Metadata element sets in the CISMeF Quality-Controlled Health Gateway

Benoit Thirion [1], Magaly Douyère [1], Lina F. Soualmia [1, 2], Badisse Dahamna [3], Jean-Philippe Leroy [4], Stéfan J. Darmoni [1, 2]

[1] *L@sTICS, Rouen University Hospital & Medical School, France*
[2] *Perception, Information and System Lab, INSA Rouen & Rouen University, France*
[3] *Computing & Network Department, Rouen University Hospital, France*
[4] *Toxicology Center, Rouen University Hospital, France*
Benoit.Thirion@chu-rouen.fr

Abstract

Background: *Quality-controlled subject gateways are Internet services which apply a selected set of targeted measures to support systematic resource discovery. Considerable manual effort is used to operate a selection of resources which meet quality criteria and to display an extensive description and indexing of these resources with standards-based metadata.* ***Objective***: *Several metadata element sets are proposed to describe, index and qualify health resources to be included in a French quality-controlled health gateway called CISMeF. The main objectives were to enhance Internet health document retrieval and navigation, and to allow interoperability with other Internet services.* ***Results***: *The Dublin Core metadata element set is used to describe and index all Internet health resources included in CISMeF. For teaching resources, some elements from IEEE1484 Learning Object Metadata are also used. For evidenced-base medicine resources, specific metadata are employed which assess the health content quality. The HIDDEL metadata set is used to enhance transparency, trust and quality of health information on the Internet.* ***Conclusion***: *Comprehensive metadata element sets can be extremely useful to describe, index and assess health resources on the Internet in a quality-controlled subject gateway. Machine-readable metadata creates an Semantic Web which is more efficient for end-users as compared to the current Web.*

Keywords:
CISMeF; Catalogue et Index des Sites Médicaux Francophones; France; health resources; Internet; medical information; quality subject gateway; quality control;

1. Introduction

Metadata is concise information concerning all types of data. On the Internet, the term metadata specifically refers to: 1) descriptive information about the Web resources used to improve information retrieval, 2) it refers to the content, structure and logistical information of all data including electronic resources, 3) it is used for data discovery and control of data, 4) it helps to enhance Internet document retrieval and navigation. There is a need for

https://doi.org/10.23106/dcmi.952107865

an interoperable infrastructure for Digital Libraries, quality-controlled subject gateways, and other Web-based services that rely on cross-institutional and cross-border co-operation. Agreement on a metadata standard which serves as a starting point for information exchange in specific domains and provides a common ground for cross-domain interoperability, is a crucial element of this infrastructure. The main metadata standard from a cross-domain perspective is the Dublin Core, now recommended across Europe for use in many sectors as the gold standard of choice to ensure interoperability between resource discovery systems on the Internet.

Quality-controlled subject gateways were defined by Koch [1] as Internet services which apply a comprehensive set of quality measures to support systematic resource discovery. Considerable manual effort is used to operate a selection of resources which meet quality criteria and to display an extensive description and indexing of these resources with standards-based metadata. Regular checking and updating ensure optimal collection management. The main goal is to provide a high-quality of subject access through resource indexing using controlled vocabularies and by offering a deep classification structure for advanced searching and browsing.

The objective of CISMeF (French acronym for Catalog and Index of French-language health resources) [2-3] is to describe and index the main French- language health resources in order to assist health professionals and consumers in their search for electronic information available on the Internet. CISMeF is a quality-controlled subject gateway initiated by the Rouen University Hospital (RUH). Its Universal Resource Locator (URL) is http://www.chu-rouen.fr/cismef. CISMeF began in February 1995. In March 2004, the number of indexed resources totalled over 13,500, with an average of 50 new resources indexed each week. Each of the following phases proposed by Koch, which characterise a typical quality-controlled subject gateway, are implemented in CISMeF: (a) selection and collection development, based on the Net Scoring, a list of 49 criteria to assess quality of health information (URL: http://www.chu-rouen.fr/netscoring) [4], (b) collection management, (c) intellectual creation of metadata (done by experts), (d) resource description (an extensive and documented metadata set), and (e) resource indexing (using a controlled vocabulary system).

The main objectives of this work were to enhance Internet health document retrieval and navigation, and to permit interoperability with other Internet services. To allow interoperability, gateways apply open standards. CISMeF uses two standard tools for organizing information: the MeSH (Medical Subject Heading) thesaurus from the US National Library of Medicine (NLM), and several metadata element sets: (a) the Dublin Core metadata format [5] to describe and index all the health resources included in CISMeF, (b) some elements from IEEE1484 Learning Object Metadata for teaching resources [6], (c) specific metadata for evidenced-base medicine resources which also qualify the health content, and (d) the HIDDEL metadata set [7] will be used to enhance transparency, trust and quality of health information on the Internet in the EU-funded MedCIRCLE project.

