

# Chapter 3

## Ontologies for Interoperability

Catherine Roussey, Francois Pinet, Myoung Ah Kang, and Oscar Corcho

### 3.1 Introduction

The goal of this chapter is to help readers understand how ontologies can be used to improve interoperability between heterogeneous information systems. We understand *interoperability as the ability of an information system or its components to share information and applications*. In the literature there is not a common agreement on which types of interoperability can be found between heterogeneous systems, but mainly classifications of the different types of heterogeneity that can be found between systems and the levels or layers where this heterogeneity has to be solved or overcome. However, this is not the purpose of this chapter. We will focus on which types of system interoperability can be resolved by ontologies, and which types of ontologies have been normally used for this purpose. About ontology types, we refer to the first ontology classification presented in Chap. 1.

Some of the illustrative examples will be taken from project presentations made in the context of the COST UCE Action C21 (Urban Ontologies for an improved

---

C. Roussey (✉)

LIRIS, Université de Lyon, France  
e-mail: Catherine.Roussey@cemagref.fr

F. Pinet

French Institute for Agricultural and Environmental Engineering, Clermont Ferrand, France  
e-mail: Francois.Pinet@cemagref.fr

M.A. Kang

Laboratory of Computer Science, Modelling and System Optimization, Blaise Pascal University, Clermont Ferrand, France

O. Corcho

Departamento de Inteligencia Artificial, Facultad de Informática, Universidad Politécnica de Madrid (UPM), Spain  
e-mail: ocorcho@fi.upm.es

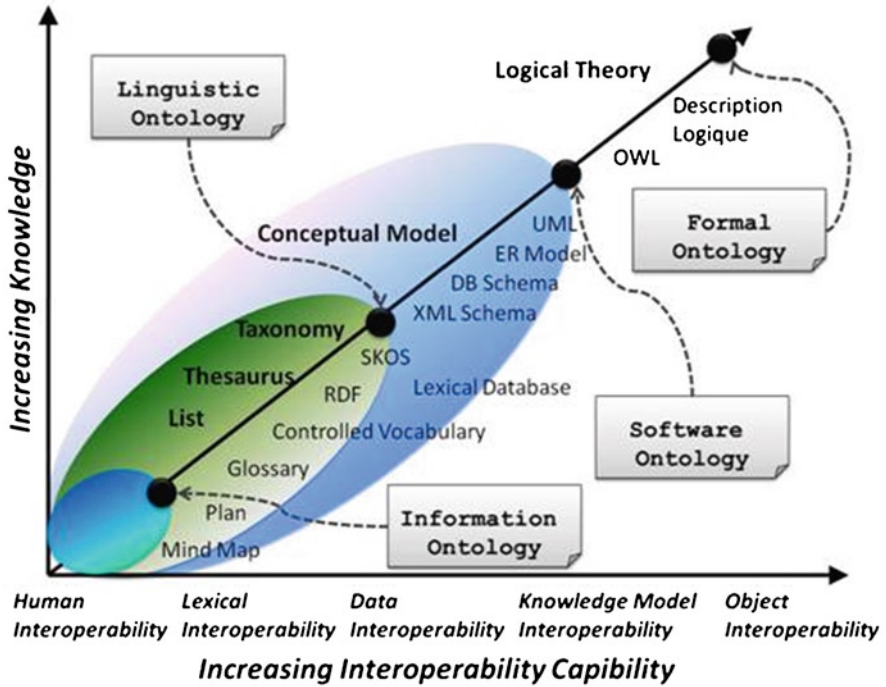


Fig. 3.1 A schematic representation of the different kind of interoperability based on our ontology classification

communication in UCE projects TOWNTOLY) or, in general, in the area of Geographic Information Systems (GIS).

As shown in Fig. 3.1, this chapter presents four kind of interoperability: lexical, data, knowledge model and object. (The human interoperability is not presented because these interactions are made only among human). 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 is in the context of these systems.

