Towards open, flexible, interoperable, semantic environments & formats

AXIS-CSRM 2019

Autonomous eXchange packages for Interoperable Systems

Conceptual Semantic Reference Model

Editor: NPO TITAN : www.titan.be

Authors:

Guy Maréchal – Roger Roberts (Titan asbl)

With contributions of Franck Casado (Semsitivity) & Guillaume Rachez (Perfect-Memory)

Secrétariat: Marie-Laure Schellings (RTBF)

The present version V 0.98 of the AXIS-CSRM documents is a 'Work in Progress'

Release V 0.99 (2019-05-11) version EN

Preamble

AXIS-CSRM proposes a conceptual modelisation of the presentation and the exchange of **Documentary Contents** between technical and human **Entities** in order of ensuring the **Interoperability** of these Entities (temporal, systemic, historical and cultural); an interoperability that unfreezes users from the constraints of proprietary systems [not captive to their suppliers

AXIS-CSRM stands for:

Autonomous e**X**change packages for **I**nteroperable **S**ystems **C**onceptual **S**emantic **R**eference **M**odel

The proposed conceptual approach is founded on powerful basic concepts: Entity / Configuration / Representation / Declaration of Existence / Knowledge base / Modelling Technology / Register / Point & Segment (static or dynamic) / Work – Edito – Artefact – Item / …

In the 1990s, Guy Maréchal, the godfather of this project, was the technical coordinator of the AMICE project that developed the CIMOSA architecture, the basis of the creation of SAP. However, important aspects of CIMOSA were not implemented in SAP due to the lack of computing power of the 1990s. AXIS-CSRM is in line with the will of the time, but with available a powerful computer and the possibility of computer modelling as close as we want the meaning that humans give to things.

The definition of AXIS-CSRM took inspiration from the ARCADE and CIMOSA projects (which have been the foundation for SAP) learning from their strengths & weaknesses. In a way, we can say that

- "AXIS-CSRM is the 2020 CIMOSA", or
- "AXIS-CSRM is the generic and semantic SAP".

ISO's standardization of the Open Archival Information System (OAIS[1]) in January 2002 weighed heavily on the vision of this project. Led by the Consultative Committee for Space Data Systems, this model is a reference outlining the functions, responsibilities and organization of a system to preserve information (particularly digital data) over the long term to ensure access to identified user communities. AXIS-CSRM defines a methodology for a concrete realization of OAIS compliant systems based on the semantic approach, such that the interoperability and, in particular, the persistence will be guaranteed by construction.

In 2006, the non-profit association Titan took the initiative to study and develop **AXIS-CSRM** within the European project "Memories" based on an expression of the needs expressed by the "Memory of the World" (MoW) of UNESCO (under the leadership of Ms Joie Springer [Programme Manager] and with contributions from the former head Abdelaziz Habib.)[2]

Subsequently, projects of the CelticPlus Cluster (MediaMap, MediaMap+), the IASA, the UNESCO and the working groups of the EBU-MIM have largely contributed to the development of the concepts. The general properties and aims of AXIS-CSRM have found an echo in the IASA-OK initiative [of the Organizing Knowledge task force of the International Association for Sound and Audiovisual Archives] and in the UNESCO Memory of the World programme.

AXIS-CSRM has been organized in three parts:

1. **Part 1:** A description of the origins of the project and an executive summary that describes in a substantial way the main part of the modelling work that has been carried out! This allows a quick visit of the results. Chapters 3, 4 and 5 aims at introducing the problematic: the objectives, the current situation (its weaknesses and the pitfalls) and the tracks for identifying valuable long-term solutions.
2. **Part 2:** The chapter 6 is devoted to the Conceptual Modelisation itself which is expressed through eight interlinked models. The three main and most innovative ones are :
    a. The “Functional Model” which generalizes and merge many of the functional standards; it explicitly links the data modelling within the functional modelling; it places as nodal the evolution: the alternation of the stable states and of the evolutionist processes of things.

[1] OAIS has been registered as an ISO standard under reference 14721: 2012

[2] . See UNESCO and TITAN collaboration during the European Memories project (FP6-IST 035300)

b. The "**Semantic Data Model**" which presents a unified frame for deciding for declaring the existence of things, for representing and managing them and their models.

c. The "**Configuration Management Model**" managing the implementations and operations of AXIS-CSRM based systems.

3. **Part 3:** The last part (chapter 7 and annexes) intends of introducing elements of the implementation of the concepts, such as they receive a concrete illustration and demonstrate the capacity of the current technologies for implementing the concepts in an efficient and easy to use manners. This last part still remains, on purpose, elementary.

Since 2008, digital technologies have powerfully increased people's global networking, new forms of communication (e-mails, social networks) and decentralization in the flow of data. In short, the 5 V of Big Data (Volume, Variety, Velocity, Value and Veracity) have taken colours! In this universal concert, the Titan non-profit organization stands out by its desire to establish a "semantic" primacy in this deluge of data. Identifying, naming organizing and indexing are the foundation of communication between Humans, between Machines and between Humans and Machines. It is not only naming the objects, but also the facts, concepts, instructions, processes and representations of these objects, facts, concepts, instructions ... are also named and identifiable.

For all these reasons, the association advocates via AXIS-CSRM:

- a semantic modelling of the modularity and interoperability of digital content allowing the migration, archiving and control of the evolution of organizations and their processes.
- the provision of tools for the construction of related knowledge bases for the sharing of metadata and models.

It is not enough to define a conceptual model for adoption! It must in addition ensure it is validated, not only for its internal coherence, but also for its ability to respond to actual or latent needs.

That semantic approach is foundations for granting the interoperability according to its four axes (temporal / technical / cultural / historical) not only between Machines, but also between Human and Machines, and, indirectly and transitively between Humans.

This project is also due to the International Federations, Associations and Library Institutions that have been working on preservation for decades (IFLA-CIDOC), as well as the LOC (Library of Congress).

The approach has been confronted with the reality of pilot projects that have implemented the most innovative elements of AXIS-CSRM, especially the Eureka Celtic(Plus) MediaMap and MediaMap+ projects. The latter was first short listed by Eureka experts and the evaluation awarded the « **CelticPlus Excellence Award 2015 of the Category: Services and Applications**», which has been delivered in Vienna in late April.

It faced other experts via presentations and discussions including four workshops of the "European Media Wrapper Round Table" [TITAN's conferences] and in the context of the '**IASA-OK** Taskforce' [**I**nternational **A**ssociation of **S**ound and **A**udiovisual archive - **O**rganizing **K**nowledge].

The first lecture of AXIS-CRM (first acronym) occurred during EMWRT IX in Amsterdam on September 13, 2013. Developments contained in this document have been discussed within the project MediaMap+ (Eureka CelticPlus) which ended in June 2014.

On March 19, 2015, a general introduction and the summary of AXIS-CSRM have been presented to Mrs. Iskra Panevska (the new head of the UNESCO's MoW program).

AXIS-CSRM was presented at a number of conferences and workshops, particularly the EBU MIM-MDN Workshop *(Metadata Developer Network)* on 9 and 10 June 2015, at the two PASIG-2016 conferences (in Prague and New York).

The definition of this project was the object of a separate report after the IASA Conference 2014 in Cape Town under the title: AXIS OK – Organizing Knowledge (including meetings of the 'Task Force' IASA -OK).

Within the framework of UNESCO, the AXIS-OK project (Organizing Knowledge) was defined as complementary to the PERSIST project (Platform to Enhance the Sustainability of the Information Society Transglobally - see: https://www.unesco.nl/digital-sustainability).

This version consolidates the two previous publications: September 2013 and October 2014 (Final review MediaMap+). These previous working versions have been invaluable for confronting the concepts with expert panels and, in addition to supporting the MediaMap + project, to carry out three focused pilot projects (in collaboration with the University of Lens[3]). This 2019 working version seems to be stabilized at the conceptual level, but it will still have to rely on exchanges during dissemination sessions and publication via audio-visual recording of tutorials.

This 2019 version of AXIS-CSRM is a working version focused on modelling (Chapter 6) where the first chapters are simply summarized to show the logic of the whole document.

A pilot project is being defined: it will illustrate how in practice to move from the conceptual level to concrete achievements. This pilot project will deliver in open source a standard knowledge base, a methodology of realization and exploitation as well as the key elements of the software developed for this pilot.

Two versions of the document exist: one expressed in French and one in English (the two official languages of UNESCO). Both versions, in their final form will be placed on the portal of the "Memory of the World" program of UNESCO, in accordance with decisions taken at the end of the IST-project MEMORIES.

This document is an invitation made to contribute and/or join for the setting-up a concrete project aiming at realizing the detailed specification of that AXIS format and illustrative samples of domain / media specific profiles and at implementing an open platform demonstrating the capabilities of the approach.

Roger Roberts
Président TITAN asbl
Président-Voorzitter de Be MoW / UNESCO

Mobile : +32 475 49 17 71
E-mail : roger.roberts904@gmail.com

Guy-Noël Maréchal
Membre de TITAN asbl
Membre de Be MoW / UNESCO
Fondateur du Comité IASA-OK
Mobile : +32 496 21 14 99
E-mail : guy.noel.marechal@gmail.com

[3] In order to validate certain concepts, the non-profit organization also initiated and managed three projects study and development with IG2I teams (Engineer in Computer and Industrial Engineering) Lens, a department of Ecole Centrale de Lille:

Acknowledgments

Thanks to all (and there are many) who contributed to the advancement of this semantic saga!

The authors wish to thank those who have helped us to lead in the development of AXIS CSRM!

- The patience and the complicity of their wives deserve more than a few words! This achievement would not have been possible without the generosity of our two 'Colette'!
- Thank you for then Franck Casado, Guillaume Rachez and Marie-Laure Schellings who really put their hands in the dirty grease; their CV and contributions are listed below.
- Then, a special thanks to those who have ensured proofreading (any remaining errors or problems remain obviously of the sole responsibility of the authors):
  - Charles Bébert (FR), Adrian Scott (UK), Jean-Louis Blanchart (BE), Colette Simar (BE), Hugues Dethienne (BE) !

Then, a thank you to those who have given us a general support and their contributions to the general analysis:

Jean-Pierre Evain (EBU-CH), Bruno Bachimont (UTC-FR), Peter Hermann (CelticPlus-DR), Jean-Hugues Chenot (INA – FR), Isabelle Canno (Radio France – FR), Abdelaziz Abid (UNESCO – TU), Kevin Bradley (IASA – AU), Pio Pellizzari (IASA-CH), Stefano Cavaglieri (IASA – CH), Ivan Herman (W3C – NL), Yves Raymond (BBC - Fr), Kara Van Malssen (AVP – US), Tony Shaw (LoC – US), Simon Rooks (BL – UK), Fran Alexander (OpenReg - CA), Laurent Boch (RAI-IT), Jan Takke (Philips – NL), Philippe Van Bastelaer (Philips – BE), Richard Ranft (BL – UK), Joie Springer (UNESCO-FR), Richard Wright (BBC-UK), Marc Leman (UKL – BE), Mark Plumbley (University of Surrey – UK), Nicholas Crofts (Olympic – UK & CH), Frank Jaeger (RTL-LU), David Butler (BBC-UK), Raoul Monnier (INA-FR), Georgios Ioannidis (EPFL - CH), Michaël Lamot (InnovIRIS - BE), Robert Meersman (VUB – BE), Youri Korsak Koulagenko (Atos – BE).

A special thanks to:

- Jakob Vlietstra (NL & USA) who has been Research Manager at Philips electronics and the project leader of the ARCADE and CIMOSA projects, to
- Michel Merten (BE) who was the CEO of MEMNON (Main contractor of MEMORIES) and
- Steny Solitude (FR) the CEO of Perfect-Memory (leading partner in MediaMap and MediaMap+) and his R&D team [Guillaume Rachez (FR) - Aymeric Brisse (FR)].

Finally, an outstanding thanks to all the members of the Non Profit Association TITAN for their persistent support and in particular:

- Jean-Louis Blanchart,
- Anne-Françoise Decelle and
- Francis Bodson.

The AUTHORS

Guy MARECHAL

I made my full 'formal' carrier with PHILIPS Electronics [At start in one of its subsidiaries, MBLE]. I ended it as Manager of the Research & Development Laboratories of "Philips Professional Systems" where I was in charge of the consumer applications of the cryptography to media and of the 'Professional Audio'. At my retirement, I established a one person consulting facility called PROSIP. I am a founding member of the TITAN NPA [Non Profit Association], member of the IASA (in charge of the creation of the IASA-Organizing Knowledge committee [IASA-OK]), administrator in the NPA of a large school of Brussels, member of the Club of Rome (*European Chapter*), Member of the Belgian Committee for the UNESCO Memory of the World programme …

In 1961, I got my Master in Electronics and in Architecture. I became assistant in one of the firsts European University Computing Centres. I joined the Advance Research Centre of Philips Brussels focusing on the parametric amplifiers (professor Vorobeitchick) and on the advanced Work flows Data Bases (professor Belevitch) with the ERAE [Entity Attribute Relationship Event] model, prelude to some aspects of AXIS-CSRM.
In 1970, I joined the applied research department for applying the parametric amplifiers to the highly reliable security system of the first generation of Nuclear Plants; then for instrumentation for textile industry; then for the SHAPE and NATO an archival system of documents using digital recording on a battery of optical disks (later it becomes one of the bases of the development of the CD-Audio).

In 1970, I joined the Computer department of Philips and became in charge of the coordination of the after sales technical maintenance [UNIDATA consortium]. In parallel I has been in charge of missions for the coordination of the advanced research in Computer Assisted Design and Production. In particular I have been the technical coordinator of the CIM-OSA project which has been a seed for the SAP business and technical leader of the ARCADE project which has been one of the seeds of the OAIS standard. I want to thank especially three persons who provide me trust, tracks and encouragement: M. Claude Fosséprez, Jakob Vlietstra and Jan Takke.

Retired, I made a consultancy for Philips, the British Library, the Library of Congress, the UNESCO, the CIRTEF, TVFI, Memnon Archiving, Perfect Memory, Belgavox, the Norwegian Institute for Sound Recording, Siemens, Ministry of Justice of Belgium, the Belgian CSA, the RTBF, …

In the context of my collaboration with TITAN, within the MEMORIES project, I designed the essence of the AXIS-CSRM concepts. That has initiated my collaboration with the Memory of the World programme of the UNESCO.

I am 'vice-chairman' of the IASA-OK Committee that I created in 2015.

E-mail: guy.noel.marechal@gmail.com

Roger ROBERTS

I was literally born on an ancient border (between Belgium and Prussia from 1815 to 1921) in 1950 in Welkenraedt, so I lived a childhood and adolescence bathed in bilingualism and multicultural (old Duchy of Limburg); to Herbesthal for primary school and in Eupen for college (now in German-speaking Community).

At 18, I undertake a candidature in Economic, Social and Political Sciences at the Catholic University of Louvain (UCL).

In 1970, my first Photokina: the technological fairs virus will never leave me (Montreux until 99, IBC since 1984, NAB since 1995, CeBit since 1999 ...)

At UCL, I pursue a licence in Social Communications. It is also the encounter with semantics (Ferdinand de Saussure, Jacques Derrida, Roland Barthes ... that will resurface during my R&D course).I finished my studies with specialization in professional degree of "Ciné/TV director" at the Institut des Arts de Diffusion (IAD) from 1972 to 1975.

From 1975 to 1982, I made many slide shows and some movies for various private companies! In particular, a "Belgium Today", an AV program show in 6 US cities to celebrate the 150th anniversary of Belgium in the United States!

In 1982 I enter at the RTBF! My career changes, I become addicted to an extension of my job: live directing for the Information Department and later for the Sports!

I met Pierre L'Hoest and Laurent Minguet who founded EVS. So I will be closely involved in the development of a digital slow motion system fitted for sports competitions. I use this LSM for the first time in 1990 in a Formula 1 Grand Prix in Francorchamps. I discovered the erratic nature of technological change and the difficulty of accrediting new solutions. In 1992, Michael Schumacher won his first F1 Grand-Prix, this was actually my last and the abandonment of my TV director career.

In 1992, I accepted the position of "Chef de Réalisation" (managing the director teams).

In 1993 I chair the working group "Technical and forward-looking" within the "Audiovisual Crossroads" initiated by the Minister Di Rupo. This was the basis for launching the Titan association (Televisual Interactive Terminal and Associated Network) a multidisciplinary non-profit organization founded in order to facilitate the transition to digital technologies (see www.titan.be). It is the meeting with Guy Maréchal (Philips), Francis Bodson (Canal+ Belgium), Jean-Louis Blanchart (FWB) that will occupy me for the rest of my professional life. This "Think Tank" coordinated a digital decoder project, modulation tests on networks, proposals for open access systems, AIME (for Intelligent Multimedia Archiving Economic) of Cirtef, technology watch activities on the topics the open exchange of media content with their associated metadata and controls (EMWRT: European Media Wrapper Round Tables).

Therefore, I was elected member of ISF (Information Society Forum: an initiative of Commissioner Bangemann), where I will meet Robert Pestel which will also influence my technological dissemination capital.

From 1995 to 2014, I am part of the team organizing the "SEFOR" (Training Seminar) on behalf of Cirtef in many countries, especially in Africa.

In 2003, I assumed the position of head of Common Cultural Facilities (archive, play-out and graphics, in addition to the *'Bureau de Réalisation'*) to the Production Department.

In spring 2005, I met Steny Solitude (Perfect Memory) with whom we share today nearly eight years of design work and prototype implementations within MediaMap and MediaMap+ projects.

April 2015, I finish my career at the DGTE, R&D Knowledge Management, and coordinate the technology watch on the exchange of multimedia content with their semantics associated in the chains of production and audiovisual exploitation.

During the AXIS-CSRM enterprise I contributed to editorial work, namely constant work on concepts and their descriptive and visual changes since the first drafts.

I am the president of the Belgian Committee for the UNESCO Memory of the World programme.

E-mail : roger.roberts904@gmail.com

The CONTRIBUTORS

Frank CASADO

I had the pleasure of meeting the two previous authors at the Memories project where I represented Memnon [4] (a major player in the market for the digitization and professional archiving of media content) as a software engineer specialized in processing and enrichment of audiovisual content.

As part of Memories and MediaMap, I developed a "Proof of Concepts" covering the most innovative elements of the AXIS CSRM approach namely the construction of interoperability by **AXE** interchange [**A**utonomous e**X**change **E**ntities] built on the semantic approach and the incorporation of profiles; the exchanges and the mechanism of the statement of existence (in particular the Axis-Foot Print), the static/dynamic fusion, the recursion representations/operations, orthogonal structuring with respect to the representations and the integration with the finite states machines.

I built this validator as a nUpplets native database that I called **ISIS** [**I**nteroperable **S**emantic **I**nformation **S**ystem]. This database was built on an 'Open Source' approach based on a software substrate also in 'Open Source'. In addition, ISIS incorporates natively the representation of finite state machines in layered mode and coupled with a generic mechanism for representing state vectors [the key to validation and exception handling]. ISIS is moreover self-extracting which immediately validated the approach of the interoperability via the communication of AXE. ISIS has been implemented in a system of semantic enrichment and structuring of audiovisual content marketed by Memnon under the name of **IPI**® [**I**ntegrated **P**roxy **I**ndexer®].

In summary, I think I have demonstrated that the implementation of the AXIS-CSRM concepts is not only possible but corresponds to *a significant breakthrough* in the way we approach content and process management [structured, unstructured and semi-structured, structured, represented as interoperable semantic networks in the framework of profiles].

I graduated from a polytechnic engineering school in 2004 and holds a master's degree in signal processing. I started my career as a developer in voice recognition technology integration (including natural language processing and dynamic human-machine dialogue management).

In 2010 I created my own company SEMSITIVITY IT service, specialized in semantic computing, expert systems, virtual reality and process control. I also work as a freelance consultant for leading Belgian broadcasters, press groups and lead an innovative project on a mobile platform using motion recognition, augmented reality and virtual reality technologies.

As part of the development of AXIS-CSRM, I realized some punctual studies (some of which are included in 'Part III').

SEMSITIVITY SPRL-S: http://semsitivity.be/

E-Mail : casado.f@gmail.com

The authors want to pay homage to Franck and to testify to a professional collaboration which has turned into a rare friendship!

[4] Memnon Archiving Services www.memnon.eu

Guillaume RACHEZ

Born on 4 May 1984 in Roubaix, the former world capital of wool work, Guillaume Rachez is a “ch'ti” at heart! He is deeply attached to his northern roots, the old remodelled manufactures, the belfries that mark the plain, the coal and its heaps, and the hop that quenches in the absence of light! Because of this, an intimate proximity with Belgium and its surrealist practices!

Fan of Star Wars since the beginning, he turns to the mathematical and computer science while being a movie passionate: films of author to the pure entertainments, there is always matter to discuss and to debate infinitely. He regularly writes reviews online. And then the guitar, today in solo, with The Edege or David Gilmoure as guitar hero or a Mark Knopfler for his placid virtuosity! Athletic (jogging), he is an active supporter of the football LOSC Lille team and the French national team (Les Bleus)!

From 2002 to 2004, studious, applied, he discovers computer developments and finds jubilant pleasure in solving problems by implementing sequence of elementary operations according to a defined process leading to a solution, alias algorithms!

In 2004-2005, a Bachelor's degree in Audiovisual and Multimedia at the University of Mont Houy (Valenciennes) completes the arid scientific training! It is the adventure of audiovisual production, the service to the development of the media, ... an understanding of the codes of this world!

In the summer of 2005, an internship to discover the reality of the field and its proprietary tools such as Avid Station Administrator, Technical Assistance to editors, Authoring DVD, video encoding, and … purchase order management.

In 2005-2006, a master's in Multimedia, in the "Dream" department, still at the University du Mont Houy, reinforces the dialectic of scientific training and the practice of audiovisual practice! It is the opening to image processing, indexing, compression, mastery of theoretical models in the service of the media!

In spring 2007, a well-structured letter of application opens the doors of SKEMA in Compiègne. It is the encounter with Steny Solitude who sponsors this UTC Spin-Off and who proclaims with aplomb aim to provide all means to produce quality video while respecting an editorial line on the basis of a Audiovisual scenario modelling. To do this, there is a team in charge of developing video processing software (indexing, encoding). All these questions are also waived by different players present in a Eureka Celtic project (MediaMap). Other contacts through meetings organized by Titan asbl (Bruno Bachimont, Guy Maréchal, Roger Roberts, ...), especially during the second EMWRT in Paris, which preaches the virtues of ISO's OAIS standard!

In 2008, from Skema to Perfect-Memory with the same objectives: to enable the exchange of information between heterogeneous and closed audiovisual systems, ontology engineering, analysis and integration of semantic technologies, development Of video software (image analysis, automatic editing), analysis of customer needs, integration of services, training on the challenges of the data web ... This is the opportunity to penetrate deeply the universe of broadcasters (Radio France, RTBF) or the institutions (Pep's : Preservation Program of the Wallonia-Brussels Federation).

With MediaMap+ here is the EBUCore live with his "Evaingelist" J-P Evain... an opportunity to rub against normalization at its highest level! Fairs (IBC Amsterdam), Conferences (the EMWRT suite, the SITEM, ...).

In August 2014, the virus of daily production takes over with the creation of a company TiteMaison Production! Identifying a real need for any structure to break through the technical barriers associated with video production, he invests in tools to democratize the use of video as a medium of valorization. Result: two years of production of corporate promotional films, educational videos, analysis of customer needs, while keeping a foot in the modelling through the developments of the Titan association (AXIS-MOM and AXIS-SOW) in Lens with IG2I teams of engineers.

January 2017, the return to front-line interoperability modelling at Perfect-Memory!

Mail : guillaume.rachez@gmail.com

Marie-Laure SCHELLINGS

I was born in Hermalle-sous-Argenteau in 1963 not far from the Meuse, just before entering the Netherlands.

After primary and secondary studies in economics at Visé and Liège, I graduated as secretary in evening classes. And I have always cultivated a passion for languages and reading.

For a whole year, I work at the secretariat of the Libraries Les Chiroux (Liege) in multitasking mode. In 1989, I returned to RTBF (television : Continuity service) in Brussels, then switch to information at the production centre in Liège!

I am attached to the department of information on Radio for many years, « Antenne-Midi », Antenne-Soir, Liège-Matin, La Première, as well as assistant and secretary. Particularly involved in Walloon and intercultural programs with Guy Fontaine (journalist and producer), I also work for Edmond Blattchen (Nom de Dieu- TV), sport with Thierry Luthers (Droit au but - La Première-Radio). In 2002, for the "radio soap opera of the summer: Dr. Renaud, Mister Renard" broadcast on French Public Radios, La Première (Belgium), France Inter, RSR (Radio Suisse Romande) and Radio Canada ... more than 10 hours with the complicity of producer Dominick Martinot-Lagarde. This made me discover a rich and friendly universe.

In 2005, I met Roger Roberts during trainings organized as part of the RTBF. He gives me the virus of his passion for technological evolution. A great professional change that I have never regretted.

In 2008, I am in charge of the administrative and logistical dimensions of the MediaMap project. During 3 years, I will take care of the supervision of the writing of the numerous project deliverables, as well as their formatting, English translation and archiving. In this project, I discover the work of Titan asbl and get to know its members.

« Never change a winning team », and so I resume in 2012 the same functions for the project MediaMap +.

In the AXIS-CSRM saga, I contributed to the transcription of sound recordings in text mode, formatting, chapter coherence, links to illustrations, modifications, FR / UK synchronization, and the many changes made to the text since 2012 until mid-2015.

Courriel : mls@rtbf.be

Table of Contents

Illustration table :

Towards an open, flexible, interoperable semantic format & environment

AXIS-CSRM 2019

Autonomous eXchange packages for Interoperable Systems

Conceptual Semantic Reference Model

Part 1: Introduction & MOTIVATIONS

The part of the version 0.98 present an abstract of the first chapters described in part 1 (release 2014)

1 Chapter 1: a Logo and an initiative of the Titan association:

The NPO TITAN conducts an initiative, called AXIS-CSRM, for defining a pivot /axis format being open, flexible and interoperable, suitable for constructing persistent archives, for enabling easy 360° publishing, for facilitating an effective interchange between independent systems or data bases, facilitating the migration of 'flat' systems to 'semantic' systems or of 'semantic' systems based on different ontologies /taxonomies and for empowering aggregation portals. AXIS-CSRM is expressed using only open representation and internationally recognized standards.

That AXIS format would be based on a core semantic profile including mainly an upper-ontology (with associated upper-taxonomy/thesaurus/terminology/configuration management); a resolvable URI allocation and management protocols and hooks for domain and media oriented specific profiles; it will be autonomous in the sense that all the profiles involved in an instance of export and import would be included in the wrapped interchange data.

The axonomic logo of AXIS-CSRM illustrates the intentions:

Figure 1-1*:* The AXIS-CSRM « axonomic » logo

AXIS-CSRM logo is gait artwork that covers the French term "sens", an illustration of the main goal:

The French term in itself is rich in meanings (a faculty to perceive sensations, an intuitive knowledge, a faculty of judgement, ...), but in this case it is indeed the « meaning » option that "prevails" with a touch of « direction to take »!

2 Chapter 2 AXIS-CSRM: Executive summary:

2.1 GENERAL OBJECTIVE

The chapter 2 introduces **AXIS-CSRM** as **Conceptual Framework** to represent **INFORMATION** in the form of **DATA,** hence to organize the **temporal, systemic, historical and cultural "interoperability's" of the signified**.

Information underpinned by these data must be such as **HUMAN** can rebuild original information, to the next; or that **MACHINES** can interchange theses data in an open, interoperable and persistent way and can perform treatments close to what humans would conduct based on the information. Beyond the writing of this document AXIS-CSRM discusses the development of these concepts, their validation and encourages the implementation of open systems based on the following concepts:

1. **From Information to Data** : To develop an architecture and IT concepts in such a way that the actors and agents can represent the information they intend to store or transmit through data encoded according to conventions representation formats.
2. **From Data to Information:** To show that this approach is powerful! It allows to symmetrically represent the data for processing by machines and perceptible expression by humans, and that the initial information may be perceived accurately to its creation by the author or third parties in an indeterminate future.
3. **From Information to Data // to 'Business' // to Data // to Information:** To show that this approach is realistic and can cover a whole chain from the representations of intentions (projects) to various forms of tenure (in human perception, economic or commercial).

AXIS-CSRM does not impose any kind of implementation but has been designed for an easy partial or full realization using the Semantic and Web 3 open and standardized technologies. The use of a semantic approach aims to bring data modelling and processes to the nearest information and events they represent. AXIS-CSRM is only a framework, that is, it provides a way to embed and attach the specific content generated by users and represented according to their formats and technologies (flat or semantics) of modelling.

2.2 UNESCO – Documentary Heritage :

This project was initiated in June 2006 as part of the project MEMORIES based on an expression of the needs expressed by the "Memory of the World" (MoW) of UNESCO (under the leadership of Ms Joie Springer [Programme Manager] and with contributions from the former head Abdelaziz Habib.)

As the main objective of AXIS-CSRM is to ensure "interoperability" through a framework that meets many **social, economic and cultural** interoperability objectives, the chapter detailing those defined by UNESCO of the conference held in **Vancouver in 2012**: **« The Memory of the World in the Digital Age »: digitization, preservation and access to the « Documentary Heritage »:**
http://www.unesco.org/new/fr/communication-and-information/events/calendar-of-events/events-websites/the-memory-of-the-world-in-the-digital-age-digitization-and-preservation/

2.3 UNESCO & AXIS-OK collaboration (Organizing Knowledge)

The definition of this project was the object of a separate report after the IASA Conference 2014 in Cape Town under the title: AXIS OK – Organizing Knowledge (including meetings of the 'Task Force' IASA -OK).

Within the framework of UNESCO, the AXIS-OK project (Organizing Knowledge) was defined as complementary to the PERSIST project (Platform to Enhance the Sustainability of the Information Society Transglobally - see: https://www.unesco.nl/digital-sustainability)

The objective of AXIS-OK is to add the "ACCESS" dimension to "PERSIST" with an approach to the representation of knowledge of documents that are easy to find and navigable for users with rights. To do this, there are representation standards (RDF, OWL, ..) and interaction services with the Open Data. The non-profit organization also has an advantage, namely an ontological basis for structuring a domain via AXIS-CSRM.

For more on the AXIS-OK initiative, see the Annex 3

2.4 Core Features of the AXIS-CSRM Approach

In order of obtaining the **interoperability by construction**, AXIS-CSRM has selected the following basic characteristics:

- The conceptual approach is declined according to eight interlinked models:
  1. The **Identification** and **Naming** model
  2. The **Interoperability** model
  3. The **Functional** model
  4. The **Data** model
  5. The **Behavioural** model
  6. The **Security** model
  7. The **Configuration Management** model
  8. The **Import – Export** model

- The Documentary Contents represent **Entities** and the **Configurations** which structures and links these Entities, their Representations in their Context.

- AXIS-CSRM proposes that Configurations be represented according to **Modelling Technologies** expressed explicitly and exclusively via object-oriented representation standards. The definitions of these Modelling Technologies are explicitly incorporated into **Knowledge Bases**. In particular, when implemented on the Internet, these representations are expressed in accordance with the standards of the **Semantic Web** and the **Internet of Things**.

- The representations of the Entities are expressed according to Modelling Technologies, as far as possible or reasonable, expressed, as for the MT of Configuration, via object-oriented representation standards; otherwise, according to Modelling Technologies adapted to the type of Entity (for example ISDN for a book) or type of representation of Entity (for example JPEG to represent a photo). In cases where secret / private / proprietary Modelling Technology are used, AXIS-CSRM recommends that at least the Entity's representation file incorporates a 'Proxy' model, represented in accordance with a Formal Modelling Technology, which is integrated into the Representation **Registry** the Entity.

- The Entities concerned can be of any kind, sometimes seen as Static Entities (Natural Objects, Artefacts, QR-Code, RF-ID, Documents, Images, Laws, Conventions, Titles, Contexts, Concepts ...) or as Operating Entities (Live Beings, Process, Audio Stream, Video Stream ...). The 'Finite State Machines' approach not only represents [in 'intent', 'operation' and 'results'] the operations of the Operating Entities, but also ensures that the Static Entities can be involved in their operations positioned in historicity.

- The consideration of an Entity begins with a Declaration of Existence ∃ of the Entity to which the various representations of the Entity and the Contexts concerned are related, in particular, the Knowledge Bases involved.

- Interoperable exchanges between technical entities are themselves documentary entities characterized by the fact that they are **autonomous**, in the sense that the knowledge base associated with the documentary content is incorporated or associated with the 'package'. In AXIS-CSRM these Interchange Entities are referred to by the acronym **AXE** [**A**utonomous e**X**change **E**ntities]. The AXEs are usually encapsulated via a particular technology adapted to the data transfer: the 'packaging' of AXES.

- The presentations in Perception and in Interaction with humans are built specifically to the characteristics of the mediating artefact.

3 Chapter 3 : The Empire of the senses

This chapter 3 is an attempt to formulate some conceptual bases of the representation, memorization and processing of the information and the communication of the data representing it. If, for simplicity, the information is not reducible to the message (which is only a representation of it), and communication is the process of transmission and interactions between the sender and the receiver, there are two distinct domains to which the acronym ITC has conferred a too highly transcendental character. A little detour through linguistics and semiology is therefore essential!

We shall here treat the distance between the “signifier” and the “signified” only in a limited way.

Here are the details of the points addressed in this chapter 3:

3.1 Underlying Concepts

This point presents the results of the work of linguists (Ferdinand de Saussure, Jacques Derrida) and Roman Jakobson (for Communication). An analysis of the representation of the link between an object and its representation, between its representation and meaning through a painting by René Magritte - The Betrayal of Images 1928-29: "This is not a pipe"!

3.2 The representation of meaning

A definition of the signifiers: ideogram - phonogram - glyph - pictogram and the signifier / signified relation resulting from the work of linguistics.

3.3 The representation of the phonograms (signifier)

Between languages that could not be written in an ISO or proprietary standard and those that were lost in format conflicts, not to mention the impossibility of multilingual files, the computer world offers ASCII (American Standard Code for Information Interchange) as a universal standard to enable digital exchanges. ASCII is sufficient to represent English texts, but it is too limited for other languages, including French and its accented letters. This point describes the whole saga to arrive at the UTF-8 (Universal Character Set - 8 bits), used by the vast majority of websites in the world, more than 80% which has the advantage of being backwards compatible with ASCII. The UTF-8 allows to code more than one million characters, spread over a maximum of 4 bytes of code.

3.4 The Semantic Web

The Semantic Web (invented by Tim Berners-Lee) is a mesh of data associated so that it can be easily processed by computer machines rather than by human operators. It is an extended version of the existing World Wide Web and represents an effective way to move from representing data in the form of information to contextual representations via knowledge bases. The Semantic Web targets the conversion of the Web from unstructured documents to a navigable web of information / data (the machine interprets terms and relationships between terms).

3.5 From Data - Information to Communication // Des Données à l’Information et à la Communication

A Data is a representation of information (facts, concepts or instructions), information is the meaning that humans assign to data by means of conventions. And finally, communication is a process by which living organisms define and share meaning. This point also describes Jakobson's 6 language functions as well as the seventh "Performative:" defining and activating a target resource to do anything, anytime! "

3.6 The Representation of the Knowledge

To meet the needs of an **open digital world** it is essential to generate **explicit knowledge** in the computer system; this digital knowledge is the basis of interoperability between heterogeneous systems.

3.7 Symbolism related to the different models in AXIS-CSRM

3.8 The modelling of Processes, Events, Agents and Traces

One of the key features of AXIS-CSRM is its desire to model processes, agents involved and generated traces.

3.9 From Intention to Publishing through Recording: the prerequisites for modelling

3.10 General Terminology

AXIS CSRM relies, as much as possible, on standardized definitions of basic concepts. In particular, ISO definitions are preferred.

3.11 AXIS-CSRM Terminology

In the specific context of AXIS-CSRM, some terms have a specific, restrictive use.

4 Chapter 4 Principles guiding the definition of AXIS CSRM :

Chapter 4 concerns the definition of the principles underlying the representation of AXIS CSRM using standards of semantic technologies, in particular Ontologies.

Here are the details of the points addressed in this chapter 4:

4.1 Interoperability concept definition

Interoperability is the capacity of a system whose interfaces are fully known to work with other existing or future systems with unrestricted access or implementation. Interoperability is a chain that links information and computer systems within an organization or with other organizations, governments, businesses or citizens.

4.2 Interoperability in space & in time

This point introduces the concepts of Persistence (interoperability in time) and Migration (evolving interoperability).

4.3 Efficiency of the pivot AXIS approach

The "upper profiles" or "upper ontologies" are relatively close to the representation of concepts well accepted and common to many different philosophies and epistemologies stabilized and accepted (at least in terms of objective knowledge)

4.4 Power of the semantic approach

4.4.1. Up to concepts
4.4.2. Many concepts are almost elements of philosophy commonly accepted and understood worldwide.
4.4.3. The semantic representation is standardized and mature.
4.4.4. The contribution of the Linked Open Data
4.4.5. The system specificity
4.4.6. The semantics allows to separate the conceptual layers orthogonally
4.4.7. Structured vs Unstructured
4.4.8. The Consistency

4.5 Definition of a standard modular format « Upper profile »

4.5.1 Exclusive use of internationally recognized open standards.
4.5.2 Hooks for "domain profiles"
4.5.3 « Configuration Management Services »
4.5.4 Volume reduction (Compactness)

4.6 General principles of modelling in AXIS-CSRM

4.6.1 “a ” Basic Epistemology modelling
4.6.2 Other general principles of modelling:

5 Chapter 5 The different standards implemented in the proposed solution:

This chapter reviews some of the international or industry standards that different information systems should adopt to exchange data. This chapter also aims to highlight the lack of expressiveness of some of these existing standards, or the difficulty of mobilizing "differentiating factors" for the targeted business.

Back to concepts, gain height leads to base interoperability regardless of the formats used locally or standard (as is done in MPEG-21, for example). The reason for going back to the concept provides a means for emancipation from the constraints of choice and representation formalism and structure. If the representation format has been changed, taking up semantics leads to locally change (no change propagation.)

This analysis primarily concerns ISO[5] (for STEP - Standard for the Exchange of Product Model Data & OAIS - Open Archival Information System), the IFLA[6] (for FRBF) and the LOC (Library of Congress for PREMIS) which have guided the developments of this project, the EBU[7], which is very active in the audiovisual field (EBUCore and CCDM Class Conceptual Data Model), the AMWA[8] / EBU collaboration in FIMS (Framework for Interoperable Media Services) and finally the Federations International Organizations, Associations and Library Institutions that have been working on preservation for decades (IFLA-CIDOC[9]): FRSAD (Functional Requirements for Subject Authority Data) and CIDOC-CRM.

Given the importance of ISO/OAIS modelling and the work of IFLA (FRBR) in the developments of this AXIS-CSRM project, a brief description of these standards is given in this document.

5.1 OAIS (ISO - Open Archival Information System)

We have already indicated how much ISO's standardization of the Open Archival Information System (OAIS) has weighed heavily on developments in this project. Initiated by the “Consultative Committee for Space Data Systems” in the late 1990s, this ISO-standardized model is a universal reference outlining the functions, responsibilities and organization of a system to preserve information. (in particular digital data), in the long term, to ensure access to identified user communities. This model was revised in 2012, including the semantic dimension in the vision. The data model of AXIS-CSRM implements fully the semantic approach introduced in the new version.

OAIS is an Information Model that processes both digital and non-digital objects simultaneously. The model must indeed be able to process existing physical objects, representation (physical objects) of the real world, but also the digital representations that describe them (digital objects). This practice makes it possible to make separate statements about an object, a document that describes it and the links that exist between objects and representations and their meanings (the signified).

In the OAIS model, the main thing is the creation of an Information Object. The diagram above (a vision specific to the Titan non-profit organization) clearly identifies the object data (bits and bytes) materialized by a representation tool (specific application) and interpreted at the level of meaning by a Knowledge Base.

In the context of “deep preservation” it is necessary both to preserve the data (on suitable media), the applications that generated these data and finally to create a knowledge base to generate the links between the data and their meaning (s). The preservation of data and applications is by no means the object of this project ... all the effort is focused on the ability to connect the world with their objects, their computer representations and their meanings. The creation of a knowledge base is nodal!

[5] ISO : International Organization for Standardization : https://www.iso.org/home.html
[6] IFLA : The International Federation of Library Associations and Institutions - https://www.ifla.org/
[7] UER/EBU : Union Européenne de Radio-télévision - European Broadcasting Union - https://www.ebu.ch/home
[8] AMWA : Advanced Media Workflow Association - https://www.amwa.tv/
[9] CIDOC : Comité International pour la Documentation (ICOM's International Committee for Documentation)

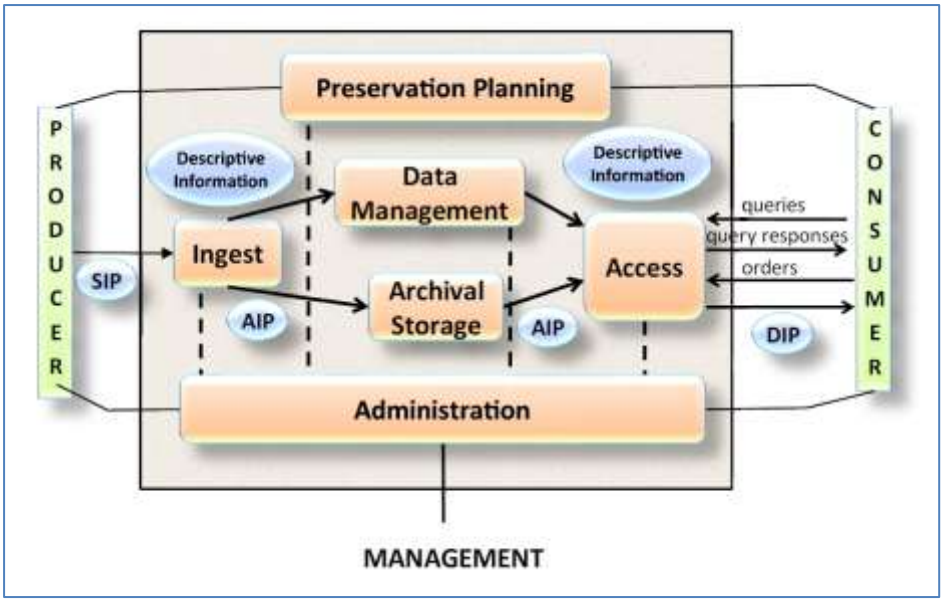


Figure 5-1: The OAIS Functional Model

1. Ingest: process of conformation of objects intended for a digital archive. The Submission Information Package (SIP) is the digital object to be archived with ancillary metadata. The Archival Information Package (AIP) is generated from the SIP as the final step in the ingestion process. The AIP contains all descriptive metadata, technical, project information, access and usage rights, processing (antivirus scans, extraction mode, ...).The ingest transfers the data of the producer (SIP) to the Archivist / Archives (AIP).
2. Archive storage: Following the ingest, the AIP is stored, maintained and retrievable from the archive centre. Archive storage includes persistent storage, regular checking of bit stream integrity, and disaster recovery.
3. Data Management: This feature supports searching and retrieving archived content using descriptive metadata.
4. Administration: Refers to day-to-day operations and maintenance of archives and coordination with other functions: archiving, user assistance, implementation and maintenance of policies and processes, etc.
5. Access: The interface that allows users to retrieve data from the archive. The information requested by the user is received as a set of broadcast information (DIP), generated from the AIP stored in the archive centre.
6. Preservation planning: archives must have a continuous digital preservation strategy (regularly updated) and be monitored regularly to detect the risks inherent in this type of activity.
7. Common services: IT services that any computer system, such as a digital archive, needs to function: hardware, software, data, processes, agents, feedback for improvements, etc.

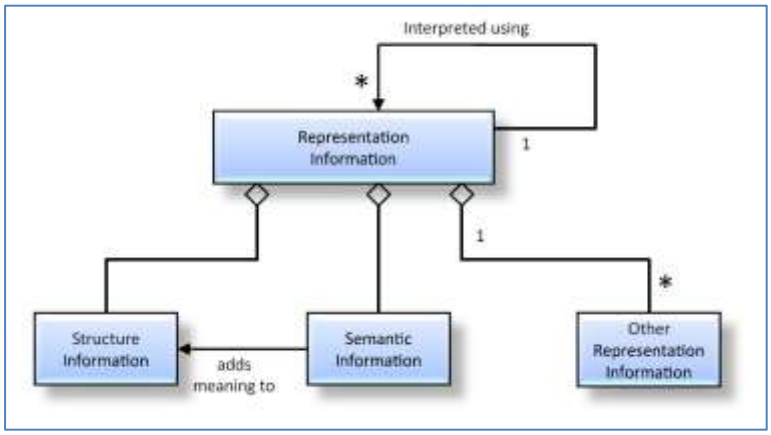


Figure 5-2: The OAIS Information Representation

The figure shows the introduction of the semantics in the 2012 version of the standards.
The OAIS standards are further introduced at Annex 5.

5.2 STEP Standard for the Exchange of Product Model Data (ISO 10303)

Source : https://en.wikipedia.org/wiki/ISO_10303

ISO 10303 is an ISO standard for the computer-interpretable representation and exchange of product manufacturing information. Its official title is: *Automation systems and integration — Product data representation and exchange*. It is known informally as "**STEP**", which stands for "Standard for the Exchange of Product model data". ISO 10303 can represent 3D objects in Computer-aided design (CAD) and related information.

The international standard's objective is to provide a mechanism that is capable of describing product data throughout the life cycle of a product, independent from any particular system. The nature of this description makes it suitable not only for neutral file exchange, but also as a basis for implementing and sharing product databases and archiving.[1]

Typically STEP can be used to exchange data between CAD, computer-aided manufacturing, computer-aided engineering, product data management/enterprise data modelling and other CAx systems. STEP addresses product data from mechanical and electrical design, geometric dimensioning and tolerancing, analysis and manufacturing, as well as additional information specific to various industries such as automotive, aerospace, building construction, ship, oil and gas, process plants and others.

5.3 IFLA Standards

5.3.1 IFLA - FRBR (Functional Requirements for Bibliographic Records)

FRBR) is a conceptual modelling of the information contained in library bibliographic records developed by an IFLA expert group from 1991 to 1997, approved in 1997 by the IFLA Standing Committee on the Cataloguing Section and published in 1998.
Roughly, FRBR organizes the different components of the bibliographic description from a logical point of view (the authorities, the subject accesses and the information on the document itself) into three groups of entities connected by relations (entity-association model):

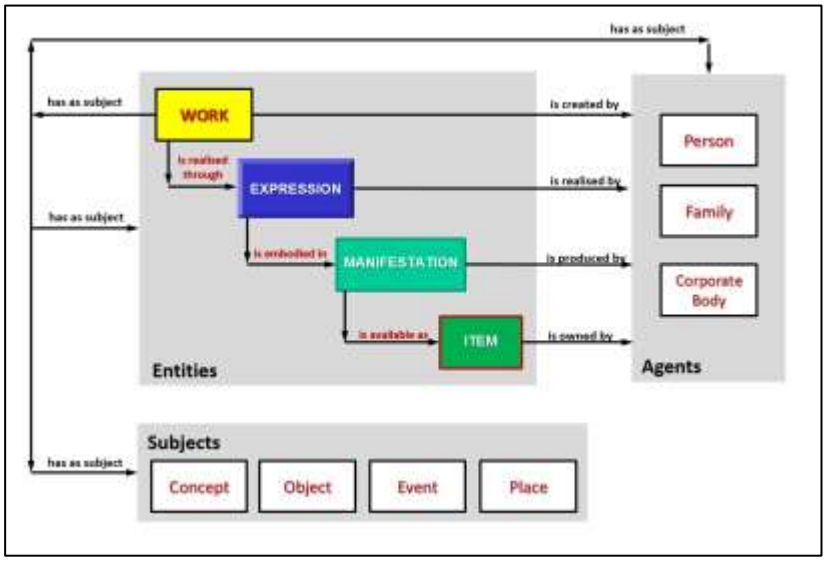


Figure 5-3: General FRBR schema

- The first group of entities includes everything about the documents and their different versions. These entities are: work, expression, manifestation, and item. The originality of the model is the notion of work, which can link such a novel and its translations or adaptations, a thing that catalogues are unable to do except with a rebound by title (if the title is the same) or by author.
- The second group of entities is the modelling of individuals and organizations who have a responsibility for creating entities in the first group. It is introduced the concept of relationship (creation, production or membership) to mark the way people operate in relation to the work-expression-manifestation-item event phase that is important.
- The third group manages entities that are the subject of work: concept, object, event, location.

Comments:

The main criticism of the FRBR model is to be centred on the description of accomplished and published works that corresponds to the status of a library (the result, not the process of gestation leading to the materialization of the work). In practice, however, this is an important step towards a library modelling concepts similar to those of the "Semantic Web." The entities of group 1, provide a model to multiple publication and description of versions, the entities of group 3 evoke ontologies describing for example the relationship between event and location.

The limitations of this approach are well known, the model is made in the conceptual sense, but to become “semantic” there must be a clear distinction between what is about the process and the result of the representation process (works).

