Guillermo Perasso. A Case for FRBR and Semantic Wikis in Enterprise Information Environments. A Master's paper for M.S. in I.S. degree. April 2013. 55 pages. Advisor: Gary Marchionini.

Wikis allow users to collaboratively create and maintain content. As a new platform for wiki sites, Semantic wikis provide additional means to annotate the content to add structure. These Semantic Wiki sites are experiencing an enormous increase in popularity because structured data is more usable and thus more valuable than unstructured data. This study proposes the use of a semantic wiki to develop an web portal to collect and organize the information maintained in an Enterprise's content management system. The ontology selected to support the conceptual infrastructure of the portal is based on a scheme inspired from an IFLA proposition known as Functional Requirements for Bibliographic Records (FRBR).

The paper introduces a model for the specification of mappings between FRBR entities and the organization's information artifacts. Also, the information system stakeholders - process owners, document creators, SMEs (Subject Matter Experts), support reps, end users, etc.- are also defined and interrelated using simple standard ontologies and vocabularies such as Friend-of-a-Friend (FOAF), RDF and OWL. The FRBR ontology and the web portal are developed using the platform provided by Semantic MediaWiki, an extension of MediaWiki, the platform for Wikipedia. The main features for standard vocabularies integration and semantic queries and searches are summarized. Finally, a number of metrics and indicators are presented as a reference for the portal's project managers and sponsors to evaluate the performance and effectiveness of the tool and help them make decisions about future initiatives, enhancements and new information requirements implementation.

Headings:

Electronic Data Processing – Ontologies

Websites – Wikis

Semantic Integration – Semantic Web

Metadata – Document Type Definitions

A CASE FOR FRBR AND SEMANTIC WIKIS IN ENTERPRISE INFORMATION ENVIRONMENTS

by
Guillermo Perasso

A Master's paper submitted to the faculty
of the School of Information and Library Science
of the University of North Carolina at Chapel Hill
in partial fulfillment of the requirements
for the degree of Master of Science in
Information Science.

Chapel Hill, North Carolina

April, 2013

Approved by:

____

Gary Marchionini

Table of Content

1. Introduction and Roadmap

This study is intended to describe a target organizational environment and its main information resources: information systems, user needs and enterprise workflows, contextual constraints and issues related to the configuration and management of these resources. A semantic web technology-based solution is presented to address the information management of the described system.

The solution proposed is based on the design of an ontology using common models, languages and vocabularies and the implementation of a web portal using a semantic wiki system. The objective is to promote collaboration and easy access to the information collected and organized by the ontology.

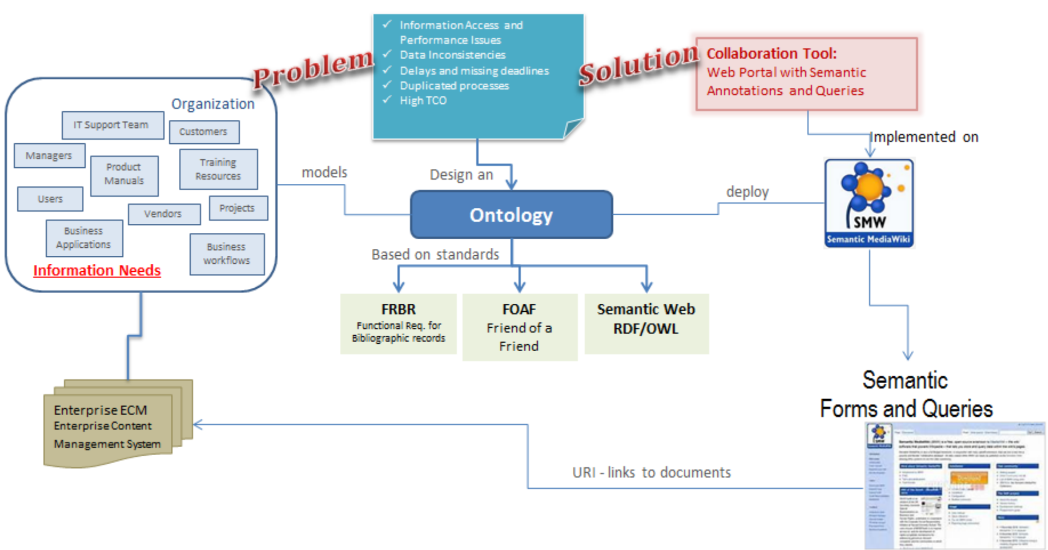

Figure 1: The solution workflow and key concept

Figure 1 shows the main concepts covered in this paper. The cornerstone of the presented process is the construction of an ontology to model the organization resources. This ontology is mostly based on the FRBR (Functional Requirements for Bibliographic Records) model and other standard languages and vocabularies from semantic web technologies, like RDF (Resource Description Framework) and FOAF (Friend Of A Friend).

The solution, a web portal, is implemented using Semantic MediaWiki. It is intended to collect information about resources (documents, files, database records) created and stored in the organization content management system (CMS). This knowledge management portal organizes available documents, their content summary and description, their location in the servers (URIs, hyperlinks) and, most important of all, the agents (employees and other stakeholders) related to these resources. Connections between agents and stored documentation are defined according to the different roles these agents play within the enterprise information landscape: creators, owners, sponsors, trainers, SMEs (subject matter experts), end users, IT support, and other relevant roles.

One of the main purposes of the ontology presented in this study is to elicit and structure data about informal relationships among agents; that is, those relationships that go beyond the organization charts and enterprise areas and departments. For instance, who are employees with certain kind of background or expertise in certain resources and software tools that could be consulted for support or advice.

2. Problem Definition: The Information Integration challenge

One of the main reasons preventing an organization to reach their data integration goals is a tendency toward information silos. Information workers (and that includes just about everyone in any organization nowadays) have to focus on solving their business problems first, and will create information artifacts that assist them to this end.

The widespread popularity of spreadsheets attests to this trend; they provide power into the hands of the information workers themselves, allowing them to organize information in a way that helps them with their own particular task.

These spreadsheets and other small information solutions are a problem, there is no way to make sure that they are consistent, or to take advantage at an enterprise scale, of information that is found in them. At best, they represent valuable information assets that are not being used to their complete potential. At worst, they are points of resistance to bringing in new, more comprehensive systems.

This situation results in the following barriers to allowing an enterprise to take full advantage of the data it has:

1. Commitment to legacy data. At all levels of the enterprise, data has been organized in a particular way for a particular purpose. Even as the needs of the enterprise outgrow this data organization, it is difficult to let go. Workers have familiarity with the information structure; they know where they can go to answer everyday questions. The difficulty with which new information can be found from old structures stymies innovation, encouraging the status quo.

2. Commitment to legacy work process. Going hand-in-hand with the legacy data is legacy work process. Innovation is difficult, and the enterprise has to keep making money, even in an outdated mode. Legacy work processes make other, non-functional, demands on data infrastructure. Enterprises are accustomed to having top-down control of corporate data. This tendency can be so strong that some enterprises will simply ignore the existence of desktop data (i.e., spreadsheets) to maintain the fiction that enterprise data is organized from the top down. There are strong drivers for maintaining top-down control, ranging from data quality (with a central curator who controls what goes into the database), to issues of privacy and security for corporate data.

3. Massive size of the indexing problem. In document-centric situations, the massive volume of documents presents a daunting challenge to any attempt at indexing. Incentive structures that focus on document creation, but not document indexing, result in a large, undifferentiated document corpus. The size of the backlog makes it difficult to get a start.

The implementation of databases and information systems such as Knowledge Management Systems (KM) and ECMS (Enterprise Content Management Systems) is not enough to cover all the information needs of the organization. Documents, spreadsheets, training material, presentations and other valuable information items are stored in these systems, but users are not provided with the proper resources to search and find these documents in a timely and effective manner.

This work presents a solution based on ontology building, semantic web technologies and wiki tools to design a knowledge portal to improve the performance of the whole organization by promoting the collaboration and data sharing among all the participants in the targeted information system.

3. Using Ontologies for Information Integration

Information integration has been recognized as a significant problem in enterprises for some years, and it is a problem of considerable economic importance, well before the Semantic Web was conceived, and before the use of ontologies were a major subject of research in Information Science.

The term ontology is originated from Philosophy and was adopted by AI (artificial intelligence) researchers to describe formal domain knowledge. Several ontology definitions have been proposed in the last decades. The most frequently cited definition is that given by Gruber in 1993, that is, ontology is defined as “an explicit specification of a conceptualization”. In other words, an ontology is a domain model (conceptualization) which is explicitly described (specified).

Within the context of this work, ontologies are defined as a computational artifact that encodes knowledge about this domain in a machine-readable form to make it available to information systems. In various application contexts, and within different communities, ontologies have been explored from different points of view, and there exist several definitions of what an ontology is. Within the Semantic Web community the most popular definition of an ontology is taken from Studer et al. **[1]**:

Ontology: *A formal explicit specification of a shared conceptualization of a domain of interest.*

An ontology used in an information system is a conceptual yet executable model of an application domain. It is made machine-interpretable by means of knowledge representation techniques and can therefore be used by applications to base decisions on reasoning about domain knowledge.

They describe a conceptualization in general terms and does not only capture a particular state of affairs. Instead of making statements about a specific situation involving particular individuals, an ontology tries to cover as many situations as possible that can potentially occur.

The Semantic Web

The **Semantic Web** is a collaborative movement led by the international standards body, the World Wide Web Consortium (W3C) **[2]**. According to this organization, "The Semantic Web provides a common framework that allows data to be shared and reused across application, enterprise, and community boundaries."

The main purpose of the Semantic Web is driving the evolution of the current Web by enabling users to find, share, and combine information more easily. Its infrastructure is built using RDF (Resource Description Framework), a general method for describing information, described in more detail below, in section 6.2.

Generic Functionalities of Ontologies in the Semantic Web

The term Semantic Technologies or Semantic Web Technologies shall denote the whole range of methods and tools typically used in applications that rely on a formal ontology (or several ontologies) and explicit metadata for information items or information systems, in order to enhance information search, integration, processing, or management, especially in distributed and open scenarios.