### 3.2 Lexical Interoperability in Document Management System

In Information Retrieval, users send a query to the system in order to retrieve relevant documents. The goal of linguistic 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. 3.2: 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 linguistic ontology links the terms “river”

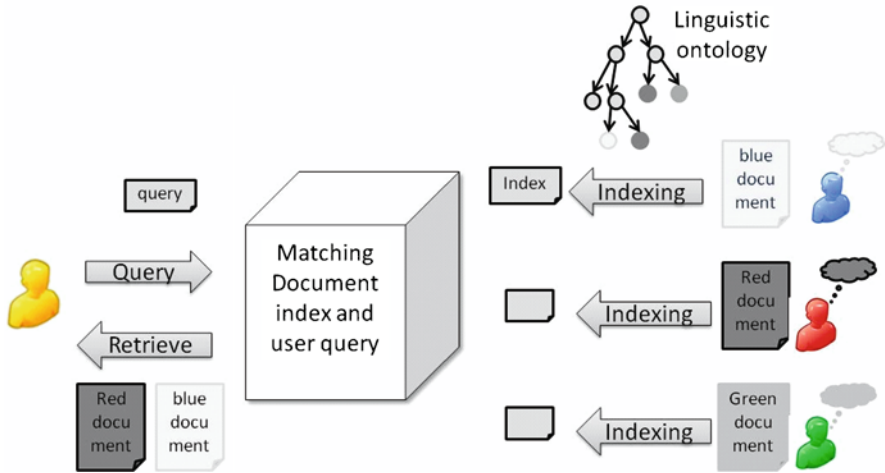
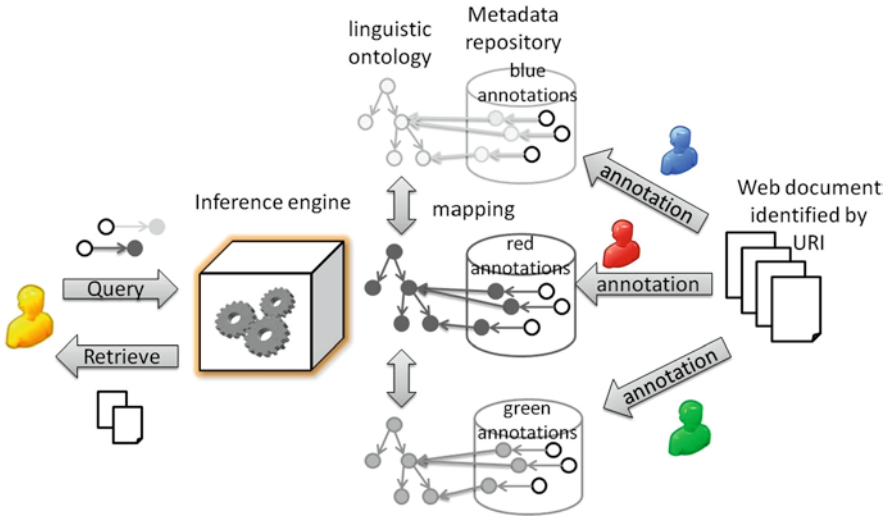


Fig. 3.2 Architecture of an information retrieval system

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. Chapter 5 complements this broad description, it explain how multilingual information retrieval system use linguistic ontology.

Linguistic 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 index and the user query. Users build their queries by choosing the appropriate terms in the linguistic 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. The scope of the linguistic ontology depends on the scope of the corpus of documents: domain, core reference or general.

Semantic Web search engines represent a new trend in Web search engines. In the Semantic Web, users can annotate web pages according to a set of domain, local, core reference, etc., ontologies, what may also include references to linguistic ontologies. Annotation is different to indexing because annotation does not refer to the whole document like indexation. Annotation process associates a piece of data (a part of web page) to its corresponding metadata (a piece of data that describes the web page part). Annotation is composed of RDF triples (subject, property, objet): the subject is a part of web document identified by a URI, the property and the objet (the associated value of the property) is defined inside the linguistic ontology. All the RDF triples and their associated linguistic ontologies compose a graph where leafs are web document parts. Notice that in Fig. 3.3 the same document can be annotated by different users using different linguistic ontologies. This collaborative annotation process can take in charge the large amount of data available on the Web.



**Fig. 3.3** Architecture of Semantic Web search engine

The Semantic Web search engine makes inferences about data and their metadata 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 linguistic ontologies in order to retrieve parts of web pages.

