

AN INFORMATION SYSTEM FOR THE BUILDING INDUSTRIES: A COMMUNICATION APPROACH BASED ON INDUSTRIAL COMPONENTS

A.F. Cutting-Decelle^{1*}, A.M. Dubois², J.E. Dubois³

^{1*} Université de Savoie/ESIGEC/LGCH, Savoie-Technolac, 73376 LE BOURGET DU LAC, France

Email: afcd@univ-savoie.fr

² CSTB, BP 209, 06904 SOPHIA-ANTIPOLIS, France

Email: am.dubois@free.fr

³ Université Paris VII/TODYS, 1 rue Guy de la Brosse, 75005 PARIS, France

ABSTRACT

This paper presents the SYDOX/MATCOMP/Xi project, funded by the French Ministry of Industry. The goal of the project was to provide the construction professionals with an on-line aid for component specification and selection at different levels of the construction life cycle. This two years project started in 1997 and involved several partners.

This paper describes the main features of the information system: databases, query and communication systems. SYDOX (Système de DONnées complexes) is aimed at defining and demonstrating a prototype to access information about MATerials and COMPonents used in construction, implemented on a WWW server. Though the objective is general, the work was focused on a restricted sub-section of the construction domain. We describe the domain and the scope of the project, the starting point and the lessons learnt from the development of the prototype. We also propose some important ideas on which this research is based.

Keywords:

Interactive information system, Product and component databases, « Smart » access, Building and construction, WWW, Communication, Product information, Complex information, Ontologies.

1 INTRODUCTION

Construction activities (CONST) have been and still are essential to our societies. Their wide scope and diversity are such that even systemic analysis does not help so far in successfully handling their complexity. Improvement and dramatic progress in the industrial processing of components (COMP) and materials (MAT) have led to new building processes. Many aspects of the CONST have greatly benefited from specific computer programming but it is probably too early to consider the early creation of a global CONST Information System (IS).

A general system SYDOX designed to deal with Complex Data and Information (DOX) is used in the project to facilitate and enhance communication between CONST experts, designers, manufacturers and end users. In this SYDOX/MATCOMP/CONST project we combined two strategies to generate an optimized representation of the MATCOMP information and its attributes, both for storage (archiving) and intelligent access.

The first one is a bottom-up creation of MATCOMP data and information representation carried out to produce a practical CONST ontology and thesaurus recognized by all experts in the field. To downsize our work we have limited the project by choosing only a few CONST components. In this paper, we analyse the problems associated with the *window* component as a typical example of a building component.

In the second strategy, we use a top-down holistic mechanism: in practice we performed some CONST simulations or specific scenarios through which we identify the part played by CONST data at different levels and times in the CONST cycle. As we foresee exchanging and using external data and information we stress the importance of standardizing materials, products and processes data in our evaluation: special attention was paid to the standardization work undertaken in the domain of the *Industrial Data* within the framework of the ISO TC 184/SC4 International Committee (ISO TC184/SC4), leading to the standards STEP (ISO, 1994), P-LIB (ISO, 1997), MANDATE (ISO, 2000) and EDI (ISO, 1999-1), and also to other kinds of regulatory documents, such as technical agreements and regulations. Attention was also paid to the work done within the framework of the IAI (International Alliance for Interoperability) that led to the development of the IFCs (Industry Foundation Classes) ([Integrated Architectures, 2000](#)).

These two approaches set up reliable tools for the definition of any manufacturing internal database subsystem allowing to make also the best use of external information located elsewhere in the heterogeneous framework of the information CONST Space. Besides, the integration of public and private information can be simplified through good decisions at the representation level.

The initial objective of improving communications between experts and companies of industrial CONST components proved to be a powerful approach to the sector, leading for the future, since their participation increases in modern building concepts. In fact in the CONST complexity the part of the industrial MATCOMP increases as components more and more become key-elements of the modern building processes. The MATCOMP/CONST system is in fact a significant contribution for the future CONST/IS mainly because the industrial contributions in Building are becoming essential.

2 THE SYDOX/MATCOMP/XI PROJECT

SYDOX/MATCOMP/Xi, as an information system (IS), is analysed through three axes:

- *a database approach*: as an IS relies on a database structuring a collection of basic information needed by the system;
- *a query approach*: to help formalizing interrogations on the DBs;
- *a communication approach*: to define the whole set of communication procedures needed by the important heterogeneity of the storage/retrieval facilities, the communication protocols and the geographical location of the computers.

2.1 General context of the proposal

The context of this proposal is related to the development of information systems (IS) more and more needed by industrial companies for acquiring, structuring, exchanging complex technical data they have to handle during the production process. A second complexity axis must however be added to the intrinsic complexity of the data: this axis comes from the relational structuring of the data, necessary to enable a selection of competitive solutions able to answer given specifications.

This information system is mainly aimed at SMEs (Small and Medium size Enterprise), since they often have to face situations for which they do not have either the necessary skill or the tools enabling a permanent follow-up of the normative texts, or a continuous updating of the technical information needed by the projects they work on.

Fundamentally, this need is common to numerous industrial sectors. The methodology carried out within this project could be applicable to different industrial sectors, with possible consequences on the level of complexity of the IS: a selection will probably have to be made between the complexity related to the degrees of freedom in the choice of candidate solutions and the technological level of the sector: high level technology and low degree of freedom or else low level technology and high degree of freedom.

The construction industry, prior to the construction on the site in itself, is characterized by:

- an increasing complexity and the acceleration of the relations among the partners;
- an increasing diversity of the information and data handled, mainly due to the development of new representation structures (message standards such as EDIFACT messages, product or production management data representations such as the ISO 10303 STEP, 15531 MANDATE and 13584 P-LIB standards) and also to the use of specification languages such as the ISO 18629 PSL standard ([ISO, 2001](#));
- the development of new software tools able to deal with more and more important amounts and diversity of information.

2.2 Objectives of the project

SYDOX (« *Système de données complexes* » - complex data management system) defines an « intelligent » software system built on an integrated knowledge based approach. The use of this system in the domain of (construction) products and materials (SYDOX/MATCOMP) must provide the user with key-information, including innovative materials and their use.

SYDOX/MATCOMP/Xi can be viewed at several levels:

- SYDOX: software system enabling an access to complex documents;
- MATCOMP: applied to the domain of materials and components;
- Xi: application to a specific sector: here the construction sector.