5.3.2 The IFLA and CIDOC developments:

Between 2015 and 2017, the FRBR conceptual model was gradually replaced by the integrated IFLA-LRM model, which harmonizes and consolidates the three models developed separately by IFLA: FRBR for bibliographic data, FRAD for authority data and Functional Requirements for Subject Authority Data (FRSAD) for material authority data.

5.3.3 IFLA - Functional Requirements for Subject Authority Data [FRSAD] General Framework

The IFLA Working Group on Functional Requirements for Subject Authority Records (FRSAR), established in April 2005 by the Classification and Indexing Section considered it essential to study users of material authority data, to identify the contexts in which the data are used, and to characterize different Use scenarios.

5.3.4 The work of CIDOC (International Committee for DOCumentation) and ICOM (International Council of Museum):

CIDOC CRM was developed by the International Committee for DOCumentation (CIDOC) of the International Council of Museum (ICOM). This semantic model constitutes an "ontology" of information relating to cultural heritage, coming from the world of museums, with multi-domain ambitions (libraries, archives and research institutions).

5.4 PREMIS (Preservation Metadata Maintenance Activity - Library of Congress)

The PREMIS Data Dictionary is an international standard for metadata in order to ensure the preservation of digital objects and their long-term use. Developed by an international team of experts, PREMIS is implemented in digital conservation projects around the world, as well as a number of open-source and commercial systems for digital preservation tools. The Drafting Committee PREMIS coordinates revisions and the implementation of the standard, which consists of a Data Dictionary, a XML schema and documents.

5.5 EBU

The European Broadcasting Union (EBU) is the world's first public service media alliance (PSM) with 117 member organizations in 56 European countries (34 additional partners in Asia, Africa, Australasia and the Americas). The metadata specifications and classification schemes of the EBU are updated under the control of the EBU's MIM (Media Information Management) and MIM-MM (Metadata Models) committees. MIM MDN (Metadata Developer Network) has extended its activities to new semantic technologies and techniques based on Artificial Intelligence.

5.5.1 EBU Core

EBUCore defines a set of concepts, relationships and properties that apply to MEDIA. This is a part of metadata that can be used to describe any multimedia content. It is based on a model of metadata, Dublin Core. The compatibility of AXIS-CSRM with the EBUCore has been checked.

5.5.2 UER/EBU - Class Conceptual Data Model [CCDM]

The CCDM specification combines several aspects from existing models and specifications into a common framework. It has been built over several EBU attempts to represent broadcasting as a simple logical model. It has benefited from EBU MIM MDN work in metadata modelling (P-META and EBUCore) and semantic web developments. The distribution part has been designed to seek maximum mapping to TV-Anytime and the "BBC Programmes Ontology".

5.5.3 EBU/AMWA : FIMS (Framework for Interoperable Media Services)

FIMS is a joint initiative of the European Broadcasting Union (EBU) and the Advanced Media Workflow Association (AMWA). It is a framework of service definitions for implementing media processing using a Service Oriented Architecture (SOA) and the promotion of interoperability and reusability of services.

5.6 ABC Ontology

The Harmony Project is an international digital library project funded by DSTC (Australia), JISC (U.K.), and the NSF (U.S.). The broad goal of the project is to research methods and models for describing the variety of rich content that increasingly populates the Web and digital libraries.

The Harmony project has similar intentions as those of the AXIS-CSRM project: it intends to provide a common conceptual model to facilitate interoperability between metadata vocabularies from different domains. A conceptual model organizes the interoperability, then constructs it through a concrete semantic representation of the model. The intention is to cover not only the static objects but also the dynamic objects as targets. The ability to model change makes ABC appropriate for describing a wide variety of entities and the relationships between them. In particular, it has been designed to model physical, digital and analogue objects held in libraries, archives, and museums and on the Internet. This includes objects of all media types – text, image, video, audio, web-pages, and multimedia. It can also be used to model abstract concepts such as intellectual content and temporal entities such as performances or lifecycle events that happen to an object. In addition the model can be used to describe other fundamental entities that occur across many domains such as: agents (people, organizations … instruments), places and times.

The data model of AXIS-CRSM is been largely inspired by the organization of the *Entities* in the ABC ontology, i.e. ***temporality, actuality*** and ***abstraction***.

See : https://www.researchgate.net/publication/2847486_The_ABC_ontology_and_model

5.7 3D Visualization & Representation

3D Visualization is a technique for effectively communicating information and data (i.e. 3D models, 3D predefined views, diagrams, animations, etc.) for displaying definition data, derived data, and other objects archived data.

Towards an open, flexible, interoperable semantic format & environment

AXIS-CSRM 2019

Autonomous e**X**change packages for **I**nteroperable **S**ystems

Conceptual **S**emantic **R**eference **M**odel

Part 2:
Semantic MODELLING

Authors : Auteurs :

Guy Maréchal – Roger Roberts (Titan asbl)

The "Functional" and "Data'' Models in Chapter 6 of this version V 0.98 have been completely revised.

6 Chapitre 6 AXIS-CSRM Modelling:

Chapter 6 deals with the definition of the concepts and principles underlying the representation in AXIS- CSRM, using the standards of semantic technologies, in particular ontologies, as reference languages, and thus interoperability and scaleability of models in their implementation (semantic databases are flexible and not fixed).

In previous chapters, the issue of interoperability was introduced along with the principles of building systems that are interoperable by design. Chapter 6 will detail its implementation in the conceptual approach proposed.

In this chapter, the term 'model' is omnipresent. In science, it means the link between a concrete and a logical reality with full use of their descriptive analogies. There is a difference in nature between the model and the reality, the model having a symbolic value.

6.1 INTRODUCTION

6.1.1 Modelling & the eight models of AXIS-CSRM:

AXIS-CSRM expresses the modelling through eight interconnected and interoperable models. Their detailed description is the subject of the next sections:

1. The '**IDENTIFICATION**' & '**DENOMINATION**' model
   It describes the capacity of coping with several identification and naming systems and suggests a reliable generic approach. Such a system must be able to identify the location of a representation, facilitate the restructuring of a website and ensure the traceability of a museum object and its property modification web page (temporary provision, assignment or sale).
2. The '**INTEROPERABILITY**' model
   It describes the semantic encapsulation mechanisms such as being capable of reliable transfers and preservation of the essence of the information. It copes with the functional interoperability (the Human face) and the technical interoperability (the machine face). Innovation is to introduce into the packaging a Project layer that has the ability to represent and deliver the "substance" in the exchanges between heterogeneous systems. If the substance has been well designed and the project specified, humans will be able to interpret, through their culture, the underlying information provided by the computer system.
3. The '**FUNCTIONAL**' model
   It describes the "circular" integration of the representation of the documents and the flows of the processes which generate them. This model generalizes and merges the FRBR and OAIS models. In particular, it is able to cover all kinds of content (not just books) and describes the life cycle of these contents (including the processes involved and their contexts). On this basis, "Working" is specialized in four processes: [Manifesting]; [Authoring]; [Expressing] and [Enjoying].
4. The '**DATA**' model
   It focuses on the modelling of the Things, of their declaration of existence and on the inner and outer structures, states and procedures of the Things attached to the modelling of the events and processes.
5. The '**BEHAVIOURAL**' model
   It covers the modelling of the processes and their perceptions
6. The '**SECURITY**' model
   It focuses on the authentication of the Agents and on the integrity of the interactions between Agents
7. The '**CONFIGURATION MANAGEMENT'** model
   It covers the management of the existence of Things and of their models and of the management of the evolutions.
8. The '**IMPORT & EXPORT'** model
   It covers the organization of the interchange of 'Packages' between domains or between sub-domains. It also covers the 'mapping' onto the ISO OAIS standard.

Note:

Whenever possible, the terminology and graphic symbols will be those applicable of the ISO standards. The main source is: https://www.iso.org/en/news/2009/11/Ref1261.html

In particular, the ISO/IEC 20944-1: 2013 standard: https://www.iso.org/standard/51914.html

In terms of Rights & Media, the IFTA and FIAT/IFTA standards will be the reference:

http://fiatifta.org/index.php/media/fiatifta-publications/

In particular:
http://www.ifta-online.org/sites/default/files/IFTA%20International%20Standard%20Terms%20-%20MILA%205th%20Edition.pdf

6.1.2 AXIS-CSRM: The Global Interoperability Model

AXIS-CSRM deals with three types of interactions: communication between **Humans**, **Human-Machine** interaction (GUI interface) and **Machines** interconnection!

The inter-operational parts may be computer systems (known by the acronym ARD [Autonomous Resource Domain]), works or people. The myth of the 'Tower of Babel' illustrates the difficulty of combining diversity and collegiality; individual identity and membership of a group. Information technology can help humans to achieve this interoperability without sacrificing the specificity of each element.

The three founding core concepts are:

1. Consider cultural, social, and productive systems and / or policies as autonomous within the organization of their specificity;
2. Organize data exchanged between autonomous systems as representations which conform as closely as possible to the underlying concepts;
3. Arrange the underlying concepts so as to present them in a manner which is both easy to access and appropriate to the specific needs of each user.

These concepts are intended to operate:

- Over distance: bridging the gap between today's independent and autonomous systems (different organizations, internal technical migrations or interoperability between different departments within the same organization, etc.).
- Over time: delivering interoperability between archives from the past and the technologies of the present/future.
- Via the widest possible variety of technical platforms, including all available forms of human interface (screen, keyboard, tablet, PDA, TV, second screen, and so on).

In this context, three components are to be considered:

- Exchange between multiple heterogeneous IT systems
- Access to computer networks by human
- The Configuration Management Services that ensure sustainability features.

To emphasize the principle of autonomy in the Alice and Bob domains, AXIS-CSRM describes them as Autonomous Resource Domains (ARD) playing different roles within the same system (Bob can represent the storage area for an Alice application.)

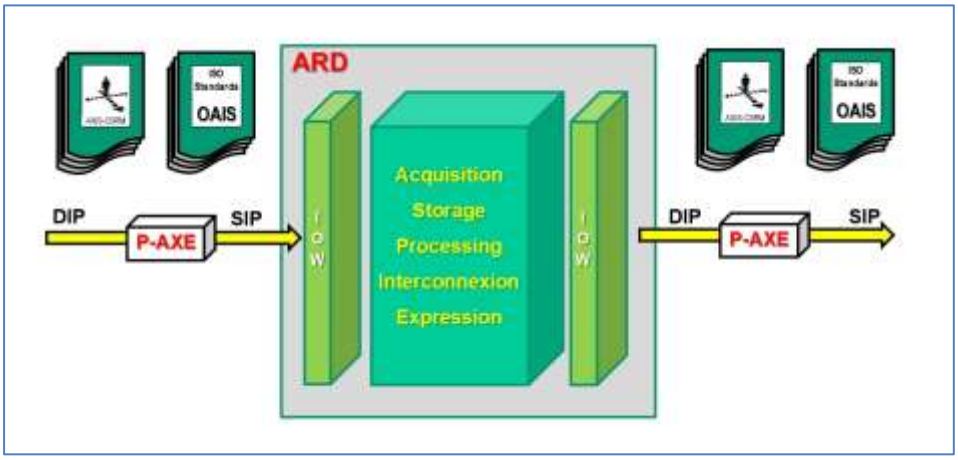


Figure 6-1: the description of an ARD Autonomous Resource Domain)

An ARD (Autonomous Resource Domain) is an IT system (one or many computers) in charge of managing the acquisition, storage, processing and exchange with other systems.

Choosing to name the ARD systems "ALICE" and "BOB" respectively copies the method used in cryptography to designate two agents that interact (whether they are machines or individuals). These agents need to be protected against external or internal attacks and ensure the integrity of data exchanged. The aim is to achieve exchange of data in a fully secure manner.

Each ARD is, of course, completely free to adopt an internal data model of its own, optimized to meet its specific requirements and give specific advantages (differentiating factors) compared to its competitors. Meanwhile, advances in computer technology and increases in the file sizes demanded by digital media content items (gigabytes for audio and terabytes for films in 4K) militate in favour of adopting internal to the ARD data models functional organization guided by the same principles as those of AXIS-CSRM. In this case, the development of IOWs will be very simple; while IOW development will be more complex if the internal models are too specialized or comprise ‘’proprietary’’ elements handicapping the capability for open semantic export.

To achieve these three aspects of the implementation, AXIS-CSRM incorporates into the ARD an import/export processing module, designated by the acronym **IOW** [**I**nter**O**perability **W**indowing]. Its role is to render the technologies and internal representations of data within the ARD independent with respect to the representation of P-AXE, to assure human expression and the call to "Configuration Management Services."

The implementation of the AXIS-CSRM approach is intended to:

1. Represent the data exchanged between machines using the concepts and technology of semantic data, exchanges being encapsulated in ‘Autonomous Packages' (where models and formats are explicitly represented in the package together with instances of these models and formats). These packages are designated by the acronym **P-AXE** [Packaged - Autonomous eXchange Entity] in the terminology of the OAIS and peripheral standards. These P-AXEs being a special implementation of SIP [Submission Information Package] and DIP [Dissemination Information package] where PDI [Package Descriptive Information] is included in the package.

- ISO 14721 OAIS : a reference model for what is required for an archive to provide long-term preservation of digital information
- ISO 16363 Audit and certification of trustworthy digital repositories – sets out comprehensive metrics for what an archive must do, based on OAIS)
- ISO 16919 Draft (2014) : Requirements for bodies providing audit and certification of candidate trustworthy digital repositories – specifies the competencies and requirements on auditing bodies)

The Machine to Machine interaction (interconnection) concerns the exchange of data:

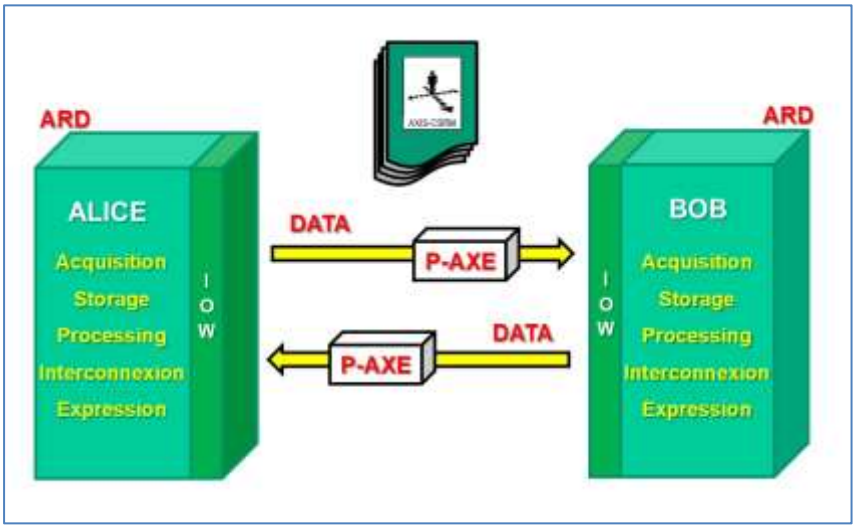


Figure 6-2: Exchanges between Autonomous Resources Domains

2. Present the human interaction or user interface in an understandable manner which is appropriate to the technologies in use; this approach is frequently called "360° Publishing", which reflects the concept of a human being henceforth surrounded by a set of tools with which he or she interacts.

In this context, Human-Machine interaction [usually called Graphical User Interface (GUI)] concerns the exchange of manifestations (perceptions in evidence):

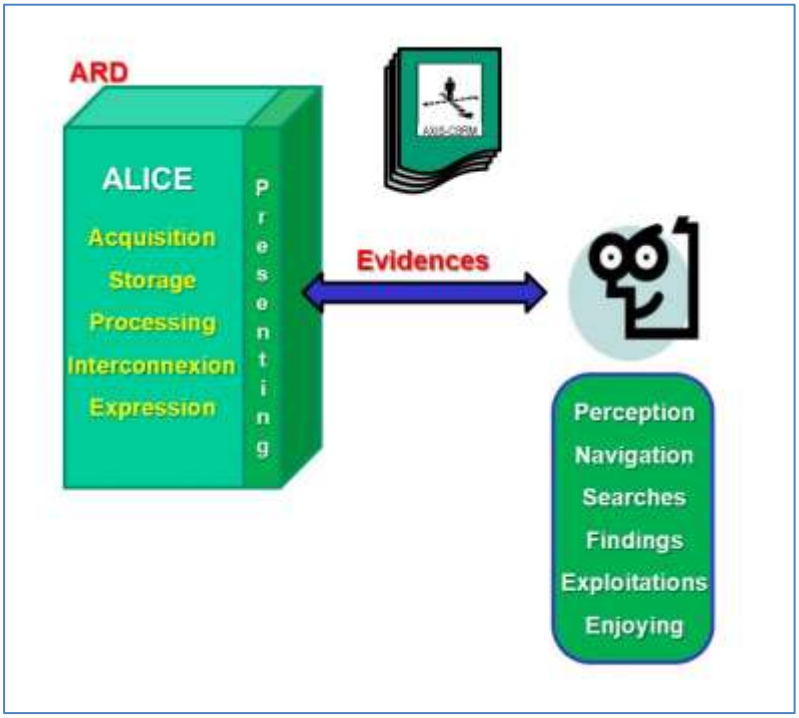


Figure 6-3: Interaction in evidence between a ARD and a citizen

By combining the two previous figures, one's obtains the general view of the approach:

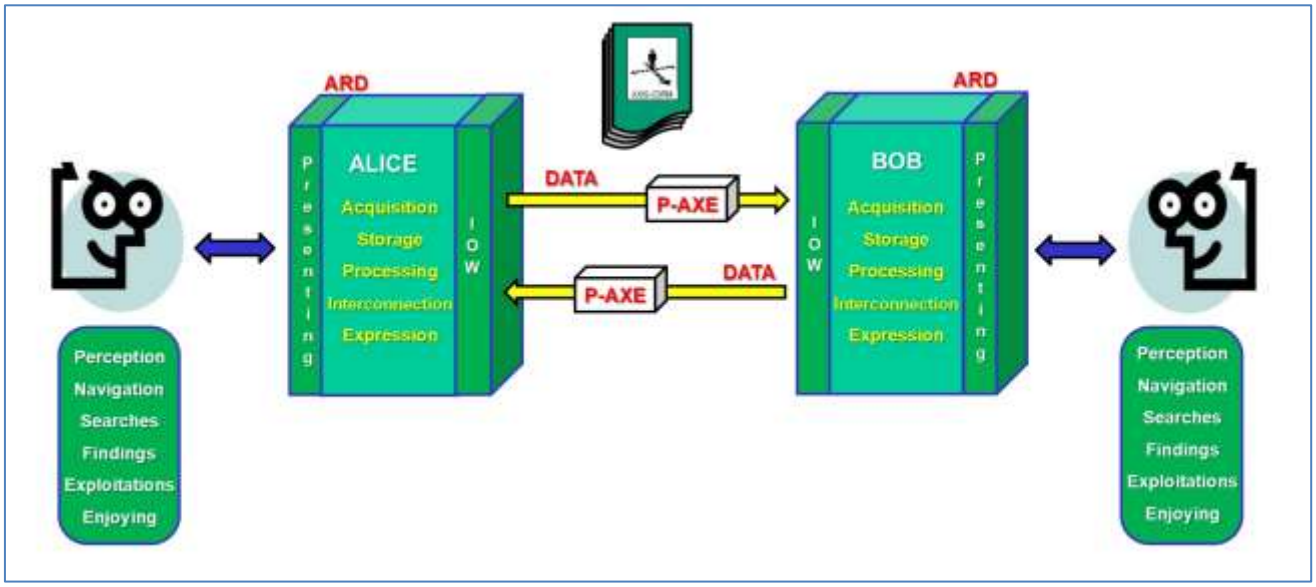


Figure 6-4: Exchanges between an Autonomous Resource Domain and a citizen

In order to ensure access by any citizen to any computer system, it is desirable to have common, neutral services to facilitate interaction and reduce conflicts. In human societies, for example, this is about legislative, executive and judicial services. For the purpose of technical interoperability, only a few specific technical and neutral services prove desirable

and in many cases may be considered implicit. They are called the "Configuration Management Services" which are necessary to ensure scaleability of the infrastructure.

3. Arrange the requirements for Configuration Management Services following protocols which are simple, configurable and capable of being normalized.

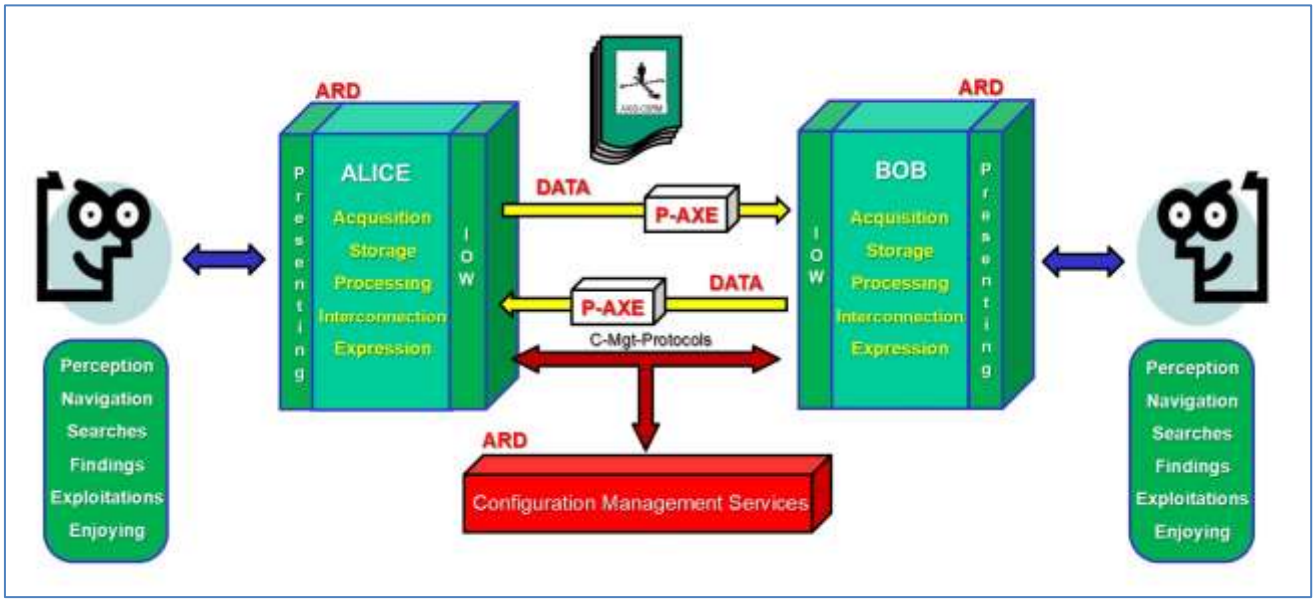


Figure 6-5: Interoperability through the « Configuration Management Services »

The system owes its coherence to the Configuration Management Services which integrate different hardware configurations in the form of files and data, the technical description and its various components, and all the changes made over the course of time. This involves the requirement to store and track the different versions or revisions of any information used by a system (hardware, software, documents, data, etc.) with the assistance of software version control, either proprietary or free. The internet offers a Configuration Management System which provides citizens with semantic navigation tools using open interfaces to access knowledge bases.

AXIS-CSRM is expressed in this document by the following figure:

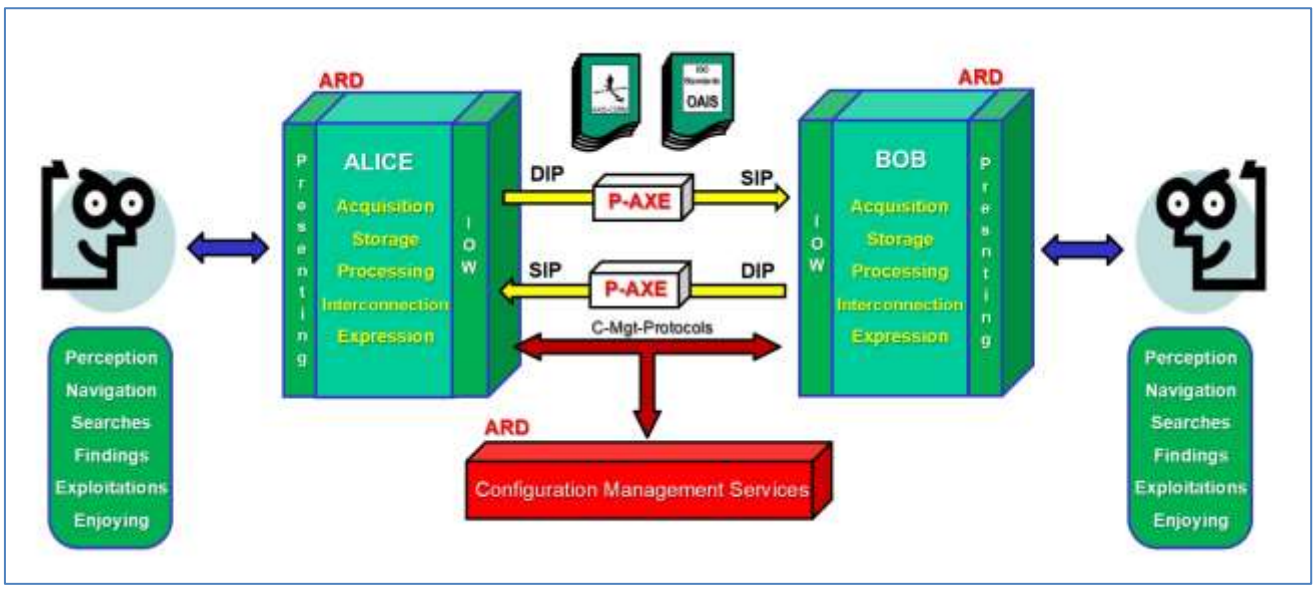


Figure 6-6: Global Interoperability in AXIS-CSRM/OAIS

By considering that "**Presenting**" and "**I**nter**O**perability **W**indowing" [**IOW**] are specializations of the general concept of "**Manifesting**", the previous figure can also be presented as follows:

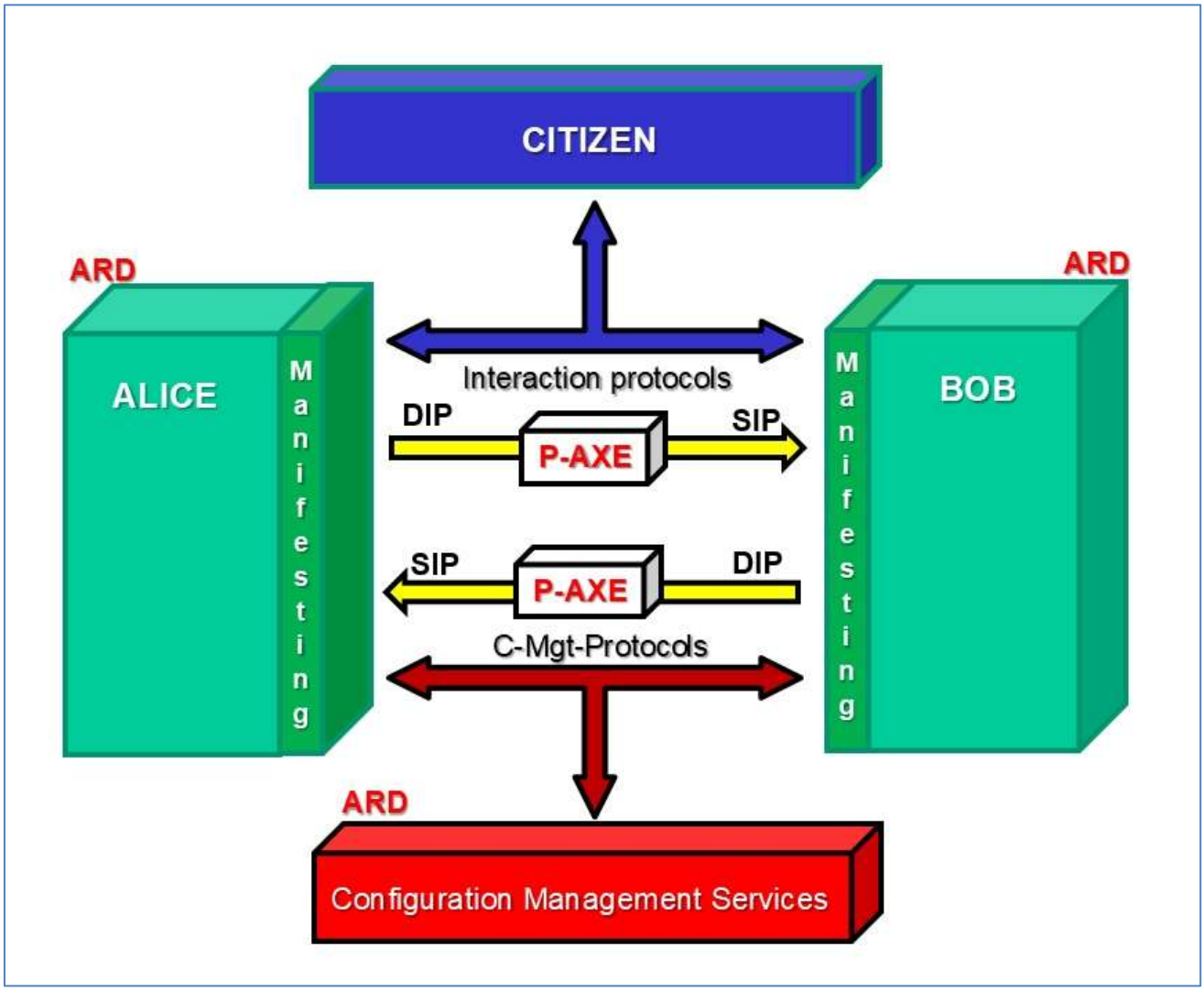


Figure 6-7: Global Interoperability in AXIS-CSRM/OAIS (Compact view)

6.2 The IDENTIFICATION & DENOMINATION Model (Recording & Representing & Locating):

Identifiers are names (terms, formal representations) chosen with the intention to be unambiguous, unique in the current context. Universal identifiers are names built to be unique in the context of their rules of construction.

In Belgium, the official identification of human beings is based on the date of birth followed by a sequence number and a validation code, all coded according to a very strict syntax example: 37.12.25-317.85 [Guy Maréchal is the 317th registered in the Register of Belgium, born December 25, year 1937; 85 is the validation code based on a cryptographic calculation]. It is significant that this example involves a Registry and Cryptography. Both concepts are essential in the development of 'good' systems for identifying and naming.

Identifications and denominations as such can't locate what they designate (objects, concepts, representations ...) but these locators can be identified too as objects themselves.

In AXIS-CSRM in every 'thing' is given a name and an identifier: both are two distinct properties of that thing. The name of a thing is rather intended to be read by humans and the identifier is rather intended for computer usage.

Identifiers are an integral part of the audiovisual universe. The unique description provided by an identifier is extremely important for a wide range of areas such as rights management. The identifier determines with a high degree of reliability that a video object is indeed unique. IDs issued by third parties (TTP) are also important resources for the statement of existence of a piece of content.

6.2.1 URI (Universal Resource Identifier)[10]

There are many mechanisms for identifying objects, some of which have been standardized as those of the "web". Basically in the web, the objects are not named, they are simply localized! This is why the concept of identification known as the URI name (Universal Resource Identifier) covers two separate realities: the naming of the URN (Universal Resource Name) object and the location of the URL (Universal Resource Locator) object.

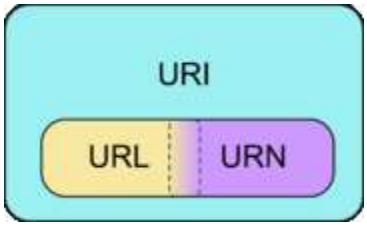


Figure 6-8: URI covers two realities URL & URN

For object identification made in a decentralized manner, a hierarchical system has been developed to identify the domains (and sub-domains) and objects in a particular field. To ensure separate identification for each object, each area has a fountain ID and a registry of distributed identifiers. Such services are entitled UIS "Unique Identification Services." UIS can serve multiple domains or sub-domains.

6.2.2 - Configuration Management Services

To be able to refer to an object independently by its URN or URL, it's necessary to use services translating an URN to an URL or vice versa! Such services are entitled "Resolving Services'' that can serve multiple domains or sub-domains. Often UIS and RS are grouped with other services under the name of "Configuration Management Services." They are the subject of a detailed definition at section 6.8.

[10] A more detailed introduction is available http://en.wikipedia.org/wiki/Uniform_resource_identifier.

Each set has appointed an URN. It can refer to an entire domain. A subdomain of the domain can be named by an URN, which is usually built starting from the domain name. The set of all services and protocols, systems associated with this URN is known as URN namespaces.

Writing these documents and standards is going on while disparate systems already existed. The current approach underlying these documents suggests the presence of a single coherent system, modelled on the consistent approach of the Web, will rise gradually, and that existing heterogeneous systems will gradually disappear[11].

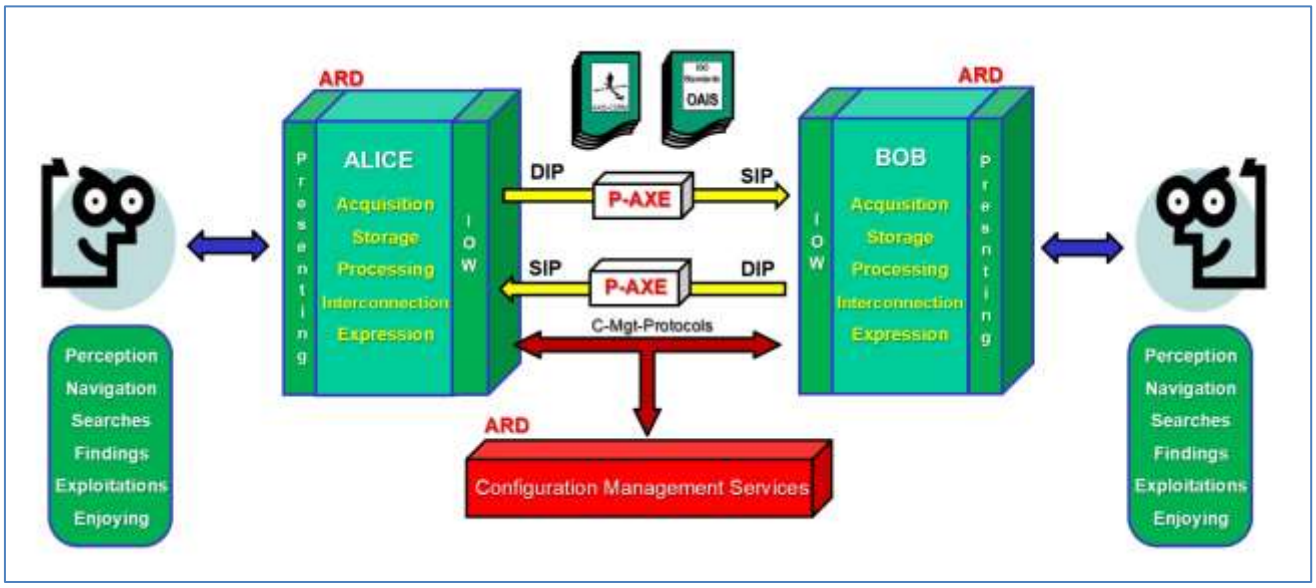


Figure 6-9: Global Interoperability in AXIS-CSRM/OAIS

The concept of ARD [Autonomous Resource Domain] was introduced in the previous section.

In AXIS-CSRM, the basic principle is that all relevant identification systems for an object can coexist. In each of the AXIS-CSRM fields, an object is identified with an ID which is specific to the domain.

When two domains exchange objects, in AXIS-CSRM, these objects are encapsulated (wrapped) in a structured representation called P-AXE (**P**ackaged **A**utonomous e**X**changed **E**ntities). In the illustration above, when Bob receives in its Interoperability Window (IOW) a P-AXE from Alice, all objects provided will be issued with a new specific and unique ID in the Bob domain. All pre-existing objects will be preserved, especially those that were in effect when the object was under Alice's management. This is so that Alice can continue to manage the object in its domain and Bob can independently evolve the file Alice sent to him. If Alice had received a copy of this object via a third party, it would have preserved all its original identifiers (including ISBN)[12].

In AXIS-CSRM, the ID concept is quite general. The data structure is based on identifiers, objects or identifiers of object models, and all the details of these models.

To be able to deal with the problem correctly, it's necessary to give a single identifier on a global level both to the objects whose existence one declares (data objects or physical objects), and to software objects which be used to model the objects whose existence has been declared.

It's necessary to differentiate the statement of existence (name and identification of an object which one claims to exist) from the representation(s) of this object which have their own names, identifications and historical changes.

[11] The obvious example is the identification of books for which the root domain is ISBN (International Standard Book Number) and has sub-domains sometimes quite inconsistent (e.g. Benelux!).
[12] This point will be clarified in the "data model". The traceability of exchanges can be provided by a service of Configuration Management.

The first criterion is to give an ID to an object (software or hardware) regardless of its current manifestation. It is necessary that the computer has the capacity to represent by a string of bits, or characters, an identifier that is associated with an object or model of that object. The objective is by no means to locate through this identifier the location of the object nor the computer resource where it resides. Nor is the identifier related to time, nor a special technology, nor to a carrier; it is simply a single identifier called URI in the web domain (URN or URL). It is essential that there exist a name independent of time, space and way in which it is materialized.

At a conceptual level, the model of representation, identification and naming will dictate the choice of a system that will ensure that the object identifier is unique, independent of where the object is located.

The UUID (Universal Unique Identifier[13]) technology was cleverly exploited by the German system (MuseumID) to keep the link with the identifier of the museum object, the bar-code printed on the existing object. This allows a large computer exploitation in the producing of the UUID, and ensures that the user can find the identifier in the museum by an automatic Resolving. A change of the identification system of the object (in a sale), may cause a change in the barcode but the UUID could remain the same. In these cases, however, despite the loss of traceability, in practice most often the UUID is also changed when the barcode is changed according to the new local rules.

The system proposed by MuseumID thus offers an ingenious and satisfactory solution to identify worldwide a museum object uniquely, using as an identifier the seed of the object owned by the museum. This identifier is such that it can be "resolved" to find the seed identification such that it can serve as a basis to generate a URL pointing to the documentation of this object. However, this system does not solve the three classic 'Resolving' cases, i.e. find where the museum object is (in which city, building, room ...), the restructuration of websites and the traceability of the museum object and its website when available, transferred or sold to another institution or person.

1. For the first case, it is enough to imagine that the Web page indicated by the URL mentioned where the work is located; Experience shows that the heads of institutions often fail to update the page or even choose not to update to avoid having to report that the work is out of the exhibition, loaned, or being restored; which could discourage some potential visitors.
2. The PURL system was designed and implemented to solve the second case (see below).
3. The third case is twice unresolved : the museum object will be re-identified in its new environment (so the link will be lost) and a new web page will be issued (with a new URL) for organized documentation in the new context.

AXIS-CSRM proposes a generalization of the naming and location of **Things**, their **Representations** as well as diarized links[14] between them. The intention is that the approach proposed is open, flexible, complete and scalable. The generalization consists in covering the expression of links (potentially changing) between all forms of materialized things and all forms of locations.

Note that in AXIS-CSRM the expression of the URL, associated with an URN (URL's associated with a URN) is always expressed in a separate document, specifically and solely for carrying the link between URN and URL (links between URN and URL's). The particular case to give the URL of the "Resolving Service" will be addressed in the data model (Section 6.5).

At the conceptual level, the exploitation of the URN specifications is essential (although this concept has been created to represent a resource in data processing within the web). The approach distinguishes strictly the identifier of an object in the conceptual sense (naming the object) of the way in which this object will eventually be physically located by a URL (http address on the web, or as a file on a CD-ROM, or as a GPS coordinate). It's a Resolving Services which will locate the resource physically, that will give the access or location.

What is important is to ensure the uniqueness of these URNs and this is done most often via a central system generating all the URN. But that central system could simply generate a root identifier, what allows the decentralization of the local generation of the individual URNs. For example, an ISAN number which is complemented by a string of characters

[13] Wikipedia: These unique identifiers are 128-bit encrypted and are produced using pseudo-random and the characteristics of a computer (hard disk number, MAC address, etc.) components. A UUID is usually in this form: E29B-11D4-110E8400-A716-446655440000)

[14] Example : the process of partitioning an audio stream into homogeneous segments according to the speaker identity.

linked to a specific country, organization or version ... This example implies a function of Configuration Management within the system to manage the URN.

There are systems that don't need a configuration management service (such as an URN fountain) to generate unique URN. This is the case of the UUID system! They are unique with a probability of "femtoscopic"[15] collusion. They are based on cryptographic mechanisms that generate sequences.

It's possible to generate the URN by a concatenation of the name of the ARD with an UUID naming the object (resource) in a mixed encapsulation mechanism. For example, when creating an URN, we write the domain identifier (eg RTBF) followed by ''UUID generated''. ''RTBF-UUID'' indicates a spontaneous RTBF origin. This example assumes that the RTBF ARD name is unique and universal, which is not obvious (eg. Radio Télévision Burkina Faso).

It is at the PROFILE level that should be reported how the URN are encoded, according to which rule, for example the RTBF domain followed by the UUID generated based on standards. This rule also provides that the same object can be indexed with several different identifications.

For a variety of applications, it is necessary to know the identification (e.g. the ISBN of a book). When treated by a user, it is perfectly possible that he uses a generated identifier (example RTBF-UUID), more pragmatic to locate, archive, identify all the translations, all editions, each with a different ISBN.

The conceptual model for identifying and naming distinguishes object and object model or parts of an object model through an URI system because it will one day be able to designate the physical materialization of these objects.

It is therefore necessary to have a mechanism that from the URN recovers the materialization of the external object (a geographical location ...) or internal object (a computer resource: bits coded on a DVD). This involves an identifying of the encapsulated object, or of the becoming physically accessible object, of the object becoming interoperable in the sense that it has a physical existence. Its identifier having characteristics such that one can access to his or their materialization(s).

This is the purpose of the Resolving Services of the Configuration Management. At each occurrence of a data object or from each statement of the existence of a physical object, the Configuration Management Services store information about the location of the physical object represented by this URI. The Configuration Management Service must be able to relay this information, or request access to one or more copies of such data objects.

There are also other types of Resolving Services to be able to answer indirect questions about the location of an object model. This is another kind of Resolving because it involves not only to find the object but also to exploit the URI associated with a model of the object.

The Resolving Services therefore carry not only the ability to transform an identifier in a location but can solve extremely complex things like finding the modelling(s) of an object. Or vice versa, having the model and finding the modelled object, or the evolution, the chaining between the models.

The identification model is therefore essential to provide a system for identifying essential to good organization properties.

AXIS-CSRM defines a way of structuring such that the properties of the organization are met: a separate URN, disjoint from the URL or the activation service that makes it possible to answer questions from Resolving Services.

The approach would be to construct the URN as a sequence of three fields:

15 The term « femtoscopic » refers to a value so small that the impact could be managed as negligible in the designated context. In such systems, either the collision is detected and surrounded *ab initio*; either the impact will be limited before detection and corrected as soon as detected. The term is derived from the mathematical and physical prefix "femto" meaning 10-15 (one part in one million of billions).

1. The name of the ARD would be built based on the domain name as used on the Web. In case one wants to identify a domain department, the name of this resource would be concatenated with the domain. The advantage of this approach is that it guarantees uniqueness without having to create a new Configuration Management Service. This part can also be used as a designation of 'namespaces' stakeholders in XML, OWL, RDF and similar expressions.

   For syntax, see: http://tools.ietf.org/rfc/rfc2141.txt

   Example: the name of the ARD may itself consist of two parts:

   - a web domain (by definition only). For example: **www.titan.be**
   - From a hierarchical chain of acronyms expressing the structure in the domain. It is not a subdomain in the Web sense: it is simply to be able to name distinctly different ARD domain associated with the unique domain on Web.
   - The concatenator character between web domain name and the various acronyms would, for example, the simple dash (hyphen) "-"

   Example of ARD name: **www.titan.be-rd-s5**

2. A concatenator character: in the example, we suggest using:
   a. the double dot ":" for URN modelling object and
   b. the underscore "_" for URN modelled objects.

3. The internal name of the object in the domain: it is built by using the UUID where uniqueness is guaranteed by a cryptographic standard approach that does not require external Configuration Management Service approach. The identifier has a fixed size. It can be presented according to various modes; the most common is a string of hexadecimal characters, which makes it readable by humans.

   Example : **16fd2706-8baf-433b-82eb-8c7fada847da**

   See : http://fr.wikipedia.org/wiki/Universal_Unique_Identifier

Following syntax and choice of example, the URN would be:

www.titan.be-rd-s5:16fd2706-8baf-433b-82eb-8c7fada847da

The concatenator is the double point character, it is the "modelling" object of an URN.

It is indeed a URN rather than a URL!

http://museumid.net/documentation-20101206 .

In practice, within an ARD, objects are identified by their own UUID preceded by concatenator distinguishing modelled and modelling. In exports, the objects inside the P-AXE are appointed by the concatenated full name. Objectively, the UUID is sufficient to identify the object (resource). However, for many reasons, the mechanism of autonomy naming ARD and naming of objects (resources) in an ARD has many advantages. A very detailed analysis is made in reference MuseumID (supra): http://museumid.net/documentation-20101206.

In another example, a user enters the "dbpedia.org/resource/Bruxelles" address for pages searches focused on the city of Brussels! The address entered is a formal representation of the concept of a resource, but it's not resolvable!

"Dbpedia.org/resource/Bruxelles" is the HTTP URI that identifies the concept of Brussels in the world of DBpedia. By inserting this URI in a browser, the Resolving Service will give here with a 303 redirection, the description of the resource in DBpedia, a description which is located by URL "dbpedia.org/page/Brussels".
Moreover, in each ARD, there are different representations of the same concept! The city of Brussels is defined differently in DBpedia or freebase ... the linking words (manually or by inference) as "sameAs" globalizing Entities from the two knowledge bases.

PURL : Persistent Uniform Resource Locator

A persistent uniform resource locator (PURL) is a uniform resource locator (URL) (i.e., location-based uniform resource identifier or URI) that is used to redirect to the location of the requested web resource. PURLs redirect HTTP clients using HTTP status codes. PURLs are used for controlling the URL resolution process, thus solving the problem of transitory URIs in location-based URI schemes like HTTP.

PURLs implement one form of persistent identifiers for virtual resources. Other persistent identifier schemes include Digital Object Identifiers (DOIs), Life Sciences Identifiers (LSIDs) and INFO URIs. All persistent identification schemes provide unique identifiers for (possibly changing) virtual resources, but not all schemes provide curation opportunities. The curation of virtual resources has been defined as, "the active involvement of information professionals in the management, including the preservation, of digital data for future use".

PURLs have been criticized for their need to resolve a URL, thus tying a PURL to a network location. Network locations have several vulnerabilities, such as Domain Name System registrations and host dependencies. A failure to resolve a PURL could lead to an ambiguous state: It would not be clear whether the PURL failed to resolve because a network failure prevented it or because it did not exist.

Wikipédia : http://en.wikipedia.org/wiki/Persistent_uniform_resource_locator

ISO standard (ISO 26324) for the DOI system
DOI registration agencies: http://www.doi.org/registration_agencies.html
DOI Handbook: http://www.doi.org/hb.html

6.3 The INTEROPERABILITY model (Temporal, Systemic, Historical, Cultural):

AXIS-CSRM distinguishes two kinds of interoperability's:

1. Technical interoperability covers properties that assure humans that machines can communicate reliably between domains.
2. The functional interoperability covering the properties which, ultimately, **ensures to humans** that they will be able of enjoying the artefacts via interoperability's in the four axes***: temporal, systemic, historical and cultural interoperability.***

6.3.1 Technical Interoperability

In AXIS-CSRM, the technical communication interoperability is presupposed set by appropriate approaches, preferably compliant with an ISO-OSI (**O**pen **S**ystems **I**nterconnection) model[16]. The technical interoperability information model of AXIS-CSRM is based on a similar methodology but whose purpose is not communication but the data representation of information. In the previous chapter it was pointed out how the information concept is linked to a human being based on the ''signified'', while the data (signifier) is the term by which the techniques try "less awkwardly possible" to represent information.

The AXIS-CSRM interoperability model is illustrated by the figure below. It has the same mechanisms as those of the OSI, namely the independence of the layers, interfaces between layers, between systems and protocols within layers.

