be crucial to analyze the continuous **human impact** on the planet and to **monitor the effects** of the European Union **policies**.
- Digital data preservation is more challenging than paper preservation and has been previously considered in several domains but has been not widely discussed for the geospatial data and there is no standardized model that has global acceptance.

Introduction

- The Institut Cartografic i Geològic de Catalunya (ICGC) want to promote a protocol for preserving digital data that could be integrated in their production processes.

- *First step*: Establish a **digital preservation metadata model** that allows documenting the relevant and necessary features to **ensure the durability of the information**.



- Based on:
  - ISO 19115 Geospatial metadata
  - ISO 14721:2012 Reference model for an Open Archival Information System (OAIS)
  - others...

Step 1: Selection of Standards

MiraMon GIS & RS 20years

GRUMETS METHODS & APPLICATIONS IN REMOTE SENSING & GIS

UAB Universitat Autònoma de Barcelona

CREAF

ICGC Institut Cartogràfic i Geològic de Catalunya

Standards selection (1/2)

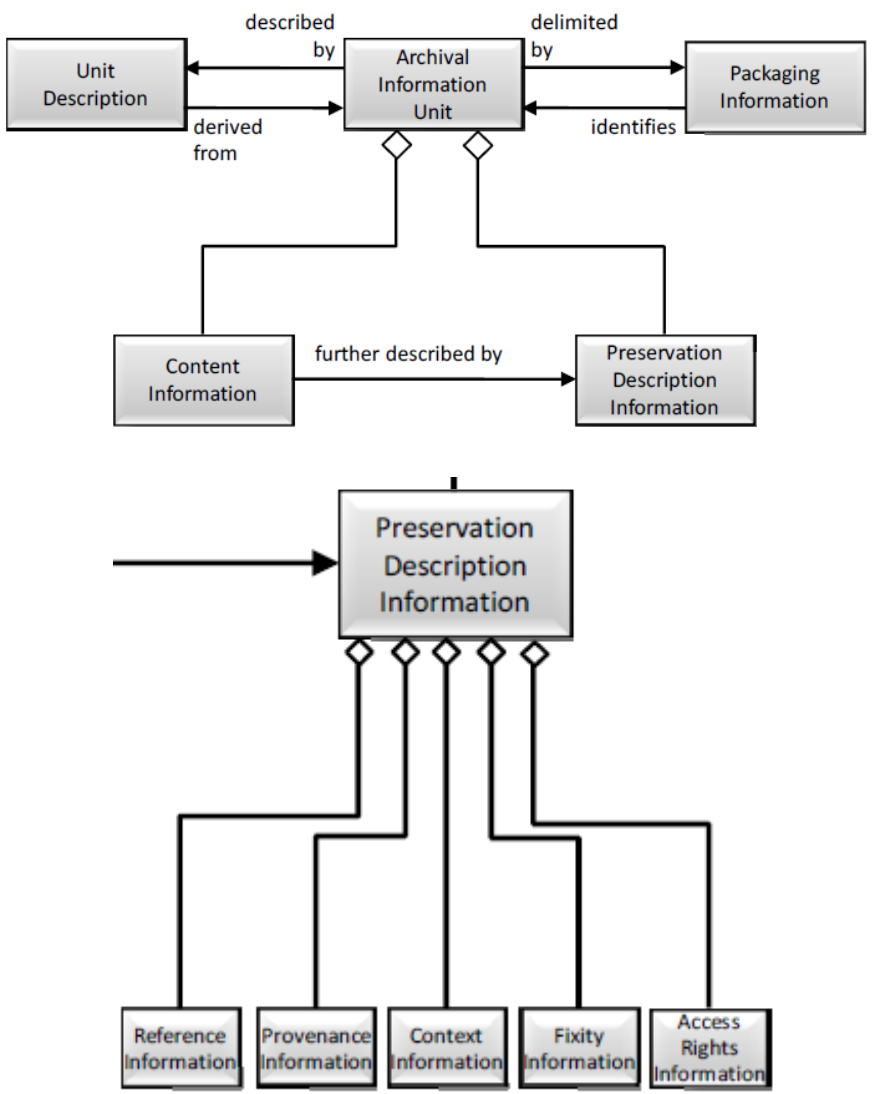
- Standards (and other initiatives) considered:
  - Preservation model Open Archival Information System (OAIS)
  - PREMIS Data Dictionary for Preservation Metadata
  - Archival Metadata Elements for the Preservation of Geospatial Datasets (GeoMAPP)
  - Methodology for the preparation of Documentary Evaluation and Access Proposals (PAAD)
    - Catalan National Commission on Access, Evaluation and Selection of Documents (CNAATD).
  - GI+100: Long term preservation of digital Geographic Information - 16 fundamental principles agreed by National Mapping Agencies and Archives.
- Precedent
  - Long-term Preservation for Spatial Data Infrastructures: a Metadata Framework and Geo-portal Implementation (D-lib Magazine).

Standards selection (2/2)

- Other initiatives checked:
  - The documentation of a new ISO / TC 211 initiative on preservation: ISO 19165 *Geographic information – Preservation of digital data and metadata*
    - Based on applying the OAIS and that recognizes the need for a package of geospatial preservation
  - Technology Watch Report: Preserving Geospatial Data
  - Digital preservation repository of the Library of Catalonia
  - Data Model for Managing and Preserving Geospatial Electronic Records by Center for International Earth Science Information Network (CIESIN)
  - Long Term Preservation of Earth Observation Space Data. European LTDP Common Guidelines
  - … papers in the Journal Of Map Geography Libraries….

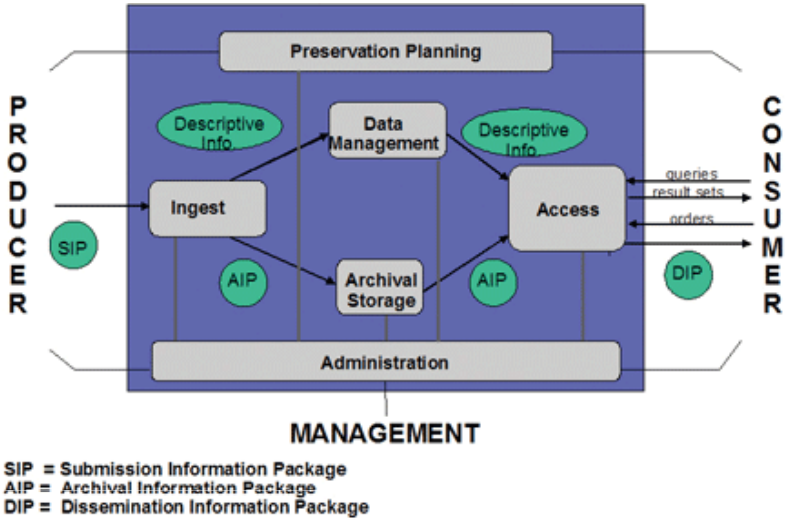
Open Archival Information System (OAIS)

- It is the **reference** standard in the preservation of digital information.
- It is prepared by the Consultative Committee for Space Data Systems (CCSD) and the International Organization for Standardization (ISO).
- The reference model (that includes the practices) dates from June 2012.



Open Archival Information System (OAIS)

- It is a conceptual model
  - Neither indicates how to implement it nor which attributes must be considered

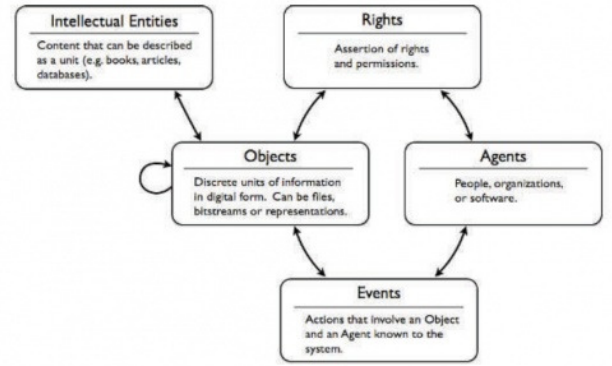


- We have taken from it:
  - The idea of the **preservation package**
  - **The classification** (the large UML containers) of the preservation metadata
  - The structure in “collections” and “units” useful for **geospatial “series”**
  - We ensured that our model is conformed to the reference framework

PREMIS Data Dictionary

PRESERVATION METADATA MAINTENANCE ACTIVITY

- Based on the OAIS, PREMIS develops a proposed core set of preservation metadata applicable to a wide range of digital preservation contexts (not just geospatial) and with the support of guidelines and recommendations for the creation, management and use. It is produced by the PREMIS Working Group.
- The model defines 5 distinct entities:
  - Intellectual Entities, Objects, Events, Rights and Agents
  - The ability to relate them



- We have taken from it the the capacity of **relating resources and properties**.

GeoMAPP

- Geospatial Multistate Archive and Preservation Partnership. Archival Metadata Elements for the Preservation of Geospatial Datasets (21-09-2011)
- Implements a metadata model according to the OAIS model for geospatial information from the point of view of American libraries
- Reaches the level of detail of element and attribute
- It is based on the FGDC metadata standards (does not incorporate the ISO 19115 model)
- We have taken many details about the **preservation strategy, the value and the justification for preservation**

Documentary Evaluation and Access Proposals (PAAD)

Generalitat de Catalunya
gencat.cat

- It is a methodology proposed by the Catalan National Commission on Access, Evaluation and Selection of Documents (CNAATD) to determine which data is worth to preserve and the general scheme of access the items it receives. It includes a form to fill (metadata) and sent it back to the commission for evaluation
- It does not consider geospatial items diferently
- We have taken the **statement and justification of the value** of the data to determine the expiration of the need to preserve (completing GeoMAPP).

GI+100: Long term preservation of digital Geographic Information

- Presents 16 principles agreed by the governmental mapping agencies and archives for the preservation of digital geographic information.
- Some are very general and others are more applied
- They are focused on geographic information.
- Our model respects these principles and includes details such as the need to document the **symbolization** and to include a **quick look**.

GI+100: Long term preservation of digital Geographic Information — 16 fundamental principles agreed by National Mapping Agencies and State Archives

by Carsten Rönsdorf, Paul Mason and Jonathan Holmes, Ordnance Survey; Urs Gerber and André Streilein, swisstopo; Marguérite Bos, Schweizerisches Bundesarchiv; Arif Shaon, Rutherford Appleton Laboratory; Kai Naumann, Landesarchiv Baden-Württemberg; Michael Kirstein, Generaldirektion der Staatlichen Archive Bayerns; Göran Samuelsson, Mid Sweden University; Marja Rantala, Maanmittauslaitos; Sidsel Kvarteig, Statens kartverk; Lynne Ralsberg and Jenny Svennewall, Lantmäteriet and Wolfgang Stößel, Landesamt für Vermessung und Geoinformation Bayern.