The aim of the project is thus the development of an « *information access point* », in the form of an Internet WWW server enabling:

- dissemination of basic knowledge essential for industrial companies which have to deal with:
 - ◆ the organization and processing of their internal data;

- ◆ the use of data coming from clients, sub-contractors, professional associations, public bodies;
- updating and dissemination of information related to the status of new standards and rules (decisions or discussions able to generate subsequent modifications);
- development of relations between « providers » and « users » of technologies and tools for complex technical data structuring and processing;
- training by means of interactive examples of the realization of products using entity-relationship or object oriented modelling: industrial subsets or components (materials, semi-finished products, structural subsets, ...).

In order to carry out an interactive information system, as quickly as possible, but also to show the feasibility of the basic concepts of the methodology, we proposed to limit the application field of the prototype to materials and industrial subsets used in the construction sector, only focusing on upstream production processes. We think that it is in this sector that companies (mainly SMEs) have to face the most important needs in terms of data structuring systems.

The objective of the project is to develop both a learning (basis of information) and a know-how (basis of methods) allowing industrials to structure their data related to materials/components (MATCOMP) according to their needs, in France and in Europe. In this project, we have also analysed the methodology leading to the elaboration of the information and communication system.

The result of the project consists in the elaboration of a software tool prototype aimed at providing an information simple, tailored to the needs of the user and to the kind of question he may ask. It is also possible to use this software as a training tool enabling the user to learn how to use these new *Information technologies*. The software tool also provides, on demand, pictorial or additional information relating to a given product.

2.3 Structure of the project and partnership

2.3.1 Organization of the project

To fulfil its role, the project has been organised through four tasks, which are:

- **Task 1: Data identification and structuring: state of the art**

whose objective is to make an inventory and a compilation of the existing data bases and information sources. The aim of this task is to prepare the elements enabling an identification and modelling of the concept of complex data related to materials and their associations as components for the other tasks. This work relies on the good knowledge of the construction domain, as provided by the Technical Research centers partners of the project.

- **Task 2: Identification of methods and tools necessary for data representation and structuring**

The aim of this task is to identify, when they exist, the corresponding software tools. One of its functions is also to help SMEs to follow the evolution of new electronic communication means and their possible use.

- **Task 3: Formal specifications of the information system**

This task is aimed at providing the main result of the project, through the generic way defined within the project and used for structuring the information.

The potential users of the IS appear through usage scenarios, defined as sets of typical requests corresponding to the type of information generally asked for by these users. Those types of potential users are grouped together in « *user classes* » whose specificity is to share the same type of queries, the same kind of information defined in the same way.

The task is also aimed at defining the specifications of the links with other networks or already existing information servers.

- **Task 4: software prototype of the information system**

The main objective of this task is to build on an interactive software system, running in a client/server mode, with several input levels, thus enabling a use by several types of users. Its aim is also to enable a short term validation of concepts and methodology defined by the other tasks of the project

2.3.2 Partnership

The different partners and their role in the project are:

- * **CODATA France: (scientific coordination)**

- quality of the information system and development of the software tool;
- important skill in the domain of data representation.

- * **University of Savoie/ESIGEC/LGCH (« *Laboratoire Génie Civil et Habitat* », Civil engineering and Housing Research Laboratory)**

- modelling and representation of complex data
- important knowledge in the domain of building product and process modelling

- * **Technical Research Centres:**

- CSTB (« *Centre Scientifique et Technique du Bâtiment* ») ([CSTB, n.d.](#));
- CTBA (« *Centre Technique du Bois et de l'Ameublement* »): for timber structures and furnitures ([CTBA, n.d.](#));

- CTICM (« Centre Technique Industriel de la Construction Métallique »): for steelwork construction ([CTICM, n.d.](#)).
 - for their overall skill of the professional actors of the sector, thus enabling formalising their needs with the terms, the vocabulary they really use.
 - their help has been of the biggest importance, particularly in the task 1, to help formalizing the lacks in terms of the access to the information met by the SMEs of the construction sector.
 - important help in the domain of the identification of the information sources, related to the different products, components and materials they work on.

3 STATE OF THE ART OF THE INFORMATION HANDLED IN THE CONSTRUCTION SECTOR

The starting point of the elaboration of an information system lies in the capacity to identify, to take into consideration and interface, when it is possible, existing document repositories or product databases, whatever their structuring and/or geographical location ([Dubois, Cutting-Decelle, Fies, Lanvin, Bonfils, Chabrolin et al., 1998](#)).

The project started with an analysis of the data intervening during the production process of a building, in several ways:

- identification of the needs of information in the building sector, through the results provided by a previous sociological study, whose aim was to analyse the ways used by professionals of the sector to access the information they need;
- identification of the professional stakes of an information related to products, materials and components:
 - ◆ economic stake: information on products and materials becomes a key-feature of the cost control;
 - ◆ role of asserting the professional identity of the user/prescriber of this information, as he becomes able to choose pertinent solutions in terms of products;
 - ◆ creation of a new skill within the company: the owner of this « new-type » information will more easily adjust to the new economic constraints.
- usual practices in terms of research of information related to products and materials: more generally, those practices are made of a combination of different information sources resulting from three main poles:
 - ◆ permanent follow-up of information about products and materials;
 - ◆ practical information related to specific points;
 - ◆ information necessary to take decisions: generally about the real cost of the product.
- legal aspects of the information: as mentioned in multiple rules, codes of practice or other legal documents whose application is mandatory.

In a second stage, we tried to locate the data related to materials and components within the overall building project. We also explored several ways of classifying information, using:

- classifications:
 - ◆ the ISO/DIS 12006-2 standard, developed within the framework of the ISO TC59/SC13/WG2, about the « *organisation of information about construction works* » ([ISO, 1999-4](#));
 - ◆ professional classifications, such as CI/SfB, EPIC, UNICLASS, FARTEC and BATIBASE (in France).
- meta-data: such as the « *Dublin Core* », or else the structuring provided by Meta Data Coalition ([Diffuse, n.d.](#)).

The inventory of the sources of information has been done through the research of a possible « categorization » of the data by means of a table whose lines and columns represent:

- **columns:** possible « *approaches* » of the construction, in terms of life cycle stage, features of the construction, leading to the following headings: *constitutive elements, general features, achievement, regulation, environment, certification, safety, logistics*;
- **lines:** « *themes* » proposed for the information sources: based on a functional approach of the building, in terms of adaptation, structure, envelope, enclosure, facilities, finishing. Those lines have then been developed further, in order to refine the research and increase the adequation between the theme of an information source and the different headings of the columns.

