

Ontologies in Agriculture

C. ROUSSEY^{1, 2}, V. SOULIGNAC¹, J-C CHAMPOMIER¹, V. ABT¹, J-P CHANET¹.

1 Cemagref, UR TSCF, 24 avenue des Landais, BP 50085, 63172 Aubière cedex, France.

2 Université de Lyon, CNRS, Université Lyon 1, LIRIS UMR5205, F-69621, France

Abstract: Ontologies have been used for the last decades for a set of tasks, one of which is focused on achieving interoperability between heterogeneous information systems. In this paper, we present different types of ontologies and we explain how next generation of information system will benefit from the use of ontologies to resolve interoperability issues.

Keyword: ontology, knowledge representation, information system interoperability.

1 Introduction

In the last decades, the use of ontologies in information systems has become more and more popular in various research fields, such as web technologies, Information Retrieval or document management system, Natural Language Processing, database integration, etc. But each of these fields use in a different manner a specific kind of ontologies, thus the word “ontology” is not clear enough to understand exactly which computer science object is behind this word !

The article achieves two goals. First it defines what exactly ontologies are with a description of their components. Secondly, it explains how next generation of information system will benefit from the use of ontologies. All the illustrative examples of information system will be taken from project related to agricultural domain.

2 Ontology Classification

We define ontologies as new computer science tools used to design phenomenon model and entities of a domain. Ontologies are used to improve the communication process between different agents (human or computer system) by sharing and reusing information, data or knowledge.

Depending on their purpose, ontologies are composed of different components: i.e. concepts, properties, instances, logical formula. Fig. 1 presents all their possible components and their relationships. The central components of ontologies are concepts. The Fig. 1 shows that concepts can be defined in different (and complementary) ways:

- by their textual definition: For example the concept “*human*” is defined by the sentence “*an individual human being*”,
- by a set of properties: for example the concept “*person*” has the property “*name*”, “*birth date*” and “*address*”,
- by a logical formula: for example the concept “*human*” is defined by the formula “*LivingEntity* \cap *MovingEntity* \cap $\forall hasChild.Human$ ”.

A concept can also be defined by the set of instances that belong to it. For example, “*Martin Luther King*” is an instance of the concept “*person*”.

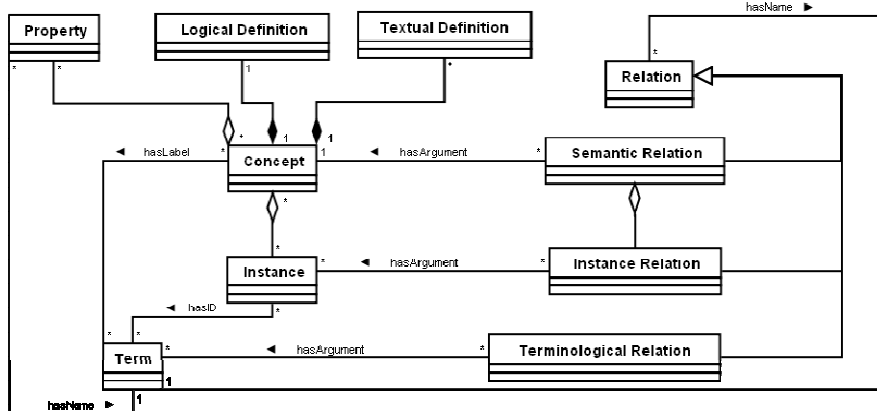


Fig. 1: UML schema representing ontology components and their relationships

Concepts, instances and properties are referenced by one or more symbols. Symbols are terms that humans can rapidly understand roughly by reading them. And finally all these ontology components are connected together thanks to relations. Semantic relations link only concepts together (for example the part of relationship indicates that a concept is composed of another one), instance relations connect only instances (for example “*Martin Luther King*” is married to “*Coretta Scott*”) and terminological relations express the relationships that terms can have (for example the term “*person*” is synonym to the term “*human being*”).

According to the usage of these components, in the following sections we present three kinds of ontologies.