Such Semantic Technologies comprise core aspects like ontology engineering and management, as well as metadata creation and management, but also contributing and underlying base technologies like natural language processing **[3]** or automated reasoning **[4]**.

Semantic Technologies in Organizations: the Corporate Semantic Web.

Behind the firewall of a company, an Intranet application may be much easier to realize than a similar Internet application, from technical and from nontechnical points of view (trust, standards compliance, incentive systems, etc.); but, nevertheless, many company-internal information landscapes provide challenging-enough problems. Some corporate Intranets today are bigger than the Internet was 10 years ago.

While Semantic Web focuses primarily on fundamental technologies, Corporate Semantic Web focuses on pragmatic aspects of transferring semantic technologies into productive usage. Besides realizing semantic applications it also includes reviewing the economic aspects (e.g. cost models) of their development and management. Thus, it can help decision makers on the strategic, tactical, and operational level to understand the impact and benefit of semantic technologies.

There are three main areas of the Corporate Semantic Web: ontology engineering, semantic applications, and collaboration. Ontology engineering considers the efficient and effective development of ontologies to lessen the costs of ontology development and maintenance. The area of semantic applications analyzes existing applications and evaluates to what extent they could benefit from semantic technologies, for instance search on the basis of background knowledge (semantic search). Collaboration focuses on the human-centered aspects of knowledge management in corporate contexts. For example, extracting explicit knowledge from the interaction of users within enterprises.

The solution presented in this study covers these three areas of the corporate semantic web:

Ontology Engineering	> Proposed Approach: FRBR, FOAF ontologies.
Semantic Applications	> Proposed Solution: Semantic Portal
Collaboration	> Proposed Approach: Semantic Wiki Platform

4. FRBR: An Ontology for Enterprise Information Resources

Identifying documents and their relationships: This work proposes a documentation identification scheme inspired from the FRBR, Functional Requirements for Bibliographic Records.

This section is extracted from the documentation provided by IFLA, the International Federation of Library Associations and Institutions **[5]**. The goal is to build an ontology for information resources within the organization based on the FRBR model, with some modifications to adapt the concepts and links to the target environment.

The Functional Requirements for Bibliographic Records (FRBR) is a conceptual model of the bibliographic universe based on an entity-relationship schema **[6]**. It describes entities, their attributes, and relationships among these entities. The model follows a hierarchical structure and proposes a more comprehensive, holistic approach compared to previous methods developed for retrieval and access in online library catalogues and bibliographic databases, as the links between entities can be navigated through the hierarchy defined by the model.

Within the FRBR model, IFLA divides entities into three groups. As described in the final report released by IFLA in 2009 **[7]:**

> *"The first group comprises the products of intellectual or artistic endeavor that are named or described in bibliographic records: work, expression, manifestation, and item (see Figure 2). The second group comprises those entities responsible for the intellectual or artistic content, the physical production and dissemination, or the custodianship of such products: person and corporate body. The third group comprises an additional set of entities that serve as the subjects of intellectual or artistic endeavor: concept, object, event, and place".*

Group 1 Entities: Work, Expression, Manifestation, Item.

The entities defined as work (a distinct intellectual or artistic creation) and expression (the intellectual or artistic realization of a work) reflect intellectual or artistic content. The entities defined as manifestation (the physical embodiment of an expression of a work) and item (a single exemplar of a manifestation), on the other hand, reflect physical form.

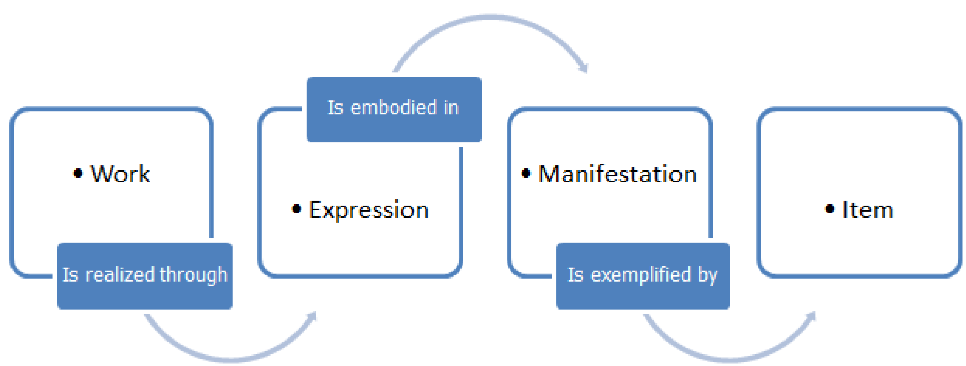

Figure 2: The FRBR Group 1 Entities.

The following key concepts and definitions are extracted and summarized from the FRBR Final Report produced by the IFLA Study Group on the Functional Requirements for Bibliographic Records [7].

Work

According to IFLA, in FRBR a work:

> *"Is an abstract entity; there is no single material object one can point to as the work. We recognize the work through individual realizations or expressions of the work, but the work itself exists only in the commonality of content between and among the various expressions of the work".* Regarding the scope of a work, it states: *"because the notion of a work is abstract, it is difficult to define precise boundaries for the entity. The concept of what constitutes a work and where the line of demarcation lies between one work and another may in fact be viewed differently from one organization to another. Consequently the bibliographic conventions established by various communities, organizations or groups may differ in terms of the criteria they use for determining the boundaries between one work and another".*

Expression

This concept is defined as:

> *"An expression is the specific intellectual or artistic form that a work takes each time it is "realized." An expression excludes aspects of physical form, such as typeface and page layout, that are not integral to the intellectual or artistic realization of the work as such".*

Manifestation

Whereas work and expression are clearly defined as abstract entities, manifestation is defined as a concrete representation, what IFLA calls a "physical embodiment":

> "*This physical embodiment of an expression constitutes a Manifestation. As an entity, manifestation represents all the physical objects that bear the same characteristics, in respect to both intellectual content and physical form. The boundaries between one manifestation and another are drawn on the basis of both intellectual content and physical form. When the production process involves changes in physical form the resulting product is considered a new manifestation".*

Item

Items is defined as a specific copy of a manifestation. According to IFLA:

> *"A single exemplar of a manifestation constitutes an Item. An item is a concrete entity. In terms of intellectual content and physical form, an item exemplifying a manifestation is normally the same as the manifestation itself".*

The FRBR proposition allows us to establish distinctions and precise relationships between the various intellectual creations - artifacts - handled during an information system project. Various terms are used by creators and publishers of intellectual and artistic entities to signal relationships between those entities. Terms such as "edition" and "version" are frequently encountered on publications and other materials, as are statements such as "based on ..." or "translated from ....". FRBR represents specifically relationships that operate between one work and another, between a work and an expression, between one expression and another, between a manifestation and an item, etc.

For the semantic portal project presented in this paper, FRBR Group 1 entities will be adapted to identify and represent user requirements, business rules, training resources, databases, and other information resources. These mappings are specified below, in section 7.

Group 2 Entities: Person, Corporate Body

The FRBR model specifies that:

> *"The entities in the second group represent those responsible for the intellectual or artistic content, the physical production and dissemination, or the custodianship of the entities in the first group. The entities in the second group include person (an individual) and corporate body (an organization or group of individuals and/or organizations)".*

IFLA also describes the links between group 2 and group 1 entities using the concept of responsibility. Responsibility relationships constitute one of the pillars of the ontology developed in this paper for the semantic portal. From the FRBR report:

> *"There is a type of "responsibility" relationship that exists between entities in the second group and the entities in the first group. A work may be created by one or more than one person and/or one or more than one corporate body. Conversely, a person or a corporate body may create one or more than one work. An expression may be realized by one or more than one person and/or corporate body; and a person or corporate body may realize one or more than one expression. A manifestation may be produced by one or more than one person or corporate body; a person or corporate body may produce one or more than one manifestation. An item may be owned by one or more than one person and/or corporate body; a person or corporate body may own one or more than one item".*

Group 3 Entities: Concept, Object, Event, Place

The entities in the FRBR third group represent, as stated in the IFLA report:

> *"An additional set of entities that serve as the subjects of works. The group includes concept (an abstract notion or idea), object (a material thing), event (an action or occurrence), and place (a location). There are "subject" relationships between entities in the third group and the work entity in the first group. A work may have as its subject one or more than one concept, object, event, and/or place. Conversely, a concept, object, event, and/or place may be the subject of one or more than one work".*

Group 3 concepts are used in the ontology presented in this paper to describe the contextual entities related to the resources represented in the core ontology: current software applications, business rules, constraints and regulations, and external entities such as contractors, external consultants, vendors, etc.

Relationships depicted in figure 3 indicate that a work may be realized through one or more than one expression. An expression, on the other hand, is the realization of one and only one work (there is a one-to-many relation linking work to expression). An expression may be embodied in one or more than one manifestation; likewise a manifestation may embody one or more than one expression (a many-to-many relation linking expression to manifestation). A manifestation, in turn, may be exemplified by one or more than one item; but an item may exemplify one and only one manifestation (a one-to-many relation linking manifestation to item).

