

A Semantic Metadata Enrichment Software Ecosystem (SMESE) Based on a Multi-Platform Metadata Model for Digital Libraries

Ronald Brisebois¹, Alain Abran¹, Apollinaire Nadembega^{2*}

¹École de Technologie Supérieure, Université du Québec, Montréal, Canada

²University of Montreal, Montréal, Canada

Email: ronald.brisebois.1@ens.etsmtl.ca, alain.abran@etsmtl.ca, *apollinaire.nadembega@umontreal.ca

How to cite this paper: Brisebois, R., Abran, A. and Nadembega, A. (2017) A Semantic Metadata Enrichment Software Ecosystem (SMESE) Based on a Multi-Platform Metadata Model for Digital Libraries. *Journal of Software Engineering and Applications*, **10**, 370-405.
<https://doi.org/10.4236/jsea.2017.104022>

Received: February 28, 2017

Accepted: April 27, 2017

Published: April 30, 2017

Copyright © 2017 by authors and Scientific Research Publishing Inc. This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).
<http://creativecommons.org/licenses/by/4.0/>



Open Access

Abstract

Software industry has evolved to multi-product and multi-platform development based on a mix of proprietary and open source components. Such integration has occurred in software ecosystems through a software product line engineering (SPLE) process. However, metadata are underused in the SPLE and interoperability challenge. The proposed method is first, a semantic metadata enrichment software ecosystem (SMESE) to support multi-platform metadata driven applications, and second, based on mapping ontologies SMESE aggregates and enriches metadata to create a semantic master metadata catalogue (SMMC). The proposed SPLE process uses a component-based software development approach for integrating distributed content management enterprise applications, such as digital libraries. To perform interoperability between existing metadata models (such as Dublin Core, UNIMARC, MARC21, RDF/RDA and BIBFRAME), SMESE implements an ontology mapping model. SMESE consists of nine sub-systems: 1) Metadata initiatives & concordance rules; 2) Harvesting of web metadata & data; 3) Harvesting of authority metadata & data; 4) Rule-based semantic metadata external enrichment; 5) Rule-based semantic metadata internal enrichment; 6) Semantic metadata external & internal enrichment synchronization; 7) User interest-based gateway; 8) Semantic master catalogue. To conclude, this paper proposes a decision support process, called SPLE decision support process (SPLE-DSP) which is then used by SMESE to support dynamic reconfiguration. SPLE-DSP consists of a dynamic and optimized metadata-based reconfiguration model. SPLE-DSP takes into account runtime metadata-based variability functionalities, context-awareness and self-adaptation. It also presents the design and implementation of a working prototype of SMESE applied to a semantic digital library.

Keywords

Digital Library, Metadata Enrichment, Semantic Metadata Enrichment, Software Ecosystem, Software Product Line Engineering

1. Introduction

With more and more data available on the web, how users search and discover contents is of crucial importance. There is growing research on interaction paradigms investigating how users may benefit from the expressive power of semantic web standards.

The semantic web may be defined as the transformation of the worldwide web to a database of linked resources, where data may be widely reused and shared [1]. Web services can be enhanced by drawing on semantically aware data made available by a variety of providers. In addition, as information discovery needs to become more and more challenging, traditional keyword-based information retrieval methods are increasingly falling short in providing adequate support. This retrieval problem is compounded by the poor quality of the metadata content in some digital collections.

SECO [2]-[17] is defined as the interaction of a set of actors on top of a common technological platform providing a number of software solutions or services [2] [3]. In SECO, internal and external actors create and compose relevant solutions together with a community of domain experts and users to satisfy customer needs within specific market segments. This poses new challenges since the software systems providing the technical basis of a SECO are being evolved by various distributed development teams, communities and technologies.

There is growing agreement for the general characteristics of SECO, including a common technological platform enabling outside contributions, variability-enabled architectures, tool support for product derivation, as well as development processes and business models involving internal and external actors. At least ten SECO characteristics have been identified [18] that focus on technical processes for development and evolution, see **Table 1**.

Table 1. SECO characteristics [18].

1	Internal and external developers
2	Evaluative common technological platform
3	Controlled central part
4	Enable outside contributions and extensions
5	Variability-enabled architecture
6	Shared core assets
7	Automated and tool-supported product derivation
8	Outside contributions included in the main platform
9	Tools, frameworks and patterns
10	Distribution channel

Gawer and Cusumano [19] have analyzed a wide range of industry examples of SECO and identified two predominant types of platforms:

1. Internal platforms (company or product): defined as a set of assets organized in a common structure from which a company can efficiently develop and produce a stream of derivative products.
2. External platforms (industry): defined as products, services, or technologies that act as a foundation upon which external innovators, organized as an innovative business ecosystem, can develop their own complementary products, technologies, or services.

Indeed, the new generation of SECO must be an integration of multi-platforms (internal and external) that allows the interaction of a set of internal and external actors.

Concurrently modern software demands more and more adaptive features, many of which must be performed dynamically. In this context, a collaborative platform is important in order to coordinate collaborative and distributed environments for development of SECO platforms.

Furthermore, as the requirement of SECO to support adaptation capabilities of systems is increasing in importance [20] it is recommended such adaptive features be included within software product lines (SPL) [21] [22] [23] [24]. The SPL concept is appealing to organizations dealing with software development that aims to provide a comprehensive model for an organization building applications based on a common architecture and core assets [20] [21].

SPLs have been used successfully in industry for building families of systems of related products, maximizing reuse, and exploiting their variable and configurable options [22].

SPL development can be divided into three interrelated activities:

1. Core assets development: may include architecture, reusable software components, domain models, requirement statements, documentation, schedules, budgets, test plans, test cases, process descriptions, modeling diagrams, and other relevant items used for product development.
2. Product development: represents activities where products are physically developed from core assets, based on the production plan, in order to satisfy the requirements of the SPL [25].
3. Management: involves the essential processes carried out at technical and organizational levels to support the SPL process and ensures that the necessary resources are available and well coordinated.

To develop and implement SPL the literature proposes several SPL frameworks [23] using a variety of CBSD approaches [26] [27] [28]:

1. COPA (component-oriented platform architecting): an SPL framework that is component-oriented.
2. FAST (family-oriented abstraction, specification and translation): a software development process that divides the process of a product line into three sections: domain qualification, domain engineering and application engineering.
3. FORM (feature-oriented reuse method): a feature-oriented method that, by

analyzing the features of the domain, uses these features to provide the SPL architecture. FORM focuses on capturing commonalities and differences of applications in a domain in terms of features and uses the analysis results to develop domain architectures and components.

4. Kobra: a component-oriented approach based on the UML features that integrate the two paradigms into a semantic, unified approach to software development and evolution.
5. QADA (quality-driven architecture design and analysis): a product line architecture design method that provides traceability between the product quality and design time quality assessment.

Semantic web [29] [30] [31] [32] [33] linked data is the most important concept to support Semantic Metadata Enrichment (SME) in a SECO architecture [34]-[40].

Today, semantic web technologies, for example in digital libraries, offer a new level of flexibility, interoperability and a way to enhance peer communication and knowledge sharing by expanding the usefulness of the digital libraries that in the future will contain the majority of data. Indeed, a semantic web engine, based on semantic web technology, ensures more closely relevant results based on the ability to understand the definition and user-specific meaning of the word or term being searched for. Semantic search of semantic web engines are better able to understand the context in which the words are being used, resulting in relevant results with greater user satisfaction. Unfortunately, in the public domain there is a scarcity of search engines that follow a semantic-based approach to searching and browsing data [33]. Furthermore, the web is currently not contextually organized.

Thus, to enrich web data by transforming it into knowledge accessible by users, we propose a multi-platform architecture, referred to as SMESE, which uses a CBSD approach to integrate distributed content management enterprise applications, such as libraries and the Software Product Line Engineering (SPLE) approach.

Our SMESE architecture includes mobile first design (MFD) and semantic metadata enrichment (SME) engines that consist of metadata and meta-entity enrichment based on mapping ontologies and a semantic master metadata catalogue (SMMC).

More specifically, our SMESE implements a new decision support process in the context of SPLE, called the SPLE decision support process (SPLE-DSP), a meta entity model that represents all library materials and a meta metadata model. SPLE-DSP allows support for metadata-based reconfiguration. It consists of a dynamic and optimized metadata based reconfiguration model (DOMRM) where users select their preferences in the market place.

The major contributions of this paper are:

1. Definition of a software ecosystem model that configures the application production process including software aspects based on a proposed CBSD and metadata-based SPLE approach.

2. Definition and partial implementation of semantic metadata enrichment using SPLE and a semantic master metadata catalogue (SMMC) to create a universal metadata knowledge gateway (UMKG).
3. Design and implementation of a SMESE prototype for a semantic digital library (Libër).

This paper proposes a semantic metadata enrichment software ecosystem (SMESE) to support multi-platform metadata driven applications, such as a semantic digital library. Based on mapping ontologies SMESE also integrates and enriches data and metadata to create a semantic master metadata catalogue (SMMC).

The remainder of the paper is organized as follows. Section 2 is a literature review. Section 3 presents the multi-platform architecture of the proposed SMESE, and Section 4, the related nine sub-systems. Section 5 presents the prototype of a SMESE implementation in an industry context. Section 6 presents a summary and ideas for future work.

2. Literature Review

A software product line (SPL) [20]-[25] [41] [42] is a set of software intensive systems that share a common and managed set of features satisfying the specific needs of a particular market segment developed from a common set of core assets in a prescribed way [21] [23]. SPL engineering aims at: effective utilization of software assets, reducing the time required to deliver a product, improving quality, and decreasing the cost of software products.

The following sub-sections present the four research axes related to our research:

1. Software product line engineering (SPLE).
2. SECO architecture using component integration and component evolution.
3. SECO architecture and SPLE.
4. Semantic metadata enrichment (SME).

The related works section is at the intersection of SPLE, service-oriented computing, cloud computing, semantic metadata and adaptive systems.

2.1. Software Product Line Engineering (SPLE)

The development of software involves requirements analysis, design, construction, testing, configuration management, quality assurance and more, where stakeholders always look for high productivity, low cost and low maintenance. This has led to software product line engineering (SPLE) [24] as a comprehensive model that helps software providers to build applications for organizations/clients based on a common architecture and core assets. SPLE deals with the assembly of products from current core assets, commonly known as components, within a component-based architecture [43] [44], and involves the continuous growth of the core assets as production proceeds.

Note that the following related works are organized according to two axes: organizational and technical.

An overview of SPLE challenges is presented in [21] [22] [24]. Metzger and Pohl [21] suggest that the successful introduction of SPLE heavily depends on the implementation of adequate organizational structures and processes. They also identify three trends expected from SPLE research in the next decade:

1. Managing variability in non-product-line settings.
2. Leveraging instantaneous feedback from big data and cloud computing during SPLE.
3. Addressing the open world assumption in software product line settings.

A survey of works on search based software engineering (SBSE) for SPLE is presented in Harman *et al.* [22] [24].

