

A Semantic Metadata Enrichment Software Ecosystem based on Metadata and Affinity Models

Ronald Brisebois

École de technologie supérieure, University of Quebec, Montreal, Canada
E-mail: ronald.brisebois.1@ens.etsmtl.ca

Alain Abran¹ and Apollinaire Nadembega²

¹École de technologie supérieure, University of Quebec, Montreal, Canada
E-mail: alain.abran@etsmtl.ca

²Network Research Lab., University of Montreal, Montreal, Canada
E-mail: Apollinaire.nadembega@umontreal.ca

Received: 22 February 2017; Accepted: 09 May 2017; Published: 08 August 2017

Abstract—Information systems need to be more flexible and to allow users to find content related to their context and interests. Metadata harvesting and metadata enrichments could represent a way to help users to find content and events according to their interests. However, metadata are underused and represents an interoperability challenge. This paper presents a new framework, called SMESE, and the implementation of its prototypes that consists of its semantic metadata model, a mapping ontology model and a user interest affinity model. This proposed framework makes these models interoperable with existing metadata models.

SMESE also proposes a decision support process supporting the activation and deactivation of software features related to metadata. To consider context variability into account in modeling context-aware properties, SMESE makes use of an autonomous process that exploits context information to adapt software behavior using an enhanced metadata framework. When the user chooses preferences in terms of system behavior, the semantic weight of each feature is computed. This weight quantifies the importance of the feature for the user according to their interests.

This paper also proposed a semantic metadata analysis ecosystem to support data harvesting according to a metadata model and a mapping ontology model. Data harvesting is coupled with internal and external enrichments. The initial SMESE prototype represents more than 400 millions of relationships (triplets). To conclude, this paper also presents the design and implementation of different prototypes of SMESE applied to digital ecosystems.

Index Terms—Metadata, metadata enrichment, metadata model, ontology, semantic metadata enrichment, software ecosystem.

I. INTRODUCTION

With more and more data available on the web, how users search and discover content or events is of crucial importance. There is growing research on interaction paradigms investigating how users may benefit from (1) the expressive power of semantic web standards, (2) the existing cataloguing models and metadata enrichments.

The semantic web may be defined as the transformation of the world wide web to a database of semantic linked resources, where data may be widely reused and shared [1]. Semantic information discovery approaches [2, 3] are now challenging traditional keyword-based information retrieval methods. This retrieval problem is further burdened by the poor quality of the metadata content in many digital collections.

Software ecosystems (SECO) [4-19] are defined as the interaction of a set of actors on top of a common technological ecosystem providing a number of software interactions or webservices [4, 5]. In SECO, internal and external actors create and compose relevant solutions together with a community of domain experts and users to satisfy customer requirements. This poses new challenges since the software systems are being evolved by various distributed development teams, communities, experts and technologies.

There is growing agreement on the main characteristics of SECO, including a common technological platform enabling outside contributions and variability-enabled architectures. Nine characteristics have been identified [20] that focus on technical processes for system development, interconnection and evolution.

Gawer and Cusumano [21] have analyzed a wide range of industry examples of SECO and identified two predominant types of platforms:

1. Internal platforms: 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: 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.

Concurrently modern software demands more and more adaptive features. The semantic web [22-26] and linked data are some of the most important concepts to support Semantic Metadata Enrichment (SME) in a SECO architecture [27-33].

Today, semantic web technologies offer a new level of flexibility, interoperability and a way to enhance peer communication and knowledge sharing. Indeed, a semantic web engine, 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 engines try to understand the context in which the words are being used, resulting in more relevant results with greater user satisfaction.

However, to enrich web data by transforming them into knowledge that may be more accessible and understandable by systems and users, this paper proposes a framework using metadata model architecture, referred as the SMESE framework (Semantic Metadata Enrichment Software Ecosystem).

The SMESE architecture includes semantic metadata enrichment engines based on a metadata model, a mapping ontologie model and a user interest affinity model. It integrates and enriches metadata.

SMESE also proposes a decision support process supporting the activation and deactivation of software features related to metadata. To consider context variability into account in modeling context-aware properties, SMESE makes use of an autonomous process that exploits context information to adapt software behavior using an enhanced metadata framework.

The multi-platform metadata model of SMESE was presented in [34] while this paper focuses specifically on the metadata and affinity models of SMESE.

The remainder of the paper is organized as follows. Section 2 presents the related works. Section 3 summarizes the multi-platform framework of the proposed SMESE, and Section 4 presents the related eight metadata and affinity models and sub-systems of SMESE. Section 5 presents the prototype of SMESE implementation in an industry context. Section 6 presents a summary and ideas for future work.

II. RELATED WORKS

This related works section is at the intersection of SECO and SME and presents the three related research axes:

1. SECO architecture using component integration.
2. SECO architecture and concepts.
3. Semantic metadata enrichments (SME).

The related works section is at the intersection of SECO and SME. First, the SECO architecture is presented, second, the concept and finally the semantic metadata enrichments.

A. SECO architecture using components integration

Software ecosystems [4-6, 12, 21, 28, 32] consist of multiple software products, often interrelated to each other by means of dependency relationships. When one product undergoes changes and issues a new release, this may or may not lead other products 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 [4] report that:

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

They define a SECO as the interaction of a set of actors on top of a common technological platform. They also identify three main perspective in a SECO architecture:

1. Software engineering: the focus is on technical issues related directly or indirectly to the technology platform.
2. Business and management: the focus is on the business, organizational and management aspects.
3. Relationships: represent the social aspect.

B. SECO architecture and concept

Christensen, Hansen, Kyng and Manikas [5] define the concept of SECO architecture as a set of structures comprised of actors and software elements, the relationships among them, and their properties.

Demir [27] also proposes a software architecture that is strongly related to a defence system and limited to military personnel. Their multi-view of the SECO architecture is described step by step.

Neves, Carvalho and Ralha [33] 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.

Alferez, Pelechano, Mazo, Salinesi and Diaz [35] propose a framework that uses semantically rich variability models at runtime to support the dynamic adaptation of service compositions. They propose that service compositions be abstracted as a set of features in a variability model.

C. Semantic metadata enrichment

Bontcheva, Kieniewicz, Andrews, and Wallis [36] 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.

Some authors focus their enrichment model on person mobility trace data [37, 40]. Krueger, Thom, and Ertl [37] 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.

Kunze and Hecht [38] 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.

