DBMS STANDARDS *

MARCH, 1981

C. THANOS

ISTITUTO DI ELABORAZIONE DELL'INFORMAZIONE, VIA S. MARIA 46, PISA-ITALY

1. INTRODUCTION

The problem of standards and standardization is a topic which is being widely discussed by scientists and researchers on the one hand and by producers and manufacturers on the other. In the academic world, the discussion is based on considerations as to whether a too hasty adoption of standards could be of hindrance to further research. The business world is concerned with the impact a new standard might have on its products. In fact, the introduction of new standards could render current products obsolete or weaken their position on the market. From the user point of view, the lack of standards is a great disavantage for several reasons, e.g. the difficulty of comparing products with dissimilar specifications, the difficulty of converting from one package to another, etc.. From the report "Experience of DBMS Usage in Europe", produced in the first phase of this project, it can be seen that 66% of the organisations interviewed are in favour of the adoption of a standardization for database systems, 25% expressed no view and only 9% declared themselves to be contrary to this approach. The same report also showed that 61% of organisations feel themselves to be "locked-in" by a particular DBMS, 16% expressed no opinion and the remaining 23% gave no reply. Given the size and the complexity of today's information systems, the introduction of standards in this field becomes especially difficult; in particular, it should be remembered that the standardization of a database system, apart from its complexity, has also political aspects. In order to justify the effort required to introduce DBMS standardization, in section 2 of this paper we will formulate a certain number of hypotheses and list some of the expected benefits.

Section 3 describes the principal areas in the field of database systems which must be taken into consideration when selecting areas for standardization.

In section 4, the suggestions made by the different national and international organisations in this field will be briefly presented and discussed.

In section 5, mention will be made of the different national and international organisations studying the problem of standardization.

Some concluding remarks are made in section 6.

2. WHY DBMS STANDARDIZATION IS NEEDED

The decision to move towards DBMS standardization may be justified on the basis of certain hypotheses, the

---

* This report is part of a more extensive project on Evaluation of Database Systems, partially financed by the European Economic Commission, in which Institutes from Germany (GMD), France (INRIA) and the United Kingdom (NCC) participated.

particular benefits to be gained and a balancing of the costs which can be saved against those involved in the implementation of standardization /3/.

2.1 Hypotheses

a) The use of DBMSs will continue to spread.
This hypothesis has been largely shown to be true by our experience gained during the first stage of this project. In fact, it has been seen that one of the main reasons which have encouraged organisations to adopt the database technology is the desire to keep pace with new technological developments rather than the immediate need for such technology, as it is felt that the database approach in the future will be the only solution to management problems. In /1/ it is also shown that 76% of the organisations interviewed intend to extend the use of the database approach to other areas of application. Finally, it must be emphasised that none of the interviewed organisations have said that they regret having introduced the database approach. Therefore, it can be reasonably presumed that hypothesis (a) is valid.

b) Databases will grow in the future in terms of size, number and complexity.
Indeed, this hypothesis seems to be quite well founded when we remember that, as is shown in /1/, while 50% of the organisations interviewed have less than 40% of their data within the DB approach, the tendency for the most part is to extend this approach to an ever increasing percentage of data. It is also reasonable to expect a growth in the complexity of the databases as, at the present, the level of the integration of the files and the applications is not high and attempts to create a corporate database are still rare /1/. On the other hand, the growth in database complexity will also be favoured by an increase in the experience of the DP community.

c) Future alterations in the architecture of the DBMS will not be revolutionary but rather evolutionary.
In the past, there have been heated discussions between scientists concerning the architecture of the DBMS. These discussions have shown that, if certain changes in the database technology are inevitable, in the near future these changes will certainly be evolutionary and standards will exist to control this evolution.

d) Changes in the hardware and software in the next few years will have only a limited effect on the main functions performed by the database systems.