Capilla *et al.* [24] provide an overview of the state of the art of dynamic software product line architectures and identify current techniques that attempt to tackle some of the many challenges of runtime variability mechanisms. They also provide an integrated view of the challenges and solutions that are necessary to support runtime variability mechanisms in SPLE models and software architectures. According to them, the limitations of today's SPLE models are related to their inability to change the structural variability at runtime, provide the dynamic selection of variants, or handle the activation and deactivation of system features dynamically and/or autonomously. SPLE is, therefore, the natural candidate within which to address these problems. Since it is impossible to predict all the expected variability in a product line, SPLE must be able to produce adaptable software where runtime variations can be managed in a controlled manner. Also, to ensure performance in systems that have strong real-time requirements, SPLE must be able to handle the necessary adaptations and current reconfiguration tasks after the original deployment due to the computational complexity during variants selection.

Olyai and Rezaei [23] describe the issues and challenges surrounding SPLs, introduce some SPLE ecosystems and compare them, based on the issues and challenges, with a view to how each ecosystem might be improved. The issues and challenges are presented in terms of administrative and organizational aspects and technical aspects. The administrative and organizational comparison criteria include strategic plans of the organization while the technical comparison criteria include requirements, design, implementation, test and maintenance. According to them, there is not a single approach that takes into account all these criteria together. Also, no single approach takes into account metadata for implementation and testing.

2.2. SECO Architecture Using Components Integration and Components Evolution

Software ecosystems (SECO) [2] [3] [4] [10] [19] [35] [39] consist of multiple software projects, often interrelated to each other by means of dependency relationships. When one project undergoes changes and issues a new release, this may or may not lead other projects to upgrade their dependencies. Unfortunately, the upgrade of a component may create a series of issues. In their systematic

literature review of SECO research, Manikas and Hansen [2] report that while research on SECO is increasing:

1. There is little consensus on what constitutes a SECO.
2. Few analytical models of SECO exist.
3. Little research is done in the context of real-world SECO.

They define a SECO as the interaction of a set of actors on top of a common technological platform that results in a number of software solutions or services where each actor is motivated by a set of interests or business models while connected to the rest of the actors. They also identify three main components of SECO architecture:

1. SECO software engineering: focuses on technical issues related directly or indirectly to the technological platform.
2. SECO business and management: focuses on the business, organizational and management aspects.
3. SECO relationships: represent the social aspect of the architecture since it is essential for SPLE actors to interact among themselves and with the platform.

2.3. SECO Architecture and SPLE

This section focuses on SECO architecture related to SPLE, beginning with an industry perspective.

Christensen *et al.* [3] define the concept of SECO architecture as a set of structures comprised of actors and software elements, the relationships among them, and their properties. They present the Danish telemedicine SECO in terms of this concept, and discuss challenges that are relevant in areas beyond telemedicine. They also discuss how software engineering practice is affected by describing the creation and evolution of a central SECO architecture, namely Net4Care, that serves as a reference architecture and learning vehicle for telemedicine and for the actors within a single software organization.

Demir [34] also proposes a software architecture that is strongly related to a defence system and limited to military personnel. Their multi-view SECO architecture design is described step by step. They begin by identifying the system context, requirements, constraints, and quality expectations, but do not describe the end products of the SECO architecture. They also introduce a novel architectural style, called “star-controller architectural style” [34] where synchronization and control of the flow of information are handled by controllers. However, a major drawback of this style is that failure of one controller disables all the subcomponents attached to that controller.

Neves *et al.* [40] propose an architectural solution based on ontology and the spreading algorithm that offers personalized and contextualized event recommendations in the university domain. They use an ontology to define the domain knowledge model and the spreading activation algorithm to learn user patterns through discovery of user interests. The main limitation of their architectural context-aware recommender system is that it is specific to university populations and does not present the actual model of the system that shows the interactions between the components and the data.

Alferez *et al.* [45] propose a framework that uses semantically rich variability models at runtime to support the dynamic adaptation of service compositions. They argue that should problematic events occur, functional pieces may be added, removed, replaced, split or merged from a service composition at runtime, hence delivering a new service composition configuration. Based on this argument, they propose that service compositions be abstracted as a set of features in a variability model. They define a feature as a logical unit of behavior specified by a set of functional and non-functional requirements. Thus, they propose adaptation policies that describe the dynamic adaptation of a service composition in terms of the activation or deactivation of features in the causally connected variability model. Unfortunately, this variability model is limited to activation and deactivation of services. Indeed, the model should allow adaptation of services or include a service interoperability protocol (SIP) rather than compositions only according to changes in the computing infrastructure.

In component based software development (CBSD), the fuzzy logic approach [27] [28] is largely used to select components. Singh *et al.* [27] explored the various measures such as separation of concerns (SoC), coupling, cohesion, and size measure that affect the reusability of aspect oriented software. The main drawback of their contribution is that the fuzzy logic rules are static. They do not propose a way to improve the rules based on developer satisfaction of the fuzzy inference system (FIS) output. In addition, their fuzzy inference system is limited to reusability of software.

2.4. Semantic Metadata Enrichment (SME)

Bontcheva *et al.* [46] investigate semantic metadata automatic enrichment and search methods. In particular, the benefits of enriching articles with knowledge from linked open data resources are investigated with a focus on the environmental science domain. They also propose a form-based semantic search interface to facilitate environmental science researchers in carrying out better semantic searches. Their proposed model is limited to linking terms with DBpedia URI and does not take into account the semantic meaning of terms in order to detect the best DBpedia URI.

Some authors focus their enrichment model on person mobility trace data [47] [48] [49] [50]. Krueger *et al.* [47] show how semantic insights can be gained by enriching trajectory data with place of interest (POI) information using social media services. They handle semantic uncertainties in time and space, which result from noisy, imprecise, and missing data, by introducing a POI decision model in combination with highly interactive visualizations. However, this model is limited to POI detection.

Kunze and Hecht [48] propose an approach to processing semantic information from user-generated OpenStreetMap (OSM) data that specifies non-residential use in residential buildings based on OSM attributes, so-called tags, which are used to define the extent of non-residential use.

Our conclusions from these related works are:

1. SPLE architecture needs to be flexible and meet administrative and organizational aspects such as the organization's strategic plans and marketing strategies, as well as technical aspects such as requirements, design, implementation, test and maintenance.
2. Researchers need to focus on real-world SECO.
3. Several proposed SECO models do not take into account autonomic mechanisms to guide the self-adaptation of service compositions according to changes in the computing infrastructure.
4. In CBSD fuzzy inference systems (FIS) have been employed to develop the components selection model, however, there is no FIS based model that proposes more than one software measure as FIS output.
5. There is no SECO architecture that takes into account several semantic enrichment aspects.
6. Current metadata and entity enrichment models are limited to only one domain for their semantic enrichment process and therefore do not involve several enriched metadata and entity models.
7. Current metadata and entity enrichment models only link terms and DBpedia URI.
8. Current metadata and entity enrichment models do not take into account person mobility trace data gathering and analysis in the enrichment process of metadata.

3. SMESE Multi-Platform Architecture

This section presents the proposed semantic enriched metadata software ecosystem (SMESE) architecture based on SPLE and CBSD approaches to support metadata and entity social and semantic enrichment for semantic digital libraries and based on an MFD approach for user interface design. Each component of the SMESE architecture is based on existing approaches (SPLE and CBSD) and an SME concept (proposed in this work) to generate, extract, discover and enrich metadata based on mapping ontologies and making use of contents and linked data analysis.

For the new generation of information and data management, metadata is a most efficient material for data aggregation. For example, it is easier to find a specific set of interests for users based on metadata such as content topics, or based on the sentiments expressed in a content. Furthermore, it is possible to increase user satisfaction by reducing the user interest gap. To make this feasible, all content needs to be enriched. In other words, specific metadata must be available including semantic topics, sentiments and abstracts. However, at the present time more than 85% of content does not have this metadata.

The SMESE multiplatform prototype includes an engine to aggregate multiple world catalogues from libraries, universities, Bbookstores, #tag collections, museums, and cities. The collection of pre-harvested and processed metadata and full text comprises the searchable content.

Central indexes typically include: full text and citations from publishers, full

text and metadata from open source collections, full text, abstracting, and indexing from aggregators and subscription databases, and different formats (such as MARC) from library catalogues, also called the base index, unified index, or foundation index.

The SMESE multiplatform framework must link bibliographic records and semantic metadata enrichments into a digital world library catalogue. SMESE must search and discover actual collections or novelties, including: works, books, DVDs, CDs, comics, games, pictures, videos peoples, legacy collections, organizations, rewards, TVs, radios, and museums.

The five levels of the semantic collaborative gateway are:

1. Meta Entity.
2. Entity.
3. Semantic metadata enrichment and creation.
4. Free sources of metadata and subscription-based metadata.
5. Content.

Figure 1 presents the entity matrix. The metadata are defined once and are related to each specific entity.

Semantic relationships between the contents, persons, organization and places are defined and curated in the master metadata catalogue. Topics, sentiments and emotions must be extracted automatically from the contents and their context:

ENTITY (NOTICES) MATRIX of the SMESE's Master Catalogue (EXAMPLE)					
<u>Calendars</u>	<u>Contents</u>	<u>Documents</u>	<u>ALRO</u>	<u>Places</u>	<u>Rewards</u>
▪ Interests	▪ Audio Books	▪ Google Doc	▪ Titre	▪ City	▪ Litterature
▪ Library	▪ Books	▪ Paint	▪ Topics	▪ Localization	▪ Movies
▪ POI	▪ Cartographic Mat.	▪ PDF	▪ Keywords	▪ POI	▪ Music
▪ Rewards	▪ Citations	▪ Powerpoint	▪ #referances	▪ ...	▪ Nobel
▪ TV Channel	▪ Comics	▪ Spreadsheet	▪ Annotations	<u>Products</u>	▪ ...
▪ ...	▪ Estampes	▪ Word	▪ ...	▪ Financial	<u>Resources</u>
<u>Collections</u>	▪ Manga	▪ ...	<u>Objects</u>	▪ Groceries	▪ Online
▪ Interests	▪ Microforms	<u>Events</u>	▪ Object	▪ Hardware	▪ Physical
▪ Library	▪ Movies (DVD)	▪ Cinemas Rep.	▪ Work of Art	▪ Natural Health	▪ ...
▪ Organizations	▪ Music (CD)	▪ Expotitions	▪ ...	▪ Pharmacy	<u>Subjects</u>
▪ Personal	▪ Musical Partitions	▪ Libër Spirits	<u>Persons</u>	▪ Software	▪ Genomes
▪ ...	▪ Old Books	▪ News	▪ Actor	▪ ...	▪ MindMaps
	▪ Photos (Image)	▪ Notifications	▪ Author	<u>Publications</u>	▪ Ontologies
	▪ Press	▪ Press Conference	▪ Celebrity	▪ Articles	▪ ...
	▪ Serials	▪ Shows	▪ Musician	▪ Education Programs	<u>WebSites</u>
	▪ Sounds	▪ Spectacles	▪ Politician	▪ Fact Sheets	▪ Homework Help
	▪ Videos	▪ Theaters	▪ Producer	▪ Questions/Answers	▪ Youth
	▪ ...	▪ TV Shows	▪ Singer	▪ Manuals	▪ ...
		▪ ...	▪ Students	▪ Monographs	<u>Works</u>
			▪ User	▪ Newsletters	▪ Concepts
			▪ ...	▪ PostCards	▪ Expressions
				▪ Posters	▪ Manifestations
				▪ Proceedings	▪ ...
				▪ Thesis	
				▪ ...	

