



2nd Year Project Report

Distribution: Public

MedIEQ

Quality Labeling of Medical Web content using Multilingual
Information Extraction

National Centre for Scientific Research "Demokritos"
Teknillinen Korkeakoulu – Helsinki University of Technology
Universidad Nacional de Educacion a Distancia
Col.legi Oficial de Metges de Barcelona
Zentralstelle der deutschen Ärzteschaft zur Qualitätssicherung in der Medizin
Vysoka Skola Ekonomicka V Praze
I-Sieve Technologies Ltd

2005107

February 2008

Project ref. no.	2005107
Project acronym	MedIEQ
Project full title	<i>Quality Labeling of Medical Web content using Multilingual Information Extraction</i>

Security (distribution level)	<i>Public</i>
Contractual date of delivery	<i>31 December 2007</i>
Actual date of delivery	<i>15 February 2008</i>
Deliverable number	<i>D3.2</i>
Deliverable name	<i>2nd Year Project Report</i>
Type	<i>Report</i>
Status & version	<i>Final</i>
Number of pages	<i>17</i>
WP contributing to the deliverable	<i>WP1</i>
WP / Task responsible	<i>NCSR</i>
Other contributors	<i>TKK, UNED, WMA, AQUMED, UEP, I-sieve</i>
Author(s)	<i>V. Karkaletsis, P. Karampiperis, K. Stamatakis (NCSR)</i>
EC Project Officers	<i>Artur Furtado</i>
Keywords	<i>Project results, 2nd year</i>
Abstract (for dissemination)	<i>This document describes the work carried out and the results obtained in the 2nd year of the project, as well as the actions foreseen for the 3rd year of the project.</i>

Table of contents

1. Introduction.....	4
2. System Integration	6
3. Quality labelling of medical web sites.....	7
4. Web Content Collection.....	9
5. Information Extraction.....	11
6. Lexical and Semantic Resources	12
7. Project Evaluation.....	13
8. Project Dissemination	14
9. Concluding Remarks.....	16

1. Introduction

The number of health information web sites is increasing day by day. It is known that the quality of these web sites, published by various providers, is very variable and difficult to assess. At the same time, it is increasing the necessity to implement control measures that give the consumers adequate guarantee that the health web sites they are visiting meet a minimum level of quality standards and that the professionals offering the information on the web site are responsible for its contents.

Different organizations around the world are currently working on establishing labelling criteria for the accreditation of health-related web content. The European Council supported an initiative within eEurope 2002 to develop a core set of “Quality Criteria for Health Related Websites”. However, self-adherence to such criteria is nothing more than a claim with little enforceability. It is necessary to establish rating mechanisms which exploit such labelling criteria.

There are two major mechanisms in medical quality labelling:

- Filtering portals: the web resources are classified according to predetermined criteria and organized in groups in order to facilitate a quick access to quality reviewed information. Examples of this mechanism are: “Catalog and Index of French-speaking Medical Sites” (CISMEF), “Organising Medical Networked Information - The UK Gateway to reliable health information” (OMNI), “Agency for Quality in Medicine” (AQUMED).
- Third party accreditation: an organization evaluates actively the quality of the web site according to a set of criteria. Compliance with those criteria is showed with a logo or trust mark on the homepage. HON Code of the Health on the Net Foundation, URAC Accreditation Program, Web Médica Acreditada are the most lively known quality seals.

The main problem that these mechanisms face is the need for a continuous review and control of the accredited or classified web sites that means a huge amount of human effort. WMA, as third-party accreditation system, for instance, periodically reviews manually the accredited web sites to renew the quality label. On the other hand, in AQUMED, as a classification system (filtering portal), web site directories are periodically updated due to the addition of new web resources and changes in the characterization of the already visited ones.

In order for the health-related content labelling mechanisms to be more effective, they must be equipped with semantic web technologies that enable the creation of machine-processable labels as well as the automation of the labelling process. Such technologies may involve web crawling techniques that allow the retrieval of new unlabelled web sites, their characterization and addition in a health portal, and/or information extraction techniques that allow the continuous monitoring of labelled web resources alerting the labelling agency in case some changes occur against the labelling criteria.

Based upon state-of-the-art technology in the areas of semantic web, content analysis and quality labelling, the EC-funded project MedIEQ (DG DANCO 2005107) aims to support labelling experts in their work of labelling health-related web content. MedIEQ will deliver tools that crawl the web to locate unlabelled health web resources in seven different European languages in order to examine their content using a set of machine readable quality criteria. MedIEQ tools will monitor already labelled health web resources alerting labelling experts in case the resources’ content is updated against the quality criteria.