Step 2: Model design

MiraMon GIS & RS 20years

GRUMETS METHODS & APPLICATIONS IN REMOTE SENSING & GIS

UAB Universitat Autònoma de Barcelona

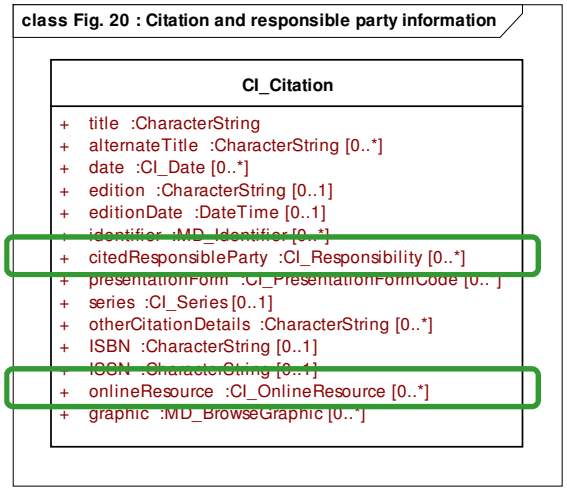
CREAF

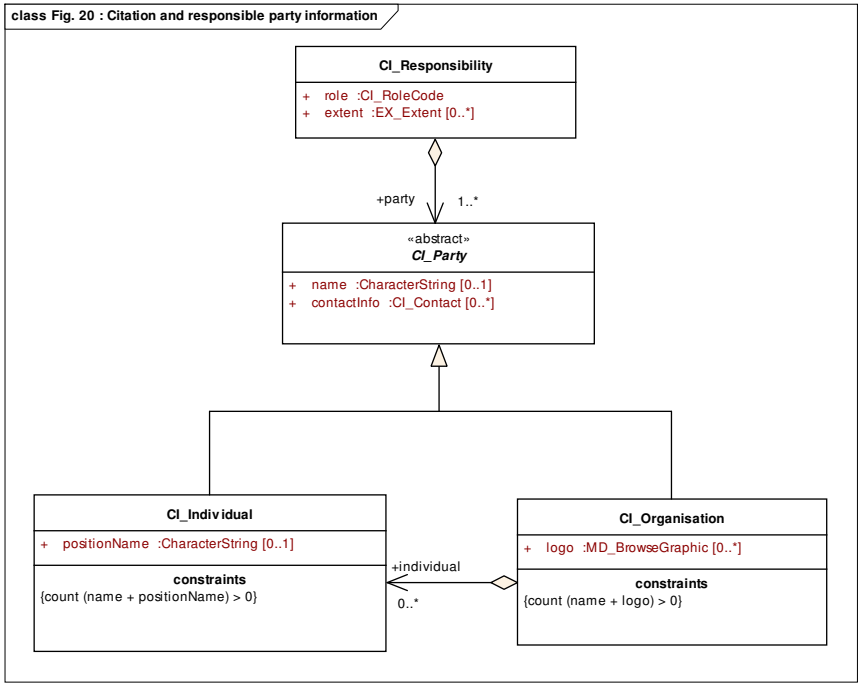
ICGC Institut Cartogràfic i Geològic de Catalunya

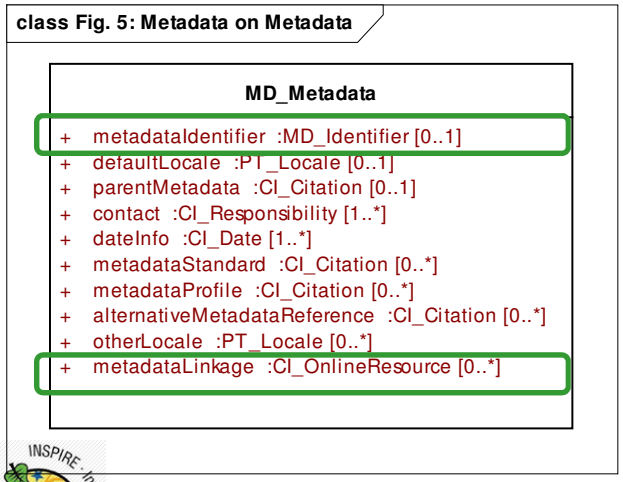
The foundations

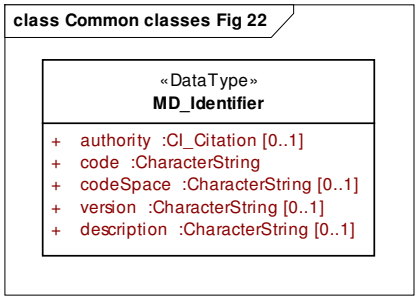
- We have chosen the ISO 19115 metadata model as a starting point.
- In order to communicate better our model, we used UML model of preservation metadata.
  - It is based on the distribution of the ISO/TC 211 UML models maintained by CSIRO (Simon Cox) in the web.
- The UML model has been extended (in the same file) with the preservation concepts, thus maintaining a link with the ISO19115-1 models.
- Reasons for using the ISO 19115-1 (instead of 19115 base line)
  - Better treatment of identifiers and citations
  - It has downward compatibility (it is not much different from the previous one)

