Federated Information Management for virtual enterprises

Garita Rodriguez, C.O.

Citation for published version (APA):
Chapter 2

Distributed Information Management Approaches for Virtual Enterprises

2.1 Introduction

The Virtual Enterprise (VE) concept represents an interoperable network of pre-existing enterprises that collaborate by means of specific IT components, towards the achievement of a common goal (see also [42]). In principle, these enterprises can function together and be regarded as a single organization, for a determined period of time, until the common objective is achieved or until the enterprises decide to dissolve their cooperation. In most cases, the driving force for enterprises to join such collaborations, derives from the emergence of a business or market opportunity, whose fulfillment would not be feasible for a single enterprise under normal circumstances. The Virtual Enterprise goal becomes viable thanks to a global management of activities and coordination of a selected set of resources and services that are made available by individual members of the VE. The VE paradigm is nowadays an active research area, for which many existing technological approaches and tools are required to be applied. In recent years, several significant research efforts and initiatives addressing this field have materialized in the form of international and European research projects and workshops/conferences [72, 70]. However, there is still a lack of common terminology regarding the concept of VE and its complete life cycle phases and elements, as well as the nature and consequences caused by the new interactions that VEs introduce in existing organizations.

Currently available Information and Communications Technology (ICT) resources and tools, such as those offered by the Internet environment, enable enterprises to share information and strengthen their interactions with other companies representing partners, clients or suppliers in different collaboration scenarios. Nevertheless, most of the current ICT developments provide solutions to only certain specific technical
problems that arise when supporting certain basic interactions among enterprises. There are still many obstacles and open issues that need to be properly addressed when supporting complex collaborations among enterprises involved in VEs. Here, one challenging case for virtual enterprise platforms is the proper sharing and exchange of information among pre-existing heterogeneous and autonomous enterprises and their internal proprietary systems. In general, without an adequate support framework for information management, it is impossible for enterprises to collaborate as a single virtual entity. Among the key problems faced in information management approaches supporting the VE domain we can emphasize the following:

- Lack of standard definitions of information models and access mechanisms. In general, the definition of common enterprise information models and data access services introduces key problems for the VE support infrastructure [140]. Existing standard representation models and data access interfaces are usually applicable to a specific field (such as ISO/STEP for technical product data representation and exchange), and in most cases they cannot be generalized to be used by VEs infrastructures in other application domains, such as tourism or agribusiness. However, VE member enterprises definitely need to adopt a common understanding in terms of semantic models and data access issues. These issues also involve aspects related to the common VE reference model, including the VE topology, VE partner roles and VE contracts.

- Support for sharing and exchange of distributed information, while maintaining the proper level of autonomy and security for each VE member. In order to successfully collaborate towards the achievement of the common VE goals, VE members are compelled to exchange and share certain information. However, the level of collaboration among the partners is in practice limited by many aspects such as mutual trust, market competitions, and bilateral contractual agreements, which need to be translated into proper authorized information access and visibility rights defined individually by enterprises. For instance, it is possible that competitor enterprises may eventually find themselves being members of the same VE. In this case, these enterprises will not be willing to expose private *know-how* information to each other. On the other hand, they may be contractually obliged to share a part of the data related to the operation of certain internal production processes. The VE information management system must address this trade-off between information sharing and protection, and must provide a flexible mechanism to define fine-grained information access rights on local information for other VE members.

- High degree of heterogeneity encountered at every VE node. Typically, when enterprises come together to form a VE, they already exist and operate using predetermined hardware and software platforms including legacy systems, which need to be partially integrated to the VE world. In order to properly support the required collaboration among VE members, the information management system must define a common approach and set of mechanisms that enterprises can adopt in order to overcome the interoperability gaps.
- Wide diversity of information technologies and tools. The VE information management platform could incorporate an extremely wide variety of technologies and tools originating from different computational fields. These technological elements need to be evaluated and harmonized in order to work together according to the ultimate objectives and functionalities addressed by the VE support platform. Examples of these related fields include workflow management techniques, distributed database systems, Web technology, middleware architectures, communication and security components, and multi-agent systems. The architecture of the VE information management component may need to be extended or adjusted in order to properly inter-operate with these related tools or applications.

- Functional generality and extensibility. The VE problem domain is deceitfully broad and complex, and the information management system should be sufficiently general and extensible to support the wide range of requirements in terms of data storage and access that are posed by different applications and VE scenarios. For example, the information management system should be able to cope (in a generic and extensible way) with new functional requirements that may be identified in the future for the support other VCL components (e.g. EDI, STEP, VE Coordination modules), end users, and different VE scenarios. Namely, it would be desirable to add new information management functionalities to the system in an orthogonal way, without investing too much effort in their development.

- Adequate performance and scalability. The information management system supporting the VE operation must provide the required level of performance and scalability, independently of the number and physical location of VE nodes, the volume of the information being exchanged, and the intensity of data access operations that need to be processed. In some VE applications, e.g. concurrent engineering, the performance aspects can become critical, and advanced techniques such as distributed shared memory and intelligent cache mechanisms may have to be considered in the implementation of the information management component (see for instance [140]).

The design and implementation of an information management system aimed at supporting VE infrastructures must address these general challenges, as well as many other specific requirements related to the particular application domain under consideration. A list of information management requirements related to VE support have also been identified in [12] and [133].

In this context, one of the objectives of this chapter is to present an analysis of several information management techniques and VE support infrastructures, that need to be considered when designing and developing an information management system for a given virtual enterprise support platform. Due to the diverse nature and complexity of the requirements described above, the intention of this analysis is to identify a set of potential technological solutions and reference infrastructures that are applicable for addressing the described information management
needs. The presented work is also useful for enterprises that are considering to join virtual organizations, since it points out some crucial ICT administration issues that will be faced by these companies.

In order to achieve these objectives, this chapter addresses two main aspects related to the VE information management:

1. It provides a survey of related information management technology, including general approaches to manage distributed information, as well as some relevant information management standards and tools.

2. It presents a study of information management approaches that have been applied in several actual VE-support research and development projects. The study of those approaches reflects the way in which existing VE infrastructures incorporate and adjust some of the standards and tools described in the first point above.

The main focus of this chapter however, resides on the second point above, for which specific relevant research projects in VE support infrastructures are described and compared according to a predefined set of information management characteristics that have been specially introduced in this chapter for their evaluation purposes [72].

The structure of this chapter is therefore organized as follows. Regarding the subject of the first point above, Section 2.2 provides a summarized description of VE related information management technology, including generic approaches for distributed information management, information representation models and standards, and other related technologies and tools. In relation to the subject of the second point above, Section 2.3 provides a survey and evaluation of existing information management approaches for VE support platforms developed in several research projects. Furthermore, for the evaluation of different approaches, a set of information management criteria is defined, and consequently, a comparison of the different infrastructures is carried out and shown in a tabular form according to this defined criteria. Finally, Section 2.4 summarizes the major conclusions and future directions of this work.

### 2.2 VE Related Information Technology

In order to address the information management requirements described in Section 2.1, there are many different technologies that can be applied. In this section, three main areas of information management technologies relevant to VE support infrastructures are introduced. The first area refers to the generic approaches for distributed information management, through which certain level of information sharing and exchange among VE nodes can be achieved. Secondly, several related information representational standards and proposals are analyzed. As will be described for this second area, a good part of the success in the adoption of a VE infrastructure by existing enterprises lies on the support of widely accepted information management and modeling standards in specific domains. The third and last related area analyzed in this section addresses several component technologies that definitely play
a crucial role in the VE information management support, including the workflow management techniques, distributed object computing, advanced Web standards and tools, secure communication techniques, high-performance computing environments, and multiagent systems.

2.2.1 Approaches for Distributed Information Management

The need to support the general information sharing and exchange among different organizations is not something particularly new nor exclusively related to the VE domain. For example, different multi-database management approaches have been historically conceived and designed in order to support interoperability and integration among independent databases, based on fundamental principles such as seamless access to distributed data, support for different degrees of software/hardware heterogeneity, interoperability mechanisms, and reinforcement of database autonomy [28].

In general, the application of heterogeneous distributed database management systems in VE infrastructures seems a natural choice because in the end, the VE itself can be regarded as a network of distributed autonomous nodes that exchange and share particular sets of local information. These advanced database management systems can provide a good base for information management in VE infrastructures, since much of their functionality can be reused in order to avoid the "reinvention of the wheel", i.e. to avoid the development of existing mechanisms for consistent management of local and distributed information in a network of heterogeneous and autonomous systems. Furthermore, aspects related to transaction management, recovery, concurrency control, and deadlock resolution techniques, which have been extensively addressed in the past by distributed database systems, can be effectively used in certain situations to support advanced collaborative applications such as those required for VE infrastructures.