2.2 Benefits

The potential benefits that can be gained from the standardization of the DBMS are listed in this section.

a) Easier data and program conversion
At the moment, the process for data and program conversion is very complicated and expensive. In fact, many organisations prefer to redesign their database and applications rather than have to resort to time-consuming operations which frequently give unsatisfactory results. The problem of conversion is however very much felt with respect to the future as a growth in operations involving the movement of data from different organisations in one country, or between organisations in different countries, or in multinational type organisations is to be expected.

b) Easier staff mobility
The costs involved in staff educa-

tion are quite high. It is clear that, if a DBMS standard is adopted, the costs of re-education will be eliminated and the temporary loss of productivity linked with staff mobility will be reduced. If more than one standard is introduced, however, the saving of education costs will be inversely proportional to the number of the standards introduced. In any case, staff mobility will always be facilitated by the introduction of standards.

c) Reduction of the proliferation of products which differ only slightly.

d) Simpler process for DBMS selection and evaluation.
At the present, the DBMS selection and evaluation process is both complex and very expensive and, consenquently, is normally conducted only superficially, as shown in /1/. Therefore, all too often, a package is acquired which does not satisfy the information needs of the organisation. It is clear that the adoption of one or of a limited number of standards will make the evaluation process simpler and more linear.

e) Fairer competition between companies producing DBMSs.
The introduction of DBMS standards will make competition between manufacturers fairer and will enable the wider participation of all manufacturers in procurement processes. The necessity to acquire each component of the product from the same manufacturer will be eliminated.

2.3 Costs Considerations

Clearly any decision to adopt DBMS standards must be based on and justified by a careful analysis of the costs involved. In fact, the following questions should be posed /4/:

a) What costs can be attributed to a lack of a standard?

b) How much would it cost to implement a standard?

c) How much would it cost to maintain a standard?

d) How much would it cost to retrofit or discard the prestandard elements?

We will briefly summarise certain hypotheses which are formulated in /4/ concerning costs, as they may be of assistance in replying to these questions.

The hypothesis of an ever increasing utilisation of the database in the near future, of course, remains valid.

It could also be hypothesised that:

- in the next 5 years 50% of the applications will utilise database system techniques as their building block;
- unless standards are implemented, system maintenance is destined to become one of the principal costs in passing from one hardware to another. The reduction of costs is dependent on the quantity of the reprogramming effort necessary and frequency of hardware/system change;
- the introduction of a DBMS standard will improve system debugging;
- the development of a DBMS standard would require the employment of about 20 people full time for three years with an additional cost of $7000/year per person.
- the maintenance of a DBMS standard would require about 20% of the work-force of the same group;
- the reeducation of the personnel would entail a period of approximately one month for 50% of the programmers.

Obviously, these costs would increase

if more than one standard was introduced.

It conclusion, we summarise here the costs and savings hypothesised in /4/ after the introduction of a DBMS standard:
Savings due to maintanance = $1460 M/yr
Savings in reeducation = $11.2 M/yr
Development costs = $3 M
Maintenance costs =$200.000/yr
One-time reeducation costs = $560 M

These figures can obviously be discussed but in any case they do show that the introduction of a DBMS guarantees substantial saving.

3. MAJOR AREAS TO BE CONSIDERED FOR STANDARDIZATION

Before discussing the components of the DBMS which can be considered to be candidates for standardization, a brief presentation of the position paper of the ANSI/X3/SPARC Study Group on Database Management Systems is necessary. In the autumn of 1972, the American National Standards Institute (ANSI) Committee on Computers and Information Processing (X3) through its Standards Planning and Requirements Committee (SPARC) created a Study Group on Database Management Systems in order to investigate the potential for standardization in the area of database management systems. The Study Group presented an interim report in 1975 /5/ and a final report in July 1977 /6/.