MedIEQ aims to advance current medical quality labelling technology capitalizing on the results of previous work on quality labelling and content analysis. The implementation of this objective will be based on the realisation of the following more specific objectives:

- Develop a schema for the quality labelling of health-related web content and provide the tools supporting the creation, maintenance and access of labelling data according to this schema;
- Specify a methodology for the content analysis of health-related web resources according to the MedIEQ schema and develop the tools that will implement it;
- Specify a methodology and develop the tools for the creation and maintenance of the multilingual resources that will support content analysis in health-related web resources;
- Develop a prototype labelling system and demonstrate it in 7 different languages and two labelling applications (third party accreditation, classification).

Sections 2-8 presents the work carried out during the 2nd year of the project, describing the project results (technical or not) towards the achievement of the above objectives. Finally, section 9 concludes presenting some remarks based on the results achieved so far.

2. System Integration

The overall objective of System Integration is to guarantee the seamless functionality of the overall system and thus provide the ability to integrate, demonstrate and evaluate the technology developed pertaining to:

- the automation of the labeling process, by bringing together the content collection and information extraction tools;
- the interfacing of content collection and information extraction tools, so as to achieve transparency to the user experience;
- the maintenance and manipulation for the lexical and semantic resources used by the extraction and collection tools.

During the reporting period, the main objective of System Integration was to implement MedIEQ 1st Integrated Prototype, integrating components/toolkits from the technical Work Packages.

During the reporting period, MedIEQ 1st Integrated Prototype, named AQUA (that is, Assisting Quality Assessment), was implemented. The 1st version of AQUA supports:

- both case studies for third party accreditation and classification
- two languages, namely, English and Spanish
- a basic sub-set of the labeling criteria.

More specifically, AQUA consists of the following components / toolkits:

1. The *Label Management Toolkit (LAM)*, which manages (generates, validates, modifies, compares) machine readable labels which conform with the RDF-CL model and the MedIEQ vocabulary;
2. The *Web Content Collection Toolkit (WCC)*, which identifies, classifies and collects on-line content relative to a number of machine readable quality criteria (according to the proposed vocabulary in the MedIEQ schema);
3. The *Information Extraction Toolkit (IET)*, which analyses the web content collected and extracts attributes for MedIEQ compatible content labels;
4. The *Multilingual Resources Management Toolkit (MRM)*, which gives access to health-related multilingual resources; input from such resources is needed in specific parts of both the WCC and IET toolkits;
5. The *Monitor-Update-Alert Toolkit (MUA)*, which handles a few auxiliary but important jobs, like the configuration of monitoring tasks, the MedIEQ database's entries updates, the alerts to labelling experts when important differences occur during monitoring existing quality labels.

The final version of AQUA is due to month 32 (end of August 2008). This will integrate the final versions of all sub-systems, will support the full set of the labeling criteria and will cover seven European languages (namely, Spanish, Catalan, German, English, Greek, Czech, and Finnish), taking into account the results of the evaluation of the 1st version of the prototype.

3. Quality labelling of medical web sites

The overall objectives of Quality Labelling of Medical Web Sites can be summarized in the following actions:

- extend the general vocabulary of the EC-funded project QUATRO in order to create a schema for medical content labeling based on the labeling criteria established in previous EC-funded projects MedCERTAIN, MedCIRCLE and WRAPIN.
- specify a methodology for exploiting the RDF labeling data by the MedIEQ labeling platform in two applications that correspond to the two examined labeling mechanisms (third party rating and classification).

During the reporting period, the main objective of Quality Labelling of Medical Web Sites was to finalize the RDF schema for medical labels, as well as, to localize and improve the functionalities of the 1st version of the Label Management Toolkit (LAM) based on the immediate feedback received from MedIEQ users (namely, AQuMED and WMA).

During the reporting period, the final version of the RDF schema for medical labels was prepared. The final version of the RDF schema for medical labels contains the description of the full set of labeling criteria, selected based on the comparison and analysis of the criteria currently used by the participating labelling agencies WMA and AQuMed, the recommendations of the European Union from “*eEurope 2002: Quality Criteria for Health related Websites*”¹ and a label agency of international reference as HON², taking also into account the relevant remarks of the advisory committee (AC) members (see the “Advisory Committee Meeting: Compiled Report”). For more details on the final set of criteria, which actually attempts to encompass the above mentioned initiatives, see Deliverable D4.2.