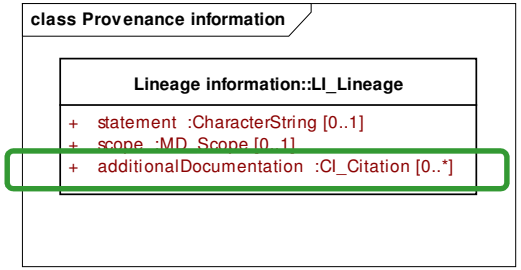
More citations in ISO19115-1



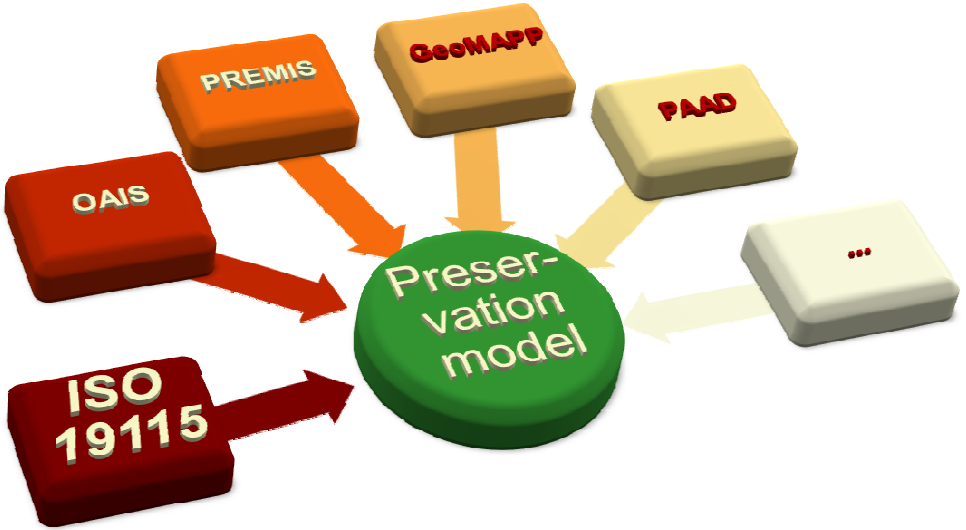






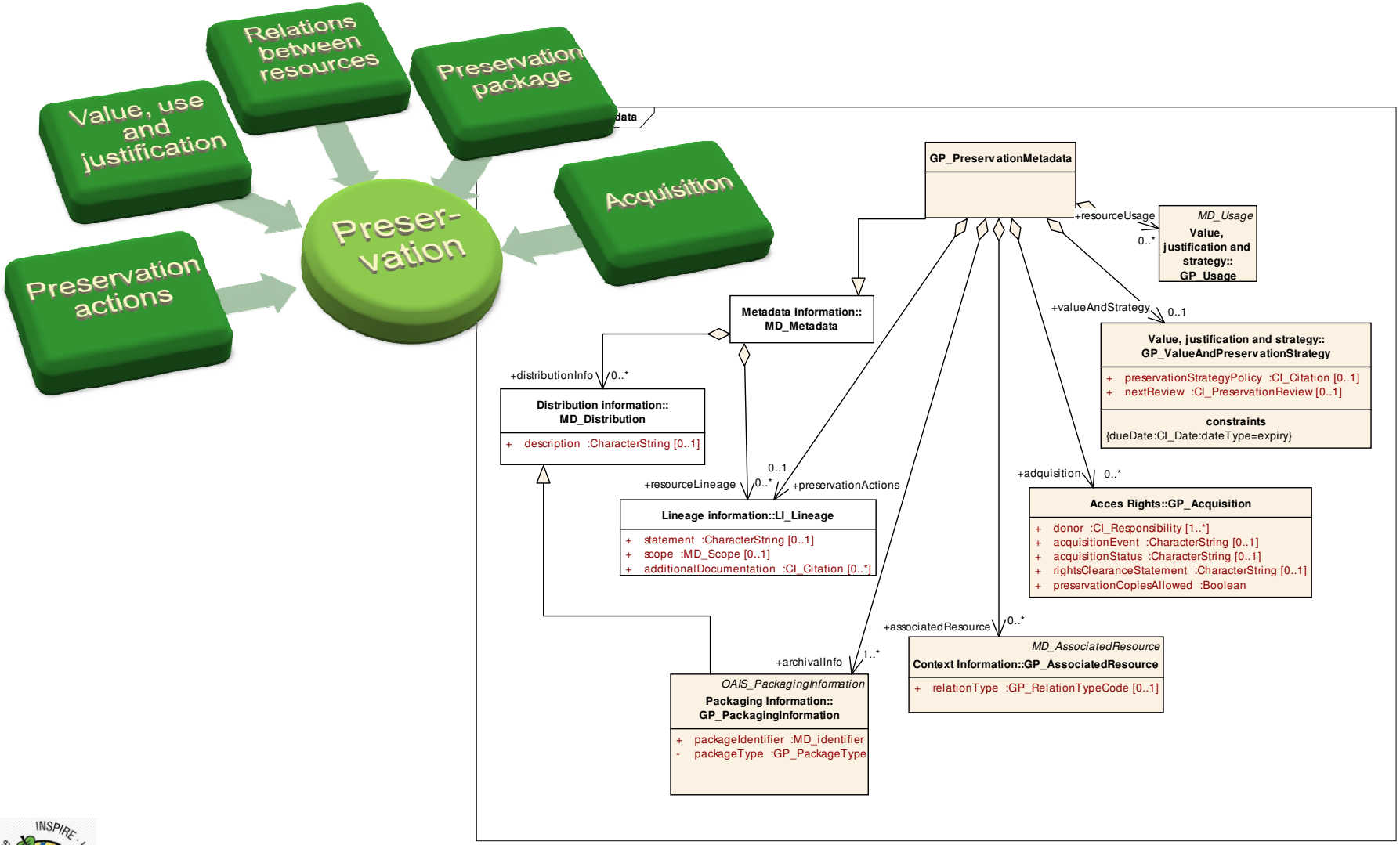


Model design rule: Do not reinvent



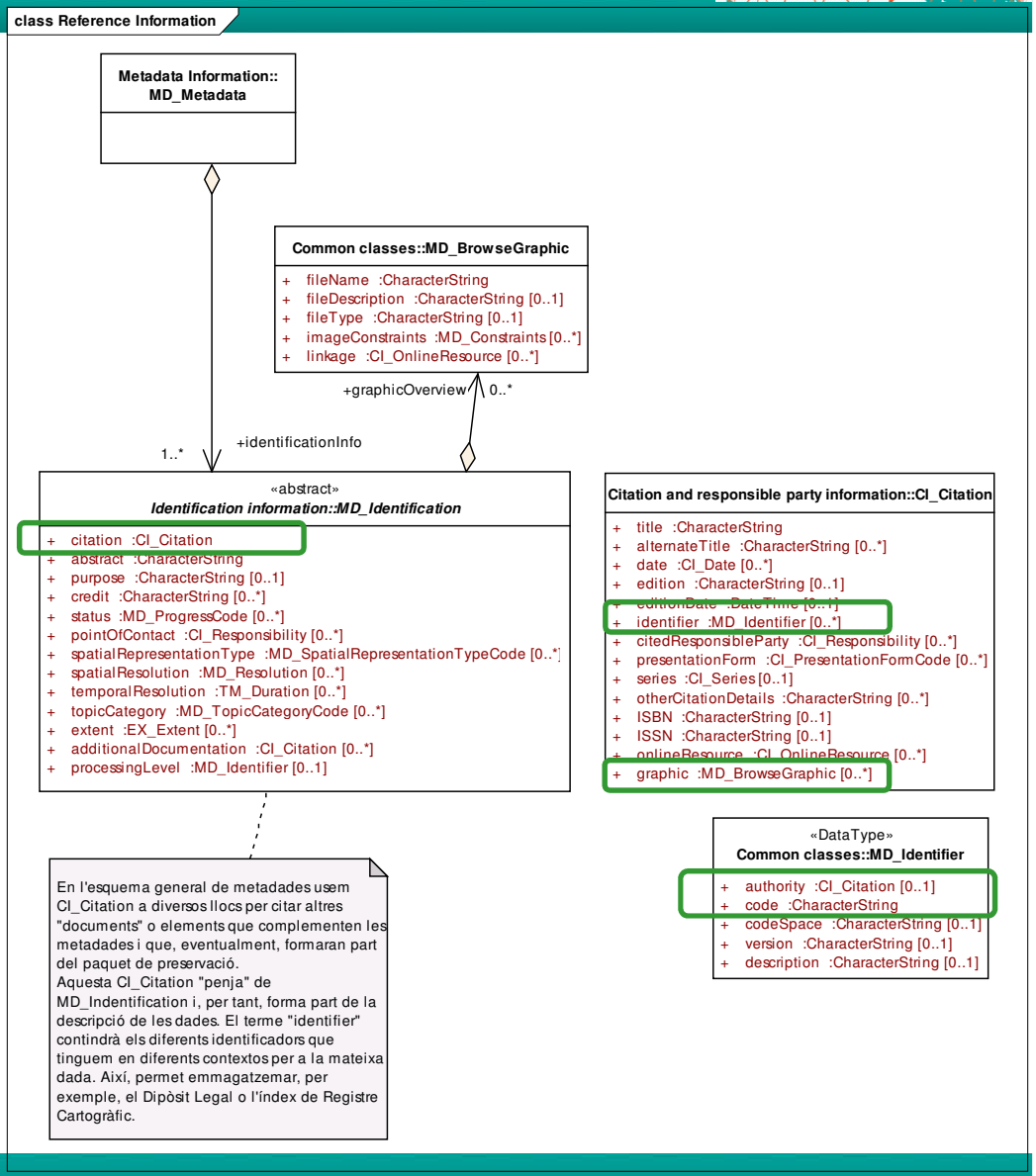
- If it exists previously in the ISO19115-1 it is adopted
  - If possible, we just give recommendations
  - Recycling of classes, e.g. DistributionInfo is used to describe the ArchivalInfo.
  - Drawback: The preservation information is "scattered" into the ISO 19115 model
    - The information is re-centralized in the preservation package part
- Only if it was not in the ISO 19115-1 a new concept and coding is proposed

Additions & model



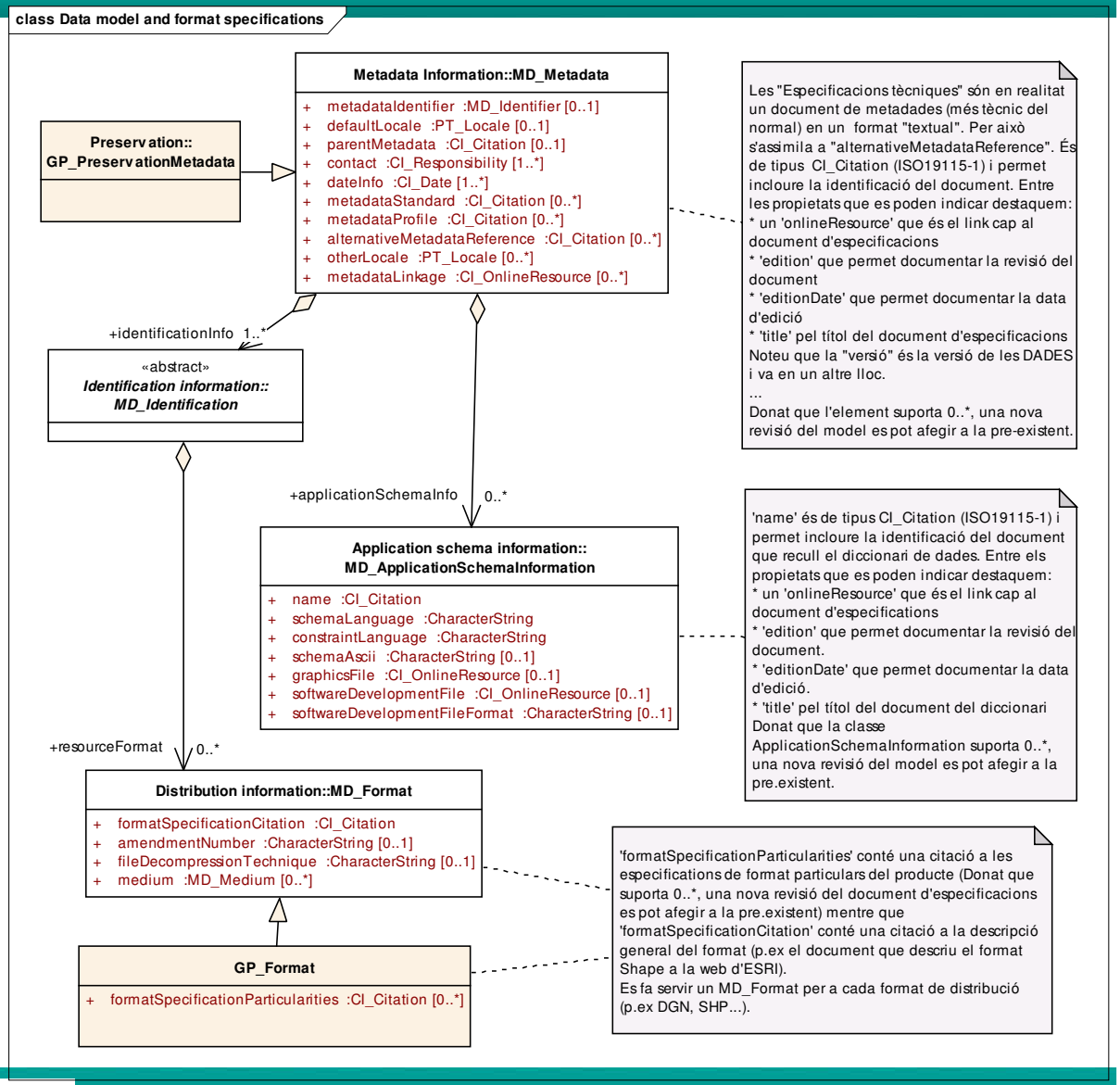
Identifiers

- Several identifiers must coexist, such as:
  - Storage ID
  - Map Library ID
  - Production ID
- According to the 16 principles, it is also important to have a QuickLook

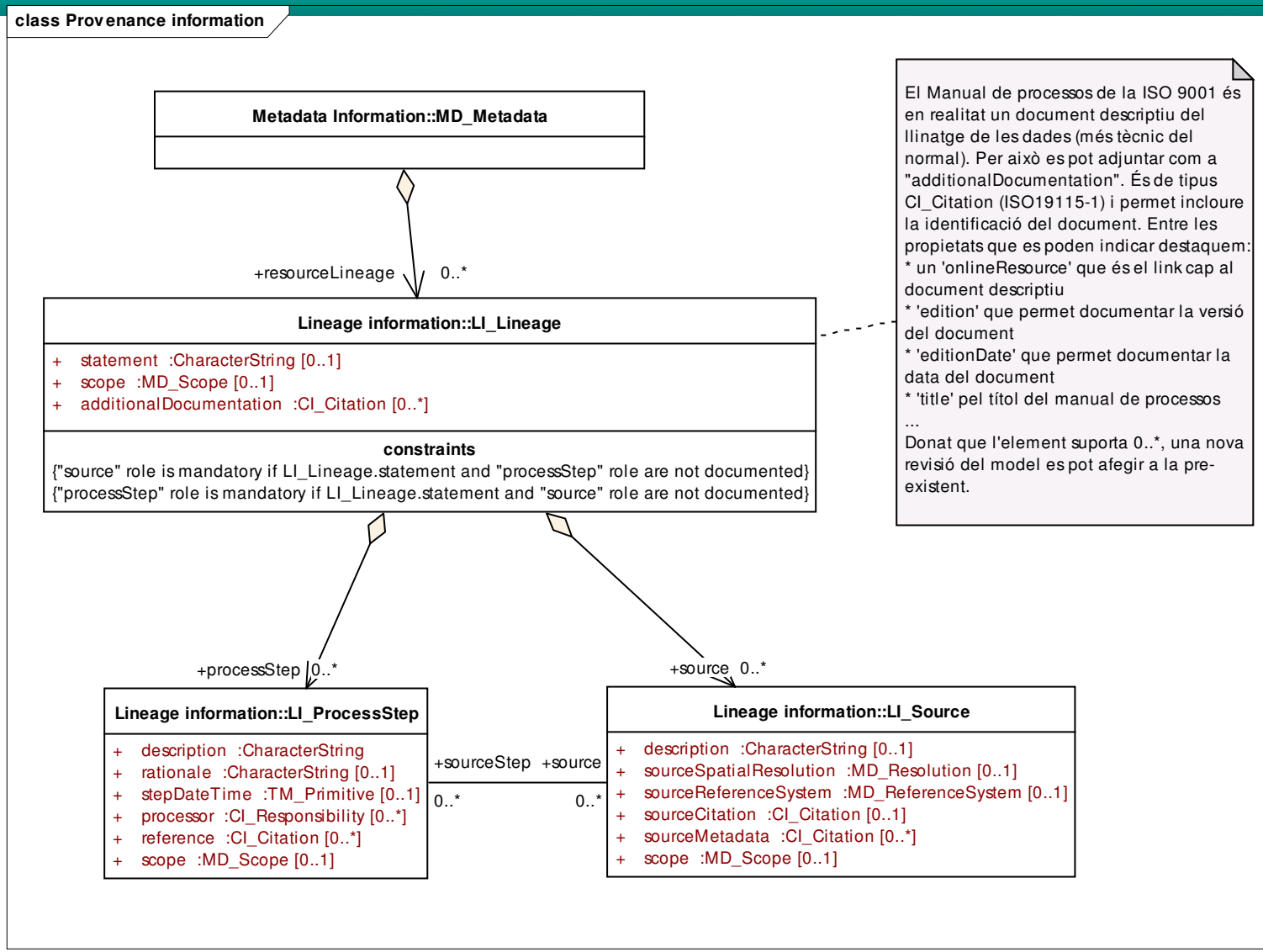


Technical documentation (1/2)