3.1 The ANSI/X3/SPARC Study Group Framework

The study Group has identified the scope of the database management system as being "records, fields, files, etc. and the descriptions for all these, and all the indices, mapping techniques, access methods, file organisations and end user languages" /6/. These offer a framework which is made up of three interlinked schemas which guarantee data independence. This approach isolates the physical description of the stored data (the Internal Schema) from the particular data structures which the application programmer uses (the External Schema) by providing a third schema (The Conceptual Schema) into which both the physical and the application descriptions can be mapped. Therefore, the Conceptual Schema supplies the necessary indirection to isolate the physical description from the application views. As the Conceptual Schema contains the definitive descriptions, both the Internal and External Schemas must contain only the description of the objects described in the Conceptual Schema. In fact, each External Schema should normally contain a part which is oriented towards a specific application. The existence of multiple and concurrent data descriptions requires processes to match the description of an object in each of the three schemas. These matching processes (or mappings) require stored representations to allow machine processing. In addition, three administrative "roles" within the enterprise are identified, the Enterprise Administrator, the Application System Administrator and the Database Administrator and have the task of supplying the necessary descriptions, mappings and views. Each interface between a role and the framework needs a language to express the information which represents the mappings and the descriptions and permits the processing by the transforms. In the following paragraphs, the main components of the framework will be discussed. Fig. 1 at the end of this report shows a condensed form of a larger, more detailed figure which is presented in the final report /6/. In this figure, the hexagons represent roles, the rectangles represent processors and a bar associated with a circled number represents an interface, the lines connecting components indicate the data flow, the control of the information, the programs and the data description. Finally, dashed boxes indicate program preparation and execution subsystems.

3.1.1 Conceptual Schema

The Conceptual Schema is a collection of objects representing the entities, properties and relationships of interest in the enterprise. Examples of these objects are records, irreducible relations, nodes in a graph structure, etc. The Conceptual Schema is described explicitly in machine readable form with a well defined and potentially standardizable language.

Managed by the Enterprise Administrator, the Conceptual Schema has the following purposes:

- provides a description of the information of interest to the enterprise;
- provides a stable platform to which Internal and External schemas may be bound;
- permits additional external schemas to be defined or existing ones to be modified or augmented, without impact on the internal level;
- allows modifications to the internal schema to be invisable at the external level;
- provides a mechanism of control over the content and use of the database.

3.1.2 Internal Schema

The internal schema is a collection of objects which are related to the objects in the conceptual schema of the database. The internal schema is oriented towards the most efficient use of the computing facility, consistent with the processing requirements of the enterprise. Provided by the Database Administrator, it reflects current storage techniques and permits mappings between these and the Conceptual Schema. It is reasonable to assume that entities and conceptual objects representing facts about entities may have a longer life than the technologies upon which hardware and software implementations are based. Therefore, the internal schema should be able to change to reflect changing technologies.

3.1.3 External Schema

Several External Schemas provided by the Application Administrator may exist, each containing descriptions of entities, properties and relationships of interest to a specific application. Each External Schema may subset the objects in the Conceptual Schema, and may reflect a different data model view of the objects. In this way, the three principal data models (hierarchical, network and relational) can have different External Schemas mapping to the same Conceptual Schema. The different external views do not influence the way in which the data are effectively stored.

3.1.4 Correspondence of the Levels

The correspondence between objects in the external, conceptual and internal schemas are established through mappings and transforms. Mappings bind descriptors in one schema to those in another. The binding may be between an external schema and more encompassing external schemas, an external schema and the conceptual schema, the conceptual and the internal schemas or between an external schema and the internal schema. Transforms which are processing functions use the information provided by the mappings to translate external-oriented requests to internal--oriented requests either directly or indirectly via the conceptual schema.

3.1.5 System Control Points

The Study Group has proposed that a "data dictionary" should be at the center of the framework and should be the container where the different schemas and mappings should be stored. The Group concluded by indicating that the most suitable subjects for standardization are the languages used in the different interfaces.