It must be stressed again that we are developing machine readable domain specific labels, based on the RDF-CL model, re-using in our vocabulary terms (criteria) already proposed by others. Our aim is NOT to suggest this vocabulary as “the only one to use” but to show instead the value of machine readable labels built on the basis of a well accepted model. Other labelling authorities, outside MedIEQ consortium, could use project’s technology to create machine readable labels and for those of the terms in MedIEQ vocabulary appearing also in their vocabularies, exploiting thus MedIEQ technology to support the process of maintaining their labels.

Building upon the experience from the EC-funded project QUATRO³, in which two MedIEQ partners were involved (NCSR, WMA), the results of the W3C Web Content Labelling Incubator⁴ and the outcome of its successor on-going W3C Protocol for Web Description Resources Working Group (POWDER⁵) (NCSR is actively involved in both initiatives), we decided to use the RDF language to create the vocabulary for the medical labels. QUATRO’s RDF-CL model is currently used but this will be converted to POWDER model as soon as this is finalised (expected in June 2008). Our aim was to re-use existing RDF vocabularies and create new Properties and Classes under a MedIEQ vocabulary only when our needs are not covered.

Furthermore, during the reporting period, the 1st version of the Label Management Toolkit (LAM) was localized in Spanish and its functionalities were improved based on the

¹ http://europa.eu.int/information_society/europe/ehealth/doc/communication_acte_en_fin.pdf

² <http://www.hon.ch/>

³ <http://www.quatro-project.org/>

⁴ <http://www.w3.org/2005/Incubator/wcl/>

⁵ <http://www.w3.org/2007/powder/>

immediate feedback received from MedIEQ users (namely, AQuMED and WMA). More specifically, LAM was extended to provide:

- a user friendlier web interface, which makes easier for users to create, update and delete labels, to monitor changes in labels and properly import and export labels to the system;
- improved editing functionalities, which make easier for users to make refined descriptions for the desired web resources;
- improved label comparison functionalities, which make easier for users to track the evolution of labels through the label versioning system already implemented in LAM;
- access per organization: unified access to the created labels for all the members of the same organization is now enabled.

Based on the evaluation results of the 1st MedIEQ Integrated Prototype, the final version of the Label Management Toolkit (LAM) will be updated and integrated in the final version of MedIEQ Integrated Prototype.

4. Web Content Collection

The overall objectives of web content collection can be summarized in the following actions:

- specify the methodology and architecture for web content collection. This includes two main components: focused crawling to locate interesting web sites and site spidering to locate interesting web pages according to the labeling criteria inside these sites;
- develop the toolkit for content collection that will be integrated in the prototype system.

During the reporting period, the main objective was to develop the final version of the Web Content Collection toolkit. The final release of the toolkit covers both case studies (third party accreditation and classification). It also covers all criteria and all project languages. However, to support this, it is necessary to have trained the corresponding modules. The training process started for the initial set of criteria for Greek, Finnish and Czech, after having collected the necessary corpora. Concerning the additional criteria, this has started for English and Spanish.

During the reporting period, the final version of the Web Content Collection Toolkit (WCC) was prepared. The Web Content Collection methodology involves the following key steps:

- *Crawling*: searching the web to identify on-line resources with health related content. Such a search is performed by the Crawler tool exploiting existing search engines (both general purpose and specialized to the health domain) and web directories.
- *Spidering*: Health-related web resources either known or identified by the Crawler, are explored. Every visited resource's content is classified according to the labelling criteria issued and fit content is locally stored.

The final version of the Web Content Collection (WCC) toolkit includes the final versions of the following components:

1. *Corpus formation tool* (CFT), for collecting corpus necessary for the training of the content classifiers. The final CFT release has been enhanced with classification capabilities (see point 2, below).
2. *Trained module generator* and *Content classification component* (TMG/CCC), for training the classifiers using the collected corpus (TMG) and also for classifying content (CCC). TMG/CCC is integrated within the Spider, for classifying content while traversing a site, within the Crawler, for making crawling more focused, and the CFT, for helping in the identification of relevant content while forming a corpus.
3. *Crawler*, for locating unlabeled web resources. The final version of the Crawler gives the possibility to the user to exploit a content classification mechanism, which makes crawling even more focused. A classifier is trained taking into account manual classification (into pos & neg samples), by the user, of the resources returned by the tool.
4. *Spider*, for navigating in a web site (already labelled or not) in order to locate interesting content according to the labelling criteria. The Spider which is included in the final WCC version has been extended to cover more languages, GR, CZ and FI, for the initial set of criteria. Concerning the rest of criteria and languages, the corpus formation process, necessary for training the classification/extraction modules, has started.

