MIE 2008 Göteborg 26-28 May
eHealth beyond the horizon – get IT there
Poster abstract

Semantic Integration of Data in an Information System for Multicenter Epidemiological Studies on Cancer

Marcos MARTÍNEZ [a, 1] , José M. VÁZQUEZ [a] , M. Gloria LÓPEZ [a] ,
Francisco M. ARNAL [b] , Benito GONZÁLEZ-CONDE [b] ,
Javier PEREIRA [a] , and Alejandro PAZOS [a]
[a] *IMEDIR Center, University of A Coruña, Spain*
[b] *University Hospital Complex Juan Canalejo, A Coruña, Spain*

Keywords: Hospital IS, Knowledge management, Epidemiological research.

Introduction and Methods

The data collected during epidemiological studies are rarely reused. This is mainly due to the frequent utilization of traditional storage supports, as well as specific formats and/or terminologies which can make it difficult to integrate the gathered data with data from other information resources. Ontologies can help to solve this problem, since they provide a common specification which can be used to overcome heterogeneity in the collected data. This work presents the data integration capabilities of an Information System (IS) for the development of multicenter epidemiological studies on cancer, which has been tested during the execution of “A Pilot Study of Colorectal Cancer in Galicia, Spain”, funded by the U.S. National Cancer Institute (NCI).

The system stores the data collected during interviews to patients and their relatives in a centralized database. In order to make these data openly available, we mapped the fields in the database to terms in the NCI Thesaurus ontology. By means of a Web service, the system uses the defined mappings to answer remote queries written in the terminology of the Thesaurus. The IS also uses the Thesaurus as a reference to query remote data sources in an integrated manner.

Results and Conclusion

Reusing the data collected during epidemiological studies would provide great benefits. Information and Communication Technologies can contribute significantly to this task, thanks to the development of ISs such as the presented one, which 1) allows to remotely access the data collected by the system and 2) makes it possible to retrieve information from remote data sources in an integrated manner. We were able to successfully map a 93% of the data collected during a real epidemiological study on colorectal cancer to terms in the Thesaurus. Now, these data can be reused. Preliminary results show that the IS is also able to effectively query remote data sources that have been prepared to be accessed by means of the NCI Thesaurus terminology.

[1] Corresponding Author: marcosmartinez@udc.es