Figure 1. Entity matrix.

1. Libraries spend a lot of money buying books and electronic resources. Enrichment uncovers that information and makes it possible for people to discover the great resources available everywhere.
2. The average library has hundreds of thousands of catalogue records waiting to be transformed into linked data, turning those thousands of records into millions of relationships.

FRBR (functional requirements for bibliographic records) is a semantic representation of the bibliographic record. A work is a high-level description of a document, containing information such as author (person), title, descriptions, subjects, etc., common to all expressions, format and copy of the work (see **Figure 2** for an FRBR framework description).

SMESE must allow users to find topically related content through an interest-based search and discovery engine. Transforming bibliographic records into semantic data is a complex problem that includes interpreting and transforming the information. Fortunately, many international organizations (e.g., BNF, Library of Congress and some others) have partly done this heavy work and already have much bibliographic metadata converted into triple-stores.

Recent catalogues support the ability to publish and search collections of descriptive entities (described by a list of generic metadata) for data, content, and related information objects. Metadata in catalogues represent resource characteristics that can be indexed, queried and displayed by both humans and software. Catalogue metadata are required to support the discovery and notification of information within an information community. Using the information from these Semantic Metadata Enrichments, the search engine, discovery engine and notification engine are able to give to the final user better results in accord with his interest or mood.

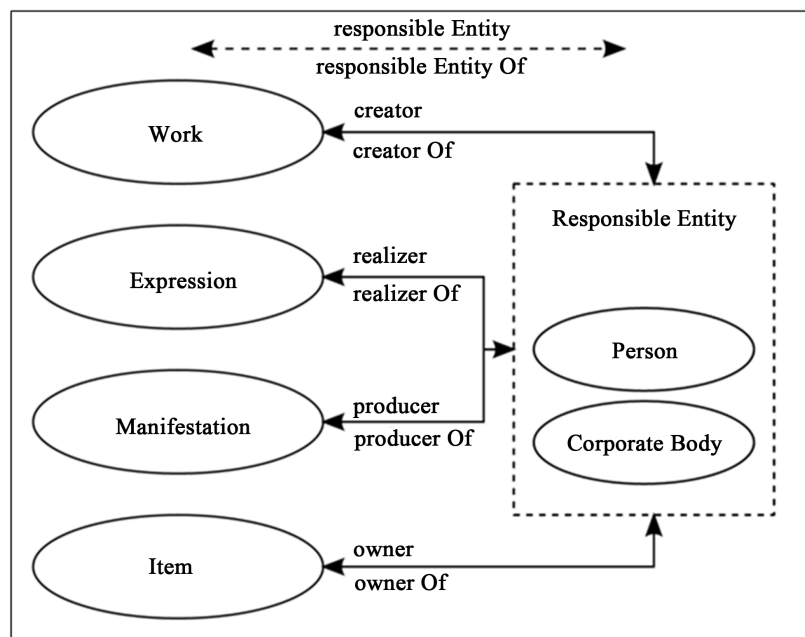


Figure 2. FRBR framework description.

SMESE must also include an automated approach for semantic metadata enrichment (SME) that allows users to perform interest-based semantic search or discovery more efficiently. To summarize, our SMESE makes the following contributions:

Definition and development of a proposed semantic metadata enrichment software ecosystem (see **Figure 3** for SMESE overview and **Appendix B** shows the detailed version).

This new semantic ecosystem will harvest and enrich bibliographic records externally (from the web) and internally (from text data). The main components of the ecosystem will be:

1. Metadata initiatives & concordance rules
2. Harvesting web metadata & data
3. Harvesting authority metadata & data
4. Rule-based semantic metadata external enrichment engine
5. Rule-based semantic metadata internal enrichment engine
6. Semantic metadata external & internal enrichment synchronization engine
7. User interest-based gateway
8. Semantic master catalogue

A. Topic detection/generation: A prototype was developed to automate the generation of topics from the text of a document using our algorithm BM-SATD (Semantic Annotation-based Topic Detection). In this research prototype, the following issues were investigated:

1. Semantic annotations can improve the processing time and comprehension of the document.

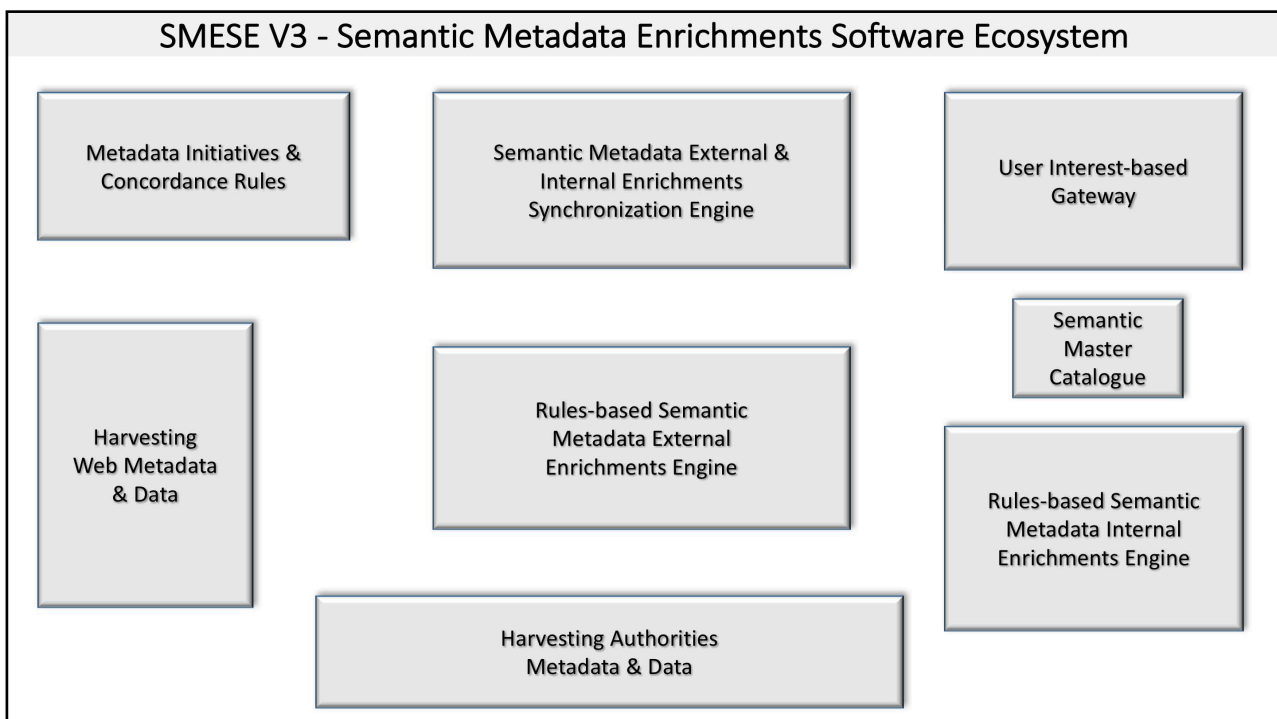


Figure 3. Semantic Enriched Metadata Software Ecosystem (SMESE) architecture.

2. Extending topic modeling into account co-occurrence to combine semantic relations and co-occurrence relations to complement each other.
 3. Since latent co-occurrence relations between two terms cannot be measured in an isolated term-term view, the context of the term must be taken into account.
 4. Use of machine learning techniques to allow the ecosystem SMESE to be able to find a new topic itself.
- B. *Sentiment/Emotion Analysis*: The prototype developed has the following characteristics:
1. Traditional sentiment analysis methods mainly use terms and their frequency, parts of speech, rules of opinion and sentiment shifters; but semantic information is ignored in term selection.
 2. Our contribution to sentiment analysis includes emotions.
 3. The human contribution to improve the accuracy of our approach is taken into account.
 4. Sentiment and emotion analysis are combined.
 5. It is important to identify the sentiment and emotion of a book taking into account all the books of the collection.
 6. The collection of documents and paragraphs are taken into account. In terms of granularity, most of the existing approaches are sentence-based.
 7. These approaches did not take into account the surrounding context of the sentence which may cause some misunderstanding with discovery of sentiment/emotion. In our approach, the surrounding context of the sentence is included.

The prototype makes use of the proposed algorithm BM-SSEA (Semantic Sentiment and Emotion Analysis). The SMEE algorithm fulfills all the attributes of **Table 2**.

Table 2. SMESE characteristics.

1	Internal and external developers
2	Evaluative common technological platform
3	Controlled central part
4	Enable outside contributions and extensions
5	Variability-enabled architecture
6	Shared core assets
7	Automated and tool-supported product derivation
8	Outside contributions included in main platform
9	Social network and IoT integration
10	Semantic Metadata Internal Enrichments
11	Semantic Metadata External Enrichments
12	User Interest-based Gateway

The SMESE extends the SECO characteristics presented in [18] from 10 to 12. See **Table 1** SECO characteristics versus **Table 2** SMESE characteristics.

More specifically, the proposed SPLE approach is a combination of FORM and COPA approaches focusing on data and metadata enrichment. Through the combination of these two approaches, the following can be taken into account:

1. Administrative and organizational aspects such as roles and responsibilities, intergroup communication capabilities, personnel training, adoption of new technologies, strategic plans of the organization and marketing strategies.
2. Technical aspects such as requirements, design, implementation, test and maintenance.

With respect to CBSE, our SMESE includes a method for selecting composer components for design of an SPLE. This method can manage and control the complexities of the component selection problem in the creation of the declared product line. Also, the SMESE architecture supports runtime variability and multiple and dynamic binding times of products.

4. Subsystems within the SMESE Multi-Platform Architecture

The following sub-sections present in more detail the nine subsystems designed for the prototype of this SMESE architecture.

4.1. Metadata Initiatives & Concordance Rules

This section presents the details of the metadata initiatives & concordance rules, specifically the semantic metadata meta-catalogue (SMMC) as shown in **Figure 2**.

Metadata is structured information that describes, explains, locates, accesses, retrieves, uses, or manages an information resource of any kind. Metadata refers to data about data. Some use it to refer to machine understandable information, while others employ it only for records that describe electronic resources. In the library ecosystem, metadata is commonly used for any formal scheme of resource description, applying to any type of object, digital or non-digital. Many metadata schemes exist to describe various types of textual and non-textual objects including published books, electronic documents, archival documents, art objects, educational and training materials, scientific datasets and, obviously, the web.

Libraries and information centers are the intermediaries between the information, information sources and users. In order to make information accessible, libraries perform several activities, one of the most important and fundamental of which is cataloguing. The technological developments of the past 25 years have radically transformed both the process of cataloguing and access to information through catalogues.

Several rules have been proposed to cover the description and provision of access points for all library materials (entities). These rules are based on an individual framework for the description of library materials. There is no ecosystem

that allows the creation of universal, understandable and readable, metadata, that would describe all entities used in a library.

The most known metadata models are:

1. Dublin Core (DC): primarily designed to provide a simple resource description format for networked resources. DC does not have any coding to provide the necessary details for the specification of a record that could be converted to any machine readable coding like UNIMARC, MARC21.
2. UNIMARC: consists of data formulated by highly controlled cataloguing codes. This format is difficult to understand and unreadable for the end user. For this reason, MARC21 was proposed.
3. MARC21: is both flexible and extensible and allows users to work with data in ways specific to individual library needs. MARC21 remains difficult to understand, however.
4. RDF/RDA: mainly in Europe, is a new model that includes FRBRized Bibliographic Records.
5. BIBFRAME: mainly in North America, is a new model that includes FRBRized Bibliographic Records.