However, most existing multi-database systems are still research prototypes, and the commercially available DBMSs provide only limited functionalities for the fundamental principles mentioned above [28]. In practice, the full use and exploitation of advanced multi-database management systems by typical potential VE member enterprises is not common, and the diversity of the used approaches make it even more difficult to generalize the application of database theoretical frameworks in this field. Furthermore, as mentioned in the introduction section, there are also other specific VE-related issues, which are not well covered by pure multi-database approaches.

Nevertheless, advanced distributed information management techniques provide a solid technological base that must be studied, evaluated, and eventually tailored to apply it to the VE infrastructure. A detailed analysis and comparison of particular multi-database approaches is outside the scope of this survey. Some classifications and taxonomies can also be found in [33, 146, 28]. In the following sections, a short description of some generic approaches for interoperable distributed information management, is provided as a reference.
Distributed Information Management Systems

In general terms, the “distributed information management systems” can be found within a wide variety of architectural approaches and may run on different hardware architectures, ranging from several tightly coupled computers with a centralized control, to geographically distributed machines. The data itself may also be represented and modeled in different ways. But the fundamental feature that makes these systems differ from a traditional centralized database system is that the data is partitioned and physically stored on different computers [67].

Despite the big diversity of possible classifications and architectures of these systems, after an analysis of the VE application domain to identify the information management requirements, the following approaches seem relevant to be considered in the design of a VE infrastructure support framework: distributed databases, multidatabases, and federated databases (see Figure 2.1). These approaches are briefly described below.

In the first approach, the most coupled version (e.g. in terms of control) of a distributed database management system involves the administration of several databases by one general management system. Usually, the management system and the depending databases have the same data model, provide the same functionality at all levels, and one global schema is kept where every database is represented. The users can submit queries applied to the global schema, and the management system is in charge of the distribution of subqueries between the components and the processing of the individual results to satisfy the global request. Since there is a high level of integration between the management system and its components, and all of them are very homogeneous regarding their functionality, this approach can be implemented with a good global performance. However, the same characteristics can also imply high local modification costs as well as loss of control at every remote component site.

On the other hand, a multidatabase system supports the operation of several databases, where each of these components is ruled by a given database management system. The management subsystems for every database may be different, and could be either centralized or distributed among several computers. If the management subsystem is the same for all databases, then the multidatabase system is homogeneous; otherwise, it is heterogeneous. Well-known commercially available relational DBMS such as Oracle and Sybase can be seen as examples of homogeneous multi-

---

**Figure 2.1:** Main distributed information management systems relevant to VEs.
database management systems. On the contrary, systems like Pegasus [144] can be catalogued as heterogeneous multidatabase language systems. In this architecture, the multidatabase system acts as a front end that interacts with the database components through a particular interface embedded in every subsystem. The component databases, despite this interaction with a global system, preserve their local autonomy. Furthermore, a specific kind of multidatabase system arises when one global schema is maintained and the local nodes must cooperate very closely in this task. ADDS [30] and Dataplex [54] are examples of this approach. However, the database manager of the global schema does not have control on the local schemas.

Finally, the federated database systems represent a variation of the multidatabase systems. In this case, instead of creating a single local schema, every node in the federation maintains local autonomy on the data, and defines on it a set of export schemas that are made available to other specific nodes. Also, every node will be able to import schemas from other nodes according to the defined access permissions. As a consequence of this general interaction, this approach allows the cooperation between the nodes in the federation to accomplish a common or global task, while the local autonomy and independence of every node is preserved and reinforced.

Analysis of closely related distributed systems

This section contains a brief overview and general description of some distributed, multidatabase, and federated database systems that are closely related to the approach presented in this dissertation (see Figure 2.2). An extensive classification and analysis of such systems is outside the aim and the scope of this chapter. However, the architecture and approach of some of these systems is described here, since it is definitely related to the design of the information management approach for VE infrastructures. A general survey of multidatabase prototypes can be found in [28].

The IRO-DB federated database system supports seamless access to distributed heterogeneous databases [60, 95]. The IRO-DB architecture consists of three layers: the local layer, supporting access to heterogeneous component database systems; the communications layer, to provide services for remote database and object access; and the interoperable layer that integrates the various local schemas into an interoperable schema. The interoperable schema combines data from the local databases and handles inconsistencies in structure, naming, scaling behavior and semantics.

![Figure 2.2: Examples of VE-related distributed database management systems.](image-url)
Based on the interoperability approach developed in IRO-DB, the Miro-Web project aims at the development of middleware components in order to provide integrated access to multiple data sources using Internet/Intranet protocols [63]. Please notice that architectures like Miro-Web can be used to support the operation of VEs in some domains, by integrating distributed enterprise data, based on Inter/Intranet protocols.

The PEER system [156, 15, 169], is designed and developed at the CO-IM group of the University of Amsterdam (but its development is not a part of this dissertation). PEER is a federated object-oriented information management system that primarily supports the sharing and exchange of information among cooperating autonomous and heterogeneous nodes. The PEER federated architecture consists of a network of tightly/loosely interrelated nodes. Both the information and the control mechanisms are distributed within the network. The interdependencies between the information stored at different nodes are established through specific federated schemas; thus there is no need to store the data redundantly in different nodes. Every node is represented by several schemas: a local schema (LOC), several import schemas (IMPs), several export schemas (EXPs) and an integrated schema (INT) [14]. The local schema is the schema that models the data stored locally. The various import schemas model the information that is accessible from other databases. An export schema models some information that a database wishes to make accessible to other nodes (usually, a node defines several export schemas). At last, the integrated schema presents a coherent view on all accessible local and remote information. A prototype implementation of the PEER system is developed using the C language in the UNIX environment, and includes two user interface tools: a Schema Manipulation Tool (SMT), and a Database Browsing Tool (DBT). The PEER architecture inspired the design of the federated information management system in VE support infrastructures such as PRODNET II [71] and MASSYVE [127].

The WebFindIt project applies an architecture that allows the description, location, and access of data in large networks of databases, towards the support of some kind of "world-wide database" [29]. In other words, the target infrastructure would provide the end-users and applications with one single logical database, which components represent database sites connected to the WWW. It is based on a previous project called FindIt that aimed at supporting interoperability among very large multi-database management systems. The WebFindIt prototype was implemented using Web, Java and CORBA technologies.

InfoWeb is a general-purpose information management infrastructure based on interoperable and independent components that are interfaced by specific middleware [104]. The main tiers of the InfoWeb systems are the client, the object server and the component databases. Part of the client layer interface is built with dynamically generated HTML pages, while more interactive interfaces are supported with Java applets. The object server acts a mediator between the client layer and the underlying databases. The object server defines an InfoWeb standard data model on the existing databases and supports the integration of legacy databases and Web server engines. For this purpose, a logical database interface has been defined, which can be used to integrate other external sources that comply with standards such as the SQL language.
and ODBC protocol. The end-user has access to InfoWeb through a Web browser interface, and all the distributed information that is retrieved from local and remote sources is presented in a uniform and consistent way. The Enterprise Intelligence System Network (EisNet) project uses InfoWeb to integrate internal and external enterprise information resources [104].

InfoSleuth is an agent-based system, aimed at the integration of heterogeneous distributed information sources using common ontologies definitions [115, 62]. The sources can be for instance, traditional structured databases or unstructured sources such as collections of text documents. The InfoSleuth agents collaborate with each other to retrieve and process information distributed in a dynamic Web-based environment. Agents can communicate among them using the Knowledge Query and Manipulation Language (KQML). The end user can issue queries using his/her own vocabulary based on the common ontologies, without being concerned about where the data is physically located. The application of InfoSleuth intelligent agent technology is being evaluated for VE environments such as EisNet (although InfoWeb is currently used).

As mentioned before, the general approaches and systems described in this section can provide a base for the development of VE information management infrastructures. Examples of other related federated and multidatabase management systems that may eventually be applied in these VE infrastructures include UniSQL/M [97, 21], Pegasus [144, 28], VHDBS [172, 171], VODAK [35, 28], and Disco [155]. Furthermore, several well-known and widely accepted relational DBMSs among small and medium enterprises in different sectors have also been used by some VE infrastructures, in order to support local and distributed information management functionalities. These DBMSs include Oracle, Microsoft Access and SQL Server among others.

The actual application of some of the systems described in this section as internal DBMS in existing VE platforms will be illustrated in Section 2.3 of this chapter.

### 2.2.2 Related Information Models and Standards

Besides the generic distributed information management approaches described in the previous section, there are a number of existing (and under development) information management models and standards that also need to be carefully considered for the design and implementation of the VE information management platform. For example, EDI and STEP technologies represent consolidated standards that are currently being used by many industries in certain commercial and manufacturing domains to support the execution of their daily business operations. The management of information represented in some of these standards can become crucial for the adoption of VE infrastructures for enterprises whose operations heavily rely on these standards.