The conclusions from these related works are:

1. Metadata-based architecture needs to be flexible and meet administrative, organizational and technical aspects.
2. Several proposed models do not take into account autonomic mechanisms to guide the self-adaptation of service compositions according to changes in the computing infrastructure.
3. There is no SECO architecture that takes into account several semantic enrichment aspects.
4. 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.
5. Current metadata and entity enrichment models do not take into account person mobility trace data gathering and analysis in the enrichment process of metadata.

III. SMESE ARCHITECTURE

This section presents the architecture of the proposed Semantic Metadata Enriched Software Ecosystem (SMESE). It is based on metadata semantic internal and external enrichments and their interoperability. Each component of the SMESE architecture is based on semantic metadata to generate, extract, discover and enrich metadata based on mapping ontologies and a user interest affinity model. SMESE makes use of contents and linked data analysis.

For the new generation of information and data

management, metadata is one of the most efficient material for data aggregation and understanding. 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 using appropriate metadata. To make this feasible, content and events need to be semantically enriched. In other words, to achieve specific searches, specific metadata must be available including semantic topics, sentiments and abstracts. However, at the present time and according to our prototype, more than 85% of the content does not have these metadata.

The SMESE prototypes include an engine to aggregate multiple catalogues or datasets from the web, libraries, universities, bookstores, #tag collections, museums, and cities. 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. They are in different formats and are also called either base index, unified index, or foundation index.

The SMESE framework enhances bibliographic records with semantic metadata enrichments. It searches and discovers actual collections or novelties, including: works, books, DVDs, CDs, comics, games, pictures, videos peoples, legacy collections, organizations, rewards, TVs, radios, museums and other events calendar. The prototype creates triplets to define relationships enriching metadata's content. To be able to map the user interest and the content metadata, the prototype includes a user interest affinity model. This model (see Fig. 1) includes:

1. An algorithm to recommend user contents or events matching his interest according to the user interest affinity model;
2. An algorithm to rank dynamically the contents or events according to the user interest affinity model;

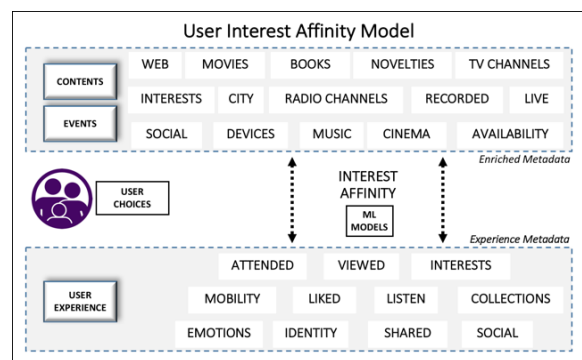


Fig.1. User Interest Affinity Model.

Semantic relationships between the contents, persons, organization and places are defined and curated in the master metadata catalogue. Topics, sentiments and emotions are extracted automatically from the contents but with respect to their context. 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 (triplets).

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. Many international organizations have partly done this heavy work and already have much bibliographic metadata converted into triple-stores but there is not a definition of a common catalogue using the same semantic metadata model for all standards.

The SMESE prototypes harvest and analyse multiple catalogues and linked open data (LOD) from libraries, universities, bookstores, #tag collections, museums, open catalogues, national catalogues to produce semantic metadata enrichments.

Central indexes typically include full text and citations from publishers, full texts, abstracting and indexing from aggregators and subscription databases, and different formats (such as MARC) from library catalogues.

The SMESE framework allows to connect bibliographic records and semantic metadata enrichments (SEM) into a unified master metadata catalogue. The next figure (Fig. 2) presents the four levels of the metadata enrichment view uses par SMESE: (1) Meta-Entity (black), (2) Entity (blue), (3) Semantic metadata enrichment (grey), and (4) Contents & Events (white).

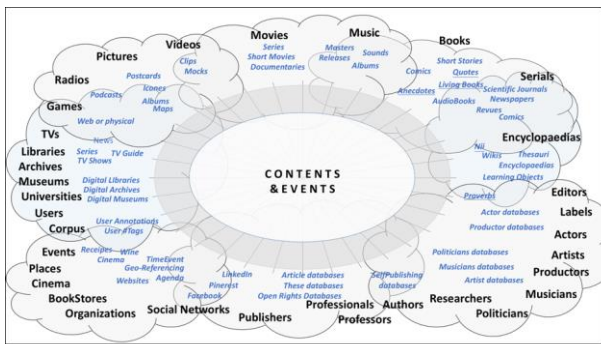


Fig.2. Metadata enrichment view.

Semantic relationships between content, persons, organizations, events and places are defined and curated in the master metadata catalogue. Topics and sentiments are extracted (where possible) from the content, its context and related objects.

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 machine. Enriched catalogue metadata are needed to support the discovery and notification of information within an information community.

SMESE includes an automated approach for semantic metadata enrichment that allows users to perform interest-based semantic search or discovery more efficiently. To summarize, SMESE makes the following contributions:

A. Architecture, prototype and analysis of SMESE – Semantic Metadata Enrichment Software Ecosystem. (See Fig. 3 Detailed of the ecosystem; Appendix A shows a more readable version).

This new semantic ecosystem SMESE has the ability to harvest and enrich bibliographic records externally (from the web) and internally (from text data). The main components of the ecosystem are (see Fig. 3 and Appendix A shows a readable version):

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

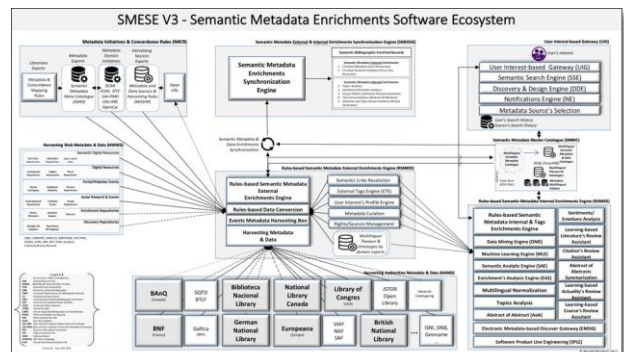


Fig.3. Detailed Semantic Enriched Metadata Software Ecosystem [34].

B. Topic detection/generation - A prototype was developed to automate the generation of topics from the text of a document using our algorithm 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.
2. Extending topic model 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 SMESE ecosystem to be able to find a new topic itself.

C. Sentiment 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; SMESE use

- semantic information to perform his analysis.
2. The collection of documents and paragraphs are taken into account. In terms of granularity, most of the existing approaches are sentence-based.
 3. In SMESE prototypes, the surrounding context of the sentence is included. The traditional approaches do not take into account the surrounding context of the sentence which may cause some misunderstanding with discovery of

sentiment.