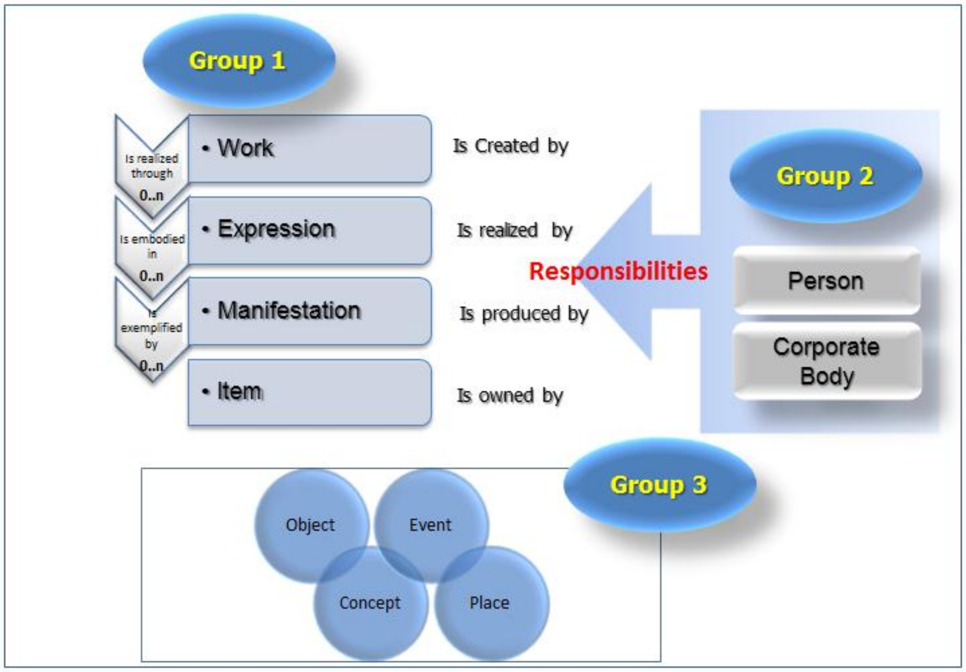

Figure 3: IFLA's FRBR Model

5. Application of Ontologies in the Semantic Web: Web Portals

A Web Portal is a unique place in the Internet or a corporate Intranet to gather and present information from diverse sources in a unified manner; typically collecting and syndicating content (streams) about one specific topic, domain, region, or company, facilitating the work of one topic-oriented community (community portal), or the collaborative effort of a team with a dedicated task (project portal). Content types may include news and up-to-date information streams, e-mail, instant messages and (multimedia) documents. Web portals often provide a consistent look-and-feel for heterogeneous input, with access control and interfaces for multiple applications, for example, information push services (like RSS feeds) or comfortable information access with mobile devices.

A summary of Web Portal typical functionalities includes:

● **Information Supply:** Users have easy means to submit information and make contributions to their communities.

● **Information Management:** Portal administrators can easily integrate new (static or dynamic) information sources, keep the content consistent, change layouts, etc. This can include the (automatic) establishment of links between content items, which are not existing at the level of the individual content items.

● **Information Browsing:** Domain-knowledge structures are the basis for navigation menus, faceted browsing, information visualization, etc.

● **Information Search:** Unified search over heterogeneous content is provided.

● **Personalization:** Individual configurations for layout, information-delivery modalities, content selection, etc.

Among the most common challenges related to knowledge transfer and management in the information technology sector are recording, reusing, locating and sharing information. The same observations apply to documentation management. Additionally, documentation usually comes with the following issues: documents may be referred by a name, vague (e.g. user’s manual) or precise (e.g. French translation of system requirements for the latest version of given software product). Referring a document (or a set of documents) by name requires identifying documents and relationships among documents.

Compared to a conventional Web Portal, a Semantic Portal can be characterized by: (1) a domain ontology used as the central, harmonized topic structure for knowledge organization, navigation, and visualization; (2) semantic search mechanisms; and (3) Semantic Web languages for internal data management – which facilitates declarative approaches for further functionalities like personalization or consistency checking of content.

6. Methodologies and Tool Support

6.1 Ontology Engineering

Ontologies constitute valuable, complex assets that are slowly, but continuously gaining recognition and use throughout a set of disciplines. The objectives pursued with their development and the development itself must be critically assessed by the organization that is pushing for their creation and maintenance. The discipline that investigates the principles, methods and tools for initiating, developing and maintaining ontologies is known as “Ontology Engineering”.

Some of the most popular methodologies for ontology engineering have been derived from several case studies of building and using ontologies in the realm of Knowledge Management (KM). Knowledge management deals with the thorough and systematic management of knowledge within an organization and between several cooperating organizations. Knowledge management is an inherently interdisciplinary subject, a major issue for human resource management, enterprise organization and enterprise culture. In this context, information technology (IT) constitutes a crucial enabler for many aspects of knowledge management and ontologies frequently turn out to be valuable assets for knowledge management in order to target core knowledge management issues such as search, information integration, or mapping of knowledge assets.

Ontologies used in various applications differ, for instance, in terms of scope, size, or expressivity. While it is often possible to reuse existing ontologies that fulfill all the requirements of a certain application, many practical scenarios demand the acquisition of new ontological knowledge or the adaptation of previously given models defined by business rules, standard vocabularies and taxonomies. Additional changes to the ontology might become necessary at runtime as the domain knowledge or user requirements evolve.

Regardless of the ontology engineering methodology to be selected, with guidelines and best practices in ontology design developed from both, practical experiences and theoretical considerations, the process must be supported by efficient software tools, including data storage and editors or ontology development environments.

Methodologies for Ontology Engineering

As stated above, an ontology engineering methodology is a set of procedures, guidelines, and best practices derived from real-world development experiences or theoretical considerations taken from disciplines like philosophy, computer science and information theory. In recent years, the interest in these ontology engineering methodologies has grown significantly, mostly due to the need of ontologies capable of supporting the operation of increasingly complex systems and information domains.

Generic Methodology

As presented in **[8]**, three phases for a generic approach for an ontology engineering methodology can be identified: Requirements Analysis, Conceptualization and Implementation.

Requirements Analysis: Usually, an ontology engineering approach starts with a detailed assessment of the requirements and information needs that come out from the underlying application scenario. The domain expert collects and specifies these requirements in an ontology requirements specification document, which serves as a basis for subsequent modeling activities and validation. For this purpose, the description of the requirements should contain information, for example, about the scope of the ontology, the contextual aspects in which the ontology will operate, its intended use, or the level of expressivity.

Conceptualization: In the conceptualization phase, the ontology's content is represented in terms of a semantic vocabulary and descriptions of the target domain of interest, which involves the choice of ontological entities and the formulation of relationships, rules and constraints. Based on the requirements specification produced in the previous phase, ontology engineers and domain experts try to confront their views about the targeted context, and achieve a common agreement upon the basic structure of the ontology. The result of this phase is an informal or semiformal specification of their shared conceptualization.

Implementation: The formalization of the ontology in terms of a concrete representation language (i.e.: FOAF, OWL, RDFS) is the final step of the core ontology engineering methodology. Choosing an appropriate ontology language most notably depends on the intended use of the ontology, knowledge and background of practitioners involved in the development process, and the required level of expressivity.

Nowadays, the state-of-the-art comprises a variety of different methodologies for specific ontology development scenarios (e.g., distributed, collaborative ontology engineering), and specific application domains, such as bioinformatics or medicine. Despite of the existence of various ontology engineering methodologies, the construction of an ontology remains a labor-intensive, time-consuming, and error-prone endeavor if is carried out entirely manually. A number of tools, languages and ontology design and construction resources are available to help practitioners with these tasks. The next section introduces some of these resources, to be applied to the semantic portal developed in section 7.

6.2 Semantic Web Languages, Vocabularies and Ontologies

In order to design and specify ontologies within a Semantic Web context, designers and practitioners should be familiar with the key components of the semantic web architecture. Modeling resources as RDF, OWL and other languages are briefly described here. These tools are used for the specification of the ontology to be deployed within the semantic portal, to describe all the concepts (classes) and relationships that provide the foundation of the proposed solution.

RDF: Resource Description Framework

Linking data distributed across the organization requires a standard mechanism for specifying the existence and meaning of connections between items described in these data. This mechanism is provided by the Resource Description Framework (RDF).

The RDF model is specified by the W3C **[9]**, and it was originally conceived as a metadata data model. It is aims at being employed as a *lingua franca*, capable of moderating between other data models that are used on the Web. Within the Semantic Web landscape, RDF is used as a general method for conceptual description or modeling of information that is implemented in web resources. RDF provides a flexible way to describe things in the world, such as people, artifacts, documents, or abstract concepts, and how they relate to other things. These statements of relationships between things are, in essence, links connecting things in the world, information is represented as node-and-arc-labeled directed graphs.

In RDF, a description of a resource is represented as a number of *triples*. The three parts of each triple are called its *subject*, *predicate*, and *object*. A triple mirrors the basic structure of a simple sentence, such as this one:

John Smith	owns	Instrument_Configuration_Glossary
Subject	**Predicate**	**Object**

- The subject of a triple is the described resource.

- The object can either be a simple *literal value*, like a string, number, or date; or the link of another resource that is somehow related to the subject.
- The predicate, in the middle, indicates what kind of relation exists between subject and object.

In the example above, the triple specifies that a person in the organization is the owner of a given resource, in this case a glossary of terms for instruments configurations. As stated, the triple could include the name of this document, or a link to the file in a file system or document repository.

Using Vocabularies to describe data

RDF provides a generic, abstract data model for describing resources using *subject*, *predicate*, *object* triples. However, it does not provide any *domain-specific terms* for describing classes of things in the world and how they relate to each other. This function is served by *taxonomies*, *vocabularies* and *ontologies* expressed in *SKOS* (Simple Knowledge Organization System), *RDFS* (the RDF Vocabulary Description Language, also known as RDF Schema) and *OWL* (the Web Ontology Language)

SKOS is a vocabulary for expressing conceptual hierarchies, often referred to as taxonomies, while RDFS and OWL provide vocabularies for describing conceptual models in terms of classes and their properties. For example, someone may define an RDFS vocabulary about Medical Devices that includes a class **Microplate System**, of which all individual pieces of equipment in the product line are members. They may also define a property **hasCapacity**, thereby allowing the administrators to publish RDF descriptions of the capacity of each equipment, that is, the number of tests or samples that a given instrument can process in one run.

Depending on the scope and domain of application, SKOS, RDFS and OWL would provide different capabilities and features for expressivity (level of detail to represent entities and relationships) to practitioners in charge of data modeling activities.

For instance, SKOS is widely used to represent thesauri, taxonomies, subject heading systems, and topical hierarchies (for instance that Microbiology *belong to* the boarder topic of In-vitro diagnostic market). RDFS and OWL, on the other hand, are used when “is-a” relationships need to be represented. These types of “is-a” connection between terms are called subsumption relationships, they are very useful when applied to a reasoning engine to infer information from the data model; for instance, Field Engineers and Application Specialists *are both Customer Service employees, these CS employees are employees, and employees are persons*. Figure 4 shows this sample hierarchy.

More information about inference and OWL in Semantic Web systems is available in **[13]**.

