From Web 1.0 to Social Semantic Web: Lessons Learnt from a Migration to a Medical Semantic Wiki

Thomas Meilender[1,2], Jean Lieber[2], Fabien Palomares[1], and Nicolas Jay[2]

[1] A2ZI, 3 route de Boncourt, 55200 Commercy, France
{thomas.meilender,fabien.palomares}@a2zi.fr
[2] Université de Lorraine, LORIA, UMR 7503, Vandœuvre-lès-Nancy, F-54506, France
CNRS, LORIA, UMR 7503, Vandœuvre-lès-Nancy, F-54506, France
Inria, Villers-lès-Nancy, F-54600, France
{thomas.meilender,jean.lieber,nicolas.jay}@loria.fr

Abstract. Oncolor is an association whose mission is to publish and share medical guidelines in oncology. As many scientific information websites built in the early times of the Internet, its website deals with unstructured data that cannot be automatically querried and is getting more and more difficult to maintain over time. The online contents access and the editing process can be improved by using web 2.0 and semantic web technologies, which allow to build collaboratively structured information bases in semantic portals. The work described in this paper aims at reporting a migration from a static HTML website to a semantic wiki in the medical domain. This approach has raised various issues that had to be addressed, such as the introduction of structured data in the unstructured imported guidelines or the linkage of content to external medical resources. An evaluation of the result by final users is also provided, and proposed solutions are discussed.

Keywords: semantic wikis, decision knowledge, medical information systems.

1 Introduction

During the two last decades, the Internet has totally changed the way information is published and shared in most of scientific areas, including medicine. First websites in web 1.0 were made of static pages and hyperlinks allowing limited interactions between editors and readers. Then, information sharing has evolved with the rising of web 2.0 by allowing users to contribute to the contents. Numerous studies have shown the position impact of such evolutions on medical information systems [11,23]. Participative web applications can be implemented and used in a collaborative way to build large databases. Finally, semantic web has appeared. Semantic web aims at creating and sharing formalized information in order to make it available for both humans and machines. Social semantic web

E. Simperl et al. (Eds.): ESWC 2012, LNCS 7295, pp. 618–632, 2012.

is considered as the merging of web 2.0 and the semantic web, i.e. a web where shared formal information is edited collaboratively.

The Kasimir research project started in 1997. It aims at providing tools to assist decision making by practitioners and, more generally, decision knowledge management in oncology. The project is conducted in partnership with Oncolor, an association gathering physicians from Lorraine (a region of France) involved in oncology. On its static website, Oncolor publishes more than 140 medical guidelines written in HTML in a web 1.0 fashion. This base is built through a consensus between medical experts and is continually updated according to the oncology state of the art and to local context evolutions. In order to facilitate the creation, maintenance and publication of guidelines, Oncolor has expressed the need for more efficient and collaborative tools. Moreover, it would be a great benefit if the knowledge contained in guidelines was formalised and made available for semantic systems, particularly for Kasimir, since knowledge acquisition is a bottleneck for building knowledge systems.

In this paper, an application of a semantic wiki approach for medical guideline edition is reported.[1] The expected benefits are twofold: first, online collaborative work is simplified by the use of wikis and second, semantic technologies allow the creation of additional services by making use of external medical resources such as terminologies, online ontologies, and medical publication websites. However, despite the effort of the semantic wiki community to simplify its systems, it is still hard for medical expert to create semantic annotations. This issue involves the need of taking into account structured and unstructured content but also, when this is possible, to include dedicated tools for formalising data. In these cases, implementation and development of semantic wiki extensions are required.

The rest of the paper is structured as follows: Section 2 describes the application context. The migration of static Oncolor website to a collaborative system is presented in Section 3, while Section 4 relates the addition of semantic annotations and services. After a report on our evaluation study in Section 5, some related work is introduced in Section 6. Section 7 is a discussion about the benefits of the system, as well as ongoing and future work.

2 Context

2.1 Application Context

Oncolor Website and Oncology Guidelines. One of Oncolor's objectives is to create and to keep up to date oncology guidelines. Clinical guidelines are sets of recommendations on treatments and care of people with specific diseases. They aim at improving treatment quality and patient support by standardising cares. They are based on clinical evidence, clinical trials and consensus between medical experts from different specialties such as oncology, surgery, etc.

More than 140 guidelines have been edited to give recommendations about treatment of many different cancers as well as typical situations such as pain

[1] `http://oncowiki.a2zi.fr`

treatment or dental care. Since guidelines are intended for both medical staff and patients, editors have exploited various kinds of formats in order to be both precise and didactic. Most guidelines follow the same structure. The first part introduces the guideline with few sentences that explain which circumstances imply the use of the guideline and the treatments that will be proposed. The next part is a textual description of clinical and paraclinical investigations that can lead to the starting point of the guideline. This starting point is often a staging step allowing to classify the patient according to international classifications. These classifications are presented as simple tables. Depending on classifications results, decision trees guide the reader to the next step that details the medical recommendation available in various formats, such as medical publications in PDF or hypertext links to distant resources. Finally, guidelines conclude with advice about medical supervision and sometimes with a lexicon of specifics terms.