### 3.2.1 Example: *URBAMET Databank*

The URBAMET databank is an example of information retrieval system based on a linguistic ontology. The documents search engine is accessible through the URBAMET thesaurus. An analysis of this thesaurus can be found in Chap. 10.

### 3.2.2 Example: *The FAO Case Study of the NEON Project*

The “NeOn – Lifecycle support for networked ontologies” project aims at using ontologies for large-scale semantic search engine applications in distributed organizations. Indeed, fisheries department has several information and knowledge organization systems describing the world’s fisheries and aquaculture. Information resources are available as parts of websites as individual documents, images, databases etc. These data sources could be better exploited by bringing together related and relevant information. To reach this goal, a set of fisheries ontologies are developed to provide semantic search information service. The set of fisheries ontologies

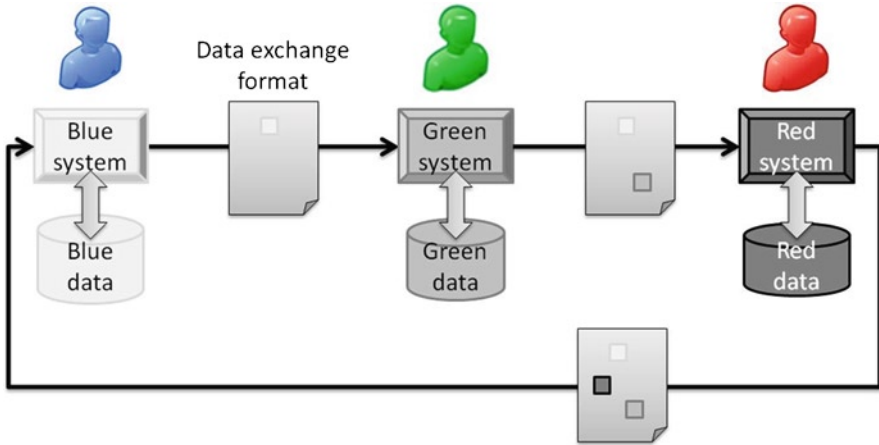
is composed of: land areas ontology, fishing areas ontology, biological entities ontology, fisheries commodities ontology, vessels ontology, gears ontology and fisheries fact sheets ontology. These ontologies are built by merging and integrating several thesaurus like AGROVOC ([AGROVOC](#)), ASFA, RTMS and others fishery glossaries. Indeed these fisheries ontologies are not purely linguistic because they also deal with structured data like database, thus in the NEON project some participants develop a new ontology model merging linguistic ontology model with software and formal ontology model (Montiel et al. [2008](#)).

### ***3.2.3 Example: The GEO Semantic Web Communities of the Italian “Three Lake Region”***

The territory of the Italian three lake region has developed a unique urbanism model characterized by combining an historical villas landscape and great naturalistic areas. In order to preserve these landscapes and to promote sustainable tourism, it is necessary to plan urban expansion for a rational use of natural and cultural resources. Sustainable tourism is a multidisciplinary domain dealing with scientific, historical, artistic and economical point of views. This requires integration and sharing of information between a numbers of local actors. (Marcheggiani et al. [2007](#)) propose a Geo Semantic Web community tools based on RDF annotations. Annotations are provided by each local actor to be accessible by all members of their communities. A community of local actors shares the same domains of interest; their centers of interest are described in a domain ontology and the related RDF annotations. Notice that a local actor can belong to several communities. Seven communities are identified: touristic system, municipalities, protected area, guide, police, Bed&Breakfast. The Geo Semantic Web community tools use Google Maps and Google Earth to visualize geographic object. A geographic object could be a point with latitude and longitude coordinate or a more complex geographic object like a polygon. To describe geographic objet the authors used two RDF ontologies the W3C Basic Geo Ontology and the RDFGeom Ontology.

### **3.3 Data Interoperability Between a Software Chain: Definition of a Data Exchange Format**

Software ontology can be used as a data exchange format recognized by different systems. As shown in Fig. [3.4](#), 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 lexical and structural agreement between each software company. The structural agreement enables each software to share the same data structure storage. Notice that usually data are stored thanks to an object oriented model. Thus concepts are object classes and instances are objects. The structural agreement is possible only if



**Fig. 3.4** A software chain using a data exchange format

a lexical agreement is reached. The lexical 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 dependant 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. Due to the fact that this ontology is used by different systems, data exchange format should be core reference ontology. Indeed, each system represents a user group task.