The prototype makes use of the proposed algorithm SSEA (Semantic Sentiment and Emotion Analysis). This algorithm fulfills all the attributes of Table 1.

The SMESE extends the SECO characteristics presented in [20] from number 10 to 12 [34]. See Table 1 SECO characteristics versus SMESE characteristics.

Table 1. SMESE characteristics [34].

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

IV. SUBSYSTEMS WITHIN THE SMESE ARCHITECTURE

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

A. Metadata initiatives & concordance rules

This sub-section presents the details of the Metadata initiatives & concordance rules, specifically the semantic metadata meta-catalogue as shown in Fig. 3.

Metadata is a 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 metadata 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, applicable 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.

Actually there is no common meta-model that allows the creation of universal, understandable and readable meta-model, that would describe all entities used in all the libraries.

The most popular metadata models are:

1. Dublin Core (DC): primarily designed to provide a simple resource description format for networked resources.

2. UNIMARC: consists of data formulated by highly controlled cataloguing codes.
3. MARC21: is both flexible and extensible and allows users to work with data in ways specific to individual library needs.
4. RDF/RDA: mainly in Europe, it includes FRBR capability.
5. BIBFRAME: mainly in North America, it includes FRBR capability.

There is no known mapping model among these that would make them interoperable. The overall challenge is to prototype: (1) a meta model of partial international standardization of entities, (2) a model of partial metadata mapping ontology and user interest affinity model.

In addition, organizations create digital collections and generate metadata in repository silos. In general, 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. Indeed, aggregators usually ignore idiosyncratic use of metadata schemas and enforce the use of designated metadata fields.

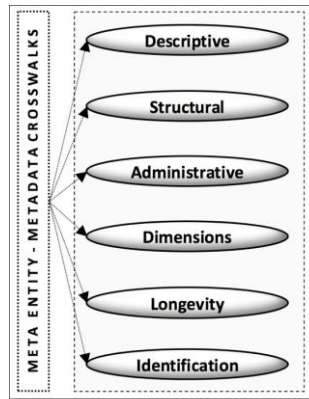


Fig.4. Semantic metadata model.

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 from different catalogues. 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. Fig. 4 shows the semantic metadata model that SMESE propose to address these issues.

In SMESE the proposed metadata are classified into six different categories:

1. Descriptive metadata: describes and identifies information resources at the local (system) level to enable searching and retrieving at the web-level, and to enable users to discover resources. 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 metadata and entities.
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 metadata: is a new classification that aim to increase user satisfaction, in terms of expected interests and emotions. Dimension metadata regroups all metadata about space, time, emotions and interests. Another example: emotions may suggest specific content to a particular user at a specific time and place. Furthermore, the source identifies the provenance and the rights relative to the creation of the metadata.
5. Longevity metadata: is a new classification that aim to manage the rights related to the content (entity).

6. Identification metadata: is a new classification that aim to manage the type of form or support of the media that contains the content (entity).

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. 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), described through Unique Resource Identifiers (URIs) [36]. In addition, International Standard Names Identifier (ISNI) is proposed by National Libraries to organize and catalogue the semantic metadata relationships, see Fig. 5 adapted from the source [41] where the symbol with three blue dots represents a semantic repository using triplets. The BNF is identifying workflows with publishers to provide them with ISNIs for new authors. ISNI represents the opportunity to help to enrich an author's metadata and the quality of the authority files. ISNI Semantic relationships allow to connect together many sources of information such as:

1. Wikipedia,
2. Wikidata,
3. Union List of Artist Names,
4. IdRef,
5. Data.bnf.fr,
6. BNF Catalog,
7. SNAC,
8. AGORHA,
9. VIAF,
10. Data.banq.ca.

Fig. 5 shows also the introduction of ISNI semantic relationships into the semantic metadata meta-catalogue of the SMESE prototype.

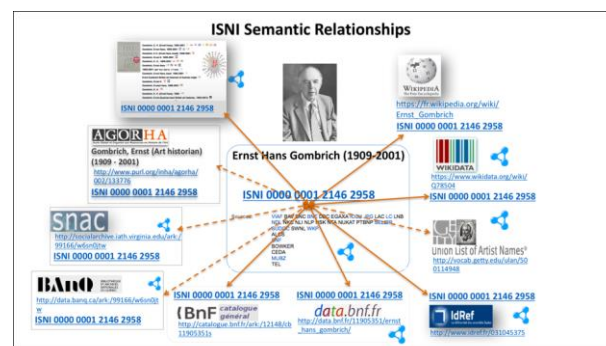


Fig.5. ISNI semantic relationships of semantic metadata meta-catalogue in the SMESE prototype (adapted from [41]).

The original contents should be enriched with relevant knowledge from the respective LOD resources. This is needed to answer queries that require common-sense knowledge, which is often not present in the original content. For example: following semantic enrichment, a semantic search for events that provide specific emotions (e.g., happiness, joy, etc.) in Montreal according to

individual interests this weekend would provide relevant metadata about events in Montreal, even though not explicitly mentioned in the original content metadata.

The semantic annotation process of SMESE creates relationships between semantic models, such as ontologies and persons. It may be characterized as the semantic enrichment of unstructured and semi-structured contents with new knowledge and linking these to relevant domain ontologies/knowledge bases. This requires the usage of ISNI, or other authority files or other techniques.

These processes extract, analyze and catalogue

metadata for topic and sentiments involved in the SMESE ecosystem. As of today, 5 million records (entity) have been harvested over a potential target of close to 500 million, see Table 2 for an overview of the detail about harvested metadata and data (p.e. papers and events) in the prototype. For each content type many metadata and data have been extracted and enriched. These enrichment processes are based on information retrieval and knowledge extraction approaches. The text is analyzed by means of extensions of text mining algorithms such as latent Dirichlet allocation (LDA), latent semantic analysis (LSA), support vector machine (SVM) and k-Means.