The questionnaire built on the basis of the previous table has then been sent to the different Technical Research Centres partners of the project, in order to be completed with the available information, together with the indication of the numbering mentioned in the table, and the forms to be completed, each of them corresponding to an intersection (line-column) of the main table.

The answers to the questionnaire mainly apply to the following materials, products and components:

- wood, and all products based on wood;
- concrete and derived products;
- steel profiles;

- regulation about the products;
- classifications, nomenclatures.

An in-depth analysis of the answers to this questionnaire highlighted several common characteristics:

- the important number and diversity of information sources;
- generally, these information sources are rarely computerized (notably for the French sources, in 1997): the main part of the information is essentially accessible through paper documents: this was the most surprising feature discovered through this project;
- however, the current situation is evolving: development of computerized catalogs, CD-ROM, use of the Internet/WWW capabilities and more generally access to Web catalogs, even though this access, most of the time, is not free;
- the big heterogeneity of the information sources, with important lacks of consistency among the information contained.

As a consequence of the low enough level of computerization of the information sources related to the products and materials used in the construction sector, it seemed to us interesting to develop a specification document about « *requirements for a SYDOX methodology for data representation* ».

Since most of the (current) catalogs of products are often paper documents, the use of a computerized version of the information contained in these catalogs by an information system such as SYDOX/MATCOMP/Xi needed a structuring conforming to the model defined within SYDOX. The aim of this report is to describe the « *SYDOX model* ».

4 THE SYDOX/MATCOMP/XI INFORMATION SYSTEM

In this section, we describe the information system as developed within the framework of the project.

We start with the presentation of the methodology used for data structuring. In the second part, we present the general structure of the SYDOX/MATCOMP/Xi information system, that is the logical schema of the database, the user interface, then the software prototype that has been developed.

4.1 Methodology used for data structuring

The methodology used for structuring the data provided by the analysis of the answers to the questionnaire is close to the structuring proposed by the *American Concrete Institute Committee 126 « Database Formats for concrete Materials Properties »*, described in the report « *Proposed format for data on cements in a material properties database* » ([Kaetzel and Galler, 1997](#)).

In the NIST document, the organization and structure provided for the common format for representing data about cements is aimed at defining a framework for cross-referencing cement properties, data and other information. This format seemed to us, at the same time, generic enough, powerful, simple and well suited to computerized databases and WWW interfacing we wished for the system to be developed to be customised to our needs.

- Basic principles for data structuring:

Information related to products and materials is represented using similarly complete sets of data, all of them together constituting the material and product database.

For structuring the information, we used the following concepts:

- **object**: represented by the item selected on a « column »: highest level of the structuring hierarchy;
- **constituent**: 1st level detailed information about the object;
- **meta-segment**: set of information of the same nature related to a constituent;
- **segment**: category of information used to subdivide the meta-segment: table of data, definitions, figures, if necessary;
- **meta-element**: collection of information of the same kind related to the same segment;
- **element**: item of information (if necessary).

The following schema (see Figure 1) gives an example of this structuring for the following concept: « *shape* » of a « *window* ».

Example of data structuring

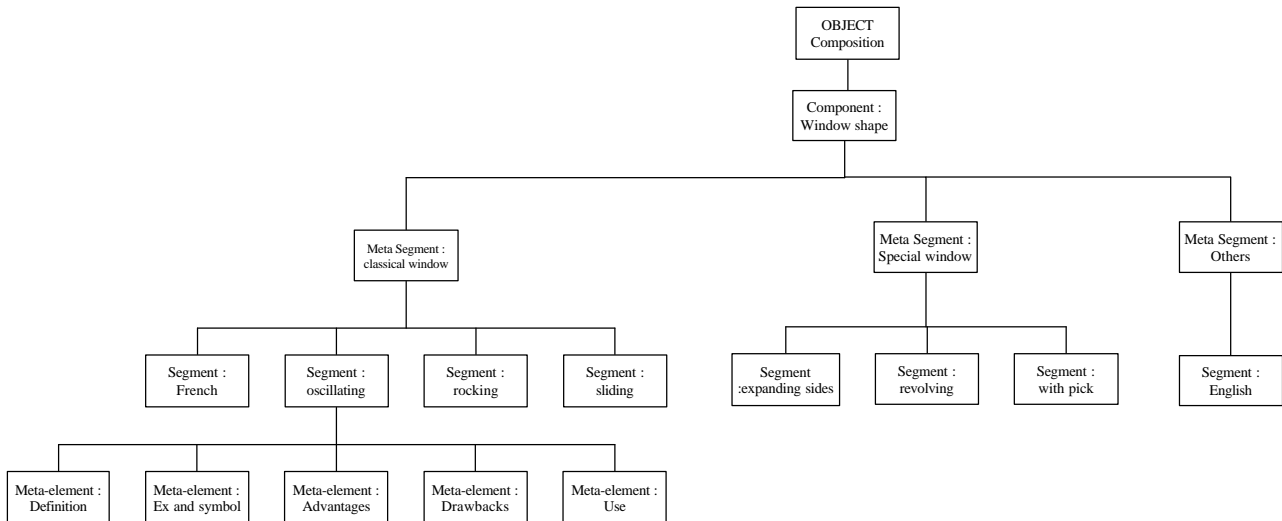


Figure 1. Example of data structuring: shape of a window

For each concept of each level, the number of related sub-concepts has to be kept low enough, in order to keep visible the possible selections (headings) on the different menus of the screen.

4.2 Structure of the SYDOX/MATCOMP/Xi information system and software prototype

The SYDOX/MATCOMP/Xi information system is an interactive, hypertext software system, accessible through HTML browsers such as Netscape™ or MS-IE™, provided that they are able to handle Java applets ([Dubois, Cutting-Decelle, Fies, Lanvin, Bonfils, Chabrolin et al, 1999](#)).