### 3.3.1 Example: Building Information Models

The Aurora is a new university building in Joensuu. During the second phase of the Aurora project, IFC classes are used as data exchange format between several design software: architecture, structure and building services. During the early conceptual stage of the project, several models called Building Information Models (BIM) are build based on IFC classes: 3D Architectural model is build by architect to create space, 3D structural model are used by fabricators and contractors to detail frame structures, The building service design model describes lighting system, the product model estimate the cost of the building process. All of these models exchange data and are associated to specific software with visualization and simulation capabilities. Using BIM and data exchange format improve the communication between stakeholders and the scheduling process. It also improves the cost estimation and the final product quality. For more details, see the case study presented in Chap. 8.

### ***3.3.2 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 are issued from thousands of organizations and public services. The SANDRE's priorities are to make compatible and homogeneous data definitions between producers, users and databanks. For example, some themes considered by SANDRE are: groundwater, hygrometry, 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 fulfill the communication needs between partners involved in the field of water" (SANDRE). One of the SANDRE's goals is to define, at a national level, a common vocabulary concerning the field of water (SANDRE's common language). To fulfill 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 software ontology focused on Water community, thus it also defined a core reference ontology about water.

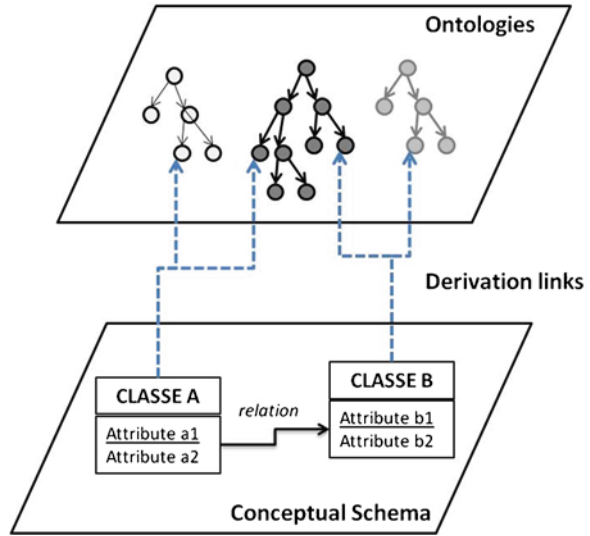
### ***3.3.3 Example: Farm Information Management Project***

The French standard proposed by the FIM project (GIEA in French; Pinet et al. 2009) describes a large number of concepts related to farms. The final goal of the standard is to provide more complete data interchange formats in order to facilitate and to improve interoperability between information systems of the French Farms (GIEA).

The first step of the project was to carry out an inventory of the various previous standardization initiatives. Then, different terms, concepts and their relationships have been identified for different main fields of Farm activities (land management, agricultural infrastructures and buildings etc.). An important part of the FIM project consists in integrating and enhancing the definition of concepts, and work on standardization already initiated by the various partners. The monitoring of these approaches and the participation in various work groups and their corresponding project committees are therefore fully integrated in the project.

A software ontology has been chosen to formalize the proposed standard. All the members of the project can propose new concepts to the developed ontology. Data interchange formats are also proposed on the basis of the vocabularies and the concepts of the ontology. The ontology is represented by UML class diagrams. UML has been chosen to model the ontology because the participants of the FIM project are familiar with UML.

**Fig. 3.5** Ontologies used during the development of information system



### 3.4 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 an engineering artifact which is a component of the information system development. Thus reusing data or knowledge may decrease cost of developing GIS project, and may improve the quality of the development process. Most part of ontologies used in this kind of interoperability system are software and core reference ontologies. 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. 3.5).