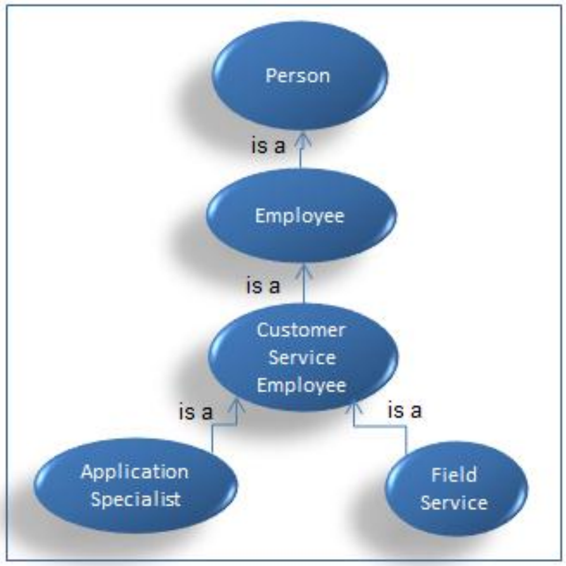

Figure 4: Sample hierarchy for subsumption relationships.

FOAF: Friend Of A Friend Vocabulary.

The FOAF ontology is not part of the standard W3C specification, like RDF, for instance. Instead, it is managed by a schema similar to those found in Open Source or Free Software project standards and maintained by a community of developers **[14].** This is a popular vocabulary on the semantic web, it uses RDF to describe the relationships people have to other people and the "things" around them. FOAF permits software agents to make sense of the thousands of connections people have with each other, their jobs and the items important to their activities; connections that may or may not be enumerated in searches using traditional web search engines.

FOAF depends on W3C standards such as RDF and OWL. More specifically:

• FOAF ontology is written in OWL.

• FOAF documents must be well-formed RDF documents.

The web portal uses FOAF to specify the links and relationships between the people associated to the different information resources supported by the application.

FOAF terms are used to identify characteristics and properties that are attached to persons playing certain roles in a given domain. Person name, email account, topic interest, publications, account, Personal Profile Document, website are some of the attributes that are available in FOAF to document the system user characteristics that could be relevant for the intended purpose of the system under design.

FOAF Properties. *foaf:knows* and *rdfs:seeAlso*

An important FOAF property is foaf:knows. It is used to describe relationships with other people, and it is very useful when it comes to build a network of people within a domain, or, even better, between different domains, using FOAF documents. It is important to note that the foaf:knows property is not symmetric; that is, that John knows Anna does not imply that Anna knows John.

As stated above, maybe the most important use of foaf:knows property is to connect FOAF records together. Often by mentioning other people (foaf:knows), and by providing a rdfs:seeAlso property at the same time, different documents can be linked together. The property rdfs:seeAlso is defined in RDF schema (RDFS, see section above), and it indicates the fact that there is some additional information about the resource this property is describing.

These two properties play a key role in the definition of the ontology that supports the semantic portal, because they allow users to dynamically expand the existing database with additional data about the information resources stored in the knowledge repository.

This capacity of the information system is sometimes called “information enrichment” and constitutes a natural expected characteristic of a good quality knowledge management application.

6.3 Wikis and Semantic Annotations: Semantic MediaWiki

Web portals are entry points for information presentation and exchange over the Internet or an organization intranet about a certain topic or field, usually powered by a community. Leveraging semantic technologies for portals and exploiting semantic content has been proven useful, especially the aspect of providing semantic data, which has gotten a lot of attention lately due to the Linked Open Data initiative **[15]**.

Wikis

One of the most successful techniques to power communities of interest on the web are wikis. In general:

- A wiki is a Web application that manages a collection of web pages, where anyone can create new pages and edit existing pages by using a simplified markup language. The most popular and common example of a wiki is Wikipedia.org.
- A wiki engine is a software system that powers the wiki site and makes it work. For instance, Wikipedia is developed using the engine MediaWiki **[16]**.

Wikis allow users to collaboratively create and maintain mainly textual, unstructured content. The main idea behind a wiki is to encourage people to contribute by making it as easy as possible to participate. The content is developed in a community-driven way. It is the community that controls content development and maintenance processes. A wiki provides the users with the means for quick and easy adding and changing of content, in the sense that they just need to know the simple wiki markup and have a web browser.

Why wiki site for a knowledge portal?

One of the main goals when designing a web portal was to have low barriers for the organization's members to contribute, extend and maintain the content. Hence, a wiki is considered as an appropriate choice.

Semantic Wikis

Despite the great success and the widespread use of wiki systems, there are some limitations that should be mentioned. These limitations can be grouped in the areas of knowledge discovery, reuse and consistency.

Today, using a given wiki site primarily means reading articles in the site. For example, in Wikipedia, there is no way to request a list of cities with at least 100 years of history, population over a given number of inhabitants, and also obtain data about the healthcare infrastructure available in the region, in spite of the fact that the information is contained in a set of pages in the wiki. This information has to be obtained through human reading and, even though the wiki engine does provide a full text search, it suffers from the ambiguity of natural language. When a search is executed, it normally produces a result list where most of the pages shown are irrelevant; users still have to go through the returned pages to look for the information they need, with a negative impact in users' productivity.

The information in these common wiki sites is not presented and stored in a machine-readable way, but rather is only accessible to human eyes.

A solution of these issues is to follow the idea that is promoted by the Semantic Web technology: add semantics into the wiki pages so that the information on these pages will be structured enough for machines to process. Once this is done, knowledge discovery will be easier for the wiki users. In order to add formal semantics into wiki pages, the wiki engine itself has to be enhanced so that it knows how to take advantage of the structured information contained in the page, a new breed of wiki site - semantic wiki – is required.

A semantic wiki can be defined as a wiki site powered by a semantic wiki engine; it enables users to add semantic markup to wiki pages, and the added structured information can then be used for better searching, browsing, and exchanging of information **[17]**.

Semantic wikis allow users to annotate the content in order to add structure. This structure allows users to consider the wiki as a semi-structured database and to query its structured content to exploit the wiki's data and to create various views on that data. This way wikis become even more powerful content management systems.

Semantic MediaWiki

Semantic MediaWiki **[18]** is a semantic wiki engine built on top of MediaWiki engine. It was originally developed by AIFB, a research institute at the University of Karlsruhe, Germany **[19]**. Over the years, it has been under constant improvement by developers around the world. Since it is a free extension of MediaWiki, it is widely available, and very popular. In contrast to regular wiki systems, Semantic MediaWiki allows people to semantically annotate the content. This free and independent annotation paradigm has the advantage of being flexible, and expandable. Moreover, it does not require the knowledge of a predefined schema. The underlying notion is that more annotations are in general better than less annotations even if they are not well organized and do not follow a predefined vocabulary or ontology.

Semantic Annotations: Forms and Templates.

In Semantic MediaWiki, templates and forms are used to restrict the wiki user to a predefined set of annotations. The combination of forms and templates allows wiki authors to have a toolset with predefined annotations aimed at improving the productivity and quality of the information contained in the repositories.

- Templates define the logic and the appearance of a part of a page. Inserting annotations in the template entails the annotation of all pages using the template with the same annotations. Consequently, changing the annotation inside a template cascades this change to all pages and thereby allows a flexible modification of the structured data.
- Forms provide a graphical user interface for using templates correctly and do not even require the usage of wiki markup.

Building a web portal with Semantic Wiki

The documentation system proposed in this paper in intended to provide an information management platform capable of delivering the following services: find, identify, select, obtain and navigate. The scoped information resources are maintained in the targeted organization's content management systems.

In the next section an example application is developed to demonstrate the tools and resources presented so far in the study. The ontology to be deployed is based on the FRBR schema to model information artifacts, users and stakeholders using modeling tools and vocabularies such as RDF, FOAF, Semantic MediaWiki, semantic forms, templates and queries.

7.Example Application: A Semantic Portal for knowledge Organization.

7.1 Case Study. Introduction and Goals. The Target Organization

The semantic portal is implemented within the context of a healthcare organization which is mainly focused on medical devices and reagent products for the in-vitro diagnostic (IVD) market. **[19]**

The term IVD covers a diverse range of products from individual reagents to testing systems that consist of reagents, instrumentation, software, as well as accessories such as dedicated software, control and calibration materials. IVDs differ from most general medical devices in that many do not come into direct contact with patients. Therefore they cannot cause direct harm to the patient if they fail to perform as intended.

An in-vitro diagnostic medical device or IVD is considered to be a reagent, calibrator, control material, kit, specimen receptacles, software, instrument, apparatus, equipment or system, whether used alone or in combination with another diagnostic product for in vitro use; and is intended by the manufacturer to be used in vitro for the examination of specimens derived from the human body. Figure 5 shows an example of a typical instrument system for microbiology, with its main components and accessory modules.

Figure 5. An in-vitro diagnostic system with components.

(http://www.biomerieux-usa.com/upload/platelet_quality_control.jpg)

The Target Information System

The organization maintains a sophisticated IT infrastructure for its servers and workstations, databases, office tools (spreadsheets, word processors), business applications, intranet and messaging systems. The main documentation repository is implemented over a commercial Enterprise Content Management System (ECMS). The members of the organization: employees, contractors, consultants; have a login account and access to these information system resources available within the enterprise IT portfolio.

In spite of the fact that these information resources are available and stored in the organization systems, users report regularly difficulties to find the right information in a timely manner. The documents stored in the databases most of the time do not contain a valid set of metadata attributes that help users find them and there are also gaps and holes in the identification of the people involved in the administration of certain documentation, owners, creators and experts.

As the target organization operates in a global marketing, one of the key goals and directives from the top management is to secure that all the workflows and processes are standardized and implemented consistently in all the countries where the company deploys medical devices in labs, hospitals, clinics and other healthcare institutions.

Information needs and issues.