2. Results

2.1. Use of Metadata

Using DCMES in CISMeF:

The Dublin Core Metadata Initiative (DCMI) is a project from the Online Computer Library Center (OCLC) and the National Center for Supercomputing Applications (NCSA). The Dublin Core Metadata Element Set (DCMES) is a 15-element set intended to facilitate the discovery of electronic resources (URL: http://dublincore.org). The DCMES is intended to be used by non-cataloguers as well as resource description specialists. CISMeF covers

https://doi.org/10.23106/dcmi.952107865

three main health areas: evidence-based medicine, education of health care professionals and students, and patient education. Specific metadata are used for these domain.

The fifteen Dublin Core elements are optional and repeatable. Resources included in CISMeF are described by 11 of 15 items taken from version 1.1 of the DCMES (URL: http://dublincore.org/documents/dces/). These are: author or creator, date, description, format, identifier, language, publisher, resource type, rights, subject and keywords, and title. CISMeF does not use the 4 other DCMES items (contributor, coverage, relation, source) [9] because they were not necessary to describe health resources to be included in CISMeF. To capture more information for each health resource indexed in CISMeF, another element set was developed locally to meet specific search and retrieval needs. The following eight fields were added in the data and metadata and are specific to CISMeF: institution, city, province or state, country, target or audience, type of access, cost and sponsorship. Some of these fields (e.g. cost) are also present in LOM.

From 1995 to 1999, CISMEF used only static HTML. As CISMeF uses the MeSH to index resources, each HTML page is based on a MeSH term and includes Dublin Core metadata. In March 2004, CISMeF used 9,903 MeSH terms (44% of the MeSH thesaurus). Since 2000, CISMeF has also included a database and a search tool which generates an HTML (or XML or RDF) page for every indexed resource.

Metadata in Education (IEEE 1484 Learning Objects Metadata (LOM)):

The IEEE 1484 Learning Object Metadata (LOM) (URL: http://ltsc.ieee.org/doc/wg12/LOM_WD6_4.pdf) version 6.4 contains around 80 elements in the following nine categories: General,. Lifecycle, Meta-metadata, Technical Educational, Rights, Relation, Annotation, Classification. LOM metadata includes the 15 DCMES elements (see URL: http://www.ischool.washington.edu/sasutton/IEEE1484.html).

CISMeF is one of the search tools of the French Medical Virtual University (FMVU) Consortium which was created to test various tools and methods required to build a virtual university (URL: http://www.umvf.org) [10]. To describe and index teaching resources, this consortium has decided to use in its search tools only the 11 elements of the LOM Educational category because they are the most specific. Also, a feasibility study has showed that: the CISMeF team spends an average of 30 minutes to describe and index a teaching resource with the Dublin Core set and needs 30 minutes more for the LOM Educational subset.

Recently, DCMES proposed a new section DC.education to develop and promote a set of basic principles for the development and application of modular interoperable metadata for dissemination to the global education and training communities. DC.education will map with several elements of the LOM Educational subset. The CISMeF metadata element set will use DC.education as soon as it will be finally approved. The FMVU Consortium is now studying the use of Sharable Content Object Reference Model (SCORM) (see Discussion).

Description of EBM (Evidence-Based-Medicine) metadata element set:

CISMeF uses two specific metadata elements for EBM resources and more broadly ‘sensitive’ information. Sensitive information is defined as information found in documents published on the Internet, which could be used in a medical decision: These two metadata elements

are: (a) indication of the level of evidence which we proposed to be the main criterion chosen for the quality of the health information content [8] and (b) the method used to calculate the level of evidence as more than twenty are currently used in the literature. CISMeF explicitly indicates if the level of evidence is mentioned for each indexed 'sensitive' document. Furthermore, this criterion is easily searchable using the Doc'CISMeF

https://doi.org/10.23106/dcmi.952107865

search tool [3].

The use of HIDDEL in CISMeF:

HIDDEL is a standard vocabulary/metadata language developed in the MEDCIRCLE project (URL: http://www.medcircle.info) [7]. HIDDEL is used to enhance transparency, trust and quality of health information on the Internet. HIDDEL is designed to be used by 1) information providers to describe and disclose properties of e-health services (self-rating) and 2) third-parties, e.g. by subject gateways, to express third-party opinions about health information providers [3].