Furthermore, the existence and current usage of these technologies can facilitate the integration of internal company information with the information managed by the VE platform. Some of the heterogeneity issues that must be addressed in order to exchange and share meaningful data among VE partners, have been addressed by these standards. Namely, enterprises which have adopted them already count
on common data models with well-defined semantics, and with a set of information management tools that allow the exchange of this information with other enterprises. Several VE support projects have also adopted standards such as STEP as their general model for information representation.

In other words, this kind of standards provide a working platform enabling the processing of information across company boundaries that also needs to be considered when designing and implementing more general and complex VE scenarios. In this context, there exist many standards, proposals and tools aiming at the definition of common data models and access mechanisms in different fields. It is a challenge for the information management system to devise a flexible mechanism that allows the proper support of the necessary management of data models standards in the context of the VE application domain.

These models and standards include the following among others (see Figure 2.3):

- Electronic Data Interchange – EDI.
- Distributed Business Processes – DBPs.
- Standard for the Exchange of Product Model Data – STEP.
- Ontology Models and Definitions.
- Data Model Standardization Initiatives.

In the next sections, each of these points is briefly described.

**Electronic Data Interchange**

EDI (Electronic Data Interchange) encompasses a set of standards that basically define a common format for representing and exchanging business data. The EDIFACT (EDI for Administration, Commerce and Transport) standard is an EDI family standard which is one of the most widely used today. In order to support the automated exchange of business documents among companies, standards such as EDIFACT define a framework that specifies conventions for character sets, data structures, data semantics and implementation guidelines, among other related aspects [79]. EDIFACT suggests structures for commonly used business messages, and specifies which

---

**Figure 2.3**: Main VE-related information representation models and standards.
part of the information is mandatory and which part is conditional. The messages are coded as character strings containing data elements that represent for instance, order items such as price, quantity, product description, and so forth. By adopting the EDI standard, companies have achieved to a certain extent their goal of having a mechanism for business data exchange that is independent of existing applications, communication media and hardware systems at each site. Nevertheless, some vendors and enterprises foresee important changes in the role that EDI may play in the future, specially regarding the emergence of new Web-based standards that may have significant advantages over EDI.

**Distributed Business Process (DBP) Models**

In general, the Business Process (BP) concept encapsulates a global goal or objective being carried out by an enterprise, such as the manufacturing of a product, provision of a service, delivery of goods, etc. In the context of VEs, the VE goals can be modeled as a Distributed Business Process (DBP) composed of a set of BPs, where each BP can be in turn carried out by a different VE member [99]. Clearly, the BPs that are being fulfilled separately by each VE member, must be periodically monitored and coordinated in order to guarantee their combined progress towards the accomplishment of the global DBP. The application of DBP models in the context of VEs has proven to be very advantageous in several projects, such as PRODNET, FETISH, and MASSYVE (to be described in Section 2.3.1). The DBP can also be modeled in terms of workflow management plans, as it will be discussed later in this chapter. Please notice that in order to enable the proper coordination and monitoring of the DBP activities, a distributed data model and access functionality must be defined and supported by the VE information management infrastructure at each VE member site [74]. Examples of currently available tools that can be applied to model and manage BPs include ARIS [22] and PROSIM/PROCAP [96], among other systems. Also, there exist several standards and working groups for the exchange of BP information, such as the WPDL (Workflow Process Definition Language) and PIF (Process Interchange Format) initiatives.

**STEP Standard for the Exchange of Product Model Data**

The Standard for the Exchange of Product Model Data (STEP – ISO 10303) has as main goal the uniform representation and management of product-related data during the whole life cycle of the product [141]. Different application protocols have been standardized in STEP in order to define application-specific product data models in different sectors such as mechanics, automotive, aerospace and manufacturing industries. The STEP EXPRESS language is an object-oriented-like language used to define data models that can be used in turn by different applications to interoperate based on such common data definitions. Data defined in EXPRESS can be exchanged through files using an ASCII format, which is determined by STEP as well. Furthermore, STEP also specifies a Standard Data Access Interface (SDAI) that regulates how applications can store and retrieve information from data repositories. In general,
the use of STEP applications promotes the sharing of a common product data model in which system independence, data consistency, and interoperability features are fully supported. As mentioned previously, several VE research projects have adopted this standard in particular VE application fields [142, 140].

Ontology Models and Definitions

In a few words, an ontology can be defined as a *shared understanding* of some domain of interest [158]. It encompasses a set of concepts and their definition within a given domain, as well as the interrelationships among these concepts. In the context of IT developments, ontologies may be used in order to address several problems that arise when a shared conceptual understanding is missing in a specific domain. These problems include for instance: inefficient communication between people and organizations, difficulties in requirement analysis and functional specifications, and operational mismatches among software systems and tools. In the case of Virtual Enterprises, ontologies can facilitate the integration of different enterprise information models by providing a common interpretation of their semantics. For example, the aim of the TOVE (Toronto Virtual Enterprise) project is to create a set of enterprise ontologies that can be used and understood by different applications [85]. The TOVE ontologies constitute an integrated enterprise model that can support other enterprise intelligent tools, such as production planning systems. Furthermore, the Enterprise Ontology defined within the Enterprise project, provides a collection of terms and definitions relevant to business enterprises, which can be used as an *inter-lingua* to integrate different software tools [159]. Also, in the FETISH project [8], an integrated knowledge based is developed to store and manage ontology definitions in the tourism domain, in order to enhance the interoperability among tourism organizations. Please notice that other information representation standards that are also described in this section, such as STEP and EDIFACT, play an equivalent role in terms of defining a common understanding for interoperability in a given domain.

VE Related Data Modeling and Standardization Initiatives

There are several other industrial consortia and initiatives aiming at standardizing information models and processes in specific fields. The results of these efforts usually take the form of XML Data Type Definitions (DTDs) and Application Program Interface (API) specifications. Although most of them cannot be regarded as a standard yet, they may eventually gain this official status. Therefore, these promising initiatives or proposals may need to be considered and supported in the context of VE infrastructures. As an example of these cases, the Open Applications Group (OAG) proposal for modeling business data is briefly described next.

The Open Applications Group is a non-profit industry consortium formed by many world-wide distinguished enterprises in the area of software component interoperability [116]. The mission of OAG is, as stated by the group itself, “to define and encourage the adoption of a unifying standard for eBusiness and Application Software interoperability that reduces customer costs and time to deploy solutions”. The OAG
application integration approach is based on the definition of a "Business Object Document" (BOD). The Business Object Document is the model used to communicate a service request from the originating business application to the destination business application. The BOD consists of two main parts or areas. The Control Area specifies items to characterize the sender application, the receiver application, and the business service request. The Business Data Area contains all the codes, sizes and values needed to support the Business Service Request. The BOD is in the end coded as a sequence of strings of characters, with some tag/delimiter strings to specify the structures and values of data. In this sense, it is similar to the kind of message formulated by an EDI application. Using the BOD, a large set of predefined services can be requested among different modules or even different enterprises. Finally, the OAG proposal considers not only the data structures passed among enterprises, but also the possible services that can be requested among enterprises, and the protocol functions that an OAG-compliant API should support.

Before finishing this section, it is important to mention that most of the information models and standards described here, have been developed with different objectives and for different target application domains. Namely, many of these technologies complement each other, and none of them can be exclusively used in order to design a generic and flexible infrastructure to support realistic VE scenarios in a given sector, i.e. without some of the others being involved. As mentioned before, it is necessary to synthesize them within a common framework that provides an adequate level of integration and interoperability of the involved standards.

2.2.3 Other related technologies and tools

In addition to the information representation models such as those described in the previous section, there are many other technologies, standards and tools that are strongly required and related to the information management aspects of VEs. Therefore, in order to guarantee the success of the VE application, these existing component technologies also need to be properly synthesized and integrated into the new VE framework [41]. The interactions between each corresponding component module and the information management module that supports the VE infrastructure, need to be carefully identified. The main component technologies required for VEs include (see Figure 2.4):

- Workflow management techniques.
- Advanced Web application support mechanisms.
- Distributed Object Computing.
- Secure and reliable communication protocols and services.
- High-performance computing environments.
- Multi-agent systems.

These component technologies and their relationship with VE infrastructures are briefly described next.
Workflows management techniques

The diverse and complex nature of the interoperation among independent enterprises towards the achievement of common VE goals, introduces a plausible need for advanced coordination mechanisms in order to support flexibility and configurability within the VE environment. Workflow management techniques represent a particular coordination approach that can be conveniently applied in VE support infrastructures at different levels (for a study of other general coordination approaches, see [125]). For instance, they can be applied to coordinate the sequenced execution of service requests among different VE infrastructure components that need to closely interoperate in order to process a given external event, such as the reception of a purchase order. Furthermore, the execution of distributed business processes representing the activities of the VE itself (see Section 2.2.2), can also be supported by workflow management techniques. In fact, workflow management systems represent an approach that has been extensively applied by several VE projects such as PRODNET [40], VEGA [176] and NIIIP [114]. In all these cases, the reference model defined by the Workflow Management Coalition (WfMC) [167] has been followed. In contrast with other coordination approaches, workflow management techniques present certain features that make their application to the VE environment very suitable and convenient [48]. For instance, workflow plans can be easily configured by human operators through friendly graphical interfaces, without any need for low-level programming tasks. Currently, there is a wide variety of available commercial and experimental workflow management systems and tools, including METEOR [147, 111], WIDE [84], and WebFlow [166], among many others. In some particular cases, it could be possible to integrate existing tools like these with a VE infrastructure. In addition, some examples of interaction scenarios that demonstrate the combined application of information management and workflow-based coordination functionalities within an actual VE environment can be found in [9, 74].