5. *Automatic Ontological Concepts Extraction Tool* (POKA) is a tool for automatic extraction of ontological concepts in text documents. In the MedIEQ framework Poka is used to find relations between health web content and medical vocabularies such as MeSH to facilitate categorization of web content. The Poka system is used as a component of the web spidering tool.

Based on the evaluation results of the 1st MedIEQ Integrated Prototype, the final version of the content collection toolkit may be refined/updated, taking into account the evaluation report, and will be delivered for integration in the final version of MedIEQ Integrated Prototype.

5. Information Extraction

The overall objectives of Information Extraction can be summarized in the following actions:

- specify the methodology and architecture for information extraction from web pages.
- develop the two versions of the toolkit for information extraction that will be integrated in the prototype system.

During the reporting period, the 1st version of the information extraction toolkit (IET) was developed. IET provides a uniform interface to multiple IE engines which are used within MedIEQ. IET also provides functionalities to define, run and monitor the progress of extraction tasks, to export and visualize extracted information and to manage extraction data models (which describe the attributes and classes to be extracted). IET now allows pipelining multiple IE engines where each engine may exploit outputs of its predecessors. Tasks may be run in different modes including the extraction mode and training mode. IET can process documents which have already been annotated (e.g. when in training mode), and is able to read both local and on-line documents.

Two extraction engines have been integrated into the IET so far. The Ex IE engine is based on extraction ontologies and allows the combination of manually encoded extraction knowledge (mostly in the form of extraction patterns and axioms) and of trainable machine learning classifiers. The second IE engine is based on Conditional Random Fields, a statistical technique that has been recently successfully applied to various IE problems.

New features introduced into the Ex IE engine during this period include:

- integration with trainable classifiers: each extraction ontology may now link to multiple classifiers which may be used to aid the extraction of all or selected extractable attributes;
- feature induction module has been added;
- extensions to the extraction ontology language: pattern syntax allows references to formatting elements, attribute inheritance, support for resolving simple co-references, attribute value transformations (via scripting), support for probabilistic axioms at both attribute and class level;
- improved instance parsing algorithm and scoring method.

Based on the evaluation results of the 1st MedIEQ Integrated Prototype, the final version of the information extraction toolkit will be developed, to be delivered for integration in the final version of MedIEQ Integrated Prototype (that is due to month 32 - end of August 2008).

6. Lexical and Semantic Resources

The overall objectives of Lexical and Semantic Resources can be summarized in the following actions:

- collect and examine existing resources, both general-purpose ones such as Wordnets for the languages that will be examined in MedIEQ as well as medical ontologies and thesauri (e.g. MeSH thesaurus);
- specify the methodology for the semi-automatic creation and maintenance of medical and general-purpose resources, as well as for their use by the various content collection and extraction tools;
- develop a toolkit for the management of medical resources and general-purpose lexicons.

During the reporting period, the final version of the Multilingual Resources Management (MRM) Toolkit was developed, after the development and the technical evaluation of the 1st version of the toolkit. MRM Toolkit is a set of tools/components for the management of linguistic resources in different languages.

The toolkit allows to import linguistic and semantic resources in pre-defined supported formats into a repository and to access them in order to manipulate their contents for the project needs. Once a given resource is imported into the repository, it is possible to browse across its contents and to create new resources by selecting and merging data.

The key system behind the MRM Toolkit is the Unified Medical Language System (UMLS), and more specifically one of its components: the UMLS Metathesaurus. UMLS is an effective way of obtaining and updating the linguistic resources needed in MedIEQ, since it is distributed as a unique package. In addition, it already contains the standard version of MeSH controlled vocabulary in most of the languages involved in the MedIEQ project (namely, English, Spanish, German, Finnish and Czech). Regarding compatibility among resources, UMLS provides the essential mapping between different resources, e.g. MeSH↔SNOMED-CT, MeSH↔ICD. In order to access, manipulate and integrate all the resources available in UMLS within the MRM Toolkit, we use the UMLSKS services, specifically the Java API provided for developing purposes. Thus, the MRM Toolkit works as an extra-layer on top of the UMLS datasets and tools in order to allow communication between the resources repositories and the other components.