As in all medical information systems, data quality in oncology is critical. Each guideline should be reviewed every second year by experts. Two kinds of editors can be identified in the reviewing process:

– Medical experts contribute with their technical knowledge. They are gathered in committees under the supervision of coordinators that make sure the guidelines are complete and the consistent. Most medical experts have poor computer skills, limited to word processing and Internet browsing.
– Oncolor staff manages communication between the committee members and creates the final guideline layout. They also check that guidelines are up to date and propose new ways to facilitate their diffusion, while public health physicians check the consistency of the information base. Most of Oncolor employees do not have more computer skills than medical experts, except for a computer graphic designer. Particularly, Oncolor does not have a webmaster in its staff.

Guidelines are made available on the Oncolor website [2], which also contains various information about local healthcare services and provides links to dedicated tools. This site also stores other Oncolor projects, including a thesaurus of pharmacological products which is closely related to oncology guidelines. It contains information about drugs used in cancer treatment.

Created in the mid 1990s, this website was completely made using a commercial WYSIWYG HTML editor. The resulting HTML code is not readable, due to successive technology evolutions. The first created pages were done using only HTML and then, in the past 15 years, CSS, Javascript and XHTML were introduced. Few pages also use ASP. All these evolutions have led to the construction of weird pages where only the visual aspect is important and in which document structure is hard to identify. Over the years, updating the website is becoming more and more complex for Oncolor staff. All the pages edited on the Oncolor website must be validated to follow the principles of HONcode certification [1] which guarantee the quality and the independence of the content.

In this context, Oncolor has been asked to integrate a collaborative tool to simplify the guideline creation and maintenance process. Moreover, it would be of great benefits for Oncolor to keep track of all changes in the guidelines. That is why the system has to propose a versioning file system and some social tools to allow communication between experts during updating process.

The Kasimir Research Project. Started in 1997, Kasimir is a multidisciplinary project also involving industrial (A2ZI) and academic (LORIA, CNAM Laboratory of Ergonomics) partners. Kasimir aims at providing software to assist decision making by practitioners and more generally decision-making knowledge management in oncology. The Kasimir project's recent work mainly focuses on semantic web as a background for formalizing, sharing, and exploiting pieces of knowledge [9]. The last version of the KatexOWL toolkit and, particularly, the framework EdHibou [4], use semantic web technologies such as SparQL and OWL for storing and exploiting pieces of knowledge. It can automatically generate simple user interfaces for decision support thanks to user-friendly forms that guide practitioners around the knowledge base.

To fill its scientific contribution, Kasimir needs to use more widely its tools by taking advantage of real world data sources. However, few guidelines are currently available for EdHibou: they need to be formalised, i.e. transformed into a knowledge base using a formalism that can be handled by an inference engine. Until now, this complex step required two experts: a medical domain expert writing guidelines and validating the final results, and a knowledge engineer formalising them. It seems that if medical domain experts could formalise the guidelines themselves in a machine-understandable way, the process would be simplified. Even if this goal seems very difficult to reach for now, it would be a good evolution if formalisation tools could help experts make simple semantic annotations.

2.2 Scientific Context

Medical Resources. To build efficient tools, it is important to take into account numerical digital resources already available. Among them, large websites reference scientific communications in the domain of medicine, such as the well-known Pubmed [3]. Pubmed provides an easily configurable search engine that can be called through distant requests. Publications are indexed using a specific controlled vocabulary, Medical Subject Headings (MeSH [20]). MeSH contains more than 25,000 descriptors, most of these accompanied by a short description or definition, some links to related descriptors, and a list of synonyms or very similar terms. In the French context, Cismef [10] uses a French traduction of MeSH to index medical online resources with a French vocabulary.

Beyond the already-cited MeSH, many controlled vocabularies have been used to structure medical applications [8]. Among resources available in French, the International Statistical Classification of Diseases and Related Health Problems (ICD-10) is probably one of the simplest. ICD-10 is a medical classification that provides codes to classify diagnoses and causes of death and is organised as a simple hierarchy. ICD-10 is widely used in medical information systems, but semantic applications generally use other vocabularies due to its lack of semantic depth. Considered the most comprehensive, SNOMED is a multiaxial, hierarchical classification system including coverage of diseases, clinical findings, therapies, procedures and outcomes. About 270,000 concepts are described by unique identifiers with several labels and can be used to describe complex situation by using semantic relations and modifiers. It is interesting to note that MeSH, ICD-10, SNOMED and other ontologies such as Galen are integrated in the terminology integration system Unified Medical Language System (UMLS).