The general structure of the software prototype corresponds to schema represented Figure 2, with the following features:

- the **database**: containing the information about products and materials;
- the **user interface**: defining the way of accessing the system, according to the type of user: user wishing to get general information on products, use by an architect who wants to get a precise detail about properties of a given product, use by a technical engineer to get technical information (e.g. thermal behaviour), use for new construction, or for refurbishing, ...
- the **information menus**: they mention the headings, considered as possible choices (and corresponding, in the database, to meta-segments, segments, meta-elements). Those headings mainly result from the analysis of the answers to the questionnaire sent to the technical research centres partners of the project at the first stage of the project;
- the **scenarios**: their aim is to pre-define « *navigation routes* » within the information system, defining a kind of « coloured roadmaps » that users must follow. Today, those « routes » are pre-determined, since it is the first validation of the system. Addition of dynamic features to the current software system would allow to decrease the rigidity of these scenarios. For the examples mentioned in the prototype, and described here, we decided to choose scenarios belonging to several stages of the life cycle of the construction in order to increase the validation of the prototype software;
- the **data dictionary**: the aim of this dictionary ([Rumble & Smith, 1990](#)) is to provide a correspondence between the concepts handled by the scenarios and the data as they have been put (and structured) in the database. For a simple correspondence, it is possible to identify a direct equivalence between the concepts. However, for more complex concepts used by scenarios, the translation is not necessarily obvious, some cases may need to build on a kind of « *grammar* » defining composition rules for the words of the dictionary, and the nature and/or level of semantics accepted, but also the way of representing this semantics.

The main developments of the software prototype have related to specific construction products such as a window (fully developed example) and possibly flooring systems, for the following reasons:

- since the aim of the project was to validate a methodology, it seemed to us better to restrict the scope to a given product, but to fully develop all the information related to this product;
- we also tried develop a second example, in order to be sure that the behaviour of the first example was not a unique case;
- the validity of the answers provided by the IS heavily relies on the amount of information put in the database. Since this information did not previously exist under this format, we have also been obliged (in 1998) to build on the contents of the database, that is to spend a lot of time in trying to fetch, collect, assemble, then structure the information according to the DB format ([Mommessin, Cutting-Decelle, Dubois & Dubois, 1999](#)).

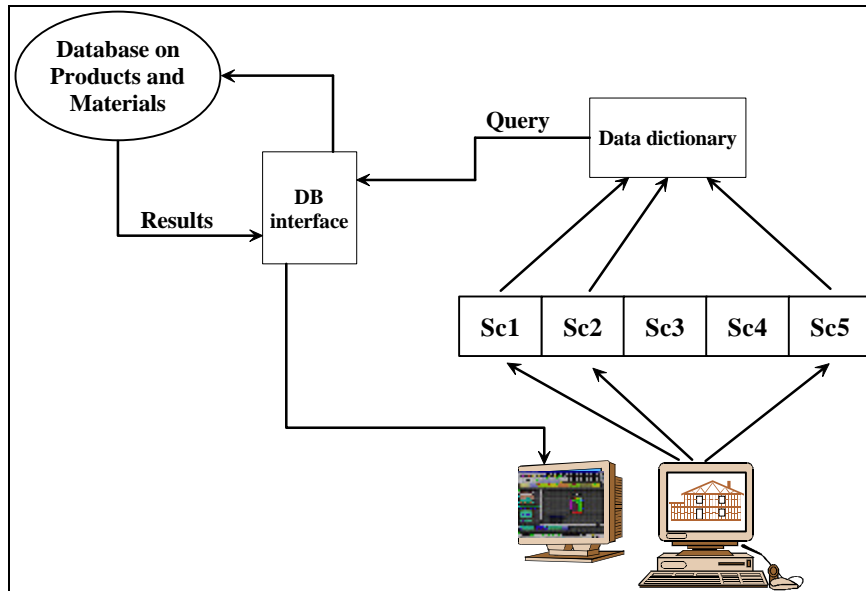


Figure 2. Structure of the SYDOX/MATCOMP/Xi information system

The Figure 3 below shows an example of the screen corresponding to the selection of the following headings (for the selection of a window):

- upper menu (corresponding to « object »): « **composition** »: (constitutive elements);
- left menu (corresponding to « constituent »): « **formes** »: shape;
- lower menu (corresponding to « meta-segment »): « **classiques** »: classical;
- right window (corresponding to « segment », or else « meta-element » or « element » depending on the structuring of the information): contents of the information about « *composition/formes/classiques* »: properties, usage, characteristics, ... in the form of text and schemas.