- The NMA (and ICGC) produces a number of documents that can be linked and preserved (and eventually packaged)
  - Technical specifications
  - Data models
  - Distribution formats
  - …

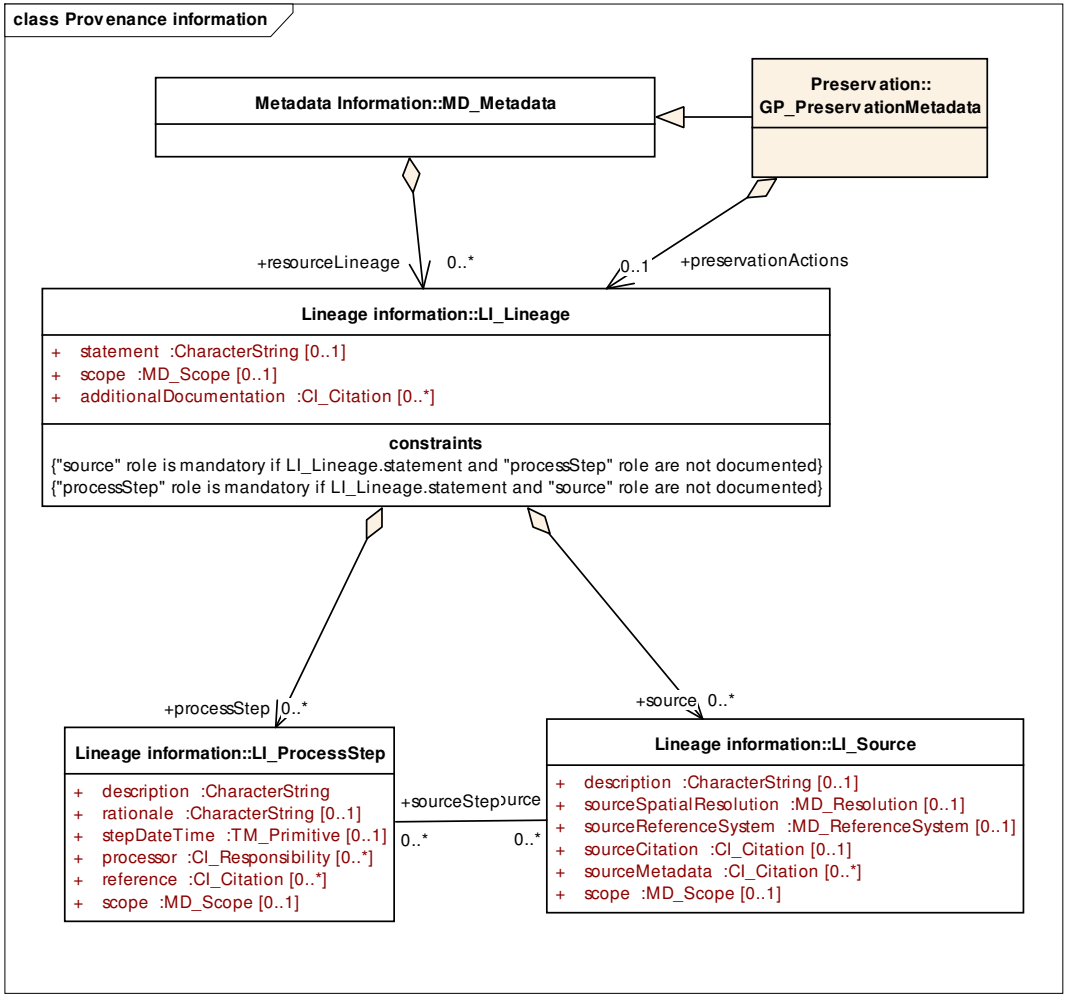


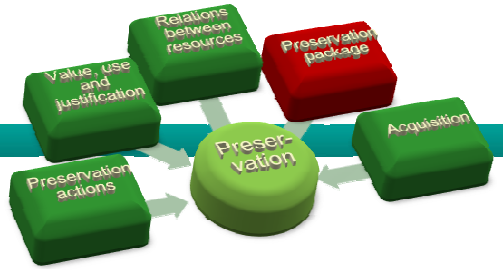
Technical documentation (2/2)



Preservation actions

- Preservation actions are incorporated after cataloging and therefore they are separated from the production lineage
  - they have the same structure





Package Information

- Based on the OAIS, archival and distribution packages are incorporated
- Based on the GeoMAPP, the derived formats are incorporated

class Archival Packaging Information

Preservation:: GP_PreservationMetadata

Metadata Information:: MD_Metadata

+archivalInfo 1..*

OAIS_PackagingInformation
GP_PackagingInformation
+ packageIdentifier :MD_identifier
- packageType :GP_PackageType

La data d'expiració del paquet es troba a GP_remove dins de GP_ValueAndPreservationStrategy

Distribution information::MD_Distributor
+ distributorContact :CI_Responsibility

+distributor 0..*

+distributionInfo 0..*

Distribution information:: MD_Distribution
+ description :CharacterString [0..1]

GP_Archive
constraints
{{role: CI_RoleCode=custodian}}

«CodeList»
GP_PackageType
+ master
+ derivate
+ masterCopy
+ derivateCopy

Aquesta aproximació permet especificar dos paquets: el paquet de distrisbució (que penja directament de MD_Metadata a través de distributionInfo) i el paquet d'arxiu (que penja de GP_PreservationMetadata a través de archivalInfo).

+distributionFormat 0..*

Distribution information::MD_Format
+ formatSpecificationCitation :CI_Citation
+ amendmentNumber :CharacterString [0..1]
+ fileDecompressionTechnique :CharacterString [0..1]
+ medium :MD_Medium [0..*]

+packageElement 0..*

+transferOptions 0..*

GP_PackageElement
+ name :CharacterString
+ format :MD_Format [0..1]
+ size :Real [0..1]
+ type :GP_PackageElementTypeCode [0..1]
+ disseminationConstraints :MD_RestrictionCode [0..*]
+ onLine :CI_OnlineResource [0..1]
+ offLine :MD_Medium [0..1]

constraints
{count(online+offline)<2}

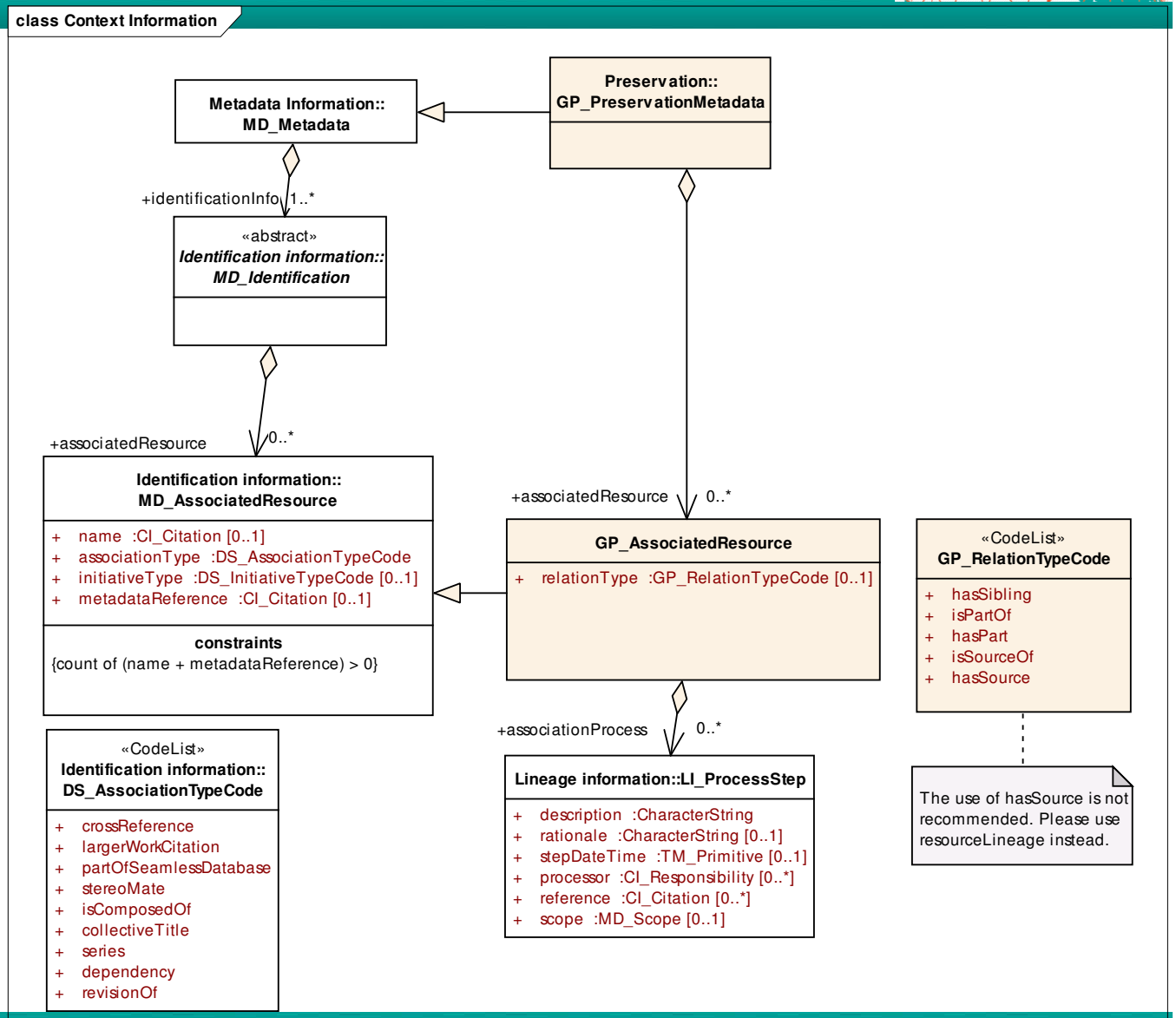
notes
Use onLine and offLine when this packageElement is stored in another package of a collection.