#### 3.4.1 Example: ODIS

Software Ontology can be a description of a generic knowledge model in order to develop new specific knowledge model dedicated to particular software able to solve a particular domain task. Each specific knowledge model based on this generic model will be easily mapped to another one which is also a specialization of the generic knowledge model. This type of system development based on generic knowledge model is called Ontology Driven Information System (ODIS) (Guarino 1998).



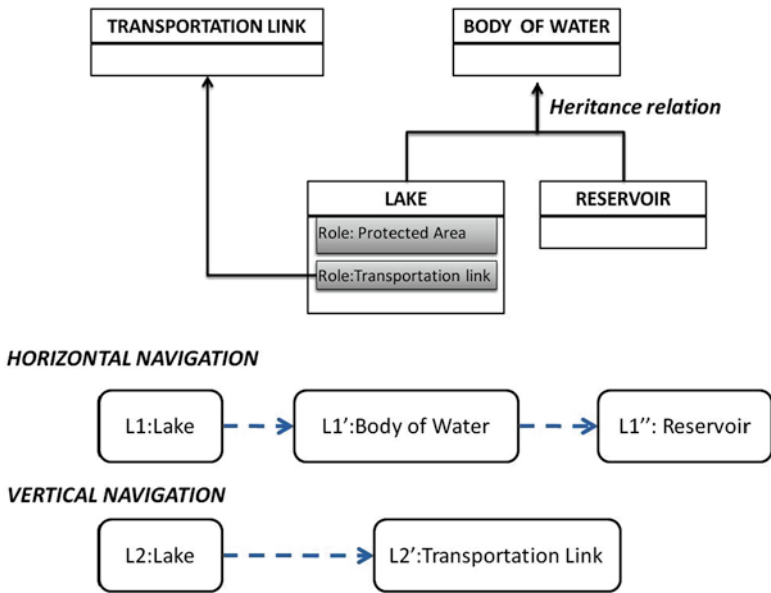


Fig. 3.6 Two examples of navigation between objects

Several software ontologies can be used to control the system development: domain ontology, task ontology, core reference ontology or foundational ontology like CityGML, geometric ontology, spatial reference system ontology or GML.

(Fonseca et al. 2000) propose an extension of this ODIS called Ontology Driven Geographic Information Systems (ODGIS). OGDIGS are built using software components derived from various ontologies. These software components are classes that can be used to develop new applications (Fonseca et al. 2000, 2002). The mapping of multiple ontologies to the system classes is achieved through object-oriented techniques using multiple inheritances. OGDIGS employs user classes that are derived through multiple inheritances from various formal ontologies to solve schematic heterogeneity. Thus a single geographic object supports multiple views; that is to say that each view is an object role containing an instance of a different parent class. The problem of the different levels of detail was approached by the introduction of a navigation mechanism that allows an object (the implementation of an ontology entity) to change its class by generalization or specialization. See for example Fig. 3.6, the object L1 instance of the class Lake, can be change to the new object L1', instance of the class Body of water. L1' has less detail than L1 but it could be change to the new object L1'' instance of reservoir class. This type of change is a vertical navigation along the hierarchic classification of user classes. Another operation called role extraction enables horizontal navigation (Fonseca et al. 2002). An object role can be automatically transformed into a new instance, acting as an independent object. Therefore, the new instance can be matched to an object associated with another entity in a different ontology. As shown in Fig. 3.6, the object L2' instance of Transportation link class, is created from the role transportation link of the objet L2 instance of Lake class.

### 3.4.2 *Example: User Adapted Interface Development*

Metral et al. (2007a) propose to use a core reference ontology to develop automatically several user specific interfaces of information system. User specific interfaces enable to access only suitable sources of information using an adapted vocabulary. A user specific interface is for example a web site.

This system manages heterogeneous sources of information like:

- Textual documents: regulation, legal text.
- GIS : cartographic system to search legal data related to parcel for example.
- Master and local plans (maps used for urban planification).
- 3D city models are used to simulate the impact of urban project or to promote this project. 3D model are communication tool that do not contains textual information.

The goal of this system is to gather all the sources of information and to adapt their presentation according to a user profile. All information is not suitable to a group of user: for example legal texts are not adapted to city inhabitants.