The MRM Toolkit has been integrated within the MedIEQ Integrated Prototype (AQUA). Based on the evaluation results of the 1st MedIEQ Integrated Prototype, the final version of the Multilingual Resources Management (MRM) Toolkit may be refined/updated, to be delivered for integration in the final version of MedIEQ Integrated Prototype.

7. Project Evaluation

The overall objectives of Project Evaluation can be summarized in the following actions:

- Specification of the evaluation strategy. Different evaluation strategies for the two labelling mechanisms will be developed. The evaluation measures will concern issues such as: labelling effort, processing time for the extraction from large collections of medical web content, effort required to customize the system, etc.
- Evaluation of the 1st and final prototypes. In addition to the evaluation at the level of individual components, systematic evaluations will be carried out to measure the overall system performance.

During the reporting period, the evaluation of the 1st AQUA prototype was conducted. The evaluation of MedIEQ at the level of its individual components (label management, web content collection, information extraction and multilingual resource management toolkits), was performed in the context of the corresponding work packages.

Since the primary goal of the evaluation of the 1st AQUA prototype was to conclude with a functional prototype that has the potential to be fully integrated within the day-to-day activities of a labelling organization, a parallel technical improvement action took place, refining given functionalities. The main objective of the extra technical improvement action was to enhance the overall system workflow, so as to better match the day-to-day practice. As a result, the scope of the evaluation itself was refined so as to provide more focused feedback towards this direction. The scope of this evaluation was the performance evaluation of AQUA on supporting the labelling process (i.e. identification of new unlabeled web resources and labelling of web resources), as well as, the usability evaluation of AQUA interface. The evaluation of the labelling performance on a real integration scenario in the usual work of a labelling organization was considered as the evaluation scope of the 2nd version of AQUA prototype.

The first prototype of AQUA supports the first set of labeling criteria, in two languages (English and Spanish). The evaluation was conducted in English by AQUMED and in Spanish by WMA.

8. Project Dissemination

The overall objective of Project Dissemination can be summarized in the following actions:

- publish MedIEQ technology encouraging as many labeling schemes as possible to participate and to engage relevant policy makers;
- encourage the practice of quality labeling agencies;
- develop partnerships for all the applications related to MedIEQ;
- contribute to international standardization bodies (W3C semantic web and content labeling initiatives);
- exploit the links of the research partners with other researchers in the area of semantic web techniques.

During the reporting period, the project consortium prepared a series of dissemination activities, including the following:

- *Conferences/Workshops*
 1. D.V. Gonzales, M.A. Mayer, A. Leis, V. Karkaletsis, K. Stamatakis, V. Metsis, P. Nasikas, M. Labský, M. Ruzicka, V. Svátek, F. López-Ostenero, V. Peinado, E.A. Cabrera, T. Honkela, M. Pöllä. “AQUA (Assisting Quality Assessment): a system based on Semantic web and information extraction technologies to support medical quality labelling agencies” , Proceedings of the 12th World Congress on the Internet in Medicine (Mednet 2007), Leipzig , Germany , October 7-10, 2007.
 2. K. Stamatakis, K. Chandrinos, V. Karkaletsis, M.A. Mayer, D.V. Gonzales, M. Labsky, E. Amigó, and M. Pöllä, “AQUA, a system assisting labelling experts assess health web resources”. Proceedings of 12th International Symposium for Health Information Management Research (iSHIMR 2007), p. 75-84, Sheffield, UK, 18-20 July, 2007.
 3. K. Stamatakis, V. Metsis, V. Karkaletsis, M. Ruzicka, V. Svátek, V. E.A. Cabrera, and M. Pöllä, “Content collection for the labeling of health-related web content”. Proceedings of the 11th Conference on Artificial Intelligence in Medicine (AIME 07), LNAI 4594, pp. 341-345, 2007. Amsterdam, 7-11 July, 2007.
 4. MA Mayer. Els continguts de Salut a Internet. Cursos d’estiu “Els Juliors 2007”: Internet i seguretat. Universitat de Barcelona, Barcelona 2-6 July, 2007.
 5. MA Mayer, A. Leis. Los metadatos y la Web Semántica al servicio de los usuarios de salud (Metadata and Semantic Web working for health consumers). XI National Internet and Information Society Conference, Málaga 14 – 16 May, 2007.
- *Book chapters*
 1. V. Karkaletsis, Stamatakis, K., Karabiperis, P., Labský, M., Růžička, M., Svátek, V., Cabrera, E. A., Pöllä, M., Mayer, M. A., Villaroel Gonzales, D., “Management of Medical Website Quality Labels via Web Mining”. Following a call for chapters it was accepted to be included in “Data Mining and Medical Knowledge Management: Cases and Applications”, P. Berka, J. Rauch, D. Abdelkader Zighed (eds), a book to be published by IGI Global Inc.