Distribution information:: MD_DigitalTransferOptions
+ unitsOfDistribution :CharacterString [0..1]
+ transferSize :Real [0..1]
+ onLine :CI_OnlineResource [0..*]
+ offLine :MD_Medium [0..*]
+ transferFrequency :TM_PeriodDuration [0..1]
+ distributionFormat :MD_Format [0..*]

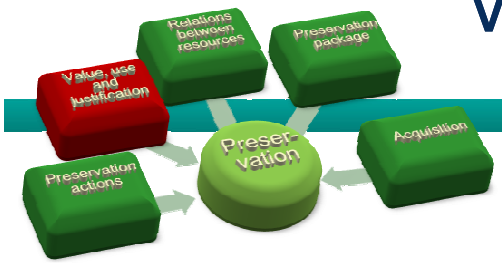
És possible tenir un packet de preservació sense elements interns. P.ex. un packet de types "derivate" en format PDF

Context information

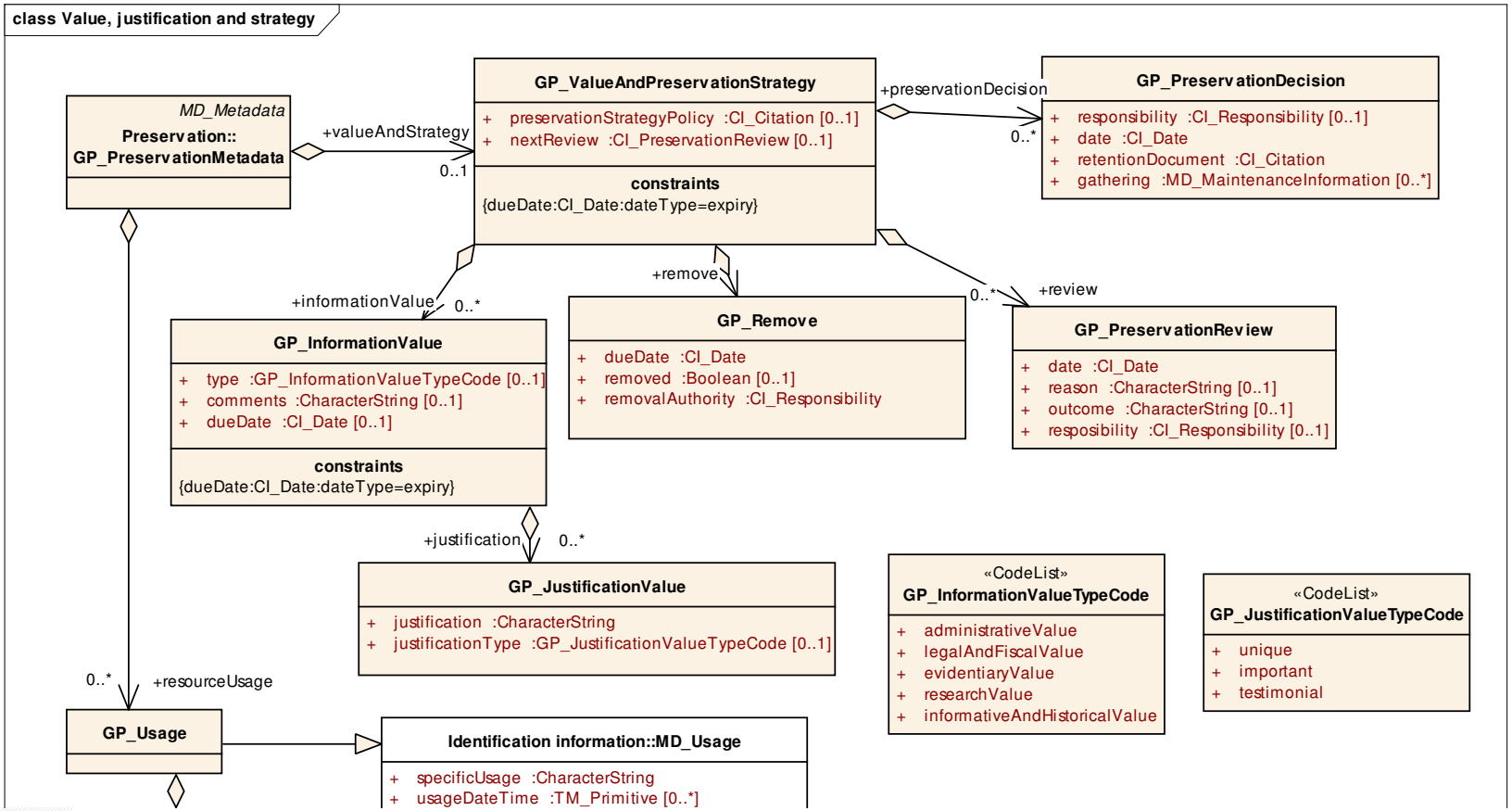
- Relates the resource with other resources
- The OAIS presents this concept and PREMIS describes it in detail