CISMeF is a member of the MedCIRCLE project which is a collaboration of trusted European health subject gateways, medical associations, accreditation, certification, or rating services, which share the common goal of evaluating, describing, or indexing health information. The MedCIRCLE project is funded by the European Union under the Action Plan for Safer Use of the Internet (URL: http://www.europa.eu.int/information_society/programmes/iap/index_en.htm). This project began in March 2002 and lasts 18 months till December 2004. As a quality-controlled subject gateway, CISMeF uses HIDDEL only as a third-party. Some elements of the HIDDEL are similar to Dublin Core (e.g. HIDDEL.Identity and DC.Author). Most of the HIDDEL elements are common with the Net Scoring previously used by CISMeF whereas some are already present in the CISMeF database (e.g. HIDDEL.policies). CISMeF focuses this rating on the main French publishers of health resources (national agencies, medical societies, universities and hospitals) which are included in the CISMeF database. In CISMeF, each publisher will have a MedCIRCLE seal with a link to the MedCIRCLE central repository where HIDDEL metadata elements are displayed. CISMeF will apply full transitivity from these publishers: each document from one MedCIRCLE rated publisher which is indexed in CISMeF will also receive the MedCIRCLE seal of the publisher with the same link to MedCIRCLE central repository. Thanks to this transitivity, 7,053 documents from these publishers (53.3% of the 13,227 resources included in CISMeF).resources have a MedCIRCLE seal in CISMeF in March 2004 [11]. The HIDDEL language is integrated in the CISMeF database and in the CISMeF pages via RDF into HTML .

2.2. Interoperability

The CISMeF metadata element sets are useful for cross-searching distributed and heterogenous subject gateways. We have successfully tested the interoperability of the CISMeF metadata element sets with the FMVU e-learning platform using the XML version of CISMeF resource pages, especially with the Grenoble Medical School (URL: http://www-sante.ujf-grenoble.fr/SANTE/corpus/corpus.htm). This Medical School is sending the CISMeF metadata sets in XML format, which are distributively filled by the authors of the Grenoble teaching resources. These metadata are automatically filled in the CISMeF database and thereafter searchable by any end-user, thanks to the interoperability. The indexing process by the CISMeF team is performed afterwards. Finally, this overall process creates a pre-CISMeF search possibility very similar to the PreMEDLINE search possibility. PreMEDLINE is in-process database for MEDLINE. It provides basic information and abstracts before a record is indexed with MeSH headings and added to MEDLINE.

These metadata elements were manually written and updated by the CISMeF team from 1995 to 1999 and currently automatically created and updated from the CISMeF database. Till 2002, the CISMeF metadata element set was mostly human-readable and not so easily machine-processable. Since August 2002, we have used Resource Description Framework (RDF) into HTML (URL: http://www.w3.org/RDF) in order to make it easily machine-

https://doi.org/10.23106/dcmi.952107865

processable and therefore to fulfil one of the main goals of this metadata element set: to become interoperable with other Internet services.

Within the MedCIRCLE project, the interoperability process consists of an exchange of RDF files, containing experts' annotations written in HIDDEL. The semantic-based Archer Annotation System deals with RDF annotations reception. Archer is a web application that allows annotating health information websites using the HIDDEL vocabulary. It is a technical platform and an organizational infrastructure that can be used by consumers, health information providers, and third party rating services. The first version of Archer was implemented as a part of MedCERTAIN, and further enhanced in the course of the successor project MedCIRCLE to allow the exchange of metadata between third party rating organizations. On another ground, through its search engine Doc'CISMeF, CISMeF provides external links to Archer backend servlets, and internal links to rated sites disclosure (see Figure 1).

RDF is a language for encoding knowledge on Web pages to be used by electronic agents searching information. Developed by the World Wide Web Consortium (W3C), it is a major building block of the Semantic Web initiative [12].

2.3. CISMeF terminology

Besides the metadata element sets described above, CISMeF uses another standard tool for organizing information: the MeSH (Medical Subject Heading) thesaurus from the US National Library of Medicine and its French translation by the French Medlars Center (French National Institute of Health).

These concepts are organized into hierarchies going from the most general ones at the top of the hierarchy to the most specific at the bottom of the hierarchy. For example, the descriptor *hepatitis* is more general than the descriptor *hepatitis viral A*. The qualifiers, also organized into hierarchies, allow specification of ~~specifying~~ which particular aspect of a descriptor is addressed~~, and then to focus on a sub-field of the keyword~~. For example the association of the descriptor *hepatitis* with the qualifier *diagnosis* (noted *hepatitis/diagnosis*) restricts the *hepatitis* to its *diagnosis* aspect.

The heterogeneity of Internet health resources and the great specificity of MeSH descriptors (which make it difficult to refer broadly to a medical specialty), led the CISMeF team to enhance the MeSH thesaurus with the introduction of two new concepts, respectively resource types and metaterms.