Thus, this system contains a core reference ontology untitled OUPP. OUPP is a global schema integrating in a common representation all the object representations, found in sources of information. An object, for example the railway station of the Lyon city, is an instance of a OUPP concept: railway station. Each source of information is linked to the instances they describe by an annotation link. Two types of annotation link exist: conceptual annotation link and instance link.

Each user group viewpoint is represented by a local ontology. Local ontology is a selection of dedicated OUPP concepts with the appropriate terminology. In this system only the linguistic part of OUPP is used. More precisely, a local concept is linked to a OUPP concept by semantic relations: equivalent relation or specialization/generalization one.

Thus thanks to the matching between local ontology and the core one, the system is able to compute all the sources of information suitable for a user group and build automatically the user adapted interface of the system (Fig. 3.7).

### 3.4.3 *Example: MDA*

(Cutting Decelle et al. 2006) presents an approach of software development known as Model Driven Architecture (MDA). MDA focuses on models (or conceptual schema) and models transformations as the primary steps in the development process. MDA prescribes three kinds of models:

- The Computational Independent Model (CIM) focuses on the environment and the requirement of the system.
- The Platform Independent Model (PIM) specifies the operation of the system independently of the platform that supports it.

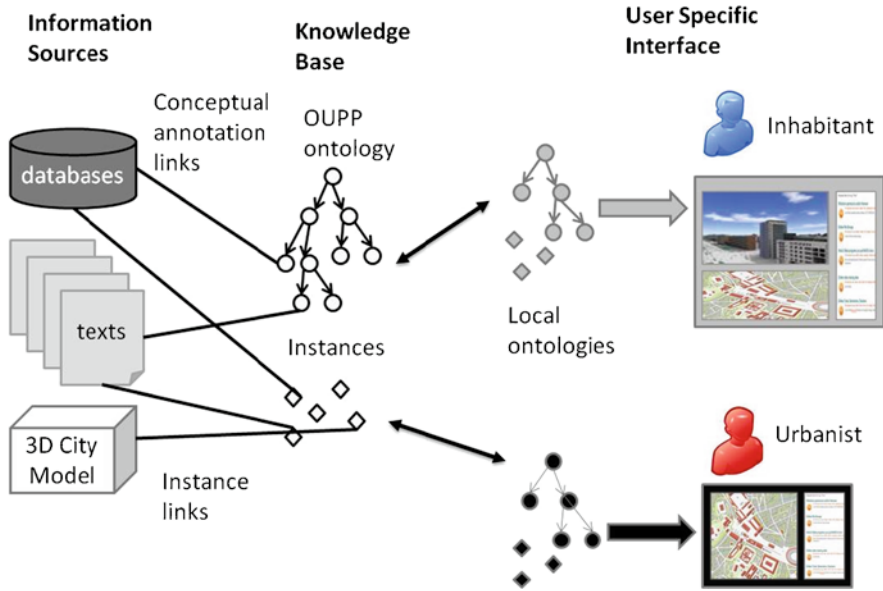


Fig. 3.7 Ontology based user specific interface

- The Platform Specific Model (PSM) focuses on the detail of the use of a specific platform by a system.

Model transformation is composed of a set of transformation rules, which specify the way a part of one model can be used to create a part of another model.

Thus, system development follows the different steps: the design of the CIM, the transformation of the CIM to PIM, the choice of the platform and the transformation of the PIM to PSM.

MDA approach allows different applications to be integrated by explicit relations between their models, thus enabling the integration, the interoperability and the evolution of supporting system.

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

### 3.5 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 a core ontology.

The goal of this core 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 core one. A wrapper is a system that abstracts data from a data source and transforms them into the common model defined in the core ontology. Wrappers play the role of a translator between the local ontology and the core one. 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.

### ***3.5.1 Example: Forum***

Another project named FORUM proposes mediation architectures to facilitate the access to different French environmental data sources ([FORUM](#)). 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 core reference ontology about environment (e.g. a global schema). The global system usually needs to access several data sources to answer the user query. Thus, the user query is rewritten in several queries by the global system; each one is dedicated to extract the needed information from a data source.