Advanced Web applications technology

Web application standards and technologies obviously play an important role in the support of VE infrastructures. For example, the development of VE infrastructures based on Internet communication protocols, Java applications, and XML, not only allows a high degree of portability, but also opens new possibilities to deal with the
associated enterprise heterogeneity problems [154]. The worldwide communication facilities provided by Internet represent a mature and accessible information transport layer that should be exploited by VE support infrastructures. For instance, Internet protocols such as TCP/IP play a fundamental role in many of the existing VE support architectures. Furthermore, many information sources and management services that are readily available through Internet are crucial to certain VE platforms, including: WWW servers, Internet clients, public forums, public information catalogues and directories, relevant mailing lists, and FTP servers. Namely, the VE creation and some aspects of the operation phases in some particular scenarios can be conveniently supported by taking advantage of information obtained from these sources and services.

**Distributed Object Computing**

Service brokerage and middleware interoperability approaches based on object-oriented technology such as CORBA (Common Object Request Broker Architecture [117]), can be applied to integrate external enterprise information sources with the VE platform in a flexible and dynamic way. For example, existing software modules running on heterogeneous environments at different enterprises, could be extended with a CORBA layer to be connected to the VE infrastructure, and to enable the interoperability and exchange of information among each other. In order to achieve this interoperability, the Object Management Group (OMG) has defined three main building blocks for the CORBA architecture: the Interface Definition Language (IDL), the Object Request Broker (ORB), and the Internet Inter-ORB Protocol (IIOP) [87]. Briefly speaking, specific application object types can be defined through IDL, which determine the operations and parameters that the object type instances will recognize (these object type definitions are independent of the programming language or system platform in which the actual object methods or services are implemented). The ORB is the middleware that allows a client object to transparently invoke a method of a server object, which can be on the same machine or across a network. Finally, the IIOP is the standard protocol that defines the way in which ORBs can communicate, by specifying for instance the target object, the operations, and the parameters representations required for each object type. As will be described in Section 2.3.1, CORBA architectures have been extensively applied by a number of VE research projects such as NIIIP and VEGA, in order to provide seamless access to data in distributed and heterogeneous environments.

**Secure and Reliable Communication Protocols**

In the design and implementation of any VE platform, the importance of a reliable, secure and efficient communication channel between any two VE members is clearly evident [120]. Probably, different layers of communication protocols need to be designed to support the information exchange among enterprises. The application of sophisticated data encryption algorithms, digital signatures, and other security mechanisms, is also essential in this kind of enterprise collaborations. Taking into account
that VE members need to protect sensitive know-how procedures and confidential information from other VE partners, it is mandatory to provide services that assure secure point-to-point communications between every two VE members. More specific functional requirements for the VE platform communication module include: communication availability and quality assurance, efficient management and monitoring of communication resources, reinforcement of information security and privacy, authentication support, maintenance of data communication logs, proper handling of time-stamps for communication events, correct incorporation of legal issues, and provision for status notifications and delivery reports [120]. In practice, these low-level communication functionalities are usually encapsulated in a specialized communication module, which makes them available as services for other components of the VE infrastructure, such as the distributed information management module. In this way, all the secure and reliable communication requirements are implicitly considered by the information management module, when sharing and exchanging information among VE members.

**High-performance computing environments**

Besides the secure and reliable communication services described before in this section, some VE scenarios may demand an information management architecture able to handle extremely large data collections that need to be accessed by geographically distributed users who in turn, want to apply computationally-intensive analysis on particular data sets. For instance, let us consider the case of VE infrastructures in the area of large-scale engineering, in which many end-users at different locations can participate in the concurrent design of a particular structure; e.g. a new building, an off-shore platform, a ship, an underground tunnel, etc. The size of the technical product data that needs to be exchanged among user sites for the execution of certain queries, may be sometimes in the order of gigabytes of data. In this kind of scenarios, the application of a high-performance distributed resource and data management architecture, such as the Data Grid, may be required for the VE infrastructure. The main objectives of the Data Grid are to integrate huge heterogeneous data archives into a distributed data management grid, and to identify services for high-performance, distributed, data intensive computing [51]. As such, this infrastructure may offer a wide variety of high-performance services that could be well used by VE information management infrastructures (see also [90]). For example, the application of the Data Grid is being considered in similar kinds of virtual collaborative environments, such as the Virtual Laboratory project [6]. Nevertheless, Data Grid technology is still not very stable at this moment, and is evolving at a high pace.

**Multiagent systems**

Multiagent approaches have been successfully applied in the past as a base for the data sharing and exchange mechanisms among autonomous and independent nodes. In fact, several advanced multidatabase database systems, such as the systems described in Section 2.2.1, have adopted agent architectures to achieve the required
level of interoperability and autonomy among distributed nodes. As mentioned before, these agent-based multiddatabase systems can offer a construction ground to implement VE information management infrastructures. Moreover, collaborative agent architectures have being proposed as a suitable framework to directly enable VEs, as well as supply chain management in extended enterprises. For instance, the approach presented in [145] proposes the use of a collaborative agent system to support VEs, and describes its main advantages in terms of openness, modularity, reconfigurability, scalability and robustness, among other points. In [128], another approach is presented in which the cooperation among autonomous agents is enhanced with federated database functionalities, in order to support distributed scheduling scenarios in VEs. In addition, mobile agent technology is also being applied to implement certain VE functionalities involving the intelligent retrieval of information from other VE nodes, such as partner advertising, negotiation and exchange of control data [53].

2.3 A Framework for Information Management Evaluation of VE Infrastructures

In Section 2.2, a set of information management technologies and standards were described, which can be applied by the VE infrastructure in order to cope with some of the challenges identified in Section 2.1. In this section, a framework for evaluation of VE support infrastructures is developed, in order to illustrate the way in which existing infrastructures incorporate and adjust some of the technologies described in Section 2.2. The resulting evaluation framework can be used as a common base for comparison of existing VE information management infrastructures [72]. The evaluation results are represented in a matrix format composed of VE information management infrastructures (e.g. related projects or other approaches), and the identified set of VE information management requirements and criteria used to compare and evaluate different platforms.

Consequently, the proposed approach for evaluation of VE information management infrastructures is described in the next sections, in terms of the following points: (1) a classification and description of specific existing VE platforms; (2) a set of required features and criteria for evaluation of VE information management infrastructures; and (3) a description of the tables that compare the specific VE infrastructures according to the identified set of features. The actual tables are included in Appendix B of this thesis.

A related VE comparison approach can be found in [157], where a framework for analysis of VE infrastructures is provided. However, this framework is based on the conceptualization of VEs as an extension of Electronic Commerce (EC) applications, and does not focus on specific information management aspects of different VE platforms.
2.3.1 Specific VE Information Management Platforms

In this section, a description of several specific VE projects and initiatives is provided. The VE platforms analyzed in this section were selected since they represent some of the most significant initiatives towards VE support, including international R&D projects, universities research efforts, and industrial attempts among other kinds of collaborations. In order to achieve a more comprehensive presentation of the initiatives, they need to be classified according to some criteria. For example, depending on the purpose of the analysis, the VE infrastructures could be classified depending on: specific application domains, e.g. industrial manufacturing, concurrent engineering; approaches to represent and manage the VE itself, e.g. one single VE vs. multiple VE's support architectures; or even the kind of consortium developing the infrastructure, e.g. European-Union supported vs. world-wide industry initiatives, etc. Given the nature and focus of this chapter, different projects are categorized according to the main technological approach used to manage and integrate the VE-related information that is actually distributed among the VE nodes. In particular, the projects are classified according to the following technological approaches for distributed data integration:

**Internet-based Integration Approaches.** In these approaches the distributed VE information is retrieved and integrated by making an extensive use of Internet information technology and resources, such as Web browsers and other client interfaces, public Web pages information, URLs as object identifiers, XML documents, etc.

**Object-Model-based Integration Approaches.** In this category, the integration of the VE distributed information mainly relies on the consistent application of object-oriented models and interoperability architectures, such as CORBA-based implementation approaches.

**Federated Database Integration Approaches.** This particular approach for distributed information integration is based on the principles of federated/distributed databases, including different levels of schema definitions that ultimately support the integration of distributed data (see Section 2.2.1).