CISMeF resource types are a generalization of the publication types of Medline. We have added types which are specific of the health resources available on the Internet, such as association, patient information, community networks. The controlled list of resource types (n=145) is available at the following URL: http://www.chu-rouen.fr/documed/typeeng.html. As with descriptors, qualifiers and publication types, resource types are organized into hierarchies. A resource type describes the nature of the resource and MeSH descriptor/qualifier pairs describe the subject of the resource. For example, in the case of a clinical guideline about carbon monoxide intoxication, 'carbon monoxide poisoning' is the MeSH descriptor and 'clinical guidelines' is the resource type.

A metaterm is generally a medical specialty or a biological science (e.g., cardiology or bacteriology) selected by the CISMeF chief librarian. For each metaterm (N=69), one semantic link was created with one or more MeSH descriptors, qualifiers and resource types. For example, the metaterm *psychiatry* is associated with the MeSH descriptors *psychiatry* and *psychiatric hospital* that belongs to a completely different tree structure within the MeSH and also with the CISMeF resource type *mental health dispensary*.

In fact, ~~the idea of creating~~ meta-terms ~~came up~~ have been created to optimize information

https://doi.org/10.23106/dcmi.952107865

retrieval in CISMeF and to ~~cope with~~ overcome the relatively restrictive nature of MeSH descriptors. For instance, the queries 'guidelines in cardiology' and 'databases in psychiatry' where *cardiology* and *psychiatry* are only MeSH descriptors get few or no answers. Introducing *cardiology* and *psychiatry* as meta-terms is an efficient strategy to get more results because instead of exploding one single MeSH tree (e.g. *psychiatry* as a MeSH descriptor), using metaterms results in an automatic expansion of the queries by exploding other related MeSH or CISMeF trees as well as the current tree. (e.g. *psychiatric hospital* as a MeSH descriptor or *mental health* dispensary as a resource type will be exploded in the case of the *psychiatry* query). The list of metaterms is available at the following URL: http://www.chu-rouen.fr/ssf/santspeeng.html.

2.4. Metadata and the Doc'CISMeF search engine

The navigation through the terminology, thanks to alphabetical and thematic indices, allows the user to know the terms that represent the concepts used in the domain and also their positions in~~to~~ the different hierarchies. Each term has its own Web page, and a set of links enables the user to retrieve, by preformatted queries, all the resources that are related to this term. He can also restrict the search according to his/her profile: resources intended ~~to the~~ for health professionals, ~~to the~~ for students,~~or to the~~ for patients and for the general public, respectively with the following resources types: “guidelines”, “education”, and “patient”. The CISMeF team also developed a MeSH navigation (in French) very similar to the MeSH navigation (in English) developed by the PubMed Web site (URL: http://pubmed.gov). This MeSH navigation is available from any MeSH term page in the CISMeF Web site. Then, the end-user may learn the various relationships existing between MeSH terms; e.g. "*abetaliproteinemia*" belongs to the following hierarchies: "*hemic and lymphatic diseases*", "*congenital, hereditary, and neonatal diseases and abnormalities*", "*nutritional and metabolic diseases*", and "*nervous system diseases*". Medical students from the Rouen Medical School are delighted to find such relationships among MeSH terms… in French.
The other and main important utility of the terminology is its exploitation by the Doc’CISMeF search engine [3]. Different search modes are available. *“Simple search”* is done via an interface in which the user can type queries in natural language (French or English, with or without accents, in capital letters or not).
“Advanced search” uses frames and drop-down lists, provides search against all CISMeF specific metadata elements (e.g. keywords, titles, year) and the use of Boolean logical operators (and, or, not). The advanced search is mainly intended for medical librarians and adequate training is provided. *“Logical search”* is done with Boolean operators and a specific query language with particular characters. The Logical search is mainly intended for medical librarians. With the logical search, every metadata from the CISMeF set is searchable (URL: http://doccismef.chu-rouen.fr/aides/aidedcacronyme.html).
~~Today~~Currently the *simple search* is based on ~~the~~ relationships between concepts. If the query (a word or an expression) can be matched with an existing concept of the CISMeF terminology which “encapsulates” the MeSH thesaurus (metaterms, descriptor, qualifier, resource type), then the result of the query is the union of the resources that are instances of the concept, and the resources that are instances of the concept it subsumes, directly or indirectly, in all the hierarchies it belongs to, as shown in the following equation:

$$\bigcup_{\mathbf{i}=1}^{4} \exp(x)$$

where i is the level of the CISMeF terminology, x is the concept and exp is the explode function. For example, if the end-user enters the term 'virology' in English or 'virologie' in