2.1 Terminological ontologies:

Terminological ontologies can be glossaries, dictionaries, controlled vocabularies, taxonomies, folksonomies, thesaurii, or lexical databases. This type of ontology mainly focuses on terms and their relationships. Unfortunately, terms are ambiguous. A concept can be referenced by several terms (for example: “computer science”, “computing”, “information technology” are synonyms) and a term can reference several concepts (for example the term “bank” can be used to reference a “river bank” or a “commercial bank”). The roles of terminological ontologies are twofold: The first one is to present and define the vocabulary used. This is achieved by a dictionary for example which list all the terms actually used in language. Secondly, terminological ontology is the result of a terminology agreement between a users’ community. This agreement defines which term is used to represent a concept in order to avoid ambiguity. This process is called vocabulary normalization. When a concept could be described by two synonym terms, the normalization process selects one of those to be the preferred label of the concept. Moreover, taxonomy and thesaurus organized their normalized vocabulary so that the a priori relationships between concepts are made explicit for human. Unfortunately the distinction between concepts and their instances are not taken in account: Instances are considered like concepts. A thesaurus has three basic relationships among terms: equivalence, hierarchical and associative. Let us point out that the last two relations hide several semantic relations. Associative relation between two terms means that there exists a semantic link between concepts labelled by these terms but no information is given on this semantic link. Hierarchical relation between two terms hide any hierarchical relation like the “instance of” relation between a concept and one of its instances, the “subset” relation between two sets, the “part of” relation between a whole and its component and so one. More information on thesaurus development are available in (ISO 2788 and ISO 5964). The Simple Knowledge Organization System (SKOS) and the Resource Description Framework (RDF) are two XML based languages specific to store this type of ontology.

2.2 Data Ontologies:

The purpose of data ontology is to model consensual and standard data model in order to facilitate data exchange between information systems. Data ontologies are built upon terminological ontologies describing the terminological agreement between data users. They provide conceptual schemata whose main focus is normally on data storage and data handling, and are used for information exchange, with the goal of guaranteeing data consistency. In data ontologies, a concept is composed of a set of properties; all concepts are also defined thanks to each other by the relations they have. These relations are also associated to integrity constraints. At execution time, data are stored in the properties of object, that is to say an instance of concept. Thus, data could be processed in various treatments (called methods). Nevertheless, data ontologies goal is not to describe particular instances during execution time. Data ontologies are normally defined with conceptual modelling languages used in software and database engineering: For example Entity-Relationship Model language or Object Model Language like the most well-known one the Unified Modelling Language (UML) (OMG. 2003).

2.3 Logical Ontologies

This type of ontologies require a clear semantics for the language used to define the concept, clear motivations for the adopted distinctions between concepts as well as strict rules about how to specify concepts and relationships. This is obtained by using formal logic (usually First Order Logic or Description Logic) where the meaning of the concept is guaranteed by formal semantics. This ontology type is the only one that contains logical definition. For example, Knowledge Bases (KB) are formal systems that capture the meaning of the adopted vocabulary via logical formulas. A logical formula is a combination of concepts and semantic relations. A KB contains more expressive components than a conceptual schema. The main purpose of logical ontologies is reasoning even if logical ontologies can store data. Compared to data ontology, data are not associated to method in order to make some calculation; data are stored in property only to be retrieved. Logical ontology does not focus on term and textual definition even if they could be store in the ontology. Terms are only used as symbol in order to help user during the handling of logical formula. There exist different formal languages used to describe logical ontology like Description Logics (DL), Conceptual Graphs (CG), First Order Logic (FOL), etc. Now, the Ontology Web Language (OWL), the standard recommended by W3C is the most used.

3 Different type of interoperability

This section illustrates how the three types of ontologies are used to solve interoperability issues across heterogeneous systems. We present four kinds of interoperability: lexical, data, knowledge model and object. For example, in the first section we provide an analysis of how these ontologies can be used for lexical interoperability in document management systems, followed by section presenting the use of ontology for overcoming differences between heterogeneous databases and knowledge bases. We will analyze their main role in the context of these systems.