### ***3.5.2 Example IGN-E Case : The Phenomen Ontology***

The National Geographic Institute of Spain (IGN-E) has in charge to manage four cartographic databases that correspond to different scale: (1:25,000), (1:50,000), (1:200,000) and (1:1,000,000). These databases present a great heterogeneity due to the difference of the information sources used to build them. IGN-E wants to integrate all these four databases in order to facilitate their maintenance and to build a common features catalogue. (Gomez-Pérez et al. 2008) propose to build a domain ontology called PhenomenOntology able to query several cartographic databases. The goal of the PhenomenOntology is to link each databases in order to query simultaneously heterogeneous databases and to keep their structure. First the PhenomenOntology is built to contain all the features types stored in each databases. Secondly, as shown in Fig. 3.8: a global system able to manage local heterogeneous system.

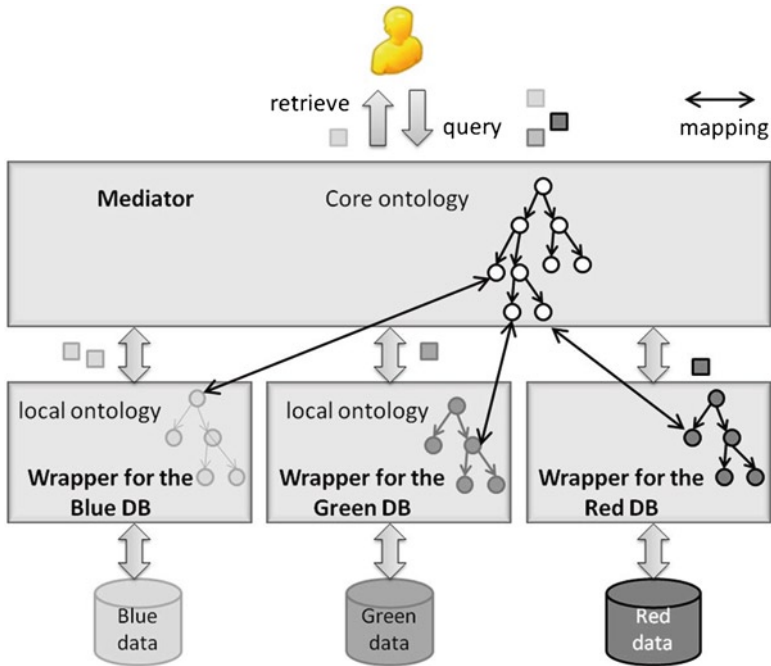


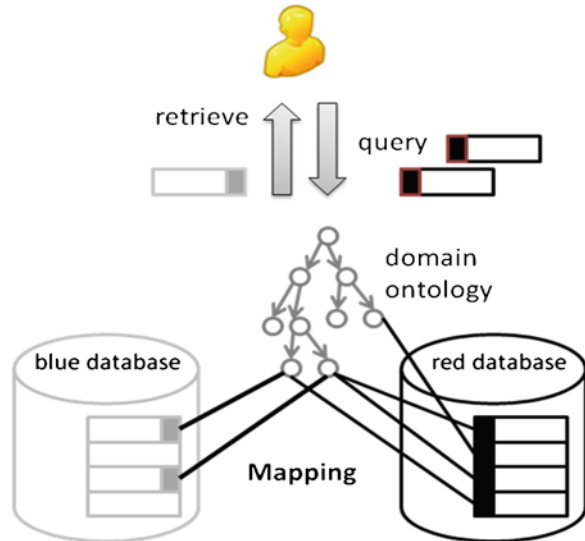
Fig. 3.8 A global system able to manage local heterogeneous system

Figure 3.9, each instance is linked to a features type by a mapping. This is a simplification of the global system presented in Fig. 3.8. This simplification is possible because all the databases share a common point of view of the domain.

### 3.5.3 Example: Integration of 3D City Models and Air Quality Models

In Chap. 7 of this book Metral and Cutting-Decelle propose to use a core reference ontology called OUPP to integrate CityGML, a 3D city model, with an air quality model. CityGML is used to visualize 3D elements and the air quality model is able to compute flow pollution. The integration of these two models enables to visualize air pollution flow in a 3D city model. CityGML and the air quality model are represented by two domain ontologies. The goal of the core reference ontology OUPP is to map equivalent concept belonging to each domain ontology. The mapping should specify how the transformation of a 3D attribute into an air quality one. Metral et al. (2007b) focuses on the extraction of street canyon, a very important air quality component, from the 3D city model.