https://doi.org/10.23106/dcmi.952107865

French, the search will be performed on the exploded metaterm 'virology', on the exploded MeSH term 'virology ' and on the exploded qualifier 'virology'. If the query cannot be matched to a concept of the terminology, the search is done over the
Metadata element 'Title' and if there is still no answers over all the fields of the CISMeF metadata set.
More recently, the CISMeF team has introduced a more complex algorithm to optimize information retrieval specifically to improve recall. The principle of the algorithm is the following: given a query Q (composed of n words $q_1 \ldots q_n$) that returns no answer because it doesn't match any term of the terminology, our objective is to find the best matching with the terminology. To do that, the query Q is segmented into $q_1 \ldots q_n$ words. The stop words (e.g. *the, before, with…*) are eliminated and a transformed query Q' is obtained. For each word q_i, the algorithm tests whether q_i belongs to the terminology. If so, the final Boolean query is: q_i[CISMeF terminology word]. If not, a Boolean query is: q_i[Title word]. If there is no answer, the final Boolean query is: q_i[all fields word]. An "*all fields*" search is performed over all the fields of the metadata (including *title, abstract, …etc.)*. All these modifications of the original query Q ~~treatments~~are performed automatically without the intervention of the user. If there is still no answer (the worst case), a full-text search is carried out in the CISMeF corpus.
Doc'CISMeF is interoperable with the Medline bibliographic database via the PubMed Web site: the CISMeF query is automatically transformed into the PubMed syntax. But as said before it is not an optimized solution and this kind of search requires a good knowledge of the medical domain, which is not obvious for any user. This algorithm is under evaluation to measure whether the returned answers, if there are some answers, correspond ~~really~~ to the initial query without noise.

2.5. CISMeF ontology: a step to a Health Semantic Web

A PhD candidate (LS) has developed the following KnowQuE (Knowledge-based Query Expansion) prototype system [23], which includes:

- ~~1.~~ a morphological knowledge base in cooperation with Zweigenbaum and Grabar, which will benefit from the UMLF consortium in charge of developing the French Specialist Lexicon [24], e.g. the query *asthmatic children* will be derived into *asthma AND child*;
- ~~2.~~ a knowledge base of association rules extracted using the data mining knowledge discovery process, e.g. breast cancer/diagnostic ⇒ mammography or hepatitis/prevention and control ⇒ hepatitis vaccines [25] ;
- ~~3.~~ a formalized CISMeF terminology using the OWL [26] language to benefit of the advantages of its powerful reasoning mechanisms [27].

3. Discussion

Several main tools could be targeted for the retrieval of health information on the Internet in ascending order: *level 1*: search engine, general or more specialised searches, such as MedHunt-Ch (URL: http://www.hon.ch/); *level 2*: catalogue and index without thesaurus, such as MedWebPlus-Us (URL: http://www.medwebplus.com/) and HealthWeb -Us (URL: http://healthweb.org/); *level 3*: catalogue and index with thesaurus, such as the Unified Medical Language System (UMLS) metathesaurus and MeSH thesaurus. The latter thesaurus is used in the following Health catalogues: DDRT (Diseases, Disorders and Related Topics) from the Karolinska Institute Library, Sweden (URL: http://www.mic.ki.se/Diseases/), CliniWeb [11] (URL: http://www.ohsu.edu/cliniweb/), Oregon Health Sciences University-USA, OMNI (Organizing Medical Networked

https://doi.org/10.23106/dcmi.952107865

Information-UK) (URL: http://omni.ac.uk/) [12] and HON (Health on the Net-Ch) from Switzerland [13]; *level 4*: catalogue and index with thesaurus, metadata, and description of sites. To our knowledge, CISMeF and Healthinsite-Au (URL: http://www.healthinsite.gov.au/) [14] have now reached level 4.

CISMeF uses DCMES differently according to the "browse" or "search" strategy chosen by the end-user. The choice of the Dublin Core was prompted by its institutional origin and its notoriety in the academic world. Several other health sites are now successfully using the Dublin Core, including the NLM (see a comprehensive list of health sites using DCMES at the following http://www.chu-rouen.fr/documed/dc.html).

The use of metadata is one main criterion in accurately assessing the quality of health information on the Internet [4]. In order to use metadata, it is necessary to properly structure information. The quality of metadata description may reflect the quality of online information. The following search on Medline using Pubmed shows that metadata is a new field of research: using the following request “metadata or meta data” in all the fields of the Medline database (2004-03-29) because metadata is not (yet) a MeSH descriptor, we found 119 references mostly published in the last four years, such as the reference which proposed specific metadata to describe video resources [15]. Metadata allow to structure information in the same way throughout several databases. Therefore, it should be easier to use these databases based on a common metadata set.