3.2 Areas for DBMS Standardization

When investigating the potential for standardization in the area of database management systems, the systems could be considered to be subdivided into functional modules. We are in agreement with the suggestions in /3/ concerning the DBMS components which should be considered for standardization. In this paper, four components are indicated:

- Data Description Facilities
- Data Manipulation Facilities
- Data Dictionary/Directory Facilities
- Query and End-User Facilties

3.2.1 Data Description Facilities

The Data Description Language (DDL) is defined as a stand-alone language which describes:

- data element attributes;
- logical relationships between data units;
- logical structure of the database;
- logical methods of data access.

The DDL does not include the definition of the physical storage media.

At the present moment, there are a large number of DBMSs in use. Each system has its own language to describe the attributes and the data relations. This large number of systems means that the exchange of data between systems is difficult as, each time, a new data description is needed for the target system. This is both expensive and time-consuming. It is, therefore, necessary to arrive at the definition of standard specifications for Data Description Languages.

3.2.2 Data Manipulation Facilities

The Data Manipulation Language (DML) is the tool which provides the programmer with access to the database and transfers data between program and database. The DML can exist either as a stand-alone language or it can rely on a host programming language which supplies the computational capabilities required for data manipulation. An application program, therefore, is written using a mixture of the host programming language statements and DML commands. The DML consists of a number of data manipulation commands and functions:

- data retrieval;
- data addition;
- data modification;
- data deletion;
- modification of data relationships.

At the present, the situation is characterized by:

- A variety of DBMSs each with its own DML;
- The data manipulation functions of each DML vary both in their syntax and semantics;
- The functional capabilities supplied by the DML are different in the various DBMSs available.

The differences in the data manipulation functions and the languages give rise to many problems, for instance the reprogramming of a large number of application programs using DML may be necessary for conversion. Problems of this type can be avoided by standardization.

3.2.3 Data Dictionary/Directory Facilities

A Data Dictionary/Directory (DD) is a software tool which is used to control the whole set of data in the database. The DD lists, describes and localises each individual data item in the database. In fact, the DD actually controls the data description rather than the data content.

The Dictionary part describes the data which are contained in the database, the Directory localises where the data are stored.

The DD is used by the DBA, the systems analyst and the software designer, as a central deposit of information concerning the data resources, it helps to plan, control and evaluate the collection, storage, and utilization of the data. The DD is also of great value as a tool in the Documentation, helping to establish standards in the naming, use and coding of data.

It can be presumed that, with the growth of databases and the development of distributed databases, the need for data dictionaries will be increasingly felt. If the interface between the DD and the DBMS was standardized, the DBMS users would be able to use any data dictionary which had the standard interface implemented without additional costs.

3.2.4 Query and End-User Facilities

At the moment, there are a large number of DBMS end-user facilities. They can be divided into the following categories:

- Report Specification Language;
- Enquiry Specification Language;
- Update Specification Language;
- Parametric Interface.

Most of these facilities can be learnt very quickly and have been designed for ad hoc activities. In this case, standardization of the syntax and semantics hardly seems justified.

4. CURRENT DATABASE STANDARDIZATION EFFORTS

In this paragraph, the proposals and recommendations made by the differrent national and international organisations responsible for the investigation of the particular areas of database management systems which seem to be most suitable for standardization are reported.

4.1 Data Descrioption Language

The Federal Information Processing Standards Task Group on Data Management System Standards (FIPS TG-24) (see 5.8) has made the following recommendations for the DDL /3/:

a) A two-part data description standard is required that contains the common description of the data element (attributes) and the facility to describe multiple data structure classes;

b) The specification of the standard description of data attributes should be similar to the attributes in the PICTURE and TYPE clauses of the CODASYL DDL.

c) The specification of the standard description of the data structures should be required to encompass current data models such as the hierarchical, network and relational models;

d) Consistent with recommendation a), the standard description of the network data structure within the DDL should be based on the CODASYL DDL.

e) Consistent with recommendation a), companion data structure descriptions for the hierarchical and relational data models should be developed.