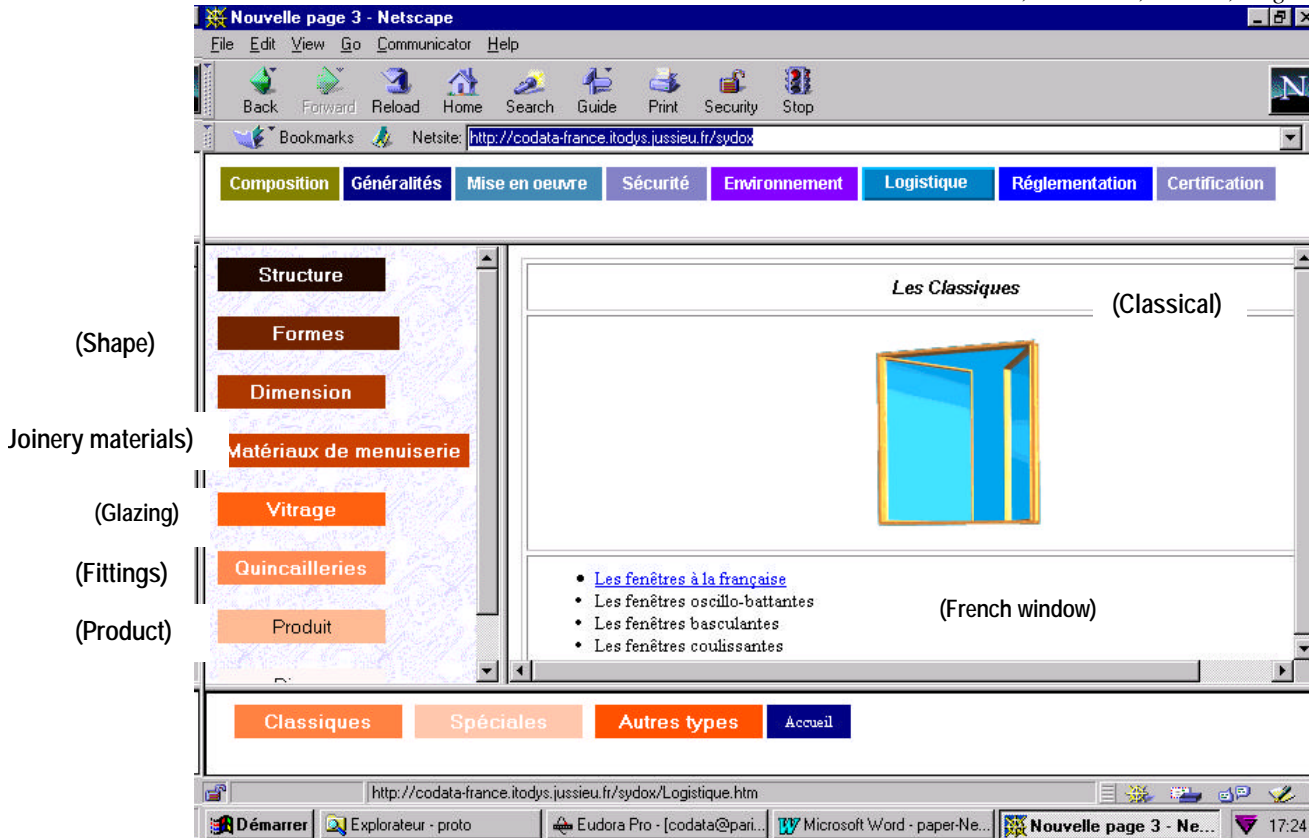


Figure 3. SYDOX/MATCOMP/Xi information menu for a given selection of headings

The information system also offers the possibility, once selected *Composition/Produit* (constitutive elements/product) then « fournisseur » heading (supplier) to get a list of products classified according to given properties, as shown on the Figure 4 below:

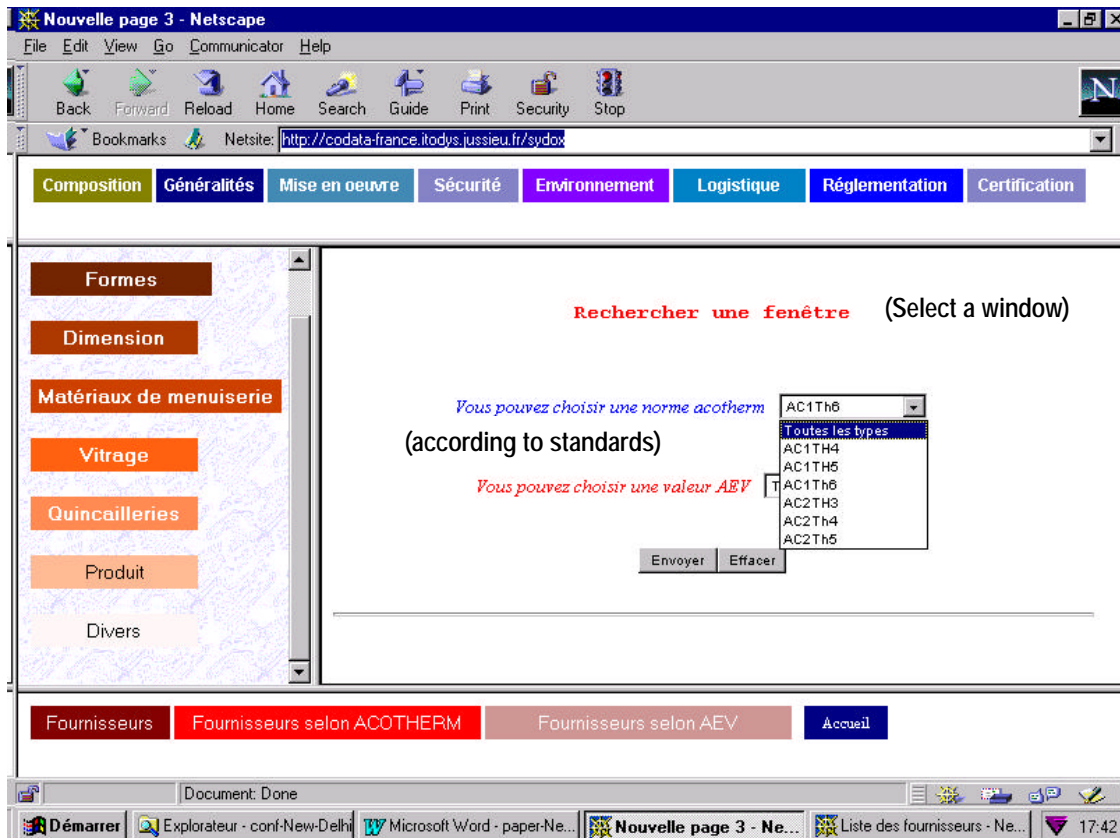


Figure 4. Selection of products according to given properties

To date, several scenarios have been developed^(*), they have been translated into the DB format (thus enabling queries using the DB format), through the use of the data dictionary and the corresponding grammar. The following scenarios are available (they are not detailed in this paper):

- for windows:

- * scenarios for an architect: new construction and refurbishing;
- * scenario for a thermician: thermal evaluation of a window.

- for floorings:

- * scenario for an architect;
- * scenario for a contractor;
- * scenario for a client.

Their aim of these scenarios is to help the user of the software system to select a component (window, or flooring system), taking into account considerations imposed or restrictions on this product due to the nature and the specifications of the construction.

5 OPERATION OF THE INFORMATION SYSTEM

It is interesting to try to analyse some of the matters arising from the development of the SYDOX software prototype:

The first comment is related to the nature of the industrial sector to which this project refers: the building and construction sector, characterized by the number and the diversity of the actors involved in the construction process.

In this sector, the expertise and the communication between the actors are essential. Consequently, a will of computerization of the sector requires work of different types:

- definition of data models for « strong » data in order to enable their circulation through various software environments;
- availability of « complex » data (as a kind of « *data self-service* ») enabling an expert interpretation of the specificity of the language used by the actors involved in the construction process;
- circulation between those data, complex data leading to the generation of « strong » data, validated by the actors, able then to enter professional domains;
- integration of all the interfaces into a same environment enabling an on-line access to the different knowledge bases.

Today, it becomes more and more necessary to adopt a pragmatic approach, that is to accept the complexity of the data as such, but also their « *evolutivity* ». This approach must also be systemic, based on classifications, thesaurus, ontologies in order to be associated to the standard data models while they are developed ([Bestougeff & Salzano, 1998](#)).

6 LESSONS LEARNT FROM THE PROTOTYPE: SYNTHESIS PHASE

Several lessons can be learnt not only from the analysis done within the framework of the project, but also from the prototype achieved. Most of them deal with the precision of the information handled or exchanged among the actors involved in construction; the diversity of this information, its heterogeneity and its complexity, the ambiguity of terms with different meanings when they are “used” by different professionals. For example, terms used in timber or steelwork construction may be different, even though those terms refer to the same piece of work described at another stage of the construction process ! Consequently, the terms describing a window may not be the same depending on whether timber or aluminium is used for the framework and also depending on professionals.