3.1 Lexical interoperability in document management system

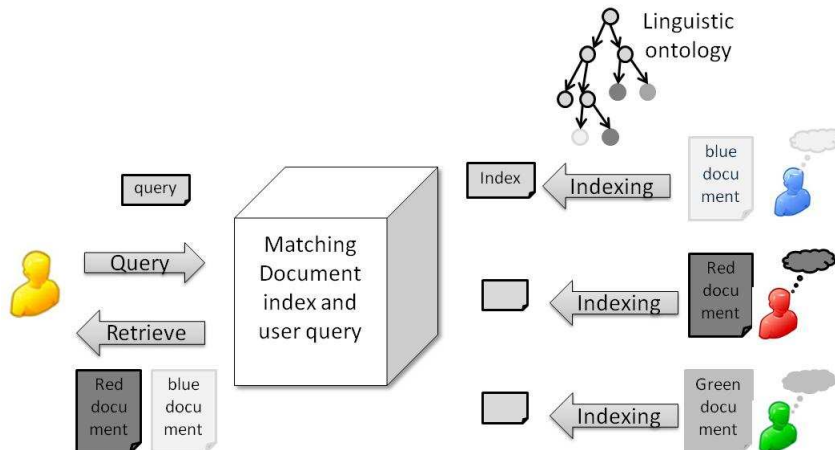


Fig. 2: architecture of an information retrieval system

In Information Retrieval, users send a query to the system in order to retrieve relevant documents. The goal of terminological ontologies in this type of system is to normalize the vocabulary used in the document to avoid lexical ambiguity. An example of lexical ambiguity is shown in Fig. 2: the green author employs the word “river” in the green document. The red author employs the word “watercourse” in his document to reference the same idea. Hopefully, the terminological ontology links the terms “river” and “watercourse” to the same concept by using a synonym link. This concept is contained in the green and red document indexes. Indexes contain the description of the document content. Thus document indexes and user queries use the same vocabulary, so the information retrieval system can compare them.

Terminological ontologies contain hierarchical links, related links and synonym links between terms. These links could be used during the matching process in order to compute a similarity degree between the document representation and the user query. Users build their queries by choosing the appropriate terms in the terminological ontology. For practical reasons, terms should be defined in the ontology not only by means of a formal definition, if any, but mainly with natural language definitions to explain the referring concept, so that humans can understand them easily.

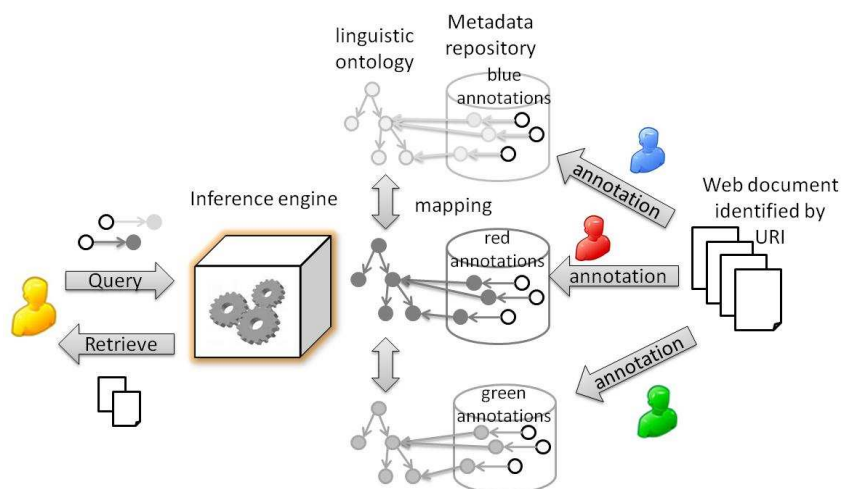


Fig. 3 : architecture of Semantic Web search engine

Semantic Web search engines are a new trend of Web search engine. In Semantic Web, users can annotate web pages according to a set of ontologies. Notice that in Fig. 3 the same document can be annotated by different users using different terminological ontologies. This collaborative annotation process can take in charge the large amount of data available on the Web. The Semantic Web search engine makes inferences about data and their meta-data in order to combine and compare them. Inference mechanisms can be more complicated than just a matching process; they can compute new metadata or check them. The final user queries the Semantic Web search engine by using its preferred terminological ontologies in order to retrieve parts of web pages.

Example: A semantic web portal in sustainable agriculture

Agriculture has to face a new challenge: sustainable development; this means among others developing technique to reduce agricultural inputs like pesticides. The main problem is the information, the know-how, the knowledge to carry on sustainable agriculture is not yet available because this knowledge may not exist or is much localized, hard to acquire and repeat (INRA, 2010). Thus the development of a semantic web portal dedicated to sustainable agriculture seems to be a good solution to increase the dissemination and the acquisition of information about sustainable agricultural techniques (Soulignac et al. 2009). Moreover the agricultural community is heterogeneous. For example in France, the agricultural organization is not reduced only to the farmers. Agriculture community is also composed of the French State actors for the initial formation, European actors for constraint specification and so one. Thus, lexical heterogeneity exists between different set of actors. The semantic web portal has two goals. The first is the retrieval of documents, using key words suggested by the user. The tool proposes query terms store in a terminological ontology. The second objective is to group farmers supporting the same problems. Thus a matching process between farmer description and problem description has to be proposed. In order to achieve all these goals the underlying ontology will be store in OWL that can save terminological and data content in a same format.