In addition, there is no mapping model among these that would make them interoperable. The overall challenge is to develop: (1) a modeling of partial international standardization of entities, (2) a modeling of partial international standardization of metadata, and (3) a modeling of partial international standardization of metadata mapping ontology.

Unfortunately, the power of metadata is limited: indeed, large national and international digital library projects, such as Europeana and the Digital Public Library of America, have highlighted the importance of sharing metadata across silos. While both of these projects have been successful in harvesting collections data, they have had problems with rationalizing the data and forming a coherent and semantic understanding of the aggregation.

In addition, organizations create digital collections and generate metadata in repository silos. Generally such metadata does not:

1. Connect the digitized items to their analogue sources.
2. Connect names to authority records (persons, organizations, places, etc.) nor subject descriptions to controlled vocabularies.
3. Connect to related online items accessible elsewhere.

Aggregators harvest this metadata that, in the process, generally becomes inaccurate. In fact, aggregators usually ignore idiosyncratic use of metadata schemas and enforce the use of designated metadata fields.

Connecting data across silos would help improve the ability of users to browse and navigate related entities without having to do multiple searches in multiple portals. The proposed model defines crosswalks that create pathways to different sources; each pathway checks the structure of the metadata source and then performs data harvesting. **Figure 4** shows the SMMC model that addresses this issue.

In SMESE the metadata is classified into six categories:

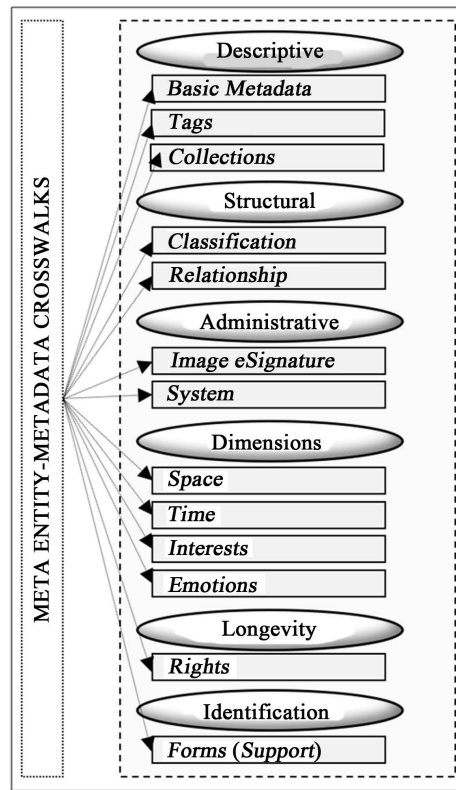


Figure 4. Semantic metadata meta-catalogue (SMMC).

1. *Descriptive metadata*: describes and identifies information resources at the local (system) level to enable searching and retrieving (e.g., searching an image collection to find paintings of animals) at the web-level, and to enable users to discover resources (e.g., searching the web to find digitized collections of poetry). Such metadata includes unique identifiers, physical attributes (media, dimensions, conditions) and bibliographic attributes (title, author/creator, language, keywords).
2. *Structural metadata*: facilitates navigation and presentation of electronic resources and provides information about the internal structure of resources (including page, section, chapter numbering, indexes, and table of contents) in order to describe relationships among materials (e.g., photograph B was included in manuscript A), and to bind the related files and scripts (e.g., File A is the JPEG format of the archival image File B).
3. *Administrative metadata*: facilitates both short-term and long-term management and processing of digital collections and includes technical data on creation and quality control, rights management, access control and usage requirements.
4. *Dimension, longevity and identification metadata*: are new classifications that aim to increase user satisfaction, in terms of expected interests and emotions. For example, dimension metadata regroups all metadata about space, time, emotions and interests. This metadata allows finding specific content. Another example: emotions may suggest specific content to a particular user at a

specific time and place. Furthermore, the source metadata identifies the provenance and the rights relative to the creation of the metadata.

4.2. Harvesting of Web Metadata & Data

The harvesting of web metadata & data sources such as:

1. Semantic digital resources
2. Digital resources
3. Portal/websites events
4. Social networks & events
5. Enrichment repositories
6. Discovery repositories

The integration of these sources in SMESE allows users to aggregate and enrich metadata and data.

4.3. Harvesting Authority Metadata & Data

This sub-section presents the details of the Harvesting of Authorities Metadata & Data.

The Semantic Multi-Platform Ecosystem consists of many authority sources, such as:

1. BAnQ (Bibliothèque et Archives nationales du Qc)
2. BAC (Bibliothèque et Archives du Canada)
3. BNF (Bibliothèque Nationale de France)
4. Library of Congress
5. British Library
6. Europeana
7. Spanish Library

The integration of these platforms in SMESE allows users to build an integrated authorities knowledge base.

4.4. Rules-Based Semantic Metadata External Enrichments Engine

This sub-section presents the details of the rule-based semantic metadata external enrichment engine.

Semantic searches over documents and other content types needs to use semantic metadata enrichment (SME) to find information based not just on the presence of words, but also on their meaning. It consists of:

1. Rule-based semantic metadata external enrichment engine.
2. Multilingual normalization.
3. Rule-based data conversion.
4. Harvesting metadata & data.

Linked open data (LOD) based semantic annotation methods are good candidates to enrich the content with disambiguated domain terms and entities (e.g. events, emotions, interests, locations, organizations, persons), see **Figure 5**, described through Unique Resource Identifiers (URIs) [46]. In addition, the original contents should be enriched with relevant knowledge from the respective

2. A multilingual normalization process.
3. Software Product Line Engineering (SPLE)
4. A topic, sentiment/emotion, abstract analysis and an automatic literature review.

These processes extract, analyze and catalogue metadata for topics and emotions involved in the SMESE ecosystem. These enrichment processes are based on information retrieval and knowledge extraction approaches. The text is analyzed making use of extension of text mining algorithms such as latent Dirichlet allocation (LDA), latent semantic analysis (LSA), support vector machine (SVM) and k-Means.

The different phases of the enrichment process by topics are:

1. Relevant and less similar documents selection phase.
2. Not annotated documents semantic term graph generation phase.
3. Topics detection phase.
4. Training phase.
5. Topics refining phase.

The different phases of the enrichment process by sentiments and emotions are:

1. Sentiment and emotion lexicon generation phase.
2. Sentiment and emotion discovery phase.
3. Sentiment and emotion refining phase.

One of the contributions of the SMESE for digital libraries is that it is not specific to one software product but can be applied to many products dynamically. In addition, it includes a semantic metadata enrichment (SME) process to improve the quality of search and discovery engines.

Indeed, our goal is to provide a SECO that offers a new way to share and learn knowledge. In practice, with the emergence of Big Data, knowledge is not easy to find at the right time and place. The proposed ecosystem uses an SPLE architecture that is a combination of FORM and COPA approaches to catalogue semantically different contents.

Furthermore, we introduce an SPLE decision support process (SPLE-DSP) in order to meet the SPLE characterization such as:

1. Runtime variability functionalities support.
2. Multiple and dynamic binding.
3. Context-awareness and self-adaptation.

SPLE-DSP supports the activation and deactivation of features and changes in the structural variability at runtime and takes into account automatic runtime reconfiguration according to different scenarios. In addition, SPLE-DSP rebinds to new services dynamically based on the description of the relationships and transitions between multiple binding times under an SPLE when the software adapts its system properties to a new context. To take into account context variability to model context-aware properties, SPLE-DSP makes use of an autonomous robot that exploits context information to adapt software behavior to varying conditions.

Furthermore, SPLE-DSP integrates the adaptation of assets and products dy-

namically. This helps products to evolve autonomously when the environment changes and provides self-adaptive and optimized reconfiguration. Additionally, SPLE-DSP exploits knowledge and context profiling as a learning capability for autonomic product evolution by enhancing self-adaptation.

The SPLE-DSP model is an optimized metadata based reconfiguration model where users select their preferences in terms of configuration of interests.

The dynamic and optimized metadata-based reconfiguration model (DOMRM) takes into account the preferences of several users who have distinct requirements in terms of desirable features and measurable criteria. For example:

1. In terms of hardware criteria, the user can select preferences in terms of memory and power consumption or feature attributes such as internet bandwidth or screen resolution.
2. In terms of software criteria, the user can select the entities and their properties, the property characteristics such as the displaying mode, and expected value type.

Indeed, when user preferences change at runtime, the system must be reconfigured to satisfy as many preferences as possible. Since user preferences may be contradictory, only some will be partially satisfied and a relevant algorithm needed to compute the most suitable reconfiguration. To overcome this drawback, we developed the use of a new metadata-based feature model, referred to as the BiblioMondo semantic feature model (BMSFM), to represent user preferences in terms of semantic features and attributes. Our BMSFM constitutes an evolution of traditional stateful feature models [51] that includes the set of user metadata based configurations in the model itself, which allows the representation of user decisions with attributes and cardinalities. More specifically, we developed a metadata-based reconfiguration model that defines all possible metadata and all possible entities that users may need in a specific domain. When a user needs new metadata, he uses the metadata-based request creation tool. The DOMRM model analyses the request and checks whether the requested metadata is relevant and does not already exist. Thus when needed the model automatically creates the new metadata and reconfigures the ecosystem which then becomes available for all users.

Figure 6 illustrates the DOMRM model we designed that is an optimized metadata based configuration for multiple users.

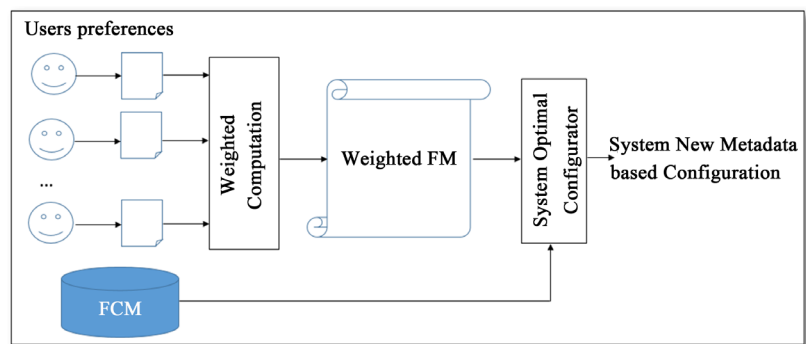


Figure 6. Optimized metadata based configuration for multiple users—DOMRM model.

When the user chooses preferences in terms of system behavior, the semantic weight of each feature is computed based on the feature configuration model (FCM). FCM represents the semantic relationship between features where each feature is active or not. In addition, FCM defines the rules that control the activation status of each feature according to its links with the other features. For example, a rule may be: feature F_i should never be activated when F_{i-1} is activated. Based on this rule, the model automatically activates or deactivates the feature.

The rules are also used to predict the behavior of the application based on the activation status of features according to user preferences. Notice that each user has his own weight per feature that is defined based on his use of the feature. This weight quantifies the importance of the feature for the user (more details about the DOMRM algorithm appear in [Appendix A](#)).