**Message-passing Approaches.** Here, the VE information is queried, retrieved, and presented to the end user or application through the exchange of specific messages among VE nodes. The messages comply with a predefined format and protocol such as email or special service request functions among tools.

**Other Integration Approaches.** This category encompasses systems that cannot be properly classified under the above categories. Some platforms may use a particular implementation approach to support certain aspects of VEs, which may not be based on one particular technology. For example, some approaches may tend to support the VE life-cycle through a set of independent tools that apply different data integration strategies at different VE phases. In this case, the application of the proposed classification method would not be very suitable.
However, in most cases it is actually possible to categorize the initiatives under one of the general approaches above.

Please notice that this general categorization does not mean that a given project classified under a given approach, does not apply the other kinds of technologies used to characterize the other approaches. In other words, it is possible that for instance, integration approaches based on object-oriented models or federated database architectures, can also take advantage of Internet facilities such as communication services. Nevertheless, it is the main technology supporting their integration approach which is used as the reference point to classify the projects in this chapter. For example, the high-level access functions offered by a given VE information management system can be based on one of the above technologies, but the other techniques may be also applied at lower implementation levels.

In the following subsections, several VE projects are classified according to the presented classification approach (see Figure 2.5).

### Internet-based Integration Approaches

The VE support infrastructures presented in this section are strongly based on the Internet environment and facilities in order to query, retrieve, and integrate the distributed VE information to be used by end-user operators and applications.

**EisNet** The Enterprise Intelligence System Network (EisNet) project has as main objective the integration of diverse internal and external enterprise information resources in order to make them accessible to users in a consistent and seamless way [104]. These resources include local databases, corporate databases, and external Web sites. The EisNet system platform uses the InfoWeb™ multimedia information management system (see also Section 2.2.1) to integrate information originating from the sources mentioned above. In this way, different databases can be wrapped and

Figure 2.5: Classification of approaches for integration of VE distributed information.
incorporated to the virtual integrated environment. The EisNet system can therefore support end-user decision-making processes by integrating information coming from distributed data sources that are accessible through Internet and intranet technologies. The physical location of the data in the network remains hidden to end users, who only need to focus on the content and the meta-schema information that is retrieved for the submitted queries. Information access mechanisms were designed for the case of the local InfoWeb-managed databases, based on a user accounts/password approach. A prototype implementation of the EisNet architecture has been developed and used by a selected community of users. Further research will evaluate the incorporation of intelligent agent technology into EisNet, using for instance the InfoSleuth agent software.

**GEN** The Global Engineering Networking (GEN) initiative has been established by a group of European organizations in order to address the cooperation challenges faced when supporting virtual engineering enterprises in the areas of building construction, mechanics and electronics [133, 132]. The architecture of the information management system in GEN is composed of GEN client modules and GEN cooperation layer (server) modules. By using the GEN clients, end users have access to both local and remote data distributed in other enterprises. The Web-based client interfaces allow the search, input and management of data and metadata information. The GEN cooperation layer serves as the gateway between the internal systems of the company and the VE environment. This layer handles all the metadata information and controls the access to the information distributed along the enterprise network. In terms of data access security, the approach is based on engineer's access rights definitions. Access to the internal information is granted only to employees of the corresponding company, and the public VE information is clearly specified as such by the end user after it is introduced into the system via the GEN client. In the GEN prototype, Java and CORBA technologies were also used, and multiple operating system platforms can be supported.

**Object model-based Integration Approaches**

The approaches presented in this section are strongly based on object-oriented technology in order to represent and manage the VE distributed information. For example, several approaches described this section are based on the CORBA reference architecture, although this is not the only possibility.

**NIIIP** The NIIIP (National Industrial Information Infrastructure Protocols) project has as main objective to develop, demonstrate, and transfer technological solutions to support Industrial Virtual Enterprises [114]. The NIIIP project is one of the most representative ones in the VE domain. The NIIIP architecture provides a set of standard protocols that will support the sharing and exchange of industrial information among systems in many application areas. For this purpose, the architecture model defines three main sets of entities: a set of resource objects that are manipulated by the NIIIP protocols; a set of interface objects, which wrap existing commercial
products so that they can provide NIIP services; and a set of protocols that determine the way in which the NIIP servers can consistently interoperate. Following this model, the resources of a VE are described as objects in a VE Global Schema. The VE schema encompasses several submodels regarding the VE organization and goals, VE gateways, protocols, resource models and security aspects. In terms of the NIIP interfaces for existing commercial systems, these are grouped into components including STEP, data management, Internet tools, user desktop, and workflow management services, among others. In relation to the NIIP protocols, these are defined and organized in functional horizontal layers. For example, the top layer includes protocols for VE creation, VE resource registration, and VE operation. A given protocol in one layer may issue requests for protocol objects in a lower layer. The core technologies that support the NIIP infrastructure are Internet communications, OMG object model, WfMC for knowledge and task management, and ISO-STEP as data modeling and access standard.

VEGA  The VEGA project aims at the support of the technical operations and business activities of VEs in the area of Large Scale Engineering (LSE) projects [176, 175]. The VEGA infrastructure incorporates four main kinds of information management technologies or standards: product data modeling (ISO STEP), middleware technology (CORBA), workflow management (as defined by WfMC), and Web-related standards (HTML, VRML, Java, etc.). Namely, the representation and management of the VE-related data in VEGA is based on the STEP standard; the global VE process is managed through workflow specifications; and the data and service interoperability issues in a distributed network are addressed using CORBA and Web technology. In terms of internal architectural components, the COAST (CORBA Access to STEP repository) component represents the system integration backbone, which provides transparent access to distributed product data specified in the EXPRESS language. The COAST layer is based on the OMG standard with specific extensions. COAST also offers several distributed transaction support services in order to ensure the safe and consistent execution of concurrent operations in the VEGA platform. The DWS (Distributed Workflow Service) component supports workflow management and monitoring of the global VE activities. Furthermore, the DIS (Distributed Information Service) supports access to distributed information on top of the COAST layer, and also works as a presentation layer for human operators. Finally, VEGA also includes a Schema Interoperability Service (SIS) component to address interoperability issues among different STEP product models.

X-CITTTIC  The X-CITTTIC (Planning and Control System for Semiconductor Virtual Enterprises) project focuses on the support for VEs in the microelectronics industry [173, 25]. This project aims at the development of software components mainly addressing the planning and the control of order flow in semiconductor virtual enterprises. These tools will provide the required connectivity between existing semiconductor manufacturing systems, in order to support the planning, scheduling and control of an order flow in the context of a global VE. The distributed information management component of X-CITTTIC offers a “virtual object world” to the other
Distribute Information Management Approaches for Virtual Enterprises

system components, which allows them to have access to both local and remote data. There are two main layers composing the information management system: a top layer with one global object server, and a lower layer with several local object servers that manage the local and private data of each enterprise [173]. The information is also classified in two categories, global and local, and for each category the information is further divided as static or dynamic data depending on the persistence of the data. The different servers can communicate with other components based on several kinds of pre-defined interoperation interfaces. The information management system was implemented using CORBA, C++ and Oracle DBMS.

PerDiS The PerDiS system targets the support for cooperative engineering applications in the VE by providing an efficient and secure platform for distributed data sharing, called the Persistent Distributed Store [140]. PerDiS is based on a distributed shared-memory store for the VE applications. This means that applications located at different sites can access distributed objects as if they were stored in local memory. PerDiS incorporates a “distributed persistence by reachability” approach, in which all objects that can be directly or indirectly accessed from a persistent root, become persistent as well and are stored on disk. As applications traverse the data structures in main memory, objects are brought into memory from the local and remote nodes. The physical location of the data itself remains hidden for the user applications. In terms of data organization, the shared store is divided into clusters of objects that are accessed together. Application programmers can allocate logically-related objects in a specific cluster. Cluster objects may point to objects located in another cluster. Advanced cooperative caching and lacy replication techniques guarantee the coherence and consistency of the data, as well as the required access performance and availability. Transaction, locking mechanisms, and check-in/out operations are provided in order to control the concurrent distributed data access. In terms of security, a trust model based on users’ tasks and roles has been defined, which allows the management of access rights to data. Furthermore, the communication channels also reinforce security by incorporating different message encryption mechanisms. In order to validate the PerDiS platform, a distributed version of the Standard Data Access Interface (SDAI) of STEP was ported to PerDiS and several applications have been developed for testing purposes.