Value, justification and strategy

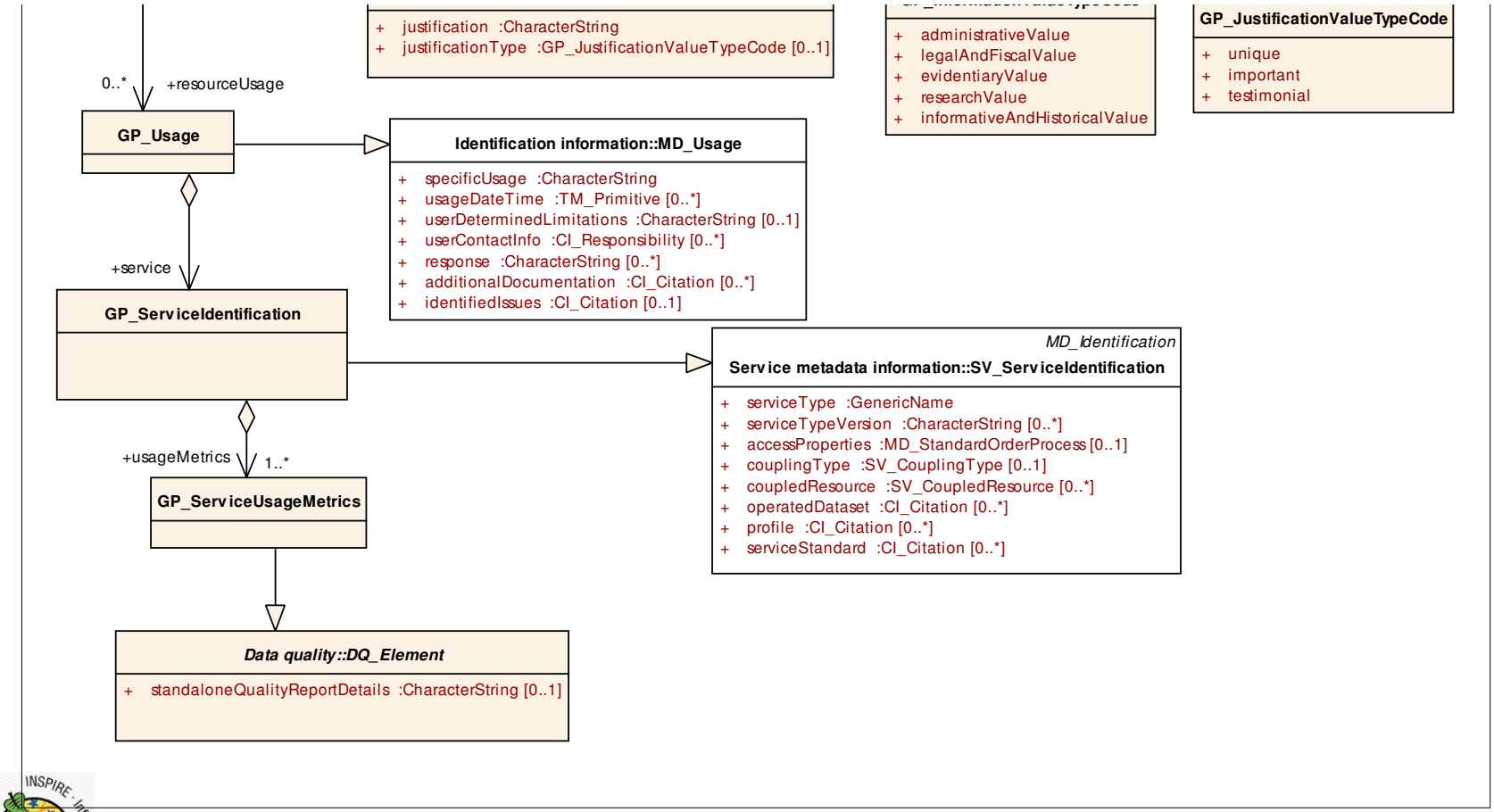


- We register the decision, the revision and the removal



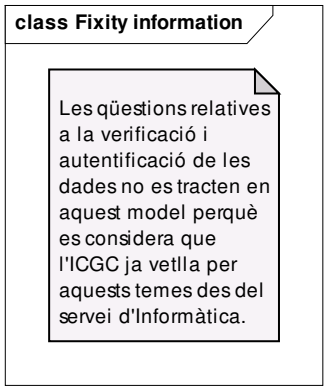
Value, justification and strategy

- We include the intensity of use of the associated Geoservices



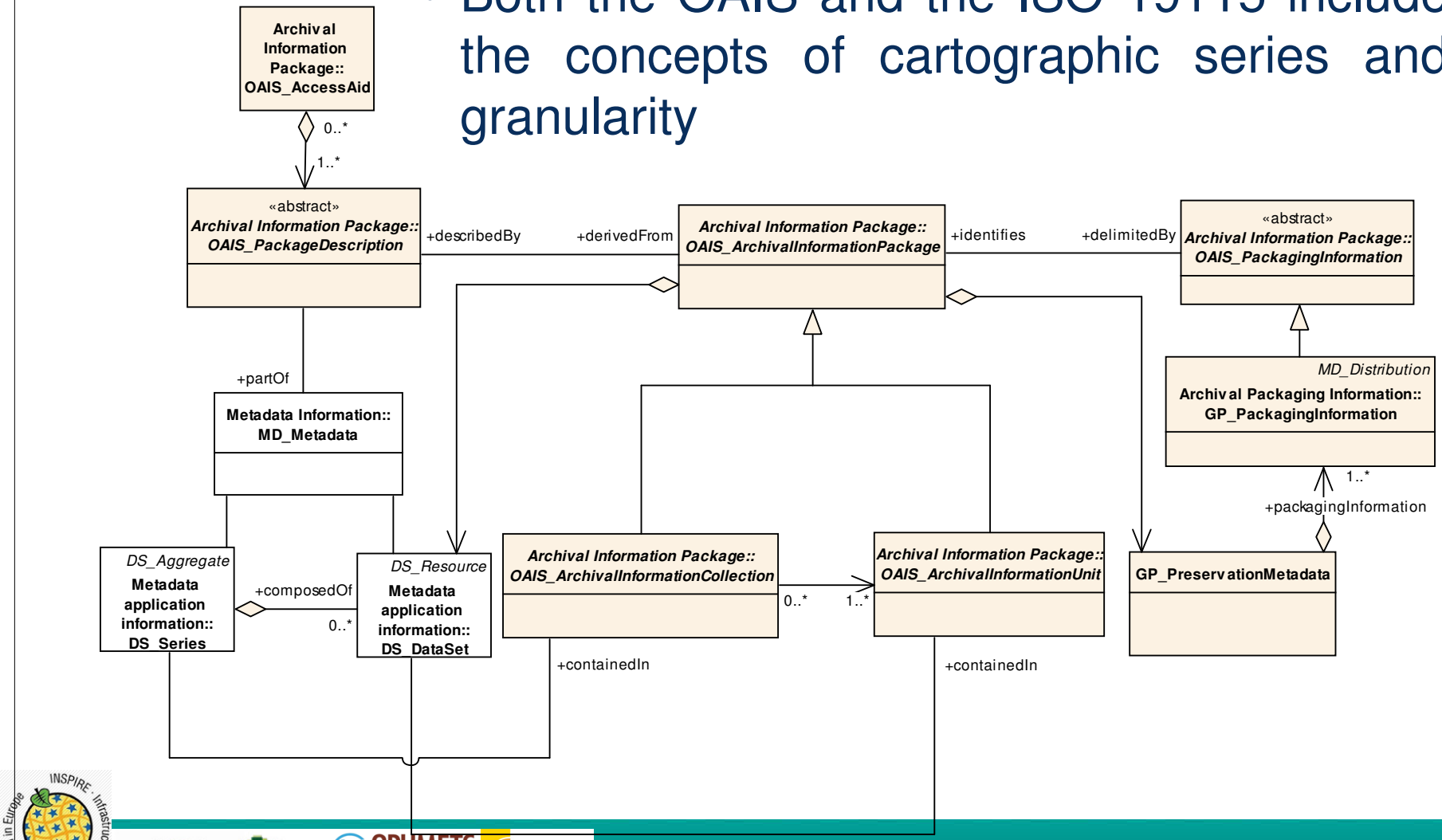
Redundancy and consistency

- Our work did not include these aspects to consider them covered by the IT department.



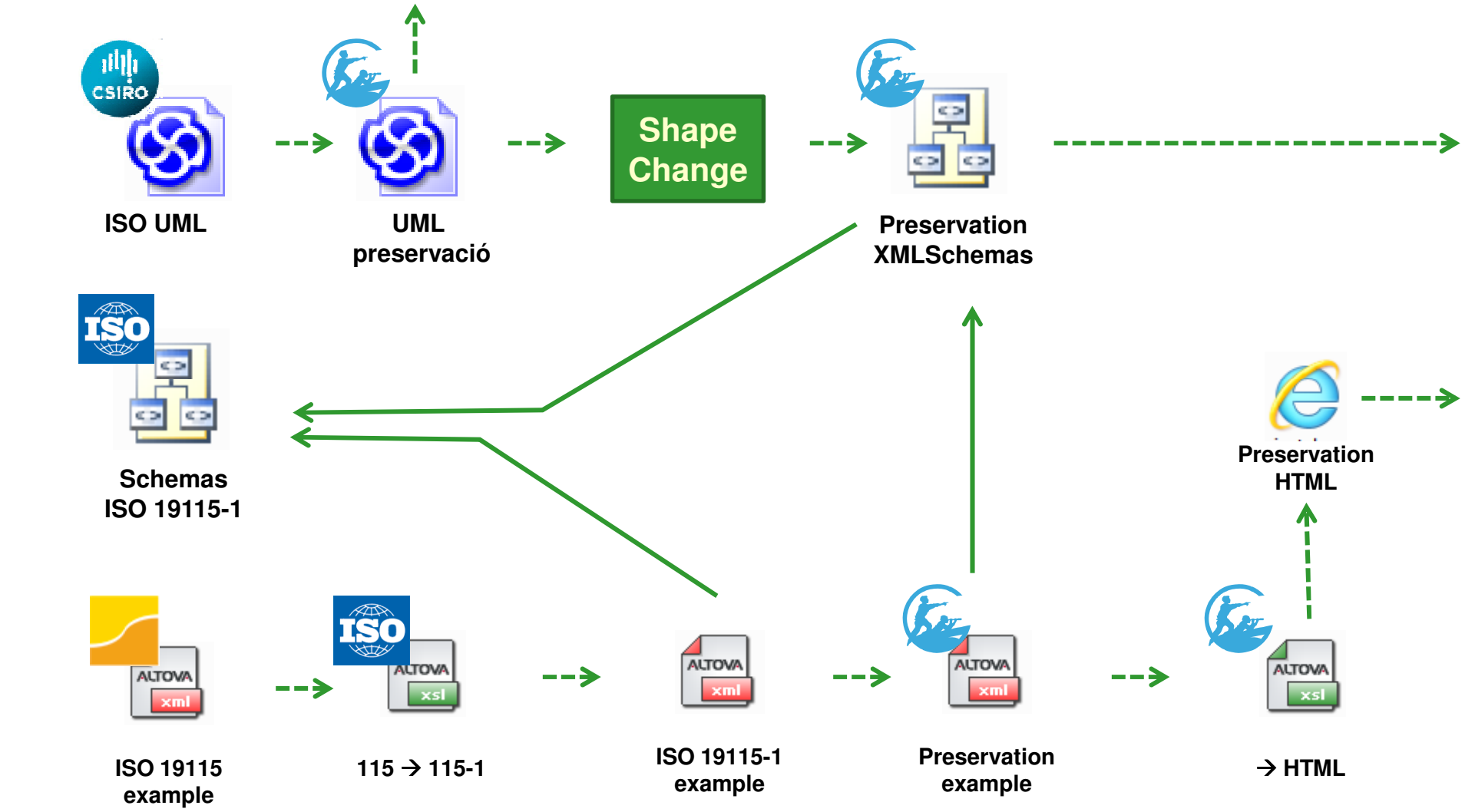
The OAIS context and the series

- Both the OAIS and the ISO 19115 include the concepts of cartographic series and granularity



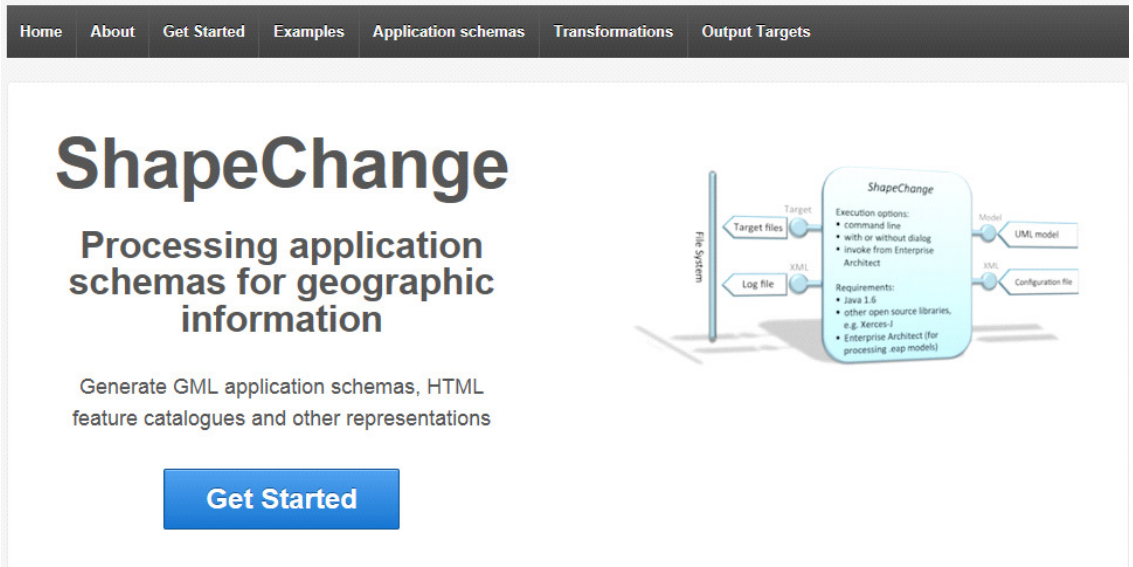
Step 3: Going into practice

Diagram of the process



Transformation of the model to XSD

- If the UML is generated following certain rules, it is possible to make an "automatic" transformation.
- Program developed by the company of Clemens Portele and that has been used to generate
  - The XSD of the INSPIRE model GML
  - The XSD of the ISO 191115-1 (that will be called 19115-3)



How to start with ISO 19115-1

- Thanks to the relationship with Ted Habermann (current Chair of the XML group of TC211) we obtained:
  - Access to the draft of the XSD schemas of the 19115-3
    - It has a modular design that requires the definition of many different interrelated namespaces
  - Access to an XSL that allows the automatic transformation of XML document s from 19115 (19139) to 19115-1 (19115-3)
- We created a new namespace "gpm" (Geospatial preservation metadata) with the provisional large name http: // www.opengis.uab.cat/preservation