Table 2. Harvesting statistics related to metadata

No	URL Sources	Status	%	Total Content	Total harvested
1	https://www.academia.edu/	h	1	15 363 735	142 217
2	https://www.researchgate.net/	h	0	100 000 000	27 984
3	http://www.opendoar.org/	h	0	235 828 824	385 155
4	http://www.amazon.com/	h	1	4 703 063	44 322
5	http://www.paulines.qc.ca/	h	76	171 120	130 000
6	http://exedre.leslibrairies.ca/	h	63	171 120	107 680
7	http://www.fnac.com/	h	67	156 224	104 109
8	http://www.librairiepoirier.ca/	f	100	176 162	176 162
9	http://www.archambault.ca/	f	100	165 405	165 405
10	http://www.renaudbray.com/accueil.aspx	f	100	347 380	347 380
11	http://www.usherbrooke.coop/	f	100	47 412	47 412
12	http://www.biblairie.qc.ca/	f	100	213	213
13	http://librairiemediaspaul.ca	f	100	29 938	29 938
14	https://www.leslibraires.fr/	f	100	888 750	888 750
15	http://www.mollat.com/	f	100	505 729	505 729
TOTAL				465 379 760	2 967 439

Status: f: finished and h: harvesting

SMESE 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.

The proposed SMESE framework uses an SPLE architecture that is a combination of FORM and COPA to catalogue semantically different contents.

SMESE also proposes a decision support process called SPLE-DSP. It supports the activation and deactivation of software features related to metadata and takes into account automatic runtime reconfiguration according to different scenarios. To take context variability into account in modeling context-aware properties, SPLE-DSP makes use of an autonomous process that exploits context information to adapt software behavior using a generic metadata model.

When the user chooses preferences in terms of system behavior, the semantic weight of each feature is computed based on the software 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 other features. For example, a rule may be: feature Fi should

never be activated when Fi-1 is activated. Based on this rule, the FCM 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 users' selections. Note that individual users have their own weight per feature, defined on the basis of that user's use of the feature. This weight quantifies the importance of the feature for the user.

B. 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.

C. Harvesting authority metadata & data

This sub-section presents the details of the Harvesting of Authority Metadata & Data are presented in Fig. 6.

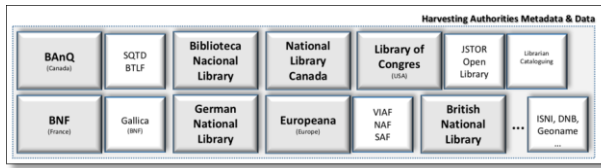


Fig.6. Harvesting of authority metadata & data.

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

D. Rules-based semantic metadata external enrichments engine

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

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 and context. The rule-based semantic metadata external enrichment engine consists of:

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

Semantic annotation methods are good candidates to enrich the content with disambiguated domain terms and entities (e.g. events, sentiments, interests, locations, organizations and persons), described through Unique Resource Identifiers (URIs) [36]. In addition, the original contents should be enriched with relevant knowledge from the respective linked open data resources (e.g. that Barack Obama is an American politician or Justin Trudeau is a Canadian politician). This is needed to answer queries that require common-sense knowledge, which is often not present in the original content. For example: following semantic enrichment, a semantic search for events that provides specific emotions (e.g., happiness, joy) in New York (or another city) according to individual interests this weekend would indeed provide relevant metadata about events in New York (or another city), even though not explicitly mentioned in the original content metadata. Furthermore, the linguistic aspect (context) of the knowledge is critical to analyse the metadata and corresponding data or content.

The semantic annotation process of SMESE creates relationships between semantic models, such as ontologies and persons. It may be characterized as the semantic enrichment of unstructured and semi-structured contents with new knowledge and linking these to relevant domain ontologies/knowledge bases. It typically

requires annotating a potentially ambiguous entity mention with the canonical identifier of the correct unique entity. The benefit of social semantic enrichment is that by surfacing annotated terms derived from the full-text content, concepts buried within the body of the paper/report can be highlighted. Also, the addition of terms affects the relevance ranking in full-text searches. Moreover, users can be more specific by limiting the search criteria to the subject or interest or emotion metadata (e.g. through a faceted search).

E. Rule-based semantic metadata internal enrichments engine

This sub-section presents the details of the rule-based semantic metadata internal enrichment engine. This sub-system includes:

1. A rule-based semantic metadata internal enrichment engine.
2. A topic, sentiment/emotion, abstract analysis and an automatic literature review.

These processes extract, analyze and catalogue metadata for topics and sentiments 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, latent semantic analysis, support vector machine and k-Means. 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 is that it is not specific to one software product but can be applied to many products dynamically. In addition, it includes two semantic metadata enrichment (SME) processes to improve the quality of search and discovery engines; the external process who analyses the context of the data while harvesting and the internal process who analyses the content of the data.

F. 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. Mainly this engine has the objective to find out the new content and context from the last harvesting.

G. User interest-based gateway

This sub-section presents the user interest-based gateway that represents the person (mobile or stationary) who interacts with the SMESE ecosystem. This engine use the metadata created by SMESE to give better results or recommendation to the users. The users and

contributors are categorized into five groups:

1. Interest-based gateway
2. Semantic Search Engine
3. Discovery
4. Notifications
5. Metadata source selection.

H. Semantic master catalogue

This semantic master catalogue (SMC) represents the knowledge base of the SMESE ecosystem based on his evolving meta model of metadata. The SMC aggregates all triplets and their relationships created by the engines of SMESE. SMC includes also all the thesaurii and ontologies for a specific domain of interest.

V. AN IMPLEMENTATION OF SMESE FOR DIGITAL ECOSYSTEMS

The proposed SMESE architecture has been implemented for some digital ecosystems. The SMESE prototypes implement partially its metadata model and framework. The catalogue contains more than 2 million items, with 18 entities and 132 defined metadata. One of the prototype identifies 1453 metadata and defines a semantic classification.

First, we defined a list of entities, called Meta Entity, which introduced 193 items. These items represent all library materials. The structure of the model allows addition of new entities as may be required. 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. 1341 metadata have been defined.

This classification allows users to search content according to their interests. Fig. 7 shows an illustration of the Metadata model; Appendix B shows a readable version.

Fig.7. SMESE prototype metadata model.

The semantic matrix model is defined for each entity based on the meta entity 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. Fig. 8 illustrates an example of a semantic metadata matrix for a specific content; Appendix C presents a readable version. The objective behind the matrix is to allow the reuse of metadata for distinct entities.

Fig.8. Entities of a SMESE semantic matrix model.

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.

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 of depth, we developed atomic metadata: these metadata allow description of the most elementary aspects of an entity. It then becomes easy to map all metadata from any schema.

This OWL file from the ontology is used by a crosswalk to automatically assign metadata value that are harvested from distinct sources.