Even though the enterprise maintains and publishes internally its organizational charts, departments, managers, leaders and team members, there is a network of informal relationships and knowledge administration that is getting lost in the daily operation of the enterprise. The following issues are detected:

- The service desks and IS teams keep receiving and answering the same type of requests over and over again, affecting the performance of the department.
- When an experienced user leaves the organization the replacement has to begin from scratch and learn the job with outdated training material and documentation about workflows and business processes. They do not know how/whom to contact for certain requests.
- Some users have a background and expertise in certain areas within the organization that are not properly identified and used in the operation. For instance, they may have experience in certain software application from a previous job or fluency in a foreign language that would be useful to support and help coworkers looking for that kind of resources.

The main purpose of the semantic portal deployed in this study is to provide a tool for these users to collect and use these "informal" links within the organization to required information resources or people available to provide any kind of support in the operation.

Due to some constraints in time and resources to cover the whole information space of the target organization, this study is mainly focused on covering those aspects related to the administration of medical devices and the operation of the technical service department. These sample models should help fulfill the goal of showing the main characteristics of ontology design and semantic portal development methods and processes.

The case study is organized as follows:

7.2 Description of the ontology engineering approach used to build the target model, using FRBR as a baseline.

7.3 Implementation of the ontology using semantic MediaWiki: Forms and Templates

7.4 Where Semantics Help: semantic queries, searches and inferences.

7.2 FRBR for Enterprise Information Resources.

The FRBR model is used to define the ontology of available information resources. See figure 6 to review the model entities and relationships explained in section 4, above. The next sections describe the mapping of standard FRBR concepts and entities with the resources identified in the target environment.

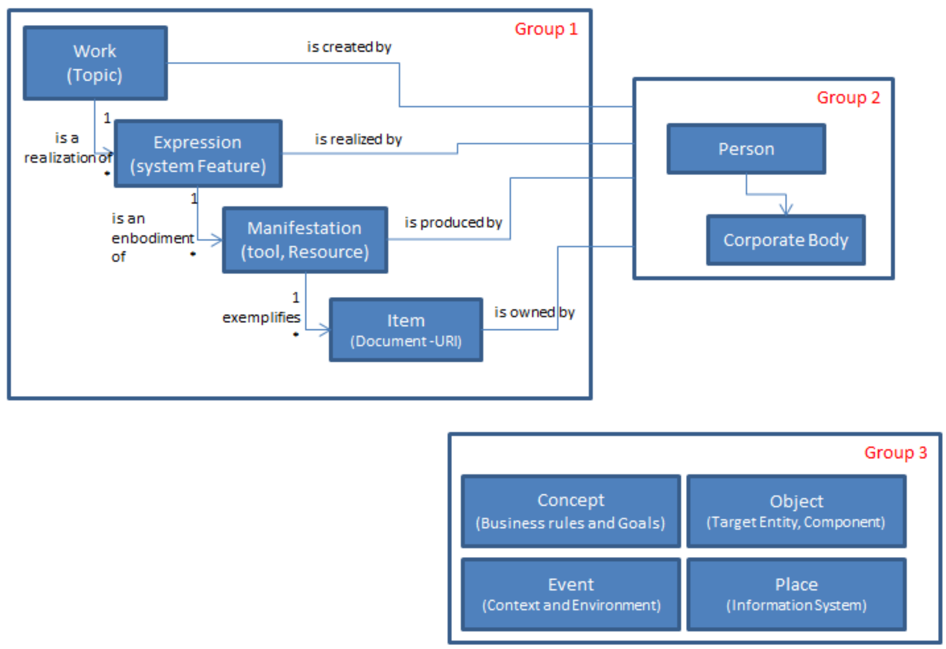

Figure 6. Groups 1,2 and 3 in the FRBR conceptual model.

FRBR Group 1 Entities

The following resources are elicited from the target environment, and should be collected and recorded in the ontology as group 1 entities (work, expression, manifestation, item).

- ✓ Training Material documentation
- ✓ Software manuals and user references

- ✓ Organization Charts
- ✓ Technical specifications for IT components.
- ✓ Working Instructions
- ✓ Position Papers
- ✓ Spreadsheets
- ✓ “How-to” guidelines and administrative workflows
- ✓ Glossary of instruments components and configurations
- ✓ Address-books with contact information for employees, consultants, vendors and contractors.
- ✓ Meeting Minutes
- ✓ IS Project Plans and Initiatives
- ✓ Internal websites

These documents are stored in the company’s content management system (CMS). Figure 7 shows a sample folder structure for the organization’s content management system. These folders and documents are available for all users that have to the organization’s network.

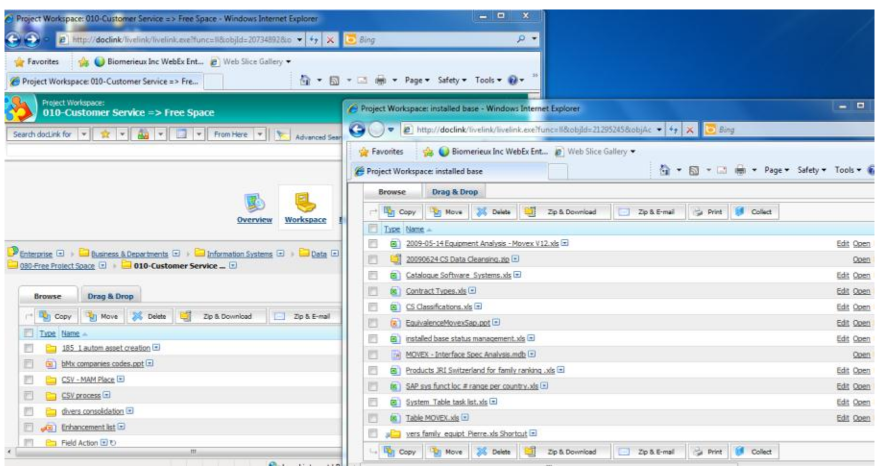

Figure 7. Sample screenshots of folder structure for a ECM system.

Each element in the ECM system can be referenced using an internal URI (Universal Resource Identifier) that is found in the address URL of the internet browser used to navigate the content of the file system. As shown in figure 8, these URIs can be found in the browser’s URL box used to navigate the ECMS within the organization’s intranet. This URI is used to access the document from the semantic portal, and it will be specified as an ITEM in the FRBR model.

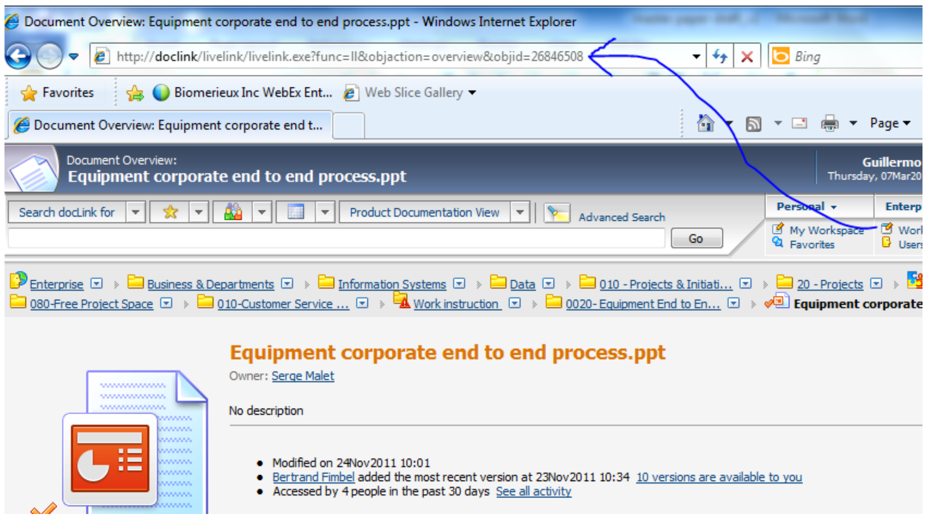

Figure 8. ECM elements are identified with an internal URI.

Building the FRBR Model

The workflow "Equipment End-to-End Process" is taken as an example for the mapping of FRBR Group 1 entities. The end-to-end process describes the installation workflow, including all the tasks, required paperwork and validations that a given piece of equipment follows from the company manufacturing facilities warehouse until the equipment is installed and made operational by the technical team in a customer site location.

As stated in section 7.1, the company promotes the standardization of these procedures at the global level, so the end-to-end process for medical devices is defined as a corporate procedure. However, due to different rules and constraints that may apply to different countries, it is necessary to define a certain level of "localization" of the procedure, still using the corporate process as a baseline, but adapting the required steps to local regulations or requirements. The most typical cases where these locations are required are related, for instance, to tax and fiscal regulations, regulatory agencies, customers and regional contractors, language translations, among others. Therefore, the schema for this process is defined as follows:

Work: used to identify corporate procedures.

Expression: used to identify localizations, versions in different languages of the corporate procedure.

Manifestation: it is the document type of class, such as a MS Word Document, a MS PROJECT document, a MS POWERPOINT presentation, SAP ERP system scripting file, etc.

Item: it is the link, the URI to the document in the company's intranet or ECM system.

Figure 9 shows the mappings between FRBR Group 1 entities and the enterprise information resources.

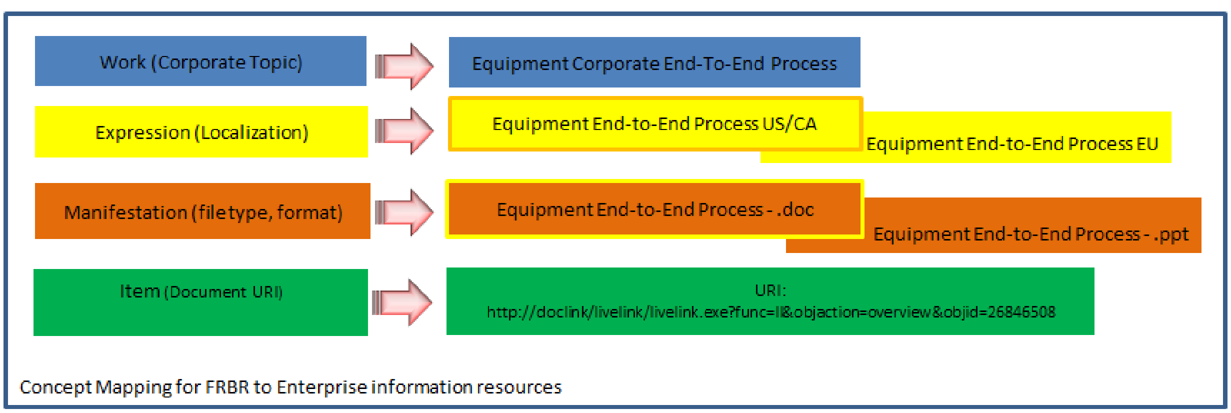

Figure 9: Mapping FRBR entities to enterprise resources.