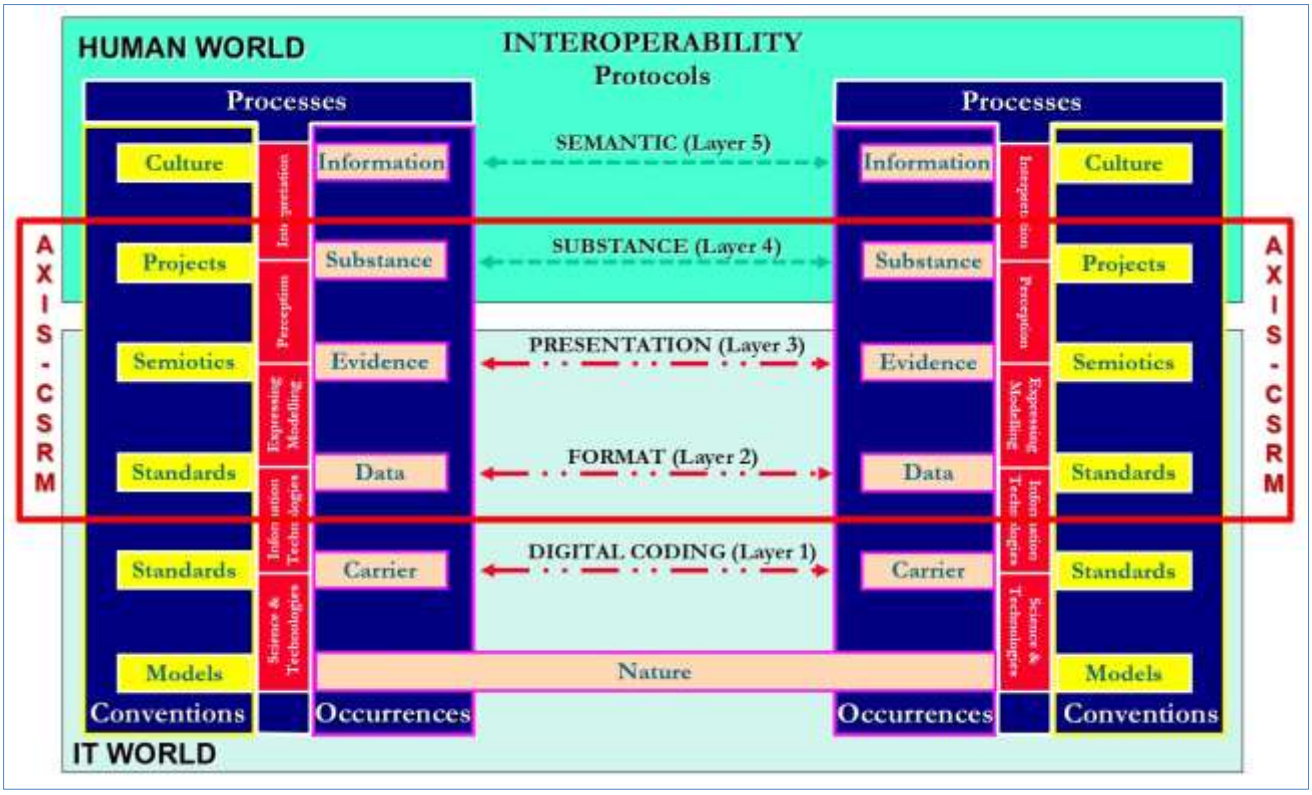


Figure 6-10: The different layers and protocols for AXIS-CSRM interoperability

The AXIS-CSRM interoperability model distinguishes 6 layers divided into Conventions and Occurrences.
The left-hand column makes a distinction between Conventions and Occurrences; for each expression of an Occurrence, it is necessary to specify the standard which allows it to be interpreted. A piece of content cannot be interpreted except by using the standard or norm which describes its format.

Both the upper (Culture) and lower parts (Nature and Digital Coding) are outside the scope of the model. The upper layer is cultural and concerns among other things natural languages. The lower layer Nature is defined by models conceived at

[16] See : https://en.wikipedia.org/wiki/OSI_model

a scientific level. And to provide a data carrier (carrier), there are many available norms/standards, for example to format hard drive or to use for optical encoding.

Interoperability between systems involves exchange protocols that are defined at level 1 (Digital coding), 2 (Format), 3 (Presentation), 4 (Substance) and 5 (Semantic).

The figure 6.7 describes six independent layers showing interoperability between machines and between humans.

"Layer 0 – NATURE'', the use of properties of the material world to serve as porters signs elementary codes of representation of information. For example, the magnetic orientation of an area of a surface can be chosen to represent a "1" or "0". Similarly, reflection of an optical area of a surface; black or white areas on a sheet of paper or on the screen of a PDA (bar-code, QR-code).
‘’Layer 1 – CODING’’ concerns the organization of a set of basic symbols to become a carrier of coherent data sets. It is this structure that is called a "Volume" so that it can be provided with a data set.

For data, "File" and "Folder" are an establish standard.

For the media data carrier aspect, the volume is essentially limited. On a magnetic disk, it’s necessary to define the formatting type (eg NTFS); block size; the volume capacity (e.g. 2 terabytes.)

See example of CD audio coding in note [17]

‘’Layer 2 – FORMAT’’concerns the organization of a set of symbols to model information through data representations. In many cases, the format specification is based on parameters that can be set. In the example JPEG format, suitable for encoding still pictures, the application can choose between representation in degrees of gray or colour, and the number of bits per pixel.

Here are the usual formats: [http://fr.wikipedia.org/wiki/Type_MIME].

Note: How:

Some formats merge the layers 1 and 2. The most significant example is the CD audio that uses optical reflection. It uses the same volume as the CD-R but without the coding organized into ‘’folders’’ and ‘’files’’. However, in many other cases this fusion of layers 1 and 2 establishes opacity of the format definition. These cases, known as "proprietary formats", give a competitive advantage to some industrials. When using such formats, there is no real open interoperability, which is hidden behind Non-Disclosure-Agreements (see Chapters 2 and 3).

In AXIS-CSRM, the merge of layers 1 and 2 (and sometimes 3) are tolerated. In this case, a consistent set of formats representation ensures the openness and independence of the three layers (0-2). Proprietary formats are linked to open core.

This approach accepts the fact that the master copy of a video can be encoded in a proprietary format. It is necessary in this case to organize an open and interoperable modelling while ensuring that interoperability with proprietary formats can be assured (by external NDA agreements or acquisitions of proprietary hardware to exploit these formats.)

‘’Layer 3 – PRESENTATION’’ is dedicated to humans and is essential for understanding the AXIS-CSRM model. This is where the computer will present an internal data format on screens via speakers or 3D printing data that has converted presentable objects viewable, audible.

Concrete layer 3 is what is expressed, how the data will be presented on a screen. The problem of presentation is fairly general: the presentation on a PDA (Personal Digital Assistant) is different from a TV or computer; it can handle more

17 The coding is directly designed for the reading reliability is maximized to offset the risk of scratches, dust, offsets and warping of the optical disc. A significant redundancy is incorporated in the standard CD-A so that the correction of errors in reading can be performed in the hardware. Coding CD-A has a specific system of tags & chunks (other than folders and files) to embed metadata and data service. The example of CD-A justifies the merging of layers 1 and 2, especially since the format specification is published and the cost of the licence to use is marginal.

than one object as a second screen connected to the TV. The sound can be declined into mono, stereo, with different sound qualities. Layer 3 provides overall consistency.

To ensure sufficient orthogonality between the FORMAT (layer 2) and the PRESENTATION (layer 3), it requires that the definition of the objects presented can be 'mapped', coupled with a specific presentation (printing a document on A4 is not the same as on an A0).

Interoperability requires that the internal coding performed at Layer 2 is such that it can be adapted to any type of external presentation system. Or vice versa for a capture at the presentation level, so that it is possible to render it independently of the layer in question.

‘’**Layer 4 – SUBSTANCE’’** concerns quality requirements for a presentation (layer 3) in line with an intention, it is a PROJECT in the CONVENTION part.

For a given SUBSTANCE, as part of a project, PRESENTATION must provide a sufficient, adequate or equivalent quality. In the case of a MPEG2 MPEG4 transcoding, the project may be a bit lower rate, a quality increased or a master/proxy synchronization!

With the same format, it is possible via the substance to offer different qualities of presentation. If for any reason, the internal coding format must be changed, it should be possible to verify that the presentation layer offers a similar substance to the original project, if not better.

So there is a whole ponder on orthogonality, consistency, between a defined project and the way to produce representations through different substances. It is not possible to display on a small screen with low resolution a HD quality. To view this object according to the substance, it is necessary to use a screen adapted to the proposed resolution.

"Layer 5 – SEMANTICS’’. This last level (5) is not in AXIS-CSRM, it does not appear within the highlighted orange area!

As in the previous chapter, the information is not accessible to computers. Information is the interpretation that humans assign to the data presentation; it’s the way that humans use to transmit modelling data to the computer for internal representation.

As by definition, this layer 5 is not accessible through information technology, it is therefore necessary to wrap the semantics into the PRESENTATION layer (3) so as to be able to transfer or preserve the information essence.

In a particular project such as the intention to archive an audiovisual production, the organization defines the archival quality (substance) and the way IT will present information so that they can be exported, wrapped or encapsulated.

It is the culture that dictates what we must keep track, the way to archive. The computing machine does not know the Why? The intention is implemented via a mechanism of data formalization.

In technologies available on the market, the three-layer "data", "evidence" and "substance" are merged into proprietary mechanisms. Often layers 1 and 2 are closed and many proprietary applications offer only not editable presentation.

To provide 360° publishing, information must be represented correctly and independently of the presentation layer, for publication on any type of device. The project to represent information quickly on a PDA is fundamentally different from a projection of the same content for a screen of 80 m2 in 4K format.

AXIS-CSRM thus concerns itself with the layers which deal with Projects/Substance, Semiotics/Evidence and Standards/Data and therefore exchange protocols for layers 2, 3 and 4.

It is at the project level, at human level, concerning the intention with respect to the substance, the uniqueness of this model can be found.

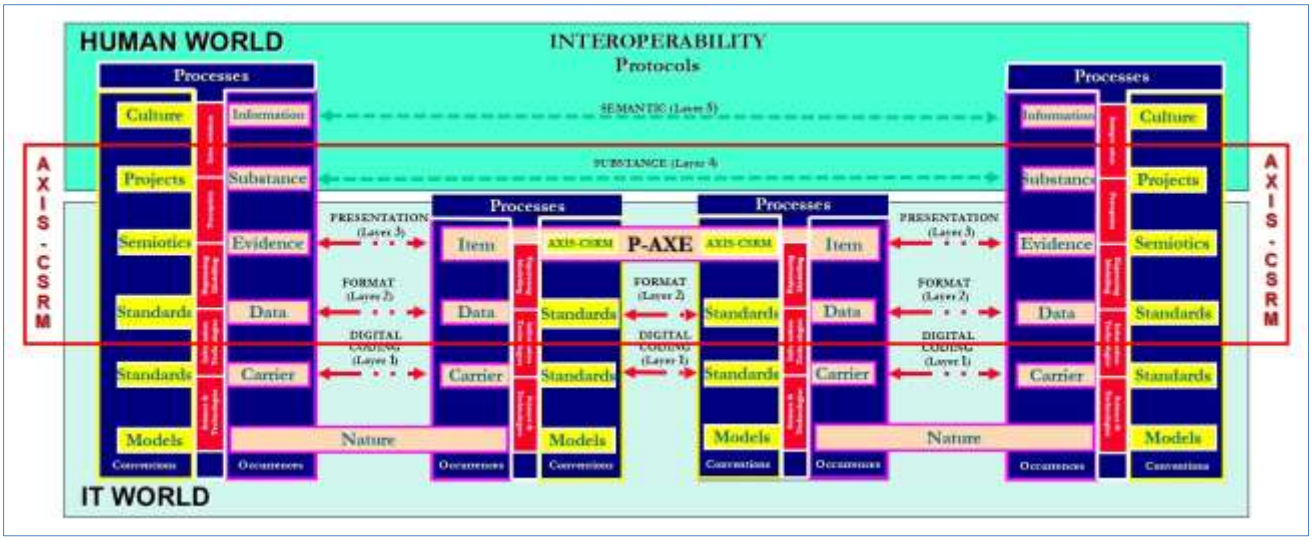


Figure 6-11: Wrapping of Projects/Substance layers

AXIS-CSRM has the ability to encapsulate the Project/Substance layer by modelling the data so that the substance that will be shown is adequate to the project defined by a human.

A major contribution of AXIS-CSRM is to introduce a PROJECT layer into the packaging of the P-AXE which does not exist in most of the other modelling systems. It thus has the ability to represent and deliver substance in the exchanges between systems. If the substance is well designed and the project specified, humans will be able to interpret through cultural means the underlying information delivered by the computer system.

These are the approaches offered by semantic technologies that allow the level of the substance layer (4) to formally represent projects in IT. So humans are able to express intentions in substances that fit their project.

6.3.2 Functional Interoperability

AXIS-CSRM addresses basically four types of **functional interoperability** (Temporal; Systemic; Historical and Cultural) seen when separate systems, called **ARD** [Autonomous Resource Domains]) are exchanging entitled Packages **P-AXE** [Packaged - **A**utonomous e**X**change **E**ntities]. These interchanges are managed and operated under protocols by citizens.

The FRAR model (Functional Requirements for Assets & Rights) that will be introduced at the next paragraph covers the fusion of " functional " and " data " models. It allows generating the **AXE** holders of such interoperability's:

- **Temporal:** the same specific signified can be represented in a scalable manner with changes in information technology and communication. This is particularly true for records management or migration technology in an organization. Temporal interoperability is '**PERSISTENCE**': "*The ability of a technology & organizational system to ensure to the citizen of today that the citizen of tomorrow will be capable of enjoying the current cultural, sociological, economic… assets*" [Abdelaziz Habib (Memory of the World) UNESCO].

- **Systemic:** a work and its context can be transmitted between two ARD without loss of meaning, although ARD use different technologies. When a temporal interoperability is ensured by construction, the systemic interoperability is also ensured: "*The ability of a technology & organizational system to ensure to the citizen of here that the citizen of there will be capable of enjoying the local cultural, sociological, economic… assets*" [Abdelaziz Habib (Memory of the World) UNESCO].

- **History:** from a work in context, the authors may want to change or derive other works, without losing the history. The persistence of the history of things is also to be ensured by construction.

- **Cultural:** a work and its context can be expressed according to various modes depending on the cultural and social contexts, in particular the language. However, traceability also allows presenting the links between the works, their

affiliations and their contexts. The place in the history of a speech (Martin Luther King on the steps of the Lincoln Center in Washington) cannot be understood without linking it to its context, the presentation of the event in which it was delivered, the events associated with it, with articles in newspapers, sociopolitical reactions, listen to this speech as it was delivered, the ability to access its transcription (including Braille) and its translation into many languages. The cultural interoperability implies the integration of the management of the things with the management of the contexts.

6.4 The FUNCTIONAL Model FRAR « Functional Requirements for Assets & Rights »:

6.4.1 Introduction / Introduction (FRBR : Functional Requirements for Bibliographic Records)

The functional model of AXIS-CSRM was built on the top a detailed analysis of FRBR conceptual model [http://www.ifla.org/VII/s13/frbr/frbr.pdf].

FRBR stands for "**F**unctional **R**equirements for **B**ibliographic **R**ecords". It fits perfectly for covering the domain of the books and edition in general. But requests quickly came for extending this limited scope! However, the FRBR standard has been extended but only for better fitting the bibliographic domain and for covering the domain of the Museum Objects CIDOC-CRM. It has been adapted for semantic representation in the conceptual model FRBRoo:

https://www.ifla.org/files/assets/cataloguing/FRBRoo/frbroo_v_2.4.pdf

The AXIS-CSRM / FRAR approach is an attempt to include FRBR into a larger scope but, while remaining semantically compatible. That's the reason why, for the reader's facility sake, the first group of FRBR concepts is reminded here after.

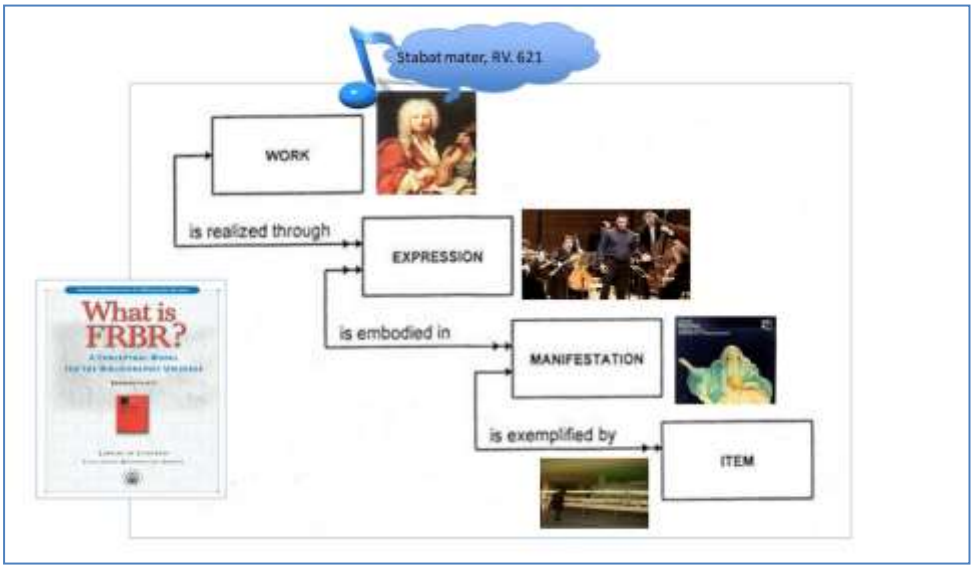


Figure 6-12: The FRBR main schema

The entities in the first group (as depicted in the Figure) represent the different aspects of user interests in the products of intellectual or artistic endeavour. The entities defined as ***work*** (a distinct intellectual or artistic creation) and ***expression*** (the intellectual or artistic realization of a *work*) reflect intellectual or artistic content. The entities defined as ***manifestation***s (the physical embodiment of an *expression* of a *work*) and ***item*** (a single exemplar of a *manifestation*), on the other hand, reflect physical form.

The relationships depicted in the diagram indicate that a work may be realized through one or more than one expression (hence the double arrow on the line that links work to the expression). An expression, on the other hand, is the realization of one and only one work (hence the single arrow on the reverse direction of that line linking expression to work). An expression may be embodied in one or more than one manifestation; likewise a manifestation may embody one or more than one expression.

A manifestation, in turn, may be exemplified by one or more than one item; but an item may exemplify one and only one manifestation. Conversely, a manifestation may be represented by one or more than one document; but a document can never represent one and only one manifestation.

Notice that the official French and English versions of the FRBR standard are not fully compatible! In the English version, an Item is an exemplification of one of the Manifestations of a Work; while in French a Document is a representation of a Manifestation. For FRAR, it is the English version of FRBR which is the reference. It is interesting to read in the text of the standard FRBRoo the discussions and difficulties encountered to perform a semantic representation of FRBR! The rigour of semantic representations comes up against the ambiguities and internal inconsistencies of the initial version of FRBR.

6.4.2 The FRAR conceptual model :

The FRAR conceptual model could be seen as an extension of the FRBR to cover **multi-domain** works where they could be reused for new creations, adapted, derived …. **FRAR** covers **the processes and their chaining involved in the creation and enjoyment of works.** Those chaining governing the progressive and loops between Work and Expression**,** Expression to Manifestation; Exemplification of one to Manifestation); covering the **links** between distinct Works; covering the internal and external **structures** of the Works and covering the **enjoying** of the Works (which is the hook for managing the Rights).

Figure 6-13 : Working and Works in the FRAR model of AXIS-CSRM

The symbolics is expressed in the previous figure, where, in particular, the processes (event [Events]) are represented by blue rectangles.

AXIS-CSRM is basically a conceptual model incorporating a **functional** and a **data** model. At the highest level, the functional model is based on "**Working**" that produces "**Works**". Ultimately the humans are interacting with artefacts such as enjoying them (with a double meaning: having **perception** on the artefact // having **rights** on the artefact).

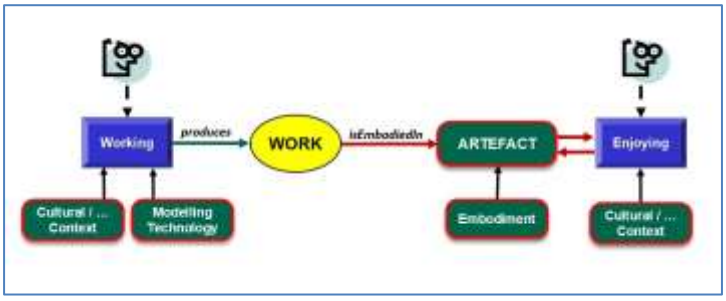


Figure 6-14: The FRAR top level schema

The name of EDITO has been chosen to designate any form of **realization** of a work, even partial and independently of the properties which will be chosen for a materialization in a form intended to be enjoyed.

- One of these forms would be, for example, the internal representation of a computer editor [called plus lion EDIDATA]. In this case, the artefact could be the artwork represented as a file stored on a volume.
- Another form would be an intentional expression (total or partial [script / synopsis / ...]) of the work.

The following three figures represent FRAR at three levels of abstraction[18].

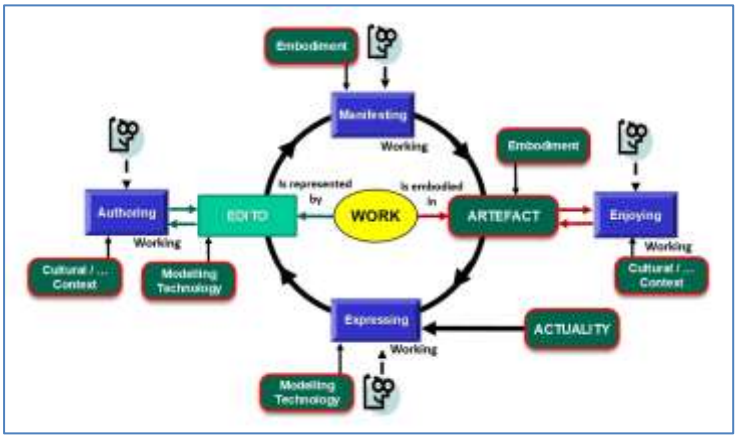


Figure 6-15: The FRAR extended schema Edito - Artefact

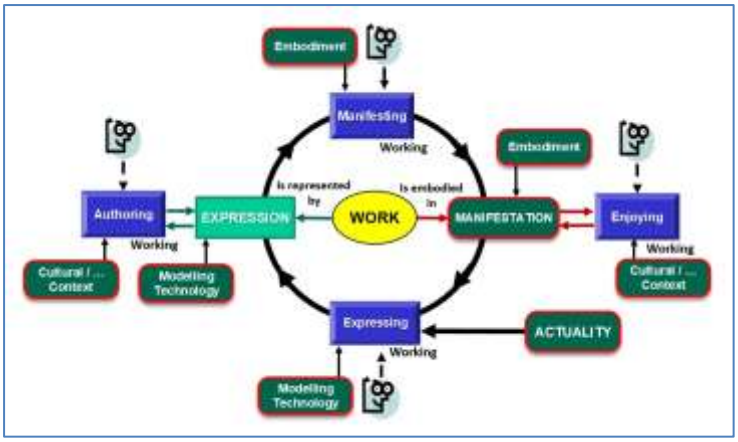


Figure 6-16: The FRAR extended schema Expression - Manifestation

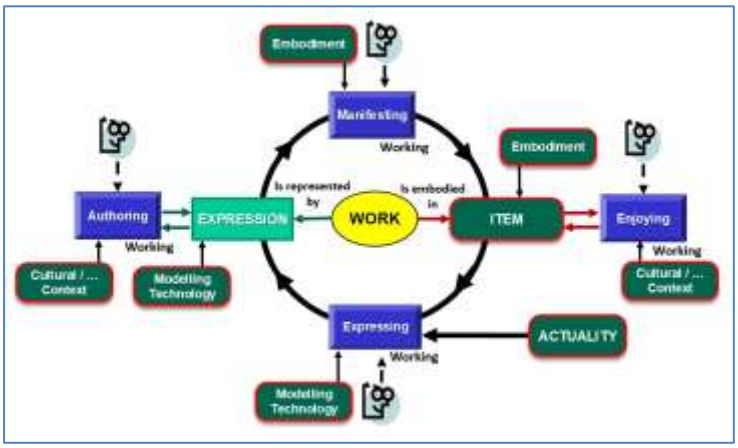


Figure 6-17: The FRAR extended schema Expression - Item

Working is specialized in four processes: **Manifesting** (Realizing a Manifestation or an Item embodying the Work), **Authoring** (Concepting, Constructing, Creating forming the Expression of the Work)**, Expressing** (Modelling an Expression of the Work from previous Works or actualities) and **Enjoying** (having fun and rights on the Work). This synthesis is shown in Figure and known under the name 'Model FRAR [**F**unctional **R**equirements for **A**ssets & **R**ights].

[18] The circular presentation of FRAR comes from a suggestion of Mrs. Isabelle Canno (Radio-France).

The first of the three figures represents the general case: the **Work** is manifested in a embodied way or accessible in perception in the form of an **Artefact**. That is the result of a process called "**Manifesting**". While the author is busy at elaborating his work (**Authoring**), the Work is called "Edito" at the general abstract level. When the Work is constructed namely from existing Works or pre-existing Artefacts (**Actuality**), the reuse process is called "**Expressing**".

The term "**Edito**" is the higher abstraction level of "**Expression**"; similarly, the term "**Artefact**" is the abstraction of "**Manifestation**" becoming "**Item**" at concrete unique level.

The functional model of AXIS-CSRM generalizes and merges the FRBR and OAIS models. In particular, it is covering any types of things (not limited to books) and the whole life cycle of Things (including the involved processes and their contexts).

The functional model introduces the **data model** where editorial representations (called **EDIDATA**) are distinguished from interoperable representations (called **ITEM**).

The **functional model** distinguishes the editorial representations (called **EDITO**) and the interoperable representations (called **ITEM**): when the 'agents' who perform the processes are humans, the 'EDITO' refer to their thoughts, motivations, intentions; While the 'ITEM' means that they express through their words, writings, gestures, interpretations (or improvisation) instrumented (i.e. using instrumental resources such as a piano, pen, paper, marble ...) …

When the 'agents' who perform the processes are data processing machines, the 'EDITO' means internal treatment data (which will be referred to as **EDIDATA** in the data model) and 'ITEM' refers to externalized representations (such content on a screen, printing on paper or in 3D...), typed computer representations (files...) and the means themselves (volumes; processors and means of communication...).

In practice, the Functional model and Data model coexist because processes are mixed.

6.4.3 The structuration of the processes (Agents & Roles) :

The processes could be further detailed namely expressed as chains of sub-processes, performed by Agents (physical Persons or Machines [acting under the control of physical Persons]) performing roles with appropriate skills and following instructions, prescriptions, directives …

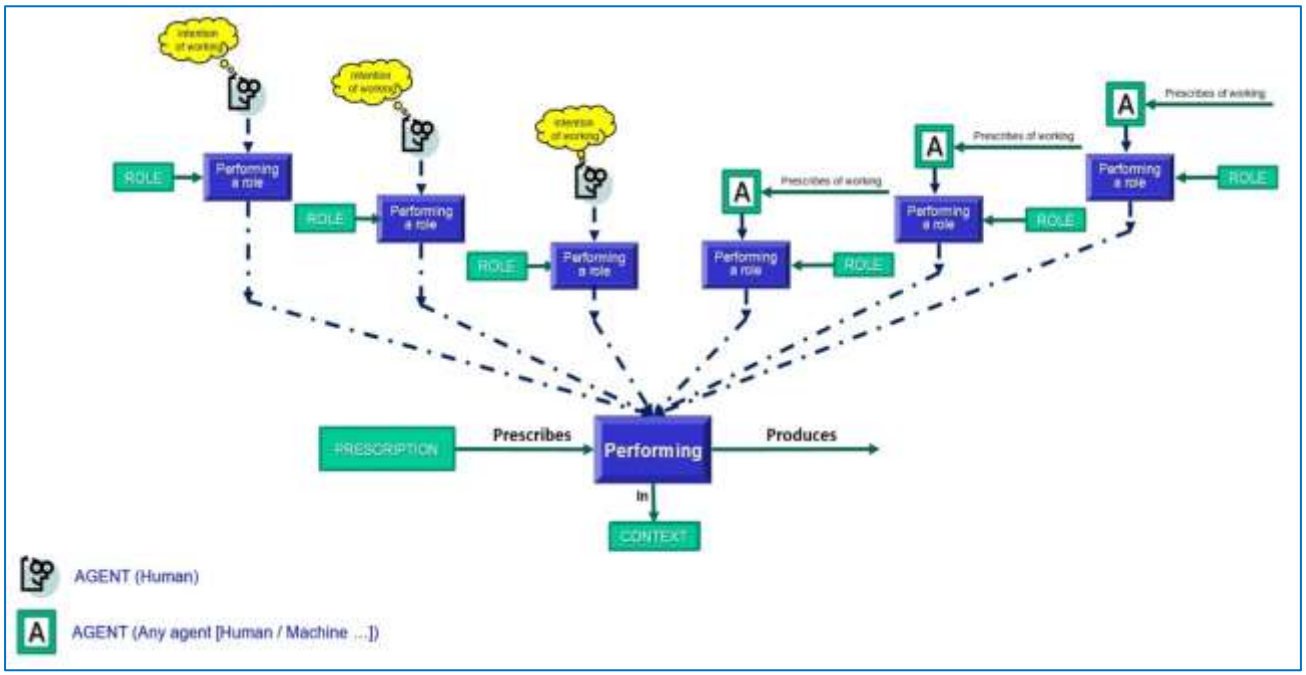


Figure 6-18: Structuring 'Performing'

In this context, FRAR can be illustrated by the following figure:

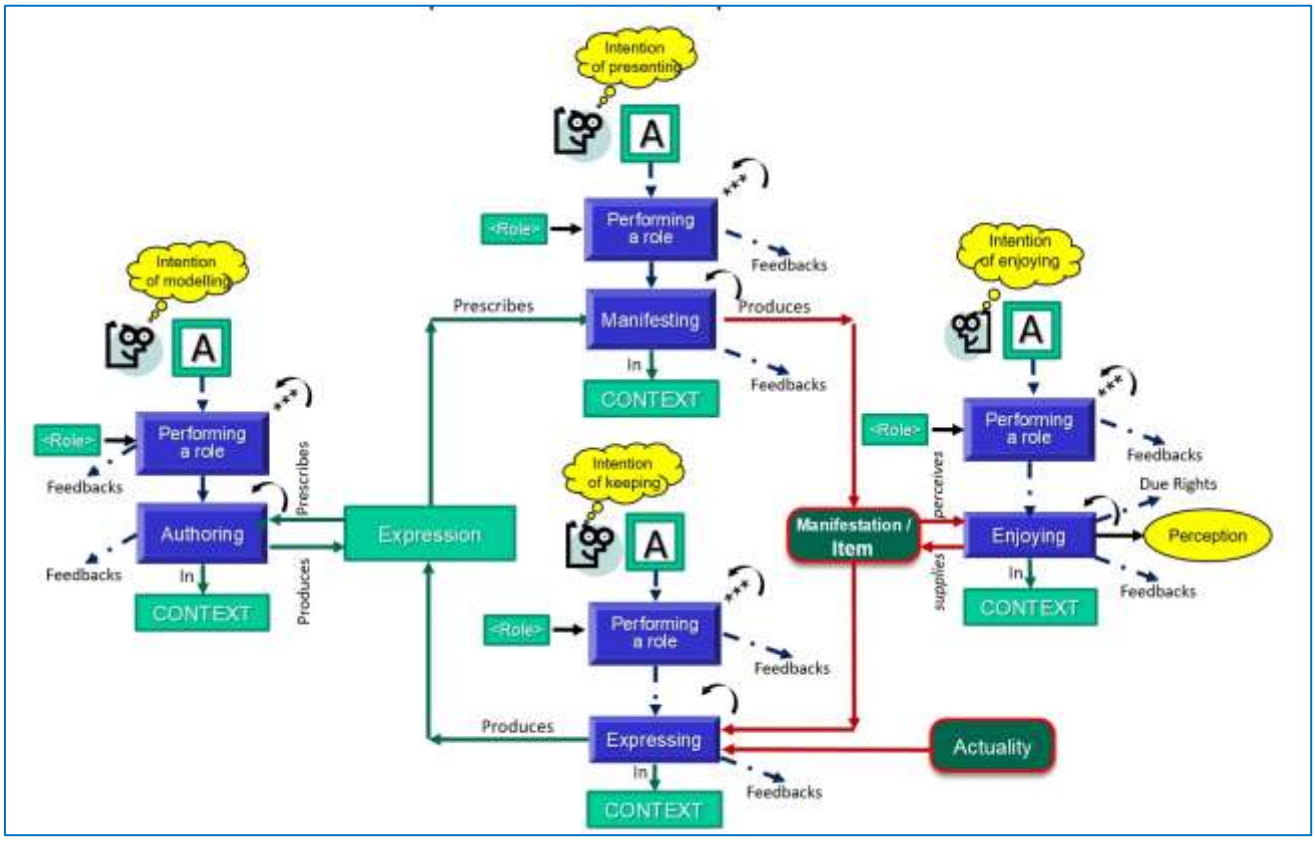


Figure 6-19: Detailed FRAR schema

6.4.4 Incorporating the Rights:

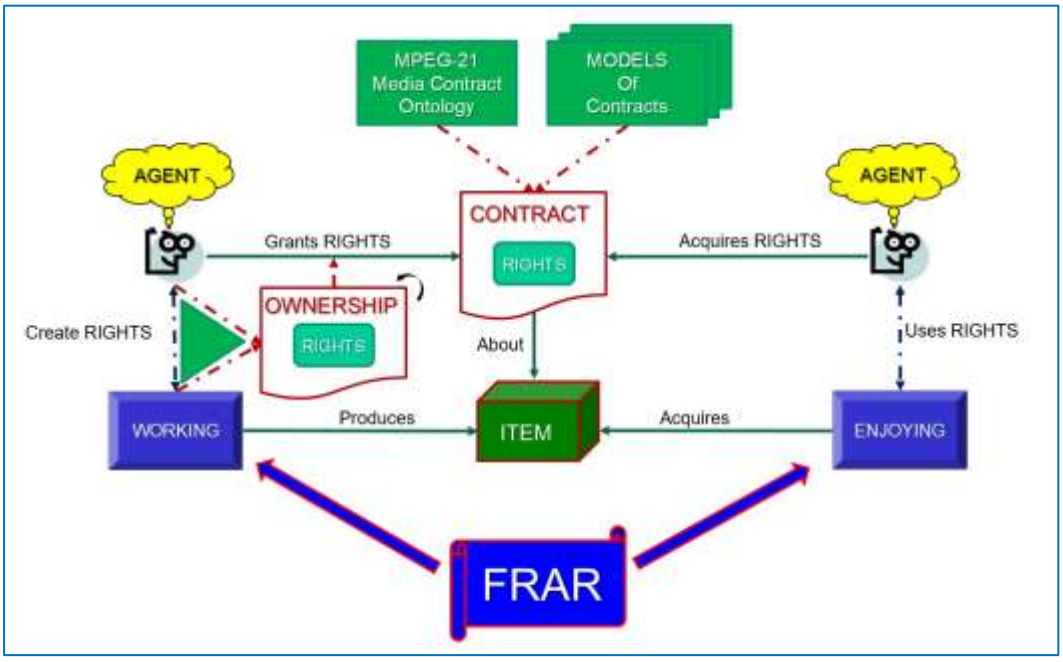


Figure 6-20: The RIGHTS in FRAR

6.4.5 FRAR and FRBR Mapping:

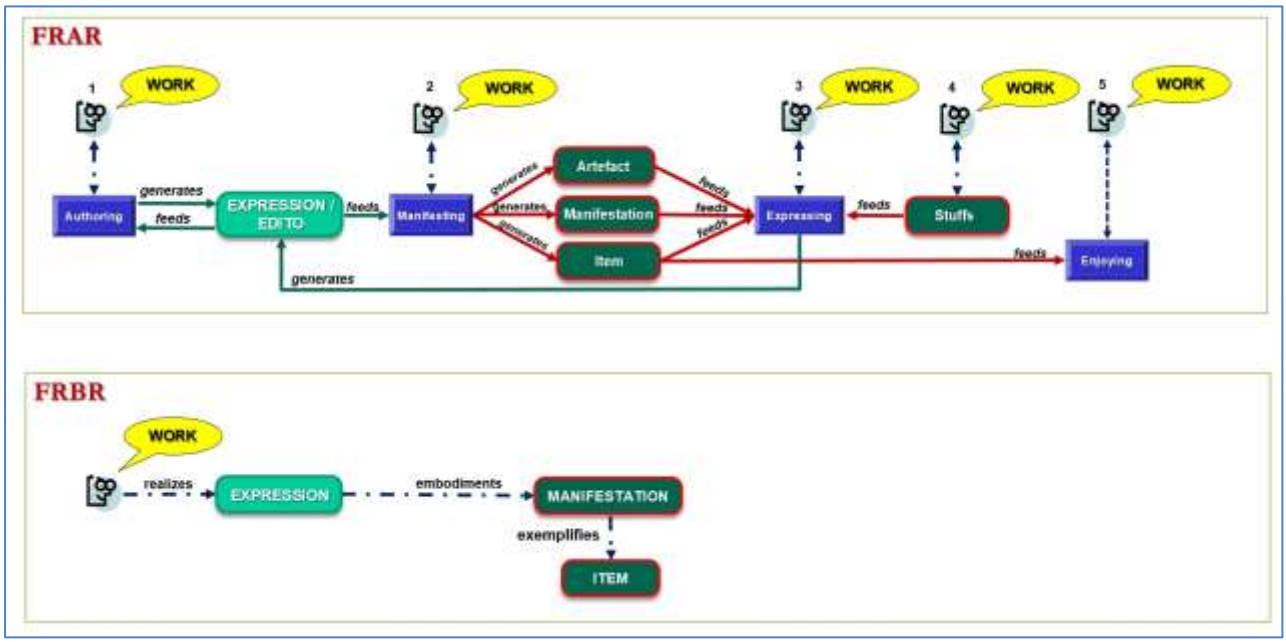


Figure 6-21: Mapping FRAR and FRBR

6.4.6 FRAR mapping on the FRBR standard:

An example of application of this functional model to the functions of the OAIS is when the 'Consumer' is (or interacts) with the 'Producer is shown in the following figure 6.20. Notice that the data are simply designated as SIP, AIP and DIP.

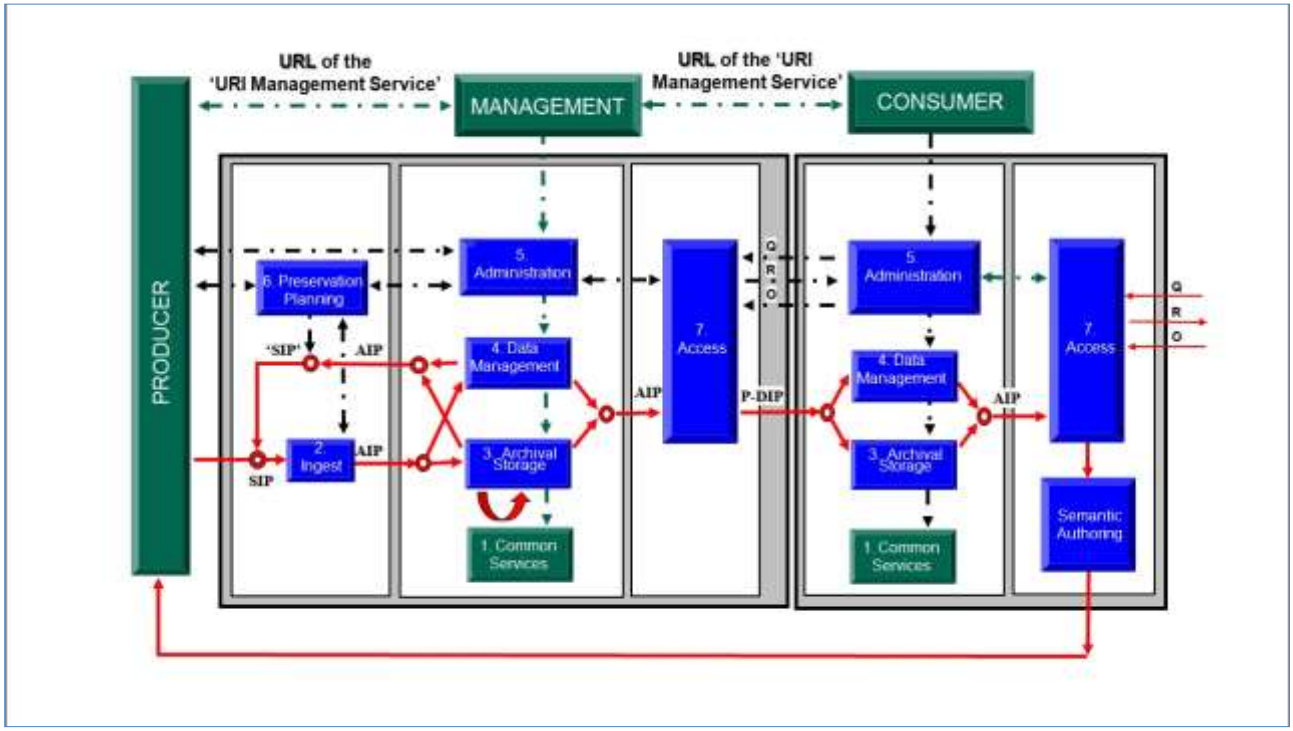


Figure 6-22 : OAIS model in AXIS-CSRM mode (SIP-AIP-DIP loops)

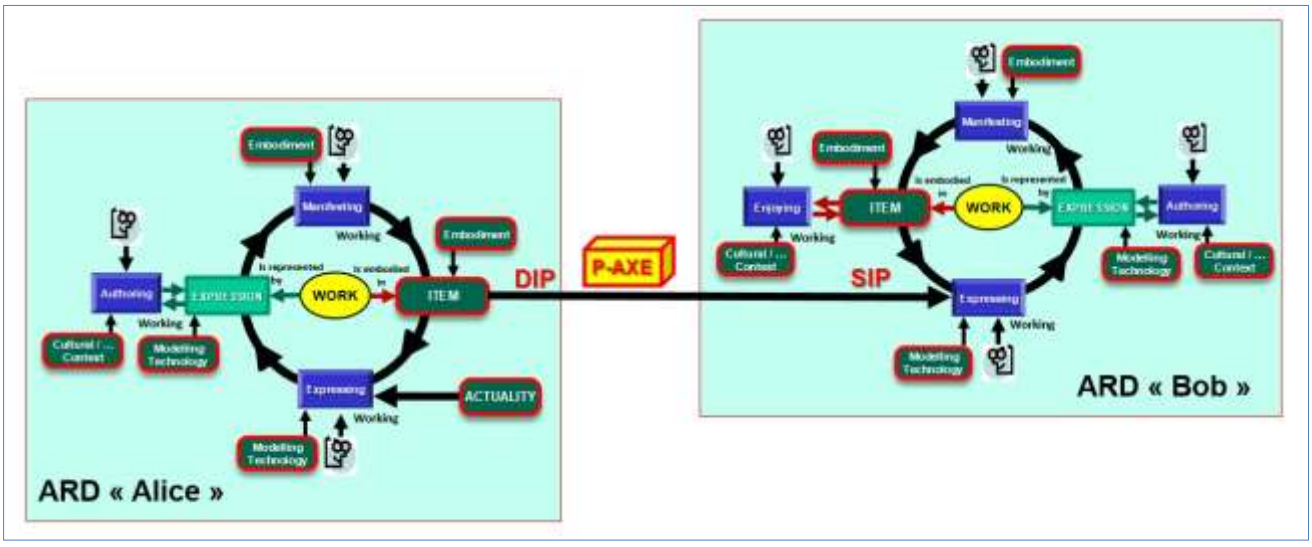


Figure 6-23 : Mapping of FRAR on OAIS in AXIS-CSRM (SIP-AIP-DIP loops)

The figure 6.21 presents two ARDs exchanging P-AXEs; each of the ARDs is shown as two FRAR loops.

6.5 AXIS-CSDM (Conceptual Semantic Data Model):

The Conceptual Semantic Data Model [CSDM] is one of the eight models of AXIS-CSRM. As its name indicates it, the CSDM is a conceptual ontology. It is a conceptual definition of **Constructs** [**Entities, Classes, Relationships, Properties** … and **Vocabulary]** which could serve of base for the definition of Upper-Ontologies according to specific ITC representation standards and through appropriate tools.

The data model is very synergistic with the functional model. As in the OAIS standard the functional model and the data model are inseparable but they each correspond to a way of modelling reality, of looking at IT resources in two quite different but complementary ways.

The data model is concerned to express how to organize data to represent concepts. This conceptual model states on the representation of the data at the semantic level, from EXPRESSIONs, MANIFESTATIONs through ITEMs.

In AXIS-CSRM there is no question of producing a data model associated with the operation itself. The focus is on the time of the data exchange. The exchange covers Packaged-Autonomous Entities eXchange (or P-AXES) or AXES because they are always 'packaged' in a specific way.

In OAIS terminology, AXES represent both DIP (Dissemination Information Package) when it is one of ARD (Autonomous Resource Domain) which sends, and is viewed as a SIP (Submission Information Package) by the domain that receives. In the opposite case, when a domain responds, it returns a DIP which is seen by the receiving domain as a SIP.

In case where the ARD requests information, the dialogue is managed by protocols. Data relating to the Configuration Management Protocol are also elements of the AXIS-CSRM data model.

Finally, it is necessary for humans (citizen) to see, hear, feel, so have perceptions of what is represented inside the computer. The very purpose of the data model is to provide the means for processing the data in order to achieve perception. In the case of a .wav file, in order to reproduce the sound through a loud speaker, it's necessary to follow a set of processes related to the use of protocols. Ditto for the representation of a MPEG2 file for a display on screen.

In AXIS-CSRM the logic aims to know how to create AXEs that can be both SIP and DIP, knowing that within the ARD there may be operational data structures which are proprietary.

During the exchange of AXEs between domains, there are still traces of the origins of the information extracted. Indeed, the protocols require the exchange of information relating to locations (URL) while in AXEs, a priori, there are only URNs. The only available URLs are those that indicate the location of the Configuration Management Services needed to find the bits in question.

Combined, the 'Semantic Conceptual Data Model' [CSDM and the 'Functional Model' [FRAR]) is the hard core, the root of the eight models of AXIS-CSRM.

It is pertinent to repeat here the ISO definitions of "Data" and "Information":

- **DATA:** A representation of facts, concepts or instructions, in a formalized manner, suitable for communication, interpretation, or processing by **human** or by **automatic means**.
- **INFORMATION:** The meaning that human assigns to DATA by means of conventions applied to the DATA.

6.5.1 Introduction:

A good "Data Model" is essential for any conceptual model related to interchange and archiving: indeed, it is the data that is exchanged between domains, what in AXIS-CSRM we call ARD, or what are the data presented in perception to humans. The AXIS-CSRM data model hardly manages only organizational data concepts, their evolution, their structure, materialization/incorporation [embodiement] (from Manifestation to Item and vice versa). It recognizes the need to declare the existence of physical things, to represent them in many ways and the need that these things could be "static" or "flow".

About stream, the poet Henry Bauchau[19] has a famous verse that has influenced many artists (and, in particular, Pierre Bartolomée[20] in his Requiem [CYP-1655]): "The song of the lark does not age! ") Indeed the sound heard when the lark sings evolves over time, while its record is static! For expressing the recording, it has to be transformed into a sound stream: just issued, the sound goes out into space. If you block your ears by listening, there is nothing left.

AXIS-CSRM is not concerned to impose a particular way of modelling the Entities. Instead, AXIS-CSRM was built to accommodate so many ways to model the Entities that imagination or need to establish. In some cases, these models will be expressed following semantic technologies; in other cases, AXIS-CSRM declares the existence and availability of a non semantically expressed model: there will be a model identification and the instantiation parameters (if any).

The Data Model AXIS-CSRM can't, strictly speaking, be considered as an "upper ontology" in that it strives to be orthogonal to what is to be declared or represented. This is its strength. It also features all the semantic hooks to integrate all appropriate ways to report something specific according to all relevant approach in the case. It is in the 'Profiles' that this particular vision will be expressed and associated with all instances of classes mobilized to formally express this vision. This section aims to introduce and motivate key concepts of this Data Model. These concepts will then be detailed in the following subsections.

The implementation of the data model will take the form of definitions and occurrence of ontology constructs: **Classes; Relationships and Properties**. It will be shown later that this implementation can be carried out on the basis of international standards. Several expressions of the implementation of the "AXIS CSDM" conceptual data model can coexist: indeed, for performance reasons (for example) certain conceptual aspects could be implemented via simpler data structures that would be supplemented by processes or inference mechanisms such as in practice the functions are fulfilled correctly but at the cost of a decrease in readability. This aspect will be illustrated later in the case of the reification of the Properties.

On the other hand, as will be seen in the Import & Export model, AXIS-CSRM recommends that for the creation of the Export Packages [the **DIP**s in the OAIS terminology] the ontology used to be as close as possible to the conceptual model, so that interoperability is preferred over performance. Clarity and readability are the guarantors of transcoding guided by semantics, that is to say that the Interoperability Wickets will have no semantic loss in the creation of AIP in the Ingest functional module (see OAIS standard **CCSDS 651.0-B-1**) from the **DIP**s mentioned above considered by the Ingest module as **SIP**s.

One of the originalities of AXIS-CSDM is to organize and structure the data by stating the existence of **Things** independently of their representations (**Artefacts)** and their activation. In turn, Artefacts can be seen as Things. It must be emphasized that Things and Artefacts can be means that can be activated to produce effects, transformations of Things and Artefacts (what are sometimes called processes).