Moreover, many semantic web systems provide freely questionable online information. For example, BioPortal [5] is a repository of biomedical ontologies whose functionalities include the ability to browse, search, and visualise ontologies. More specialised, DrugBank [25] provides an annotated database of drugs and drug target information. Many other resources are available, such as Bio2RDF [6], which allows an access to Pubmed with linked data, or LinkedCT [12] which indexes clinical trials. The information resources cited above and many more can be interlinked by using DBpedia [7].

Wikis and Semantic Wikis: The Migration Process. Traditional wikis are usually based on a set of editable pages, organised into categories and connected by hyperlinks. They became the symbol of interactivity promoted through web 2.0. One of the founding principles of wikis, which is also the principal vector of their popularity, is their ease of use even by persons that lack considerable computer skills. Wikis are created and maintained through specific content management systems, the wiki engines, while *wikitexts* enable structuring, layout, and links between articles. At this point, an idea has emerged: to exploit stored pieces of knowledge automatically.

Indeed, a limit use to the wikis is illustrated by the querying of the data contained in their pages. The search is usually done through word recognition by strings, without considering their meaning. For example, the system cannot answer a query like: "Give me the list of all currently reigning kings." The solution used in Wikipedia is a manual generation of lists. However, the manual generation of all the lists answering queries users may raise is, at the very least, tedious, if not impossible. This has motivated the introduction of a semantic layer to wikis. Moreover, it would be interesting if information contained in wikis were available through external services.

Semantic wikis were born from the application of wiki principles in the semantic web context. A semantic wiki is similar to a traditional one in the sense that it is a website where contents are edited in a collaborative way by users and are organised into editable and searchable pages. However, semantic wikis are not limited to natural language text. They characterise the resources and the links between them. This information is formalised and thus becomes usable by a machine, through processes of artificial reasoning. Thus, semantic wikis can be viewed as wikis that are improved by the use of semantic technologies as well as collaborative tools for editing formalised knowledge.

Semantic wikis corresponds to both Oncolor and Kasimir needs: guidelines can be written in a collaborative way and semantic technologies allow to formalise and extract structured content.

3 From Web 1.0 to Web 2.0

3.1 Choosing a Semantic Wiki Engine

The first part of the migration was choosing the most adapted semantic wiki engines. Whereas many semantic wiki engines have emerged for the last 10 years, only four open source projects seem active at this time: AceWiki [17], KiWI [22], Ontowiki [13], and Semantic Mediawiki [16]. AceWiki uses ACE (*Attempto Controlled English*), a sub-language of English that can be translated directly into first order logic. However, Oncolor guidelines are already written in French and the development of a controlled language for French medical guidelines that covers all the contents would be tedious. Ontowiki and KiWI focus on RDF triple edition by proposing dedicated interfaces such as dynamic forms. Their approaches are very strict and do not seem reconcilable with importation of unstructured contents. Moreover, no large scale implementation of these engines can be found and, their development and user communities are limited. So, less extensions are available and the support is weak.

Semantic Mediawiki (SMW) seems to be the best solution. SMW is an extension of Mediawiki, the engine used by Wikipedia. For the sake of simplicity for users, it integrates the RDF triples editing in its wikitext. In this way, it enables the creation of typed links that can also be used for indicating the attributes of the page. Another interesting point of SMW is its popularity: there is a large community of developers around it, and this community produces many extensions, such as editing forms, the integration of an inference engine, etc. For instance, the Halo extension[2] proposes forms, an auto-completion system, the integration of a SPARQL endpoint and much more. The only limitation for our migration is that SMW does not provide extensions that allow to draw the trees that are frequently used in the guidelines, but we have developed a decision tree editor, as will be discussed further. Tutorials and community support make the installation of SMW simple. Less than one hour is needed to install it for anybody with average computer skills.

3.2 Importing Guideline Content

Once the semantic wiki had been installed, a specific skin that corresponds to Oncolor graphics standards has been built to customise the application. The next part of the work was to import guidelines in the wiki. However, in order to correspond to wiki syntax, content had to be formatted into wikitext. For each guideline, the HTML content was extracted and HTML pages were merged when guidelines did contain more than one page. The table of contents was automatically extracted and marked up when possible. However, the state of the HTML code made impossible to systemically identify document structure. It can be noted that the migration would have been simpler if CSS had been used from the start. Then, unnecessary content such as browsing elements and

[2] http://www.projecthalo.com/

Fig. 1. An excerpt of guideline in the wiki

JavaScript functions was removed. A parser was also used to transform HTML into wikitext when simple tags were detected (images, tables, etc.). Moreover, by using a parser and context analysis, specific fields were identified. The objective was to identify interesting information about a guideline such as the date of its last update or keywords. Moreover, by examining website folder structure, an anatomical classification of the guideline has been identify. This classification was reused as a base for guideline categorisation in the wiki.