4.6. Semantic Metadata External & Internal Enrichments Synchronization Engine

This sub-section presents the semantic metadata external & internal enrichment synchronization engine which represents which processes to synchronize and which enrichments to push outside the ecosystem.

4.7. User Interest-Based Gateway

This sub-section presents the user interest-based gateway (UIG) that represents the person (mobile or stationary) who interacts with the ecosystem.

The users and contributors are categorized into five groups:

1. Interest-based gateway (mobile-first),
2. Semantic Search Engine (SSE),
3. Discovery,
4. Notifications,
5. Metadata source selection.

4.8. Semantic Master Catalogue

This sub-section presents the semantic master catalogue (SMC) that represents the knowledge base of the SMESE ecosystem.

5. An Implementation of SMESE for a Large Semantic Digital Library in Industry

The proposed SMESE architecture has been implemented for a large digital library. The product In Média V5 was implemented with a global metadata model defined with all the known entities and constraints. The catalogue contains more than 2 million items, with 18 entities and 132 defined metadata. SMMC identifies 1453 metadata and defines a metamodel that consists of a semantic classification of metadata into meta entities.

In addition to semantic web technologies, the characteristics and challenges of SMESE for large digital libraries are:

1. Automatic cataloging with the least human intervention.
2. Metadata enrichment.
3. Discovery and definition of semantic relationships between metadata and records.
4. Semi-automatic classification of bibliographic records.
5. Semantic cataloging and validated metadata making use of a multilingual thesaurus.

First, we defined a list of entities, called Meta Entity, which introduced 193 items. These items represent all library materials. In addition, the structure of the model allows addition of new entities as may be required. **Figure 7** shows the SMESE meta-entity model where for each entity there is: an ID, property Name, description, labels in different languages, and the domain that represents the logic group of the entity; for reason of formatting, **Appendix C** shows a readable version. The domain may be “user” as response value for a metadata. In this implementation, all instances of the entities of the domain can be the response value. The ID allows the user to uniquely identify the entity whatever the language, the source of entities or the metadata model (DC, UNIMARC, MARC21, RDA, BIBFRAME).

Next, the list of metadata is defined. 1341 metadata are defined. Each metadata entry has the following additional metadata called Meta Metadata: ID, related Content Type, is Enrichment, is Repeatable, thesaurus, type, and source Of Schema, which are defined as follows:

1. “source Of Schema” represents the origin.
2. “id” allows unique identification of the entity.
3. “property Name” is a comprehensive term that defines this metadata.
4. “UNIMARC”, “MARC21”, “property Name” allow users to create a mapping between them to make them interoperable.
5. “UNIMARC” and “MARC21” are codes such as 300\$abcf.
6. “Expected type” represents the type of value that may be assigned to the metadata as response.
7. “isRelated” denotes that the response of the metadata is an entity where the identity is given by “related Content Type”.
8. “thesaurus” mentions the thesaurus name that is used to control the metadata integrity.
9. “type” allows classification of the metadata as “descriptive”, “structural”, “administrative”, “dimension”, “longevity” or “identification”.

This classification allows users to do meta research. **Figure 8** shows an illustration of the Meta Metadata model; **Appendix D** shows a readable version.

The semantic matrix model is defined for each entity based on the metaentity and metadata model. This semantic matrix model allows users to define a metadata matrix for each entity where a metadata matrix denotes the logical subset of metadata of metadata model that describes a given entity. **Figure 9** illustrates an example of a semantic metadata matrix for a specific content; **Appendix E** presents a readable version. The objective behind the matrix is to allow the reuse

META ENTITY													
Id.	Oldproperty Name	Property Name	Name	Description	BM label				Client Label		Source	Port Folio Entity Name	Domain
					Fr	En	Sp	Gr	Ne	Fr			
E1-n	Book	book	Book	C'est les contenus du type livre	Livre	Book	"	"	"	Livre	Book	BNF	biibliomondo
E2-n	Serial	serial	Serial	C'est les contenus du type périodique	Périodique	Serial	"	"	"	Périodique	Serial		biibliomondo
E3-n	Audio	audio	Audio	C'est les contenus du type audio	Audio	Audio	"	"	"	Audio	Audio		biibliomondo
E4-n		comic	Comic	C'est les contenus du type bande dessinée	Bande dessinée	Comic	"	"	"	Bande dessinée	Comic		biibliomondo
E5-n		Digital resource	Digital Resource	C'est les contenus du type ressource numérique	Ressource Numérique	Digital Resource	"	"	"	Ressource Numérique	Digital Resource		biibliomondo
E6-n		document	Document	C'est les contenus du type document	Document	Document	"	"	"	Document	Document		biibliomondo
E7-n		game	Game	C'est les contenus du type jeu	Jeu	Game	"	"	"	Jeu	Game		biibliomondo
E8-n	Image	image	Image	C'est les contenus de type Image	Image	Image	"	"	"	Image	Image	BNF	biibliomondo
E9-n	Musical Score	Musical score	Musical Score	C'est les contenus du type partition de musique	Partition de musique	Musical Score	"	"	"	Partition de musique	Musical Score		biibliomondo
E10-n	Video	video	Video	C'est les contenus du type vidéo	Vidéo	Video	"	"	"	Vidéo	Video		biibliomondo
E11-n	FRBR Work	Creative work	Creative work	The most generic kind of creative work, including books, movies, photographs, software programs, etc	Oeuvre	FRBRWork	"	"	"	Oeuvre	FRBRWork	BNF	biibliomondo
E12-n		manufactation	Manufactation	C'est la manifestation d'une oeuvre	Manifestation	Manifestation	"	"	"	Manifestation	Manifestation		biibliomondo
E13-n		expression	Expression	C'est l'expression d'une manifestation	Expression	Expression	"	"	"	Expression	Expression		biibliomondo
E14-n		concept	Concept	C'est une concept	Concept	Concept	"	"	"	Concept	Concept		biibliomondo
E15-n		city	City	C'est les sous divisons d'un canton	Ville	City	"	"	"	Ville	City	BM	places
E16-n		Postal Address	Postal Address	C'est les contenus de type Adresse	Adresse postal	Postal address	"	"	"	Adresse postal	Postal address	BNF	places
E17-n		Place of interest	Place Of Interest	C'est un lieu qui suscite un intérêt particulier; par exemple, un site touristique, un musée, monument,...	Place d'intérêt	Place of interest	"	"	"	Place d'intérêt	Place of interest	BM	places
E18-n		country	Country	C'est un pays	Pays	Country	"	"	"	Pays	Country	BNF	places
E19-n		region	Region	C'est les sous divisions d'un pays; c'est le cas des régions en France, des États au États-Unis ou des provinve au Canada C'est les sous divisions d'un regions; par exemple les comtés au États unis ou les départements en France	Région	Region	"	"	"	Région	Region	BM	places

Figure 7. SMESE Meta Entity model.

META METADATA																						
Id.	Unimarc	Marc21	Old property Name	Property Name (BanQ)	Property Name (Excel)	Description	BM label			Client Label		Expected Type	Is Related	Related Content Type	Short Notice	Enriched	Data Location	Hidden	Repeatable	Thesaurus	Type isUrl	Source Of Media Schema Version
							Fr	En	Sp	Ne	Gr											
M1-m	200\$acdehi	OP0245	OP0245	title	title	It is the main title of a content	Titre	Title												Desc	BM	
M2-m	700\$abcdfg/720\$af4	OP01XX	OP01XX	author	author	The author of this content. Please note that author is special in that HTML 5 provides a special mechanism for indicating authorship via the rel tag. That is equivalent to this and may be used interchangeably	Auteur	Author		Auteur	Author	Text		X		Local			Personnes ou Organisations	Desc	BNF	
M3-m	701\$abcdfg/721\$af4			creator	creator	The creator/author of this Creative Work. This is the same as the Author property for Creative Work	Auteur	Author		Compositeur/Interprète	Composer/Performer	Text		X		Local			Personnes ou Organisations	Desc	BNF	
M4-m				promoter	promoter	C'est le organisateur du contenu; c'est le cas des événements	Organisateur	Promoter				Text				Local			Personnes ou Organisations	Desc	BM	
M5-m	215\$acde	OP0300	OP0300	determs: description	description	A short description of the item.	Description	Description		Description	Description	Multiline Text				Local		X		Desc	BNF	
M6-m				Image	Image	An image of the item. This can be a image file or a URL or a fully described Image Object.	Imagette	Thumbnail		Imagette	Thumbnail	Image		X		BM.com		X		Desc	BNF	
M7-m	101\$af	LO0546	LO0546	determs: language	language	C'est la langue (parlée ou écrite) dans laquelle le contenu est créé ou dans laquelle une personne s'exprime; par exemple, un livre rédigé en français aura comme langue le français	Langue(s)	Langue(s)		Langue(s)	Langue(s)	Text				Local		X	Langues	Desc	BM	
M8-m	540\$af	OP0740	OP0740	Title Added Uncontrolled Related	Title Added Uncontrolled Related	C'est d'autres titres du contenu qui pourraient être utilisés dans d'autres catalogue	Autres titres	Other titles		Autres titres	Other titles	Text				Local				Desc	BM	
M9-m	500\$abhijklmnrsvw	OP0240	OP0240	Title Uniform	Title Uniform	C'est le titre unifié du contenu	Titre uniforme	Uniform title		Titre uniforme	Uniform title	Text				Local				Desc	BM	
M10-m				Also Available In	Also Available In	C'est les autres langues dans lesquelles le contenu a été aussi produit ou dans lesquelles une personne s'exprime	Disponible en d'autres langues	Available in other languages		Aussi disponible en	Also available in	Text				Local		X	Langues	Desc	HEC OIC	

Figure 8. SMESE metadata model.

of metadata for distinct entities. This extends the search range for entities, facilitates the search for users in terms of search criteria and increases the probability of achieving satisfying results.

After the definition of entities of collections and harvesting of metadata from the dispersed collections, a metadata crosswalk is carried out. This is a process in which relationships among the schema are specified, and a unified schema is developed for the selected collection. It is one of the important tasks for building “semantic interoperability” among collections and making the new digital library meaningful.