The following two figures illustrate this CSDM root schema, first by an example; then by generalization:

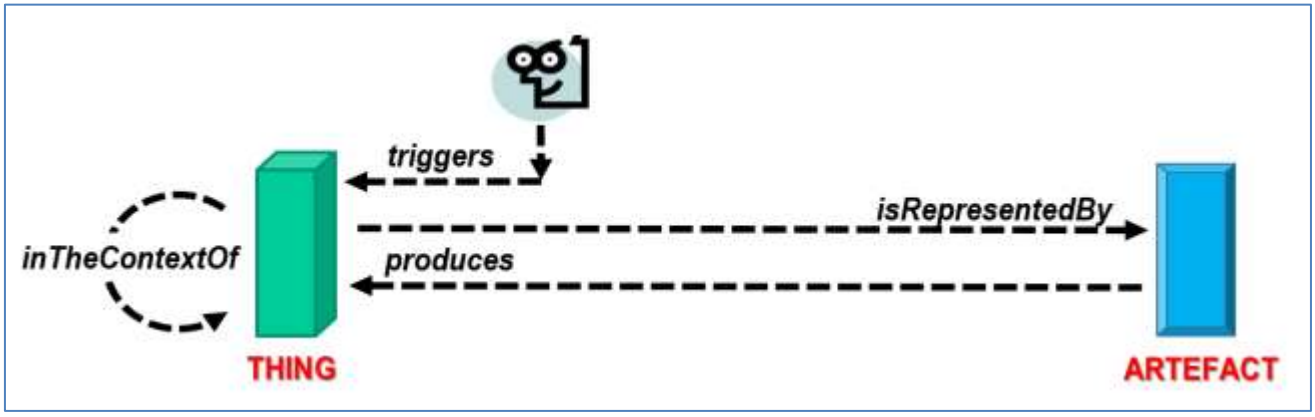


Figure 6-24: The CSDM root ontology schema (example)

[19] Henry Bauchau, born in Mechelen on January 22, 1913 and died in Louveciennes (France) on September 21, 2012, is a Belgian psychoanalyst, also a French poet, playwright and novelist

[20] Knight Pierre, Georges, Edouard Bartholomée, born August 5, 1937 in Brussels, is a Belgian conductor, composer and pianist.

The example shows that a [Thing] can be represented [isRepresentedBy] by an [Artifact]; that a [Thing] can be activated [triggers] by a human action; that an Artefact can be such that if adequate resources are available [inTheContextOf] and activated [triggers] a [Thing] can be produced [produces].

The following figure generalizes the example to express that the relationship 'triggers' of the example is a particular occurrence; likewise for 'produces' or for 'inTheContextOf' or for 'isRepresentedBy'.

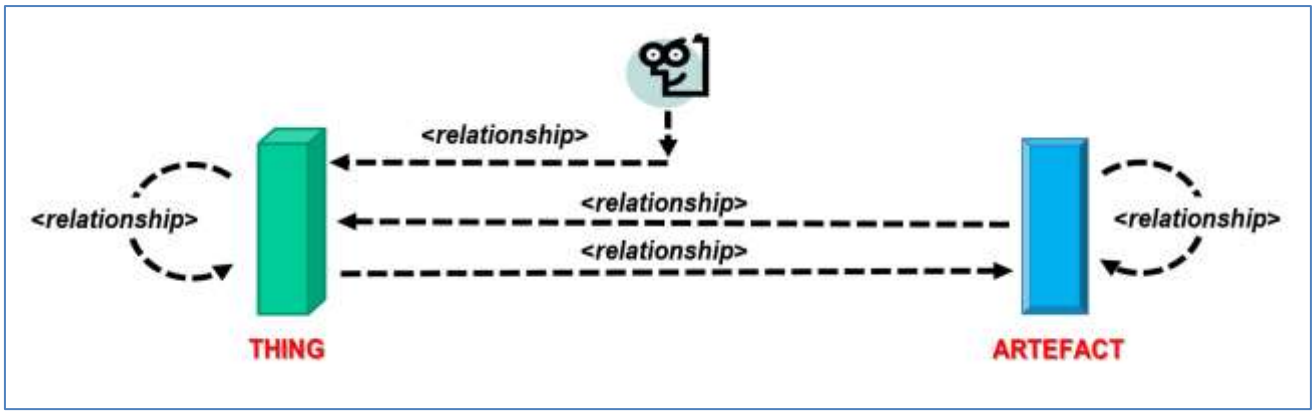


Figure 6-25: The CSDM root ontology schema (generalization)

For occurrences, it is necessary to replace the generic expression <relationship> with the term of the selected Relationship.

The AXIS-CSDM allows to model the definition of the existence of Things and to model them according to several Modelling Technologies. The definition (Ontology, associated Knowledge base …) of these Modelling Technologies [MT] is outside the scope of the CSDM. The illustrations of the concepts will induce the existence of these trivial cases.

The AXIS-CSRM/CSDM is a common framework. Users are invited to, and should, further enrich the model with classes and properties fitting their needs more specifically.

The AXIS/CSDM has been purposefully designed as being directly linked to the other AXIS-CSRM models, in particular the Functional model and the Identification model.

The AXIS-CSRM/CSDM specification also allows reconciliation & semantic migration of existing, possibly inconsistent, metadata Data bases. Dedicated Metadata in several data bases could be migrated and semantically “aligned”: an example is given in the case of Metadata based on the Dublin Core elements.

The second of the main originality of AXIS-CSDM is to organize and structure the **Constructs** data allowing the expression of two distinct structures:

1. The first concerns the modelling of Things: the **Entities**

2. The second concerns the structuring of Things: the **Configurations**

The following figure illustrates this approach:

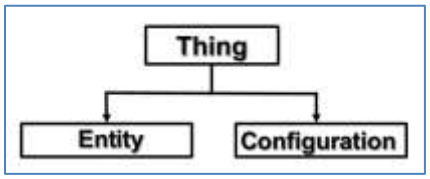


Figure 6-26: The root Classes of CSDM

The Class concept will be explained in 6.5.2.B.The next figure illustrates the roles and coupling of the Entity and Configuration concepts in order of declaring the existence of Entities and of Models of linking the Models themselves and the inner and outer structures.

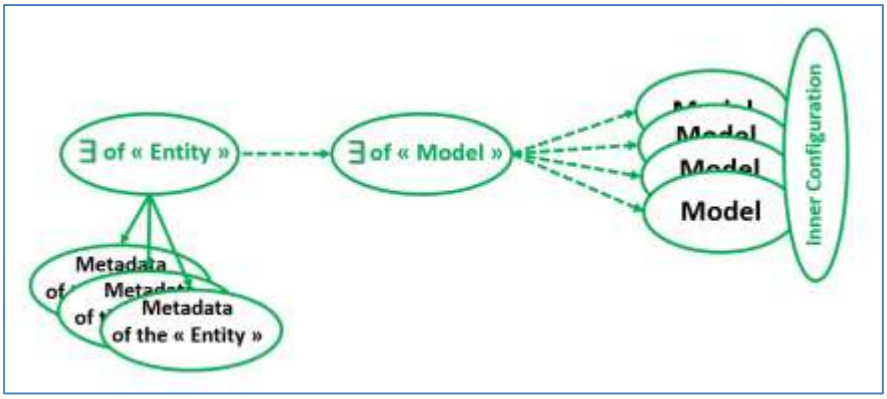


Figure 6-27: The CSDM root structure of the configuration models

The modelling of Thing is represented by the **black colour**! On the one hand, the models [Model] which represent Things and, on the other hand, the Metadata sets [Metadata] which represent the data which are necessary to model the statement of existence of Things. AXIS-CSDM is not concerned with imposing modelling ways! In fact modelling Things are completely free. AXIS-CSDM simply recognizes the existence of modelling ways that are called "Modelling Technology".

The structuring of Things is expressed by the **green colour**. The data representing the structures are disjoint from the data that represent the models. AXIS-CSDM proposes a Upper Ontology to represent the structuring of Things. This structuring always begins with a statement of the existence of Things and a statement of the existence of Models. At the detailed level of structuring, AXIS-CSDM proposes a mechanism of representation of internal structure or synchronization between the models. This is represented in the figure by the ellipsis labelled 'Inner Configuration'.

After defining configuration constructs in AXIS-CSDM, "Entities" are the things we are talking about at some point in an “Autonomous Exchange Entity”. The root structure becomes:

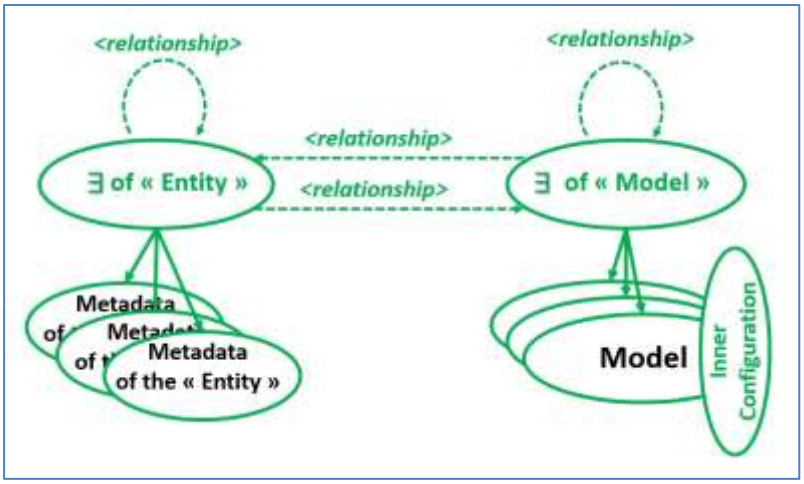


Figure 6-28: The root structure of the configuration of the models of "Entities" in CSDM

Notes on AXIS-CSDM implementations:

As mentioned above, the implementation of AXIS-CSDM involves defining ontologies such that AXEs can export DIPs in formats as close as possible to the AXIS-CSDM conceptual Constructs. This is why conversely AXIS-CSRM strives to present Conceptual Constructs in such a way that such mapping of the conceptual to the reality is the easiest.

As has been said also, Constructs allowing to express the 'Configurations' of Things are fundamental! Such an ontology is conceptually introduced in the following sections and its expression in OWL is presented in the appendix.

To achieve examples, it is necessary to choose 'Entities' ontology. The one that is conceptually introduced in the following sections and its expression in OWL is presented in the appendix. It should be emphasized that AXIS CSDM is open to other ways of modelling Entities.

The representation of DIPs closer to the concepts leads to larger packages than those that could be generated from AIP! But this would imply that it is assumed that recipients who would take these DIPs as SIPs have the same tools, resources, performance compromise contexts and targeting on a Designated Community. This point will be illustrated in the section dealing with the 'Reification' of 'Properties'.

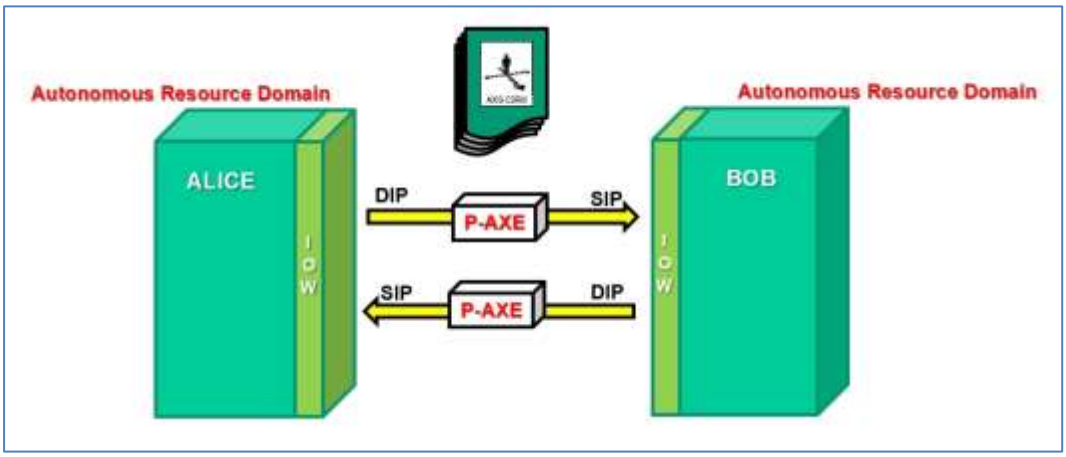


Figure 6-29: The interoperability between two independent ARD

Section 6.9 will be devoted entirely to detailing the structure of P-AXES and how to generate them. The autonomous character of the P-AXE implies that encapsulation would not only incorporate the data representing the transmitted contents, but also the data representing the data models defining the representation technologies used to represent the contents and, when applicable, the tools applicable (opening and expression tools for content representations).

The figure illustrates the transfer of a P-AXE ["Package" encapsulating an AXE (Autonomous eXchange Entity)] from the Autonomous Resource Domain "ALICE" to the Autonomous Resource Domain "BOB"; and vice-versa! ALICE, from its AIP [Archival Information Packages], through its IOW [Inter Operability Wicket], issues a P-AXE formatted according to AXIS-CSRM; the IOW of BOB processes the P-AXE to check it and to include it in its AIP database. For ALICE, the P-AXE is a DIP; for BOB, the P-AXE is a SIP [Submission Information Packages].

The implementation of the Conceptual Semantic Data Model consists in expressing the concepts by a concrete mapping onto standards such as RDF/OWL and Web-3/Linked Open Data. That does not imply that the mapping has to cover the full AXIS Conceptual Data Model, nor construct the IT ontology with a "one to one" mapping.

The representation of the 'Simple Knowledge' (Thesaurus & Taxonomies) could similarly be expressed with SKOS. Such a typical example of the tools is Protégé: https://protege.stanford.edu/

The AXIS-CSRM/CSDM specification combines several aspects from existing models and specifications into a common framework. In particular:

- The EBU-Core EBU-CCDM: see http://tech.ebu.ch/
- The FRBR: https://www.ifla.org/publications/functional-requirements-for-bibliographic-records
- The CIDOC-CRM: http://www.cidoc-crm.org/
- The FRSAD: https://www.ifla.org/files/assets/classification-and-indexing/functional-requirements-for-subject-authority-data/frsad-final-report.pdf
- The ABC Ontology: see http://dcpapers.dublincore.org/pubs/article/view/655/651

In annex, an introduction of the CIMOSA&SAP architectures are presented. Guy-Noël Maréchal has been the scientific coordinator of the CIMOSA, project sponsored by the EU, which has largely influenced the creation of the SAP business.

It is important to mention that SAP at its start has not covered the genericity[21] nor the recursivity[22] of the management of the processes. All these limitations and pitfalls are solved in AXIS-CSDM.

6.5.2 Basic conceptual constructs and their graphical representations

6.5.2.1 The concept of 'Construct' :

In CSDM, three types of **Constructs** are considered: the '**Classes'**, the '**Properties**' and the **Relationships**.

- The **Classes** are the Construct carrying the declaration of existence of Things, of the Models of Things and of the definition of the Concepts.
- The **Properties** are the Constructs carrying the documentation of the classes, in a way, the relations to the metadata of the Classes
- The **Relationships** are the constructs carrying the relations between the Classes.

These three types of Constructs make it possible to semantically model the 'declarations of existence' of the Things, the 'models' of the Things and the definitions of the concepts.

The AXIS-CSDM approach does not require any specific structuring and definition of the Constructs, but rather a conceptual approach based on a set of Classes, Relationships and Conceptual Properties which, for specific applications, could be expressed in contexts adapted to the targeted contexts, in particular for performance reasons. This point will be analysed and illustrated in more detail below.

The structuration and definition of the Constructs is not obvious and has philosophical background.

In order to facilitate understanding and elaboration of particular expressions, an exemplary structuring of concepts via Constructs is presented below.

6.5.2.2 The concept of Classes :

The basic step in any process of describing a conceptual model such as AXIS-CSRM is to generically formulate the content and internal representation of the model. It is to formulate a form of dictionaries (at the service of the Human as the Machine), which is not built on the syntax of terms, but on the semantic interconnection of concepts.

A concept associates a double semantic content: on the one hand, the set of predicates whose conjunction defines the "Being" [23] in evidence (the concept of a triangle); and on the other hand, the set of objects in the case of a triangle that can be referenced via the linguistic expression of the term triangle.

For linguistics, it is the concept of Signified (the mental representation) which is declined in "Denotation" and "Connotation".

In logic, intension (or "understanding") and extension are two ways of defining a concept:

- Intension is based on the set of characters that define a concept, like the set of predicates that belong to a concept. Intension defines the necessary and sufficient conditions for a "Being" (e.g., Belgium) to belong to the extension of a term ("Country").
- The extension is the description of all the "beings" to which the intension (the definition) may apply, which are endowed with the form meant / required by the concept.

[21] Genericity (or generic programming) consists of defining identical algorithms operating on data of different types.
[22] Recursivity is an approach whose description leads to the repetition of the same rule: a dictionary of definitions is a good example.
[23] In AXIS-CSRM, we use the term « Being » (in Uppercase) for naming the things (in lower case) or concepts (in lower case) that **are**,... and that independently of the fact that we perceive, conceive them!. When we perceive them or when we can conceive them, we can decide to explicitly declare their existence in order to being able of describing, representing, sharing, acting ....

The expression of a conceptual representation as a representation suitable for processing by the Information Technology systems implies the selection of a technological language. As an example, the Resource Description Framework [RDF] organizes the definitions via 'Schemas': it designates these "Being" as being CLASSES, they are sets of instances (the Country concept is constituted of France, Italy, the USA ...). A CLASS is thus the formal description of a set of objects having semantics and common characteristics. It is a discrete entity with an identity, properties, state and behaviour that can be invoked.

This approach is equivalent to the notion of **Class** used in the semantic web modelling (see RDF and OWL Primers), also referred to as 'Business Objects' or 'Concepts' in certain projects, see also:

- http://protege.stanford.edu/publications/ontology development/ontology101.pdf
- W3C's Media-Ontogy (http://www.w3.org/ns/ma-ont.rdf) which is based on the EBU-CCDM.

A class in RDF Schema corresponds to the generic concept of a type or category of Things (close to the notion of CLASS in object-oriented programming languages). RDF CLASSES can be used to represent virtually any category of "Things", such as abstract concepts, web pages, people, types of documents, databases …. CLASSES are built from descriptions that constrain the conditions of membership of an instance to a CLASS or a SUB / CLASS to a CLASS.

Since AXIS-CSRM claims to define Concepts, the choice of a form of representation is nodal in order to be able to formulate in a generic way the content and the internal representation of the model.

The graphical representation of Classes in CSDM is a rectangle with the name of the Class included:

Illustration: the CLASS "PhysicalPerson" represented according to the model below:

Figure 6-30: The representation of the "PhysicalPerson" Class

The CLASS "PhysicalPerson" collectively refers to all humans capable of acting. The CLASS "PhysicalPerson" can be conceived as a refinement of a (Super) CLASS which is "Person".

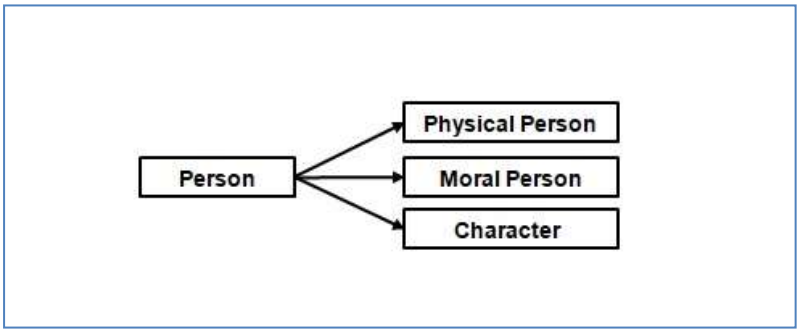


Figure 6-31: The specialization of the "PhysicalPerson" Class

The arrows indicate that "PhysicalPerson" is a "Person" specialization. Similarly other CLASSES may be related to "Person" such as:

- "MoralPerson" representing organizations and companies.
- "Character" the fictional characters (roles in movies, novels, in a drama, ...)

For AXIS-CSRM, every instance of PhysicalPerson is subject to the context of "human rights", that is, subjects of duties and rights and, as such, responsible for their actions in each of the specific contexts in which they are performed.

For the purposes of the conceptual model, and as other CLASSES of resources are likely to act, it is convenient to create a super class of agents, called "Agent" whose CLASSES "ActivatedResource" (representing objects authorized to perform tasks like robots, ...) and "Person" are refinements of Agent:

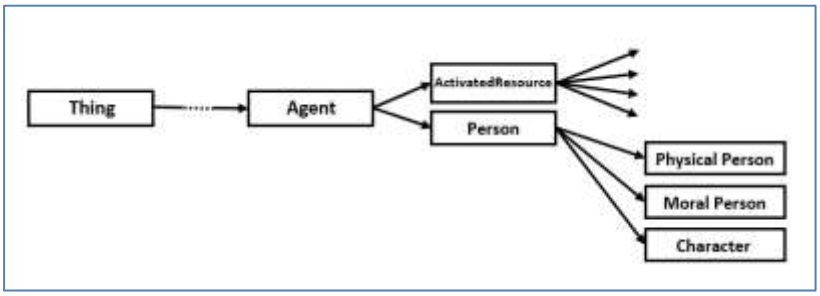


Figure 6-32: From 'Thing' to 'PhysicalPerson'

The figure above illustrates progressive specialization from "Thing" to "PhysicalPerson" and associated CLASSES. The dotted arrow between "Thing" and "Agent" means that the specialization occurs through a chain of CLASSES some of which are not indicated for the sake of schema simplification.

As other classes of resources are capable of acting, it is convenient to create a super Class of agents, named **Agent!** The Class **PhysicalPerson** is a *refinement* of the Class **Agent.** Other Classes being refinements of Agent could be:

- The Class **MoralPerson** representing organizations and enterprises capable of acting
- The Class **Character** representing fictional physical persons (roles in theatre, movies, novels …)
- The Class **ActivatedResource** representing artefacts empowered to perform tasks (robots ...).

Since the objective is to build a dynamic dictionary based on the semantic interrelationships of the instances, it is necessary to define the properties or relations in the form of binary relations (Transitive, Symmetric, ...) in order to have a generic representation.

6.5.2.3 Classes in CSDM:

In the CSDM, the Classes are refinements of the "Thing" Class. The Classes inherit of the characteristics of their super classes. Each Class has a unique name. Each instance of a Class is assumed to have one Identifier, unique within the local context; the instances could also have one or many Names which are not necessarily unique in the Context. The Names of the Classes are strings of characters **starting by uppercase letters**. The approach is usual for the naming convention in the definition of ontologies. Classes are considered as resources in the ICT world.

Note: The level of detail of the ontology: illustration by the Class "PhysicalPerson"

One of the Classes of CSDM named "**PhysicalPerson**" refers collectively to all the humans. In general we understand what is designated by humans and we found it convenient to name the Class 'PhysicalPerson'.

However, for some "Designated Communities" it is required to model it with details and refinement, for example for distinguishing those humans able to act from those disabled and qualify the disease par properties. Each instance of PhysicalPerson is subject in the context of the "Human Rights", i.e. subjects of Duties and Rights depending of its capacity of acting and, as such, responsible for his (her) acts in each of the specific contexts where they are performed.

If one intends to represent the life cycles of the human being, it will be required to consider creating Classes such as **Embryos** and **Corpses** as specializations of the Class **HomoSapiens**, which is a specialization of the Class **Nature**.

The case of AnonymousPhysicalPerson will be treated further later in the text.

In some of the illustrations and examples, the genericity is required. In that case, the name of the Class is not mentioned, but simply the fact that the Class construct is used. In that case the symbolism used is:

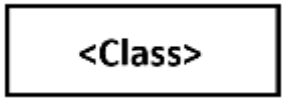


Figure 6-33: « CLASS » Generic symbolism

It means that to use it is required to substitute the string **<Class>** by the actual name of the Class, in the example by **PhysicalPerson**.

6.5.2.4 Super Classes in CSDM: Entities and Configurations:

In the Conceptual Semantic Data Model [CSDM] of AXIS-CRSM, the **Thing** Class is first specialized by two high-level classes, the **Entities** and the **Configurations**:

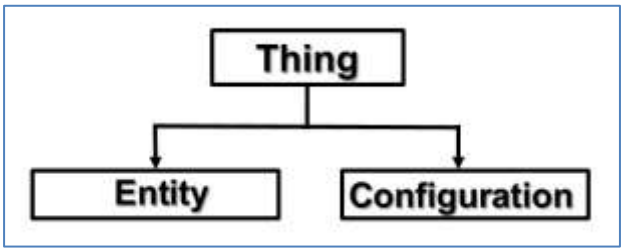


Figure 6-34: Specialization of 'Thing' to 'Entities' and 'Configuration'

The "Configuration" Class

The Class **Configuration** which covers the constructs of the CSDM of AXIS-CSRM for expressing the managerial Classes used for expressing the structure within and between the Entities; for managing their life-cycles; for bundling Entities and Configuration; for constructing the Rights, the Privacy, the Presentations and other similar managerial purposes; for controlling the imports and exports; for holding the definition of the Ontologies, the formats, the thesaurus, the taxonomies and other profiles constructs.

The top level diagram of the **Configuration** is as follows:

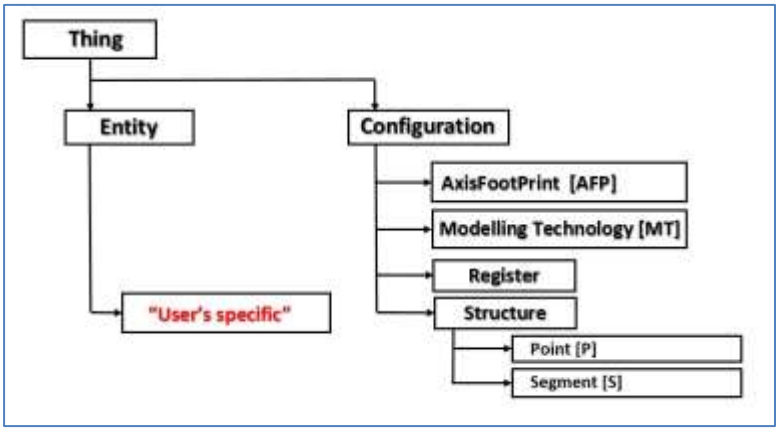


Figure 6-35: Top-level diagram of the 'Configuration'

These classes are direct mapping of the underlying concepts. Each of the Classes shown on the figure will be further defined in the last part of the 6.5 Section and illustrated through examples.

The “Entity” Class

The "**Entity**" class covers all things that are potential subjects for the declaration of existence; of representations and expression of semantic relations between Entities. For the AXIS-CSRM data model, it is up to the users to structure the Entity Class according to their specific needs. On the other hand, the AXIS-CSRM data model details the "Configuration" class so that the classes defined there can manage the instances of the constructs available in "Entity". Nevertheless, the AXIS-CSRM data model presents an illustrative "Entity" structuring in order to be able to present examples of combinations of "Entity" constructs with "Configuration" constructs. This illustrative structuring can obviously be exploited as it is when one prefers to avoid defining a completely personalized semantic structure, to confine oneself to defining only what is really specific and to choosing what one imports as ontologies.
Here is the top level diagram of the **Entities** that will be used in the illustrations:

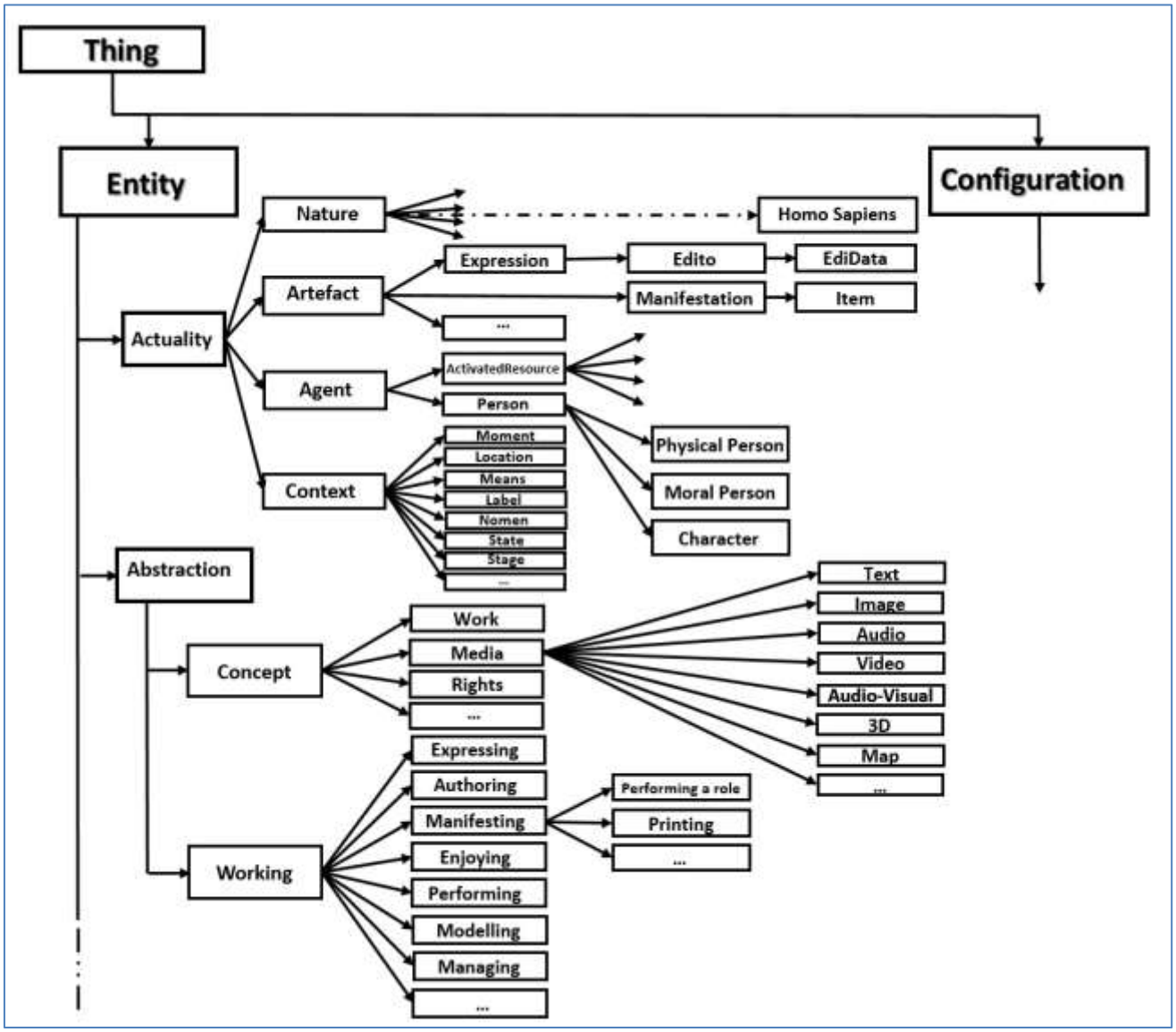


Figure 6-36: Top-level diagram of the ‘Entities’

Each of the Classes shown in the figure are illustrations of the underlying concepts. They will be further defined at section 6.5.5 and illustrated through the examples.

The understanding of the next sections implies further introduction to the concept of "**Modelling Technology**".

6.5.2.5 The Modelling Technologies :

The concept of "**Modelling Technology**" is very general in conceptual models. When implemented, for pragmatic reasons, the Class 'Modelling Technology' will be expressed through refinements as illustrated below.

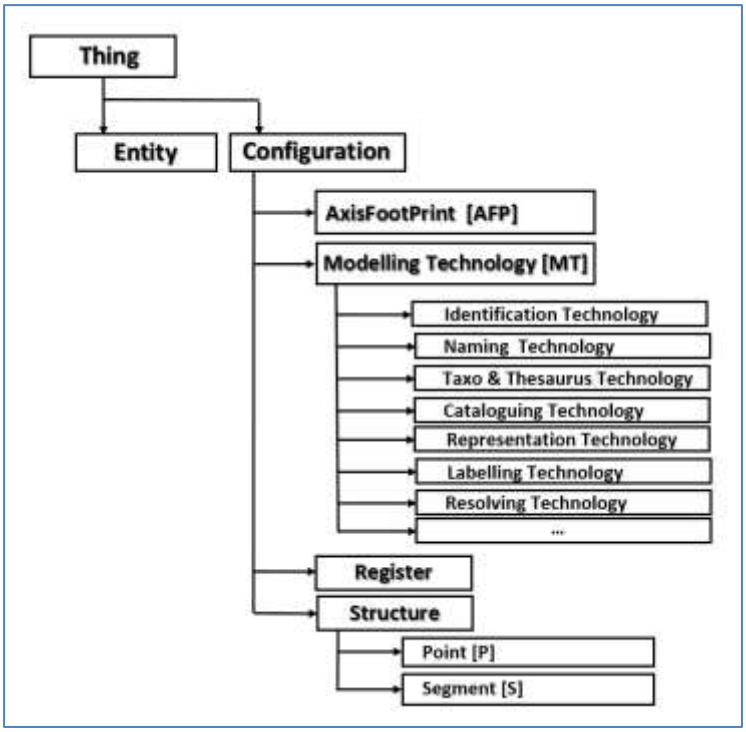


Figure 6-37: The main 'Configuration' classes with Modelling Technologies

These refinements of the Class 'Modelling Technology' will be used when introducing the Class 'Axis_Foot_Print' at the section 6.5.3.3.

But, for simplicity, in most of the cases the generic reference to Modelling Technology will be used, instead of the contextual applicable refinement.

6.5.2.6 Instances of Classes and use of the Modelling Technologies

At usage, the Classes are instantiated: it means that an instance of a Class receive an Identifier such that instances of Classes become resources:

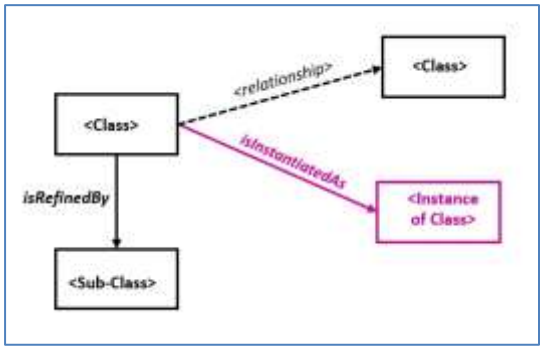


Figure 6-38: Expressing Relationships and Instantiations of Classes

The example below represents graphically one example of an instance of the Class PhysicalPerson:

Figure 6-39: Instance of the Class 'PhysicalPerson'

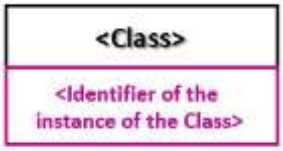


Figure 6-40: Generic representation of an instance of a Class

There are various ways for expressing the Identifiers of the instances of the Classes. The constraint is that each Identifier has to be unique within the local Context. Each of the ways of expressing the Identifier has to have a managerial way for generating it (protocol) and a way of representing the Identifier as a string of characters (coding syntax). Each of these ways (protocol / tools / syntax / ...) is assumed to be formally expressed using a semantic modelling up to the level of detail ultimately or required possible. That will be further detailed later.

Figure 6-41: Codage suivant une Technologie de modélisation Coding according to Modelling Technology

The string **97F3-4AEF-55B7** expresses the value of the URI. The violet colour expresses that the string represents an instance of the Class 'PhysicalPerson'. The black arrow entitled ***accordingTo*** represents the relationship to the Class ***BC-4x3*** meaning that the coding of the URI has been expressed according to the rules and protocols defined by **BC-4x3** a specialization of the Class 'Modelling Technology'. The uniqueness of the names of the Classes and the URI of the instances has to be ensured within the local context (the **ARE)**.

6.5.2.7 Properties of Classes and use of the Modelling Technologies /// Propriétés des classes et utilisation des Technologies de Modélisation

In the CSDM, the Properties are conceptual constructs dedicated to define the intrinsic characteristics of instances of Classes. The Properties are Typed (e.g. 'Moment' or 'Place'… 'Birthdate'…). The instance of a Property is represented according a Modelling Technology (e.g. 'bitrate' expressed as an integer; or the 'CityName' expressed as a string of characters; or a 'Date' as the concatenation of three Properties ['Year' – 'Month' – 'Day']). Some of the Properties could be expressed in more than one human language. In the CSDM, the Modelling Technologies are assumed to exist in a specific context: a Conceptual Model simply has to assume the existence of the Modelling Technologies able to define the formats, coding and protocols.

In the previous example, an instance of the Class 'PhysicalPerson' has received the ID: 97F3-4AEF-55B7: it is a persistent way of identifying somebody.

The first obvious intention is to name the instances of Things:

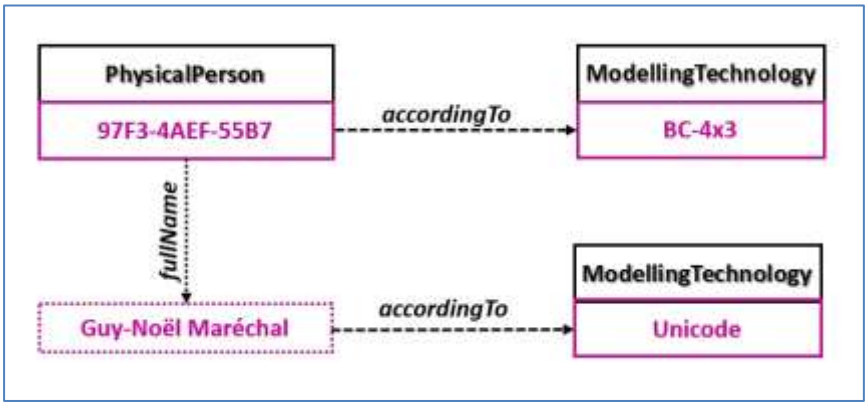


Figure 6-42: The Human view vs the Machine view

The Property '**Name**' will associate a name at the Identifier of the instance of the Class '**PhysicalPerson**'. For human understanding, the name 'Guy-Noël Maréchal' is semantically self-speaking while, for the IT system, 97F3-4AEF-55B7 has the appropriate characteristics for correct processing.

The Modelling Technology '**Unicode**' is appropriate for illustrating the concept of Modelling Technology in Conceptual Models: it has all the characteristics required for coding strings of characters and link to the semantic *rules and protocols.* See the Annex 4 for more details.

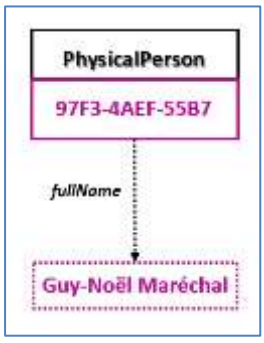


Figure 6-43: Values of the Properties of Classes

In the figures, the link to the Properties is represented by fine dotted lines and the 'Property Name' is presented in a rectangle with dotted borders; the **Value** of the **instance** of the Property is represented in violet.

However, the allocation of value to a property conceptually has to cover the possibility of expressing the same reality or the same target according to various Modelling Technologies (or parametric expression of framed Modelling Technology) or to various contexts (natural languages; moments; usages …).

The expression of a **Moment** and **Places** are a good illustration of the various Modelling Technologies:

the Birth date of Guy-Noël Maréchal could be expressed according to the ISO 8601 (selecting the variant YYYY-MM-DD, where YYYY represents the year of birth; MM the month of birth; and DD the day of birth): the instance **1937-12-25.** But that date refers to a unique moment only if the 'Place' of birth is known; in the case, in Belgium. Further it refers to the "Gregorian Calendar".

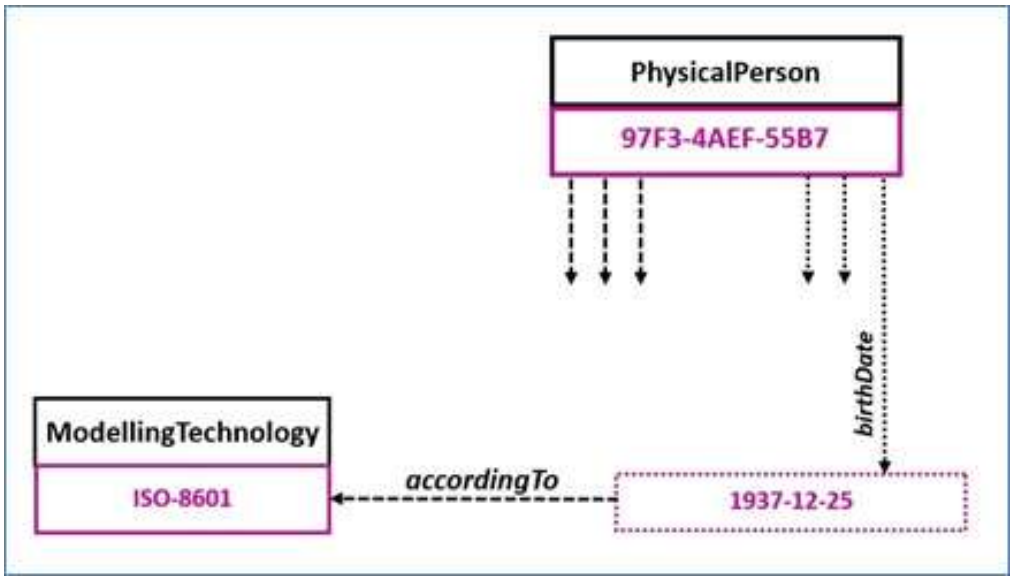


Figure 6-44: Coding of the Values of the Properties of Classes

Figure 6-45: Compact representations

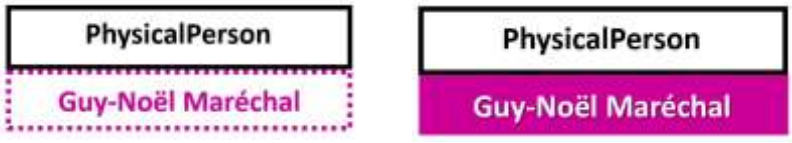


Figure 6-46: Two ultra-compact equivalent representations

The expression of the date of birth could be made according to another calendar: If the reference has to be given to the Hegira Calendar (current version 03), in Brussels, it will become **'21-10-1356'** or **'21 Shawwal 1356'**.

The calculation of the expression of the date following the Hegira calendar can be done from the date value in the Gregorian calendar. In this case, we could limit ourselves to the fact that AIPs only represent dates in Gregorian mode and that the

value of the date could be presented according to other calendars chosen by the Designated Target Community. However, in many cases, the expression of the value of a property cannot be inferred by calculation. The legal identifier of a person is defined differently according to the country.

In Belgium, each citizen has a unique legal identifier: it is formatted according to the following rule: a sequence of five fields, with the syntactic separators depending on the context:

1. The year of birth coded according to **YYYY**
2. The month of birth coded according to **MM**
3. The day of birth coded according to **DD**
4. The identifier of the Registration book of the Public authority: in the case **317**
5. The sequence number of the registrations of the year of birth on the registration book: in the case **85**

On the identity cards the syntactic separators are as illustrated: **1937.12.25-317.85!**

On the passports the representation is **1937-12-25&317.85**

6.5.2.8 Reification of Properties

The **reification** of Properties consists of creating a carrier of different representations of Properties aimed at expressing the same meaning. The bearer will be an instance of Class. In the case of Properties, reification is purely related to the consistent expression of properties; It will be seen later that the reification of relations / relationships is conducted with a view to extending the signified in a radically new way while remaining compatible and evolving.

The word 'reification' is derived from the Latin where 'res' means 'thing': in these cases, therefore, we look at a Property as a 'thing' and thus elevate it to the rank of Class.

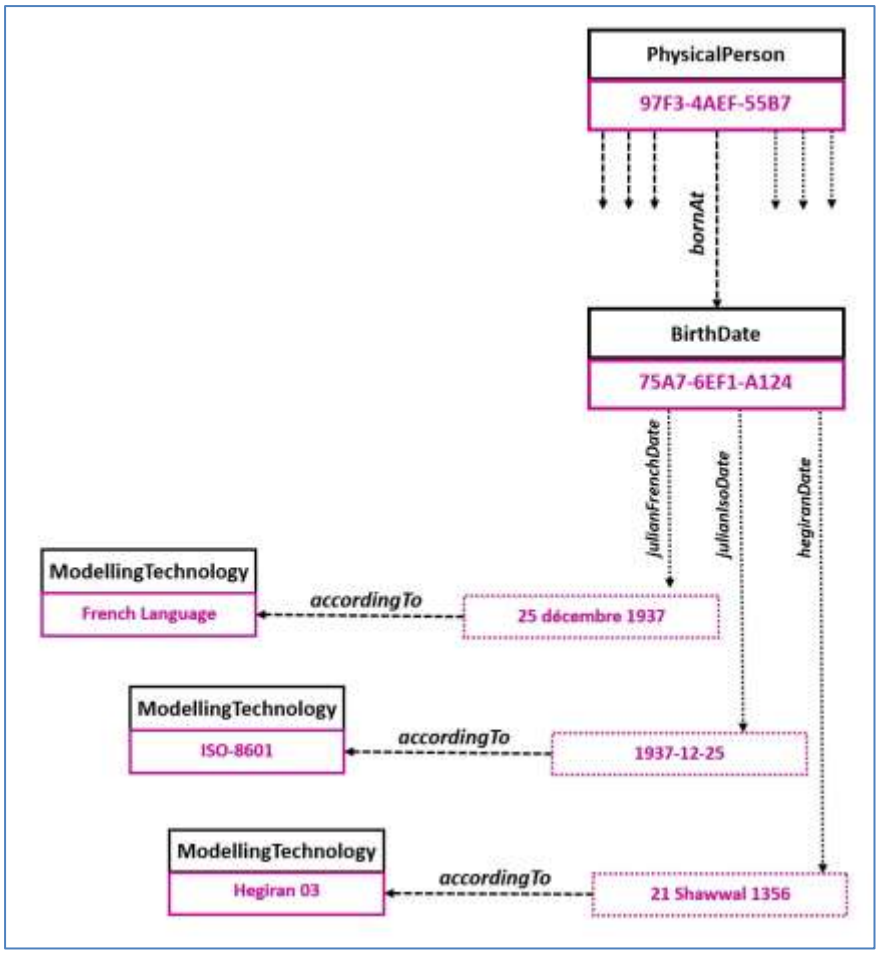


Figure 6-47: Reification of '*birthDate*' in 'BirthDate' through ISO 8601 and other Modelling Technologies

The following figure illustrates the systematic use of reification:

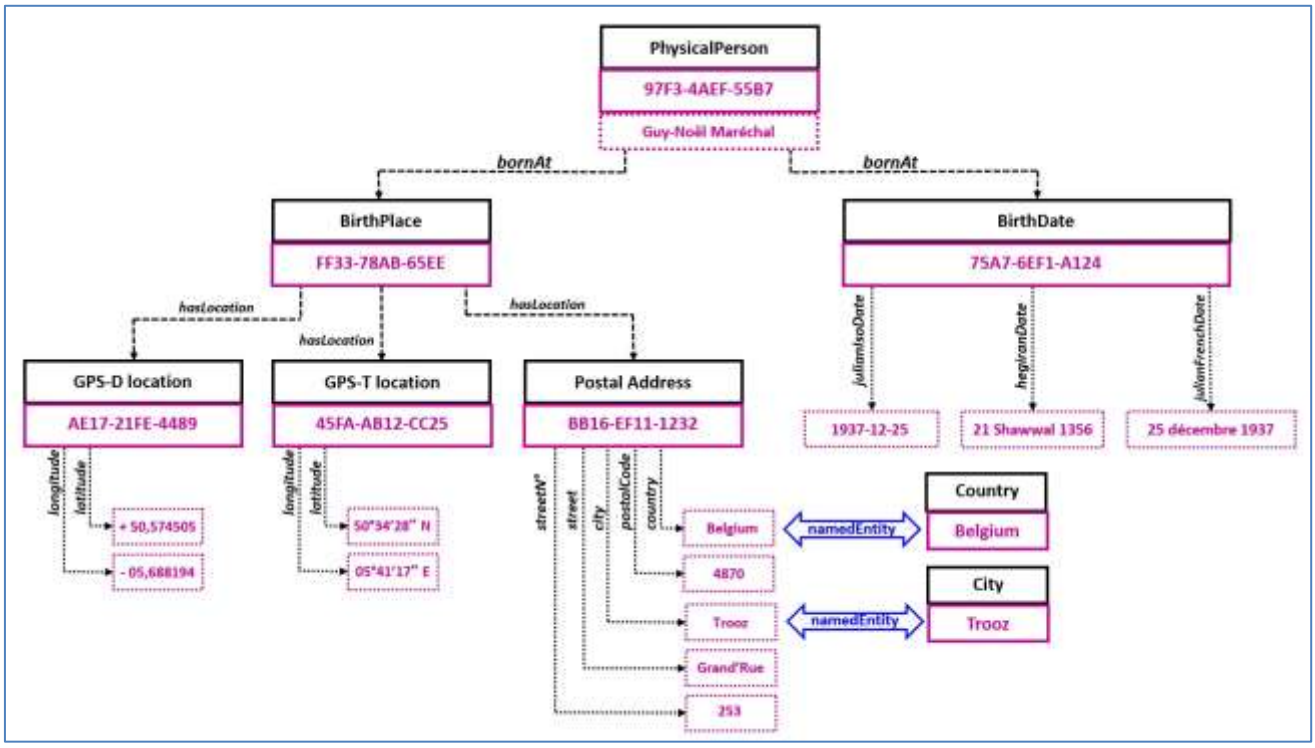


Figure 6-48: Illustration of the reification of Properties

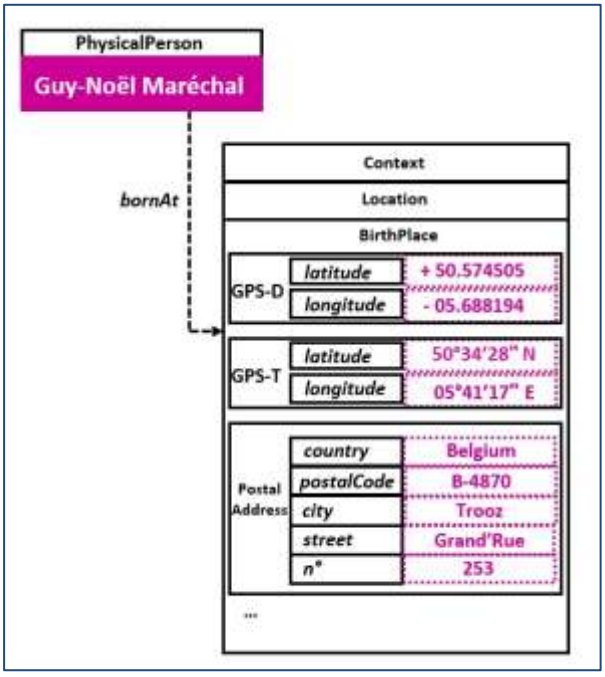


Figure 6-49: Compact representation of the reified Properties

The expression of **Names** is a good illustration of the influence of the Context. The name of the famous painting "**Mona Lisa" of "Leonardo da Vinci**" is usually known as "**La Joconde** " in French-speaking texts.