Relationships between Work Entities

In order to manage factors such as complexity or clarity; a work, defined as a corporate-based document, can be split and handled in different work entities. FRBR provides a number of relationships to specify the links between related works. For instance, in this case of end-to-end process for medical devices management, a different work can be defined based on the status of the transaction for an instrument installation: placement, leasing, rental, sale or any other type of agreement or contract negotiated with customers. Figure 10 shows the hierarchy defined to manage the different "flavors" defined for the work "equipment corporate end-to-end process".

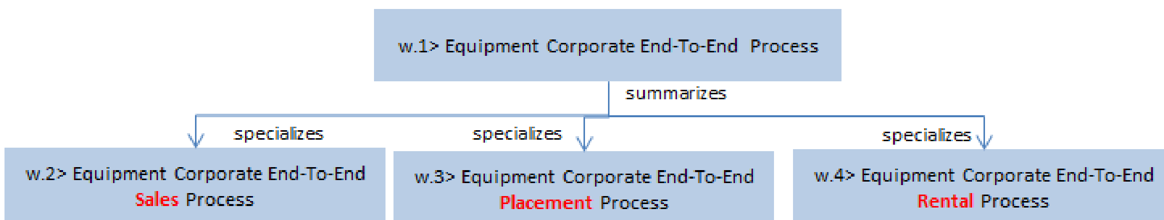

Figure 10: Relationships among Work entities.

The same types of relationships can be defined for expressions (a localization for a country is related to the location of another country for certain class of instrument), and manifestations (different file formats for a .doc document, such as different versions of MS Word, for Windows, Mac, etc.)

FRBR Group 2 Entities

Group 2 entities in FRBR are used to specify people responsible for the creation, content administration and support of the information resources identified in the domain. The following roles for employees and stakeholders within the organization are identified:

- ✓ End Users: employees accessing the information resources for the organization regular operation.
- ✓ Managers: Area and department leaders.
- ✓ Key Users: users with special knowledge and access rights, normally in charge of provide a first level of support to end-users.
- ✓ System Administrators: special users with rights to configuration features of existing systems and applications.
- ✓ IS Support: employees providing end-user support, helpdesk, assistance with IT resources and applications.
- ✓ Sponsors: Managers for projects and initiatives with the IS domain.
- ✓ SME (Subject Matter Experts), or domain expert is a person who is an expert in a particular area or topic within the business.

The key roles for the ontology presented in this work are the end-users and the SMEs, as the main goal of the semantic portal is to enhance the enterprise communication channels and promote the interaction between those users who have questions (information needs) and the ones who can provide answers (SMEs).

These Group 2 entities are represented in the portal using the FOAF vocabulary (Friend of a Friend, see section 6.3). FOAF links persons with entities defined in FRBR Group1 and, most importantly, specifies relationships between people playing different roles. The goal is to allow users to search and query in the portal the list of people capable of providing certain level of support for selected business processes and resources. Figure 11 shows FOAF terms mapping between FRBR entities from group 1 and group 2.

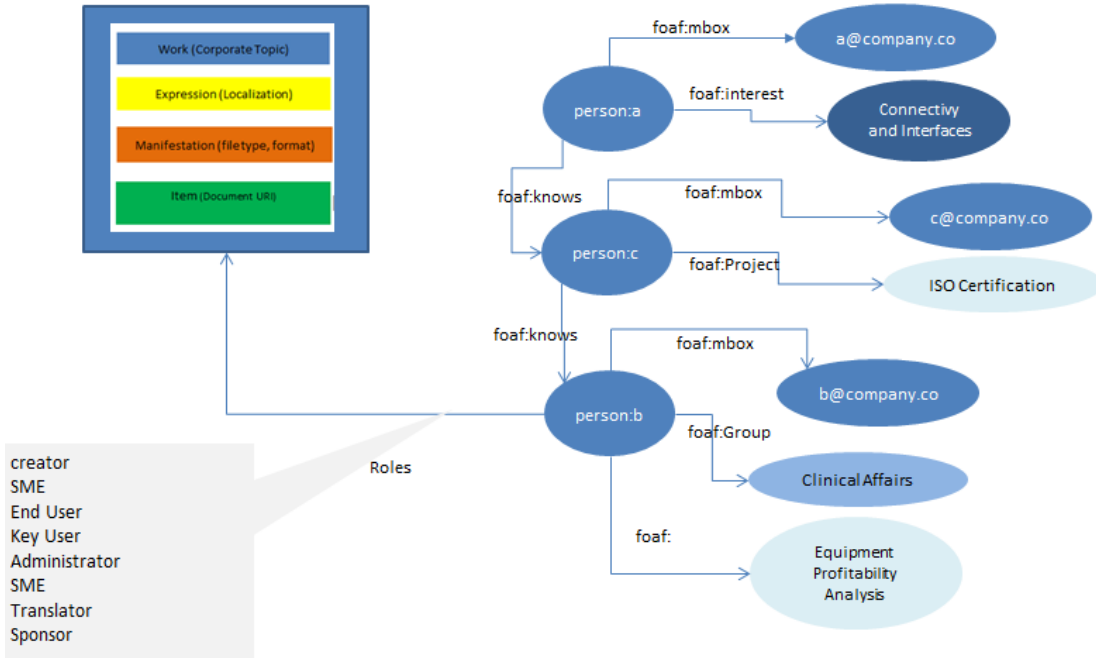

Figure 11. Roles and Responsibilities for Group 2 entities.

FRBR Group 3 Entities

The entities in the third group represent an additional set of entities that serve as the subjects of works. The group includes concept (an abstract notion or idea), object (a material thing), event (an action or occurrence), and place (a location).

Figure 12. FRBR Group 3 entities with enterprise mappings.

This set of entities are used to provide contextual information to entities described in group 1 and group 2. Figure 12 shows the entities within FRBR group 3 and the intended mapping terms within the enterprise information domain.

Concepts:

IFLA Definition **[7]**:

> *"Encompass a comprehensive range of abstractions that may be the subject of a work: field of knowledge, processes, techniques, practices, etc."*

Example: For the equipment corporate end-to-end process case, a number of concepts can be linked to the work that defines the corporate process. The concepts can also represent formally specified user requirements (URS: user requirements specifications) and business rules.

Work: Equipment Corporate End-to-End Process

Concept 1: Commercial Operations and Logistics Departments.

Concept 2: USR 5.001.23 Installed base management requirements.

Objects:

IFLA Definition:

> *"Encompass a comprehensive range of material things that may be the subject of a work"*.

Example: For the equipment corporate end-to-end process case, the most important objects are: Medical Device, Instrument Configuration, Component, Purchase Order, Delivery Note, and other forms and paperwork included in the workflow for the process.

Places:

IFLA Definition:

"Encompass a comprehensive range of locations, for the purposes of this study places are treated as entities only to the extent that they are the subjects of a work".

Example: For the equipment corporate end-to-end process case, the subjects that can be linked to the process are the business applications used to record the transactional records of the whole workflow. For instance:

Place 1: (subject): ERP system SAP (need reference)

Place 2: Company headquarters in Marcy, France.

Events:

IFLA Definition:

> *"Encompass a range of actions and occurrences that could initiative the execution or the application of a work. A event is also defined as some change in the system's environment, normally followed by a response, a set of actions performed by the system whenever a certain event occurs"*.

Example: For the equipment corporate end-to-end process case, the events that could trigger the execution of the procedure are:

Event 1: A Customer submits a formal request for a new equipment and the request is approved.

Event 2: The regulatory office (for instance, the FDA in USA) cleared a new device for marketing.

Figure 13 shows the complete set of entities for FRBR and the mapping to information resources design for the case study, all entities related to the process: Equipment Corporate End-to-End Process.

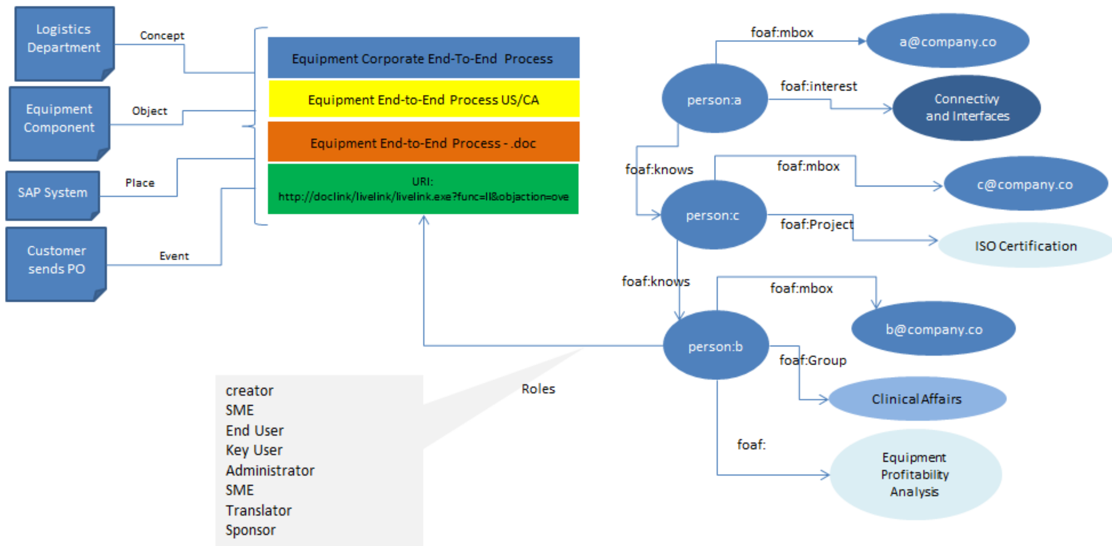

Figure 13. The Complete FRBR Mapping to enterprise entities.