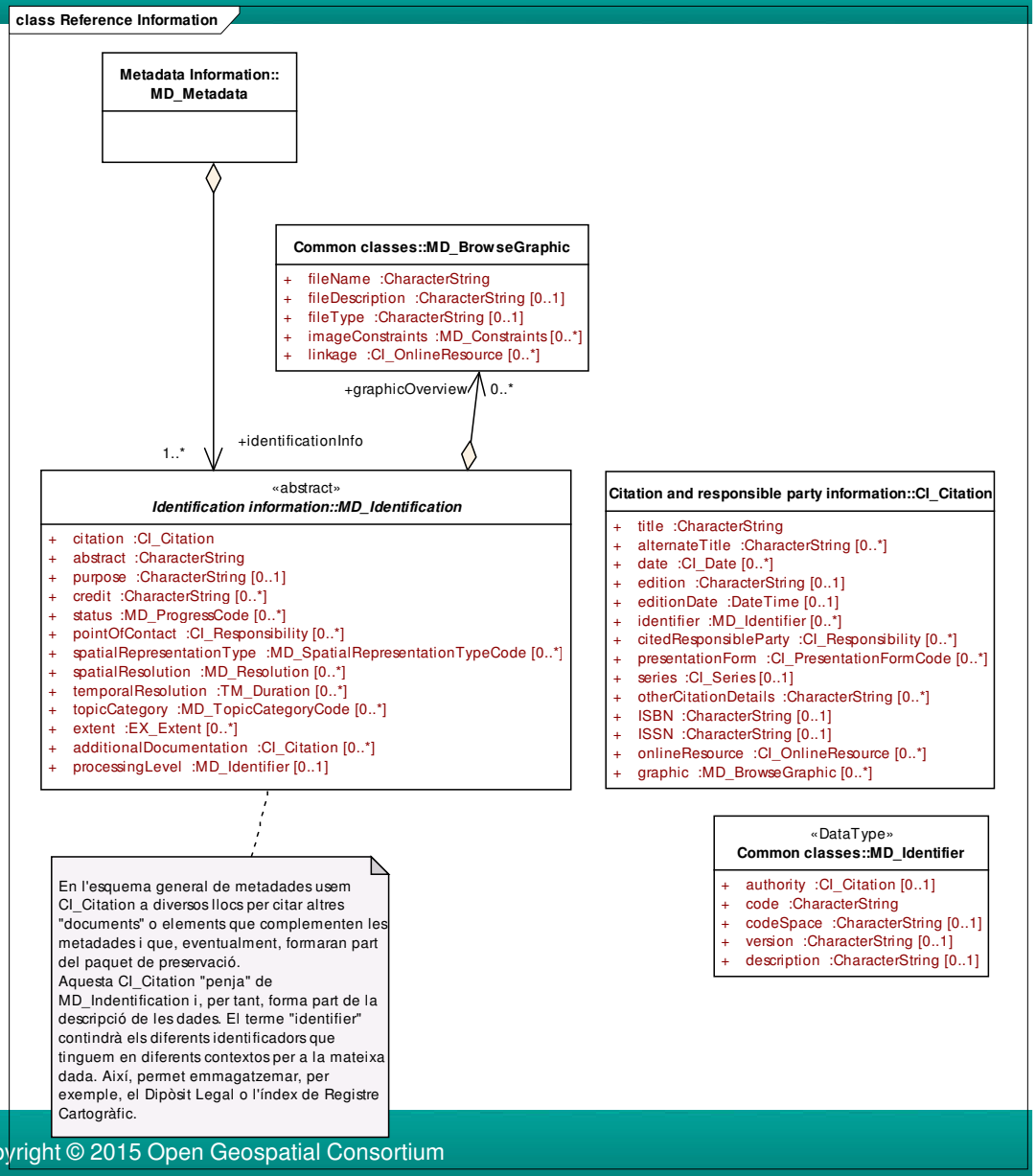
XML document: preservation parts

```
<gpm:GP_PreservationMetadata xmlns:gpm="http://www.opengis.uab.cat/preservation" xmlns:cat="http://www.isotc211.org/2005/cat/1.0/201
    xsi:schemaLocation="http://www.isotc211.org/2005/gco ../../../ISO_19139_Schemas/gco/gco.xsd
                        http://www.opengis.net/gml/3.2 ../../../ISO_19136_Schemas/gml.xsd
                        http://www.isotc211.org/2005/gts ../../../ISO_19139_Schemas/gts/gts.xsd
                        http://www.isotc211.org/2005/msr/1.0/2013-06-24 ../../../xsd20130624/ISO19115-3/msr/1.0/spatialRepresentation.xsd
                        http://www.isotc211.org/2005/mco/1.0/2013-06-24 ../../../xsd20130624/ISO19115-3/mco/1.0/constraints.xsd
                        http://www.isotc211.org/2005/mrc/1.0/2013-06-24 ../../../xsd20130624/ISO19115-3/mrc/1.0/content.xsd
                        http://www.isotc211.org/2005/mrd/1.0/2013-06-24 ../../../xsd20130624/ISO19115-3/mrd/1.0/distribution.xsd
                        http://www.isotc211.org/2005/mrl/1.0/2013-06-24 ../../../xsd20130624/ISO19115-3/mrl/1.0/lineage.xsd
                        http://www.isotc211.org/2005/mdb/1.0/2013-06-24 ../../../xsd20130624/ISO19115-3/mdb/1.0/mdb.xsd
                        http://www.isotc211.org/2005/mas/1.0/2013-06-24 ../../../xsd20130624/ISO19115-3/mas/1.0/applicationSchema.xsd
                        http://www.isotc211.org/2005/mpc/1.0/2013-06-24 ../../../xsd20130624/ISO19115-3/mpc/1.0/portrayalCatalogue.xsd
                        http://www.opengis.uab.cat/preservation ../GP_Metadata.xsd">
  <mdb:metadataIdentifier>
    <mcc:MD_Identifier>
      <mcc:code>
        <gco:CharacterString>of5mv60sd0f279110s1r100ca5</gco:CharacterString>
      </mcc:code>
    </mcc:MD_Identifier>
  </mdb:metadataIdentifier>
  <mdb:defaultLocale>
```

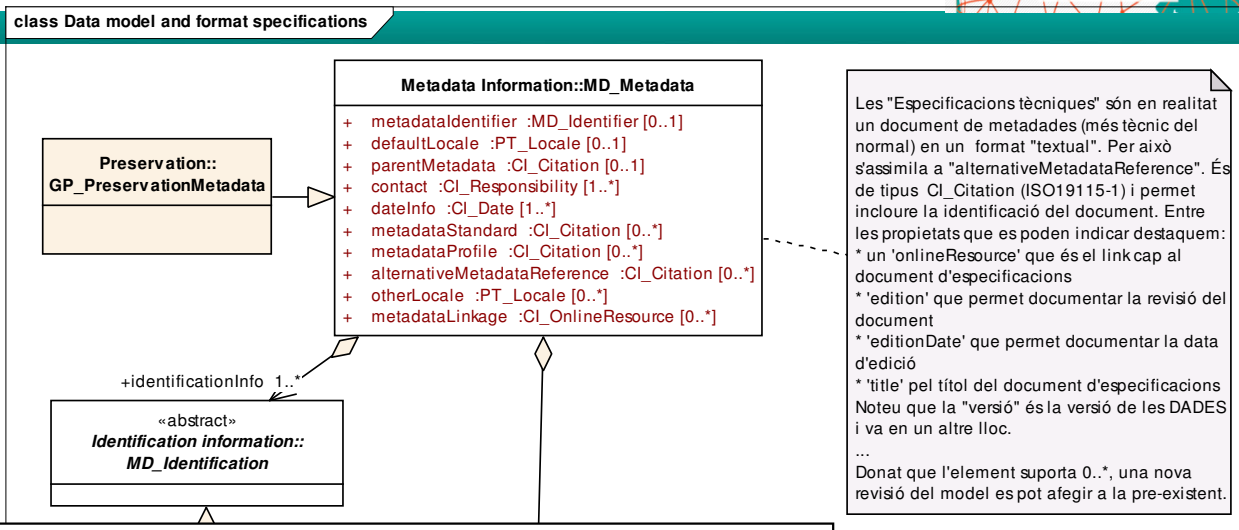
Identifiers

Reference information

Identifier: GPM_of5mv60sd0f279110s1r100ca5
Description: storage code
Responsible organization: IT department
Responsible person: J. Fuentes
Position: Head of the IT department
Phone: +34935671500



Technical documentation

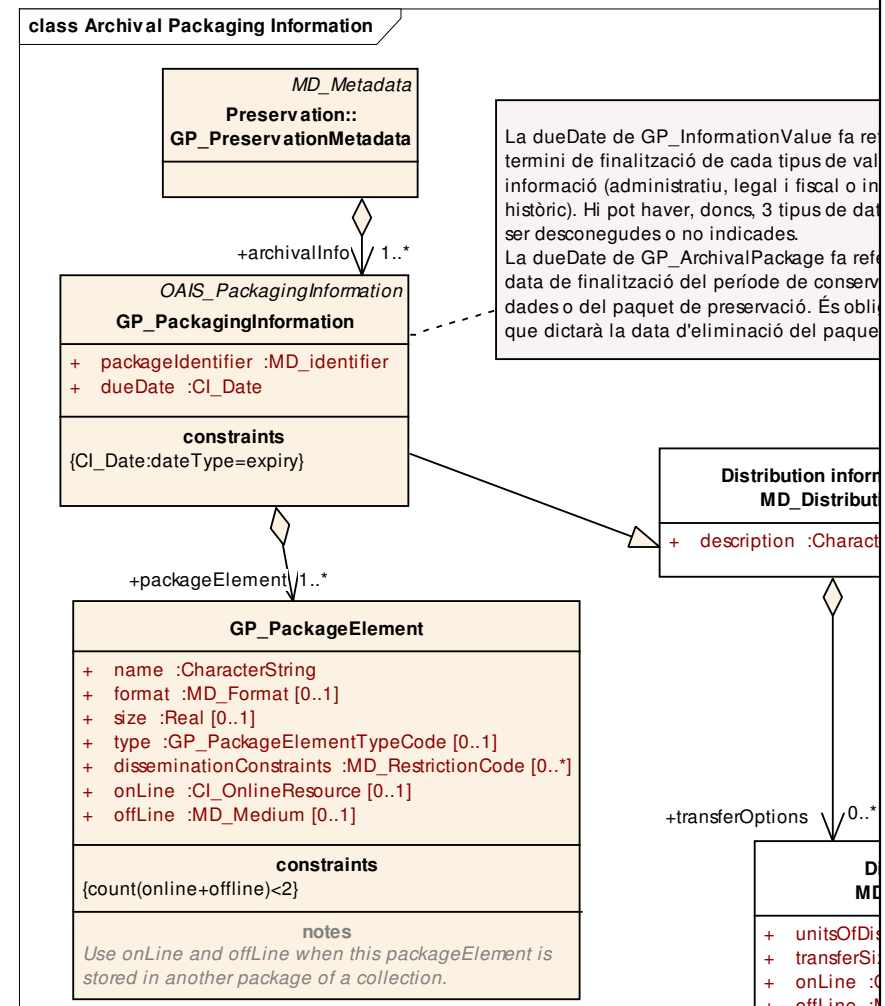


Attached documentation

Title: TECHNICAL SPECIFICATIONS for the Ortofoto de Catalunya 1:5000 (OF-5M) v6.1
Date: 2012-12-31T00:00:00
Data type: revision 1.0
Edition: 6.1
Related organization: Institut Cartogràfic de Catalunya
Role: creator

Title: MrSID FORMAT SPECIFICATIONS for the Ortofoto de Catalunya 1:5000 (OF-5M) v6.1
Date: 2011-06-30T00:00:00
Data type: revision 1
Edition: Format implementation 1
Related organization: Institut Cartogràfic de Catalunya
Role: creator

Packed information



Package information

Package identifier: GPM_of5mv60sd0f279110s1r100ca5
Responsible organization for the identification: ICGC-Map Library
Responsible person for the identifcation: Carme Montaner
Position: Head of the Map Library
Contact phone: +34935671500

Description: This is the preservation package for the orthophoto of5mv60sd0f279110s1r100ca5

Responsible person for the preservation: J. Fuentes
Position: Head of the IT department
Role: Storage responsible
Contact phone: +34935671500

Date: 2019-12-31T00:00:00
Data type: Limit for preservation

Storage formats

Package components

File name: of5mv60sd0f279110s1r100ca5.xml
File size: 0.006MB
Information type: metadata
Usage restrictions: Intellectual property rights

Context information

Context information

Resource name: Base topogràfica de Catalunya 1:5000 – Full "Monistrolet" (279110)