This metadata element set is useful for cross-searching distributed and heterogeneous subject gateways and for the creation of meta-catalogs or meta-gateways, such as Renardus [16]. The aim of the EU-funded Renardus project (URL: http://www.renardus.org) is to provide users with integrated access through a single interface to high-quality Internet resources permitting to search and browse records from existing distributed subject gateways across Europe.

This concept of meta-gateway could be applied in the medical field with the creation of a health meta-gateway including the following health gateways sharing the same thesaurus (MeSH): CISMeF, CliniWeb, DDRT, HON, MedWebPlus, and OMNI. These gateways should share the same metadata element set. Dublin Core and HIDDEL could be the minimum common element set as the MedCIRCLE project expects formal standardization of this vocabulary in collaboration with standardization organizations and committees (TC251/CEN/ISO). HIDDEL could play a decisive role in demonstrating and ensuring interoperability of rating services and will enable harvesting and dissemination of third-party ratings. In addition to CISMeF, Mallet et al. [17] and Boulos et al. [18] previously proposed a specific health metadata element set. It is essential to find a common health metadata set used at least by the main health gateways to become interoperable. Fortunately, the CEN/ TC251 (European Standardization of Health Informatics) has developed a current project "Metaknow - Metadata for medical knowledge resources" with the aim of establishing a small set of medically relevant metadata items suitable to apply to medical knowledge (URL: http://www.centc251.org/WGII/N-01/WGII-N00-05.pdf).

To complement the interdisciplinary nature of this work, our approach will be to reach out to other health professionals that have terminologies already in place. Experts in the field of nursing language should be included among the metadata creators. The two CISMeF specific metadata elements for EBM resources should also be extended to those recently proposed by Sakai [19]. In contrast, even using the LOM Educational subset alone was time consuming (doubling the time to describe and index a teaching resource). Instead of using the entire LOM metadata set, as the DC.Education subset is not yet finalized, the FMVU consortium planned in 2004 to use the Sharable Content Object Reference Model (SCORM v2), that enables Web-based learning management systems (LMS) to find, import, share, reuse, and export learning content in a standardized way.

https://doi.org/10.23106/dcmi.952107865

The existing UMVF XML based exchange process will be recasted to use SCO (Sharable Content Object) providing : content (resources), properties (metadata), and internal navigation. The LMS deals with SCOs and inter-SCOs navigation, using an application program interface (API).

Metadata also allows to coherently structure any institutional Web site as shown by Davenport et coll. [20] with the National Institute of Environmental Health Sciences Web site. Information professionals such as librarians have a main role in that task including training. These professionals should be part of every editorial board of institutional and academic Web sites.

Architectural characteristics and technical tools supporting implementation of the Dublin Core metadata standards should be considered, with the objective of contributing to the global development of specifications and identify the crucial elements that need to be in place to support the deployment of the metadata standards in a way that semantic relationships can be expressed and implemented in a machine-readable way, thereby supporting the vision of the Semantic Web by providing practical elements for its implementation, i.e. as proposed in health by HealthCyberMap (URL: http://healthcybermap.semanticweb.org/) using UMLS [16] or CISMeF using a semi-formal ontology based on MeSH [27].

4. Conclusion

To help healthcare professionals and health consumers to more easily locate high-quality health information on the Internet, catalogues must use standard tools especially metadata to describe, index and qualify Internet health resources. Machine-readable metadata builds a Semantic Web that will be more useful for end-users than the current Web and will require a concerted use of metadata.

5. References

[1] Koch T. Quality-controlled subject gateways: definitions, typologies, empirical overview. *Online Information Review* 2000: 24 (1) piii: 24-34.

[2] Darmoni SJ, Leroy JP, Baudic F, Douyère M, Piot J and Thirion B. CISMeF: a structured Health resource guide. *Methods Inf Med* 2000: 39 (1) piii: 30-5

[3] Darmoni SJ, Thirion B, Leroy JP, Douyere M, Lacoste B, Godard C, Rigolle I, Brisou M, Videau S, Goupy E, Piot J, Quere M, Ouazir S and Abdulrab H. A search tool based on 'encapsulated' MeSH thesaurus to retrieve quality health resources on the Internet. *Med Inform Internet Med* 2001: 26 (2) piii: 165-78

[4] Centrale Santé. Net Scoring : criteria to assess the quality of Health Internet information. 19 Sep 2001 [Web document, accessed 3 March 2004]. Available from Internet: <http://www.chu-rouen.fr/netscoring>.

[5] Weibel SL and Koch T. The Dublin Core Metadata Initiative. *D-Lib Magazine* 2000. Available from Internet: <http://www.dlib.org/dlib/december00/weibel/12weibel.html>.

[6] IEEE1484 – IEEE Learning Technology Standards Committee (LTSC). Available from Internet: <http://ltsc.ieee.org/>.