A total of 94,015,090 metadata records were collected from these different sources:

1. From Discogs (www.discogs.com) for music, we collected 7,983,288 entities: 2,621,435 music releases, 4,466,660 artists and 895,193 labels.
2. From ResearchGate (www.researchgate.net) for academic papers, we collected 86,031,802 entities: 77,031,802 publications and more than 9,000,000 researchers.
3. From academia (www.academia.edu) for academic papers, we collected 143,277 entities: 135,101 publications and more than 8,175 researchers.
4. From TV hebdo (www.tvhebdo.com) for TV channel programs, we collected 268,147,499 entities: 385 TV channel and 268,147,114 TV programs.
5. From OpenDOAR (www.opendoar.org) for scientific contents, we collected 235,828,824 entities: 96,265,327 thesis and 139,563,497 publications.

SMESE now contains more than 4.3 billion triplets and is growing.

VI. SUMMARY AND FUTURE WORK

In this paper, we proposed a design and implementation of SMESE, a semantic enriched metadata software ecosystem including a user interest affinity

model. The SMESE prototype, integrates data and metadata enrichment to support internal and external metadata enrichments.

SMESE also includes a decision support process. It supports the activation and deactivation of software features related to metadata. To take context variability into account in modeling context-aware properties, SMESE makes use of an autonomous process that exploits context information to adapt software behavior using a generic metadata model. When the user chooses preferences in terms of system behavior, the semantic weight of each feature is computed based on the software feature configuration model. Individual users have their own weight per feature, defined on the basis of that user's use of the feature. This weight quantifies the importance of the feature for the user according to their interests.

We also presented our implementation of SMESE including the semantic metadata model. The ontology mapping model was then implemented to make the models interoperable with existing metadata models.

This paper proposed a semantic metadata enrichments software ecosystem to support multi-platform metadata driven applications. SMESE integrates data and metadata based on mapping ontologies in order to enrich them and create a semantic master metadata catalogue. SMESE prototype represents more than 400 million relationships (triplets).

The major contributions of this paper are as follows:

1. Definition of a metadata-based software ecosystem:

- a. Enhancing the SECO characteristics from 9 to 12.
- b. The use of a LOD-based semantic enrichment model for semantic annotation processes.
- c. A repository of 43 thesaurii included in RAMEAU for semantical contextualization of concepts.
- d. An extended latent Dirichlet allocation algorithm for topic analysis.

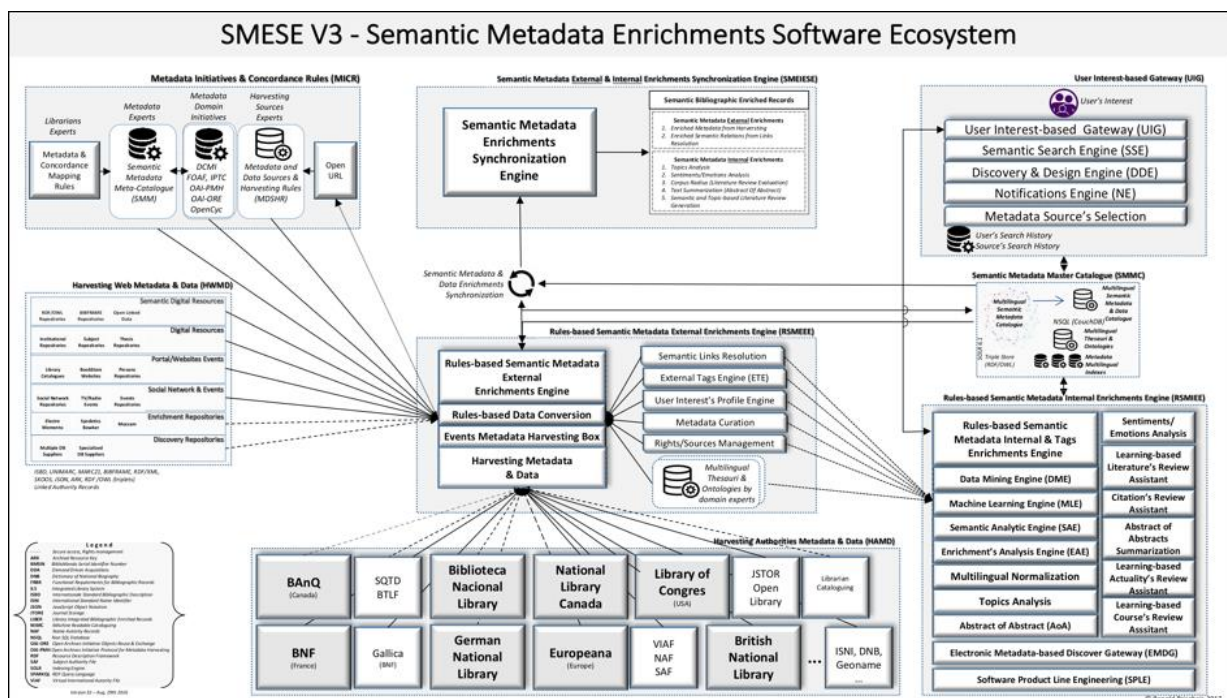
2. Prototype of SMESE ecosystem for harvesting data and metadata and generating semantic metadata enrichments.
3. Prototype of a user interest affinity model.
4. The design and implementation of an SMESE prototype for different standards in digital ecosystems.

Future work related to SMESE ecosystem will include:

1. Some enhancements to be able to enrich metadata semantically, including evolving user interest.
2. Further evaluations of the affinity model with different prototype and datasets.

Exploring text summarization and automatic literature review as metadata enrichments. The semantic annotations could be used to enrich metadata and provide further data to improve the user interest affinity model.