In May 1978, The Conference on Data Systems Languages (CODASYL) published revised specifications for the CODASYL Data Description Language (DDL).
These specifications are reported in a Journal of Development /8/.

The scope and purpose of these specifications are defined in the introduction:

"This Journal of Development contains a specification of a language to describe the structure and contents of a database. This description is called a schema. The schema language represents one of several languages which data base designers, implementators and users will employ. Other languages include procedural programming languages such as COBOL and FORTRAN, end user oriented languages, data manipulation languages, data storage description languages, and languages to control the allocation of resources and the execution of work (data processing) on a computer system. In this Journal, the term "program" or "application program" should be understood to mean restrictions written either in a procedural programming language or in an end user language. Programs typically interface to or reference the following elements when used with a data base controlled by a schema language:

- A subschema language to describe a subset of the schema which is of interest to a particular application program. A sub-schema enables an application program to deal with a subset of the data in the database. The subschema may also vary in certain respects from the schema with respect to particular elements in the database.
- A data manipulation language (DML) used at execution time to handle all program interfaces to the database.

The subschema language specifications and the DML specifications are outside the scope of this Journal. A data storage description language defines how data described in the schema may be organized in terms of an operating system independent of storage environment. Such a description is known as a storage schema. The storage schema has no effect on the logical results of application programs but only affects their performance. The separation of the storage schema from the schema insulates application programs from changes made to the data base for reasons of efficiency.

In order to control the resources of the data base system, it is necessary for the storage space described in the storage schema to be known to the operating system. This may be achieved through implementation of dependent language facilities. In order to execute work on a computer, including processing a database, it is necessary to have a language to specify control of the work to be done. Such a language is commonly called a job control language. Specifications for a job control language are not included in this Journal.

The schema language is a specification of a common data description language (DDL) which is independent of, but common to, the other languages required for a data base system.

It is expected that the schema DDL will have a significant impact on the development of functionally compatible database management systems and will increase the portability of programs between different computer systems.

While the widespread adoption of the DDL will not of itself fully achieve this objective, it does lay the foundation for the development and adoption of common subschema and data manipulation languages. The net result would be an increase in the portability of source programs. The portability of physical database representations is not dealt with in this Journal".

The American National Standards Institute (ANSI) committee on Computers and Information Processing (X3) initiated a committee for data description language specifications (H2). This committee has released a working document (9/10/79) /10/ which has been circulated to receive comments and suggestions. The 1978 CODASYL Data Description Language specifications have been used as the basis of this document.

The International Standards Organization (ISO) Technical Committee 97/Sub-Committee 5 in a recent resolution has decided to create a group of experts to examine this document and the ANSI proposal for an International Standard Data Definition Language.

Finally, in 1979, the ISO/TC97/SC5/WG3 produced a position paper aimed at defining concepts for conceptual schema languages /9/. This paper has been distributed for comments. The introductory chapter describes the roles and objectives of the conceptual schema. The discussion is based on a view of the conceptual schema, which is best described by the terms: object system, abstraction system, information base. The object system is the name given to that part of the world on which information is held in the information base. (The term information base is used instead of data base to avoid any connotation regarding the form of representation of the aquired information.) The object system is perceived as being structured: entities of the object system form classes according to common characteristics (these classes may be composed of people, departments, etc.). They obey general rules (people are assigned to one department at most, etc.). The combination of classifications and general rules constitute the abstraction system.

The conceptual schema is regarded as a model of the abstraction system. It will simultaneously describe the structure of the information base being a model of the object system. The report differentiates between two types of modelling: static and dynamic. Dynamic modelling is directed towards the description of possible changes in the object system and the information base while static modelling is restricted to the description of the valid states.