The most frequent issues regarding mapping and crosswalks are: incorrect mappings, misuse of metadata elements, confusion in descriptive metadata and administrative metadata, and lost information. Indeed, due to the varying degrees of depth and complexity, the crosswalks among metadata schemas may not-necessarily be equally interchangeable. To solve the issue of varying degrees

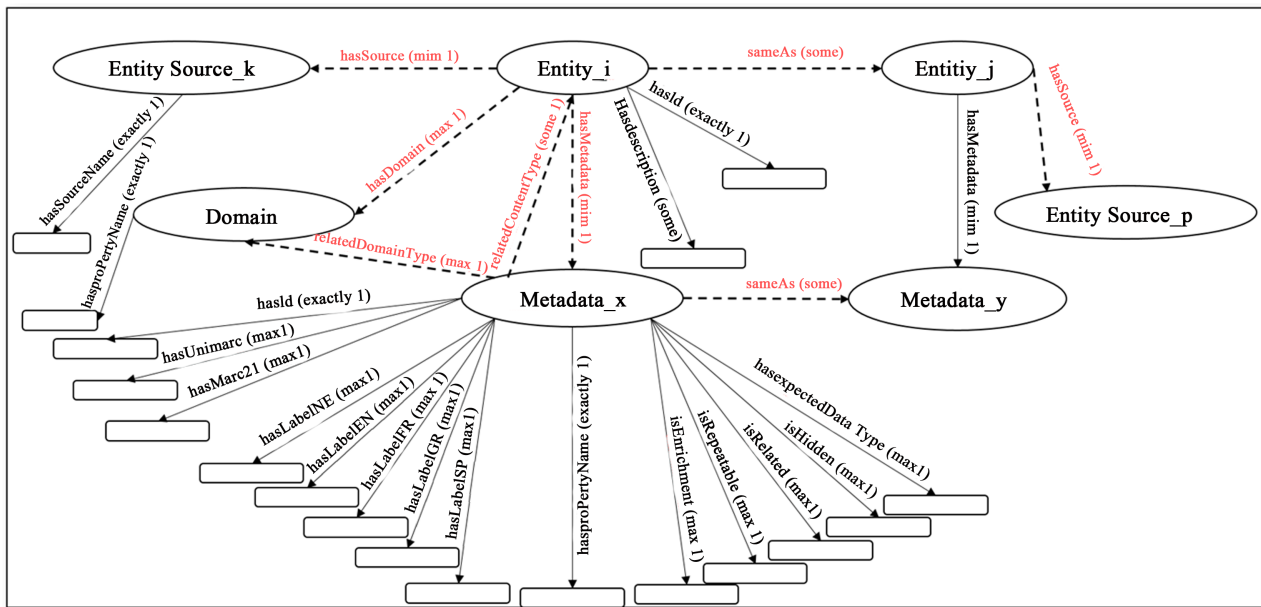


Figure 10. Ontology mapping model.

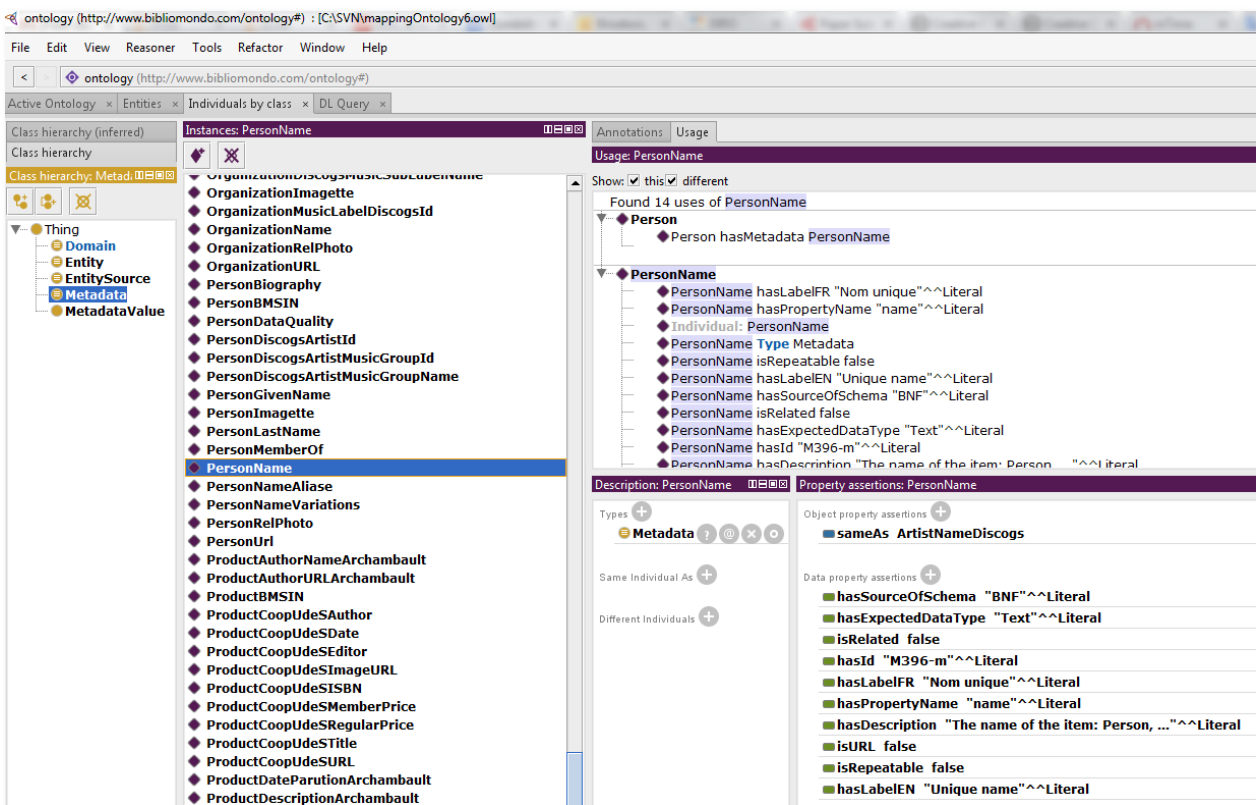


Figure 11. Ontology mapping implementation using Protégé.

- A total of 94,015,090 metadata records were collected from these two sources:
1. From Discogs, we collected 7,983,288 entities: 2,621,435 music releases, 4,466,660 artists and 895,193 labels.
 2. From researchGate, we collected 86,031,802 entities: 77,031,802 publications

and more than 9,000,000 researchers.

3. In fact, SMESE contains more than 3.4 billions triplets and growing.

6. Summary and Future Work

In this paper, we proposed a design and implementation of a semantic enriched metadata software ecosystem (SMESE).

The SMESE prototype, which was implemented at BiblioMondo, integrates data and metadata enrichment to support specific applications for distributed content management. To perform this integration, SMESE makes use of the software product line engineering (SPLE) approach, a component-based software development (CBSD) approach and our proposed new concept, called semantic metadata enrichment (SME) with distributed contents and mobile first design (MFD). In this implementation, the SPLE architecture is a combination of FORM and COPA approaches.

We also presented our implementation of SMESE for digital libraries. This included SPLE-DSP, a new decision support process for SPLE. SPLE-DSP consists of a dynamic and optimized metadata based reconfiguration model (DOMRM) where users select their preferences in the market place. SPLE-DSP takes into account runtime variability functionalities, multiple and dynamic binding, context-awareness and self-adaptation.

We also implemented the Meta Entity that represents all library materials and meta metadata. The ontology mapping model was then implemented to make our models interoperable with existing metadata models such as Dublin Core, UNIMARC, MARC21, RDF/RDA and BIBFRAME.

The major contributions of this paper are as follows:

1. Definition of a software ecosystem architecture (SMESE) that configures the application production process including software aspects based on CBSD and SPLE approaches.
 - a) The use of a LOD-based semantic enrichment model for semantic annotation processes.
 - b) The integration of National Research Council of Canada (NRC) emotion lexicon for emotion detection.
 - c) A repository of 43 thesaurus included in RAMEAU for semantical contextualization of concepts.
 - a. An extended latent Dirichlet allocation (LDA) algorithm for topic modeling.
2. Definition and partial implementation of semantic metadata enrichment using metadata SPLE and an SMMC (semantic master metadata catalogue) to create a universal metadata knowledge gateway (UMKG).
3. The design and implementation of an SMESE prototype of for a semantic digital library (Libër).

This paper proposed a semantic metadata enrichments software ecosystem (SMESE) to support multi-platform metadata driven applications, such as a semantic digital library. Our SMESE integrates data and metadata based on mapping ontologies in order to enrich them and create a semantic master metadata

catalogue (SMMC).

Within the SPLE context, SPLE-DSP is used by SMESE to support dynamic reconfiguration. This consists of a dynamic and optimized metadata based reconfiguration model (DOMRM) where users select their preferences within the market place. SPLE-DSP takes into account runtime metadata-based variability functionalities, multiple and dynamic binding, context-awareness and self-adaptation. Our SMESE represents more than 200 million relationships (triplets).

Future work will include:

1. An enhanced ecosystem of connecting engines and rule-based algorithms to enrich metadata semantically, including topics and sentiments/emotions.
2. Evaluation of the performance of an implementation of the SMESE ecosystem using different projects, comparing results against existing techniques of metadata enrichments.

Exploring text summarization and automatic literature review as metadata enrichment, the semantic annotations could be used to enrich metadata and provide new types of visualizations by chaining documents backward and forward inside automated literature reviews.

References

- [1] Lacasta, J., Nogueras-Iso, J., Falquet, G., Teller, J. and Zarazaga-Soria, F.J. (2013) Design and Evaluation of a Semantic Enrichment Process for Bibliographic Databases. *Data & Knowledge Engineering*, **88**, 94-107.
- [2] Manikas, K. and Hansen, K.M. (2013) Software Ecosystems—A Systematic Literature Review. *Journal of Systems and Software*, **86**, 1294-1306.
- [3] Christensen, H.B., Hansen, K.M., Kyng, M. and Manikas, K. (2014) Analysis and Design of Software Ecosystem Architectures—Towards the 4S Telemedicine Ecosystem. *Information and Software Technology*, **56**, 1476-1492.
- [4] Shinozaki, T., Yamamoto, Y. and Tsuruta, S. (2015) Context-Based Counselor Agent for Software Development Ecosystem. *Computing*, **97**, 3-28.
<https://doi.org/10.1007/s00607-013-0352-y>
- [5] Jansen, S. and Bloemendal, E. (2013) Defining App Stores: The Role of Curated Marketplaces in Software Ecosystems. In: Herzworm, G. and Margaria, T., Eds., *Software Business. From Physical Products to Software Services and Solutions. 4th International Conference, ICSOB 2013, Potsdam, Germany, 11-14 June 2013*, Springer, Berlin, Heidelberg, 195-206.
- [6] Urli, S., Blay-Fornarino, M., Collet, P., Mosser, S. and Riveill, M. (2014) Managing a Software Ecosystem Using a Multiple Software Product Line: A Case Study on Digital Signage Systems. *40th EUROMICRO Conference on Software Engineering and Advanced Applications*, Verona, 27-29 August 2014, 344-351.
<https://doi.org/10.1109/seaa.2014.23>
- [7] Albert, B.E., dos Santos, R.P. and Werner, C.M.L. (2013) Software Ecosystems Governance to Enable IT Architecture Based on Software Asset Management. *7th IEEE International Conference on Digital Ecosystems and Technologies (DEST)*, Menlo Park, CA, 24-26 July 2013, 55-60. <https://doi.org/10.1109/dest.2013.6611329>
- [8] Musil, J., Musil, A. and Biff, S. (2013) Elements of Software Ecosystem Early-Stage Design for Collective Intelligence Systems. *Proceedings of the 2013 International*