APPENDIX A: FIG. 3. - SMESE FRAMEWORK: SEMANTIC ENRICHED METADATA SOFTWARE ECOSYSTEM



APPENDIX B: FIG. 7. - SMESE METADATA MODEL

Id.	Unimarc	Marc21	propertyName(Excel)	Description	BM label Fr	BM label En	BM label Sp	BM label Ne	BM label Gr	MetaData type	Expected response	isUri	Repeatable	Indexed	relatedContent type	nestingRelation	Thesaurus	Enriched	isTranslatable	Data Location	Mixing	Hidden (Form level)	isNotDisplayed (bibliographic record level)	Short Notice	Onlet Edit	Source Schema	InMedia Version
M1-m	200\$acdehi	130\$anplskfdmor 245\$ahbcnp 247\$abfgnp 700\$t	title	It is the main title of a content	Titre	Title	Título				Multilingue											X	X		Desc	BM	
M2-m	7XX	100\$aqbcdnplskfg u 110\$aqbcdnplskfg u 111\$aqbcdnplskfg u 971\$a4 3100\$a4	author	content. Please note that HTML 5 provides a special mechanism for indicating authorship via the rel tag. That is equivalent to this and	Auteur	Author	Autor				Text		X	X	domain/personality		Personnes ou Organisations			Local				X	Desc	BNF	
M3-m	700\$abdcfg4 710\$abdcfg4 711\$abdcfg4 720\$a4 721\$a4	100\$aqbcdnplskfg u 110\$aqbcdnplskfg u 111\$aqbcdnplskfg u 971\$a4 3100\$a4	creator	The creator/author of this CreativeWork. This is the same as the Author property for CreativeWork	Auteur	Author	Autor				Text		X	X	domain/personality		Personnes ou Organisations			Local				X	Desc	BNF	
M4-m	7xx	1XX	promoter	du contenu; c'est le	Organisateur	Promoter	Promotor				Text		X	X	sonality		nes			Local					Desc	BM	
M5-m	206\$abdcdef 230\$a 215\$acde	255\$abdcdef 256\$a 300\$abcceg	description	C'est une texte qui décrit le contenu. An image of the item. This can be a image file or a URL or a fully described image object	Description	Description	Descripción				Multilingue		X						X	Local		X	X		Desc	BNF	
M6-m			image								Image		X							BM.com					Desc	BNF	

APPENDIX C: FIG. 8. - EXAMPLE OF A SMESE SEMANTIC MATRIX MODEL

Id.	Unimarc	Marc21	PropertyName(Excel)	Cartographic (Périodique)	Serial (Périodique)	Physical	Image	Event	Book chapter	PressRelease	Law (loi et règlements)	Research Report (Rapport de recherche)	MDAdm (Rapport de recherches)	Organization Report (Rapport d'organisme)	Thesis	Conference Paper	Conference Proceedings	General Assembly	Speech (Discours)	Periodic allsue (Numéro de périodique)	Periodical Article (Article de périodique)	
M1-m	200\$acdehi	130\$anplskfdmor 245\$ahbcnp 247\$abfgnp 700\$t	title	3	3	13	13		6	3	4	6	3	41	30	3	35	35	1	1	3	9
M2-m	7XX	100\$aqbcdnplskfg u 110\$aqbcdnplskfg u 111\$aqbcdnplskfg u 971\$a4 3100\$a4	author	4	4	126	241		134	72		3	3	4		4	42	42	2	2		4
M3-m	700\$abdcfg4 710\$abdcfg4 711\$abdcfg4 720\$a4 721\$a4	100\$aqbcdnplskfg u 110\$aqbcdnplskfg u 111\$aqbcdnplskfg u 971\$a4 3100\$a4	creator		3	3	5	3	48	69		72		58	95	89	89	3	3		100	
M4-m	7xx	1XX	promoter					47														
M5-m	206\$abdcdef 230\$a 215\$acde	255\$abdcdef 256\$a 300\$abcceg	description	15	15	39	35	50	7	13	41	88	5	114	21	5	127	132	62	57	4	134
M6-m			image																			

REFERENCES

[1] J. Lacasta, J. Nogueras-Iso, G. Falquet, J. Teller, and F. J. Zarazaga-Soria, "Design and evaluation of a semantic enrichment process for bibliographic databases," *Data & Knowledge Engineering*, vol. 88, pp. 94-107, 2013. doi:http://dx.doi.org/10.1016/j.datak.2013.10.001

[2] Junaid Rashid, and M. W. Nisar, "A Study on Semantic Searching, Semantic Search Engines and Technologies Used for Semantic Search Engines," *International Journal of Information Technology and Computer Science(IJITCS)*, vol. 8, no. 10, pp. 82-89, 2016. doi:http://dx.doi.org/10.5815/ijitcs.2016.10.10

[3] Hepsiba Mabel V, and J. Selwyn, "A Review on the Knowledge Representation Models and its Implications," *International Journal of Information Technology and Computer Science(IJITCS)*, vol. 8, no. 10, pp. 72-81, 2016. doi:http://dx.doi.org/10.5815/ijitcs.2016.10.09

[4] K. Manikas, and K. M. Hansen, "Software ecosystems – A systematic literature review," *Journal of Systems and Software*, vol. 86, no. 5, pp. 1294-1306, 2013. doi:http://dx.doi.org/10.1016/j.jss.2012.12.026

[5] H. B. Christensen, K. M. Hansen, M. Kyng, and K. Manikas, "Analysis and design of software ecosystem architectures – Towards the 4S telemedicine ecosystem," *Information and Software Technology*, vol. 56, no. 11, pp. 1476-1492, 2014. doi:http://dx.doi.org/10.1016/j.infsof.2014.05.002

[6] T. Shinozaki, Y. Yamamoto, and S. Tsuruta, "Context-based counselor agent for software development ecosystem," *Computing*, vol. 97, no. 1, pp. 3-28, 2015. doi:http://dx.doi.org/10.1007/s00607-013-0352-y

[7] S. Jansen, and E. Bloemendal, "Defining App Stores: The Role of Curated Marketplaces in Software Ecosystems," *Software Business. From Physical Products to Software Services and Solutions: 4th International Conference, ICSOB 2013, Potsdam, Germany, June 11-14, 2013. Proceedings*, G. Herzwurm and T. Margaria, eds., pp. 195-206, Berlin, Heidelberg: Springer Berlin Heidelberg, 2013. doi:http://dx.doi.org/10.1007/978-3-642-39336-5_19

[8] S. Urli, M. Blay-Fornarino, P. Collet, S. Mosser, and M. Riveill, "Managing a Software Ecosystem Using a