Federated Database-based Integration Approaches

The approaches for integration of VE distributed information presented in this section, have as common feature the application of federated database concepts in the design and implementation of the distributed information management system. In these cases, the generic federated database approach for information integration has been specifically tailored in order to cope with the specificities of the VE application domain. As will be illustrated next, the federated architecture approach has proven to adequately facilitate and support the sharing of information between enterprises in the corresponding VE scenarios, while at the same time providing the necessary information access rights to ensure their own autonomy and information privacy.
PRODNET The project PRODNET II (Production Planning and Management in an Extended Enterprise) had as its main objective the development of a reference architecture and a support infrastructure for industrial VEs [41]. In the general PRODNET architecture, every enterprise in the network of potential VE-members is considered to be a “node”, where each node may play different roles such as a material supplier, producer, or final consumer of goods or services. Every PRODNET node is composed of three main modules: an Internal Module, a PRODNET Cooperation Layer (PCL), and an Advanced VE Coordination Functionalities (ACFs) module. The Internal Module consists of the company’s internal systems, including its Production Planning and Control systems (PPC), and other engineering and information management systems. The Advanced Coordination Functionalities (ACFs) provide some additional features to extend the scope of the PCL, including the advanced monitoring and coordination of VE-related activities, and supporting tools for partner search and logistics operations. The PCL provides the necessary functionalities to support the inter-operation between nodes in the network, and it consists of several internal components, some of which are briefly described below. For example, the Local Coordination Module (LCM) executes and controls the activities of the PCL (defined as workflow plans), according to the policies of each enterprise. The STEP and EDI components support the exchange of technical product data and commercial order-related data respectively. The PRODNET Communication Infrastructure (PCI) module supports the actual communication channel between a given node and other nodes in the VE network. Finally, the Distributed Information Management System (DIMS) supports the PRODNET VE information management requirements, and provides an interoperability platform to exchange and cooperatively manage large amounts of data between VE members. The DIMS approach for distributed information management provides an innovative mechanism for defining access rights and visibility levels among VE member enterprises [65, 74]. This approach is based on the federated/distributed database principles [11, 71], and has its roots in the PEER federated system (see Section 2.2.1).

MASSYVE The MASSYVE (Multi-agent Agile manufacturing Scheduling Systems for Virtual Enterprises) project investigates the application of multi-agent systems in agile scheduling, and its extension towards the operation in a virtual enterprise environment [130, 128]. In this project, the HOLOS framework [129] is applied as a base for advanced scheduling. In HOLOS, a given multi-agent scheduling system is composed of a set of distributed nodes with particular capabilities, that need to exchange and process information in order to address the global scheduling problem. The HOLOS agents are configured to exchange information about production orders related to a global production schedule in a given shop floor. The information integration approach to support multi-agent systems is based on the PEER federated information management framework (see Section 2.2.1), as well as the distributed information management system developed for the PRODNET project. The integration of the multi-agent system architecture with a federated database mechanism in MASSIVE has brought several advantages in terms of for instance, the proper support for different levels of data access rights that need to be configured for other agents.
in the community. This functionality also supports the proper degree of autonomy required by each VE member enterprise.

**FETISH** The main objective of the Federated European Tourism Information System Harmonization (FETISH) project is to build a “federation” of both basic and value-added services in the tourism sector [8]. Examples of these tourism services include hotel reservation and booking facilities, car rental, organization of holiday packages, mobile access to travel information, etc. In order to be integrated with the federation, the actual tourism services are extended with Java/Jini proxies developed by the corresponding service provider companies themselves, according to predefined FETISH specifications. The service proxies can be physically stored at different FETISH nodes and a distributed lookup service mechanism has been implemented to allow end users and client applications to search and retrieve the service proxies that match a desired criteria. Therefore, external applications can use the matching proxies in order to have seamless access to the actual tourism services. The FETISH project has its roots in two other European projects: PRODNET II and InTouriSME. Basically, the VE conceptual framework, advanced workflow coordination, and federated distributed information management capabilities developed in PRODNET II are being applied to the specificities of the tourism sector in FETISH. On the other hand, the InTouriSME project (now EnjoyEurope.com) has built a common entry point to access European tourism information via the Internet, for which 40 European regions agreed on a common metadata definition for tourism information sets. These common data definitions are also being used in FETISH. In terms of information management, the requirements being analyzed in the project include the support of: catalogues for common service interfaces and data definitions, federated database features, and distributed business process support. The system infrastructure is developed using Java and Jini [24] as core technologies.

**Message-passing Integration Approaches**

This section includes examples of projects that use some kind of message-passing approach to integrate physically distributed VE-related information.

**COWORK** The COWORK (Concurrent Project Development IT Tools for Small and Medium Enterprises Networks) project aims at the development of a software infrastructure to enable Small and Medium Enterprises (SMEs) in the mechanical sector to cooperate in a distributed engineering environment [19, 78]. The objective is to promote the systematic application of concurrent engineering and co-design techniques in the target SMEs in order to achieve a significant performance improvement. The main information models that support the COWORK application are: the standardized product model, for representation of the product data handled by a single node in the network; the design process model, to describe the process and activities around the product design; and the SME Competence Model (SCM), which includes company profile information for partner search and selection [56]. Each node in the network has a database with product and process information, and possibly a local
SCM data storage (however, a centralized server was foreseen for the SCM repository). Nodes in the network are classified as main contractors (MC) and co-designers (CDs). Each node maintains information about which node is its MC, and which nodes are its subordinate CDs. Every change in the project data is stored until it is decided to be forecasted to its MC and some of its CDs (this propagation actually triggers a recursive notification mechanism). The design process model in COWORK is based on a workflow approach; however, some conceptual modifications and extensions were applied to the general workflow management model. Finally, network interoperation in COWORK is achieved by means of an Internet-based message passing mechanism.

**LOGSME** The main objective of the LOGSME project is the development of an open platform and appropriate protocols and tools to support VE environments in the area of food supply chain [89, 105]. The designed platform would provide SMEs in this sector with a set of reliable and low-cost modules to support their operation within several food supply chains. These modules include for example, the LOC.CPR module, which provides visibility of stock levels and replenishment orders for suppliers, clients, and distributors in a given supply chain. The LOG.MPT module supports basic functionalities for production planning in the food industry, including bills, inventory and production orders. The LOG.Forecast component provides facilities to carry out advanced forecasting techniques without needing the presence of expert staff. LOG.Warehouse supports applications that manage transactions on a logical warehouse based on movements generated by externally generated orders. The LOG.Simulator module is an independent tool that demonstrates the benefits that SMEs can obtain through integration across the supply chain. Finally, LOG.INFO is responsible for the local and remote connectivity between LOGSME modules and legacy systems. It provides access to information located at the local and remote VE partner nodes. Data access requests coming from other modules can be processed locally and/or forwarded to other remote nodes. Once the requested data is available, it is made accessible to the original applications through a database view (e.g. an Access database record set). Information from other SMEs in the supply chain is exchanged through some form of standard messages. For instance, external data requests are received from other enterprises via e-mail or HTTP communication protocols. The data results for external requests can comply with different data formats, which are configurable depending on the request made. This functionality is very important in order to allow enterprises to participate in multiple VE chains. The LOGSME architecture is based on standards such as DCOM and CORBA.

**Other Integration Approaches**

This section covers projects which approach for integration of distributed VE-related information is not mainly based on Internet, object-oriented models, federated databases nor message passing. For instance, the VIRTEC architecture is based on a Virtual Shop System (VISHOF) in order to manage the enterprise competence information that is required to support specific decision making processes during the VE creation phase.
VIRTEC  VIRTEC is a Brazilian project aimed at the support of the concept of Virtual Organization (VO) applied to the Brazilian SMEs environment [32, 143]. VIRTEC is based on a conceptual framework called Global Virtual Business (GVB). In this context, a VO is a network of potential partners with the purpose of setting up Virtual Enterprises. Besides the VO concept, the main business entities in the GVB framework are the Virtual Enterprise (VE), the Virtual Industry Cluster (VIC), and the Virtual Enterprise Broker (VEB). VICs are well-defined groups of companies with different expertise that are willing to cooperate with other partners when a good market opportunity materializes. The VEB is the entity responsible for instantiating VEs by using services provided by the VICs entities. A VE is then a cooperation formed within a VO, where the partners exploit a given business opportunity. Currently, the VIC members in VIRTEC include nine SMEs providing different technological products and services, and the VEB is represented by the NUMA research institute, which initiated the project itself. In terms of the information management approach, the VO infrastructure in VIRTEC is supported by the Virtual Shop Floor (VISHOF) information management system. The main use of VISHOF in VIRTEC is to manage the competence information that is required to support the VEB operation when the VE is being set up. VISHOF can also provide access to information about the shop-floor resources related to VIRTEC members in a specific region. The VISHOF architecture is based on Internet mechanisms and exchange information among enterprises.

Before finishing this descriptive analysis of specific VE information management platforms, it must be mentioned there are many other relevant initiatives and projects related to VE support infrastructures, including for instance:

- PLENT – PLanning small-medium Enterprise NeTworks [102].
- CALS – Continuous Acquisition and Lifecycle Support.
- FREE – Fast Reactive Extended Enterprise [124].
- GERAM – Generalized Enterprise Reference Architecture Methodology [92].
- MARVEL OUS – Maritime Industry’s Virtual Enterprise Linkage [109].
- CROSSFLOW – Cross-Organizational Workflow for Virtual Enterprises [83].