3.2 Data interoperability between software chains: definition of a data exchange format.

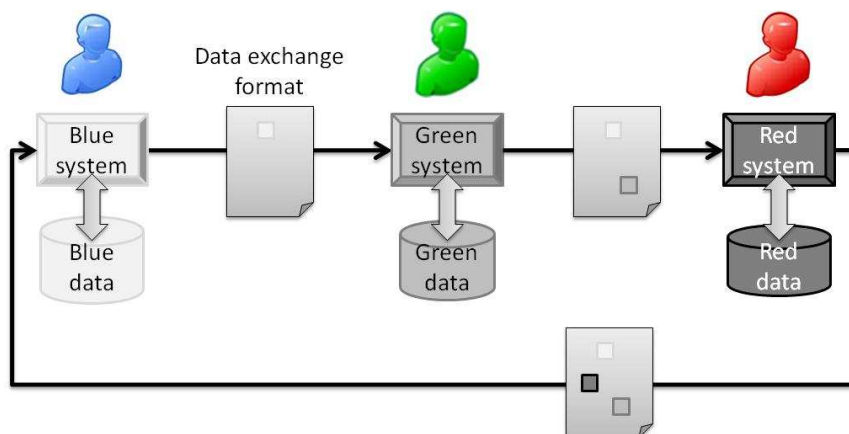


Fig. 4: a software chain using a data exchange format.

Data ontology can be used as a data exchange format recognized by different systems. As shown in Fig. 4, the output of a blue system stored in this format can become the input of the red or green system. Data exchange format is the result of a terminological and structural agreement between each software company. The structural agreement enables each software to share the same data structure storage. The structural agreement is possible only if a terminological agreement is reached. The terminological agreement signifies that the same

name is used to reference similar classes or property in the different systems. The internal model of each system is not dependent on the data exchange format. That is to say the data associated to an object in the data exchange format, can be stored in several objects inside the blue system. Inversely an object of the blue system can be built by analyzing several objects of the data exchange format. The only constraint about data exchange format is that all the data useful by another system should be defined in the data exchange format.

Example: OTAG project

The OTAG project, supported by the European Union, focuses on improving innovative economically viable mechanisms, methods, and geo-technologies for recording reliable and accurate data on beef production and developing an operational geo-decisional system under control conditions to track and trace the mobility, provenance, and state of beef cattle using emerging geospatial and geo-communication technologies (Visoli et al. 2009). It is mainly applied to traceability of beefs in Brazil. In case of sanitary alerts, the system should be able to determine the animals which have been in contact with a diseased animal. A lot of data are integrated in the geo-decisional system and the global system is composed of several systems which communicate between each other: each system communicate with XML messages. In this project the ontology used is a data one: the objective was to produce a common vocabulary for all the system and for all the actors (i.e. farmers, consumers, researchers), and to produce the XML messages between each system layers for data exchange. The ontology was based on a UML model and has been built iteratively thanks to meetings and discussions between the different actors (Visoli et al. 2008). The main concepts are farm, farmer, cattle area, paddock, animal, sanitary event and geo-locations of bovine. These concepts come from practices used in Brazilian beef cattle production systems.

Example: French Data Reference Centre for Water

For example, the French Data Reference Centre for Water (SANDRE in French) is in charge of developing a common language for water data exchange (SANDRE). In France, data related to water and hydrology is issued from thousands of organizations and public services. The priorities of SANDRE are to make compatible and homogeneous data definitions between producers, users and databanks. For example, some themes considered by SANDRE are: groundwater, hygrometry, waste water treatment, water pollution etc. SANDRE proposed "a common language concerning data involved in the French Water Information System. Specific terms relevant to water data are clearly defined and data exchange specifications are also produced to fulfil the communication needs between partners involved in the field of water" (SANDRE). One of the goals of SANDRE is to define, at a national level, a common vocabulary concerning the field of water (SANDRE common language). To fulfil this task, data models have been developed. They are associated to data dictionaries that gather all the definitions of data relevant to a topic concerning water. XML-based exchange formats have been also proposed. These XML format could be considered as data ontology focused on Water community.