Despite of all our efforts, the layout of the imported guidelines had to be checked then. Due to the critical nature of the information, this checking was done by Oncolor staff. On average, a person needed half a day to check each guideline.

Additionally, the Oncolor thesaurus of pharmacology was imported. As its content is closely related to guidelines, it was important to let it available in the same information system. One page per described drug was created. In this case, the simplicity of the HTML pages made the migration easier.

To migrate guidelines, Mediawiki import capacities were used. They allow to import wikitext content from text files. In the wiki, some templates were built to highlight the fields previously identified. An excerpt of a resulting page is shown in Figure 1. All the guidelines are presently in the wiki.

3.3 User Right Management

In the usual philosophy of wikis, everybody can edit pages, even anonymously. Although the importance of the information availability for the public, medical data are critical and the guidelines must be approved by Oncolor experts to be in public access. Moreover, if an expert modifies a guideline, the modification has to be checked by the coordinator in charge of the guideline. During the revision period, modifications are numerous and each of them implies a complete review of the guideline and its layout. To allow private modifications, a special namespace has been created, that can be viewed as a workspace for the experts. Final versions are shown on the main namespace, and each guideline has an equivalent in the new namespace where experts can add their contributions. When a guideline is considered as correct and complete by the coordinator in the workspace, the page is simply copied to the final location in the main namespace.

According to this revision process, three kinds of users have been identified:

– anonymous users, that can read pages of the main namespace,
– medical experts, that can read pages of the main namespace and edit pages in the workspace,
– administrators, that can edit all pages, even wiki system pages.

3.4 KcatoS, a Decision Tree Editor

Decision trees were imported from the previous website as bitmap pictures. At this point, guideline updates can also be simplified by proposing an online editor. KcatoS is a Mediawiki extension that allows the collaborative drawing of decision trees. KcatoS decision tree language is a graphical representation based on a small set of geometrical figures connected by directed edges. This representation was directly inspired by the graphics standards of Oncolor. Indeed, guidelines use visual representations that can mostly be viewed as trees. An advantage to use these graphics standards is that Oncolor experts already know them. We want to preserve Oncolor's graphic semantics in order to facilitate the understanding of guidelines by physicians.

From a semantic point of view, each kind of node has its own meaning; e.g. rounded rectangles represent medical situations, etc.

4 Introducing and Exploiting Formalised Knowledge

4.1 Extracting Decision Knowledge from Decision Trees

Most of the time, decision trees can be considered as structures from which a meaning can be extracted. In order to avoid ambiguities and to guarantee guideline consistency, classical syntactical rules of decision trees are used. A syntactic module can be used to check if the edited tree respects the rules. Thus, KcatoS can propose an export algorithm that allows to transform decision trees into OWL.

KcatoS's export algorithm defines two classes: `Situation` and `Recommendation`. The first one represents some patient information while the second one represents the description of the decision proposed by the system. These classes are linked by the property `hasRecommendation`. This means that for each situation there is a recommendation that is associated to it.

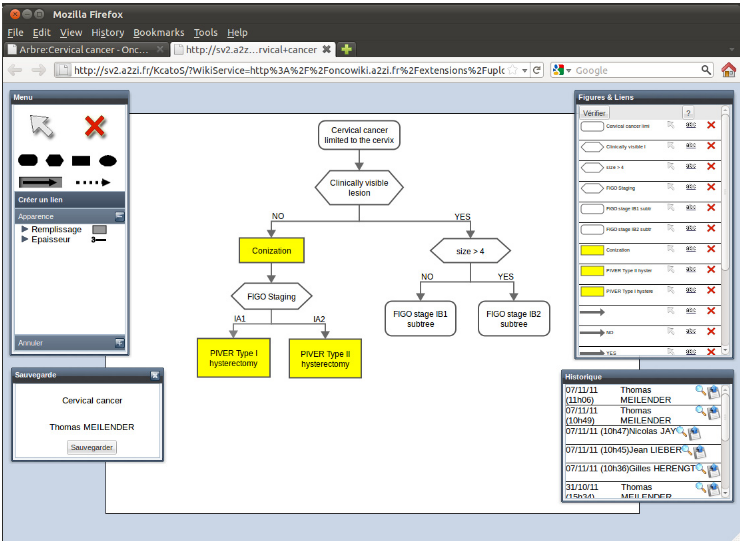


Fig. 2. The KCATOS decision tree editor interface

A tree is read using depth-first search. Each node is transformed using rules which take into account the shape and its ancestors.

The export algorithm creates many concepts and properties. Including all of them in the semantic wiki would decrease the ease of navigation because it would lead to the creation of numerous pages. In order to avoid these page creations, translated trees are stored in a specific file and linked to the wiki. Thus, created ontologies are made available for other semantic web applications. From a technical point of view, OWL API [15] is used to perform the export.