Nevertheless, the detailed study and analysis of these and other initiatives is outside the scope of this work.

2.3.2 Characterization and Evaluation of VE infrastructures

Based on the requirements and challenges identified in Section 2.1, and the VE related information management techniques identified in Section 2.2, a basic set of required features and criteria for evaluation of VE information management infrastructures has been defined. In Table 2.1, this set of characteristics and features are represented
as the critical points to be used for evaluation and comparison of information management approaches adopted in different projects. This criteria has been applied to the projects described in Section 2.3.1 and the results are reported in Appendix B.

The tables included in this appendix represent a classification and characterization of several projects based on the criteria defined in the previous section. The tables included in Appendix B depict the main features and capabilities that are commonly provided by VE information management infrastructures. These tables also identify certain features that are not supported by most of the systems analyzed here.

In particular, Table B.1 of this appendix shows the information and characterization of the projects EisNet and GEN, which represent the Internet-based approaches for integration of VE information according to the classification strategy introduced in Section 2.3.1.

Tables B.2 and B.3 describe the infrastructures which distributed information management approach is based on an object-oriented model, including the NIiIP, VEGA, X-CITTIC and PerDiS projects. Table B.4 characterizes the projects PRODNET, MASSYVE and FETISH in which University of Amsterdam (UvA) has been a partner and has developed the information management functionalities for their VE infrastructure. In these projects, the VE distributed information management approach is based on a federated database architecture, also designed and developed at UvA.

Finally, Table B.5 includes the characterization of the COWORK and LOGSME projects, which are mostly based on a message-passing approach to integrate the VE distributed information.

In Section 2.4 of this chapter, some general conclusions regarding the analysis of the information gathered and presented in all these tables will be included. Furthermore, a detailed evaluation of the PRODNET VE information management approach in relation to the other VE infrastructures, is provided in the next section.

**Detailed Analysis of the PRODNET DIMS approach**

The general evaluation criteria introduced in the previous section can be used as a starting point to study in more details the specific information management capabilities of a given VE infrastructure. As an example, the Distributed Information Management System (DIMS) of the PRODNET project, which was briefly introduced in Section 2.3.1, is further analyzed in this section as it is the subject of this dissertation. Here, the analysis is carried out based on the identification of three main classes of DIMS features, namely: distinctive features that are unique to the DIMS approach; common features that are supported by DIMS as well as by other VE information management systems; and at last, features which are not directly addressed by the DIMS approach.

As can be noticed in the tables of Appendix B, the information management approach of the DIMS is supported by a federated information management architecture. The DIMS federated approach represents the major distinguishing characteristic in relation to the other projects. Similar architectures are only identified in the other projects in which University of Amsterdam has also been in charge of the VE informa-
<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>General System Information</strong></td>
<td></td>
</tr>
<tr>
<td>Organization</td>
<td>Main developer/funding provider organization, company, consortium, etc.</td>
</tr>
<tr>
<td>Sector</td>
<td>Field of application (e.g. manufacturing, tourism, engineering)</td>
</tr>
<tr>
<td>Platform</td>
<td>Computer system platform and/or operating system</td>
</tr>
<tr>
<td><strong>Data Management Framework for Information Sharing/Exchange</strong></td>
<td></td>
</tr>
<tr>
<td>Data Models</td>
<td>Standard or proprietary data models applied (e.g. EDI, STEP)</td>
</tr>
<tr>
<td>Internal DBMS</td>
<td>Use of base underlying DBMS (e.g. Oracle, Sybase, O2, Access)</td>
</tr>
<tr>
<td>DB access method</td>
<td>Support mechanism for basic data access (e.g. ODBC, JDBC, HTTP, CGI)</td>
</tr>
<tr>
<td>High-level functions</td>
<td>Provision of specific high-level information management functions developed on top of basic mechanisms such as ODBC</td>
</tr>
<tr>
<td>Transactions Mgmt.</td>
<td>Support for local and/or distributed transaction management (e.g. transaction process monitors, transaction services, rollback mechanisms)</td>
</tr>
<tr>
<td>Middleware</td>
<td>Use of some kind of middleware technology (e.g. CORBA, RPC, RMI)</td>
</tr>
<tr>
<td>External Systems Integration</td>
<td>Definition of a communication protocol for external systems in order to support interoperability with existing legacy systems (e.g. specific APIs; online pull/push information strategies)</td>
</tr>
<tr>
<td>VE Information Access Rights</td>
<td>Access rights and visibility levels for protected information access among VE members (e.g. account/password approach for registered users; mechanisms to define access rights per VE partner role)</td>
</tr>
<tr>
<td>Federated/Distributed DB architecture</td>
<td>Explicit use of federated/distributed database concepts (e.g. local/import/export/integrated schemas)</td>
</tr>
<tr>
<td><strong>Specific VE management functionalities</strong></td>
<td></td>
</tr>
<tr>
<td>VE creation</td>
<td>Support for dynamic creation/initialization of multiple VEs collaborations in terms of, for instance, VE partners and their roles</td>
</tr>
<tr>
<td>VE operation</td>
<td>Support for advanced specific information management functionalities related to the VE operation cycle using VE topology information (e.g. special query functions based on VE structure; explicit management of VE and VE partners identifiers to retrieve specific VE information)</td>
</tr>
<tr>
<td>VE dissolution</td>
<td>Support for &quot;disassembly&quot; of the VE structure and partner relationships, evaluation of partner obligations, gathering of partner performance and historical information to be used when creating future VEs</td>
</tr>
<tr>
<td><strong>Other characteristics</strong></td>
<td></td>
</tr>
<tr>
<td>Directory Mgmt.</td>
<td>Directory information management for partners, products, etc.</td>
</tr>
<tr>
<td>Workflow Mgmt.</td>
<td>Application of workflow techniques to support information management activities</td>
</tr>
<tr>
<td>Safe Comm. Techniques</td>
<td>Support for communication techniques to ensure safe and reliable data exchange (e.g. encryption, digital signature, firewalls, access logs)</td>
</tr>
<tr>
<td>Internet Data Access</td>
<td>Application of Internet technology and tools to support high-level aspects of VE information management (e.g. use of friendly Internet client interfaces, Internet browsers, use of Java/XML tools)</td>
</tr>
<tr>
<td>Main special features</td>
<td>Provision of specific support for other kinds of special information management functionalities including: data warehousing, main-memory caching, metadata information, directory services, etc.</td>
</tr>
</tbody>
</table>

Table 2.1: Key characteristics of VE information management systems.

In the case of DIMS, the federated database architecture has proven to provide particularly attractive features for handling some open-issues associated with the management of information in VEs, including the support for:

- Seamless access to VE distributed information. The DIMS federated architecture defines a "VCL Integrated Schema" that is shared by all nodes and that is queried by end-users and client applications. Through this integrated schema, enterprise users and systems are able to access both its local information and
the information that is imported from other enterprises. Thus, the two kinds of information are seamlessly integrated into one coherent schema for the sake of enterprise's convenience of access and retrieval of information.

- Advanced and secure mechanisms for data sharing and exchange among VE nodes. The data sharing and exchange mechanisms offered by DIMS rely on federated query processing techniques, which provide protected access to proprietary VE information while hiding the data location details from end users and applications. Furthermore, the federated query mechanism also avoids the need for data centralization or redundancy. Consequently, the up-to-date data can always be accessed by the end-user or application queries if necessary.

- Definition of fine-grained VE access rights and visibility levels. The management of a hierarchy of federated export schemas by DIMS, supports a flexible and granular definition of information access rights among VE partner enterprises. The access rights configuration is done independently at each VE member node for every other VE partner (or group of partners), and can be based for example, on mutual trust or on VE contract clauses.

- Specific high-level information management functionalities for VE creation and operation phases. The DIMS offers a wide set of high-level functions that store and manage enterprise information according to a reference VE topology model, which is instantiated during the VE creation phase. Namely, most of these functions explicitly receive VE and VE partner identifiers as parameters, in order to reinforce the consistent application of VE paradigm.

The preceding points represent some of the novel features introduced in the DIMS architecture that are not directly incorporated in the other VE infrastructures. In addition, the DIMS properly supports functionalities that can also be identified in some of the other VE infrastructures:

- Integration of different information representation models and standards. The DIMS integrated database schema supports the representation of enterprise data in compliance with some of the standards and models presented in Section 2.2.2, including STEP, EDI and DBP models. External applications can store and access data using these formats through DIMS client interfaces.

- Interoperability with internal company systems. To properly support the integration with internal company systems, the DIMS module is extended with an interoperability layer (currently based on RPC mechanisms), through which services can be reciprocally requested and answered. This layer copes with some of the associated heterogeneity problems among DIMS and other external systems.