- Workshop on Ecosystem Architectures*, Saint Petersburg, 19 August 2013, 21-25. <https://doi.org/10.1145/2501585.2501590>
- [9] da Silva Amorim, S., Almeida, E.S.D. and McGregor, J.D. (2013) Extensibility in Ecosystem Architectures: an Initial Study. *Proceedings of the 2013 International Workshop on Ecosystem Architectures*, Saint Petersburg, 19 August 2013, 11-15.
- [10] Mens, T., Claes, M., Grosjean, P. and Serebrenik, A. (2014) Studying Evolving Software Ecosystems based on Ecological Models. In: Mens, T., Serebrenik, A. and Cleve, A., Eds., *Evolving Software Systems*, Springer, Berlin, Heidelberg, 297-326. https://doi.org/10.1007/978-3-642-45398-4_10
- [11] dos Santos, R.P., Esteves, M.S., Freitas, G. and de Souza, J. (2014) Using Social Networks to Support Software Ecosystems Comprehension and Evolution. *Social Networking*, **3**, 108-118. <https://doi.org/10.4236/sn.2014.32014>
- [12] Robillard, M.P. and Walker, R.J. (2014) An Introduction to Recommendation Systems in Software Engineering. In: Robillard, P.M., Maalej, W., Walker, J.R. and Zimmermann, T., Eds., *Recommendation Systems in Software Engineering*, Springer, Berlin, Heidelberg, 1-11. https://doi.org/10.1007/978-3-642-45135-5_1
- [13] Park, J.-G. and Lee, J. (2014) Knowledge Sharing in Information Systems Development Projects: Explicating the Role of Dependence and Trust. *International Journal of Project Management*, **32**, 153-165.
- [14] Lim, S.L., Bentley, P.J., Kanakam, N., Ishikawa, F. and Honiden, S. (2015) Investigating Country Differences in Mobile App User Behavior and Challenges for Software Engineering. *IEEE Transactions on Software Engineering*, **41**, 40-64. <https://doi.org/10.1109/TSE.2014.2360674>
- [15] Henderson-Sellers, B., Gonzalez-Perez, C., McBride, T. and Low, G. (2014) An Ontology for ISO Software Engineering Standards: 1) Creating the Infrastructure. *Computer Standards & Interfaces*, **36**, 563-576.
- [16] Di Ruscio, D., Paige, R.F., Pierantonio, A., Hutchinson, J., Whittle, J. and Rouncefield, M. (2014) Model-Driven Engineering Practices in Industry: Social, Organizational and Managerial Factors That Lead to Success or Failure. *Science of Computer Programming*, **89**, 144-161.
- [17] Ghapanchi, A.H., Wohlin, C. and Aurum, A. (2014) Resources Contributing to Gaining Competitive Advantage for Open Source Software Projects: An Application of Resource-Based Theory. *International Journal of Project Management*, **32**, 139-152.
- [18] Lettner, D., Angerer, F., Prahofner, H. and Grunbacher, P. (2014) A Case Study on Software Ecosystem Characteristics in Industrial Automation Software. *Proceedings of the 2014 International Conference on Software and System Process*, Nanjing, 26-28 May 2014, 40-49. <https://doi.org/10.1145/2600821.2600826>
- [19] Gawer, A. and Cusumano, M.A. (2014) Industry Platforms and Ecosystem Innovation. *Journal of Product Innovation Management*, **31**, 417-433. <https://doi.org/10.1111/jpim.12105>
- [20] Andrés, C., Camacho, C. and Llana, L. (2013) A Formal Framework for Software Product Lines. *Information and Software Technology*, **55**, 1925-1947.
- [21] Metzger, A. and Pohl, K. (2014) Software Product Line Engineering and Variability Management: Achievements and Challenges. *Proceedings of the on Future of Software Engineering*, Hyderabad, 31 May-7 June 2014, 70-84.
- [22] Harman, M., Jia, Y., Krinke, J., Langdon, W.B., Petke, J. and Zhang, Y. (2014) Search Based Software Engineering for Software Product Line Engineering: A Survey and Directions for Future Work. *Proceedings of the 18th International Software Product Line Conference*, Vol. 1, Florence, 15-19 September 2014, 5-18.

- <https://doi.org/10.1145/2648511.2648513>
- [23] Olyai, A. and Rezaei, R. (2015) Analysis and Comparison of Software Product Line Frameworks. *Journal of Software*, **10**, 991-1001. <https://doi.org/10.17706/jsw.10.8.991-1001>
- [24] Capilla, R., Bosch, J., Trinidad, P., Ruiz-Cortés, A. and Hinchey, M. (2014) An overview of Dynamic Software Product Line Architectures and Techniques: Observations from Research and Industry. *Journal of Systems and Software*, **91**, 3-23.
- [25] Krishnan, S., Strasburg, C., Lutz, R.R., Goseva-Popstojanova, K. and Dorman, K.S. (2013) Predicting Failure-Proneness in an Evolving Software Product Line. *Information and Software Technology*, **55**, 1479-1495.
- [26] Quadri, A. and Abubakar, M. (2015) Software Quality Assurance in Component Based Software Development—A Survey Analysis. *International Journal of Computer and Communication System Engineering*, **2**, 305-315
- [27] Singh, P.K., Sangwan, O.P., Singh, A.P. and Pratap, A. (2015) A Framework for Assessing the Software Reusability using Fuzzy Logic Approach for Aspect Oriented Software. *International Journal of Information Technology and Computer Science*, **7**, 12-20. <https://doi.org/10.5815/ijitcs.2015.02.02>
- [28] Yadav, H.B. and Yadav, D.K. (2015) A Fuzzy Logic Based Approach for Phase-Wise Software Defects Prediction Using Software Metrics. *Information and Software Technology*, **63**, 44-57.
- [29] Rettinger, A., Losch, U., Tresp, V., D'Amato, C. and Fanizzi, N. (2012) Mining the Semantic Web. *Data Mining and Knowledge Discovery*, **24**, 613-662. <https://doi.org/10.1007/s10618-012-0253-2>
- [30] Jeremić, Z., Jovanović, J. and Gašević, D. (2013) Personal Learning Environments on the Social Semantic Web. *Semantic Web-Linked Data for Science and Education*, **4**, 23-51.
- [31] Khriyenko, O. and Nagy, M. (2011) Semantic Web-Driven Agent-Based Ecosystem for Linked Data and Services. *3rd International Conferences on Advanced Service Computing*, Rome, 2011, 110-117.
- [32] Lécué, F., Tallevi-Diotallevi, S., Hayes, J., Tucker, R., Bicer, V., Sbodio, M. and Tommasi, P. (2014) Smart Traffic Analytics in the Semantic Web with STAR-CITY: Scenarios, System and Lessons Learned in Dublin City. *Web Semantics: Science, Services and Agents on the World Wide Web*, **27-28**, 26-33.
- [33] Ngan, L.D. and Kanagasabai, R. (2013) Semantic Web Service Discovery: State-of-the-Art and Research Challenges. *Personal and Ubiquitous Computing*, **17**, 1741-1752. <https://doi.org/10.1007/s00779-012-0609-z>
- [34] Demir, K.A. (2015) Multi-View Software Architecture Design: Case Study of a Mission-Critical Defense System. *Computer and Information Science*, **8**, 12-31. <https://doi.org/10.5539/cis.v8n4p12>
- [35] Aleti, A., Buhnova, B., Grunske, L., Koziolk, A. and Meedeniya, I. (2013) Software Architecture Optimization Methods: A Systematic Literature Review. *IEEE Transactions on Software Engineering*, **39**, 658-683. <https://doi.org/10.1109/TSE.2012.64>
- [36] Ginters, E., Schumann, M., Vishnyakov, A. and Orlov, S. (2015) Software Architecture and Detailed Design Evaluation. *Procedia Computer Science*, **43**, 41-52.
- [37] Yang, C., Liang, P. and Avgeriou, P. (2016) A Systematic Mapping Study on the Combination of Software Architecture and Agile Development. *Journal of Systems and Software*, **111**, 157-184.
- [38] Oussalah, M., Bhat, F., Challis, K. and Schnier, T. (2013) A Software Architecture for Twitter Collection, Search and Geolocation Services. *Knowledge-Based Systems*,

37, 105-120.

- [39] Capilla, R., Jansen, A., Tang, A., Avgeriou, P. and Babar, M.A. (2016) 10 Years of Software Architecture Knowledge Management: Practice and Future. *Journal of Systems and Software*, **116**, 191-205.
- [40] de M. Neves, A.R., Carvalho, Á.M.G. and Ralha, C.G. (2014) Agent-Based Architecture for Context-Aware and Personalized Event Recommendation. *Expert Systems with Applications*, **41**, 563-573.
- [41] Horcas, J.-M., Pinto, M. and Fuentes, L. (2016) An Automatic Process for Weaving Functional Quality Attributes Using A Software Product Line Approach. *Journal of Systems and Software*, **112**, 78-95.
- [42] Ayala, I., Amor, M., Fuentes, L. and Troya, J.M. (2015) A Software Product Line Process to Develop Agents for the IoT. *Sensors*, **15**, 15640-15660.
<https://doi.org/10.3390/s150715640>
- [43] Mück, T.R. and Fröhlich, A.A. (2014) A Metaprogrammed C++ Framework for Hardware/Software Component Integration and Communication. *Journal of Systems Architecture*, **60**, 816-827.
- [44] He, W. and Xu, L.D. (2014) Integration of Distributed Enterprise Applications: A Survey. *IEEE Transactions on Industrial Informatics*, **10**, 35-42.
<https://doi.org/10.1109/TII.2012.2189221>
- [45] Alférez, G.H., Pelechano, V., Mazo, R., Salinesi, C. and Diaz, D. (2014) Dynamic Adaptation of Service Compositions with Variability Models. *Journal of Systems and Software*, **91**, 24-47.
- [46] Bontcheva, K., Kieniewicz, J., Andrews, S. and Wallis, M. (2015) Semantic Enrichment and Search: A Case Study on Environmental Science Literature. *D-Lib Magazine*, **21**, 1-18. <https://doi.org/10.1045/january2015-bontcheva>
- [47] Krueger, R., Thom, D. and Ertl, T. (2015) Semantic Enrichment of Movement Behavior with Foursquare—A Visual Analytics Approach. *IEEE Transactions on Visualization and Computer Graphics*, **21**, 903-915.
<https://doi.org/10.1109/TVCG.2014.2371856>
- [48] Kunze, C. and Hecht, R. (2015) Semantic Enrichment of Building Data with Volunteered Geographic Information to Improve Mappings of Dwelling Units and Population. *Computers, Environment and Urban Systems*, **53**, 4-18.
- [49] Fileto, R., Bogorny, V., May, C. and Klein, D. (2015) Semantic Enrichment and Analysis of Movement Data: Probably It Is Just Starting! *SIGSPATIAL Special*, **7**, 11-18. <https://doi.org/10.1145/2782759.2782763>
- [50] Fileto, R., May, C., Renso, C., Pelekis, N., Klein, D. and Theodoridis, Y. (2015) The Baquara2 Knowledge-Based Framework for Semantic Enrichment and Analysis of Movement Data. *Data & Knowledge Engineering*, **98**, 104-122.
- [51] Trinidad, P. (2012) Automating the Analysis of Stateful Feature Models. PhD Dissertation, University of Seville, Spain.

Appendix A: Dynamic and Optimized Metadata-Based Reconfiguration Model (DOMRM)

This Appendix presents the details of the DOMRM model. The main idea behind DOMRM is the more a user uses a specific feature, the more his weight for this feature increases. The weight $UjFi$ of user j for feature i is given by:

$$UjFi = \frac{n(Uj, Fi)}{\sum_{k=1}^P n(Uk, Fi)} \quad (1)$$

where $n(Uj, Fi)$ denotes the number of times user j used the feature i .