4.2 Using Semantic Tools of Wiki

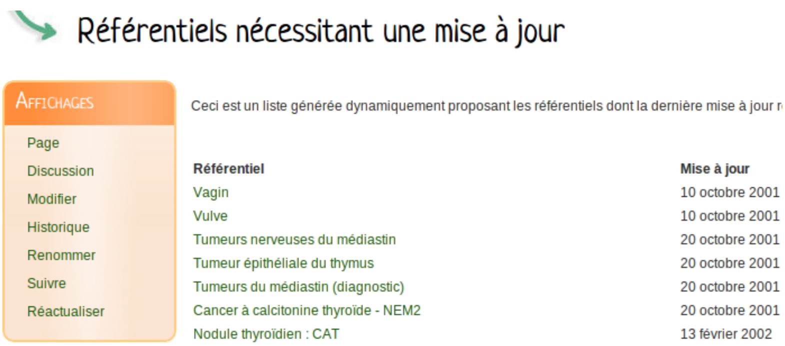
Extracting the whole semantics of a guideline is a tedious job that has to be done by a medical expert with skills in knowledge engineering. As Oncolor does not have this kind of specialist in its staff, formalising the guidelines would be a great investment. Moreover, it is still difficult for non-specialists to understand the benefits that semantics could bring to medical information system. That is why the key idea of the project is to insert step-by-step useful semantic annotations into the guidelines in order to increase Oncolor interest in the semantic web technologies. The first way to introduce semantics is to exploit identified fields extracted during the guideline migration. To improve their visualisation and their update, SMW templates and queries mechanisms were used.

SMW proposes many ways to edit semantic annotations. The more basic way to create annotations is wikitext, which can be improve thanks to templates. Templates are generic pre-developed page layouts that can be embedded in several wiki pages. They can also manage variables that are instantiated in the corresponding page. For instance, a template is used to generate the box in the top right corner of the page shown in Figure 1. The template used to create this box is generic enough to be applied to all guideline pages, and its use allows flexible modifications. As template use is simple (and can be further simplified by associating forms to them), they provide a simple way to create annotation fields that can be filled by any users without specific skills.

Then semantic annotations can be exploited by SMW inline query engine. Using a simple query language, semantic search can be done directly in a page and results are displayed as tables, lists, etc. Combined to templates, semantic queries are a simple way to create dynamic content relying on semantic annotations.

```
{{#ask:[[Category:Guideline]] [[last Update::<{{#time:d F Y|2 years ago}}]]
|?last Update
| sort = last Update
| format=template
| [...]
}}
```

(a) Excerpt of inline query that requests the guidelines that are out-of-date (translated from French).



(b) Results of the query.

Fig. 3. An excerpt of inline query that requests the guidelines that are out-of-date, and the wiki page that contains the result

A use of templates and inline queries is shown by the management of the dates in the guidelines. Every guideline has at least one date that indicates the date of the last validated update. This date is entered in a template in which it is associated with a property which links the date to the guideline. Then, a maintenance page is created to highlight the guidelines that are out-of-date. The query is shown in Figure 3(a) while its result, that can be seen in Figure 3(b), is displayed as a table thanks to specific templates. Moreover, another query is added in the template present on each guideline which shows a warning if the guideline has to be updated.

Templates are also used to link guidelines to external publication resources. To create the link, the first step was to define a common vocabulary between guidelines and publication website. Then, templates were designed to allow easily semantic annotations in guideline using MeSH vocabulary. Cismef, which indexes a large amount of medical publication in French, already indexes Oncolor's guidelines using terms from the MeSH thesaurus. These terms were imported in the wiki as a base that can be freely edited. As PubMed also uses this thesaurus to index this document, requests to PubMed and Cismef can be automatically built using templates and inline queries. Each request is dedicated to the guideline it belongs to and provides publications that are indexes by the same terms. Thus, it provides a bibliography tool useful for staff and provides further information to the reader.

4.3 Querying Resources of Web of Data

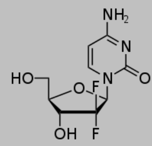
To show another view of the semantic web, we tried to investigate on external structured data sources that could bring additional information to the wiki. So, an extension was created to query external sources using SPARQL. In this part, pharmacology thesaurus was used. The idea was to explore external resources by building SPARQL requests based on the name of the drug studied in a current page. The target of the searches was Drugbank, specialised in drug description, and DBpedia, a generalist knowledge base. Thus, for most of the drugs, we get additional information in the semantic web. An example is shown in Figure 4. However, most information are in English and we deplore the lack of available French information source. This module is no longer online pending the Oncolor board is approval of the use of external data sources and the validation of the ones that can be exploited.