- Multiple Software Product Line: A Case Study on Digital Signage Systems." pp. 344-351, 2014. doi:http://dx.doi.org/10.1109/SEAA.2014.23
- [9] B. E. Albert, R. P. d. Santos, and C. M. L. Werner, "Software ecosystems governance to enable IT architecture based on software asset management." pp. 55-60, 2013. doi:http://dx.doi.org/10.1109/DEST.2013.6611329
- [10] J. Musil, A. Musil, and S. Biffl, "Elements of software ecosystem early-stage design for collective intelligence systems," in Proceedings of the 2013 International Workshop on Ecosystem Architectures, Saint Petersburg, Russia, 2013, pp. 21-25. doi:http://dx.doi.org/10.1145/2501585.2501590
- [11] S. d. S. Amorim, E. S. D. Almeida, and J. D. McGregor, "Extensibility in ecosystem architectures: an initial study," in Proceedings of the 2013 International Workshop on Ecosystem Architectures, Saint Petersburg, Russia, 2013, pp. 11-15. doi:http://dx.doi.org/10.1145/2501585.2501588
- [12] T. Mens, M. Claes, P. Grosjean, and A. Serebrenik, "Studying Evolving Software Ecosystems based on Ecological Models," *Evolving Software Systems*, T. Mens, A. Serebrenik and A. Cleve, eds., pp. 297-326, Berlin, Heidelberg: Springer Berlin Heidelberg, 2014. doi:http://dx.doi.org/10.1007/978-3-642-45398-4_10
- [13] R. dos Santos, P. , M. Esteves, S. , G. Freitas, and J. de Souza, "Using Social Networks to Support Software Ecosystems Comprehension and Evolution," *Social Networking*, vol. 3, no. 2, pp. 108-118, 2014. doi:http://dx.doi.org/10.4236/sn.2014.32014
- [14] M. P. Robillard, and R. J. Walker, "An Introduction to Recommendation Systems in Software Engineering," *Recommendation Systems in Software Engineering*, P. M. Robillard, W. Maalej, J. R. Walker and T. Zimmermann, eds., pp. 1-11, Berlin, Heidelberg: Springer Berlin Heidelberg, 2014. doi:http://dx.doi.org/10.1007/978-3-642-45135-5_1
- [15] J.-G. Park, and J. Lee, "Knowledge sharing in information systems development projects: Explicating the role of dependence and trust," *International Journal of Project Management*, vol. 32, no. 1, pp. 153-165, 2014. doi:http://dx.doi.org/10.1016/j.ijproman.2013.02.004
- [16] S. L. Lim, P. J. Bentley, N. Kanakam, F. Ishikawa, and S. Honiden, "Investigating Country Differences in Mobile App User Behavior and Challenges for Software Engineering," *IEEE Transactions on Software Engineering*, vol. 41, no. 1, pp. 40-64, 2015. doi:http://dx.doi.org/10.1109/TSE.2014.2360674
- [17] B. Henderson-Sellers, C. Gonzalez-Perez, T. McBride, and G. Low, "An ontology for ISO software engineering standards: 1) Creating the infrastructure," *Computer Standards & Interfaces*, vol. 36, no. 3, pp. 563-576, 2014. doi:http://dx.doi.org/10.1016/j.csi.2013.11.001
- [18] D. Di Ruscio, R. F. Paige, A. Pierantonio, J. Hutchinson, J. Whittle, and M. Rouncefield, "Model-driven engineering practices in industry: Social, organizational and managerial factors that lead to success or failure," *Science of Computer Programming*, vol. 89, pp. 144-161, 2014. doi:http://dx.doi.org/10.1016/j.scico.2013.03.017
- [19] A. H. Ghapanchi, C. Wohlin, and A. Aurum, "Resources contributing to gaining competitive advantage for open source software projects: An application of resource-based theory," *International Journal of Project Management*, vol. 32, no. 1, pp. 139-152, 2014. doi:http://dx.doi.org/10.1016/j.ijproman.2013.03.002
- [20] D. Lettner, F. Angerer, H. Prahofner, and P. Grunbacher, "A case study on software ecosystem characteristics in industrial automation software," in Proceedings of the 2014 International Conference on Software and System Process, Nanjing, China, 2014, pp. 40-49. doi:http://dx.doi.org/10.1145/2600821.2600826
- [21] A. Gawer, and M. A. Cusumano, "Industry Platforms and Ecosystem Innovation," *Journal of Product Innovation Management*, vol. 31, no. 3, pp. 417-433, 2014. doi:http://dx.doi.org/10.1111/jpim.12105
- [22] A. Rettinger, U. Losch, V. Tresp, C. D'Amato, and N. Fanizzi, "Mining the Semantic Web," *Data Min. Knowl. Discov.*, vol. 24, no. 3, pp. 613-662, 2012. doi:http://dx.doi.org/10.1007/s10618-012-0253-2
- [23] Z. Jeremić, J. Jovanović, and D. Gašević, "Personal learning environments on the Social Semantic Web," *Semantic Web - Linked Data for science and education*, vol. 4, no. 1, pp. 23-51, 2013. doi:http://dx.doi.org/10.3233/SW-2012-0058
- [24] O. Khriyenko, and M. Nagy, "Semantic Web-driven Agent-based Ecosystem for Linked Data and Services." pp. 110-117, 2011
- [25] F. L'écuyer, S. Tallewi-Diotallewi, J. Hayes, R. Tucker, V. Bicer, M. Sbodio, and P. Tommasi, "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*, vol. 27-28, pp. 26-33, 2014. doi:http://dx.doi.org/10.1016/j.websem.2014.07.002
- [26] L. D. Ngan, and R. Kanagasabai, "Semantic Web service discovery: state-of-the-art and research challenges," *Personal and Ubiquitous Computing*, vol. 17, no. 8, pp. 1741-1752, 2013. doi:http://dx.doi.org/10.1007/s00779-012-0609-z
- [27] K. A. Demir, "Multi-View Software Architecture Design: Case Study of a Mission-Critical Defense System," *Computer and Information Science*, vol. 8, no. 4, pp. 12-31, 2015
- [28] A. Aleti, B. Buhnova, L. Grunske, A. Koziolok, and I. Meedeniya, "Software Architecture Optimization Methods: A Systematic Literature Review," *IEEE Transactions on Software Engineering*, vol. 39, no. 5, pp. 658-683, 2013. doi:http://dx.doi.org/10.1109/TSE.2012.64
- [29] E. Ginters, M. Schumann, A. Vishnyakov, and S. Orlov, "Software Architecture and Detailed Design Evaluation," *Procedia Computer Science*, vol. 43, pp. 41-52, 2015. doi:http://dx.doi.org/10.1016/j.procs.2014.12.007
- [30] C. Yang, P. Liang, and P. Avgeriou, "A systematic mapping study on the combination of software architecture and agile development," *Journal of Systems and Software*, vol. 111, pp. 157-184, 2016. doi:http://dx.doi.org/10.1016/j.jss.2015.09.028
- [31] M. Oussalah, F. Bhat, K. Challis, and T. Schnier, "A software architecture for Twitter collection, search and geolocation services," *Knowledge-Based Systems*, vol. 37, pp. 105-120, 2013. doi:http://dx.doi.org/10.1016/j.knosys.2012.07.017
- [32] R. Capilla, A. Jansen, A. Tang, P. Avgeriou, and M. A. Babar, "10 years of software architecture knowledge management: Practice and future," *Journal of Systems and Software*, vol. 116, pp. 191-205, 2016. doi:http://dx.doi.org/10.1016/j.jss.2015.08.054
- [33] A. R. d. M. Neves, Á. M. G. Carvalho, and C. G. Ralha, "Agent-based architecture for context-aware and personalized event recommendation," *Expert Systems with Applications*, vol. 41, no. 2, pp. 563-573, 2014. doi:http://dx.doi.org/10.1016/j.eswa.2013.07.081
- [34] R. Brisebois, A. Abran, and A. Nadembega, "A Semantic

