ABSTRACT

A wide variety of distributed applications are nowadays emerging in diverse domains. These applications deploy various database systems for the management of their information, in which the diversity stems from the specific information management requirements and the objectives targeted by these applications.

Existing applications differ in their main characteristics and required features. On one hand, they differ in their distributed/centralized architecture, their size, complexity, and the type of data they handle. On the other hand, their requirements depend on the global functionalities that they need to provide and on the required level of interoperation with other sites. The used database management systems (DBMSs) are at best chosen to meet the specific characteristics and requirements of every application environment. However, currently available DBMSs lack the possibility to be efficiently used for all types of applications. Some DBMSs better suit smaller applications, while others are more dedicated to complex environments and focus on the management of, for example, multimedia information and large data sets. Thus, any attempt in the direction of forcing different applications to use the same database system for the management of all their information services is unrealistic. Even within the same environment, in certain complex applications, the use of more than one DBMS cannot be avoided.

Furthermore, from the application case (Healthcare) described in the thesis, it is clear that in today’s organizations, new and existing applications require access to data stored in several preexisting databases detained at several local and remote sites. Therefore, a main criterion required by most complex organizations, is the provision of collaboration
possibilities and information integration mechanisms among distributed, heterogeneous, and autonomous systems. The development of an application case provides interoperability and information integration among distributed systems, via the deployment of database standards and emerging Internet technologies. It is one of the most challenging approaches in the area of integrating heterogeneous information from autonomous sites.

In Order to satisfy the new information management requirements of advanced and complex organizations, a strong information integration system must be designed and developed, serving the need for information integration and interoperation among these organizations.

In this context, the work described in this thesis focuses on the design and development of a Generic Information Exchange (GIE) System. The tool supports a wide variety of existing applications with efficient means for their interconnection and interoperation, while preserving their heterogeneity, distribution, and full autonomy.

i. Heterogeneity refers to the fact that each database may apply its own distinct DBMS, and data representation is heterogeneous in terms of structures and semantics at every site.

ii. Distribution refers to the storage and processing of information from distributed data sources, located on different host computers.

iii. Autonomy refers to the fact that each database within the federation community is an independent database system. Typically, a local database is pre-existing to the creation of a cooperation network and has its own administration policies, and users community.
The distinctive features of the GIE System integration approach resides in: (a) the specific combination of database standards and Internet middleware with the fundamental research approaches, and (b) the way in which they are deployed and inter-linked within the specific components of the GIE. These two considerations make the GIE approach distinct from all other existing federated/integrated approaches, and introduce GIE as a generic solution providing a flexible architecture, and an open facility for integration/ interoperation among heterogeneous, distributed, and autonomous sites.