5 Evaluation

To carry out the evaluation, the opinions of the users have been investigated. People asked were the four main contributors from Oncolor staff: two public health physicians, a computer graphic designer, and a medical secretary.

The first interesting point is that, before the beginning, the only thing they knew about wikis was Wikipedia and none had ever contributed to a wiki. Despite this, three contributors thought that less than one day of self-training is needed to learn wikitext and to be an efficient contributor. The only difficulties are related to particular layouts (tables and references) and wiki advanced functions dealing with user management and page history. The only reluctance to migrate to a wiki was guideline quality. They agreed a concern with that the old system was time-consuming, but it had the advantage to produce high quality guidelines. Experiments were led to update Oncolor's old website and semantic wiki with the same modifications. They show that the quality did not suffer of the change and that the efficiency of updating has been increased by the semantic wiki.

Gemcitabine



Description (DrugBank)	Gemcitabine is a nucleoside analog used as chemotherapy. It is marketed as Gemzarï¿½ï¿½ by Eli Lilly and Company. As with fluorouracil and other analogues of pyrimidines, the drug replaces one of the building blocks of nucleic acids, in this case cytidine, during DNA replication. The process arrests tumor growth, as new nucleosides cannot be attached to the "faulty" nucleoside, resulting in apoptosis (cellular "suicide"). Gemcitabine is used in various carcinomas: non-small cell lung cancer, pancreatic cancer, bladder cancer and breast cancer. It is being investigated for use in oesophageal cancer, and is used experimentally in lymphomas and various other tumor types.
Mechanism of action	Gemcitabine inhibits thymidylate synthetase, leading to inhibition of DNA synthesis and cell death. Gemcitabine is a prodrug so activity occurs as a result of intracellular conversion to two active metabolites, gemcitabine diphosphate and gemcitabine triphosphate by deoxycitidine kinase. Gemcitabine diphosphate inhibits ribonucleotide reductase, the enzyme responsible for catalyzing synthesis of deoxynucleoside triphosphates required for DNA synthesis. Gemcitabine triphosphate (diflurorodeoxycytidine triphosphate) competes with endogenous deoxynucleoside triphosphates for incorporation into DNA.
Toxicity	Myelosuppression, paresthesias, and severe rash were the principal toxicities, LD_{50}=500 mg/kg (orally in mice and rats)
Pharmacology	Gemcitabine is an antineoplastic anti-metabolite. Anti-metabolites masquerade as purine or pyrimidine - which become the building blocks of DNA. They prevent these substances becoming incorporated in to DNA during the "S" phase (or DNA synthesis phase of the cell cycle), stopping normal development and division. Gemcitabine blocks an enzyme which converts the cytosine nucleotide into the deoxy derivative. In addition, DNA synthesis is further inhibited because Gemcitabine blocks the incorporation of the thymidine nucleotide into the DNA strand.
Indication	For the first-line treatment of patients with metastatic breast cancer, locally advanced (Stage IIIA or IIIB), or metastatic (Stage IV) non-small cell lung cancer and as first-line treatment for patients with adenocarcinoma of the pancreas.
Sources	http://www4.wiwiss.fu-berlin.de/drugbank/resource/drugs/DB00441 http://dbpedia.org/resource/Gemcitabine

Fig. 4. Example of data that can be imported from DBpedia and Drugbank about *Gemcitabine* using SPARQL queries

Our panel cited the main advantages they see in using a wiki. They have agreed that wikis are collaborative tools that allow more reactivity and more flexibility in the update process. It has also been said that wikis improve conditions of employment by allowing distant work, which was impossible with the previous system. Moreover, they recognised that the wiki increases the quality of the editing process and of the guideline themselves by allowing the standardisation of the guideline and by simplifying the work on its layout.

In our system, the preferred contribution is the query to medical publication websites Pubmed and Cismef which propose automatically a bibliography related to a guideline. The previous system did not permit that kind of function that has been judged very useful. It is really important for the project that Oncolor staff appreciated this contribution that is relying on semantic web technologies. Moreover, all participants declared that they are interested in using MeSH annotations and want to lead further this experimentation.

6 Related Work

It already exists many medical wikis (e.g. medical portal of Wikipedia, `http://wikisr.openmedicine.ca`, `http://askdrwiki.com`, `www.ganfyd.org`, etc.) but only few of them use semantic web technologies. OpenDrugWiki [18], which also uses SMW, is a wiki used as a back-office system for editing, merging from different sources, and reviewing information about drugs.

The closest semantic wiki to the one introduced in this paper is probably CliP-MoKi [21]. CliP-MoKi is a SMW-based tool for the collaborative encoding in a distributed environment of cancer treatment protocols. The wiki mainly relies on semantic forms and focuses totally on structured content while our project aims at migrating already existing unstructured data.