Metadata Enrichment Software Ecosystem (SMESE) based on a Multi-platform Metadata Model for Digital Libraries,” *Journal of Software Engineering and Applications (JSEA)*, vol. 10, pp. 370-405, 2017. doi:<http://dx.doi.org/10.4236/jsea.2017.104022>

- [35] G. H. Alf ez, V. Pelechano, R. Mazo, C. Salinesi, and D. Diaz, “Dynamic adaptation of service compositions with variability models,” *Journal of Systems and Software*, vol. 91, pp. 24-47, 2014. doi:<http://dx.doi.org/10.1016/j.jss.2013.06.034>
- [36] K. Bontcheva, J. Kieniewicz, S. Andrews, and M. Wallis, “Semantic Enrichment and Search: A Case Study on Environmental Science Literature,” *D-Lib Magazine*, vol. 21, no. 1-2, pp. 1-18, 2015
- [37] R. Krueger, D. Thom, and T. Ertl, “Semantic Enrichment of Movement Behavior with Foursquare-A Visual Analytics Approach,” *IEEE Transactions on Visualization and Computer Graphics*, vol. 21, no. 8, pp. 903-915, 2015. doi:<http://dx.doi.org/10.1109/TVCG.2014.2371856>
- [38] C. Kunze, and R. Hecht, “Semantic enrichment of building data with volunteered geographic information to improve mappings of dwelling units and population,” *Computers, Environment and Urban Systems*, vol. 53, pp. 4-18, 2015. doi:<http://dx.doi.org/10.1016/j.compenvurbsys.2015.04.002>
- [39] R. Fileto, V. Bogorny, C. May, and D. Klein, “Semantic enrichment and analysis of movement data: probably it is just starting!,” *SIGSPATIAL Special*, vol. 7, no. 1, pp. 11-18, 2015. doi:<http://dx.doi.org/10.1145/2782759.2782763>
- [40] R. Fileto, C. May, C. Renso, N. Pelekis, D. Klein, and Y. Theodoridis, “The Baquara2 knowledge-based framework for semantic enrichment and analysis of movement data,” *Data & Knowledge Engineering*, vol. 98, pp. 104-122, 2015. doi:<http://dx.doi.org/10.1016/j.datak.2015.07.010>
- [41] A. ANGJELI, “ISNI, For an worldwide Identification Ecosystem,” *Institut National de l’Histoire de l’Art (INHA)*, 2016

Authors’ Profiles



Ronald Brisebois is currently a PhD student at the  cole de Technologie Sup rieure (ETS) – Universit  du Qu bec (Montr al, Canada). He received a B. Science in Physics at University of Montreal in 1983, a BA in Computer Science at University of Quebec in 1985 and his MBA at Hautes  tudes Commerciales - HEC (Business School) in 1989. From 1989 to 1993, Ronald Brisebois was a professor of Software Engineering at the University of Sherbrooke. His PhD research focus on semantic web, artificial intelligence, autonomous software architecture, new generation software designing, enriched metadata modeling and software engineering.

Renowned entrepreneur in the field of information technology, Ronald Brisebois has held management positions in various top-level firms (Caisses populaires Desjardins). In 1991, he was a professor at the University of Sherbrooke; in 1992, he founded his first company. Cognicase Inc. quickly became one of the largest players in the information technology field in Canada. In 2003, Ronald created Mondo/Stellar, one of the leading providers of integrated solutions for public libraries, academic institutions, specialized and consortia systems worldwide.



Dr. Abran holds a Ph.D. in Electrical and Computer Engineering (1994) from  cole Polytechnique de Montr al (Canada) and master degrees in Management Sciences (1974) and Electrical Engineering (1975) from University of Ottawa (Canada).

He is a professor at the  cole de Technologie Sup rieure (ETS) – Universit  du Qu bec (Montr al, Canada). He has over 20 years of experience in teaching in a university environment as well as 20 years of industry experience in information systems development and software engineering management. His research interests include software productivity and estimation models, software engineering foundations, software quality, software functional size measurement, software risk management and software maintenance management. He has published over 400 peer-reviewed papers. He is the author of the books ‘Software Project Estimation’, ‘Software Metrics and Software Metrology’ and a co-author of the book ‘Software Maintenance Management’ (Wiley Interscience Ed. & IEEE-CS Press).

Dr. Abran is also the 2004 co-executive editor of the Guide to the Software Engineering Body of Knowledge – SWEBOK (see ISO 19759 and www.swebok.org) and he is the chairman of the Common Software Measurement International Consortium (COSMIC) – <http://cosmic-sizing.org/>. A number of Dr. Abran research works have influenced international standards in software engineering (i.e., ISO 19761, ISO 19759, ISO 14143-3, etc.).



Dr. Apollinaire Nadembega is currently a guest member of the Network Research Laboratory (NRL) of the University of Montreal. He received his B. E degree in Information Engineering from Computer Science High School, Bobo-Dioulasso, Burkina faso in 2003, his Master’s degree in computer science from the Arts and Business Institute, Ouagadougou, Burkina faso in 2007 and his Ph.D. degree in mobile networks from the University of Montreal, Montreal, QC, Canada in 2014. The primary focus of his Ph.D. thesis is to propose a mobility model and bandwidth reservation scheme that supports quality-of-service management for wireless cellular networks. Dr. Nadembega’s research interests lie in the field of artificial intelligence, machine learning, networking modelling, semantic web, metadata management system, software architecture, mobile multimedia streaming, call admission control, bandwidth management and mobile cloud computing.

From 2004 to 2008, he was a programming engineer with Burkina Faso public administration staff management office.

How to cite this paper: Ronald Brisebois, Alain Abran, Apollinaire Nadembega, “A Semantic Metadata Enrichment Software Ecosystem based on Metadata and Affinity Models”, *International Journal of Information Technology and Computer Science (IJITCS)*, Vol.9, No.8, pp.1-13, 2017. DOI: 10.5815/ijitcs.2017.08.01