7.3 Designing the Portal with Semantic MediaWiki.

The most common application of Semantic MediaWiki and wiki systems in general is collaborative knowledge management, e.g. for communities such as semanticweb.org. In this section the portal is built using Semantic MediaWiki **[18]** and in particular its features exploiting semantic technologies. Figure 14 shows a draft of the main page of the targeted semantic portal under development.

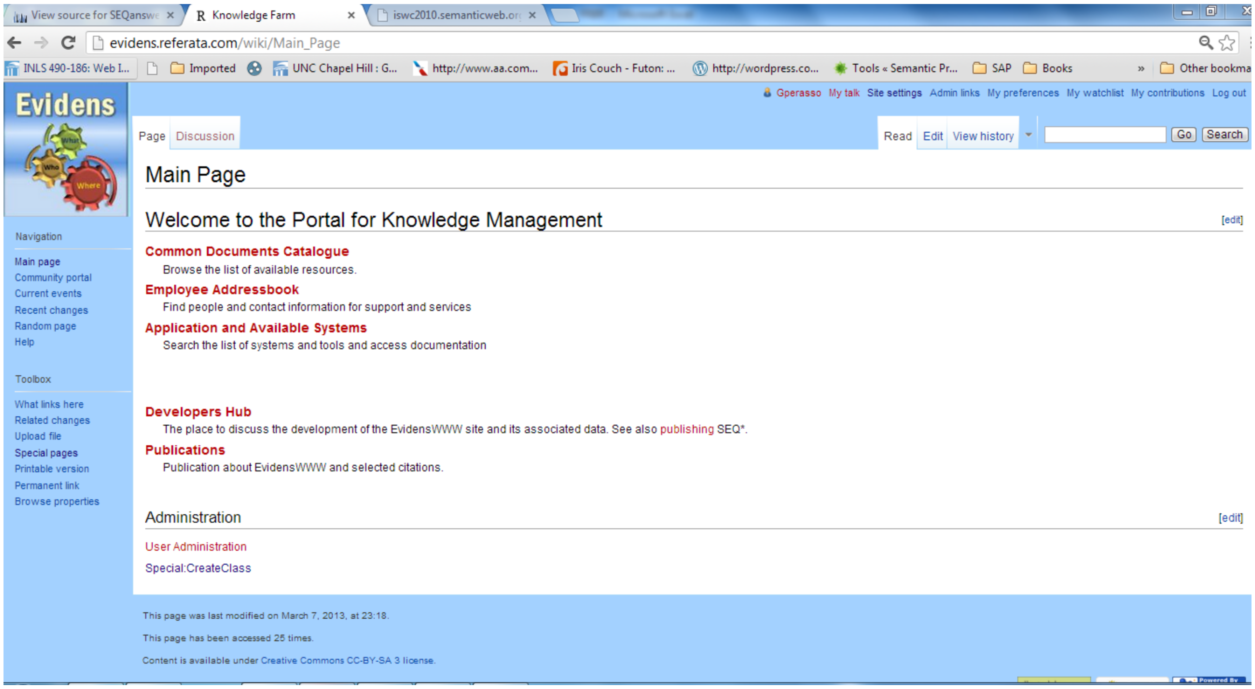

Figure 14. Semantic MediaWiki Portal sample screen.

Adding Semantics to a Wiki Site

Any given wiki document has two elements, hyperlinks and texts. Semantic annotation in wiki means to annotate any link or text on the page so as to describe the meaning of the hyperlink or the text. The result is that the added annotation turns links and text into explicit property; i.e. value pairs with the page being the subject. Besides the properties on links and text, the category system can be viewed as an existing semantic component. The reason being that it does add structural information to the wiki documents, and it is also used in semantic wiki's RDF/OWL exported files. Therefore, to have a complete picture about semantic components in a semantic wiki engine, we will start with the category system.

<u>Namespace and Category System</u>

Category system is an existing semantic component in MediaWiki, to understand the category system, we need to understand the concept of namespace first. By default, a wiki powered by MediaWiki will have 18 namespaces, and they are as follows:

- Main namespace, which is also the default namespace if a given page title does not have any prefix.
- 15 additional namespaces, each having a specific prefix.
- 2 pseudo-namespaces.

The main reason of having the namespace system is to better structure the content for a given wiki project. The following is to name a few benefits:

The namespace system can be used to separate the main content from the rest. This separation will make the wiki management become easier. For example, the main namespace (and a few others) will form a core set that is open for public, allowing users to view and edit them frequently. Meanwhile, this core set is actively policed by the wiki community; any inappropriate content will be quickly removed. The policing rules for other namespaces are generally more relaxed. Another example would be the search function. For instance, for most wiki engines, searching can be limited to any subset of namespaces, which will be helpful to users.

The namespace system can be used to group the content about the same subject and to form a set of relatively unrelated content items. For example, namespaces are stored as folders on the host file system, which gathers the content files of the same subject inside one single directory. This makes the administrator’s work much easier. For a specific wiki project, the namespace structure will most likely be customized to fit the need for that project. For example, not only each project will have its own naming method for these namespaces, but also the number of namespaces can be different.

The Semantic Portal system administrator will define the namespaces to be created for the wiki site. Different namespaces for global documents and localizations by country or region will be required.

Ontology integration into the portal. Defining classes, forms and templates

A set of forms and templates are created in Semantic MediaWiki to define the structure of the information to be collected and annotated for the users in the portal. The following main templates are identified:

- ✓ Resources
- ✓ Employees
- ✓ Target Applications

In Semantic MediaWiki, a template can be created using the "Special.CreateClass" feature. Figure 15 shows the template to be populated by the site administrator to create a new class in the portal.

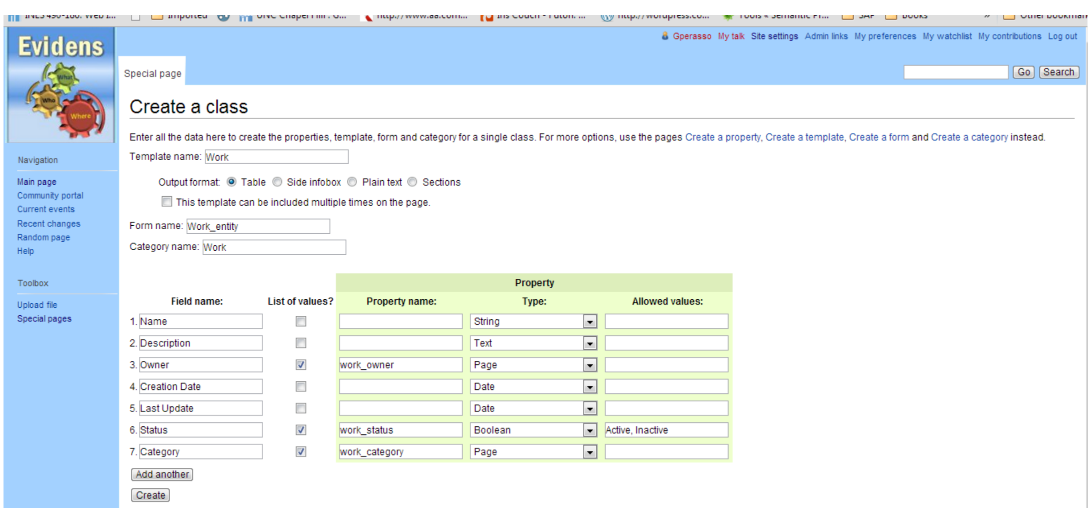

Figure 15. Create a Class for Semantic Annotations.

User Roles

Often times, when a wiki system is proposed for a portal implementation, the first concern from the sponsors and managers is the fear that everybody can edit it, even anonymously, compromising the quality of the data contained in the repositories. In order to handle these concerns, four different user groups are created. These user group definitions are taken from the standard groups used in the most widespread wiki sites.

The anonymous web surfers can only read regular pages. The authenticated users may also read pages in other namespaces and in addition are allowed to edit pages, except for pages in the template and form namespace. The latter can only be manipulated by admins. The fourth group are bureaucrats, which have the same rights as admins, but in addition they can appoint and withdraw the admin right.

7.4 Where Semantics Help - Features of Semantic MediaWiki

Key capabilities of semantic data are summarized below, showing the main features for data extractions in semantic searches and queries, and what are the most common, standard resources supported by Semantic MediaWiki (SMW) to make use of these semantic functionality deployed upon the standard wiki pages.

Inline Queries

The biggest advantage of SMW, beside its flexible annotation paradigm, is the possibility to reuse data across the platform by querying it from other pages. These inline queries allow users to request sets of data or just single property values and display them on a page in various result formats, such as tables, list, charts, maps, etc. This reuse of data avoids data redundancy, e.g. the information about a person, like name, email, or phone number, is entered once on the page about this person and then later this information is queried and displayed on pages about projects, documentation, etc., where this person is involved in. If the data changes on the source page, the data on the requesting page changes accordingly when the inline query is executed again. Inline queries create dynamic pages.

For instance, to retrieve from the portal all the localizations (FRBR:expressions) for the corporate procedure (FRBR:work) Equipment Corporate End-to-End process, a query should be stored in the portal with the following format (simplified version for demo purposes).

```
{{#ask: [[Category: Working_Instruction]]
   [[Work::Expression || objid=26846508 ]]
        | ?Name
        | ?Descripcion
        | ?Country
        | sort= CreatedDate
        | format= template
        | template=WorkingInstructionsList
}}
```

These commands will return all the expressions, that is, localizations, recorded in the system for the selected work (the object number of the stored Equipment end-to-end process) in a dynamic page created by the Inline Query.

Exploiting the Semantics for Search

One certain advantage of having the content of the portal available in a structured form is the ability to exploit it for search. The approach allows to use keywords as the means to express an information need, because most users are used to this common search paradigm. These keywords are then transformed into interpretations using the structured data of the wiki as the search space. The interpretations are shown to the user, who can select the interpretation fitting best to his information need and further refine it in the next step. In contrast to the inline queries, which use a simple, but formal query syntax and are therefore inadequate for ad-hoc search, this approach is suitable for end users and exploits the semantic annotations.