Semantic wikis have already been experimented in various domains. Particularly, the building of a semantic portal for the AIFB Institute described in [14] shows how important the technical settings are for increasing wiki performances and how difficult it is to find the right balance between structured and unstructured data. This last issue has also been tackled in [24].

7 Lessons Learnt and Future Work

In this paper, a migration from a web 1.0 website containing medical data to a semantic wiki has been described. The first step was the migration of data from an HTML website to a collaborative solution, Semantic Mediawiki. The second step consisted in adding a semantic layer to show the benefits that semantic web technologies could bring.

Among the difficulties we have met, the analysis of the HTML version of the guidelines was hard because of the use of invalid code. This is the result of the use of different HTML editors that follow the evolution of the standard over a decade. It appears that a correct use of HTML and CSS would have simplified the migration, particularly the identification of tables of content and specific fields. Moreover, medical information is critical and its migration implies a long work of verification by medical experts. According to Oncolor members, about 70 days of work were necessary to check and correct all the guidelines.

Once the semantic wiki has been installed, the use of traditional wiki tools for edition was easily learnt by Oncolor staff. However, we have noticed that the creation and the use of semantic annotations remain difficult for non knowledge expert although semantic wikis seem to be a simple approach. For example, SMW inline query language is hard to handle for non computer specialists and template construction also requires computer skills. Some tools have yet to be implemented to improve this aspect in the philosophy of semantic forms and the Halo project.

Another problem was to find the right balance between structured and unstructured data. The advantage of structured data is the typing that enables to easily reuse data in the semantic web context. However, structured data are still difficult to edit and exploit, as shown in the context of semantic wikis. Moreover, most of existing information sources are unstructured, and tedious work would be necessary to transform them. This job would be expensive and time-consuming so its benefits have to be shown first to non semantic web experts. Our methodology was to add semantic annotations step-by-step to improve the semantic wiki quality. Until now, our work has consisted in showing the improvements so that future developments will be upon Oncolor request.

Introducing structured information yields benefits when it is done in accordance with already existing resources. In the medical domain, numerous thesauri and information sources have been created, and it is hard for no medical specialists to determine which ones can be used. This choice has to be made according to the goal of the application with the approval of medical specialists. For instance, it was hard to determine which thesaurus will be used to index guidelines. We finally have chosen MeSH upon Oncolor request, although SNOMED or UMLS seem more complete and CIM-10 seems more simple. The reason was that the link to medical publication websites is useful for editors and provides additional information for the readers.

Finally, the use of data from semantic web is a major concern in the medical domain, due to the critical nature of the data. Using external resources seems to cause a kind of reluctance in clinicians. Each source has to be first approved by medical authorities before it can be exploited by a medical system. Particularly, all sources must at least follow the principle of the HONcode certification.

Currently, our work focuses on minor technical adaptation of the wiki to Oncolor needs. Our next task will be to increase gradually the semantic annotation's presence. The long-term goal is to obtain a structured knowledge base that contains all the information provided by oncology guidelines. For such a project to be successful, several issues have to be taken into account. The project must be able to rely on several medical experts to structure and check information. From this point of view, Oncolor will have a crucial role of support to play and so, their satisfaction is really important. Moreover, to complete the formalisation, resources that are more expressive than MeSH will be needed. SNOMED or UMLS seem to be better options. Finally, the scale of this final ontology will require significant improvement in ontology engineering tools, particularly for the edition and the maintenance.

References

1. Honcode, `http://www.hon.ch/` (last consulted: December 2011)
2. Oncolor website, `http://www.oncolor.fr` (last consulted: December 2011)
3. Pubmed, `http://www.ncbi.nlm.nih.gov/pubmed/` (last consulted: December 2011)
4. Badra, F., d'Aquin, M., Lieber, J., Meilender, T.: EdHibou: a customizable interface for decision support in a semantic portal. In: Proc. of the Poster and Demo. Session at the 7th International Semantic Web Conference (ISWC 2008), Karlsruhe, Germany, October 28 (2008)
5. Bail, S., Horridge, M., Parsia, B., Sattler, U.: The Justificatory Structure of the NCBO BioPortal Ontologies. In: Aroyo, L., Welty, C., Alani, H., Taylor, J., Bernstein, A., Kagal, L., Noy, N., Blomqvist, E. (eds.) ISWC 2011, Part I. LNCS, vol. 7031, pp. 67–82. Springer, Heidelberg (2011)
6. Belleau, F., Nolin, M.-A., Tourigny, N., Rigault, P., Morissette, J.: Bio2rdf: Towards a mashup to build bioinformatics knowledge systems. Journal of Biomedical Informatics 41(5), 706–716 (2008)
7. Bizer, C., Lehmann, J., Kobilarov, G., Auer, S., Becker, C., Cyganiak, R., Hellmann, S.: Dbpedia - a crystallization point for the web of data. Journal of Web Semantics: Science, Services and Agents on the WWW 7(3), 154–165 (2009)
8. Bodenreider, O.: Biomedical ontologies in action: Role in knowledge management, data integration and decision support. In: IMIA Yearbook Medical Informatics, pp. 67–79 (2008)