**Fig. 3.9** The IGN-E case of heterogeneous databases



### 3.6 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 interoperability issues and how different type of ontology 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. In the next chapters you will find some more detailed descriptions about ontology usage and construction.

### References

- AGROVOC: A Multilingual Agricultural Thesaurus. [ftp://ftp.fao.org/gi/gil/gilws/aims/references/flyers/agrovoc\\_en.pdf](ftp://ftp.fao.org/gi/gil/gilws/aims/references/flyers/agrovoc_en.pdf)
- Cutting Decelle, A.F., Bourey, J.P., Grangel, R., Young, R.I.M.: Ontology based communications through model driven tools: the MDA approach feasibility of the approach in urban engineering projects. In: Teller, J., Lee, J., Roussey, C. (eds.) *Ontologies for Urban Development: Interfacing Urban Information Systems*. Studies in Computational Intelligence 61, University of Geneva 6, 7 Nov 2006, pp. 143–156. Springer, Berlin/Hiedelberg (2007). ISBN 978-3-540-71975-
- Fonseca, F., Egenhofer, M., Davis, C., Borges, K.: Ontologies and knowledge sharing in urban GIS. *Comput. Environ. Urban Systems* **24**(3), 232–251 (2000)

- Fonseca, F., Egenhofer, M., Davis, C., Câmara, G.: Semantic granularity in ontology-driven geographic information systems. *AMAI Ann. Math. Artif. Intell Spec Iss Spatial Temp. Granul.* **36**(1–2), 121–151 (2002)
- FORUM Project Web Site: <http://www.lirmm.fr/FORUM/>
- GIEA (FIM) Web site: <http://www.projetgiea.fr>
- Gomez-Pérez, A., Ramos Gargantilla, J.A., Rodríguez Pascual, A., Vilches Blázquez, L.M.: The IGN-E Case: Integrating Through a Hidden Ontology. *Lecture Notes in Geoinformation and Cartography*, pp. 417–434. Springer, Berlin/Heidelberg (2008)
- Guarino, N.: Formal ontology and information systems. In: Guarino, N. (ed.) *Formal Ontology in Information Systems, Proceedings of FOIS'98*, Trento, Italy, 6–8 June 1998, pp. 3–15. IOS Press, Amsterdam (1998)
- Marcheggiani, E., Nucci, M., Tummarello, G., Morbidoni, C.: Geo semantic web communities for rational use of landscape resources. In: *Proceedings of the 2nd Cost Action C21 – Towntology Workshop Ontologies for Urban Development: Conceptual Models for Practitioners*, Turin, Oct 2007
- Metral, C., Falquet, G., Karatzas, K.: Ontologies for the integration of air quality models and 3D city models. In: *Proceedings of the 2nd Cost Action C21 – Towntology Workshop Ontologies for Urban Development: Conceptual Models for Practitioners*, Turin, Oct 2007 (2007a)
- Métral, C., Falquet, G., Vonlanthen, M.: An ontology-based system for urban planning communication. In: Teller, J., Lee, J., Roussey, C. (eds.) *Ontologies for Urban Development: Interfacing Urban Information Systems. Studies in Computational Intelligence 61*, University of Geneva 6, 7 Nov 2006, pp. 143–156. Springer, Berlin/Heidelberg (2007b). ISBN 978-3-540-71975-
- Montiel-Ponsoda, E., Aguado de Cea, G., Gómez-Pérez, A., Peters, W.: Modelling multilinguality in ontologies. Poster at *Coling Conference*, Manchester
- Pinet, F., Roussey, C., Brun, T., Vigier, F.: The use of UML as a tool for the formalisation of standards and the design of ontologies in agriculture. In: *Advances in Modeling Agricultural Systems*, 18 pp. Springer, New York (2009)
- SANDRE: SANDRE Web site. [http://sandre.eaufrance.fr/rubrique.php3?id\\_rubrique=60&lang=en](http://sandre.eaufrance.fr/rubrique.php3?id_rubrique=60&lang=en)