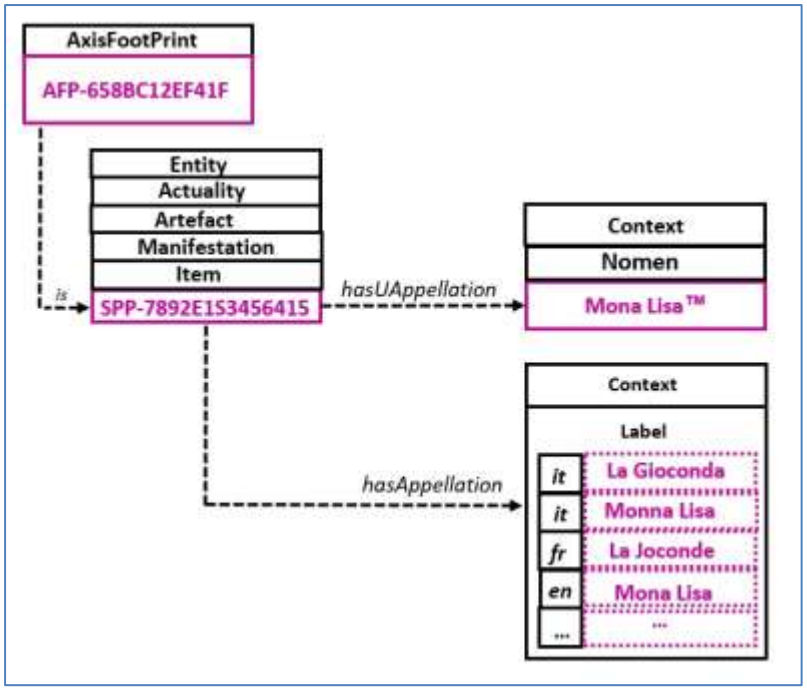


Figure 6-50: Name of Physical Person: multiple syntaxes & languages

Sometimes, the public authorities are changing the names of street for pragmatic reasons (avoiding the same name for different streets after the fusion of municipalities) but also for political reasons: The Paris's "**Rond-Point des Champs Élysées**" has been renamed "**Place Charles de Gaulle**". Constructs able to cope with these requirements (whenever required) will be introduced later.

The Property Construct can be mapped to the dataProperty construct used in the semantic web modelling (see RDF and OWL Primers). In particular, CSDM hide the constructs of OWL for defining the 'data formats' by an abstract general construct called "Modelling Technology".

In the graphical representation, CSDM represents the "Object Property" construct of OWL by a black rectangle and its link to the instance of the Class by a dotted black arrow (meaning '*has*'). That conceptual approach allows to use the same symbolism for expressing the Classes and their Instances as for expressing the Properties and their Instances. The graphical expression of the data structures could be simplified or presented in a more compact way when there is no need of the full representation for the intended message.

The few following examples are self-explaining:

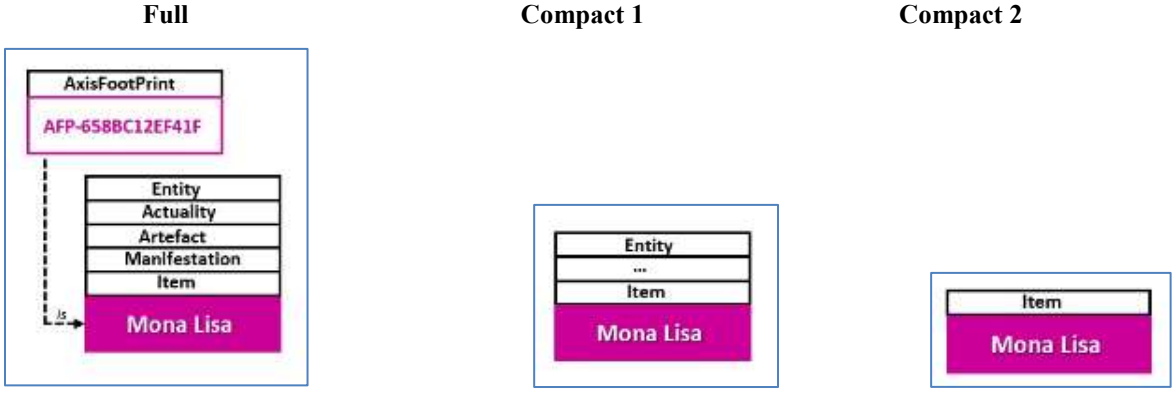


Figure 6-51: Simplification of representations

The "Compact 1" should be used when the reference to the chain of Classes has importance for the illustration; while the 'Compact 2" should be used when the instance of the Name property is enough.

6.5.2.9 Relationships :

The Relationship construct expressed a typed way of linking instances of Classes:

Example of the expression that one specific physical person is the father of another physical person will be represented by the Relationship ***isFatherOf*** .

In the CSDM, the Relationships express the types of links between Classes and/or between instances of Classes.

The names of the Relationships are strings of characters presented as italics letters starting by a lower-case letter and allowing upper case letters within the string for convenient reading and understanding the meaning expressed by the Relationship type. The name of the Relationship is mentioned on the top of a black dashed arrow. If links are shown in both directions, with the same meaning, then only one dashed line is drawn and the name is mentioned only once.

The approach is close to the naming convention in the definition of Ontologies.

The following example is self-explaining:

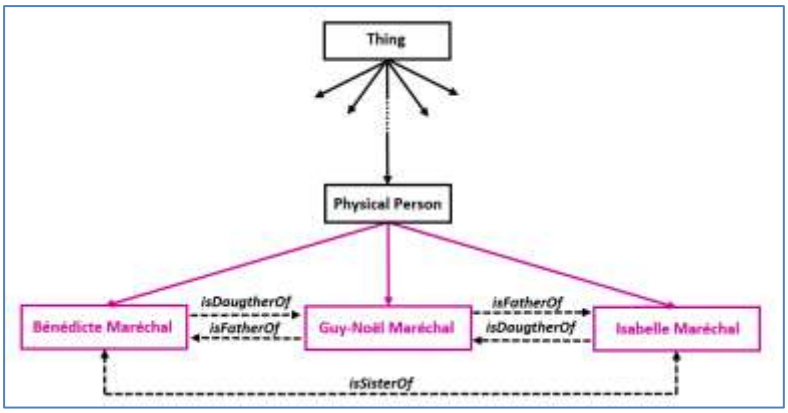


Figure 6-52: Graphical 'Relationships' representation

The Relationships '*isSisterOf*' and '*isDaughterOf*' can be inferred from the Relationships '*isFatherOf*'. The Relationship '*isSisterOf*' is reflexive, which means that one arrow represents both sides.

Reification of Properties

The reification of Properties consists of creating a carrier of different representations of Properties aimed at expressing the same meaning. The bearer will be an instance of Class. In the case of Properties, reification is purely related to the consistent expression of properties; It will be seen later that the reification of relations / relationships is conducted with a view to extending the signified in a radically new way while remaining compatible and evolving.

6.5.2.10 Reification of Relationships

The **reification** of Relationships is to create a carrier of models representing the processes and means implemented to achieve what is expressed by the Relationship!

The bearer will be an instance of Class. In the case of Relationships, reification is carried out in order to extend the signified in a radically new way while remaining compatible and evolving. This far outweighs the motivations and expressive abilities associated with the reification of the Properties.

As an example, we can see the functional model FRAR as a reification of the FRBR model:

Figure 6-53: FRBR (Group 1 Entities)

The functional model FRAR (Entities of group 1) not only reifies "realizes", "embodiments" and "exemplified"; but still differentiates the various Agents and Contexts involved; make possible the cases where the processes are not complete, introduces the possibilities of loops, derivations, fusion starting from several 'Works', creation of variants ...

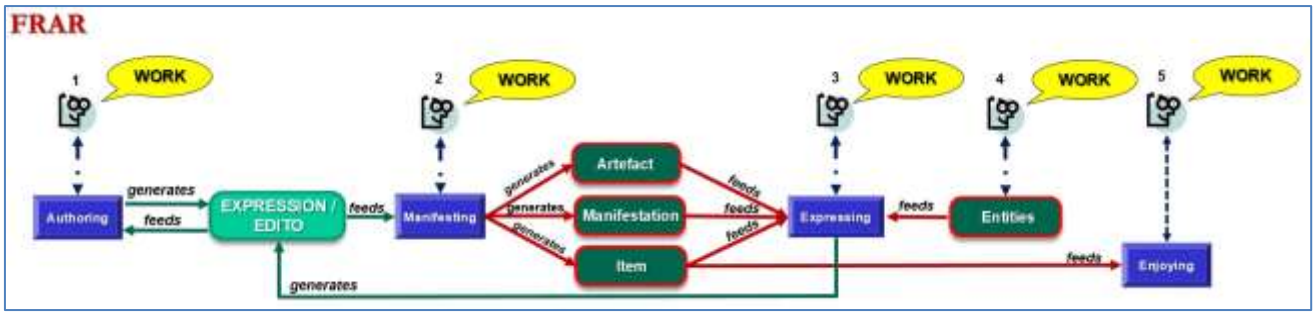


Figure 6-54: FRAR (Group 1 Entities)

The processes by which the FRBR author creates an 'Expression' of 'Work' (what FRBR expresses in the form of the Relationship 'realizes'), becomes a model carried by constructs related to an instance of the Class' Authoring Or the Expressing Class according to the phases of the processes of creation of existence, expression of the authors and modification of contents or formats.

Similarly for "*Embodiments*" which is reified via the "*Manifesting" class.*

The Reification makes it possible to identify the Agents involved at each step, which leads to bearers of the rights associated with the agents and associated ancillary processes.

Moreover, if the archives had representations of "Manifestations" in the FRBR line, the production of FRAR-compliant migrated representations could be made structurally compatible and semantically identical to the source. Enrichment would be expressed by adding structures and explicit expression of processes.

At the same time, enrichment itself can be documented to ensure traceability; and the enrichment can concern the reification of the Properties and their enrichment.

A more complex reification will be introduced further on the "Point" and "Segment" classes: it combines a reification of Properties with a reification of Relationships.

6.5.3 Other Conceptual constructs of the Entities and their graphical representations:

6.5.3.1 General Modelling of the « Entities »:

Section 6.5.3 enlarges the modelling of “Entities” and their structuring via the "Configuration" constructs.

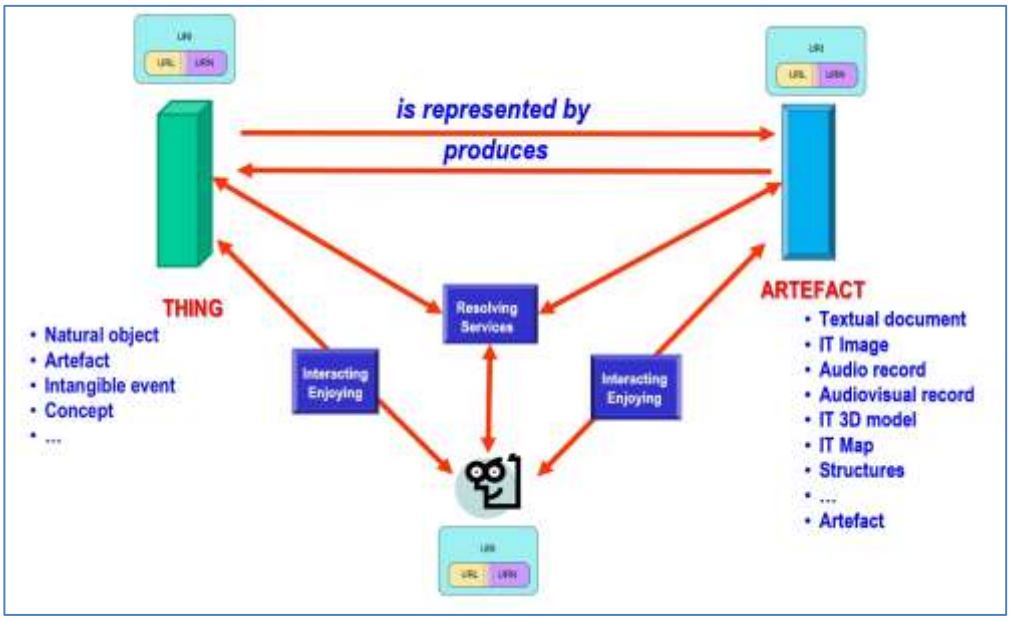


Figure 6-55: Top level general modelling of AXIS-CSRM

As was introduced in the previous section, the data model is built on two types of constructs: "**Entity**" and "**Configuration**".

The term 'construct' refers to the conceptual means used to construct a representation of something.

In AXIS-CSRM, the general conceptual approach of the data model is organized as follows:

1. Statement of existence of an instance of an "Entity"
2. To this Statement of existence, attachment of one (or more) group(s) of "Metadata"
3. Statement of existence of an instance of a "Model" of an Entity
4. To this Statement of existence, hanging of one (or more) model(s) of that Entity
5. Expression of the links between these statements of existence

This approach is illustrated by the example shown below. The intention is to represent a real Physical Object in the computer system.

- The Object can be an Artefact (in this case Lincoln's statue, or a book for example), a Natural Object (a protected natural park, ...) or a Physical Person.
- This Object is associated with an I & T Token [Identification] and / or Traceability Token. This Token would be, for example, a barcode or QR-code and / or an RF-Id badge (see next figure)
- The vertical arrow represents the physical link that is established between the Object and the I & T Token: this can take the form of a Token pasted on the Object; variant, in front of a natural landscape, on a panel; ... For a Natural Person, this is the Passport or the Identity Card.

The figure shows the declaration of existence of this object and its modelling in the computer system: in the first place the declaration of the Knowledge Base [KB] associated with the object, in the centre the declaration of existence of the representation of the object and the same for the Token (Note: the Knowledge Base associated with the Token is not shown).

Each Entity (Knowledge base - Object - Token) is described by specific Metadata and has a declaration of Existence of the models of these Entities. These new entities federate the different models.

The figure finally shows the links between the Entities (Knowledge Base, the object representation and the I & T Token) in Outer Configuration and the links between the various models representing the Entities (Inner Configuration).

<The illustration will have to be further explained>

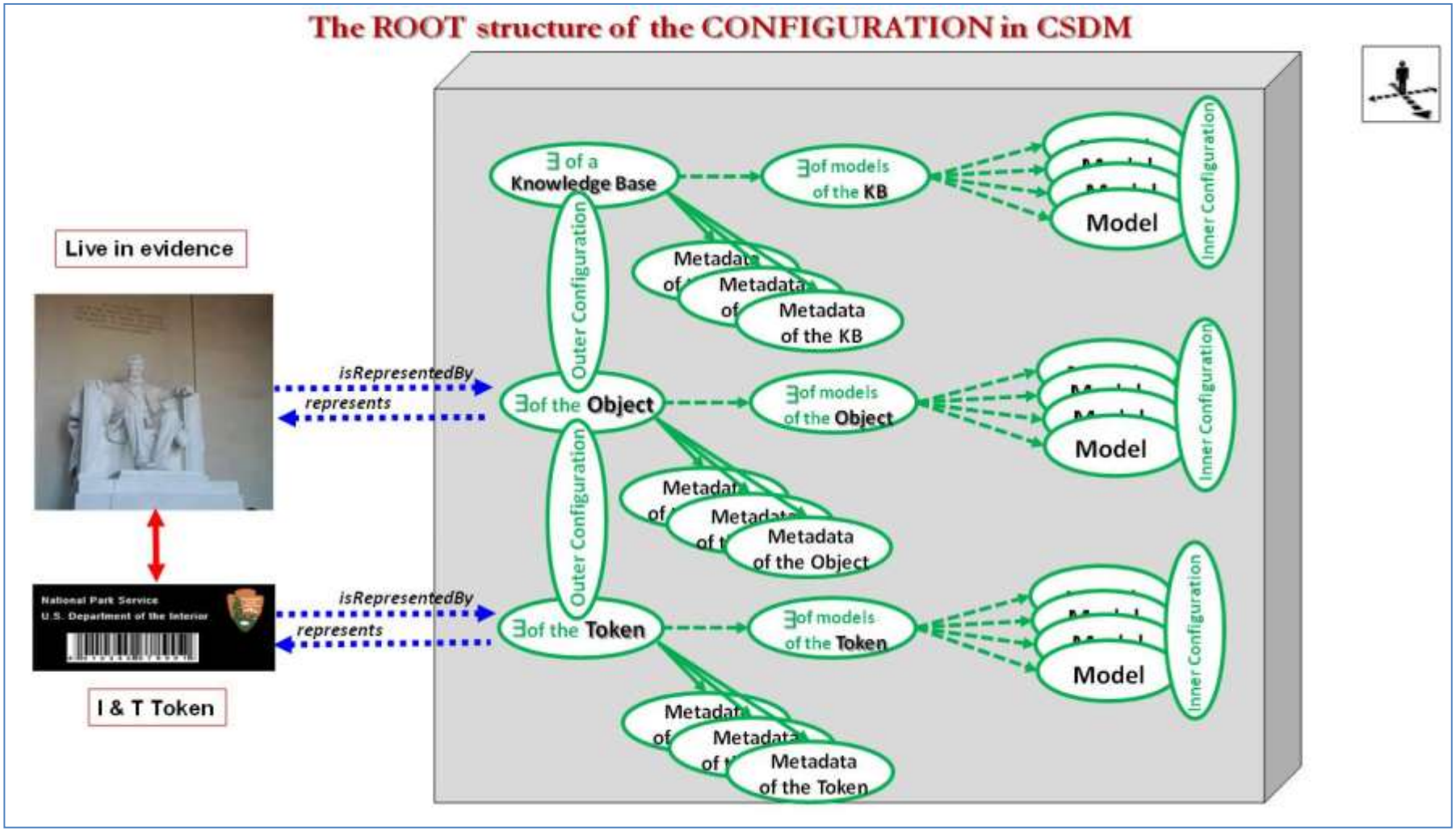


Figure 6-56: Example of application of the approach for the Data Model

The following figure shows the interaction between Models, Humans and the real world:

The human sees “in evidence”:

- Perception [A]: the Object;
- Perception [B]: the I & T Token 'Identification & Traceability Token';
- [C] manipulates the I & T Sensor (protocol agent [C]) to allow the IT system to interact with the I & T Token [4];

- [D] and finally interact with the 'Graphical & Audio & Video Human Interface' to search, find, activate ... this IT System. These interactions concern not only the Representations of the Objects but also the access to the Objects themselves (location and rights).

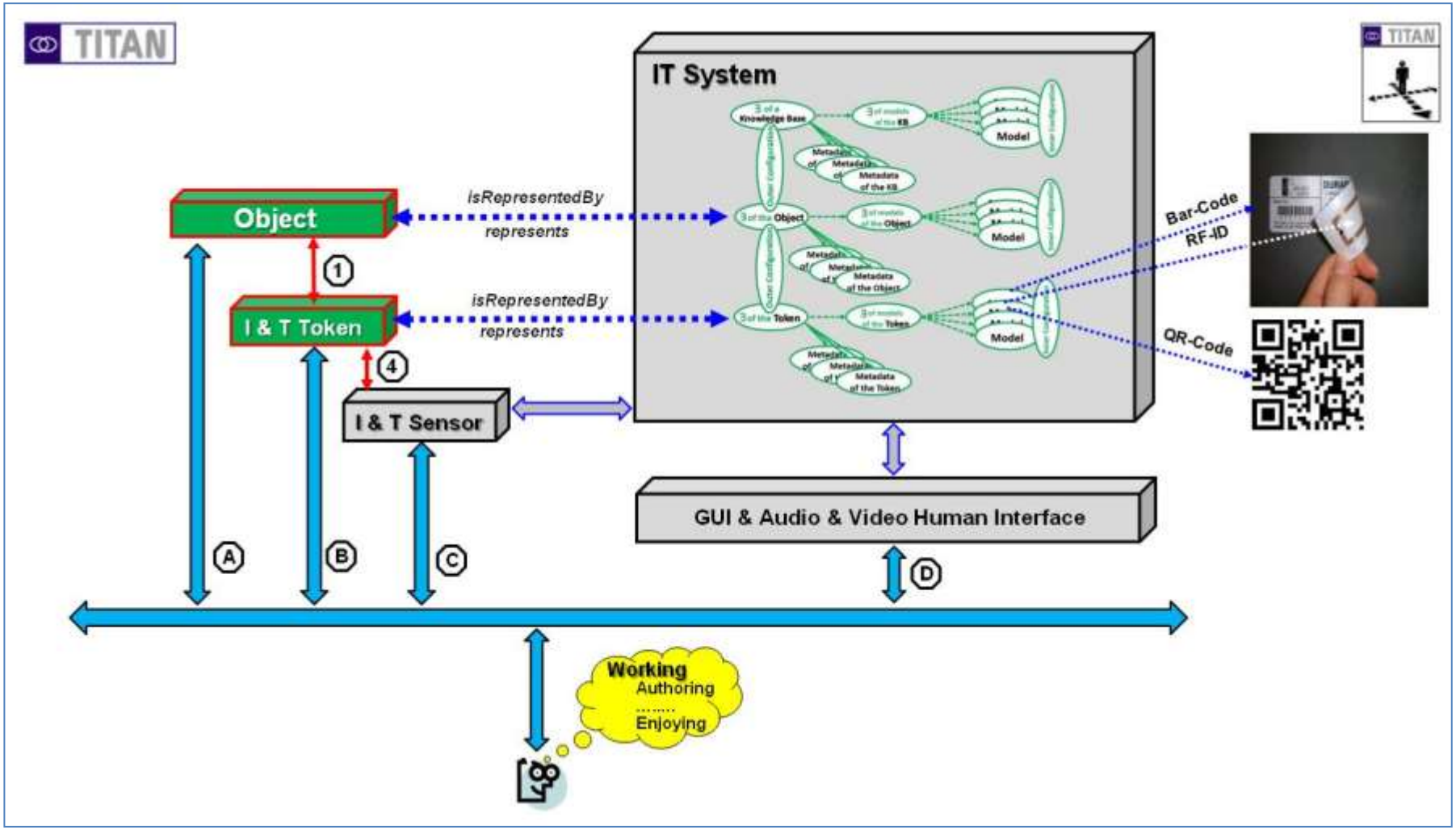


Figure 6-57: Example of application of the approach in the IT context

The conceptual approach of the modelling has been represented at the section 6.5.1 by the following figure:

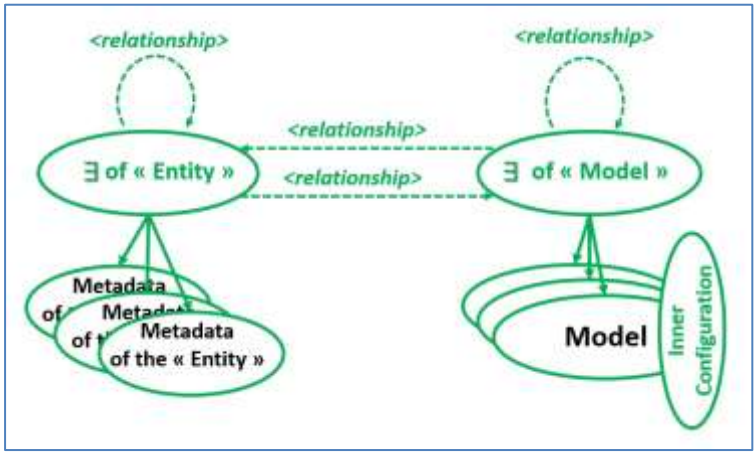


Figure 6-58: General conceptual approach of the Data Model

This figure can be synthesized according to the following figure where <**ID**> indicates the general case and **a** the particular occurrence!

Figure 6-59: Compact representation of the Data model for the A Entité

As was introduced earlier, this approach allows the representation **of Treatment executed by processes between two States**. It is a **representation of particular stages of Classes** that are specializations of the Class "**Working**", which is a specialization of the "**Abstraction**" Class, which is a specialization of the "**Entity**" class. "**Working"** is an abstraction, but the models represent the States and the Treatments that are processed between these States.

The following figure illustrates the interaction of "A" with "B" through the execution of the "P" process between two of these states:

Figure 6-60: the interaction of "A" with "B" through the execution of the "P"

This diagram represents, for example, that Mrs. "B" listens to the piano concerto shown on CD-A "A", thanks to the "P" Process. This "P" listening process, in the example, would involve a CD player, speakers, ... and have activated the reading at 10h22 which will end at 11h17.

The symbol ⟷ represents the interaction.

The approach also obviously covers the evolutions of a particular "Entité" by a chaining of states. This chaining can be **temporal** (from birth to death, for example) or **cyclical** (mass production of objects, for example) or any other chaining

usually referred to as "Finite State Machines". There are many ways to represent these machines! The best known is that of the "Petri nets". AXIS-CSRM is not involved in these ways and merely refers to them via the link to a "Modelling Technology".

The following figure shows the transition from "A" from its state "A (n)" to its state "A (n + 1)" via the "P" process.

The symbol ➡ represents the chaining.

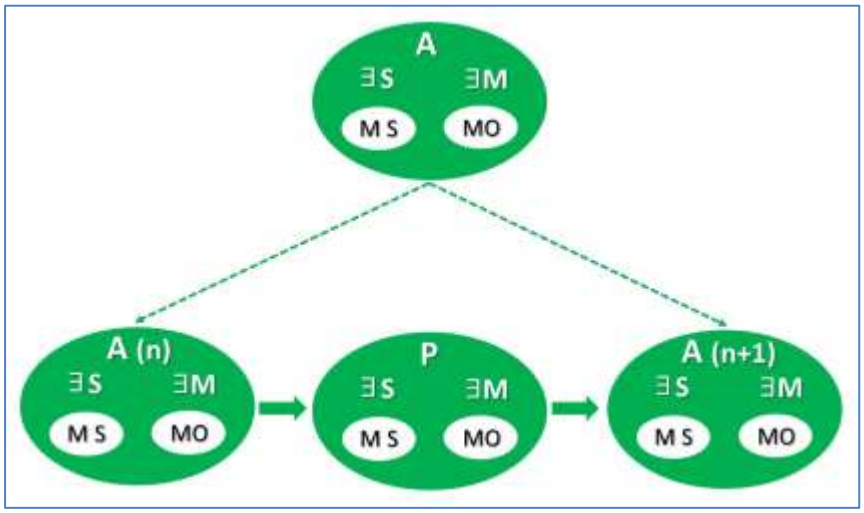


Figure 6-61: the transition from "A" from its "A (n)"state to its "A (n + 1)" state through the "P" process

6.5.3.2 Further on graphical representations:

The graphical expression of the data structures could be further simplified or presented in a more compact way when there is no need of the full representation for the intended message.

A compact representation of the Classes and Properties can be combined through the self-explaining examples:

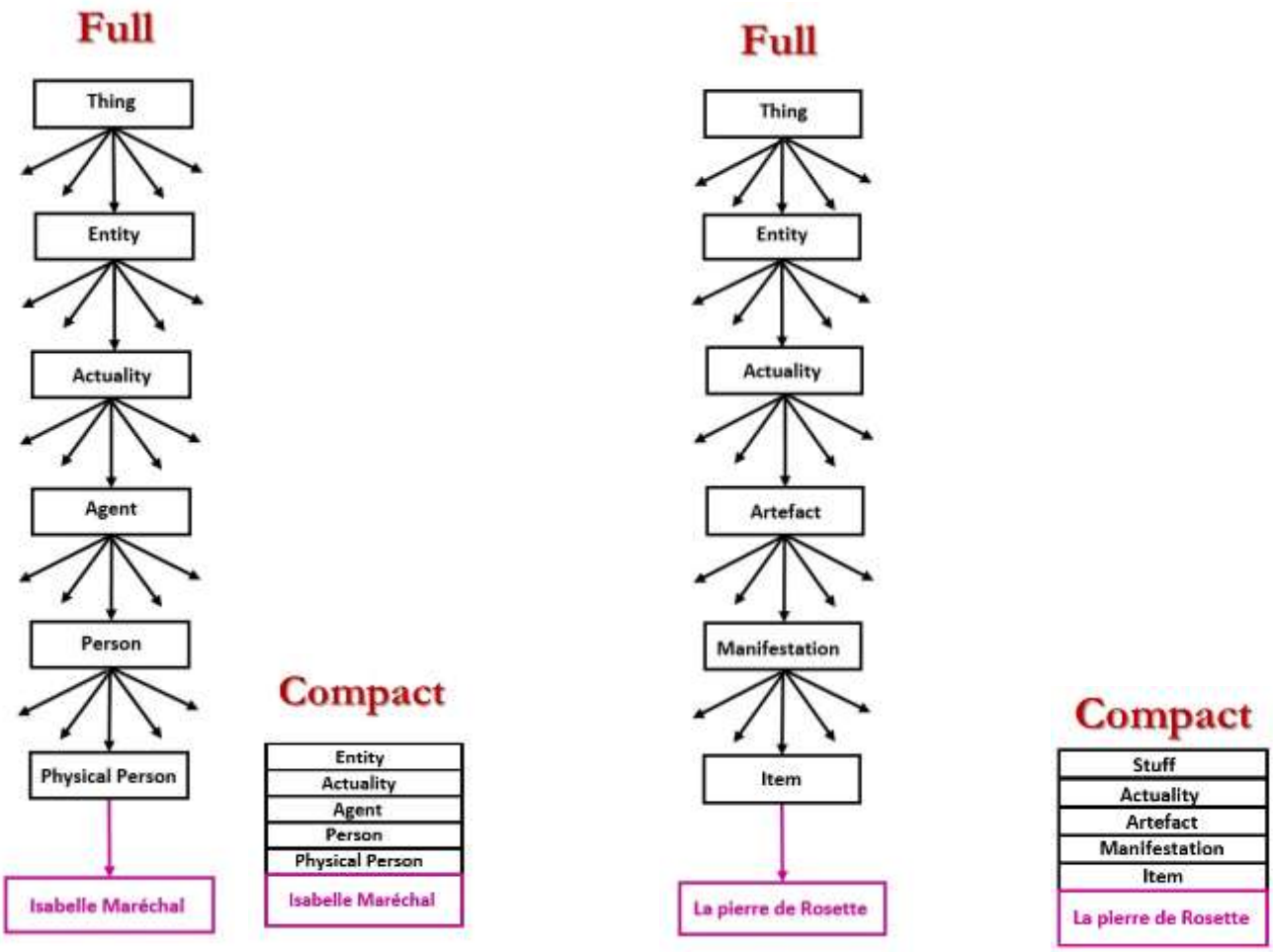
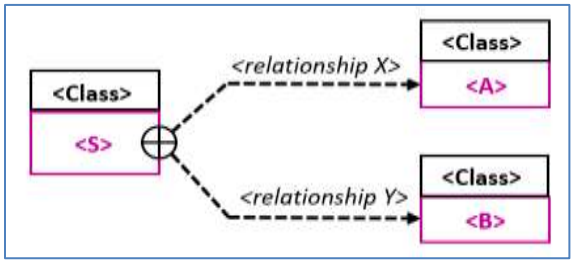


Figure 6-62: Compact representations of refinements of Classes:

The **Exclusive-Or** symbol is represented by the mathematical sign: ⊕



The Exclusive-Or symbol indicates that at the end of arrowed line represents a situation where any instance of the Entity located at the start of the line can be associated, through one of the relationships, with one, and only one, instance of the Entity linked by the arrowed line.

Figure 6-63: The relationship will be between <S> and <A>; or between <S> and <B>

When the association links two entities through a list of Relationships, one, and only one, of these have to be selected.

The illustration express that the instance <S> can be associated either with the instance <A> through the Relationship <X>; either with the instance <B> through the Relationship <Y>.

A similar case refers to the exclusive selection of Classes:

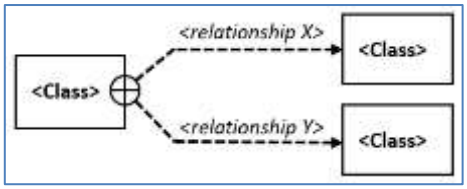


Figure 6-64: exclusive selection of a Class

A similar case refers to the exclusive selection of Relationship. The figure below expresses that the instances of the association imply to select either the Relationship "X" or whether the Relationship "Y". A typical case will be introduced later in the context of the Class '**Register**' where it is required to select the Relationship '**owns'** or the Relationship **'uses'**.

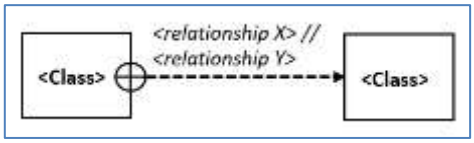


Figure 6-65: exclusive selection of a Relationship'

A dotted line with double arrows at the end represents a relationship where any instance of the Entity located at the start of the line can be associated with one or more instances of the Entity linked by the double arrow.

The **link to the location of a resource** will be represented by the following symbol:

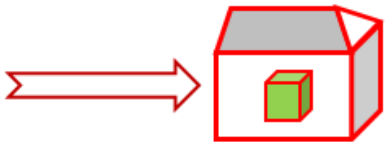


Figure 6-66: Representation of a Resource location

The **link to the location of a file on a resource** will be represented by the following symbol:

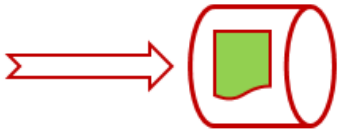


Figure 6-67: location of a file on a Resource

The following symbol represents an '**Agent**'. The dashed arrow illustrates the interaction (for example, the triggering of a process).

Figure 6-68: the symbol of an Agent

The following symbol represents a '**Processing**':

Figure 6-69: The symbol of a 'Processing'

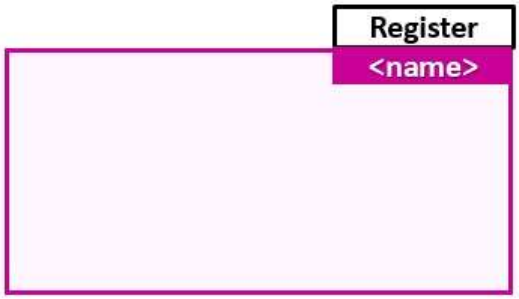


Figure 6-70: The cluster symbol of a 'Register'

6.5.3.3 Statement of Existence of "Entities" & their "Models":

The Statement of the existence of "Entities" and of their "Models" is expressed via the Class "**AxisFootPrint**" [abbreviated "AFP"]. This Statement will be the pivot, the support of their 'registration' / 'cataloguing'; their representations; their location and the location of their representations; the representation of the history, stages of the 'life' of things; ...

This Construct is a very general bearer. It is the support of what, **in the context, is by nature persistent** for Entities and links to the various representations of Entities in their life cycle and the links and evolution of the links between Entities. As introduced in the previous section, the Entities which are declared to exist are regarded as stable or factual: States of Entities or the Treatments carried out by Entities between two States are represented. This means that what is considered stable in one context may be different in another context. A particular Physical Person remains from birth to death; but if it is described in the marital social context, the relationship "is the spouse of" exists only from the moment of the marriage at the time of the death or at the time of the divorce. The Statement of existence of a natural person will therefore rely on stable or factual elements, such as the 'Date of Birth', the 'Name of Civil Status', the reference to the register of the 'Place of birth' ....

As an example, in the statement of existence of Mr. Wolfgang Amadeus Mozart we mention his date of birth "1756-01-27" to "Salzburg" and died the "1791-12-05" to "Vienna" ; but it will not be mentioned that he was a "composer of music", if the intention is to describe in detail his life.

The following figure represents the detail of the expression of the statement of existence of **a painting** known as Leonardo da Vinci under various names, including "Mona Lisa" and "La Joconde": the target of the declaration of existence is the painting! Not the model represented on the painting!

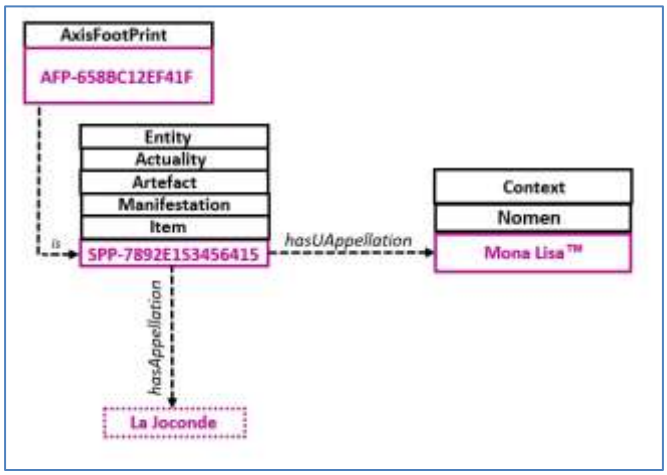


Figure 6-71: the "Mona Lisa" and "La Joconde" Statement of Existence

- The instance of the class "AxisFootPrint" has a value of "AFP-658BC12EF41F" for the URI.
- The class instance that is declared to exist is an "Item" that has a URI value of "SPP-789E1S3456415"
- Two of the metadata are represented:
  - The Property "Unique Name" which has the value "Mona Lisa".
  - The "Nomen" class instance which imagines that this Item has been afforded legal protection expressed by the reference to a Trade-Mark.
- Other metadata should be added to mean that it is an oil painting, painted on wood panels ... but the illustration is enough for itself.

The preceding figure can be represented more compactly or simplified:

Figure 6-72: Compact representation of the statement of existence

The AFP Construct is not applicable inside an instance of a "Modelling Technology".

The need to add an explicit statement of the existence of Entities is not obvious! But the need for a Construct capable of managing the variety of data types associated with something quickly showed the relevance, appropriateness and flexibility of introducing this 'AFP' class 'AxisFootPrint'.

The following figure will introduce this problem and explain the uses of the 'AFP' class. It should be remembered that the 'Modelling Technology' Class may have refinements, as introduced in the section 6.5.2.2.

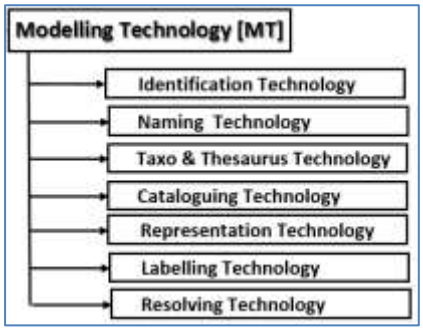


Figure 6-73: the 'Modelling Technology' Class refinements

The general pattern of use of the 'AFP' Class is shown in the following figure:

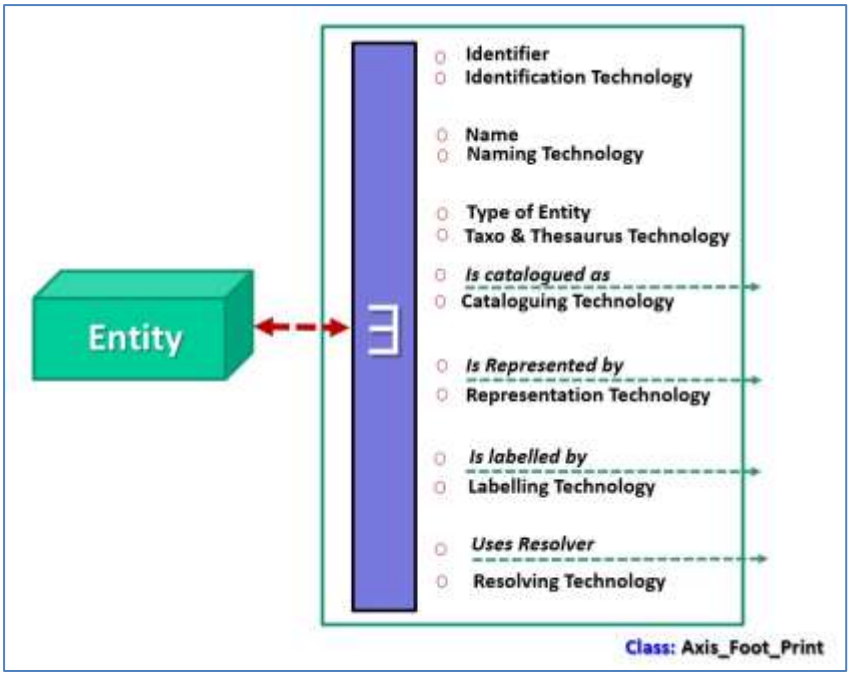


Figure 6-74: the general AFP Class

The mathematical symbol ∃ means the Statement of the existence of an Entity.

Instances of the 'AFP' class are the bearers of links to various ways of expressing data about Entities.

The following figures express the fact that AFP's can be the bearer of the relations that exist between Entities.

The following figure illustrates two types of links: the presence of an Entity in a catalogue (Like 'Parts' in a 'Register'); the existence of several representations of the same Entity.

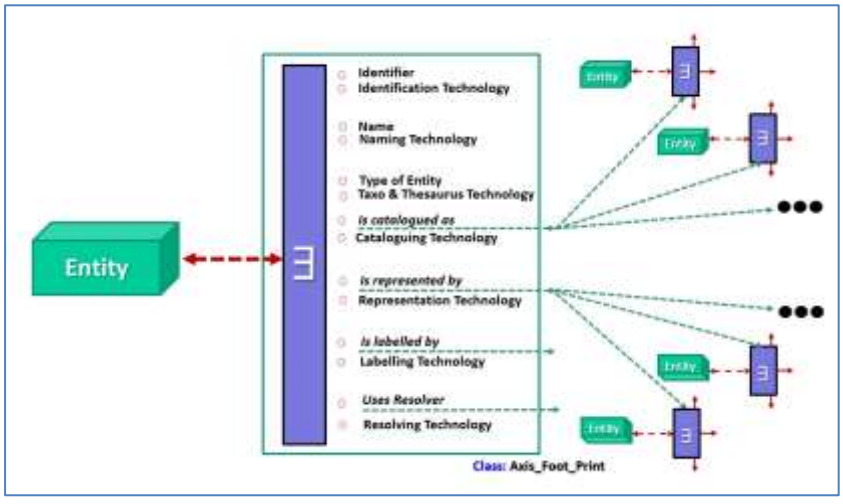


Figure 6-75: existence of several representations of the same Entity

The following figure illustrates that the Entity concerned is equipped with three types of labels (Barcode // QR code // RFID) and expresses the links to the data blocks defining them.

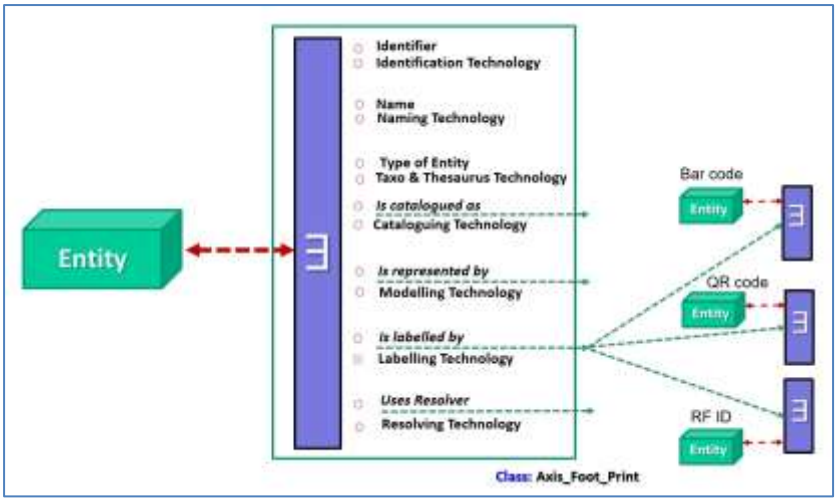


Figure 6-76: three types of Labels (Bar code // QR code // RF ID)

The next figure shows that the involved Entity has cataloguing metadata according to several standards.

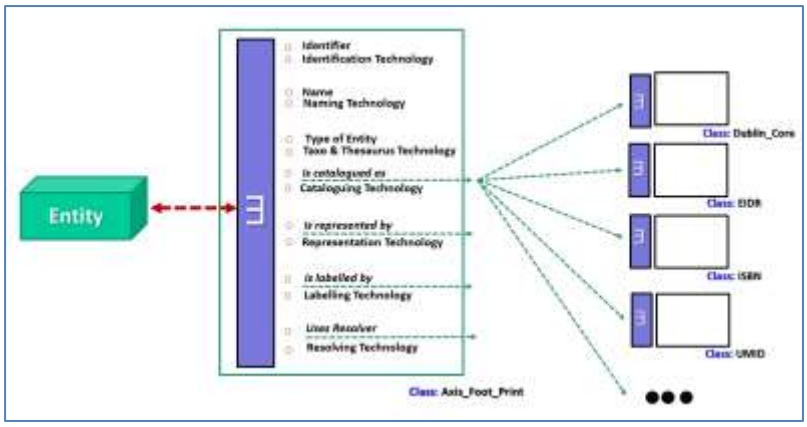


Figure 6-77: Cataloguing metadata according to several standards

The following figure illustrates the flexibility of the concept.

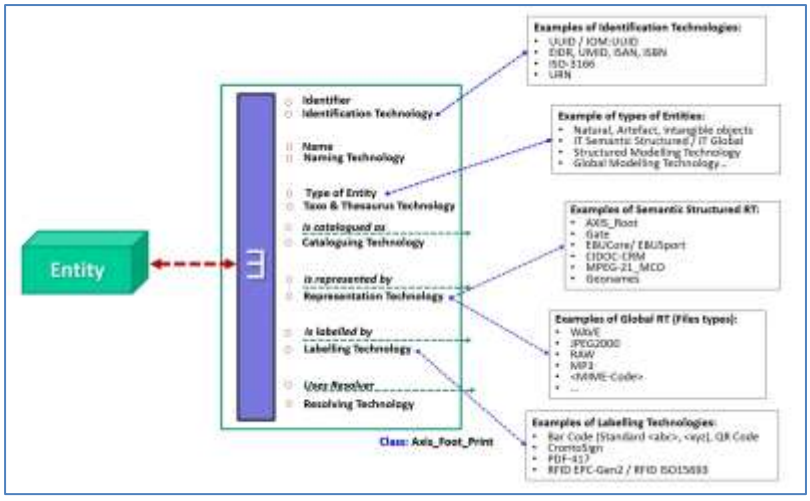


Figure 6-78: examples of the flexibility of the concept

6.5.3.4 The Registers:

AXIS-CSRM organizes the fact of being able to make coherent sets of data representing the statements of existence of Entities, representing Entities and their links. The registers bind Entities together by specifying which one is the 'conductor' in the management of these Representations. This property makes it possible to manage the Rights, the exports and the evolutions in the management itself of the life-cycle of the resources carrying the databases, in particular the migrations.

The next figure illustrates the aim of bundling instances of Constructs pour expressing the grouping of a coherent set as to show the processes involved in the video editing.