9. D'Aquin, M., Brachais, S., Lieber, J., Napoli, A.: Decision Support and Knowledge Management in Oncology using Hierarchical Classification. In: Kaiser, K., Miksch, S., Tu, S.W. (eds.) Proc. of the Symp. on Computerized Guidelines and Protocols, CGP 2004, Prague, Czech Republic. Studies in Health Technology and Informatics, vol. 101, pp. 16–30. IOS Press (2004)
10. Darmoni, S.J., Thirion, B., Leroyt, J.P., Douyère, M., Lacoste, B., Godard, C., Rigolle, I., Brisou, M., Videau, S., Goupyt, E., Piott, J., Quéré, M., Ouazir, S., Abdulrab, H.: A search tool based on 'encapsulated' MeSH thesaurus to retrieve quality health resources on the internet. Medical Informatics and The Internet in Medicine 26(3), 165–178 (2001)
11. Giustini, D.: How Web 2.0 is changing medicine. BMJ 333, 1283–1284 (2006)
12. Hassanzadeh, O., Kementsietsidis, A., Lim, L., Miller, R.J., Wang, M.: Linkedct: A linked data space for clinical trials. CoRR, abs/0908.0567 (2009)
13. Heino, N., Dietzold, S., Martin, M., Auer, S.: Developing Semantic Web Applications with the OntoWiki Framework. In: Pellegrini, T., Auer, S., Tochtermann, K., Schaffert, S. (eds.) Networked Knowledge - Networked Media. SCI, vol. 221, pp. 61–77. Springer, Heidelberg (2009)
14. Herzig, D.M., Ell, B.: Semantic MediaWiki in Operation: Experiences with Building a Semantic Portal. In: Patel-Schneider, P.F., Pan, Y., Hitzler, P., Mika, P., Zhang, L., Pan, J.Z., Horrocks, I., Glimm, B. (eds.) ISWC 2010, Part II. LNCS, vol. 6497, pp. 114–128. Springer, Heidelberg (2010)
15. Horridge, M., Bechhofer, S.: The owl api: A java api for owl ontologies. Semantic Web 2(1), 11–21 (2011)
16. Krötzsch, M., Vrandecic, D., Völkel, M., Haller, H., Studer, R.: Semantic wikipedia. Journal of Web Semantics 5, 251–261 (2007)
17. Kuhn, T.: How controlled english can improve semantic wikis. In: Proc. of the 4th Workshop on Semantic Wikis, European Semantic Web Conference 2009. CEUR Workshop Proceedings (2009)
18. Köstlbacher, A., Maurus, J., Hammwöhner, R., Haas, A., Haen, E., Hiemke, C.: Opendrugwiki – using a semantic wiki for consolidating, editing and reviewing of existing heterogeneous drug data. In: 5th Workshop on Semantic Wikis Linking Data and People, SemWiki 2010 (May 2010)
19. Lange, C., Schaffert, S., Skaf-Molli, H., Völkel, M. (eds.): 4th Semantic Wiki Workshop (SemWiki 2009) at the 6th European Semantic Web Conference (ESWC 2009), Hersonissos, Greece, June 1. CEUR Workshop Proc., vol. 464. CEUR-WS.org (2009)
20. Nelson, S.J.: Medical terminologies that work: The example of mesh. In: Proc. of the 10th Int. Symposium on Pervasive Systems, Algorithms, and Networks, December 14-16, pp. 380–384 (2009)
21. Rospocher, M., Eccher, C., Ghidini, C., Hasan, R., Seyfang, A., Ferro, A., Miksch, S.: Collaborative Encoding of Asbru Clinical Protocols, pp. 1–8 (2010)
22. Schaffert, S., Eder, J., Grünwald, S., Kurz, T., Radulescu, M., Sint, R., Stroka, S.: Kiwi - a platform for semantic social software. In: Lange, et al. [19]
23. Schreiber, W., Giustini, D.: Pathology in the era of Web 2.0. American Journal of Clinical Pathology 132, 824–828 (2009)
24. Sint, R., Stroka, S., Schaffert, S., Ferstl, R.: Combining unstructured, fully structured and semi-structured information in semantic wikis. In: Lange, et al. [19]
25. Wishart, D.S., Knox, C., Guo, A., Cheng, D., Shrivastava, S., Tzur, D., Gautam, B., Hassanali, M.: Drugbank: a knowledgebase for drugs, drug actions and drug targets. Nucleic Acids Research 36(Database-Issue), 901–906 (2008)