Concerning water pollution and waste water treatment, SANDRE ontology contains also data about organic matter spreading: Organic spreading of organic effluent can impact the quality of surface water body and ground water body. Moreover organic spreading is also used to recycle the sewage sludge of Waste Water Treatment Plants. In 2004, two communication scenarios using SANDRE ontology has being proposed: The first scenario is the description of parcels suitable for organic spreading and the second one describes the realized campaigns of organic spreading. The data contain the localization of the parcel and the type of organic product used during the organic spreading. All these data will improve the communication process between the different actors concerning with organic spreading: water agency, land agency, farmers. Thus, the French environmental ministry is implementing these two communication scenarios with an operational project called Sigemo/Sillage (Soullignac et al, 2006) in order to acquire, update and retrieve data concerning organic spreading.

3.3 Knowledge model interoperability for life cycle system (object type interoperability).

This kind of interoperability is proposed by Fonseca (Fonseca et al. 2000). The goal is not to exchange directly data or to query heterogeneous data source but to focus on how to design, implement or update easily an information system by using set of ontologies. Ontologies become engineering artefacts which are components of the information system development like a design pattern. Thus reusing data or knowledge may decrease cost of developing GIS project, and may improve the quality of the development process. Most parts of ontologies used in this kind of interoperability system are data ontologies based on UML models. Moreover all the systems design with the same ontologies will interoperate more easily because they are based on the same assumption about physical world perception. The use of ontology, translated into an active geographic information system component leads to what Fonseca call Ontology-Driven Geographic Information Systems (ODGIS)

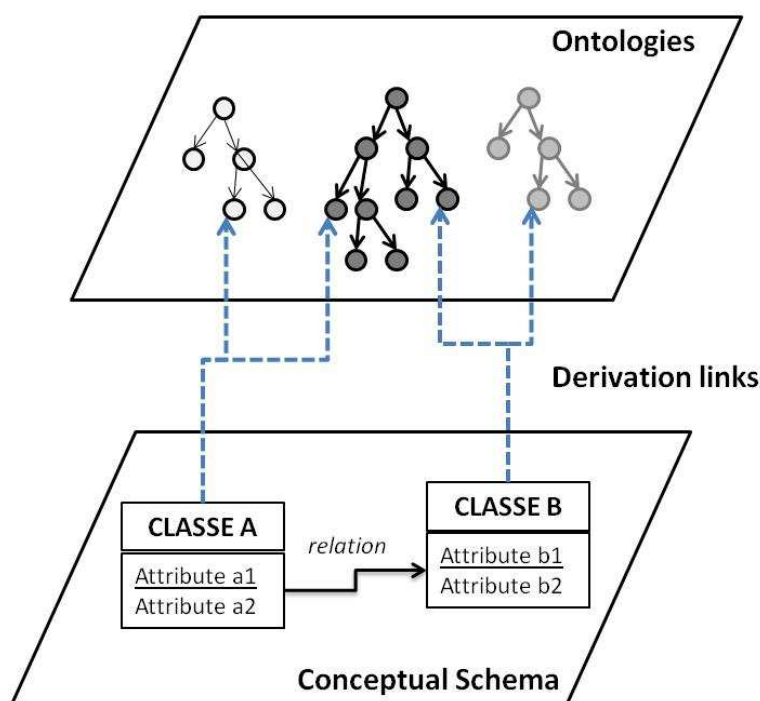


Fig. 5 : Ontologies used during the development of information system

Data ontologies can be used to annotate part of the models between different applications. So, mapping between models will easily be identified.

Example: CEMagriM conceptual framework

Farm enterprises are complex systems that are not easily designed. Thus as far as possible we need to reuse existing model: to facilitate knowledge capitalization, business process re-engineering and integration, change management in Information System. The prospective project called CEMagriM¹ (Cemagref Enterprise Modelling in Agriculture Integrated Methodology) will propose an integrated methodology for modelling farm enterprises. The aim of this project is to offer a generic conceptual framework to represent farm enterprise at a business level. This conceptual framework leans on Enterprise Modelling Constructs from the indus-

¹ If you need more information about CEMagriM, please contact Vincent Abt (Vincent.abt@cemagref.fr)

trial sector (Abt *et al.*, 2009) and notably on research activities in Enterprise Interoperability (<http://interop-vlab.eu>). This conceptual framework can be used to define modelling constructs (design pattern) for semi-formal enterprise modelling languages, but also to define a knowledge model at a high level of interoperability. This conceptual framework, based on an UML model, offers in this way a set of generic concepts such as process, procedure, operation, task, object, management unit (declined in working unit, biophysics unit and environment unit), etc.