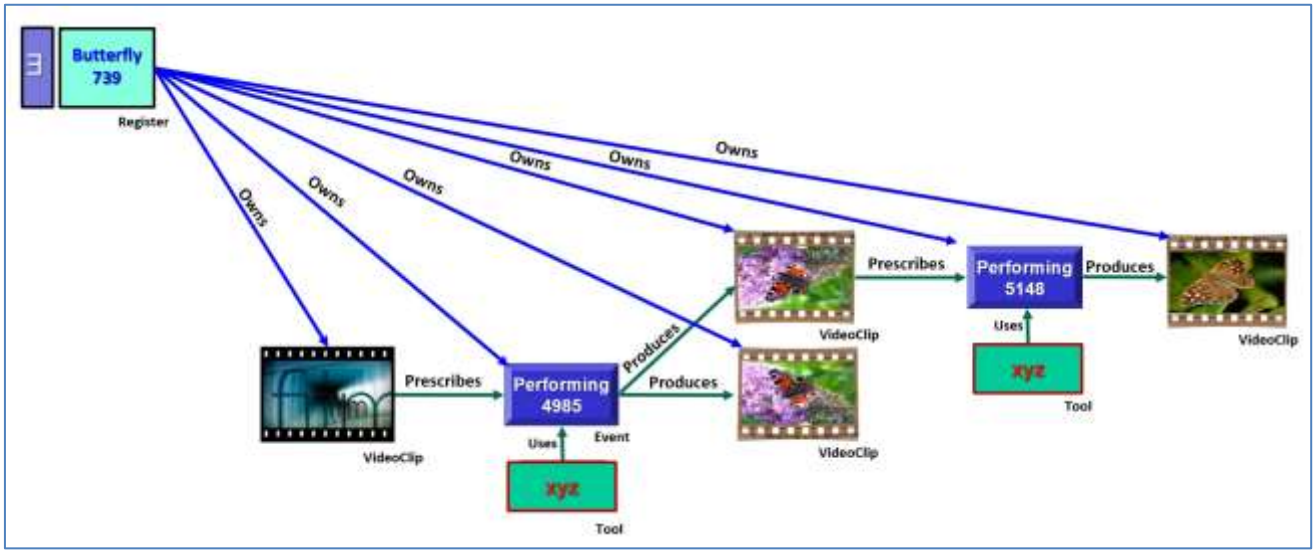


Figure 6-79: Explicit Registers (Owns)

Another example concerns the definition of the selection of what has to be extracted to obtain autonomous and semantically coherent "packages", what is easily expressed by the mechanism of the Registers.

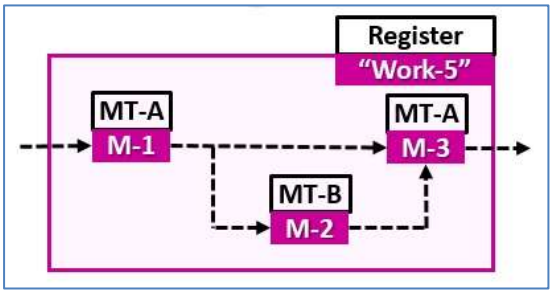


Figure 6-80: Cluster symbolism (Registers)

The figure 6-72 illustrates the symbolism for representing the **bundling** of the instances of **several Models** [each with all their instances of Classes, Properties and Relationships defined by one M**odelling Technologies]** and the links between these models as appropriated for representing (in the example) the '**Work-5**'. It is the way of bundling all the models and their relations pertaining to one intention.

It must be said that the types of 'Relationships' which link the instances of Registers to instances of classes can only be of three kinds:

1. Has
2. Owns
3. Uses

A class instance can be linked to a registry only by a single "Owns".

Imported Constructs instances are placed in an instance of the registry specialization called "Proxy".

The following three figures illustrate typical uses of the "Register" concept.

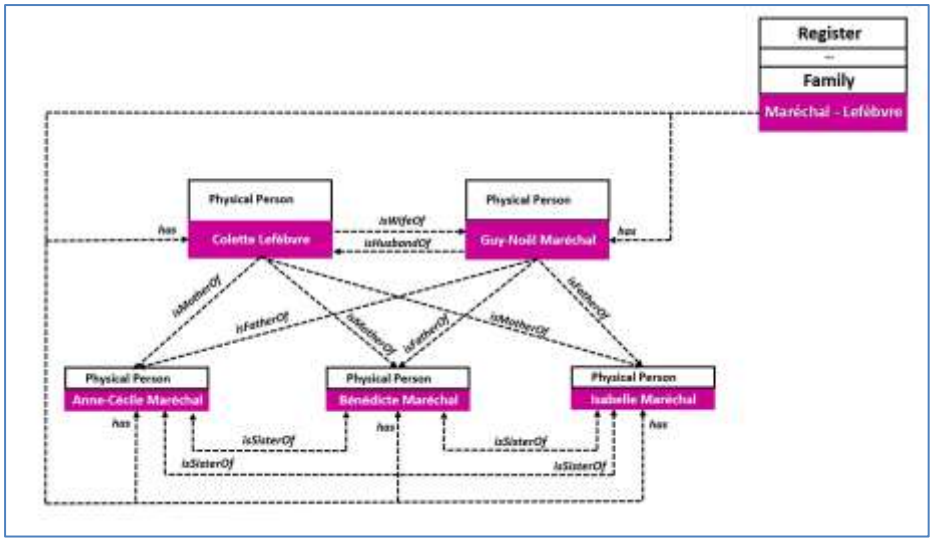


Figure 6-81: Illustration of the Concept of Register

The figure shows the existence of an instance of the "Family" Class (a specialization of the "Register" Class). This instance is called "Maréchal-Lefèbvre"! The figure shows the five 'members' of this Register in the form of Relationships of the type 'has'.

- It should be noted that the instance of the Class "PhysicalPerson" called "Colette Lefèbvre" can also be a member of another "Family" Register called "Lefèbvre-Nols".
- It should be noted that the instance of Class "PhysicalPerson" called "Guy-Noel Maréchal" can also be a member of another "Family" Register called "Marshal-Rennotte".
- It should also be noted that one can also attach to these records the scans of the 'Marriage Booklet' and the resolving model to know in which place this 'Marriage Booklet' is located.

The figure also shows the Relationshipships between these five members: "*isWifeOf*" ... "*isSisterOf*". This set allows inferences to be made.

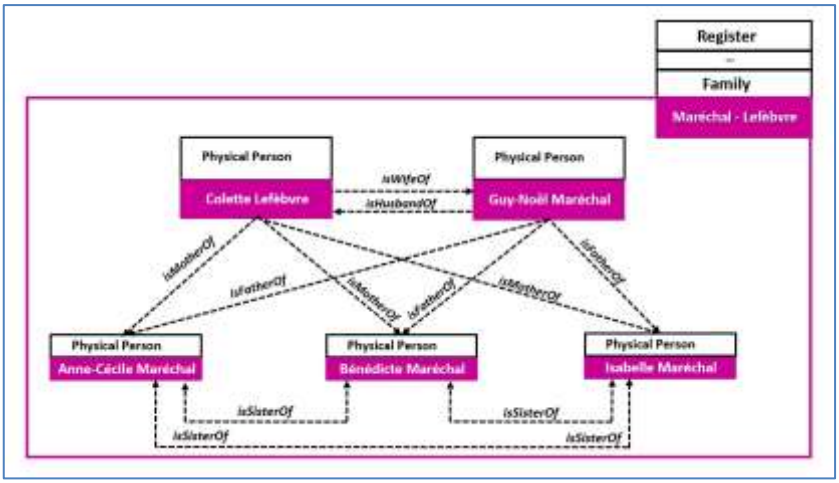


Figure 6-82: Illustration of the cluster symbolism for the Registers

This figure has the same signified as the previous figure but omitting to represent the relationships of the type "*Has*". The purple box brings together all members of the registry.

The following figure describes another register than the previous one, an identity card and its representation. This other register is a 'proxy' of the official register kept by the communal authority; which is the entity producing the Identity Card (as an object) which deposits it in use to its holder (in the example, Guy-Noël Maréchal).

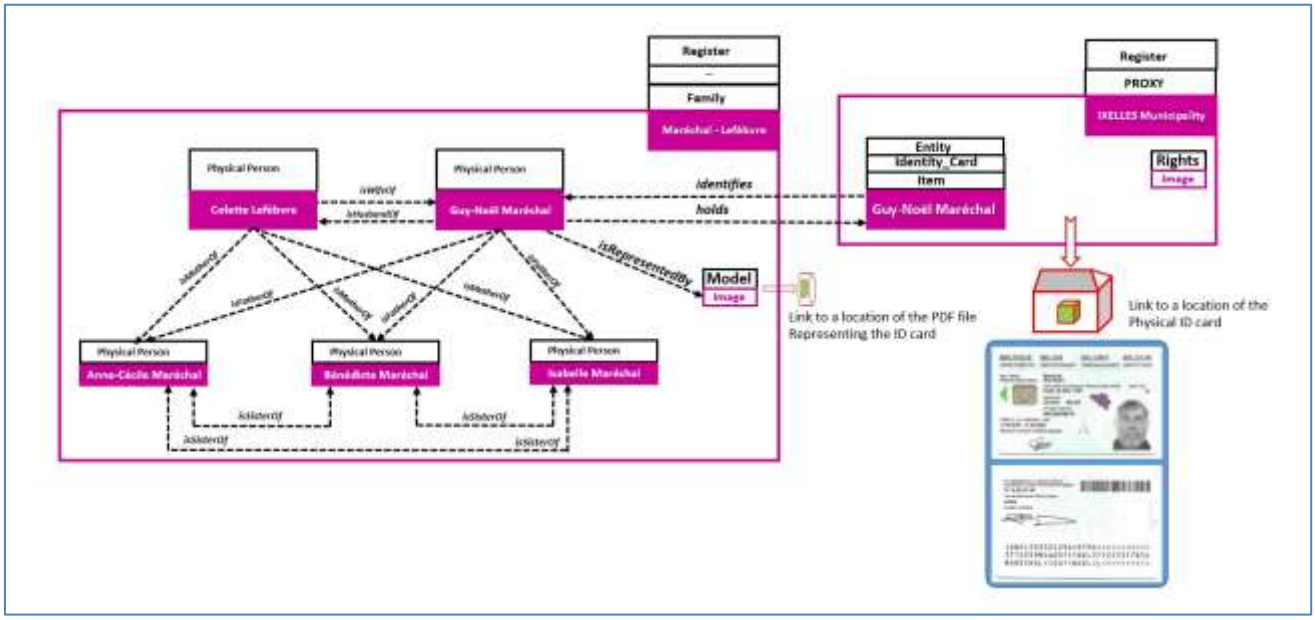


Figure 6-83: Illustration of flexibility of the Register concept

6.5.3.5 "Finite State" modelling: "Segments" & "Points":

In 6.5.3.1 [General Modelling of the Entities], the concept of "Finite State" modelling was introduced to take into account the management and representation of processes and their interactions with "Entities" and models ".

With regard to the "life-cycle" of the processes, one usually distinguishes several phases:

- Process definition
- Presentation of this definition to resources (human and / or machines) and a context capable of transforming this definition into an action. This definition is called "Prescription".
- Activation (manual or automatic) of the Resource in such a way that the process executes, realizing the treatment that was planned or having difficulties to carry out this treatment and or not doing it at all.
- Usually, the process begins with identifying and recording a starting 'State Vector' and an End Process 'State Vector'. These state vectors control the sequence of processes (via conditional activation) following a programmed sequence of processes [Work-Flow]. In particular, in case of problems, an Exception Handling process will activate either the stop of the treatments, or an adequate backup treatment.
- These processes lead to results in various forms: these activation of adequate resources, according to associated protocols, will be able to generate an artefact flow (the sound coming from a loudspeaker ...) or an artefact object (a 3D printing ...) or modifying a "Entity" or creating a new model or ....
- Recording Work-Flow and State Vectors for archives, traceability, quality analysis ....

It should be noted that computers can only represent **facts, concepts** or **instructions** in order to declare the existence of things or to model them in as many states of their evolution as necessary, desired, relevant, realistic. ... in the current context. IT thus manages this capacity according to the so-called 'Finite State' approach and is itself a finite state machine.

So that we can represent and manage the evolutions of Entities (in time and in space) and their links, it is necessary that the computer 'Constructs' representing these evolutions have the same power of representation as the 'Constructs' necessary to represent the Entities themselves. The power and the properties of representations of spatio - culturo - socialo - temporal evolutions of Entities find their source in the modelling of processes such as the "Petri nets" or extensions to first-order logic as predicate calculus.

In AXIS-CSRM, the carriers of this modelization are the "Point" and "Segment" Classes. A "Point" will be associated with "States" (Space, Time, Concepts ...) and a "Segment" will be associated with the representations of Entities (whether they are considered as "static" or that they are associated with the representation of 'processes').

Usually these concepts are represented by combinations of Properties and Relationships. In AXIS-CSRM these concepts are reified doubly: both carriers of the existence of a state or a process and carriers of models defining these states and processes.

- The "**Point**" is the holder of the existence of a "State" and of the models representing it
- The "**Segment**" is the holder of the sub-models of the representation of an "Entity".

The "**Points**" and "**Segment**" express the structure linking the sub-models and the states. They are linked by relationships such as 'isFollowedBy' or 'triggers' which constitutes the expression of the structure (temporal or topological or …) linking the Models and the states. When the model relates to a static "Entities" the segment will be qualified of '**stable**; while when the model relates to a dynamic "Entity", the segment will be qualified as "**evolving**".

An example:

The model of a sound could be represented as an MP3 file; it will be held by a '**stable segment**'
The model of the process of generating 'sound in perception' from that MP3 file will be held by an '**evolving segment**'
The activation parameters and the contextual aspects will be held by a '**point**' linked to the 'stable segment' for expressing the 'prescriptive' nature of the MP3 file and linked to the 'evolving segment' for expressing the 'activation' of the process.

The concepts of "**Point**" and "**Segment**" are expressed by Classes! They can be used after that the models have been made available! This means that the approach is 'upward compatible':

- These concepts could be linked in hierarchy through the relationships between 'Points'.
- These concepts could be the used as the structure of Workflows, Petri-Nets and similar modelling of processes
- These concepts could be used for managerial representations of the projects

Examples will be presented below illustrating these capabilities and modalities. The "Point" is the generalization of the event concept and the "Segment" the generalization of the concept of the situation to cover representations of the processes themselves. For example, the "Point" may represent a "milestone" in the programming of a project. The "Segment" can be oriented (to represent graphs, for example): it is the relationships that link the ends of the "Segment" to "Points" or "Segments" that express these potential vectors.

The "Segment" and "Point" classes are also the carriers of the synchronizations between details within the Models (for example, the synchronization between video and sound), as well for the synchronization of the audio of the comments made by a journalist at one place commenting a filmed sportive event occurring at another place.. Likewise "Segment" and "Point" can be the bearers of the management of resource evolutions that serve the data. In the figure below, the various 'Fragments' are attached to 'Points' [punctual fragments] or to 'Segments'[period fragments].

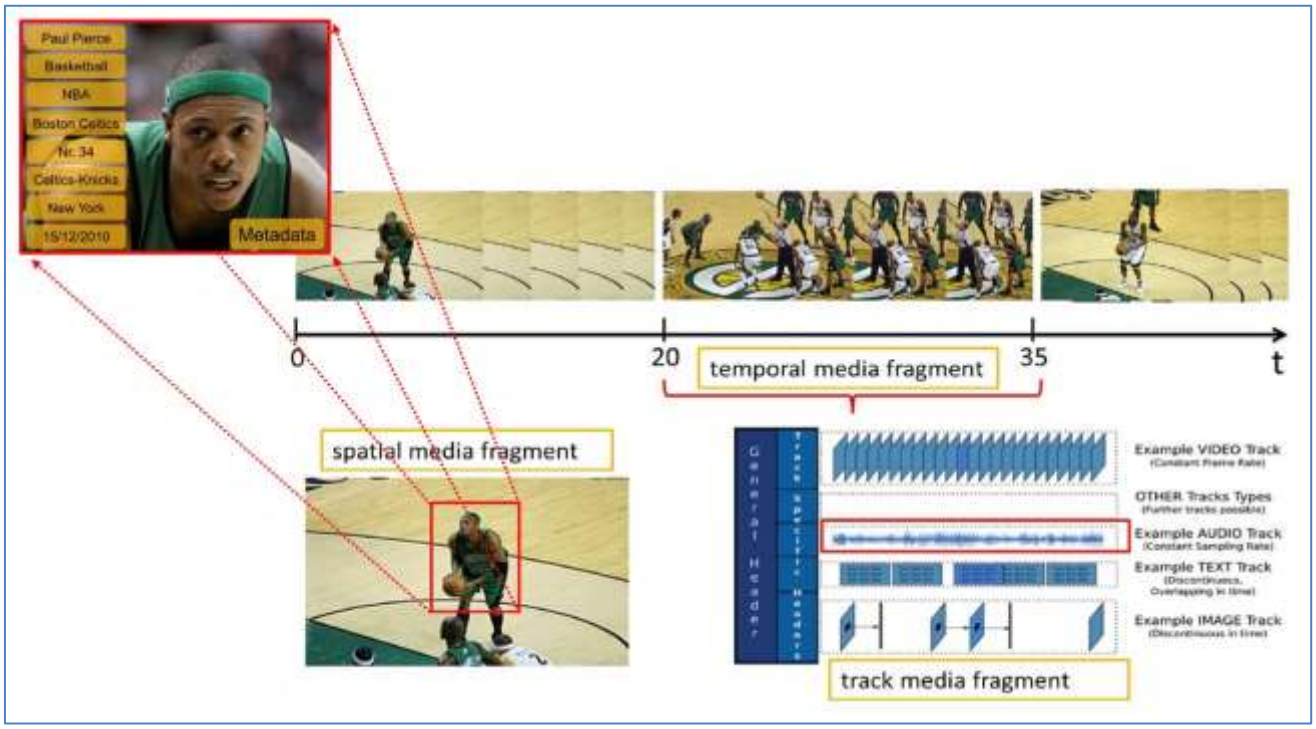


Figure 6-84: Example of use of 'Point' and 'Segment' for structuring media

The following figure shows the structural articulation of the constructs. The examples presented in the next sections will illustrate typical cases of use of these Constructs:

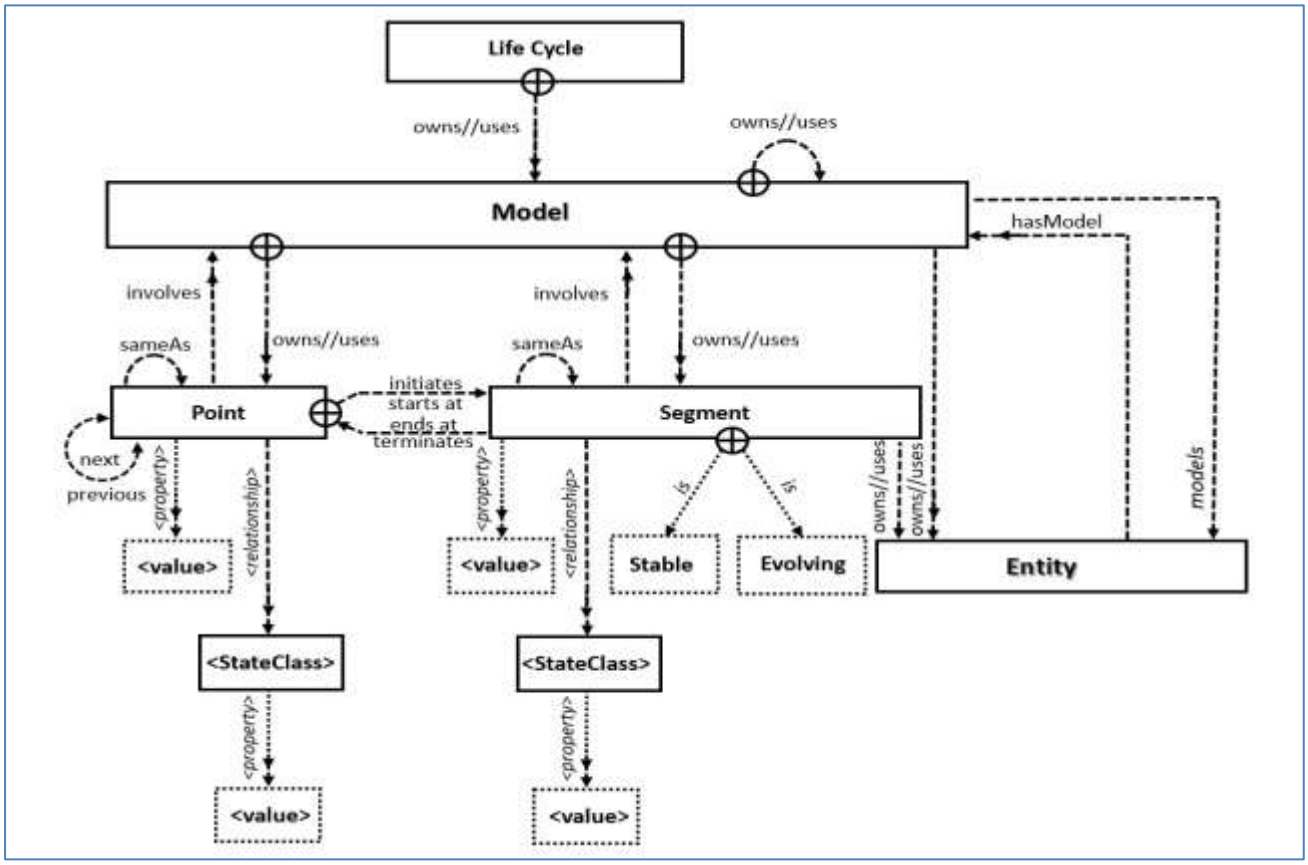


Figure 6-85: Top level data model around the classes 'Point' and 'Segment'

6.5.3.6 « Entities » and « Metadata » Modelling:

The Entities are not independent of each other and this world is constantly changing. Usually, and particularly in the world of archives (the OAIS standard is a significant example) each thing is supposed to be provided with metadata and a representation media file. In fact, representations are more detailed metadata. The deep nature of the world is evolutionary in a space-time and all the Entities that we recognize as disjoined are connected to each other. Moreover, the statements of the existence of Entities and the models of Entities are not stable (errors, contextualization, enrichment of representations and links ...). We are therefore led to model Entities in various ways.

The concept of "**Modelling Technology**" has already been used to define the formalism used, first to code the **URIs** identifying instances of Classes; then to represent and code the instances of the '**Properties**' and finally in the context of statements of existence.

The concept of "**Modelling Technology**" will be used to define the formalism used to represent the "Entities" and the "Metadata" that are hooked to the statements of existence.

The "Metadata" modelling.

The modelling of "Metadata" has been met in a simple way in the case of Properties instances directly associated with instances of Classes of Entities. The representation of metadata via ontologies is well known and available in open source. The key is to choose only the Properties and Classes that have a stable character in the given context.

An example is illustrated by the following figure:

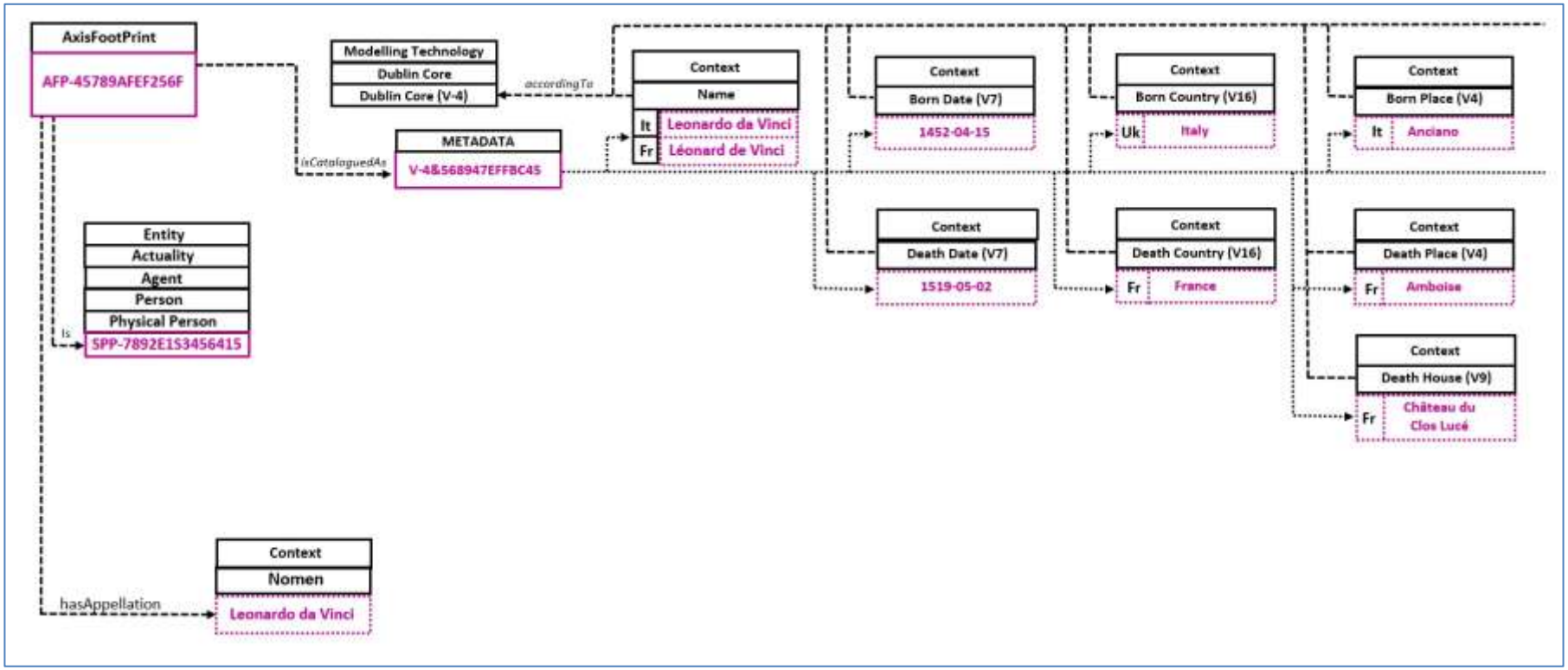


Figure 6-86: Example of a statement of existence of 'Leonardo da Vinci'

A second way of declaring existence in the local context is to add the reference to the Baptismal Register.

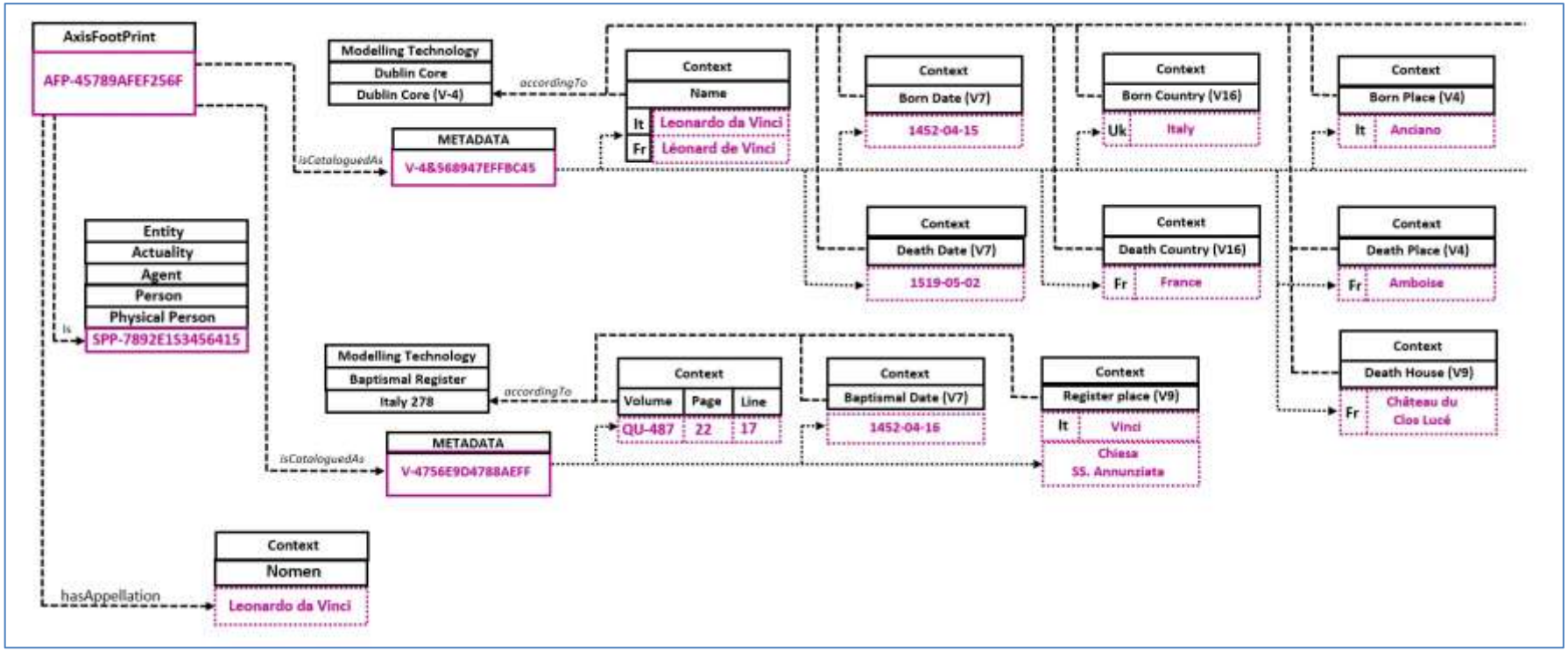


Figure 6-87: Example of two different ways of declaring the existence of 'Leonardo da Vinci'

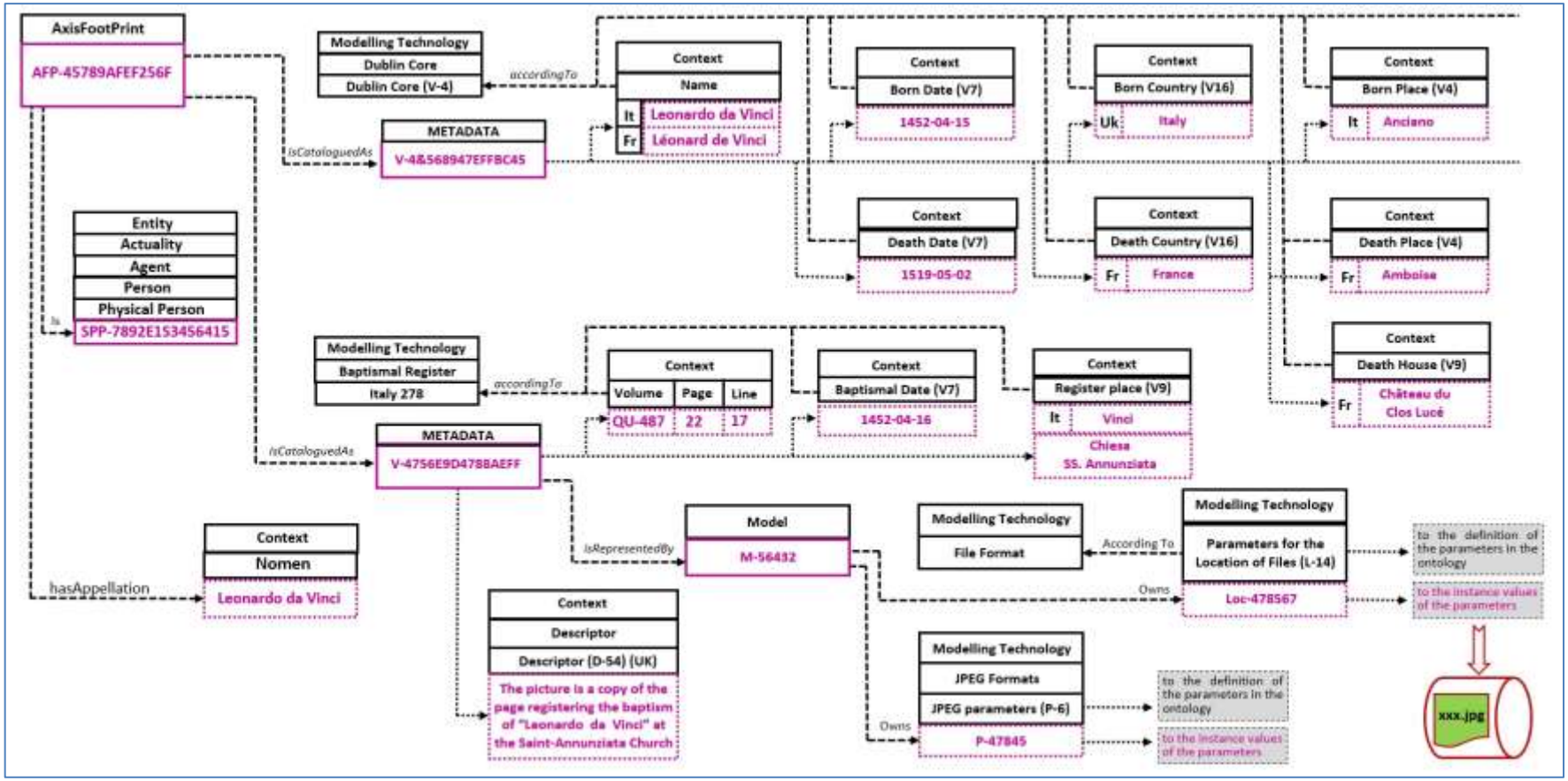


Figure 6-88: Example of linking to the representation of the page of the Baptismal Register

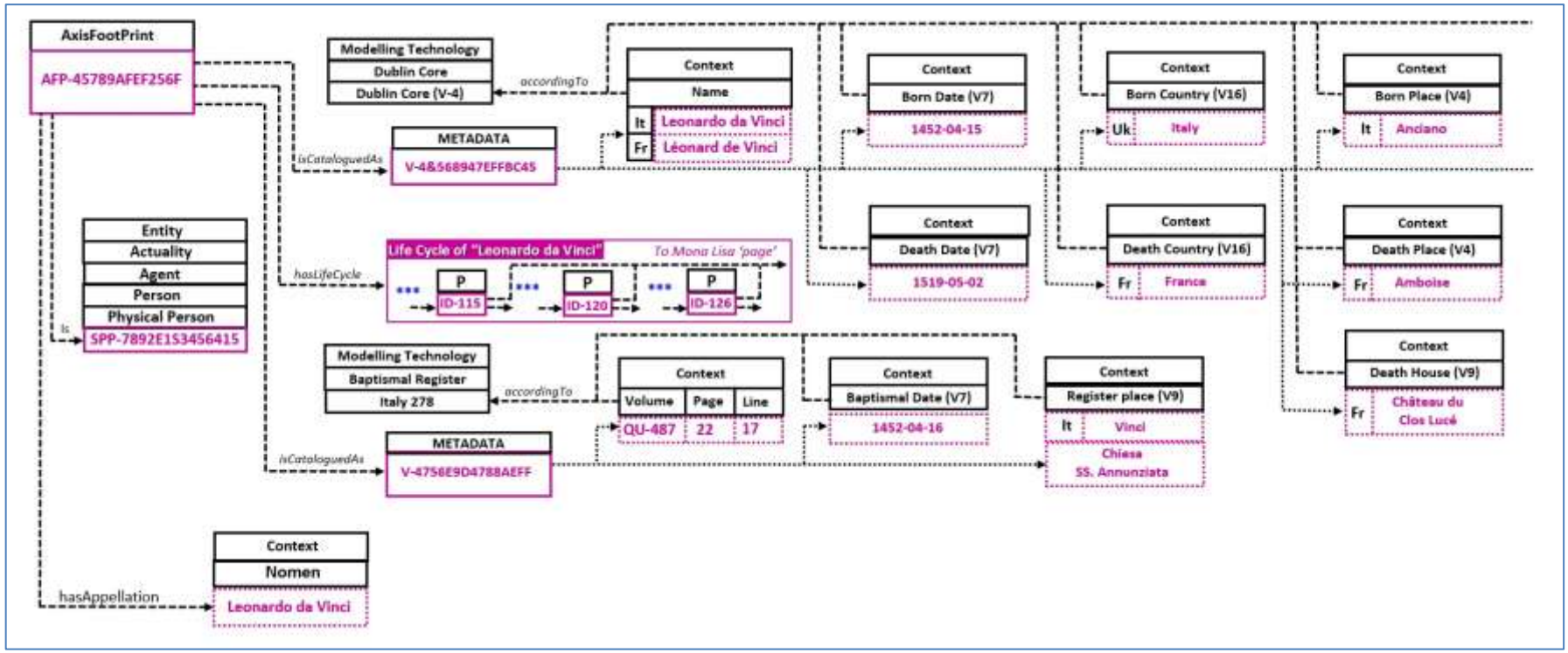


Figure 6-89: Example of linking the 'Register' containing the 'Life Cycle' to the person 'Leonardo da Vinci'

6.5.3.7 The modelling of « Entities »:

The next figure will be progressively explained:

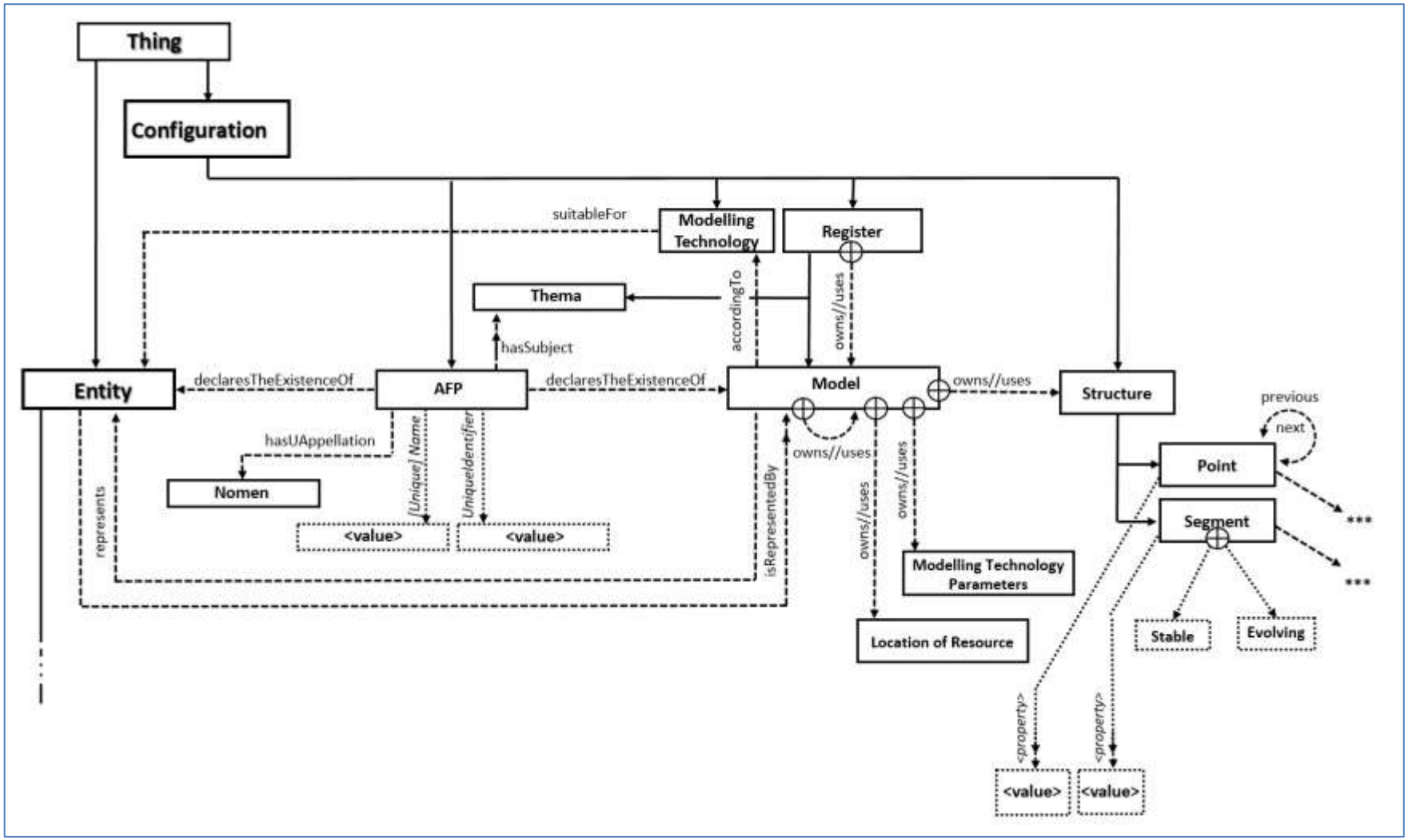


Figure 6-90: The modelling of « Entities »

The following figure illustrates the modelling of the Life Cycle of an Entity with Points and Segments:

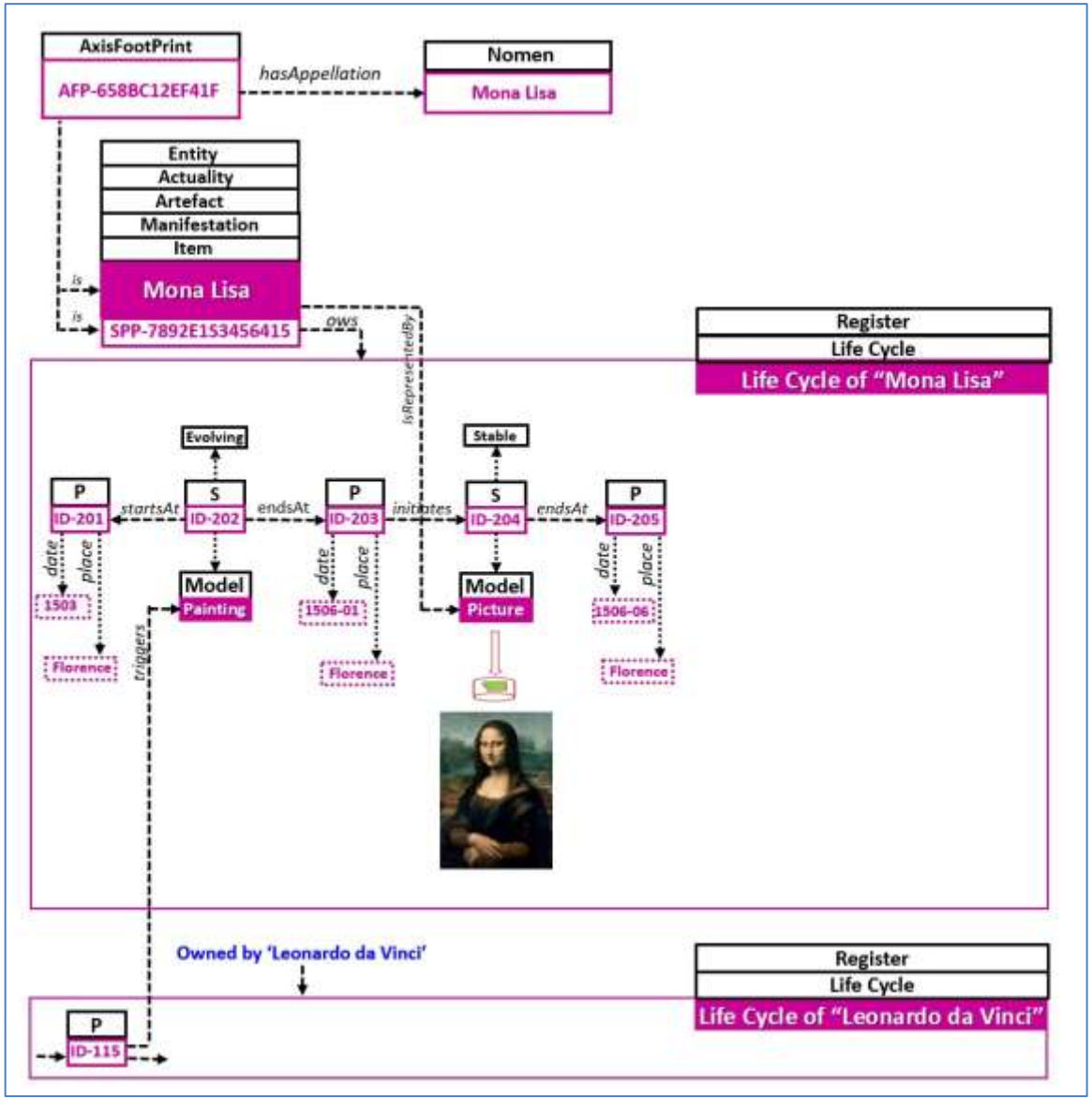


Figure 6-91: The modelling of the Life Cycle of « Entities »

This figure will be completed and explained further in the sections that follow.

For Entity modelling, two cases can be distinguished:

6.5.3.8 Full Semantic Models:

These are cases where the definition of "Modelling Technology" can be expressed exclusively using the three types of ontological Constructs (Classes, Relationships and Properties).

An example of such a case would be the complete semantic transcription of a play: the ontology defining this way of modelling the complexity of such a text is quite immediate to imagine. An illustrative example of such a case will be proposed in one of the following sections.

Another example would be to represent a **picture** (inspired by the ADOBE DNG file format [an open version of RAW format]) as:

- an instance of the **Photo Properties** Class agglomerating a group of metadata expressed as Properties: Device / Focal / Aperture / Location Coordinates / Moment / Shutter Speed / ...
- an instance of a Matrix class specialization that would represent a pixel array ["Pixel Matrix"]: for example, the 4K TV, cinema and photo format has a resolution of 4,096 pixels per line and 2,160 pixels - lines per matrix. The colour of each pixel would be coded by three numbers, for example:

  Green: 16 bits - Blue: 16 bits - Red: 16 bits

  Such a resolution therefore gives 4096 x 2160 = 8,847,360 pixels by 'Photo'; or 8,847,360 x (3x16) = 4,246,733 280 bits or even consuming six Bytes per pixel (48 bits to encode 48 bits) 53,084,160 MB per photo.

- An instance of a '**Matrix**' class specialization that would represent an **'Iconic Vignette'** of the photo as a low-resolution pixel matrix ('Pixel Matrix'): for example, 256 pixels per line and 135 lines per matrix. The colour of each pixel would be coded by three numbers, for example:

  Green: 4 bits - Blue: 4 bits - Red: 4 bits

  Such a resolution therefore gives 256 x 135 = 34,560 pixels per 'Vignette'; either 34,560 x (3 x 4) = 414,720 bits or even consuming two Bytes per pixel (16 bits to encode 12 bits): 34,560 x 2 = 69 kB per Vignette.

- An instance of a specialization of the 'Matrix' class that would represent the colour calibration table

6.5.3.9 Hybrid Semantic Models:

These are cases where the definition of "Modelling Technology" is expressed using the three types of ontological Constructs (Classes, Relationships and Properties) to refer to a resource external to the Context Database (usually to a File located on a Volume, to give the values of the format parameters associated to this file, to define the resources able to open this file, to give the links to the definition of this format of resource [file ...] external (typically the reference to an ISO or W3C standard or ...) and to ensure the indirection towards this resource via a Resolver (the resources can change places or be replaced or ...).

There are many reasons to keep hybrid models: partially non-semantic representations. The most obvious is that of efficiency! In the previous example, the representation size of an image is 53 MB; if it is represented as a JPEG file it will typically be five times less heavy without significant losses [10 MB] and twenty times less heavy with a still reasonable loss of quality [3 MB].

Another reason is that in many cases the originals of the files exist and must be saved for legal reasons.

For security reasons, models are saved in multiple copies at different locations. Some of the models are analogue avatars, artefacts from models ... Some of the models are represented in proprietary formats. It is particularly relevant here to mention the resources (CODEC or others) able to open them. Alternatively or in addition, the expression of the link to 'Proxy' of suitable quality may offer a palliative in case of impossibility to exploit the original ...

In many cases, the models not only cover representations of "Entities" but also the procedures for accessing an '**Actuality**'. A very high resolution picture, representing "The Mona Lisa", does not replace the act of going to the Louvre to see the painting. In addition, access to a very fragile or very valuable incunable requires skills, precautions, traceability ... which must be codified.

Examples will illustrate these concepts in the sections that follow.

The representation of a model always passes by the use of an instance of the Class '**Model'** which is a specialization of the class '**Register**'. This point will be explained in the next section.