6.1 Ways of using the information system

Since the IS may be operated in several ways, the information contained in the answers must be adapted to the type of questions asked.

There are two main types of interrogations of the IS : direct requests and scenarios:

- *direct question* (or request for further information) *about a specific technical or regulatory text or else a given product, through a key-word based access*: most of the time the answer provided is simple (schema, and/or text) -- provided that the question is non-ambiguous and well-posed; in that case an answer is available from the database - no transformation is needed;

^(*) The authors wish to thank S. Genthon and E. Resident for their contribution to this work ([Genthon & Resident, 1998](#))

- *direct request for a list of specific information*: list of windows of some kind, list of rules applicable to construction on seismic zones, etc.;
- *scenarios*: in this approach, the end-user of the IS is identified by specific categories of professionals of the construction sector. The categories selected in the feasibility study leading to the prototype were: an architect (new construction, or refurbishment) and a thermal engineer. Through the use of the scenario, the end-user gets only the kind of information suited to the profile he has declared (e.g. the thermal characteristics of the window but not the rules related to the categories of materials if this window has to be used for refurbishment – for a profile of thermal engineer).

6.2 Problems encountered

Several problems were encountered during this project. Most of them can broadly be split into two categories: problems related to the non-availability of the information required by the end-user, and those related to the rigidity of the scenarios (proposal only of pre-defined “tracks”).

A particular category of problems was related to the simultaneous management of the different ontologies defined behind the objects or components proposed in the menus.

The first two categories of problems we met are:

- *non-availability of the information required*: an example is provided by the case of an end-user, non necessarily expert in thermal behavior, but aware of the regulations in this domain, asking for the value of the GV coefficient (overall thermal loss applicable to housing buildings) of the building within which he lives, on the basis of the thermal conductivity coefficients related to the windows, floors, walls, etc. The evaluation of the GV coefficient is not trivial at all, since it takes into account the inner volume of the building, the total of heating losses due to ventilating and the overall loss coefficient related to the different partition walls. To date, the information system is not able to provide the user with this kind of complex information on the sole basis of the data contained into the database;

- *structure of the scenarios*: as they have been proposed for the prototype, the scenarios are rigid, this means that the user is obliged to follow the pre-defined path proposed, answering a series of very *precise* questions aimed at proposing a possible solution to his problem. To date, the scenarios do not handle questions closer to the reality of the problems met by professionals (or close to the initial design phase of the project), since they are not capable of dealing with the incompleteness and inconsistency in the “*human*” way of reasoning (the “*human*” user of the information system is not always capable of providing the system with precise, complete or perfectly right data). Sometimes also, this information is not yet defined, according to the stage of the project they refer to.

Besides, the scenario cannot evolve, e.g. suggest other questions on the basis of the previous answers (no reasoning embedded).

To date, the existing concept of “scenario”, as dealt with by the project does not enable the development of scenarios involving more than one kind of actor simultaneously, such as possible exchanges between the architect of a building and the structural engineer, during the detailed design phase, and/or with the thermal engineer in charge of the HVAC (Heating, Ventilating and Air Conditioning) facilities. All of them do not consider, for example a wall (or a partition-wall) in the same way, nor with the same functionalities (separation of spaces, aesthetics, loading -- whence the thickness, or the acoustical behavior of the concrete wall, given its thickness). For these three actors, the wall does not at all appear as the same “*object*”, even though, once completed, there is only one physical element in the building.

In other words, the current structure of the scenarios does not allow either knowledge or knowledge management to be embedded within the scenarios.

- *Heterogeneity of the information handled*: another kind of problem encountered in this project is related to the apparent heterogeneity of the information handled by the information system. This heterogeneity (in contents and structure) appears at two levels, which can be described as follows:

- *structuring of the information mentioned in the menus*: for example, on the upper menu, eight possibilities are proposed to the user. Each of them has a specific structure of the data or items of information contained. According to the “*object*” selected in this menu, the *kind* of information retrieved may be of several types: plain text (e.g. rules), figures, text and figures, hypertext, etc. The information needed is then generally accessed through the left menu, whose content varies, depending on the *object* of the upper menu (see Figure 1). All this means that the information system must be able to handle, at the same time, eight different ontologies related to the topics the *objects* refer to. The problems we faced was finding a way of managing the coordination among these ontologies: one of them was related to codes of practice and regulations applicable to construction in France, another to some “*rules of thumb*” whose knowledge is necessary to install e.g. a window and the other ones to safety rules, commonalities and general features, certification and logistics. Of course the concepts provided by these ontologies are different and they are expressed in neither similar ways nor with similar terminology. This way of managing several ontologies, simultaneously, is not a trivial issue (Bestougeff & Salzano, 1999), but it has been made more difficult in this project by the fact that the granularity of the information expressed in the ontologies is not the same. For example, some “*rules of thumb*” to be followed in installing a window for refurbishing are sometimes far more detailed than the mere list of the components (as provided by the manufacturer of the window) the

user of the IS will have to buy. The terms used to describe them may also be different, since a piece may have a completely different name according to the role this piece will play, once placed on the wall or on the roof. We have to keep in mind the fact that during the initial phases of a project, the future location of the pieces is not yet defined.

- *means of accessing the information system*: this problem can be seen as a second level of heterogeneity of the information. It can be defined as follows: when the end-user enters the information system, access is made through a first ontology, belonging to the list mentioned in the upper menu. The problem is: "which access ?" -- That is, in other words, do we have to consider the possibility of managing all the possible accesses, or do we have to favour a specific kind of access, corresponding, for example, to most users of the IS ? Again, this problem is not trivial, particularly given the difference of granularity between the ontologies. In our opinion, we have to favour a specific access, through the ontology whose granularity is the smallest, since it seems to be easier to propose and then define correspondences to other ontologies with less precision.

In terms of navigation among the ontologies, it is important to define criteria: at the lowest level, through the use of the smallest possible items of information, that is through pre-defined "key-words", or (pieces of) sentences, or else according to correspondence rules defined on trees homothetic with respect to proportions of granularity. We have, however, to pay attention to the fact that some terms have legal meanings, and some of their homonyms do not have. That is, they are not really homonymous from a legal point of view.

The last part of this study suggests a new way of approaching the development of scenarios, and more generally the use of the IS, on the basis of ontology engineering.

