



# Use Case 6 in Health Integrating dialysis and transplantation data for strategic decisions in healthcare

KW Partner: Agence de Biomédecine

## 1. Overview

### Challenge

Better patient care, better understanding of diseases and sound decision making in public health require accessing large amounts of data distributed in heterogeneous sources.

### Solution

A system, which allows semantic integration of the data A query engine based on a shared ontology Reformulating queries over the sources

### Why a Semantic solution ?

Local centers use different schema and different terminologies for encoding the data. The current approach is to store the data in a common repository, asking the centers to enter their data in a standardized way through a Web server. Since data are not automatically integrated from the already existing information systems, clinicians have to register patient's data several times locally, regionally, nationally and in various forms.

### Key Business Benefits

Prevent duplication of work for clinicians  
Assure quality and consistency of data  
Easier evolution and maintenance

### Business Partners

Agence de la Biomédecine (Abm) and University Rennes 2

### Keys components

#### Existing Software

DIADEM: Dialysis patient information system  
CRISTAL: Transplant information system  
ABM-METATHESAURUS: Abm reference ontology

#### Research and development

Semi-automated creation of ontology  
Semantic mapping  
Semantic query engine  
Reformulation of queries

#### Technology locks

A rule language extending OWL for expressing "mapping" between the local and global vocabulary and expressing "queries"  
Query-answering with ontology and rules

The screenshot displays the DIADEM web application interface. At the top, it shows the user is connected as 'John Bitume' and provides navigation options like 'Recherche' and 'Administration'. The main content area is divided into several sections:

- Messages Cristal:** A table listing patient messages with columns for Date, Patient / Equipe, and status icons.
- Dossiers à traiter:** A section for processing records, including a table with columns for Numéro, Nom / Prénom, and actions like 'Dossier à corriger'.
- Dossiers à valider:** Sections for records needing validation, with columns for Numéro, Nom / Prénom, and Date de dernier suivi.
- Dossiers validés incomplets:** A section for records that are validated but incomplete.
- Dossiers validés avec enquêtes en cours:** A section for records with ongoing surveys.

Below these sections, there are filters for 'Équipe d'origine' and 'Région d'origine'. The bottom part of the screenshot shows a search results page with a list of medical codes and their corresponding descriptions, such as '04N1: Néphropathies glomérulaires primitives'.

This health case deals with the integration of patient data in a central repository at Agence de la Biomédecine (Abm), in which high level of quality of data are considered strategic asset for epidemiological studies and public health decisions around organ transplantation and dialysis. la de la Biomédecine is a state agency in France, dealing with public health issues including organ, tissue and cell transplantation. Abm is responsible of the organ allocation policy and of the allocation system. Various studies and real experiments are carried out from the patient observed data always aiming at improving efficacy and equity of the French allocation system.

In this case, many different information are gathered from multiple sources of data. Some data lie in the transplant or dialysis databases that already exist in the different hospitals, dialysis, kidney transplant centers. Other data are directly entered by the centers into the Abm general repository in filling a standard form through a Web server.

The goal of the system is to enable uniform querying and automatic storing of all the data in a unique centralized repository.

*The usage scenario is to promote the semantic integration of heterogeneous data through a shared representation and to promote ontology-based query engine for use in biomedical applications.*

From a biomedical standpoint, semantic integration is crucial in many domains where better patient care, as well as better understanding of diseases and sound decision making in public health or epidemiological studies require to query large amounts of data from heterogeneous sources. In the domain of terminal organ failure and transplantation, many countries are developing regional or national registries. There are also European registries promoted by scientific societies (e.g. ; ERA-EDTA registry for kidney, ELTR for liver transplantation). They contribute to the scientific analysis of transplant outcomes and provide important data that can improve patient care and/or influence local, regional, and national transplant policies. Such registries are populated by data from several national or local information systems. Different terminologies are used in the different centers and in the different countries. The absence of coordination makes that clinicians are obliged to register patient's data for regional, national and international authorities several times and in various forms. The encoding is not always consistent. Semantic integration may considerably facilitate the acquisition of reliable patient's data at different levels.