Inferencing

In semantic wiki, a semantic search is to find the requested pages based on the semantic annotations added by the users. It requires that semantic information be entered onto the page document ahead of the time when the search is conducted. On the other hand, semantic inferencing refers to the fact that users can retrieve information that was not added explicitly by the users but derived by the semantic wiki engine.

In the case of the Equipment Corporate End-to-End process, the ontology structure allows the portal to infer, for instance, that a certain document (a manifestation), belongs to a localization of the corporate procedure for Canada, and the portal will be able to list all the employees located in the company's offices in Canada who are related to the creator of the localized file, who work on the same area, or department and contact them for validation or support purposes.

7.5 Case Study Conclusions

This paper is intended to show how to apply the wiki paradigm of collaborative editing to a web portal using semantic technologies, and how free, unconstrained annotations can be combined with predefined annotations in order to allow flexible and expendable structured data. How the semantic data is used and taken advantage of by Semantic MediaWiki's features is illustrated using examples based on workflows and documentation for instrument management in an organization operating in the medical diagnostic field. Taking everything in consideration, the study shows how Semantic MediaWiki can be used as a successful portal platform providing the advantages of semantic technologies.

A further study would work on assessing feedback collected from the portal users and measuring the performance of the application, analyzing the volume of data traffic, workload and frequency of use and configuration updates and required support.

8. Future Developments

As future developments for the information system discussed in this paper, the efforts should be focused on the alignment of the proposed solution to the organization’s strategy and goals. Economic aspects of the methodology and resources put in place to build the platform, performance, usability and data quality metrics should be implemented to provide the management and project sponsors with concrete indicators to measure of the overall performance of the solution and make decisions about the organization’s mid and long term strategy for their knowledge management systems.

Exploring economic aspects of ontology engineering

A core requirement for the usage of ontologies within enterprises is the availability of proved and tested techniques which guarantee an efficient engineering of high-quality ontologies. Besides feasible technological support this includes in equal measure integrating ontology engineering within the more general framework of enterprise information architectures, and taking into account the economics of ontology engineering projects, in particular issues of cost effectiveness and profitability.

Sponsors and project managers in charge of initiatives involving ontology engineering activities should work on building upon a solid case for the importance of cost-related measures as decision support in planning and controlling. Approaches should be developed for reliably assessing the costs of building ontologies, and the usage of cost-related information to quantifiably support decisions arising during the life cycle of an ontology and to optimize the operation of associated processes.

Implementing indicators for performance, usability and data quality

<u>Performance for Semantic MediaWiki:</u>

Although the usefulness of the features provided by Semantic MediaWiki gets the interest of many potential users, often skepticism about SMW's resource requirements, its stability and scalability are expressed. A successful implementation of the portal will be reflected in the "data traffic" generated on the server and backend systems that support the portal. An increase in the number of pages and forms contained in the semantic wiki environment, as well as the number of users accessing the portal regularly to create annotations and run queries will potentially undermine the performance of the solution and affect negatively the level of satisfaction of users. The support team should take in consideration these potential issues and put together a contingency plan to cover the IT infrastructure needs that would be generated by the growth in the number of users, content and the frequency, quantity and complexity of the queries and searches executed by the users.

Usability indicators may also be considered to provide the project management team with some kind of concrete evidence regarding usage statistics of the portal. These are some examples of indicators that can be defined and measured for assessment analysis of the portal:

- Number and frequency of users accessing the portal.
- Most “popular” pages by number of visits.
- Most “popular” users by activity on the portal and number of references to information objects in the database.
- Number of pages created in a period of time.
- Growth ratio of the portal based on pages added, annotations recorded by the users, number of queries.

The Wiki tool provides a number of useful resources and functions to simplify the process of collecting and recording these data. Statistics, log files and history records are available in the administrator links of the wiki site.

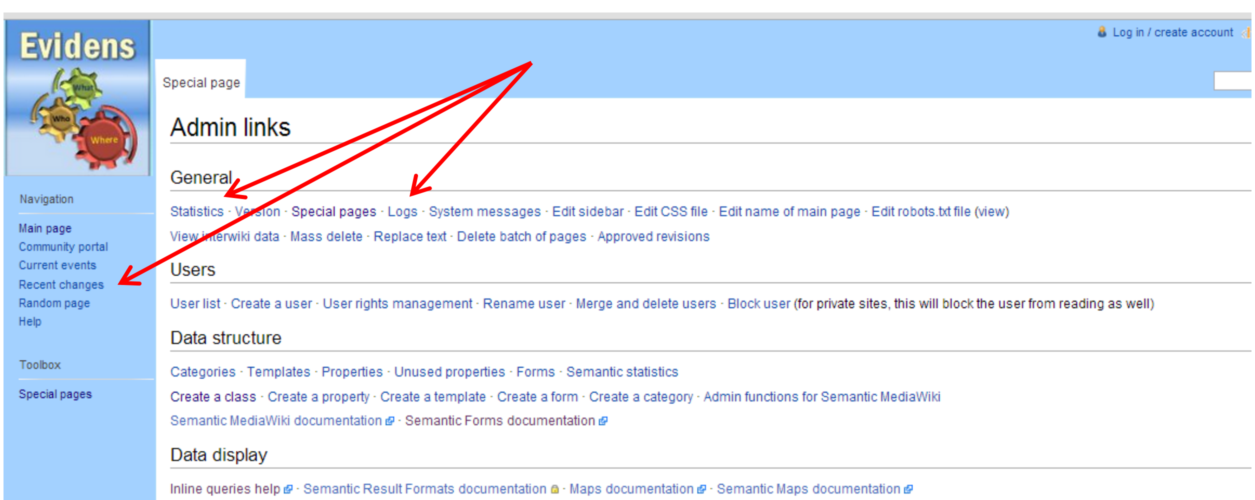

Figure 16. Log records and usage indicators in MediaWiki

9. Conclusions

The paper shows how to apply the wiki paradigm of collaborative editing to a web portal using semantic technologies within an enterprise content management system. The underlying ontology is based on the IFLA's Functional Requirements for Bibliographic Records model. The FRBR proposition includes a description of the conceptual model (the entities, relationships, and attributes) and a four-level classification to define a document identification scheme that allows the semantic wiki, through usable semantic forms and templates, to act as a library catalogue, where users can find, identify, select, obtain, and navigate resources.

A further study would work on assessing economic aspects of the proposed platform, and collecting feedback from the portal users to measure the performance of the application, analyze the volume of data traffic, workload and frequency of use and configuration updates and required support.

10. Bibliography

[1] Studer, R., Benjamins, V., and Fensel, D. (1998). Knowledge Engineering: Principles and Methods. Data and Knowledge Engineering, 25(1-2):161 – 197
[2] World Wild Web Consortium – W3C - http://www.w3.org/.
W3C Semantic Web Activity: http://www.w3.org/2001/sw/
[3] Wikipedia: Natural Language Processing NLP:
http://en.wikipedia.org/wiki/Natural_language_processing
[4] Wikipedia: Automated Reasoning: http://en.wikipedia.org/wiki/Automated_reasoning
[5] IFLA. International Federation of Library Associations and Institutions. http://www.ifla.org/
[6] Wikipedia: Entity-Relationship Model: http://en.wikipedia.org/wiki/Entity-relationship_model
[7] IFLA FRBR: FUNCTIONAL REQUIREMENTS FOR BIBLIOGRAPHIC RECORDS,
http://www.ifla.org/VII/s13/frbr/ (2009)
[8] Grimm, S., Abecker, A., Volker, J., Studer, R. (2011). Ontologies and the Semantic Web. Handbook of Semantic Web Technologies. John Dominguez, Dieter Fensel & James A. Hendler (eds.), 2011, pp 507-579
[9] RDF. Resource Description Framework. http://www.w3.org/RDF/
[10] SKOS. Simple Knowledge Organization System. http://www.w3.org/2004/02/skos/
[11] RDFS: Resource Description Framework Schema. http://www.w3.org/TR/rdf-schema/
[12] OWL: Web Ontology Language. http://www.w3.org/2004/OWL/
[13] Horrocks, I., Patel-Schneider, P., (2011). KR and Reasoning on the Semantic Web: OWL. Handbook of Semantic Web Technologies. John Dominguez, Dieter Fensel & James A. Hendler (eds.), 2011, pp 365-398.
[14] FOAF. "Friend of a Friend" Project. http://www.foaf-project.org/
[15] Linked Open Data Initiative. http://linkeddata.org/
[16] MediaWiki. A free software open source wiki package, originally for use on Wikipedia.
http://www.mediawiki.org/wiki/MediaWiki
[17] L. Yu, A Developer's Guide to the Semantic Web. Chapter 9. Semantic Wiki. Springer-Verlag Berlin Heiderberg 2011.
[18] Semantic MediaWiki http://semantic-mediawiki.org/
[19] Institute of Applied Informatics and Formal Description Methods (AIFB).
http://www.aifb.kit.edu/web/Semantic_Web/en
[20] Wikipedia: IVD – In-Vitro Diagnostics Market. http://en.wikipedia.org/wiki/In_vitro_diagnostics.
[21] Guarino, N.: Formal ontology and information systems, preface. In: Guarino, N. (ed.) Proceedings of the First International Conference on Formal Ontologies in Information Systems (FOIS 1998), Trento, pp. 3–15. IOS Press, Amsterdam (1998)
[22] Studer, R., Grimm, S., Abecker, A. (eds.): Semantic Web Services: Concepts, Technologies, and Applications. Springer, Berlin (2007).
[23] D. Herzig and B. Ell, "Semantic MediaWiki in Operation: Experiences with Building a Semantic Portal," in ISWC2010, ser. LNCS. Springer, 2010, vol. 6497.
[24] Ribaud, Vincent, and Philippe Saliou. "Using a semantic wiki for documentation management in very small projects." *Metadata and Semantic Research* (2010): 119-130.
[25] Krotzsch, M., Vrandecic, D., Volkel, M.: Semantic mediawiki. In: Proceedings of the 5th International Semantic Web Conference (ISWC2006). Volume 4273 of LNCS., Springer (2006) 935-942