7 KNOWLEDGE MANAGEMENT WITHIN SCENARIOS THROUGH ONTOLOGY ENGINEERING: A NEW SEMANTIC APPROACH OF THE CONSTRUCTION INDUSTRY

One of the problems we faced in this project lies in the research of correspondences between concepts handled by various professionals of the construction sector, these concepts being different both in their meaning (semantics), but also in the "granularity" of the item of knowledge they refer to. These concepts may also be different depending on when they occur during the different phases of the construction project life cycle (for example some temporary parts needed only during the on-site construction, are added, later on, to other structural elements of the building where they become permanent: this is a frequent feature of steelwork construction).

This aspect has not been the subject of an in-depth development in this study, because, at the beginning of the project, we have had to face other (and unexpected !) problems, notably the low level of computerisation of the information related to the products and components used in construction. Besides, the aim of the project was mainly to show the *feasibility* of an approach in terms of information related to products and components in construction.

The approach we propose here focuses on the problem of developing correspondences between the items of knowledge, of whatever kind, contained within the information exchanged in the construction industry, as dealt with in this project. These correspondences are expressed by means of ontologies. Here, the role of the scenarios is two-fold: on the one hand, they contribute to help the *engineering* of the ontologies, thus providing a bottom-up approach of the problem (from the reality to its conceptualisation); on the other hand, once defined and integrated into the system, they provide the user of the IS with a *user-friendly* interface, *de facto* perfectly suited to his needs.

Of course, the process of engineering an ontology does not stop at the mere development of a set of informal questions related to the topic the scenario deals with. In fact, this process ends with the translation of this informal knowledge into formal specifications able to be implemented on the *physical* information system.

For any given ontology, the goal is to agree upon a shared terminology and set of constraints on the objects in the ontology ([Gruninger & Fox, 1995](#)). We must agree on the purpose and ultimate use of our ontologies. We must also provide a mechanism guiding the design of ontologies, as well as providing a framework for evaluating the adequacy of these ontologies. Such a framework allows a more precise evaluation of different proposals for an ontology, by demonstrating the competency of each proposal with respect to the set of questions that arise from the applications. These justify the existence and properties of the objects with the ontology and also contribute to validate the approaches proposed through the scenarios through cross-referencing of the dual approach: from the applications to the system and vice versa.

7.1 Ontology engineering

The process of engineering an ontology starts with *scenarios* ([Gruninger & Fox, 1995](#)) describing the applications which the system is to support. The scenarios here represent different points of view (single or multiple) the end-user wants to adopt.

The procedure for developing and engineering an ontology can be described as follows:

- definition of a set of informal competency questions that represent the type of information provided by the model to each application;
- then, an initial terminology comprising objects, attributes and relations is designed;
- the terminology and questions are then refined, according an interactive process, until a precise and unambiguous set of terms is created.

Motivating Scenarios: The development of ontologies is motivated by scenarios that arise in the applications. In particular, such scenarios may be presented by construction partners as problems they encounter in their relationships with the other partners. The motivating scenarios often have the form of story problems or examples which are not adequately addressed by existing ontologies. A motivating scenario also provides a set of intuitively possible solutions to the scenario problems. These solutions provide a first idea of the informal intended semantics for the objects and relations that will later be included in the ontology.

Informal Competency Questions: Given the motivating scenario, a set of queries arises which place demands on an underlying ontology. We can consider these queries to be requirements that are in the form of questions that an ontology must be able to answer. These are the informal competency questions since they are not yet expressed in the formal language of the ontology.

Ideally, the competency questions should be defined in a stratified manner, with higher level questions requiring the solution of lower level questions. The ontology would not be considered as well designed should all competency questions have the form of simple lookup queries; there should be questions that use the solutions to simple queries.

These competency questions do not generate ontological commitments; they are instead used to evaluate the ontological commitments that have been made. They evaluate the expressiveness of the ontology required to represent the competency questions and to characterize their solutions.

7.2 From informal queries to formal specifications

Once informal competency questions have been posed for the proposed new or extended ontology, the terminology of the ontology must be specified using first-order logic.

An ontology is a formal description ([Uschold, 1996](#)) of objects, properties of objects and relations among objects. This provides the language that will be used to express the definitions and constraints in the axioms. This language must provide the necessary terminology to restate the informal competency questions.

The first step in specifying the terminology of the ontology is to identify the objects in the domain of discourse. These are represented by constants and variables in the language. Attributes of objects are then defined by unary predicates, relations among objects, which are defined using n-ary predicates.

The axioms in the ontology specify the definitions of terms in the former and constraints on their interpretation; they are defined as first order sentences using the predicates of the ontology. It is important to understand the significance of using axioms to define the terms and constraints for objects in the ontology. Simply proposing a set of objects alone, or proposing a set of ground terms in first-order logic, does not constitute an ontology. Axioms must be provided to define the semantics, or meaning, of these terms.

It is also important to realise that this is not the implementation of the ontology; it is the specification of the ontology. The implementation of the ontology has to be translated into formal representations such as KIF ([Genesereth & Fikes, 1992](#)), ([KIF, 1999](#)) or Z ([ISO, 1999-3](#)).

The process of defining axioms is perhaps the most difficult aspect of defining ontologies. However, this process is guided by formal competency questions. As with the informal competency questions, the axioms in the ontology must be enough to express the competency questions and to characterize their solutions; without the axioms we cannot express the question or its solution. Further, any solution to a competency question must be entailed by or consistent with the axioms in the ontology alone. If the proposed axioms are insufficient to represent the formal competency questions and characterize the solutions to the questions, then additional objects or axioms must be added to the ontology until it is sufficient. This development of axioms for the ontology with respect to the competency questions is therefore an iterative process.

There may be many different ways to axiomatize an ontology, but the formal competency questions do not generate these axioms. We use them to evaluate the completeness of the sets of axioms in any particular axiomatization. In this sense, we can compare the expressiveness of different sets of axioms using the competency questions. If there is a competency question that one set of axioms can represent and another cannot, then the first set is more expressive. If two different axiomatizations

can represent a competency question and characterise their solutions, then they are equivalent with respect to the question, and any comparison must use other criteria. Once the *competency questions* have been formally stated, we must define the conditions under which the solutions to the questions are complete. This forms the basis for *completeness* theorems for the ontology.