Identifier: bt5mv20mm1f279110c1r040

Identifier description: Resource identifier

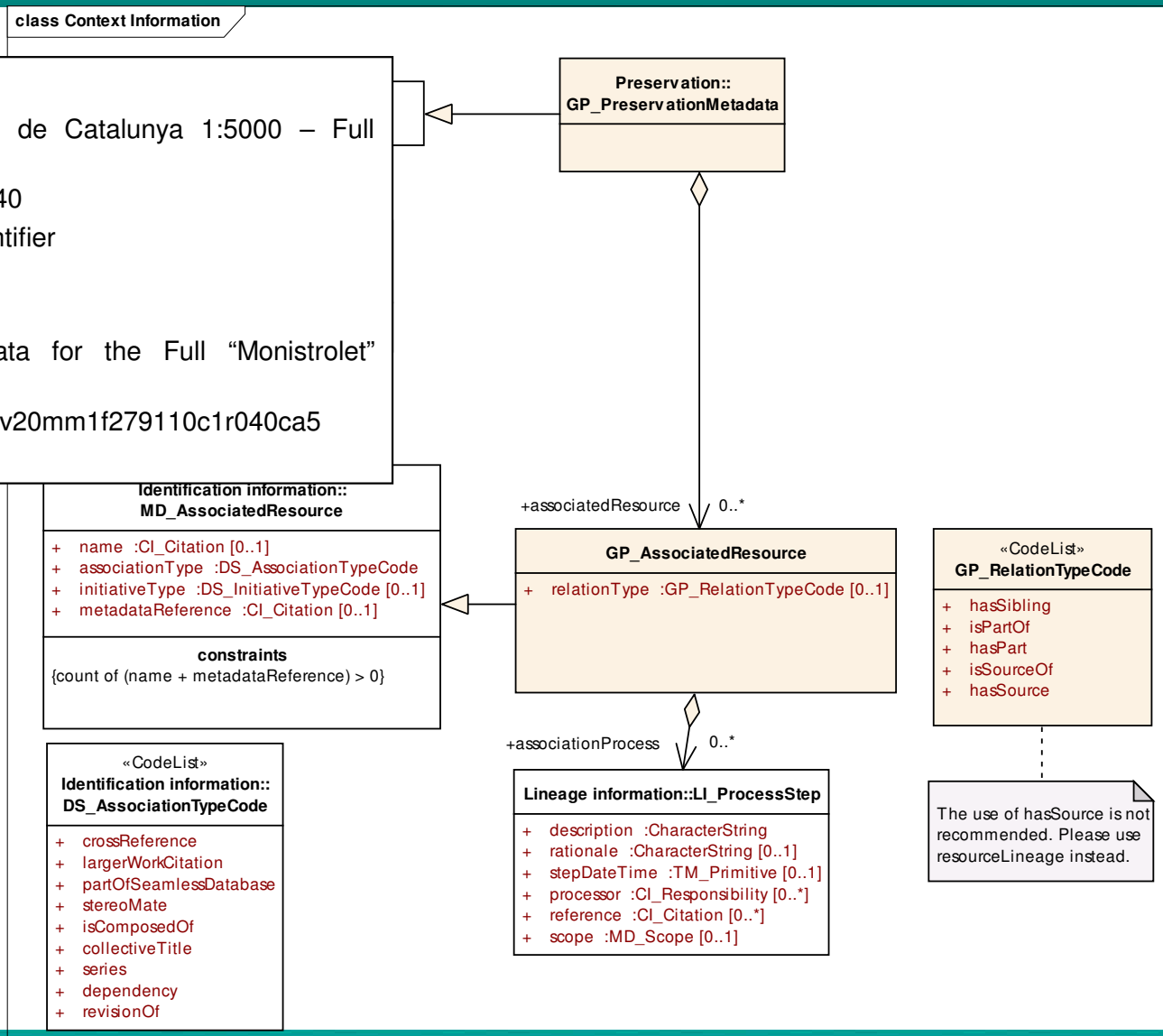
Association type: Reference

Initiative type: campaign

Metadata document title: Metadata for the Full "Monistrolet" (279110)

Metadata document identifier: bt5mv20mm1f279110c1r040ca5

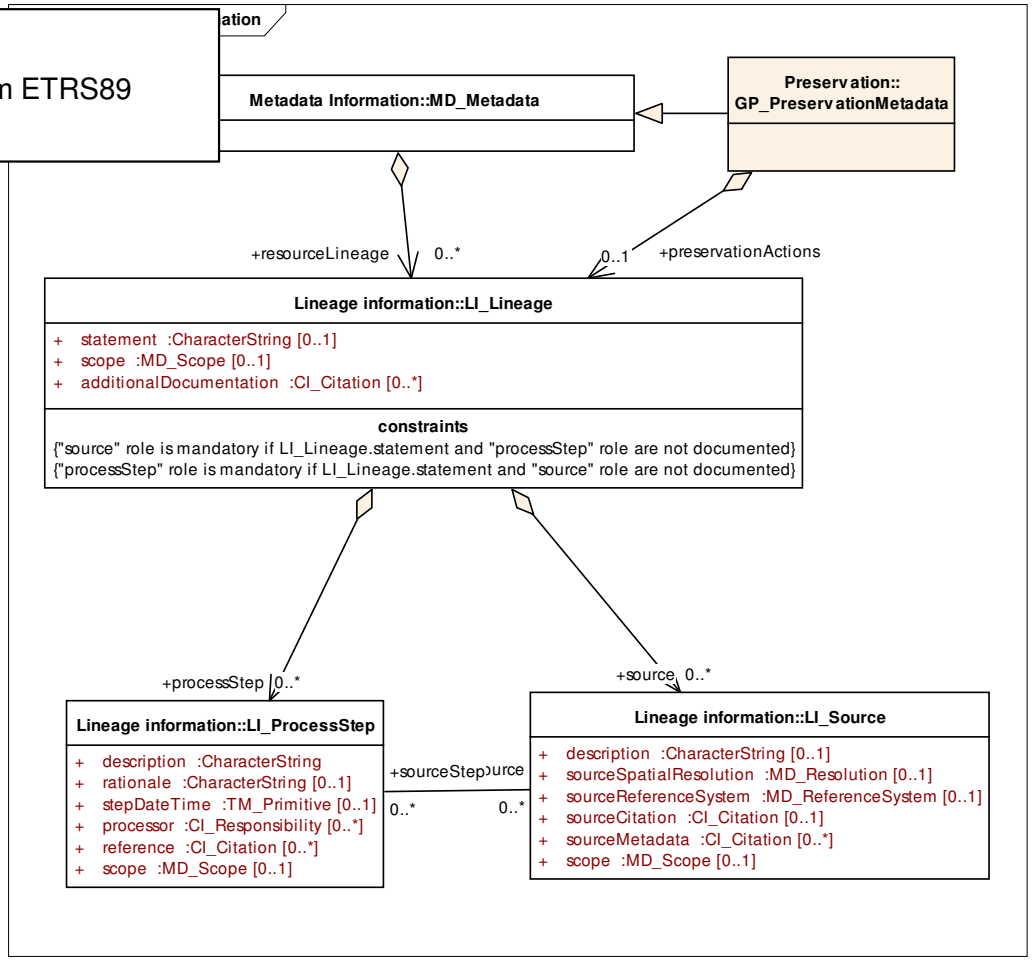
Identifier description:



Preservation actions

Preservation actions

Action description: File transformed to the new datum ETRS89

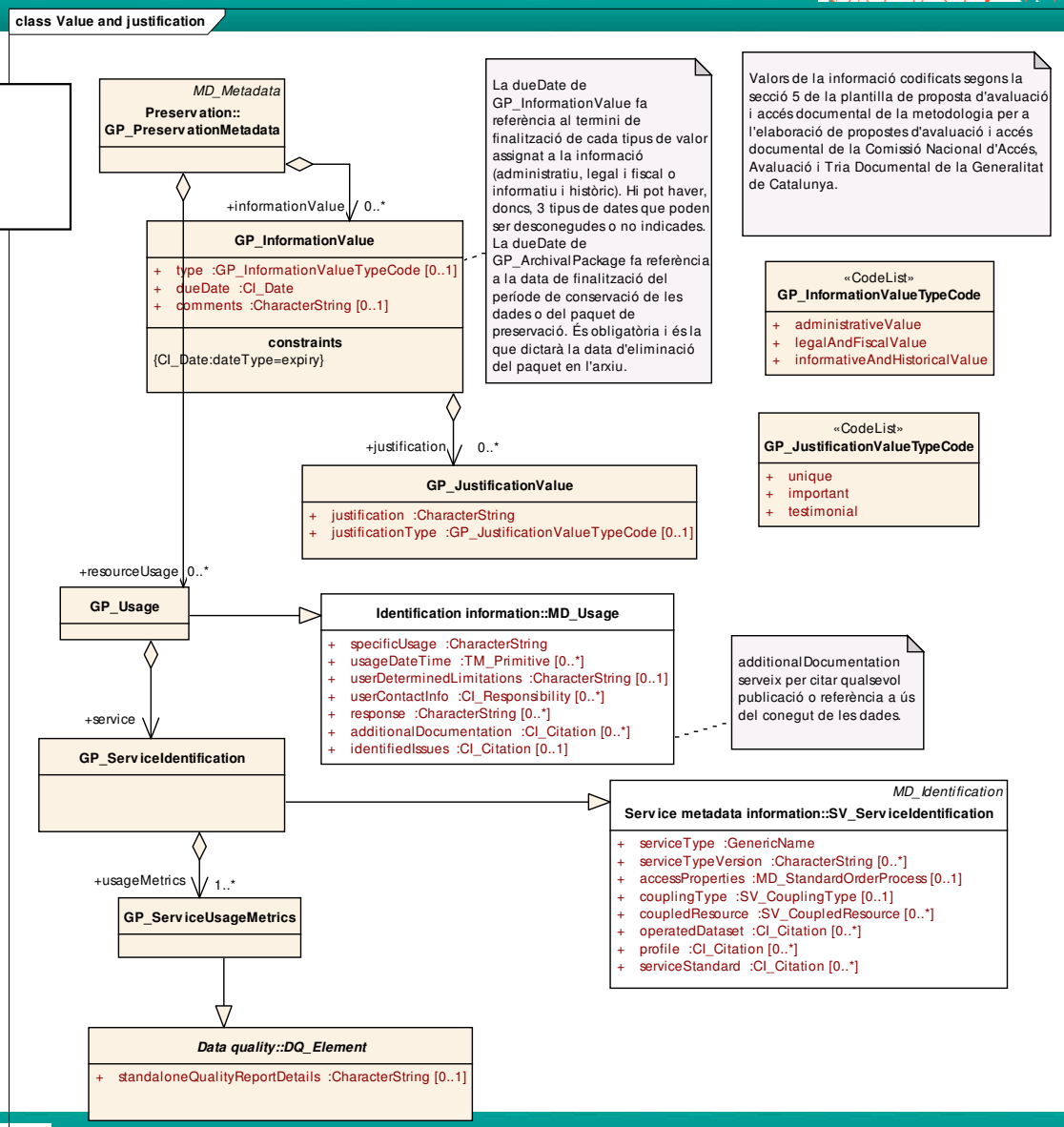


Value, justification and use

Value and justification

Value type: administrative

…



Conclusions

- We have defined a model that adopts the ISO19115-1 and makes minimal extensions to cover data preservation
- The model has been tested with real examples

Since we detect a need for this, we want to...

- Influence the new preservation ISO 19165 standard and group.
- Reinvigorate the Data Preservation group in OGC
- Influence the preservation subgroup of the Catalonia Cartographic Roadmap (Pla Cartogràfic de Catalunya)
- Help into the implementation in the ICGC.

Thanks

Joan Masó (CREAF)