- Interoperation with workflow management systems. The DIMS works in close co-operation with a workflow management engine, i.e. the LCM component, in order to coordinate the execution of different PRODNET architecture components. Together, the DIMS and LCM modules represent a coordination and information management kernel within the PRODNET infrastructure.
Incorporation of safe communication techniques. By using the services offered by the PRODNET Communication Infrastructure (PCI) module, the DIMS implicitly applies safe and reliable communication mechanisms when transmitting data among VE member enterprises. The functionalities supported by PCI include: enterprise/users authentication, dynamic communication resource management, wide variety of message formats, logged information for auditing or diagnosis purposes, support for X509 authentication standard, and file signature and ciphering options.

Finally, the comparison tables also point out some features that were only partially addressed and not directly considered by the DIMS development, in relation to the other VE infrastructures. In fact, most of these features correspond to functionalities that are either not relevant and applicable to the federated architecture designed for the PRODNET Virtual Enterprises, or that are identified as work for future extensions. These aspects include:

- Directory management. The need for directories of public information related to VE support was identified quite early in the requirement analysis phase of the DIMS (see [10]). For example, for every enterprise it would be convenient to keep a directory with information describing the company profile and the role that it would like to take in potential VEs. This information would be made available for all other nodes in the network. The support for this kind of directories was not fully implemented in the PRODNET DIMS, mainly because it was out of the scope of the main objectives of the project. However, directory information management capabilities have been incorporated within an extended DIMS architecture as will described in Chapter 5.

- Distributed transaction management. For some VE infrastructures, there is a prominent need to support advanced distributed transaction management mechanisms. For example, in the VEGA approach described in Section 2.3.1, distributed transaction management is fundamental in the context of concurrent engineering environments, in order to ensure that design modifications applied to the same model by different engineers, will be carried out in a controlled fashion, and that some resolution or warning mechanism will be activated when conflicts arise. In other cases, distributed transaction management can be required to support the execution of distributed workflow plans involving multiple enterprises, which may need to be rolled backed in case one enterprise is not able to fulfil a given process activity at some point. Even though the described scenarios represent important information management issues in certain VE environments, they are certainly not applicable to the PRODNET environment and therefore, the support for advanced distributed transaction management was not a topic to be addressed in DIMS. However, the support for local transaction management was required and is provided.

- Advanced Internet data access. The development of Internet client applications and tools to access data stored in the DIMS was not strictly mandatory in the
PRODNET project, mainly because the PRODNET Cooperation Layer software was conceived to be installed locally at each enterprise, and therefore the VE information can be accessed by end users and applications through specific DIMS tools included in the local layer. In this way, given the situation of the small and medium enterprises targeted by PRODNET, they were not obliged to depend on an Internet connection to conduct businesses in a VE environment. Instead, the communication module would be able to exchange messages with other nodes using alternative means when necessary, such as communications via modems (although TCP/IP protocol is also supported). In other application domains, Internet technologies and related standards have been extensively applied in other PRODNET follow-up projects such as FETISH, including Java application developments and XML support facilities. The necessary extensions to the PRODNET DIMS architecture to incorporate these facilities will be addressed in Chapter 5.

Some of these extensions to the PRODNET DIMS platform as well as several additional functionalities related to other application domains, are further addressed in Chapter 5 of this thesis. Please also notice that the main achievements and conclusions about the VE information management approach followed by DIMS will be described in Chapter 6.

2.4 Conclusions and Future Directions

In this chapter, several information management challenges for the specific field of VE support infrastructure were introduced. In order to harness the involved complexity in this application, a survey of VE related information management techniques was presented, including: (1) distributed information management techniques, (2) related information representation models and standards, and (3) other technologies and tools closely related to VE support. This kind of analysis provides the necessary base for the proper design and development of the information management system for different VE support platforms.

Furthermore, a set of international VE projects were described and classified in terms of the main approach being used for integration of the VE distributed information. The projects were also analyzed against a specific criteria for comparison and evaluation of their different features. The results of this analysis were arranged and presented in a tabular form. The definition of the criteria parameters and the evaluation of the information management approaches proposed by some important research projects in the VE application field, represent one of the major contributions of this survey. Namely, the presented analysis of the projects constitutes a representative survey of the state of the art regarding the application of information management standards and technologies in existing VE support platforms. Also, the evaluation framework introduced in this chapter, can be used to characterize and compare other existing and future information management platforms for VEs.

As main conclusions in relation to the analysis of VE related information management techniques addressed in Section 2.2, the following points can be highlighted:
• Existing heterogeneous multi-database systems and approaches provide a solid base to support the required information management functionalities in VE infrastructures. These systems offer advanced mechanisms for management of information distributed in a network of heterogeneous and autonomous nodes, which could be enhanced to support VE networks to some extent. However, there are several specific VE-related issues that are not well covered by pure multi-database approaches.

• A crucial issue in the design of the VE platform is the proper incorporation of existing official and de facto standards for data models and information exchange mechanisms in the target application domain. Relevant standards should be evaluated and supported in order to reinforce the openness and interoperability of the VE infrastructure.

• Many other constituent technologies complementary to the information management are readily available to support the implementation of different aspects of the VE paradigm. For example, technologies such as workflow management, Internet, and distributed object-based architectures play an important role in the design and implementation of VE information management facilities. However, they need to be properly integrated and harmonized through a common VE infrastructure.

Regarding the characterization and evaluation of the VE information management approaches presented in Section 2.3, there are some global conclusions that can be drawn by analyzing the comparative tables:

• The characterization and comparison of existing VE information management approaches is a difficult task to achieve, mainly due to the wide diversity of VE definitions and models being used and applied by the described projects.

• In most cases, the approaches for distributed information management found in the evaluated projects are based on the common acceptance and consistent application of specific information models and data access functions used by enterprises within a given domain. This situation contrasts with more "theoretical" approaches based on a generalized use of multi-database systems supporting mechanisms such as bilateral (or multi-lateral) negotiations for heterogeneous database schema integration. In other words, the application and support of generic multi-database schema integration approaches is rarely encountered in VE infrastructures.

• Nevertheless, it should be noticed that some projects have successfully adopted and customized a federated database architecture in order to provide seamless access to VE distributed information, and to offer a flexible mechanism to define information access rights among VE member nodes. In these particular cases, the federated database components have been specifically tailored to handle the complex interoperability and functionality requirements set by the VE paradigm.
• Furthermore, a high percentage of the reviewed projects have focused on VE domains related to industrial manufacturing and concurrent engineering. This may be due to the fact that enterprises in those domains have faced for many years the need to support some kind of VE collaborations. Also, common data models and access mechanisms, such as STEP and EDI, have been extensively used in this context for a long time. Of course, other application domains, such as tourism and agribusiness are also being addressed by some VE projects, and it is expected that the VE paradigm will be incrementally applied to other areas.

In terms of future directions, the following issues can be mentioned based on the analysis of the described projects:

• It is clear that there is still a sensible need for a common reference model for the VE paradigm, encompassing different functionalities along the full VE life cycle. Currently, there are too many independent efforts with distinct points of view, and the definition of a common VE conceptual framework has not been achieved yet.

• In terms of Web technology, it can be noticed that at present, only a few projects have taken advantage of XML as a standard to represent and exchange data between VE member nodes. This can be explained considering that most of these projects started long before XML became widely recognized. However, a more extensive support of XML by VE platforms is foreseen in the near future.

• In order to facilitate and promote the data exchange among enterprises, there is also a trend towards the establishment of commonly-agreed data documents, based on for instance, standardized Data Type Definitions (DTDs) in a given application domain. Here, the use of XML will again play an important role. Furthermore, the definition of these DTDs in relatively new application domains, may be better supported in the future by formal ontological methodologies.

• Also, given the wide variety of existing interchange data formats (even considering a given sector), it remains a challenge for the VE platform developers to design and implement an information management mechanism that can accommodate different formats in the most flexible and generic way. A more extensive use of metadata models may contribute to the design of extensible VE information management systems.

• The concept of access rights and visibility levels in VEs is also an issue that needs to be better supported and reinforced by most VE infrastructures. Most of the analyzed VE information management systems do not support a proper level of granularity and flexibility regarding the definition of access rights on shared information among VE members. Usually, the need to support autonomy and to protect internal enterprise information is well identified, but the implemented mechanisms do not address properly this requirement.

Other VE related challenges and future developments have been identified in [38], including: improved mechanisms for integration of enterprise applications; standard
models for certain types of information such as quality-related data; better support for creation and dissolution phases of VEs; definition of VE contract frameworks; application of telework techniques; and development of generic software installation and configuration tools.

As a final remark in relation to the complexity associated with the VE paradigm and its enabling technologies, it is also foreseen that enterprises will more frequently need to use information management surveys and evaluation frameworks similar to the work presented in this chapter, in order to successfully join and adapt to existing VE platforms.