8 CONCLUSION

Several trends appear for the future of the SYDOX/MATCOMP/Xi information system. Among them, let us mention:

- the extension of the software prototype to other products than those dealt with today, in order to increase the completeness of the system, but also the validity of the implementation already made. This will need important work in terms of the information to be included in the database;
- the development of less rigid scenarios, designing questions (asked by the user of the software) that deal with more realistic problems faced by professionals of the sector, through the use of engineering based ontology techniques and the use of the PSL language;
- interfacing with professional information providers outside the information system, notably by means of current (and future) Internet facilities;
- the follow-up of standardization work in the domain of product data exchange (e.g. ISO 10303 STEP and ISO 13584 P-LIB standards for product model data and part library representation), electronic data interchange (EDI) and PSL. This is to enable the use of these standards, notably for setting up relations between SYDOX and distributed information systems;
- work on the possible transfer of the SYDOX/MATCOMP/Const software system to other industrial sectors facing the same (or similar) problems for handling numerous and diverse product and component data.

To conclude : SYDOX/MATCOMP/Xi constitutes an interesting approach for structuring complex information, ‘structuring’ meaning to enable smart and efficient access to information bases of whatever kind.

From that point of view, if the method (and/or the methodology) developed within the project proves to be efficient, this software system could be of a great interest to a lot of people, particularly in the building industry, since all professionals involved in construction experience similar kinds of information problems.

Maybe the problem is more crucial in construction, where communication mainly relies on humans, where teams are never the same or involve the same partners, where discussions often need important expert knowledge, and above all, where nearly always need compromise to reach agreement for the problems to be solved.

We also cannot forget the fact that the products and components (for most of them) evolve rapidly and it becomes more and more urgent to adapt the software tools to this evolution.

9 THANKS

The authors want to thank Karim FRAOUA, Sandrine GENTHON and Emilie RESIDENT for their contribution to this project.

10 REFERENCES

Bestougeff, H. & Salzano, G. (1998) *Report n° 98- 15 Recherche de correspondances en vue de la fédération de bases de données relationnelles*, ISIS – UMLV, Université de Marne la Vallée, France.

Bestougeff, H. & Salzano, G. (1999) *Report n° 99- 19 Construction et génération d’ontologies pour fédérer des bases de données relationnelles*, ISIS – UMLV, Université de Marne la Vallée, France.

CSTB (CSTB, n.d.) Homepage of Centre Scientifique et Technique du Bâtiment. Available at: <http://www.cstb.fr/>

CTBA (CTBA, n.d.) Homepage of Centre Technique du Bois et de l’Ameublement. Available at: <http://www.ctba.fr/index2.html>.

CTICM (CTICM, n.d.) Homepage of Centre Technique et Industriel de la Construction Métallique. Available at: <http://www.cticm.fr>.

Dubois, A.M., Cutting-Decelle, A.F., Fies, B., Lanvin, J.D., Bonfils, P., Chabrolin, B., Dubois, J.E., Messabih, A., Mommessin M. & Hamchaoui, M. (1998) *SYDOX/MATCOMP/Xi, Identification et structuration des données, Report of the SYDOX project, n°.1.1*, Ministère de l'Industrie DGSI / SERICS, Paris, France.

Dubois, A.M., Cutting-Decelle, A.F., Fies, B., Lanvin, J.D., Bonfils, P., Chabrolin, B., Dubois, J.E., Messabih, A., & Mommessin, M., (1999) *SYDOX/MATCOMP/Xi: Final report of the SYDOX Project*, Ministère de l'Industrie DGSI/SERICS, Paris, France.

Diffuse Project (n.d.) Standards and Specification Lists, Retrieved from the *Diffuse Project* web site; <http://www.diffuse.org/standards.html>.

Genesereth, M. & Fikes, R., (1992) *Knowledge Interchange Format (Version 3.0) - Reference Manual*, Computer Science Dept., Stanford University, Stanford, CA, USA.

Genthon, S. & Resident, E., (1998) *Aide à la conception de bâtiments: produits et composants utilisables, identification de scénarios de conception et traduction en termes d'information sur les gammes de produits et composants, Travail de Fin d'Etudes (Internal Report)*, University of Savoie (ESIGEC), Chambéry, France.

Gruninger, M. & Fox, M.S., (1995) The logic of enterprise modeling. In P. Bernus and L. Nemes (Eds): *Modeling and methodologies for enterprise integration*. London: Chapman & Hall.

Integrated Architectures (2000) Homepage of Integrated Architectures. Available from <http://www.iai.com>

ISO (1994) *ISO IS 10303-1, Industrial automation systems and integration - Product data representation and exchange, Part 1: Overview and fundamental principles*, Geneva: ISO.

ISO (1997) *ISO IS 13584-1, Industrial automation systems and integration - Parts library, Part 1: Overview and fundamental principles*, Geneva: ISO.

ISO (1999-1) *ISO CD 16668, BSR Rules, guidelines and methodology*. Available from <http://www.ubsr.org>

ISO (1999-2) *ISO WD, Knowledge Interchange Format, Part 1: KIF-Core, ISO/JTC1/SC32/WG2*, Geneva: ISO (number not yet given)

ISO (1999-3) *ISO CD 13568-2, Z notation*, ISO JTC1/SC22/WG19, Final CD, Geneva: ISO.

ISO (1999-4) *ISO DIS 12006-2, Organisation of information about construction works – Part 2: Framework for classification of information*, Geneva: ISO.

ISO (2000) *ISO FDIS 15531-1, Industrial automation systems and integration - Manufacturing management data exchange: overview and fundamental principles (MANDATE)*, Geneva: ISO.

ISO (2001) *ISO CD 18629-1, Industrial automation systems and integration – Process specification language: Part 1: Overview and basic principles*, Geneva: ISO.

ISO TC184/SC4 (n.d.) Homepage of ISO TC184/SC4. Available from <http://www.tc184-sc4.org/>

Kaetzel, L.J. & Galler, M.Z. (1997) NISTIR 6034 Proposed format for data on cements in a material properties database. Retrieved August 22, 2002, from the *NIST* Website: <http://fire.nist.gov/bfrlpubs/build97/art028.html>.

Mommessin, M., Cutting-Decelle, A.F., Dubois & A.M., Dubois, J.E., (1999) *SYDOX/MATCOMP/XI: An information system adapted to the needs of construction actors, CIB W78 Workshop, 8dbmc*, Vancouver, Canada

Rumble, J.R. & Smith, F.J., (1990) *Database systems in science and engineering*, New York: Adam Hilger Eds

Uschold, M. (1996) *AIAI-TR-197 Building Ontologies: towards a unified methodology*, AIAI, University of Edinburgh, UK