- *Informational Events / Lectures*
 1. V. Karkaletsis (NCSR), Invited talk for MedIEQ technology at the Workshop “Language Technology in Biomedicine”, organized by the Greek R&D project IATROLEXI, December 10, Athens, Greece.
 2. K. Chandrinos (I-sieve), Presentation of AQUA at the industry stakeholder meeting organized by the W3C working group “Protocol for Web Description Resources (POWDER)”, July 11, Washington DC.
 3. MA Mayer (WMA), Invited lecture in the Workshop “Advances and Innovations in information searching and retrieval”, in the Health Technology Assessment International (HTAi) Conference in Barcelona, Spain. June 17-20, 2007.
- *Workshops organisation*
 1. Submission of a proposal for the organization of a Workshop on Describing Medical Web Resources (DRMed@MIE 2008), in the context of the 21st International Congress of the European Federation for Medical Informatics (MIE-2008), Göteborg, Sweden May 25-28, 2008.
- *Preparation of Dissemination Material*
 1. New project leaflet prepared and distributed by project partners

At the beginning of the 3rd year of the project, the partners have already scheduled certain dissemination activities, for which they are currently working on, and they are exchanging views for more activities in the coming months.

9. Concluding Remarks

MedIEQ is a technology project employing semantic web technologies for the description of web resources, content analysis technologies for collecting domain-specific web resources and extracting information from them. These technologies can also be applied in other domains. What we aim in this project is to examine their applicability in a specific area with special importance that of assisting the labelling of health related web resources.

For this purpose, we are developing machine readable domain specific labels. The vocabulary used in these labels is based on the criteria that are currently used by the participating labelling agencies WMA and AQuMed, the eEurope criteria guidelines and a labelling agency of reference as HON. Our aim is not to suggest this vocabulary as “the only one to use” but to show instead the value of machine readable labels. So, the emphasis is on the technology and not on the terms (criteria) included in the vocabulary. The terms selected capture important aspects of health related content and form the case study for MedIEQ technology partners.

At the end of the project’s second year, the following new results have been achieved:

- The 1st version of the MedIEQ Integrated Prototype (AQUA) has been prepared. This prototype version supports case studies for third party accreditation and classification, two languages, namely, English and Spanish and a basic sub-set of the labeling criteria.
- The final version of the MedIEQ labelling schema has been prepared, covering the full set of the labelling criteria used in the MedIEQ machine readable labels implementing the adopted RDF-CL model.
- The final version of the Web Content Collection (WCC) toolkit has been developed and integrated within AQUA. WCC provides tools for web crawling and spidering of web sites.
- The 1st version of the Information Extraction toolkit (IET) has been developed and integrated within AQUA. IET provides tools for extracting from the textual content of web resources important information, according to the criteria specified in the MedIEQ schema.
- The final version of the Multilingual Resources Management toolkit (MRM) has been developed and integrated within AQUA. MRM provides tools for accessing, editing, and exploiting UMLS resources in the project languages.
- The evaluation of the 1st version of the MedIEQ Integrated Prototype was conducted. The scope of this evaluation was to evaluate the performance of AQUA on supporting the labelling process (identification of new unlabeled web resources, labelling of web resources) as well as to evaluate the usability of AQUA interface.

We would also like to stress the fact that MedIEQ is progressing in parallel with other relevant projects and initiatives, with which it has strong links, and which can help significantly to promote its work. MedIEQ machine readable labels are based so far on the RDF-CL model proposed by the EC-funded project QUATRO, in which two MedIEQ partners were involved. This model was studied in the context of the W3C Web Content Labelling Incubator Group (WCL) which led to the W3C working group POWDER (Protocol for Web Description Resources) aiming to produce a content labelling model as a W3C recommendation. MedIEQ will use this model, at its final stage, for creating machine readable labels. Furthermore, MedIEQ will exploit QUATRO tools to actually read the labels once created.

In general, there is a growing interest for semantic web applications and MedIEQ is such an application in a domain presenting significant interest, both from a technological and a societal point of view, due to the large number of health-related web resources and the impact of their content to the society. MedIEQ consortium aims to develop technology that will support the work of labelling authorities, increasing the number of labelled health-related web sites and improving their monitoring.