From a technological point of view **(i)** semantic integration is based on a *global ontology* (or mediated schema) complying with the Abm view of the domain. This ontology defines all the concepts and properties of dialysis and transplantation. It provides the vocabulary for posing queries and for defining mappings to the local sources **(ii)** data are stored in the existing *local sources*. Each source uses its own ontology (schema) **(iii)** The content of the sources is defined in terms of views over the global ontology. A set of *mappings* relate the global ontology to the sources ontologies **(iv)** A *query engine* exploits the global ontology and the mappings to reformulate queries in terms of the sources specific ontologies. The main challenges concern knowledge representation issues: extending current standards of Web Ontology Languages with a rule language while (1) maintaining decidability for subsumption and satisfiability, (2) and for the reformulation of queries

## **2. Current Practices and Technologies**

A solution mainly adopted for data integration is datawarehouses (DW). Since data are duplicated, DW can be quite efficient, as they provide high performance to access data. However, they are often not a satisfying solution, because they are not well appropriate to evolutive data and to add new sources either. Extensibility and real-time data limitations, are prohibitive not only for this application but for many biomedical applications. For example, Genomics is a fast-moving field. Web sources are multiple, with huge and constantly evolving content. New specialized databanks with specific terminology frequently appear. More flexible recent approaches like mediators or peer-to-peer seem more appropriate. However, they raise other difficulties, because biomedicine has specific features and requirements. Indeed, data should be *integrated*, but also very often *secured*, and/or *historised* and *aggregated*.

### 3. System requirements Analysis

The different local centers and Abm do not use the same schema for their information system. Moreover, the data are often encoded according to different terminologies. This situation occurs very frequently in biomedicine, where multiple terminologies have been developed. This case makes the following problematic stand out:

1. A first requirement is the *semantic integration* of information at two levels, both of the local sources schema and of the stored data.
2. Systems have to respect *confidentiality* of private nominative patient data.
3. Since each authority should be able to justify its decision at any time, data have to be *archived* and *historicized*.
4. The data have to be *aggregated* according dimensions relevant to the studies.

Such requirements are not specific to Abm. Querying heterogeneous data while respecting confidentiality and having to justify the decisions or to aggregate data, are encountered in many biomedical applications. As datawarehouses are well adapted for confidentiality, historization, aggregation, while mediators or peers are better suited to semantic integration, evolutivity, inferences, a new hybrid architecture combining datawarehouse and mediator advantages may be an interesting direction to investigate for biomedical applications.

### 4. Review of the current systems

*Agence de la Biomedecine*, former Etablissement français des Greffes (EfG), was created in 1994 as a state agency in charge of public health issues including organ, tissue and cell transplantation. Abm is responsible for the registration of patients on the national waiting list, the management of this list, the allocation of all organs, and the evaluation of retrieval and transplantation activities by organ and transplantation units.

To fulfill its missions, *Agence de la Biomedecine* maintains a national information system CRISTAL. Organ procurement organisations register data about donors and the transplantation teams record patient data at the time of registration on the waiting list, and of transplantation if any. CRISTAL is used at the individual level for organ allocation. CRISTAL is also used for evaluation studies, and strategic decisions in the field of organ failure public health policies. Therefore, CRISTAL needs to integrate periodically data coming from multiple sources: patient data coming from transplant teams or dialysis centers, information coming from organ procurement organisations, and geographical information about the health offer organisation.

*Agence de la Biomedecine* now also manages DIADEM a patient information system centralizing all the data of dialyzed patients.

*Agence de la Biomedecine* also supports the Renal Epidemiology and Information Network (REIN) devoted to the follow-up of all patients treated by dialysis, connecting the cohorts of dialyzed and transplanted patients to offer a complete registry for all patients with end-stage renal disease in France.

*Agence de la Biomedecine* is also developing a terminological server that covers end stage diseases and organ transplantation.