An illustration of the hybrid case is given in connection with the previous example; the link to a model of Leonardo da Vinci's Life Cycle and a representation of Leonardo da Vinci's Baptism Registry page are presented here:

The following figure shows the articulation that will be explained at the section 6.5.4.1

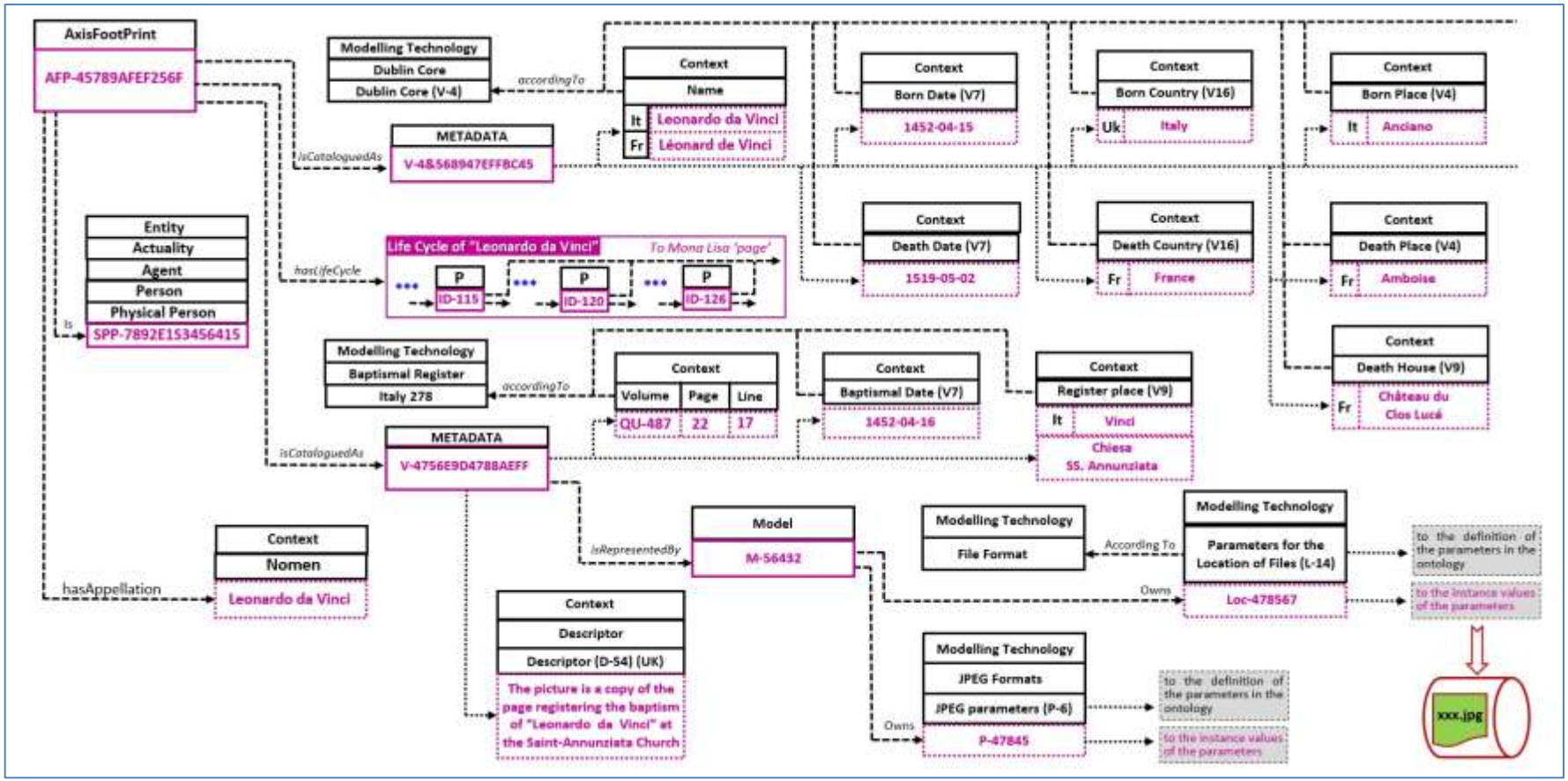


Figure 6-92: The integrated model of the statement of existence of 'Leonardo da Vinci'

6.5.3.10 Mapping the « Data Model » and the « Functional Model »:

These two models are inter-correlated!

The classical modelling is limited to declaring the existence of 'Entities' via metadata and, when this lends itself to memorizing "EdiData" in the form of attached files. In AXIS-CSRM, the statement of existence of 'Entities' has its own Constructs and likewise for the representation of the 'Entities'!

The term "Construct" refers to all the ways in which ‘Things’ can be represented in AXIS-CSRM : essentially, 'Classes', 'Properties' and 'Relationships', their specializations and their occurrences.

In addition, the representation of the 'Entities' can cover the associated processes (in particular the processes that led to the existence of the 'Entities', their 'Enjoying' and also the processes that led to the evolution of the 'Entities' of which declared existence). It is even possible to declare the intention of existence of an 'Entity' before it exists, either in reality or in model: it is an ‘Entity’ involved in the declaration of existence of the intention of creating the ‘Entity’. It is the ‘Project’s modelling.

Here's how to represent the Data Model of one of the Functional Model figures. The simplest, but also the most abstract, is represented by:

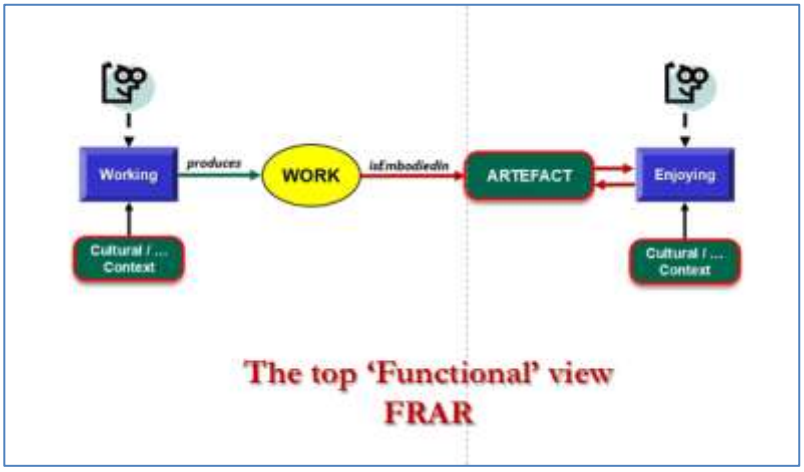


Figure 6-93: the Data Model of one of the Functional Model figures

This data model will be broken down into three steps:

I. The modelling of a <Working> via the process of 'Working' and the intervention of the ‘Agent': a register; two points ; a segment, a model of the 'Working' process, two states and Relationships.

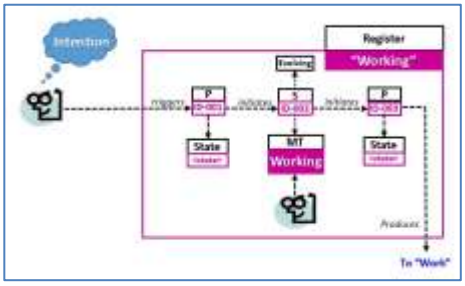


Figure 6-94 : the 'data' modelling of a 'Working'

II. The modelling of a <Work> by a model, two points, a segment and Relationships: a register; two points ; a segment, the model of an artefact 'Work', two states and Relationships.

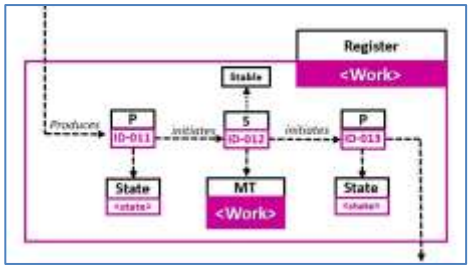


Figure 6-95 : the 'data' modelling of a 'Work'

III. The modelling of a <Enjoying> by two models (the 'Enjoying' process and the 'Agent'), two points, a segment and Relationships: a register; two points; a segment, the model of a process' Enjoying, two states and Relationships.

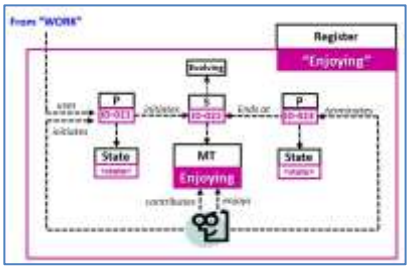


Figure 6-96: the 'data' modelling of an 'Enjoying'

The overall modelling becomes:

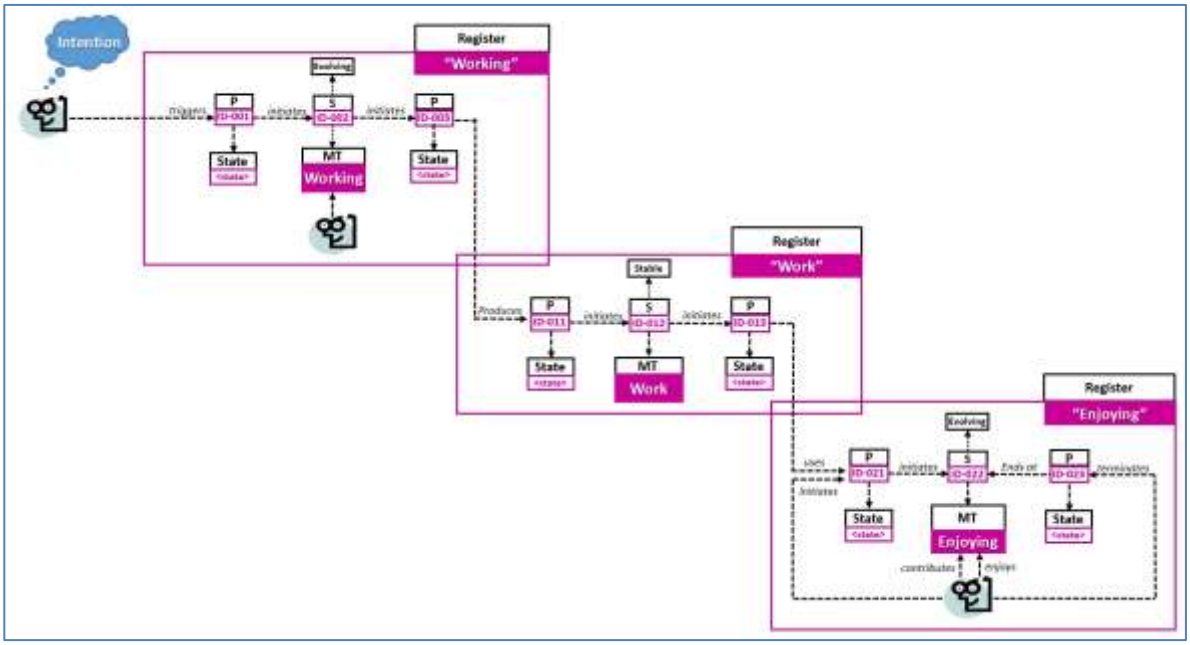


Figure 6-97: the 'Working', 'Work' and 'Enjoying' modelling

6.5.3.11 Actuality

AXIS-CSRM organizes the separation of everything that has to do with the "stable" things of reality.

<W I P >

6.5.3.12 Working

AXIS-CSRM organizes the separation of everything that has to do with the " evolutionary things" of reality.

<W I P >

6.5.3.13 Concepts

AXIS-CSRM organizes the separation of all aspects that are not real. For example, the concept of **"Work"** in the sense of FRBR.

<W I P>

6.5.4 Illustrative examples:

6.5.4.1 Life Cycle of the 'Mona Lisa' painting by Leonardo da Vinci

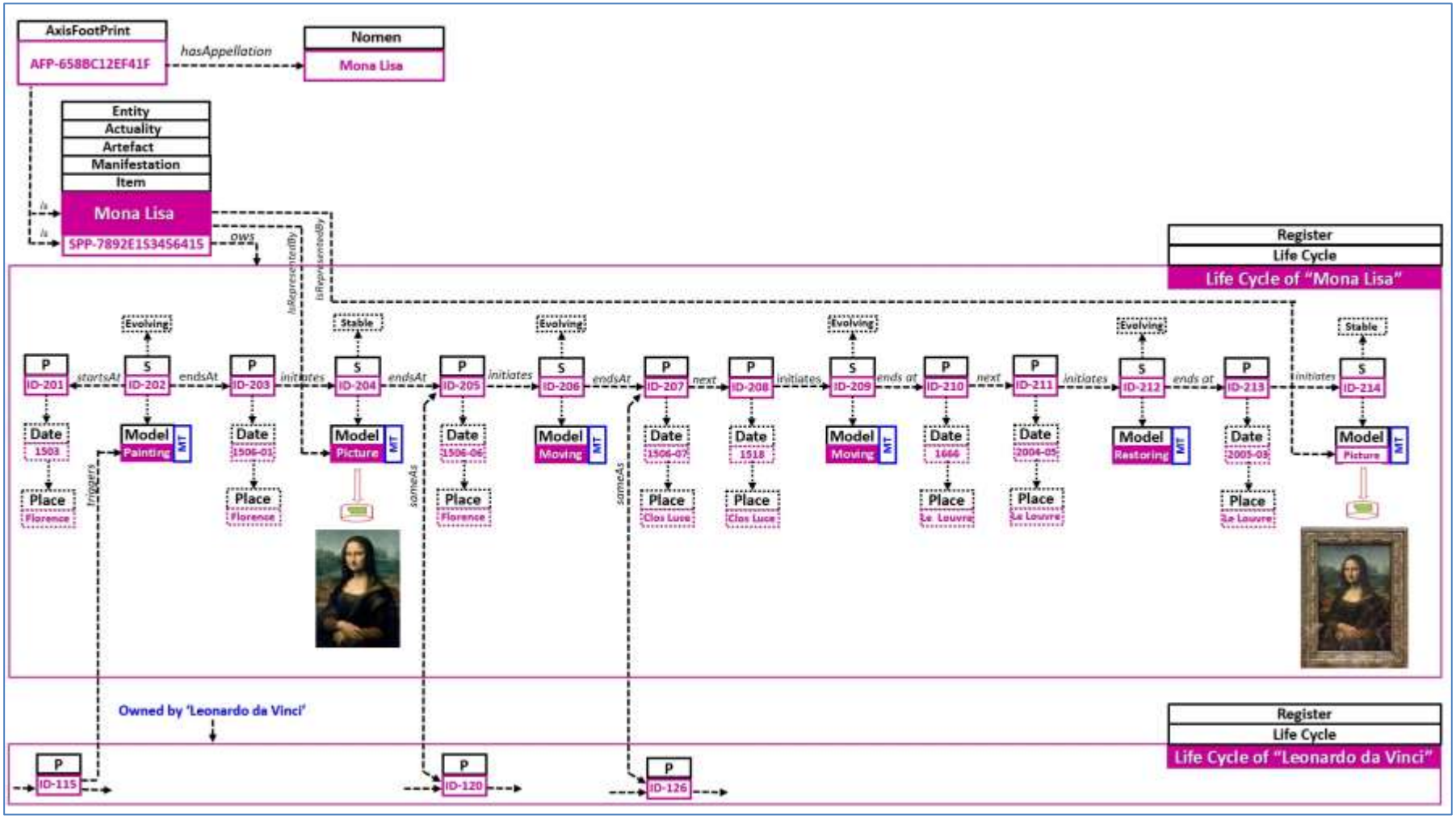


Figure 6-98: Typical modelling of a Life Cycle (1)

1. The figure starts by illustrating the trigger [from ID-115 in the Life Cycle of the life of M. Leonardo da Vinci] of the process of making the painting [ID-202] of a young lady …
2. The second state is the availability of the painting [ID-203]
3. During six months (from [ID-203] to [ID-205]) the physical painting will stay [ID-204] in Florence.
4. When Leonardo da Vinci agrees [ID-120] with the king ‘François I’ to move [ID-206] in his castle of ‘Clos Luce’, in ‘Amboise’, he comes with the painting ([ID-126] and [ID-207])
5. During 12 years [ID-208], the painting remains at the ‘Clos Luce’. Then, it moves [ID-209] to many places up to ‘Paris’ at the palace of the Louvre in 1666 [ID-2010].
6. The painting is presented in the ‘Museum Louvre’ up to May 2004
7. The restoration [ID-212] last up to March 2005 [ID-213]
8. Since then, the painting is again visible [ID-214] at the Louvre Museum in a high quality and protection.

6.5.4.2 Localization of the 'Mona Lisa' painting in the Louvre Museum

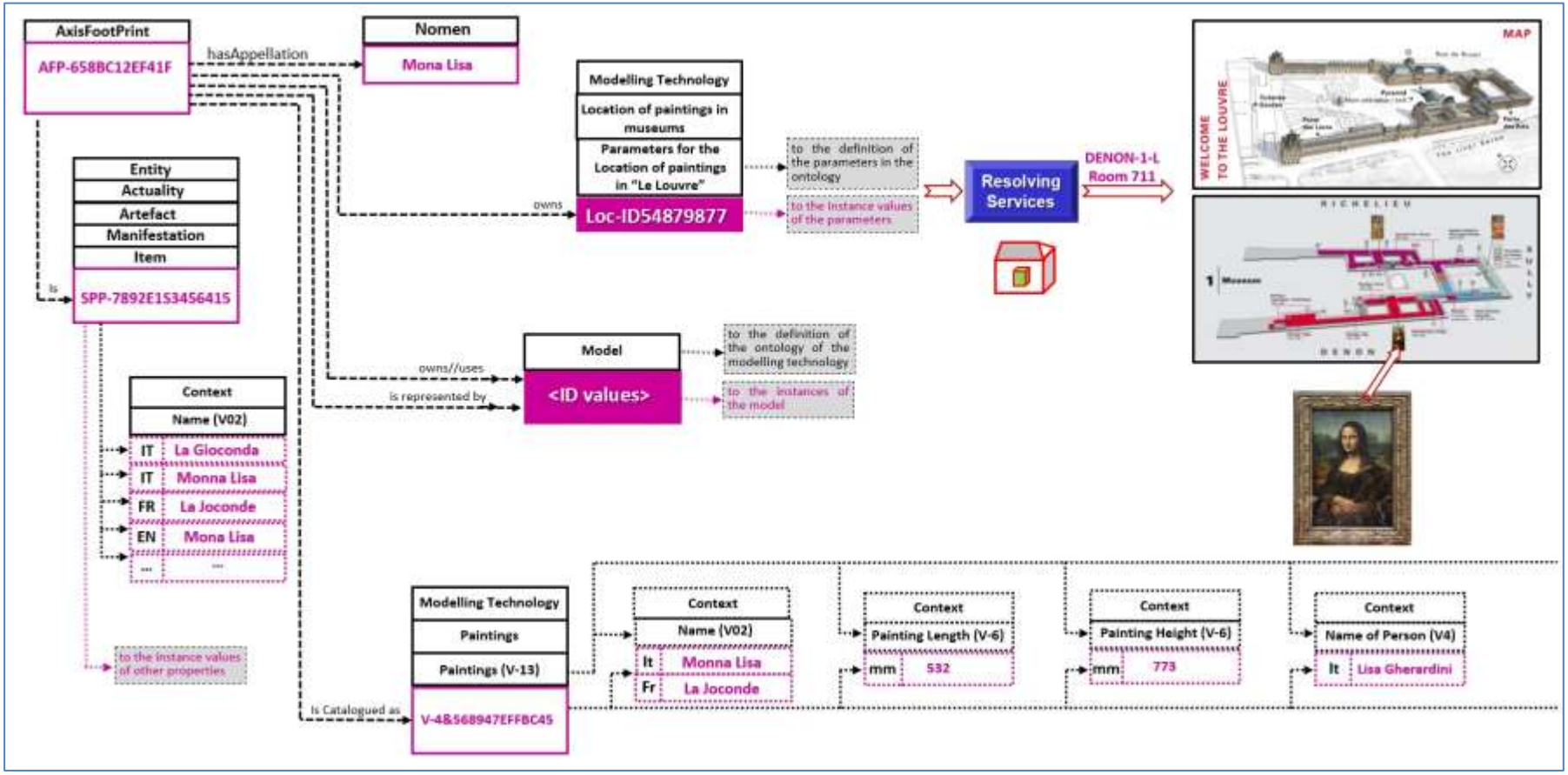


Figure 6-99: Typical modelling of a Life Cycle (2)

The figure illustrates one way for representing the data required for resolving the location of the painting in the Louvre Museum.

6.5.4.3 Exhibition in the lobby of a hotel of a Photography issued from the documentation of the basketball game

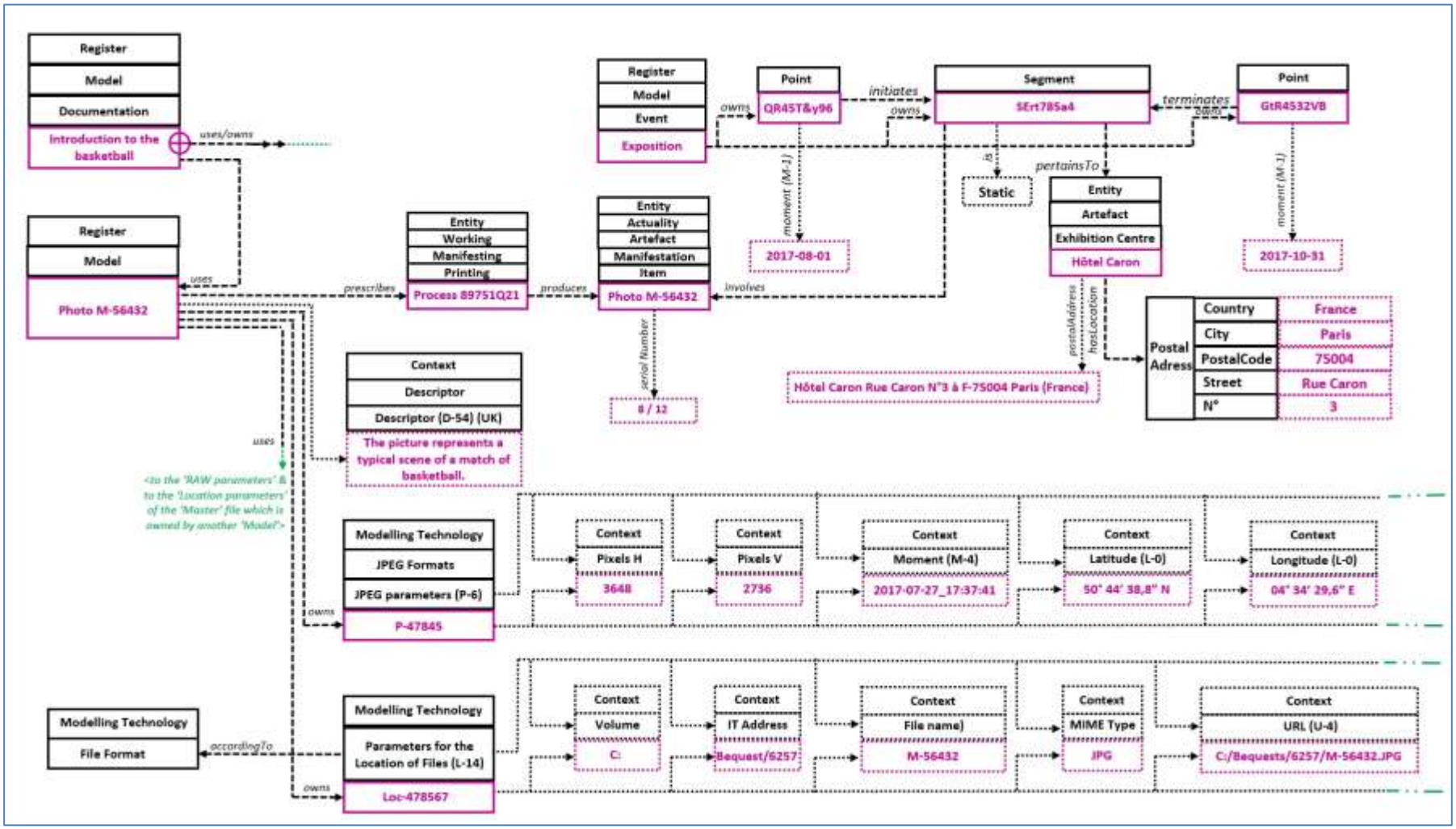


Figure 6-100: Typical modelling of an Exhibition of a object (here a photography) during a limited time.

The figure illustrates the existence of a photography and of its exhibition in the lobby of a hotel in Paris. The photography is also documented as a file.

6.5.4.4 Romeo & Juliet: structure of the Shakespeare Opus

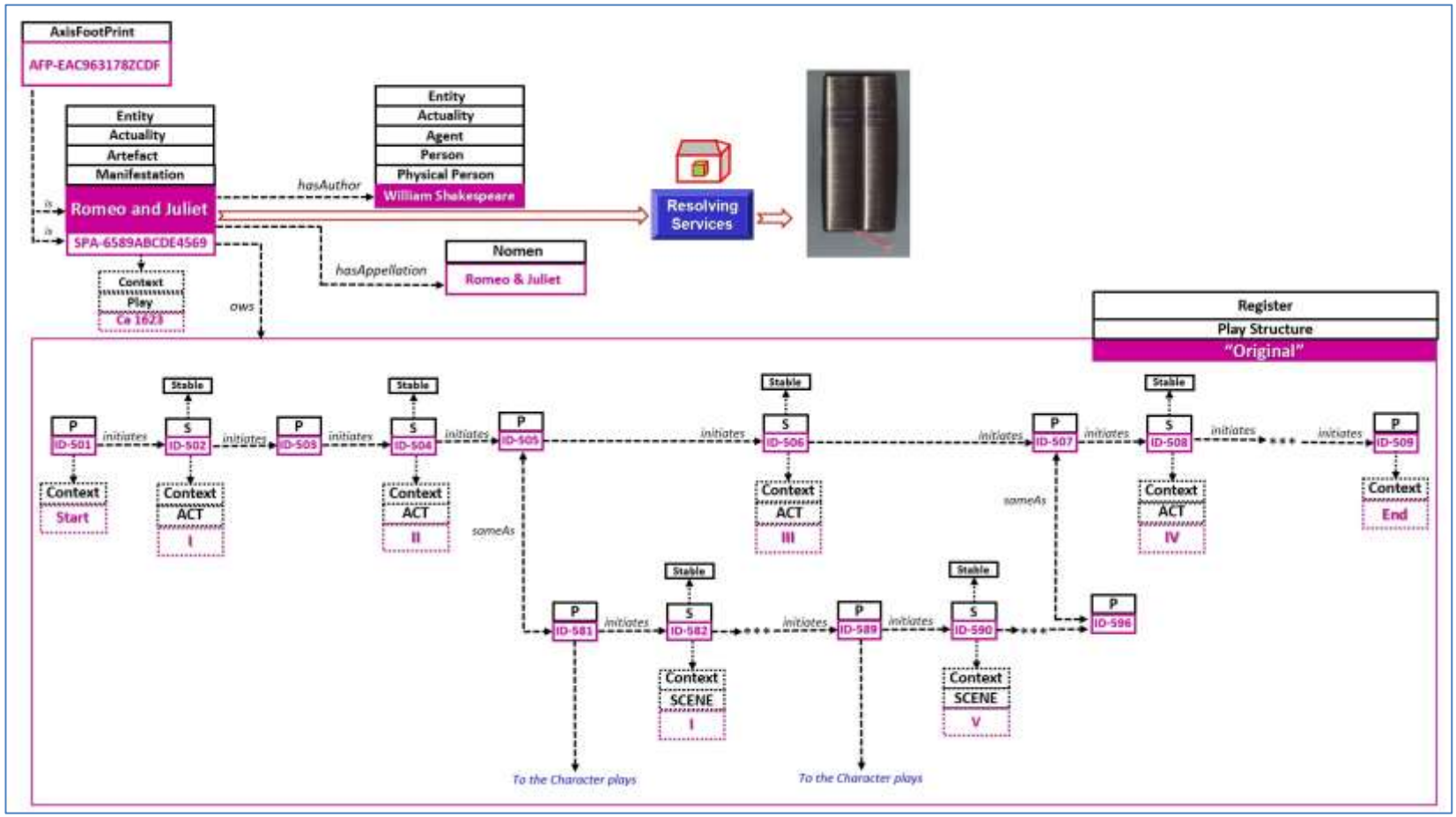


Figure 6-101: Typical modelling of the structure of a theatre opus

The figure illustrates the existence of a sequence of Scenes ([ID-582] to [ID-590] in the Act N°III [ID-506] in the Manifestation of the Work "Romeo and Juliet" [SPA-6589ABCDE4569] by 'William Shakespeare' [a Physical Person]. The data to be performed by the "Resolving service" are not shown and the unique source for the examples is excerpts of a specimen book [Item] in two volumes of the PLÉADE collection [ISBN 2-07-011723-5] from the Gallimard edition 2002.

6.5.4.5 Romeo & Juliet: detailed structure of an excerpt of the Scene V of the Act III

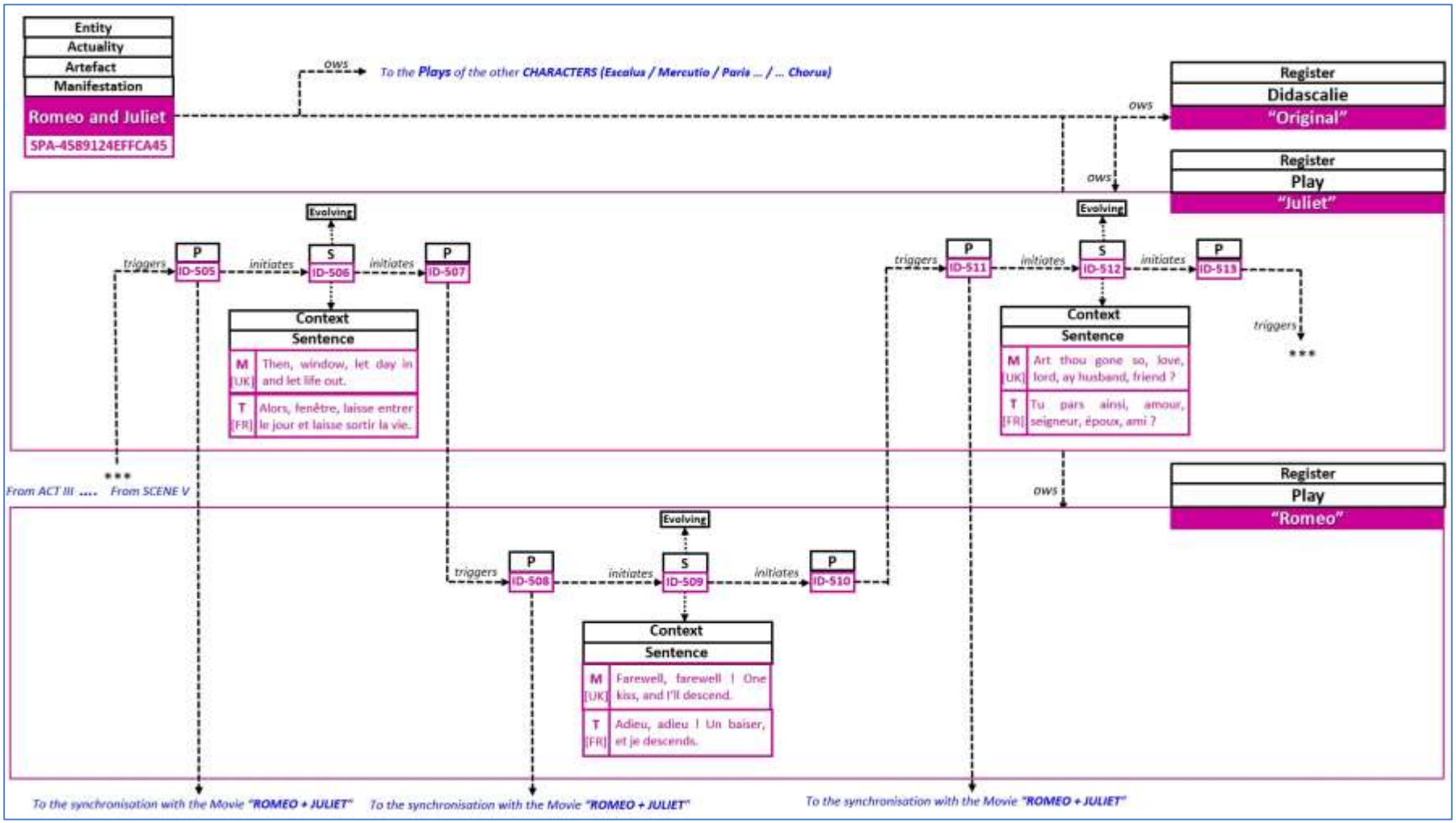


Figure 6-102: Typical modelling of the partition of the Roles and Texts of the Characters in the general structure

The figure illustrates the detailed expression if the partition of the Roles and Texts of the Characters linked in the general structure of the Opus.

6.5.4.6 Romeo & Juliet illustration of the Thesaurus of the Characters

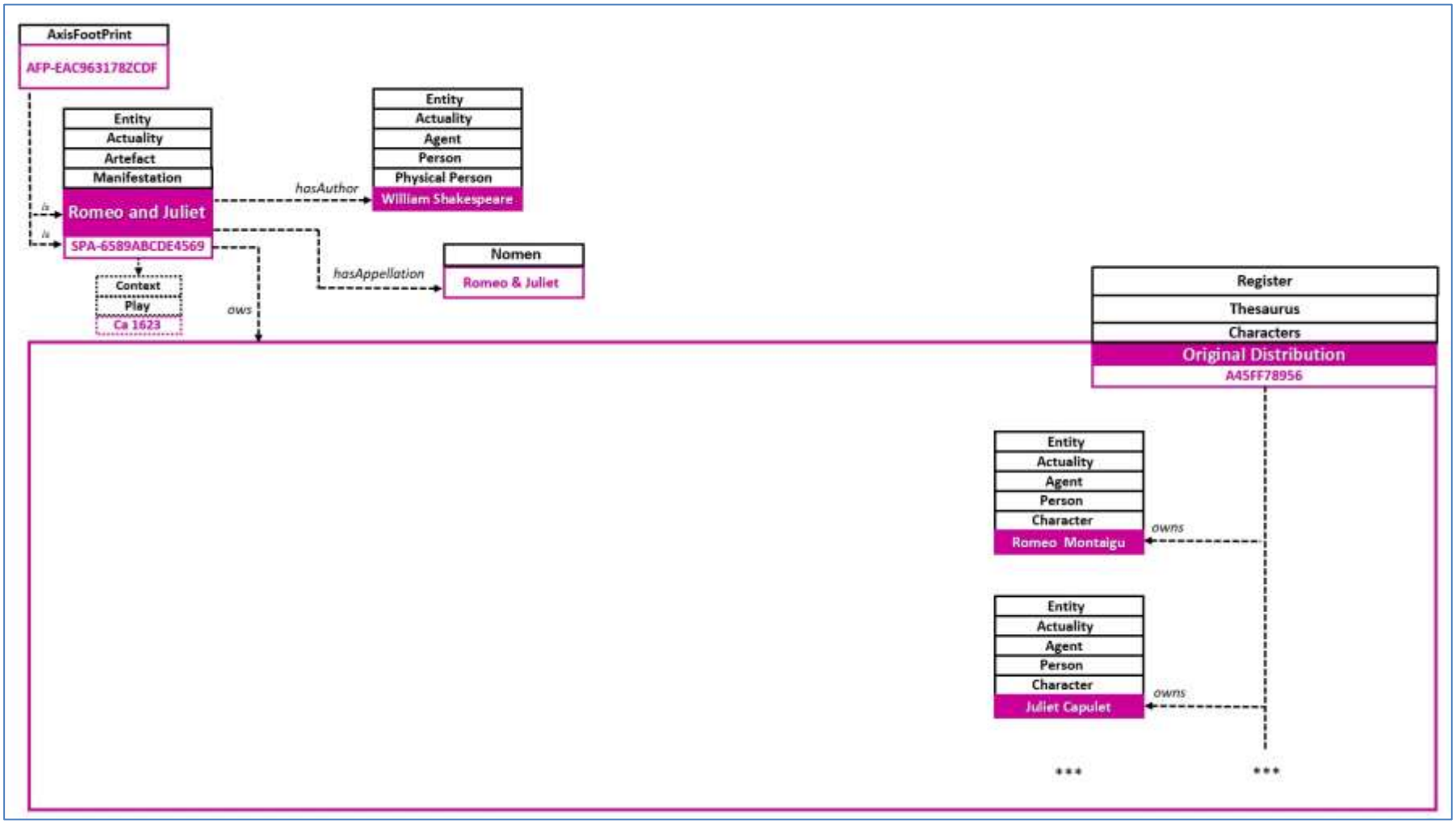


Figure 6-103: Typical modelling of the Thesaurus of the Characters

The figure illustrates the listing of the Characters.

6.5.4.7 Romeo & Juliet illustration of the Casting of the Characters for the production of a Movie

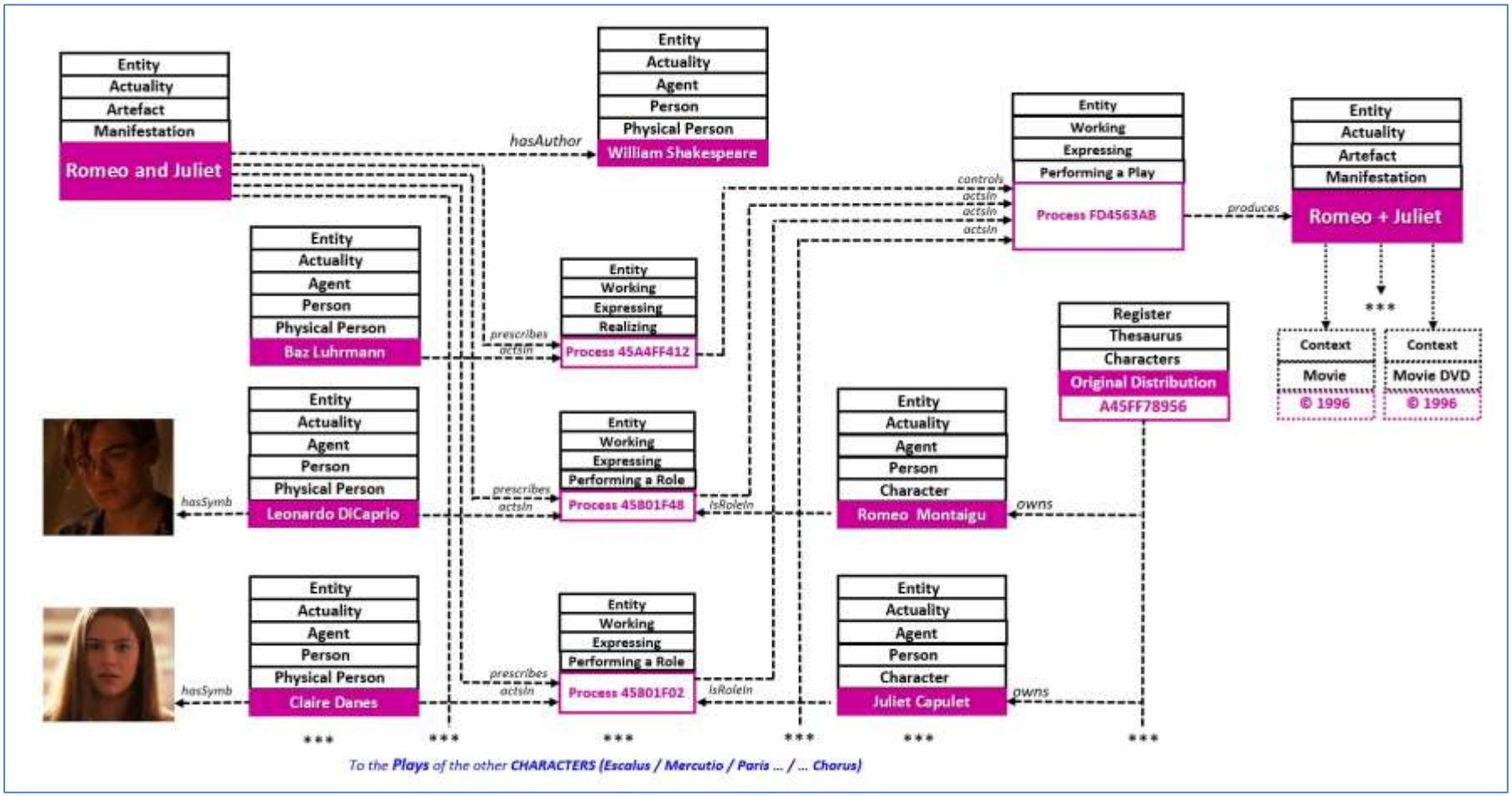


Figure 6-104: Typical modelling of the Casting of the Characters

The figure illustrates the casting of the Characters: Leonardo DiCaprio plays the role of Romeo and Claire Danes the role of Juliet. The movie is also issued as a DVD.

6.5.4.8 Romeo & Juliet: illustration of the structuring of the DVD

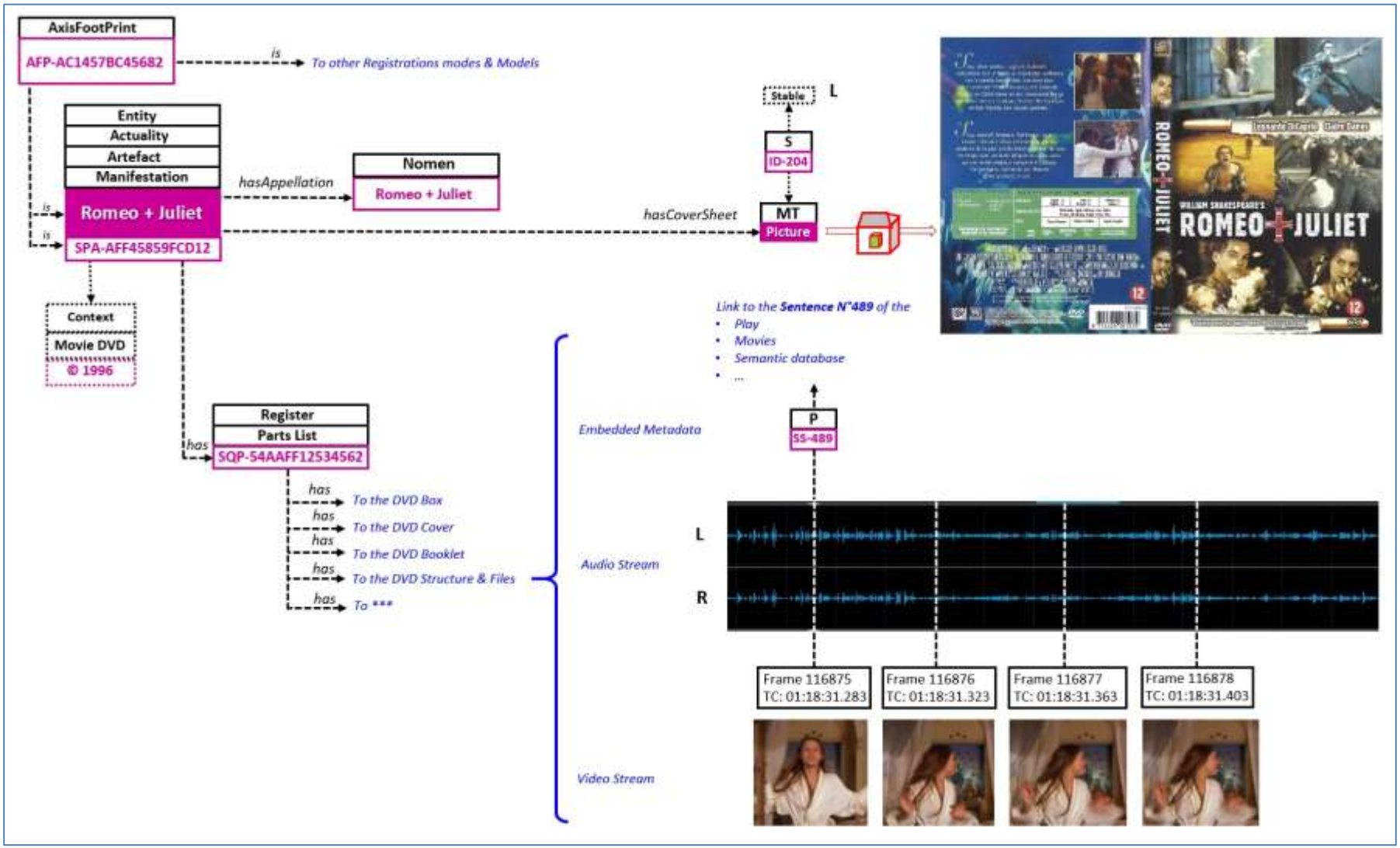


Figure 6-105: Typical modelling of the structure of a DVD

The figure illustrates the structure of the model of the DVD with the details on the scene presented at the section 6.5.4.5.

6.5.4.9 Romeo & Juliet: illustration of the synchronization of the whole structure of the DVD

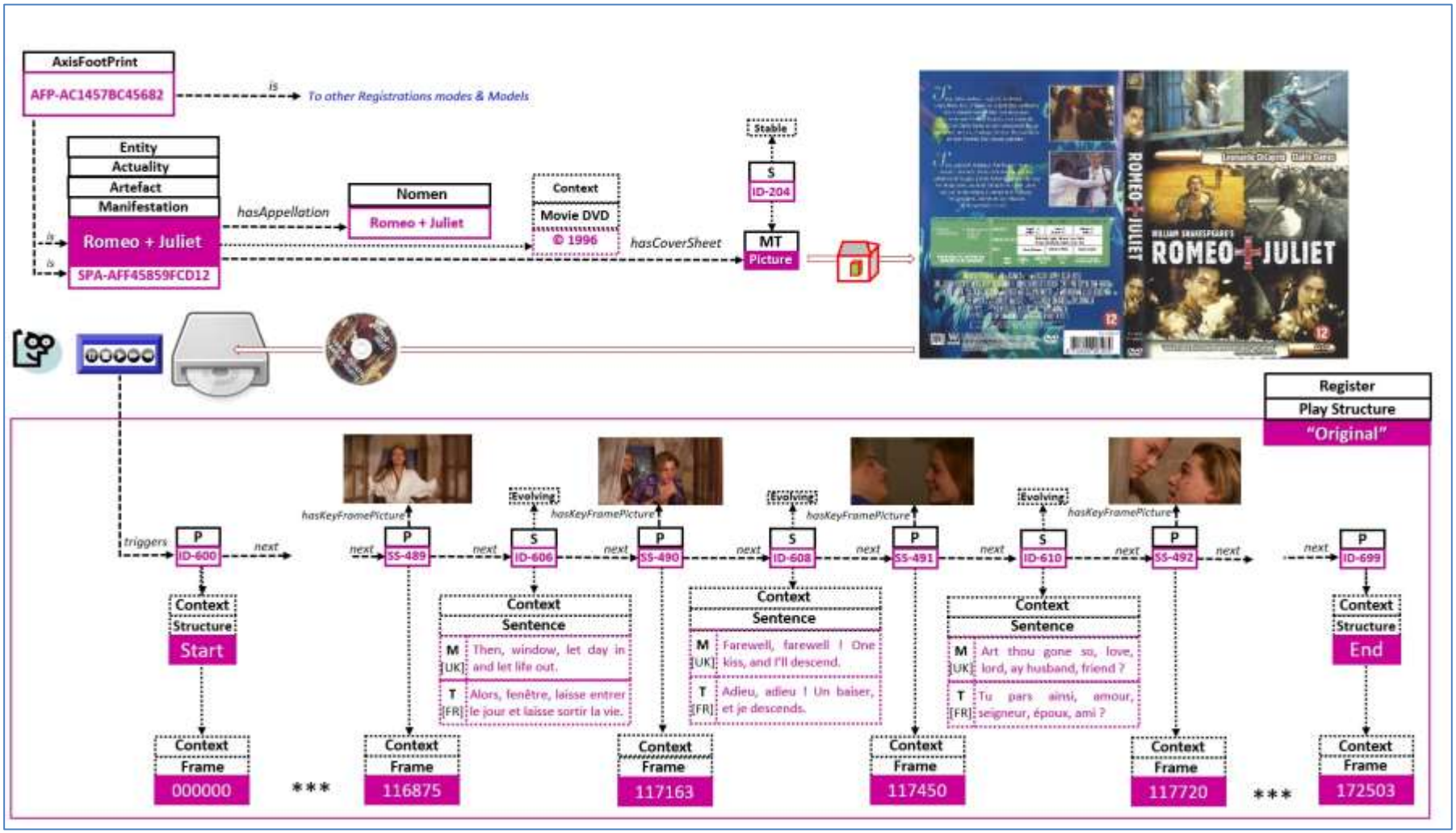


Figure 6-106: Typical modelling of synchronization

The figure illustrates the structure of the general synchronization with the details on the scene presented at the section 6.5.4.5. and the structure of the DVD.

6.5.4.10 Romeo & Juliet illustration of the synchronization of the characters, casting and zones in pictures

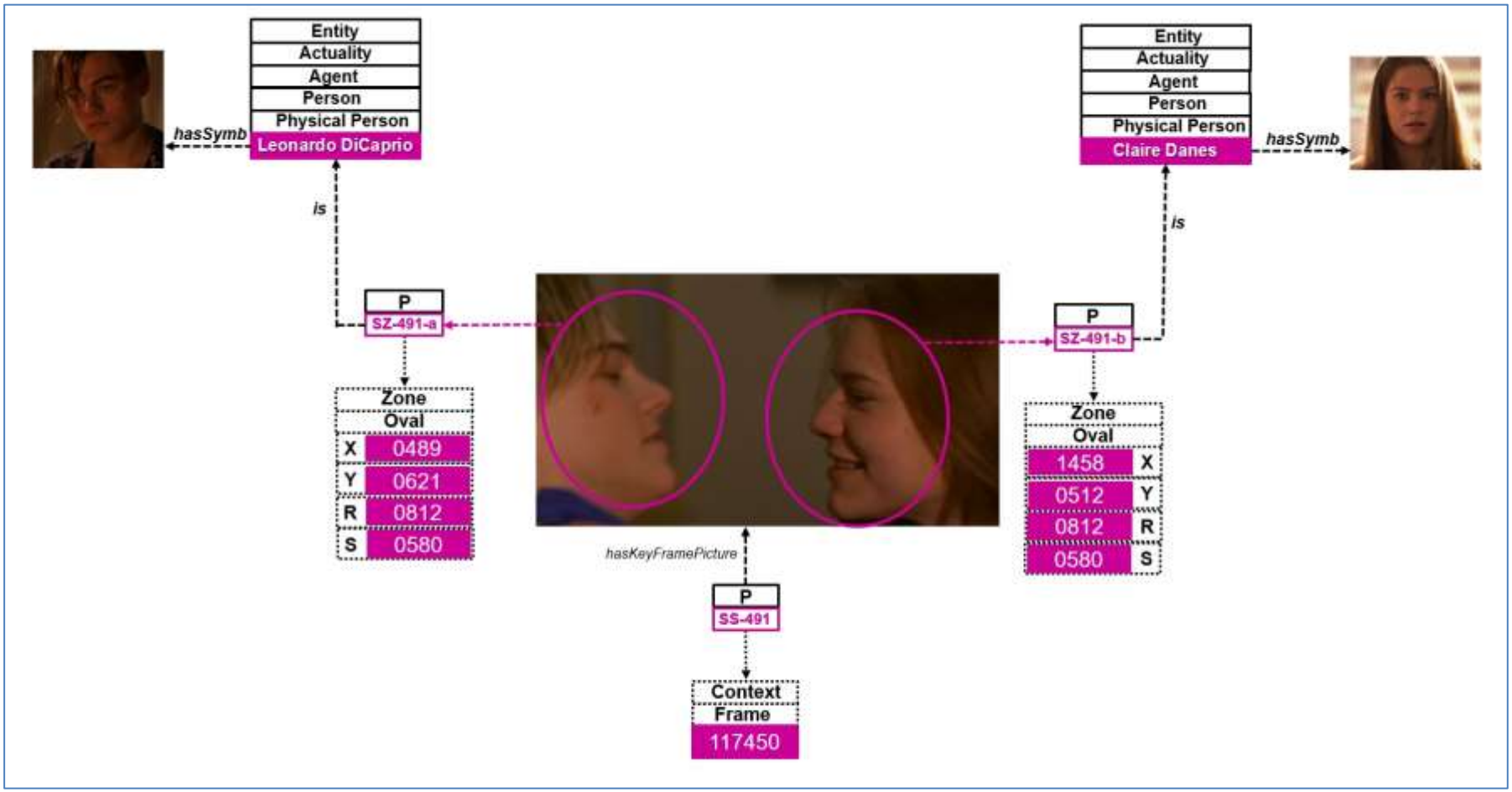


Figure 6-107: Typical modelling of zones in pictures

The figure illustrates the structure of the general modelling of the zones in pictures.