Example: GIEEA project, Farm Information Management Project

The GIEA Project ²(project about management of farm information resources), leaded by APCA (Assemblée Permanente des Chambres d'Agriculture) from 2003 to 2006, federates the actors of agriculture field on the exchange of computerized data to make easier the communication between information systems (Brun *et al.*, 2005). The main result of this project is a dictionary of approximately 200 concepts. This dictionary defines a set of organized concepts and establishes the relationships between these various concepts. It can be regarded as a data ontology. This ontology is represented by UML class diagrams (Pinet *et al.*, 2006).

The project GIEEA, which is the follow-up of GIEA, started in 2009, is managed by the firm ACTA Informatique and is focused on the integration of environmental data in the dictionary (inputs, water, energy). ACTA Informatique and its partners carried out an inventory of decision-making aid tools used in agriculture. The new concepts identified in these tools will enrich the original dictionary GIEA on the environmental aspects. In order to improve the integration of these new concepts, the GIEA data ontology formalized in UML was transformed into a logical ontology formalized in OWL language. This UML to OWL transformation generates a reusable ontology for information systems modelling software. The goal of the logical ontology is to check the constraints (topological constraints for example) during the integration of new concepts.

3.4 Object interoperability: a global system related to heterogeneous local systems.

This type of system interoperability enables several heterogeneous systems to have a common user interface for querying. The global system is composed of global ontology. The goal of this global ontology is to unify and gather the different representations of real objects or phenomenon stored in each local system. The specific domain model of each local system is represented by a local ontology. This local ontology can be a specification of the global one. Wrapper is the system that provides a way to abstract the data from a data source and transform them in the common model define in the global ontology. Wrappers play the role of a translator between the local ontology and the global one. In order to match element of different ontologies and check coherency, reasoning service are mandatory thus all ontologies are logical ones. Thanks to these wrappers, the mediator is able to identify each different representations of the same real object stored in a data source. Thus the mediator can query each local data source by using the associated wrapper and gather all the result. Mediator decides how to access each data sources and in which order, normally by making a query planning step. Moreover in this type of architecture, the local system is still available for local users.