[7] Eysenbach G, Yihune G, Lampe K, Cross P and Brickley D. A metadata vocabulary for self- and third-party labeling of health web-sites: Health Information Disclosure, Description and Evaluation Language (HIDDEL). *Proc AMIA Symp* 2001 piii: 169-73.

[8] Darmoni SJ, Haugh MC, Lukacs B and Boissel JP. Quality of health information about depression on internet : Level of evidence should be gold standard. *Br Med J* 2001: 322 piii: 1366.

https://doi.org/10.23106/dcmi.952107865

[9] Darmoni SJ, Thirion B, Leroy JP and Douyere M. The use of Dublin Core metadata in a structured health resource guide on the Internet. *Bull Med Libr Assoc,* 2001: 89 (3) piii: 297-301.

[10] LeBeux P, Le Duff F, Fresnel A, Berland Y, Beuscart R, Burgun A, Brunetaud JM, Chatellier G, Darmoni SJ, Duvauferrier R, Fieschi M, Gillois P, Guille P, Kohler F, Pagonis D, Pouliquen B, Soula G, Weber J. The French Virtual Medical University. In: Proceedings of MIE 2000, Sixteenth International Congress of the European Federation for Medical Informatics, Hanover, Germany Stud Health Technol Inform. 2000;77:554-62.

[11] Dahamna B, Darmoni SJ, Roth-Berghofer TR, Köhler C, Mayer MA, Noelle G, Eysenbach G. Trust heritage in a Quality-Controlled Health Gateway. In: Medinfo 2004, Eleventh World Congress on Health and Medical Informatics, 2004 (in press).

[12] Berners-Lee T, Hendler J and Ora L. The Semantic Web. Scientific American. May 2001. Available from Internet: <http://www.scientificamerican.com/article.cfm?articleID=00048144-10D2-1C70-84A9809EC588EF21&catID=2>.

[13] Hersh WR, Brown KE, Donohoe LC, Campbell EM and Horacek AE. CliniWeb: managing clinical information on the World Wide Web, *JAMIA* 1996: 3 (4) piii: 273-80.

[14] Norman F. Organising Medical Networks' information (OMNI). *Med Inf* 1998: 98 (23) piii: 43-51.

[15] Boyer C, Baujard O, Baujard V, Aurel S, Selby M and Appel RD. Health On the Net automated database of Health and medical information, *Int J Med Inf* 1997: 47 (1-2) piii: 27-9.

[16] Deacon P, Smith JB and Tow S. Using metadata to create navigation paths in the HealthInsite Internet gateway. *Health Info Libr J.* 2001: 18 (1) piii: 20-9.

[17] Shotton DM, Rodriguez A, Guil N and Trelles O. A metadata classification schema for semantic content analysis of videos. *J Micros* 2002: 205(Pt 1) piii: 33-42.

[18] Neuroth H and Koch T. Metadata Mapping and Application Profiles. Approaches to providing the Cross-searching of Heterogeneous Resources in the EU Project Renardus. DC-2001 (International Conference on Dublin Core and Metadata Applications. Tokyo). October 2001, pp. 122-29. Available from Internet: <http://www.nii.ac.jp/dc2001/proceedings/abst-21.html>.

[19] Malet G, Munoz F, Appleyard R and Hersh W. A model for enhancing Internet medical document retrieval with "medical core metadata". *J Am Med Inform Assoc* 1999: 6 (2) piii: 163-72.

[20] Boulos M, Roudsari A and Carson E. Towards a semantic medical Web: HealthCyberMap's tool for building an RDF metadata base of health information resources based on the Qualified Dublin Core Metadata Set. *Med Sci Monit,* 2002: 8 (7) piii: 24-36.

[21] Sakai Y. Metadata for Evidence Based Medicine Resources. DC-2001 (International Conference on Dublin Core and Metadata Applications. Tokyo). October 2001, pp. 81-5. Available from Internet: <http://www.nii.ac.jp/dc2001/proceedings/abst-12.html>.

[22] Davenport Robertson W, Leadem EM, Dube J and Greenberg J. Design and Implementation of the National Institute of Environmental Health Sciences Dublin Core Metadata Schema. DC-2001 (International Conference on Dublin Core and Metadata Applications. Tokyo). October 2001, pp. 193-99. Available from Internet: <http://www.nii.ac.jp/dc2001/proceedings/abst-29.html>.

https://doi.org/10.23106/dcmi.952107865

[23] Soualmia, L.F., Barry, C., and Darmoni, S.J. Knowledge-Based Query Expansion over a Medical Terminology Oriented Ontology. In: Springer's Lecture Notes of Computer Science, M. Dojat, E. Keravnou and P. Barahona (Eds.). Artificial Intelligence in Medicine. Proceedings of the 9th Conference on Artificial Intelligence in Medicine in Europe, AIME 2003, pp209-213.