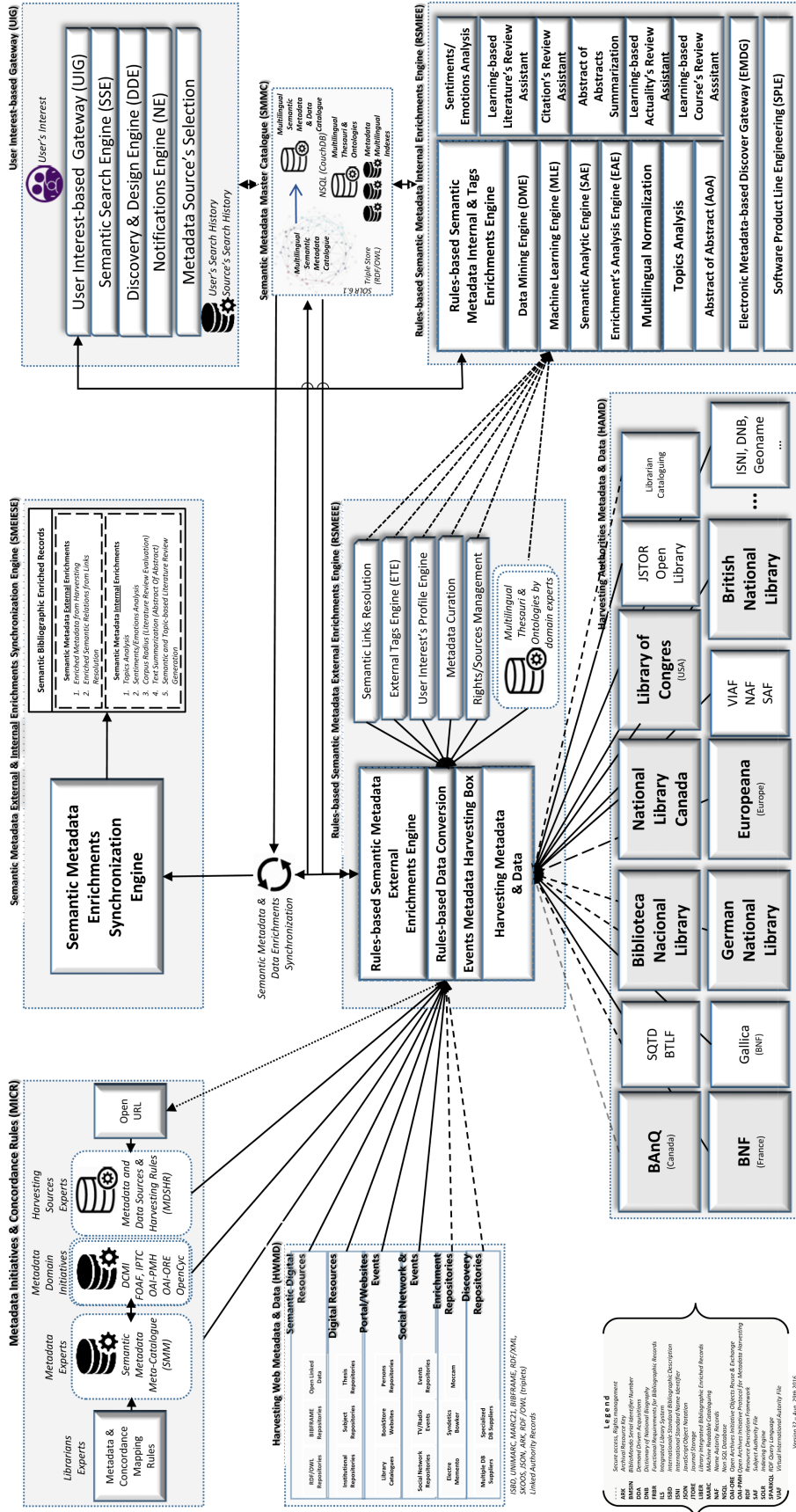
Making use of user weight per feature and their preferences, the feature weight that determines its activation or not is computed. Considering that US is the set of users who have selected a feature Fi (activation of feature), and UR is the set of users who have removed that feature (deactivation of feature), the value 1 is assigned when a user activates the feature, and -1 when he removes it. Let $c(Uj, Fi)$ be the choice of user j for the activation status of feature Fi . The weight of feature Fi can be defined using the following formula:

$$w(Fi) = \begin{cases} 1 & \text{whether } 0 < \sum_{Uk \in US \cup UR} [c(Uk, Fi) \times UkFi] \\ -1 & \text{whether } 0 > \sum_{Uk \in US \cup UR} [c(Uk, Fi) \times UkFi] \end{cases} \quad (2)$$

The computed weight of each feature allows one to define the weight FM that is used by the system optimal configurator with the FCM to generate the new configuration of the system for all users. When the feature weight is negative and the FIS rules allow de-activation, the feature is deactivated and when the feature weight is positive and the FIS rules allow activation the DOMRM model activates the feature. The activation status of the feature is not modified when the feature weight is null and the current activation status is conserved.

Appendix B: Proposed SMESE (Semantic Metadata Enrichments Software Ecosystem)

SMESE V3 - Semantic Metadata Enrichments Software Ecosystem



© Ronald Brisebois 2017

Appendix C: Figure 7. SMESE Meta Entity Model

META ENTITY													
Id.	Oldproperty Name	Property Name	Name	Description	BM label			Client Label		Source	Port Folio Entity Name	Domain	
					Fr	En	Sp	Gr	Ne				Fr
E1-n	Book	book	Book	C'est les contenus du type livre	Livre	Book	"	"	"	Livre	Book	BNF	bibliomondo
E2-n	Serial	serial	Serial	C'est les contenus du type périodique	Périodique	Serial	"	"	"	Périodique	Serial		bibliomondo
E3-n	Audio	audio	Audio	C'est les contenus du type audio	Audio	Audio	"	"	"	Audio	Audio		bibliomondo
E4-n		comic	Comic	C'est les contenus du type bande dessinée	Bande dessinée	Comic	"	"	"	Bande dessinée	Comic		bibliomondo
E5-n		Digital resource	Digital Resource	C'est les contenus du type ressource numérique	Ressource Numérique	Digital Resource	"	"	"	Ressource Numérique	Digital Resource		bibliomondo
E6-n		document	Document	C'est les contenus du type document	Document	Document	"	"	"	Document	Document		bibliomondo
E7-n		game	Game	C'est les contenus du type jeu	Jeu	Game	"	"	"	Jeu	Game		bibliomondo
E8-n	Image	image	Image	C'est les contenus de type Image	Image	Image	"	"	"	Image	Image	BNF	bibliomondo
E9-n	Musical Score	Musical score	Musical Score	C'est les contenus du type partition de musique	Partition de musique	Musical Score	"	"	"	Partition de musique	Musical Score		bibliomondo
E10-n	Video	video	Video	C'est les contenus du type vidéo	Vidéo	Video	"	"	"	Vidéo	Video		bibliomondo
E11-n	FRBR Work	Creative work	Creative work	The most generic kind of creative work, including books, movies, photographs, software programs, etc	Oeuvre	FRBRWork	"	"	"	Oeuvre	FRBRWork	BNF	bibliomondo
E12-n		manufactation	Manufactation	C'est la manifestation d'une oeuvre	Manifestation	Manifestation	"	"	"	Manifestation	Manifestation		bibliomondo
E13-n		expression	Expression	C'est l' expression d'une manisfestation	Expression	Expression	"	"	"	Expression	Expression		bibliomondo
E14-n		concept	Concept	C'est une concept	Concept	Concept	"	"	"	Concept	Concept		bibliomondo
E15-n		city	City	C'est les sous divisons d'un canton	Ville	City	"	"	"	Ville	City	BM	places
E16-n		Postal Address	Postal Address	C'est les contenus de type Adresse	Adresse postal	Postal address	"	"	"	Adresse postal	Postal address	BNF	places
E17-n		Place of interest	Place Of Interest	C'est un lieu qui suscite un intérêt particuliié; par exemple, un site touristique, un musée, monument,...	Place d'intérêt	Place of interest	"	"	"	Place d'intérêt	Place of interest	BM	places
E18-n		country	Country	C'est un pays	Pays	Country	"	"	"	Pays	Country	BNF	places
E19-n		region	Region	C'est les sous divisions d'un pays; c'est le cas des régions en France, des États au États-Unis ou des provinve au Canada C'est les sous divisions d'un regions; par exemple les comtés au États unis ou les départements en France	Région	Region	"	"	"	Région	Region	BM	places

Appendix D: Figure 8. SMESE Metadata Model

META METADATA																							
Id.	Unimarc	Marc21	Old property Name	Property Name (BanQ)	Property Name (Excel)	Description	BM label					Client Label		Expected Type	Is Related	Related Content Type	Short Notice	Enriched	Data Location	HiddenRepeatable	Thesaurus	Type isUrl	Source Of Media Schema Version
							Fr	En	Sp	Ne	Gr	Fr	En										
M1-m	200\$acdehi	OP0245	OP0245	title	title	It is the main title of a content	Titre	Title				Titre	Title	Text							Desc	BM	
M2-m	700\$abcdfg4/ 710\$abcdefgh4/ 720\$a4	OP01XX	OP01XX	author	author	The author of this content. Please note that author is special in that HTML 5 provides a special mechanism for indicating authorship via the rel tag. That is equivalent to this and may be used interchangeably	Auteur	Author				Auteur	Author	Text		X	Local			Personnes ou Organisations	Desc	BNF	
M3-m	701\$abcdfg4/ 711\$abcdefgh4/ 721\$a4			creator	creator	The creator/author of this Creative Work. This is the same as the Author property for Creative Work	Auteur	Author				Compositeur/ Interprète	Composer/ Performer	Text		X	Local			Personnes ou Organisations	Desc	BNF	
M4-m				promoter	promoter	C'est le organisateur du contenu; c'est le cas des événements	Organisateur	Promoter				"	"	Text			Local			Personnes ou Organisations	Desc	BM	
M5-m	215\$a4e	OP0300	OP0300	dcterms: description	description	A short description of the item.	Description	Description				Description	Description	Multiline Text			Local	X			Desc	BNF	
M6-m				Image	Image	An image of the item. This can be a image file or a URL or a fully described Image Object.	Imagette	Thumbnail				Imagette	Thumbnail	Image			X	BM.com	X		Desc	BNF	
M7-m	101\$a	LO0546	LO0546	dcterms: language	language	C'est la langue (parlée ou écrite) dans laquelle le contenu est créé ou dans laquelle une personne s'exprime; par exemple, un livre rédigé en français aura comme langue le français	Langue(s)	Language (s)				Langue(s)	Language(s)	Text			Local	X	Langues	Desc	BM		
M8-m	540\$a4e	OP0740	OP0740	Title Added Uncontrolled Related	Title Added Uncontrolled Related	C'est d'autres titres du contenu qui pourraient être utilisés dans d'autres catalogue	Autres titres	Other titles				Autres titres	Other titles	Text			Local				Desc	BM	
M9-m	500\$abhijklmn qrsuvw	OP0240	OP0240	Title Uniform	Title Uniform	C'est le titre unifié du contenu	Titre uniforme	Uniform title				Titre uniforme	Uniform title	Text			Local				Desc	BM	
M10-m				Also Available In	Also Available In	C'est les autres langues dans lesquelles le contenu a été aussi produit ou dans lesquelles une personne s'exprime	Disponible en d'autres langues	Available in other languages				Aussi disponible en	Also available in	Text			Local	X	Langues	Desc	HEC OIC		