² Web site: <http://www.projetgiea.fr>

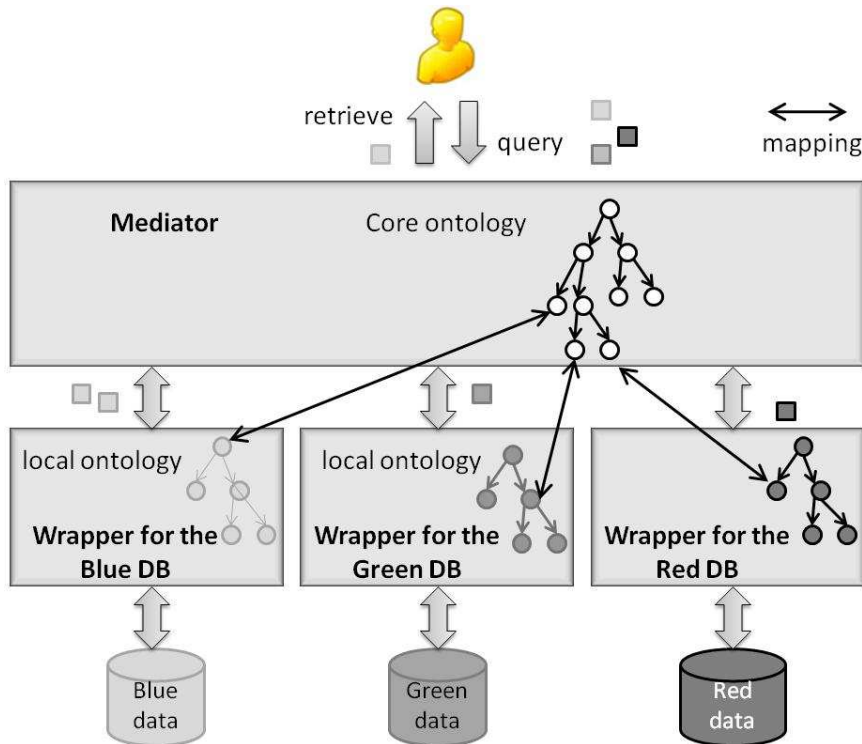


Fig. 6: a global system able to manage local heterogeneous system

Example: FORUM project

The FORUM₃ project proposes mediation architectures to facilitate the access to different French environmental data sources. In France, environmental data are handled by a large number of stakeholders for different purposes: evaluate the environmental quality, find the better place for a new infrastructure, evaluate the impacts of a human activity, etc. Mediation architectures can be used to solve the problem of accessing these heterogeneous data. The user query is based on a global ontology about environment. The global system usually needs to access several data sources to answer the user query. Thus, the user query is re-written in several queries by the global system; each one is dedicated to extract the needed information from a data source.

Conclusion

Ontologies have been used for the last decades for a set of tasks, one of which is focused on achieving interoperability between heterogeneous systems.

We have presented a new vision of different types of ontologies that have been considered in the literature for interoperability. Moreover we have described the main components of each type of ontologies and how these components can be used in the task of interoperability.

Our description is not exhaustive, and other types of interoperability could be found, but our aims are to show that for each type of interoperability there are different approaches to be taken into account. This survey is useful when approaching an interoperability problem and having to select the resources to be used for solve it.

References

Abt V., Vigier F., Schneider M. (2009) Enterprise Business Modelling Languages Applied to Farm Enterprise: a Case Study for IDEF0, GRAI-Grid and AMS Languages, in Papajorgji P. et al., *Advances in Modelling Agricultural Systems*, Springer-Verlag, p. 167-191.

³ FORUM Project Web Site <http://www.lirmm.fr/FORUM/>

Brun T., Poyet P., Bopp M., Vigier F., (2005) « Towards an agricultural ontology in France: Contributions of the farm Information Management Project (GIEA) », EFITA/WCCA 2005 Joint Conference, Vila Real, Portugal, July 25-28, p. 1296-1302

Fonseca, F., Egenhofer, M., Davis, C., and Borges, K (2000): Ontologies and Knowledge Sharing in Urban GIS. *Computer, Environment and Urban Systems*, 24(3): p. 232-251.

INRA (2010): *Ecophyto R&D Quelles voies pour réduire l'usage des pesticides*. Paris: 8 pages.

ISO 2788 (1986) *Documentation -- Guidelines for the establishment and development of monolingual thesauri*

ISO 5964 (1985) *Documentation -- Guidelines for the establishment and development of multilingual thesauri*

Object Management Group (2003). *Unified Modeling Language*, version 1.5, Mars 2003 - <http://www.omg.org/docs/formal/03-03-01.pdf>

Pinet F., Ventadour P., Brun T., Papajorgji P., Roussey C., Vigier F., (2006) « Using UML for Ontology Construction: a Case Study in Agriculture », Seventh Agricultural Ontology Service (AOS) Workshop on "Ontology-Based Knowledge Discovery: Using Metadata and Ontologies for Improving Access to Agricultural Information". Fifth International Conference of the Asian Federation for Information Technology in Agriculture (AFITA), Macmillan Edition, India Ltd, Bangalore, India, 9-10 November 2006, p. 735-739.

Soullignac, V. ; Barnabe, F. ; Rat, D. ; David, F.(2006). SIGEMO : un système d'information pour la gestion des épandages de matières organiques. Du cahier des charges à l'outil opérationnel. (SIGEMO: computerized system for the management of the spreading of organic products. From functional specifications to the operational tool.) *Ingénieries - E A T, METHODO / SYNERGIE*. n°47, p. 37-42

Soullignac, V., J.-P. Chanet, Paris J.L., Devise O., Gondran N. (2009): Knowledge management and innovative design, state of the art. 11th International Conference on the Modern Information Technology in the Innovation Processes of the industrial enterprises, Bergamo, Italy.

Visoli M., Ternes S., Pinet F., Chanet J.P., Miralles A., Bernard S., De Sousa G. (2009): *Computational architecture of OTAG project*. Proceedings of the 7th EFITA Conference, Wageningen, Netherlands, 6-8 July 2009, p. 165-172.

Visoli M., Ternes S., Chanet J.P., Pinet F., De Sousa G., Miralles A. (2008): *OTAG information system*. Clermont-Fd: Cemagref - Embrapa, 2008, 15 pages.