[24] Zweigenbaum, P., Baud, R., Burgun, A., Namer. F., Jarrousse, E., Grabar, N., Ruch, P., Le Duff, F., Thirion, B., and Darmoni, S.J. Towards a Unified Medical Lexicon for French. In: Proceedings of MIE 2003, Eighteenth International Congress of the European Federation for Medical Informatics, G. Surjan, R. Engelbrecht, P. McNair Eds, IOS Press Publisher, Stud Health Technol Inform. 2003;95:415-420.

[25] Soualmia, L.F., Darmoni, S.J. Combining Knowledge-based Methods to Refine and Expand Queries in Medicine. FQAS 2004, Flexible Query Answering Systems, to appear in Lecture Notes in Artificial Intelligence.

[26] Horrocks I., Patel-Schneider P.F., and van Harmelen F. From SHIQ and RDF to OWL: The making of a web ontology language. Journal of Web Semantics, 1(1):7-26, 2003.

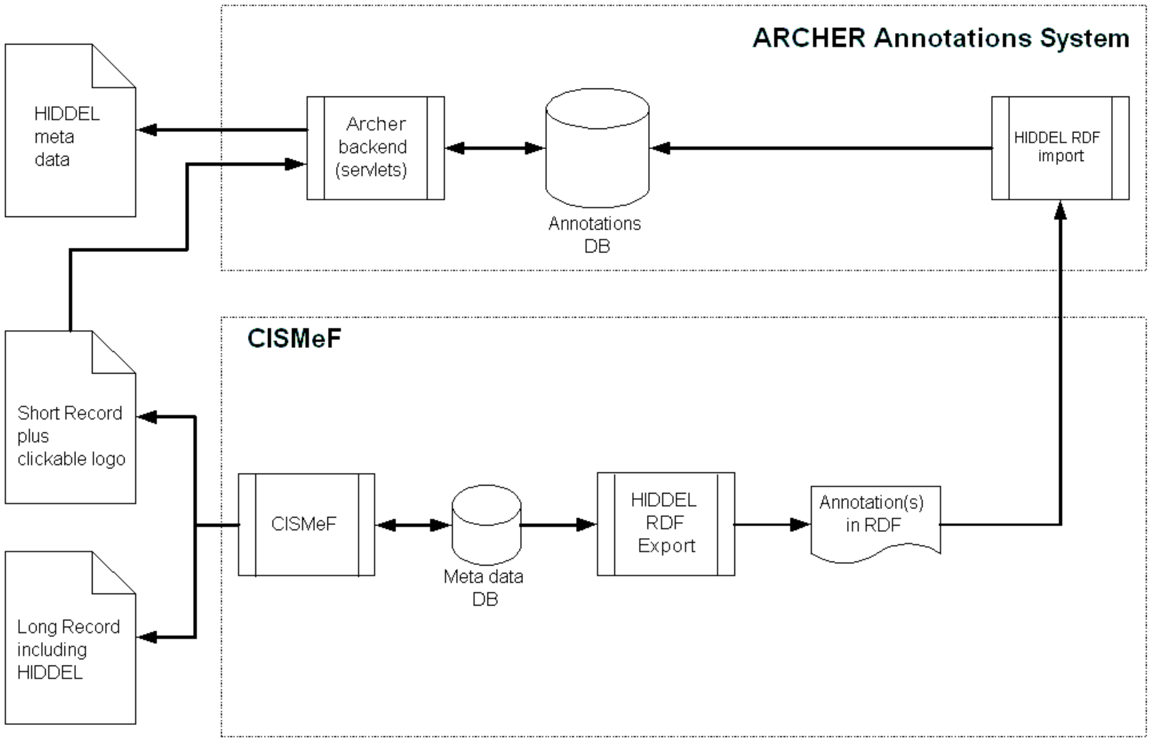
[27] Soualmia, L.F., Golbreich, C., Darmoni, S.J. Representing the MeSH in OWL: Towards a Semi-Automatic Migration. KR-MED 2004, International Workshop on Formal Biomedical Knowledge Representation, in press.

6. Address for correspondence

Darmoni Stéfan J., Public Health Department, 1 rue de Germont 76031 Rouen Cedex, France
Tel: +33.232.88.88.29; Fax: +33.232.88.88.32
E-mail: Stefan.Darmoni@chu-rouen.fr

https://doi.org/10.23106/dcmi.952107865

Figure 1: Interoperability between CISMeF and Archer (MedCIRCLE consortium)



https://doi.org/10.23106/dcmi.952107865