6.5.4.11 Romeo & Juliet: illustration of the synchronization of the whole structure of the Book

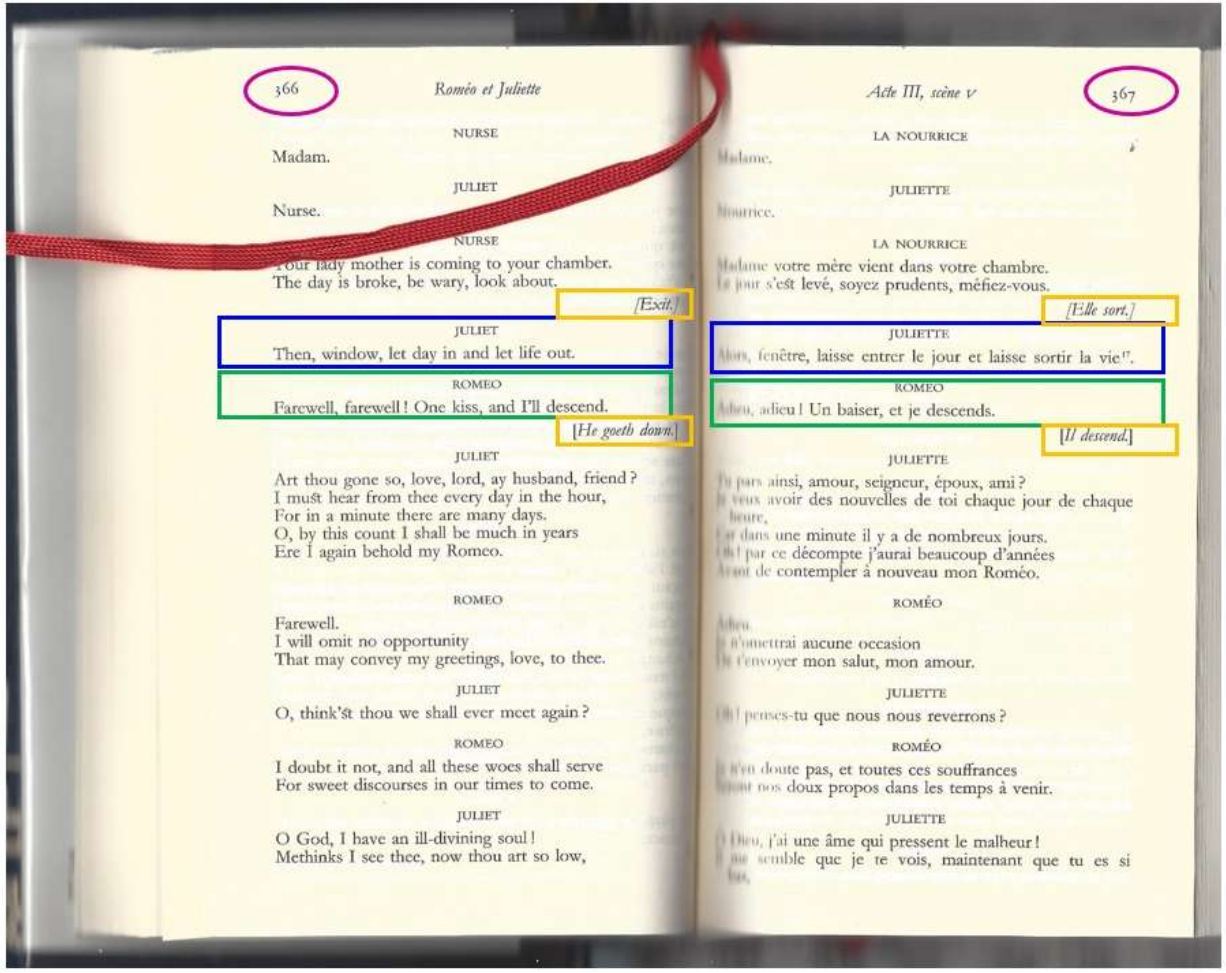


Figure 6-108: Typical modelling of synchronization with the Book

The figure illustrates the structure of the general synchronization with the details on the scene presented at the section 6.5.4.5. and the structure of the Book

6.5.5 Details of the Constructs:

6.5.5.1 Introduction to the formal definition of Constructs

<Here will come a motivation for the selection of the Modelling Technology that should be used for representing the Constructs (Ontology, Root Knowledge Base …). Currently the choice of the W3C pile of standards is obvious [in particular, the use of the RDF, OWL … languages]>.

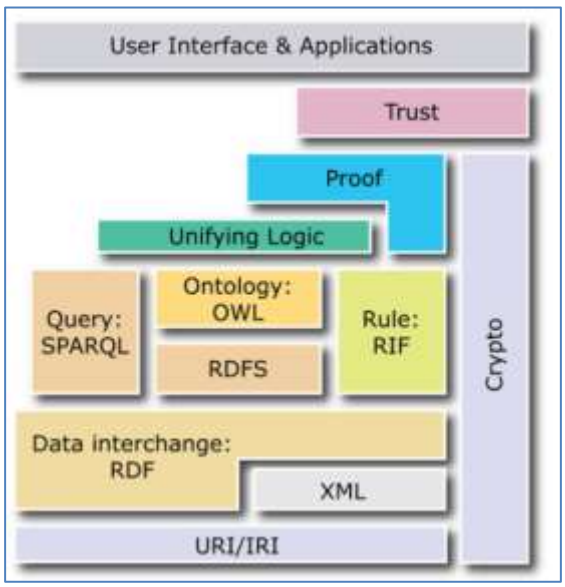


Figure 6-109: The pile of W3C Semantic Web Standards

6.5.5.2 Formal Definition of the AXIS-CSDM Constructs

- **Classes** <to place here such a formal representation>
- **Relationships** <to place here such a formal representation>
- **Property** <to place here such a formal representation>
- **Root Knowledge Base:** <to place here such a formal representation>

6.5.5.3 Suggestion for selecting fitting Modelling Technologies.

<to place here such suggestions of formal & informal MT>

6.5.5.4 Examples of use of the Constructs

<Here will come examples of use of the Constructs: all the cases introduced at the Section 6.5.2 and samples of the cases introduced at the sections 6.5.3 and 6.5.4>.

.

6.5.6 Notes

A rédiger ...

6.6 The BEHAVIOURAL Model

Summary

In the AXI-CSRM data model covers processes internal to ARDs. The behavior model covers process flow protocols, especially those needed for interactions between ARDs, citizens, and Configuration Management services. It's basically workflows.

AXIS-CSRM makes explicit this founding approach.

6.7 The SECURITY Model

Summary

In the security model, AXIS-CSRM focuses on the registration of agents, their mandates and the assurance of their authenticity when they act. It also covers the assurance of the integrity of interactions between agents and the integrity enforcement pursuant to the set (does the process performs what it's supposed to do and nothing else?). Similarly for the integrity and authenticity of AXEs or their components.

AXIS-CSRM makes explicit these aspects.

6.8 The CONFIGURATION MANAGEMENT Model

Summary

The 'Configuration Management' model details the management of the creations and operations of ARD, of the declarations of existence of the Entities, of the Models and of the Knowledge Bases. It also covers the third-party services essential to the proper operation (such as a system for generating unique identifiers and names and the versioning in the life cycle). The 'Resolving services' could be implemented within the ARD or as 'Trusted Third Party services'.

AXIS-CSRM makes explicit these aspects.

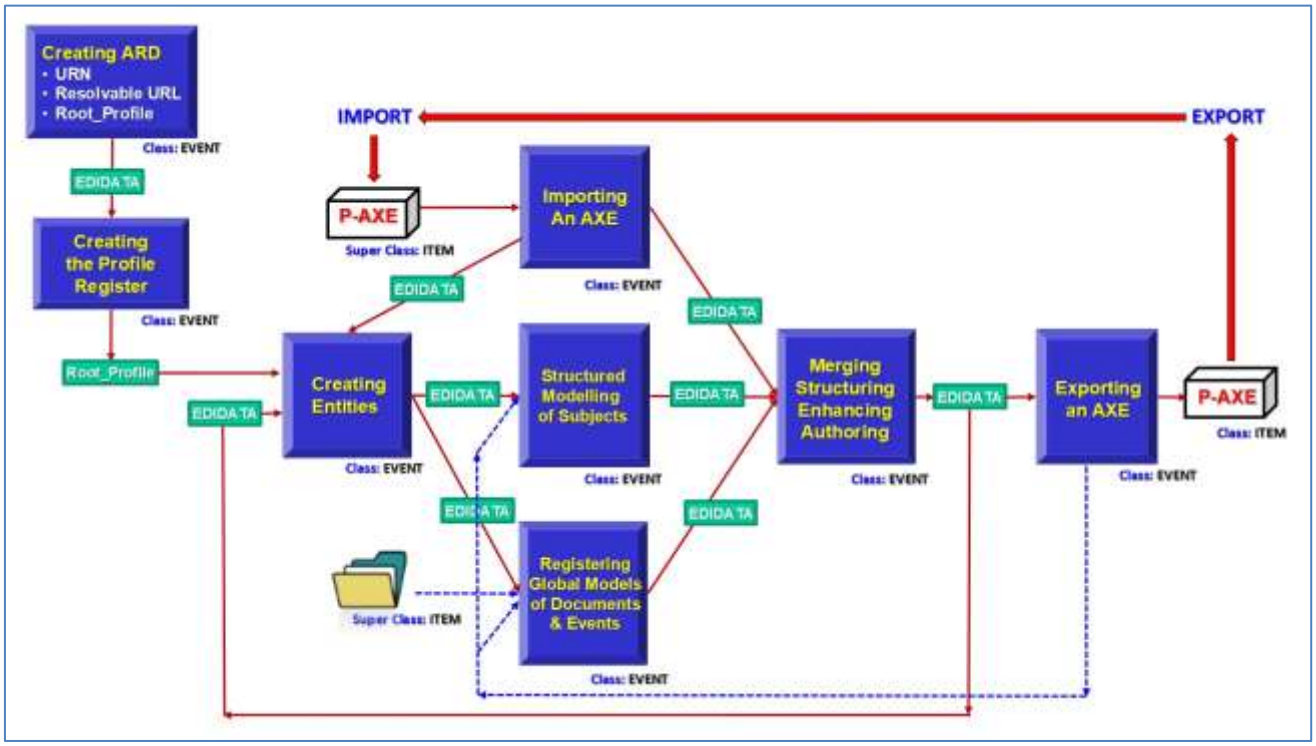


Figure 6-110: Typical flow of the initialization and operations of an ARD

The next figure illustrates the complex chaining required for step by step controlling the migration from “Flat” models towards “Networked semantic” models.

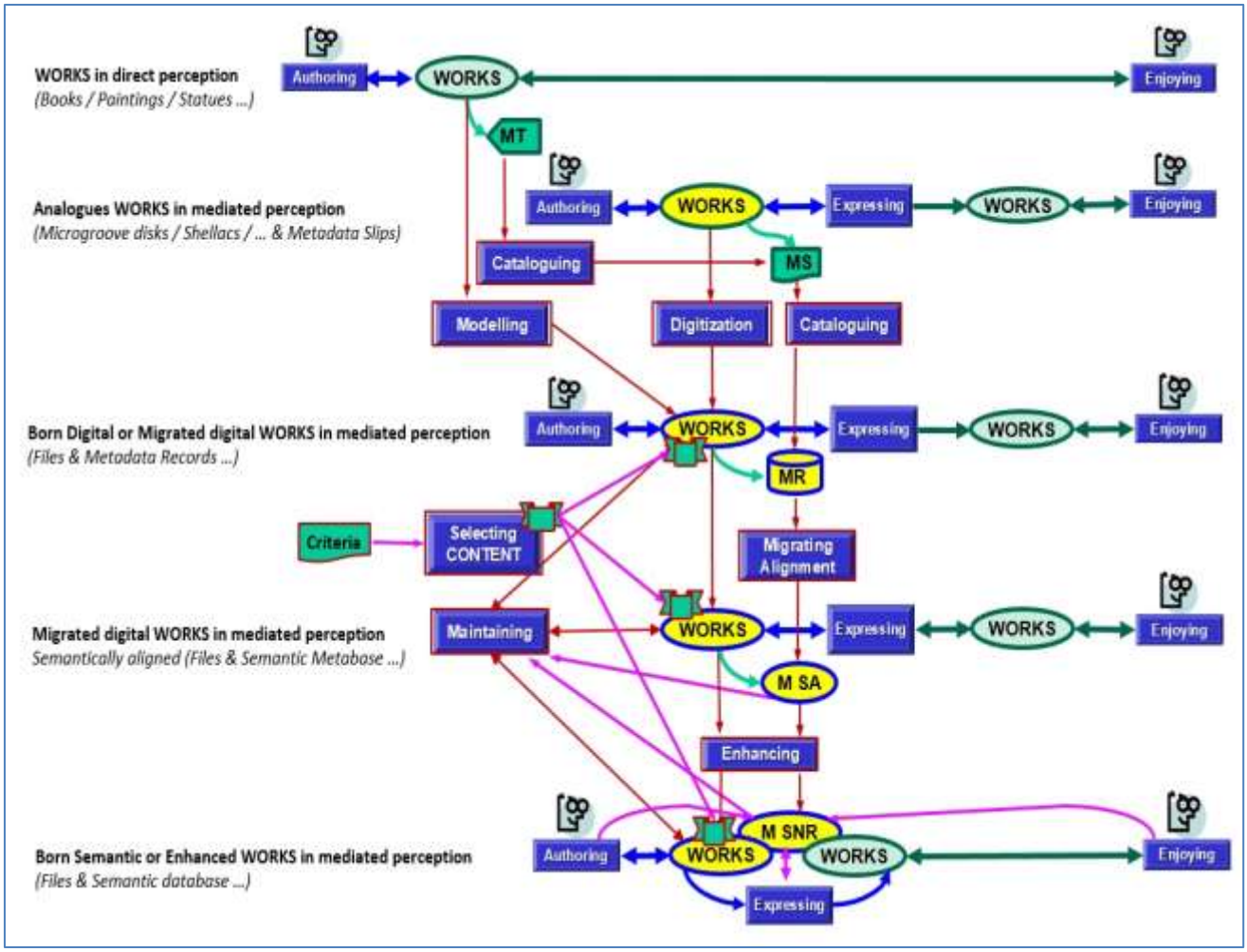


Figure 6-111: Typical flow of migration from “Flat” models to “Networked semantic” models

6.9 The IMPORT & EXPORT Model

Summary

The 'Import & Export' model addresses the problem of encapsulation (the construction of 'Packages' in the terminology of the OAIS) so that the AXE are (as the name suggests) open and autonomous. Although the exported content is only part of the semantic database of an ARD (including a portion of the associated knowledge base), nevertheless Recipient(s) could incorporate the data in its ARD, linking and integrating them into its database, while having clarified the rights linked to this transfer.

AXIS-CSRM makes explicit these issues by addressing the expected functions for 'InterOperability Windows [GIO / IOW].

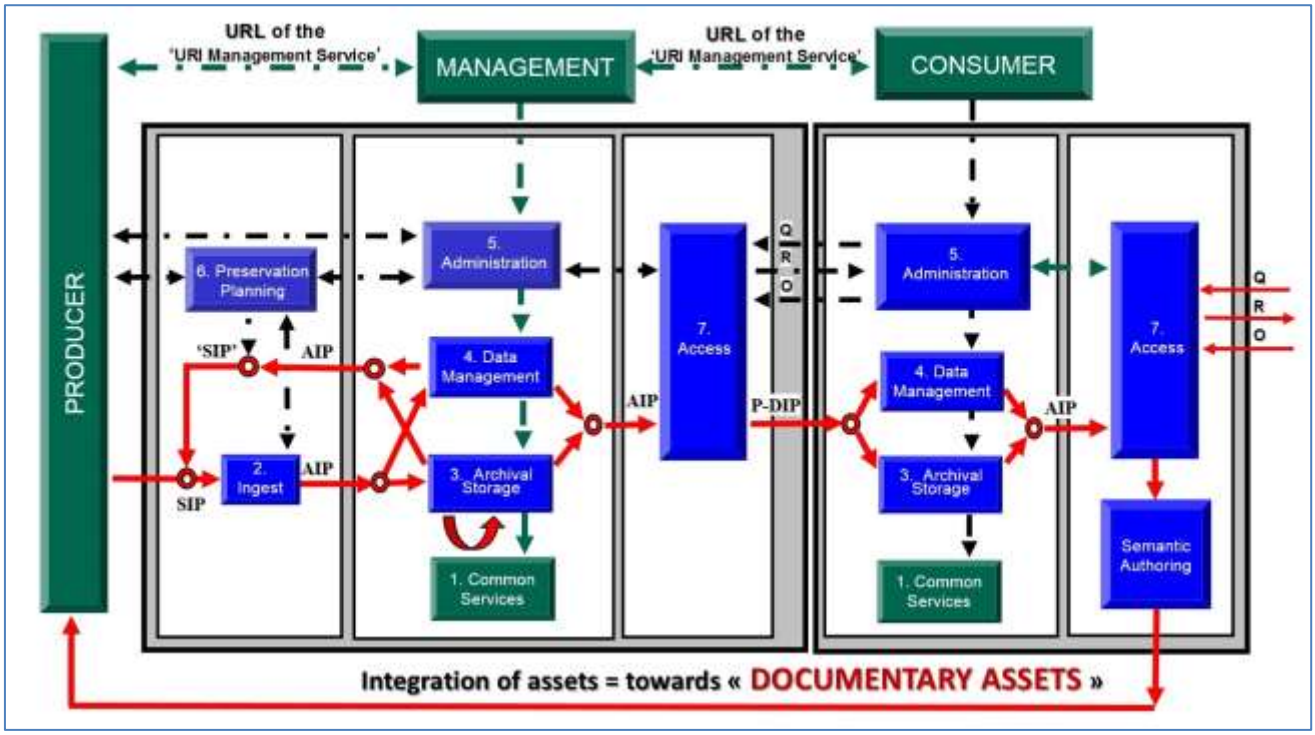


Figure 6-112: Typical flow in the OAIS standard

Notice that the SIP and DIP are expressed as P-AXE in AXIS-CSRM!

<W I P >

7 Chapitre 7 Tracks for implementations

Chapter 7 aims to provide directions for implementation, both to concretely illustrate the concepts and to show that current technologies make possible implementations that are efficient and simple to use. In this release, this third part is still unfinished, waiting for a concrete project.

Various avenues are followed to ensure that such a project can emerge.

In addition to the experience gained during the implementation of the MediaMap + project, three research projects were conducted with the University of Lens.

The model and approach were compared with operational models to check compatibility: in particular the mechanisms of rights management and identification. Here also there is a positive validation. Three projects & developments have been led by several teams of IG2I (Institut de Génie Informatique et Industriel) Lens and the Titan npo:

- 2013 – 2014: **AXIS-TEXT** – Team Aslan with the collaboration of Semsitivity (Gondwana and MediaMap+ projects)
  - Master the textual processing via an open application (GATE: General Architecture for Text Engineering de Sheffiel University - see https://gate.ac.uk)
  - Provide a semantic representation of textual documents
  - Link knowledge base vocabulary

- 2015 – 2016: **AXIS-MoM** (Museum Object Modelization): Team “Lookout” supported by of Mariemont Museum (Fed. Wallonie Bruxelles)
  - Improve Museum Object Modelling, enable the declaration of existence.
  - Design an application to connect the visitor to the Museum's network
  - Navigate all spatialized Linked Open Data, implement the Internet of the Objects (IoT) concepts

- 2016 – 2017: **AXIS-SOW** (Structuring Objects and Wrapper): Team “Nova Team” with RTBF, Memnon, SGT and Perfect Memory
  - Capture, Interpret and export the data flow.
  - Provide an open environment to access, navigate and export the data
  - Implement a semantic GUI to experience the augmented heterogeneous, aggregated assets.

8 ANNEXES

8.1 ANNEXE 1 : CIMOSA - European Enterprise Integration Concept[24]

CIMOSA OVERVIEW *by Gary Rathwell © reserved*

CIMOSA [25] was developed for ESPRIT (European Strategic Program for Research and Development in Information Technology) by AMICE (a consortium of 30 major European vendors and users of CIM systems coordinated by Philips [Jakob Vlietstra & Guy-Noël Maréchal] & CapGemini-Sogeti [Bernard Lorimy]: it includes namely IBM [Kurt Kosanke], AEG [Winfried Seifert], HP, DEC, Siemens, Fiat, Daimler-Benz. Original funding was provided by the European Common Market (EEC) through its ESPRIT project.
CIMOSA defines a model-based enterprise engineering method which categorizes manufacturing operations into Generic and Specific (Partial and Particular) functions. These may then be combined to create a model which can be used for process simulation and analysis. The same model can also be used on line in the manufacturing enterprise for scheduling, dispatching, monitoring and providing process information.

- Generic functions are performed in every enterprise independent of its size, organization and business area. Examples include: Control of work flow, administration of information, integration of resources and management of communications. Generic functions should be performed by Generic system services.
- Specific functions are dependant upon the individual enterprise. Examples include design of products and production processes, generation of production plans, scheduling of production, shipment of products, maintenance of equipment, processing of orders, accounting etc. Specific functions may be performed by machines, humans, and computers.

CIMOSA separates functions using two interrelated concepts.
- The CIMOSA Modelling Framework in which 'Specific' and 'Generic' functions are clearly separated.
- The CIMOSA Integrating Infrastructure supporting execution of Generic functions and linking specific functions. It is effectively the communication system which interconnects all of the functions in the CIM system.

CIMOSA defines four Modelling Views of the Enterprise Functions:
- The Function View describes work flows
- The Information View describes the Inputs and Outputs of Functions
- The Resource View describes the structure of resources (Humans, machines, and control and information systems)
- The Organization View defines authorities and responsibilities

These Views are reflected in the Services of the CIMOSA Integrating Infrastructure (IIS):
- Business Services define an Interpreter for the Function View. This allows control of work flow as defined in the Function Model.
- The Information Services define a set of Generic functions for the handling of the information defined in the Information View.
- The Dialogue Services define an Interpreter for dialogue control programs. These make the link between the Function Model, executing in the Business Services and the heterogeneous set of resources defined in the Resource View.
- The System Management Services provide Generic functions to allow system managers to intervene as defined in the Organization View, e.g. to change, release, activate, start, stop etc. models, both off-line and on-line.
- The Common Services are used for communication handling providing the location transparency required for the interactions between the IIS Services.

The Generation and Execution of CIMOSA Models

The CIMOSA Modelling Framework provides the user with architectural constructs and guidelines for the structured description of business requirements and their translation into CIM system design and implementation.

The Derivation Process guides the user through the three modelling levels: from the definition of enterprise business requirements (Requirements Definition) through the optimization and specification of the requirements (Design Specification) to implementation (Implementation Description). On each modelling level the enterprise is analyzed from different viewpoints (Modelling Views).

[24] https://en.wikipedia.org/wiki/CIMOSA#AMICE_Consortium
[25] The **CIMOSA** acronym has been based on « **C**omputer **I**ntegrated **M**anufacturing - **O**pen **S**ystem **A**rchitecture »

To reduce modelling effort CIMOSA defines three levels of genericity from purely generic to the highly particular.

The first "Generic" Level is a reference catalogue of basic CIMOSA architectural constructs (building blocks) for components, constraints, rules, terms, service function and protocols.

The second Partial Level contains a set of partial models applicable to a specific category of manufacturing enterprises.

The third "Particular" Level is related to one particular enterprise and is defined in the Instantiation Process by the modeler using already prepared building blocks from the Generic and Partial Level and developing new particular enterprise specific components.

After the modelling process is finished, the CIM-OSA model of an enterprise can be released (Integration Process) for the execution in the IIS environment.

It is not currently possible to generate cost models with CIMOSA. The Institut for Applied Informatics is developing an Economic View, which will facilitate economic evaluation of alternative models.

NOTE:

The famous SAP Enterprise has largely been based on the CIMOSA architecture. However, important aspects of the CIMOSA model have not been implemented in SAP, mainly due to the insufficient power of the IT systems in the years 1990.

In a way, we could say that

- **“AXIS-CSRM is the CIMOSA of the years 2020”, either**
- **“AXIS-CSRM is the generic & semantic SAP”**

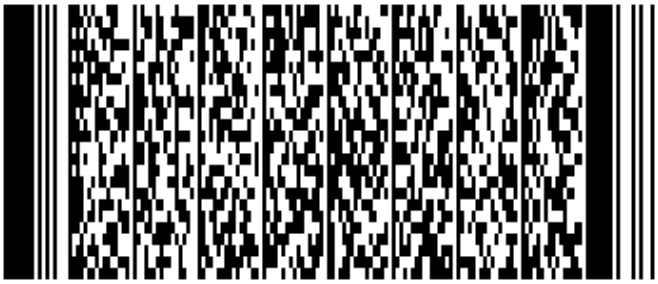
8.2 ANNEX 2 : The concept of Volume

The notion of VOLUME is very general. In the field of computer science, it is customary to use this term to designate a resource capable of storing data organized in files and incorporated into folders. But this vision is too restrictive and one can also imagine memorizing data on microfilm or even on sheets of paper!

Imagine a case where we want to transmit data via a fax.

The natural solution would be to represent the data according to graphic coding standards such as QR-Code or Macro-PDF-417 [N x 2710 characters].

One A4 sheet of paper has a dimension of 29.73 mm x 21.02 mm [based on a unit of one metre square, where the sides are in √2 report is 1189.21 mm x 840.90 mm] and a minimum size of the black dot at 600 dpi of 0.04 mm on an A4 sheet is 29 730 points x 21023 points = 630 Megapixels. With 9 pixels per reliable bit and 8 bits per character it becomes: 8 Mega octets.



The QR-code coding reduces further the efficiency. In practice, on an A4 page it is possible to store in a reliable way about 4 Mo!

When using the Macro-PDF-417, in practice it is similarly possible to store 4 Mo in a highly reliable way using the built-in highly efficient error correcting code system.

8.3 ANNEX 3: UNESCO : Introduction to AXIS-OK and the PERSIST projects

(WIP)

The project PERSIST (Platform to Enhance the Sustainability of the Transglobally Information Society) stimulates the debate between the institutions, the authorities and the ICT industries, while ensuring the preservation and the access to the digital information in the long term .

The goal of AXIS-OK (Organizing Knowledge) aims to add the "**ACCESS**" dimension to "PERSIST" with an approach to the representation of knowledge of documents that are easy to find and navigable for users with rights. To do this, there are representation standards (RDF, OWL, ..) and interaction services via Open Data.

Titan has also developed an ontological basis for structuring a domain (see : AXIS-SOW).

Through this potential, our objective is to implement a prototype on a few examples chosen jointly from the list of 'Documentary Heritages' in the UNESCO Memory of the World register.

(WIP)

8.4 ANNEX 4: Introduction to the UNICODE standard

Unicode is a computer standard that allows text exchanges in different languages, at a global level. It is developed by the Unicode Consortium, which aims at the coding of written text by giving to any character of any writing system a name and a numerical identifier, in a unified way, whatever the computer platform or the software used.

This standard is related to ISO / IEC 10646 which describes an equivalent character map. The latest version, Unicode 10.0, was released in June 2017.

Completely compatible with the universal character set (JUC) of ISO / IEC 10646, the Unicode standard extends it by adding a complete model of representation and word processing, giving each character a set of standardized properties or informative, by describing the semantic relations precisely that can exist between several successive characters of a text, and by standardizing processing algorithms that preserve as much as possible the semantics of the transformed texts. Unicode aims to make the same text usable identically on completely different computer systems.

The Unicode standard consists of a 128 172-character directory, covering a hundred entries, a set of code tables for visual reference, a coding method and several standard character encoding, an enumeration of character properties (uppercase, lowercase letters, symbols, punctuation, etc.) of a set of computer data reference files, and a number of related elements, such as normalization rules, decomposition rules , sort, render, and bidirectional display order (for correct display of text containing both right-to-left characters, such as Arabic and Hebrew, and from left to right) .

8.5 ANNEX 5: Introduction to the OAIS standard

We have already indicated how much ISO's standardization of the Open Archival Information System (OAIS) has weighed heavily on developments in this project. Initiated by the “Consultative Committee for Space Data Systems” in the late 1990s, this ISO-standardized model is a universal reference outlining the functions, responsibilities and organization of a system to preserve information. (in particular digital data), in the long term, to ensure access to identified user communities. This model was revised in 2012, including the semantic dimension in the vision. The data model of AXIS-CSRM implements fully the semantic approach introduced in the new version.

OAIS is an Information Model that processes both digital and non-digital objects simultaneously. The model must indeed be able to process existing physical objects, representation (physical objects) of the real world, but also the digital representations that describe them (digital objects). This practice makes it possible to make separate statements about an object, a document that describes it and the links that exist between objects and representations and their meanings (the signified).

In the OAIS model, the main thing is the creation of an Information Object. The diagram above (a vision specific to the Titan non-profit organization) clearly identifies the object data (bits and bytes) materialized by a representation tool (specific application) and interpreted at the level of meaning by a Knowledge Base.

In the context of “deep preservation” it is necessary both to preserve the data (on suitable media), the applications that generated these data and finally to create a knowledge base to generate the links between the data and their meaning (s). The preservation of data and applications is by no means the object of this project ... all the effort is focused on the ability to connect the world with their objects, their computer representations and their meanings. The creation of a knowledge base is nodal!

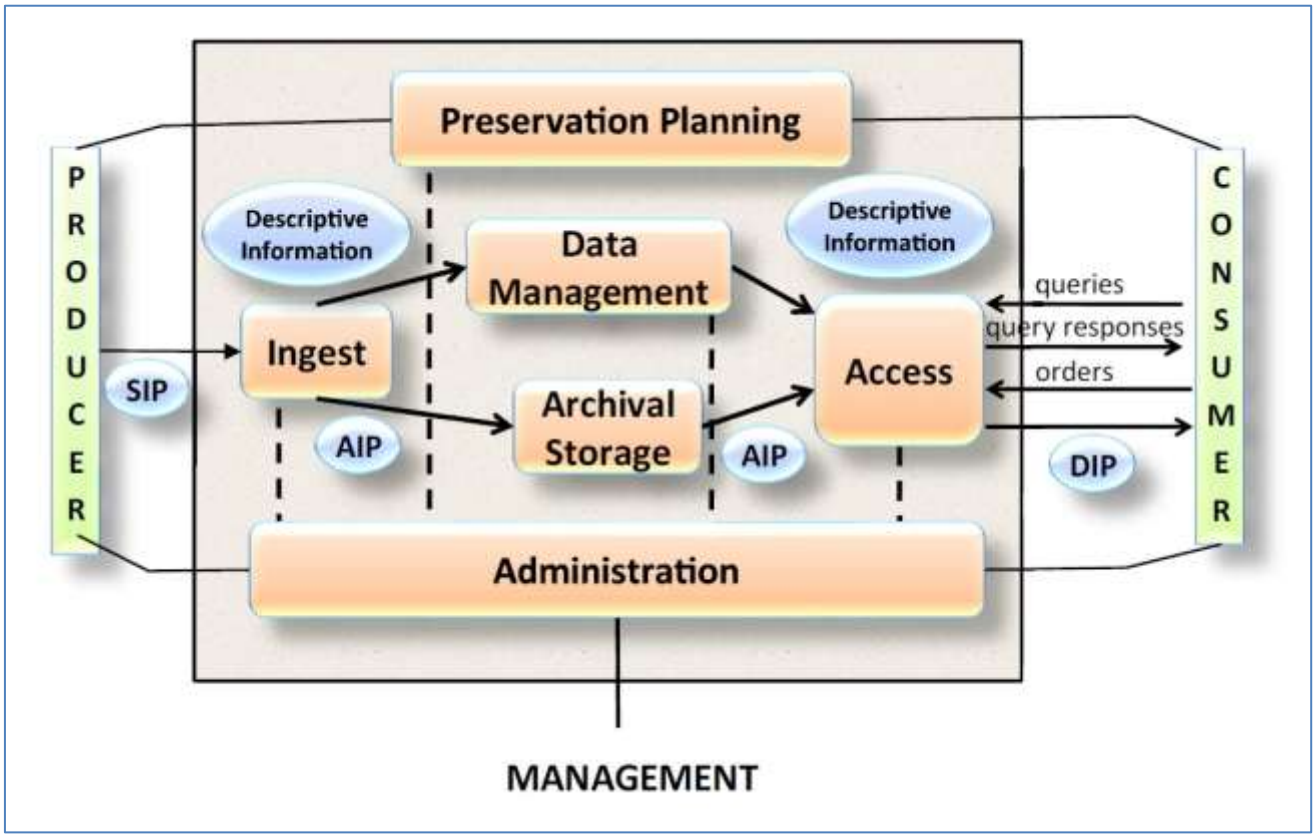


Figure 8-1: The OAIS Functional Model

1. Ingest: process of conformation of objects intended for a digital archive. The Submission Information Package (SIP) is the digital object to be archived with ancillary metadata. The Archival Information Package (AIP) is generated from the SIP as the final step in the ingestion process. The AIP contains all descriptive metadata, technical, project information, access and usage rights, processing (antivirus scans, extraction mode ....).The ingest transfers the data of the producer (SIP) to the Archivist / Archives (AIP).
2. Archive storage: Following the ingest, the AIP is stored, maintained and retrievable from the archive centre. Archive storage includes persistent storage, regular checking of bit stream integrity, and disaster recovery.
3. Data Management: This feature supports searching and retrieving archived content using descriptive metadata.
4. Administration: Refers to day-to-day operations and maintenance of archives and coordination with other functions: archiving, user assistance, implementation and maintenance of policies and processes, etc.
5. Access: The interface that allows users to retrieve data from the archive. The information requested by the user is received as a set of broadcast information (DIP), generated from the AIP stored in the archive centre.
6. Preservation planning: archives must have a continuous digital preservation strategy (regularly updated) and be monitored regularly to detect the risks inherent in this type of activity.
7. Common services: IT services that any computer system, such as a digital archive, needs to function: hardware, software, data, processes, agents, feedback for improvements, etc.

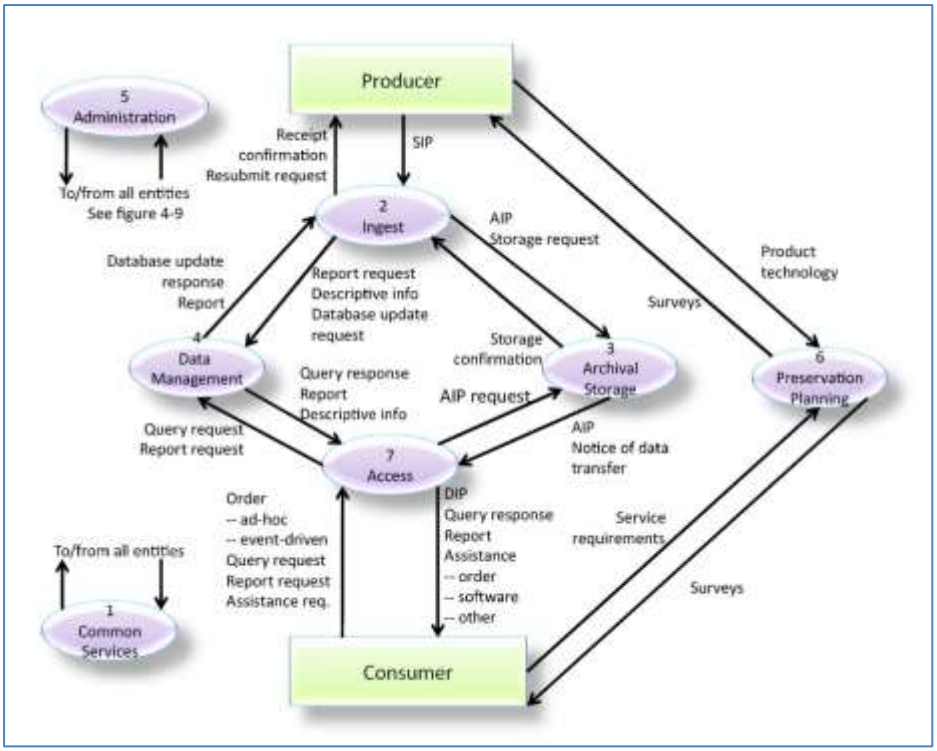


Figure 8-2: The OAIS Data Flow Diagram

The flow of data between functional entities of the OAIS is illustrated by this figure. It describes the most important data streams. Administrative data flows, which are typically background activities, are not represented. The data flows associated with common services are implicit in the illustrated functions and are therefore not displayed.

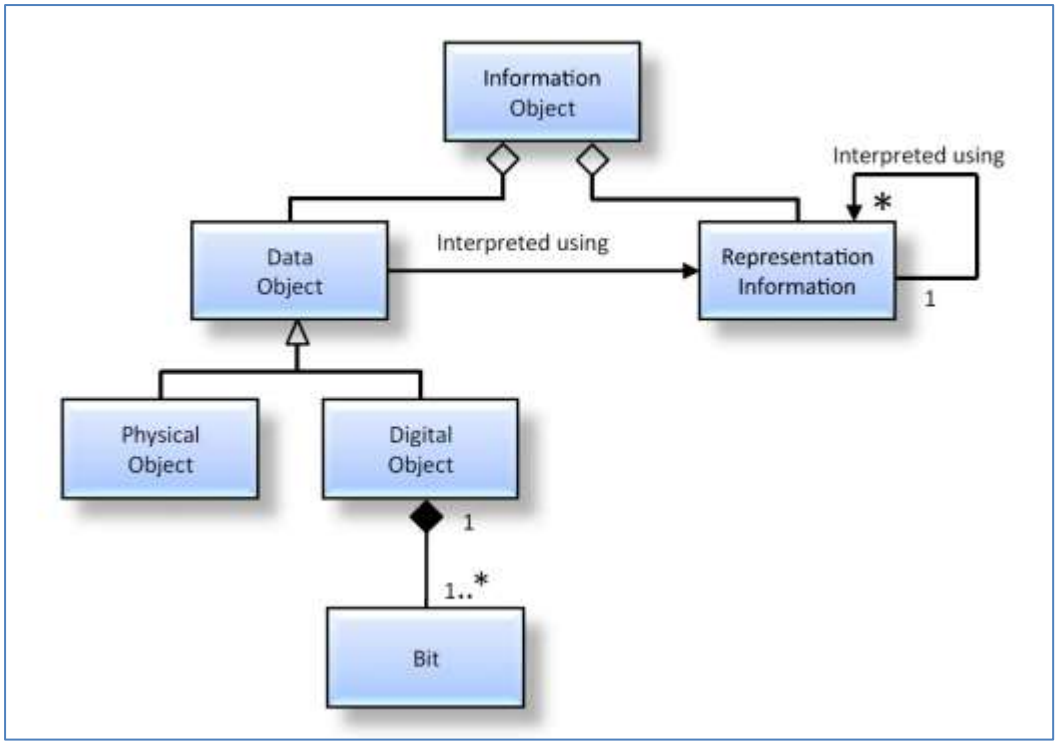


Figure 8-3: The OAIS Information Object

Information Object is a basic concept of the OAIS Reference Model of information being a combination of Data and Representation Information. The Information Object is composed of a Data Object that is either physical or digital, and the Representation Information that allows for the full interpretation of the data into meaningful information (semantic). This model is valid for all the types of information in OAIS.

The Digital Object is composed of one or more bit sequences. The purpose of the Representation Information Object is to convert the bit sequences into more meaningful information. It does this by describing the format, or data structure concepts, which are to be applied to the bit sequences and that in turn result in more meaningful values such as characters, numbers, pixels, arrays, tables, etc. These common computer data types, aggregations of these data types, and mapping rules which map from the underlying data types to the higher level concepts needed to understand the Digital Object are referred to as the Structure Information of the Representation Information object. These structures are commonly identified by name or by relative position within the associated bit sequences. This type of additional required information is referred to as the Semantic Information.. It will include special meanings associated with all the elements of the Structural Information, operations that may be performed on each data type, and their interrelationships.

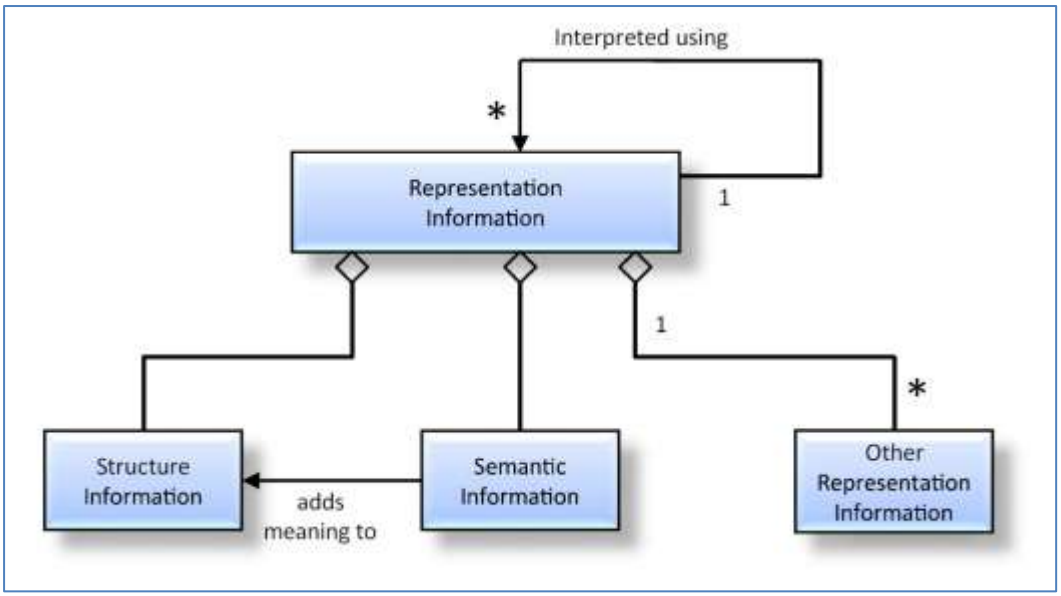


Figure 8-4: The OAIS Information Representation

Figure 5-4 emphasizes the fact that Representation Information contains both Structure Information and Semantic Information, although in some implementations the distinction is subjective. It is useful to remember that the Semantic Information associated with parts of some digitally encoded information is independent of the format. For example, the meaning of numbers in a data file is independent of whether they are encoded as scaled integers or as IEEE Reals; the meaning of words in a document is independent of whether the document is Word or PDF. This figure also shows that Representation Information may contain Other Representation Information. This indicates that the taxonomy of Representation Information presented here is far from complete. For example, software, algorithms, encryption, written instructions and many other things may be needed to understand the Content Data Object, all of which therefore would be, by definition, Representation Information, yet would not obviously be either Structure or Semantics. Information defining how the Structure and the Semantic Information relate to each other, or software needed to process a database file would be regarded as Other Representation Information.

Structure Information, Semantic Information and Other Representation Information are both subtypes and components of Representation Information. Representation Information is an Information Object that may have its own Data Object and its own Representation Information associated with understanding each Data Object, as shown in a compact form by the ‘interpreted using’ association. The resulting set of objects can be referred to as a Representation Network.

As an example, ISO 9660 describes text as conforming to the ASCII standard, but it does not actually describe how ASCII is to be implemented. It simply references the ASCII standard which is additional Representation Information that is needed for a full understanding. Therefore the ASCII standard is a part of the Representation Net associated with ISO

9660 and needs to be obtained by the OAIS in some form, or the OAIS needs to track the availability of this standard so that it may take appropriate steps in the future to ensure its ISO 9660 Representation Information is fully understandable.

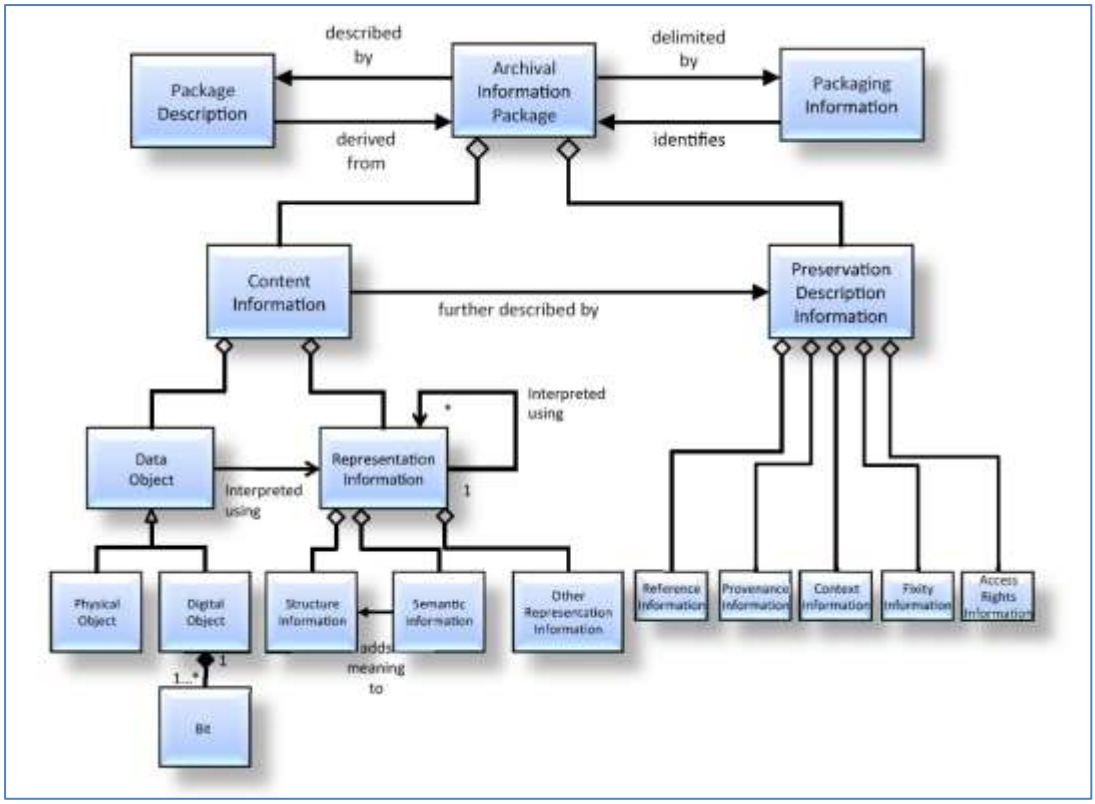


Figure 8-5: The OAIS Archival Information Package

For the "modelling" part, the ISO standard recognizes three types of formats for the representation of the contents (SIP - AIP - DIP) and specifies what must be represented: the Content Information, the Preservation Description Information (the origin, the context, the identification and integrity of the published content), Packaging and Content Description. On the other hand, there is no proposal formulation for formats.

In addition to this ISO/OAIS model, it was very important for the project to have a reference diagram showing how representations and processes are intertwined to build a functional model that takes into account the creation, manufacture and publication content. Moreover, it was necessary to inscribe this vision in the universe of **Open Data** (Data to which everyone should be able to access and that everyone should be able to use and share). It was therefore necessary to design an open system that imports / exports the contents between multiple information systems (or archiving systems) and that is able to manage the transmission of the signifier (the flat/table representation).

Colophon

TITLES Level 1: **Calibri-Title / 14 / Bold / Dark blue**

TITLES Level 2: **Times New Roman / 12 / Bold / Dark blue**

TITLES Level 3: **Calibri-Title / 12 / Bold / Dark blue**

Normal Titles 1: **Times New Roman / 14 / Bold / Black**

Normal Titles 2: **Times New Roman / 12 / Bold / Black**

Normal Texts: Times New Roman / 10 / Normal / Black

XXXXXX