The second chapter of the report describes three approaches towards static modelling at some depth:

- Entity-Attribute Relationships (EAR)
- Binary Relations (BR)
- Interpreted Predicate Logic (IPL)

The basic concepts of the EAR-approach are: entities, attributes of entities, relationships among two or more entities. The BR approach resembles the EAR-approach but does not distinguish between attributes and entities. In addition, only relationships between two entities are allowed.

While these first two approaches have been discussed widely elsewhere, the third seems to be less well known. The information base (and therefore the Conceptual Schema) is perceived as a collection of assertions, which could be expressed by declarative sentences of some formal language. This approach is based on methods developed by mathematical logic. It seems especially attractive when dealing with constraints.

The third chapter of this ISO paper deals with dynamic modelling. Dynamic concepts have rarely been discussed previously in the context of the conceptual schema. It is perhaps a striking feature of this report that it includes this topic. The report differentiates between two approaches: the transitional and the operational approach. The first one is state oriented. A transition is defined as a pair of valid states: the before-state and the after-state. The operational approach is process oriented, its fundamental concepts being basic operations (e.g. "insert information"). This part of the report is likely to trigger off a debate as to what extent dynamics should be reflected in the conceptual schema and by which means this topic could be described.

4.2 Data Manipulation Language

The Federal Information Processing Standards Task Group on Database Management System Standards (FIPS TG-24), has formulated the following recommendations for the DML:

a) Develop multiple DML Standards specifications;

b) - For each standard host programming language (e.g. COBOL, FORTRAN) develop immediately a standard DML specification for each category identified by TG-24 (Categories are grouping of DBMS based on certain technical criteria. The criteria used by TG-24 to categorize DBMS, were the common data manipulation functions and data structures).
   - As a short range goal, develop a single standard DML specification for a given category that interfaces with all standard host programming languages.
   - As a long range goal, develop a single standard DML specification containing the functionality of all categories.

The ISO/TC97/SC5 in a recent resolution aimed at assisting the WG3 in its work, have decided to create a new group which will examine proposals for the international standardization of data base facilities (DML). Lastly, CODASYL have developed detailed specifications for a FORTRAN Data Manipulation Language and a COBOL Data Manipulation Language. The FORTRAN specifications were published in January 1977 /11/ while the COBOL specifications were to be published in April 1978.

4.3 Data Dictionary/Directory Facility

The FIPS Task Group has made the following recommendation for the DD:

a) Data Dictionaries must be able to produce the Standard DDL attribute description as recommended in 4.1. TG-24 took no position on the standardization of data dictionaries but referred only to data dictionaries with an interface between the data dictionaries and the DBMS which must be standardized.

b) The data dictionary must be designed so as to combine standard data attribute descriptions with data structure descriptions to generate the DDL for one or more DBMS.

The British Computer Society has created a Working Group on Data Dictionary Systems which in 1977 published a report /12/.

4.4 Query and End User Facilities

The FIPS Task group 24 has made the following recommendation for the Query and End User Facilities:

a) Standardization of syntax and semantics of end-user facilities is not required. Such facilities are easily learned and are problem and subject--matter dependent, and there are a diversity of end-user facility "styles".

b) Guidelines should be developed to aid the specification of requirements of end-user facilities.

5. NATIONAL AND INTERNATIONAL STANDARDS BODIES

In this paragraph a brief description will be given of the national and international Working Groups which are actually active in the DBMS Standards.

5.1 International Standards Organization (ISO)

ISO is divided into technical commitees (TC), which are divided into subcommittee (SC). Sub-committees may in turn establish working groups (WG). The responsability for data processing rests with TC 97 (Computer and Information Processing). This committee has assigned the work in standards in the field of data bases to SC5 (Programming Languages). In November 1977, SC5 established the working group WG3 (Data Base Management Systems) with the scope "to prepare for standardization in the area of data base management". Thomas B Steel Jr. was appointed as convenor. This group suceeded an informal study group which was formed three years previously. WG3 and its forerunner are the most important groups dealing with DBMS standards. Neverthless, there is some interference with other groups within ISO. SC1 (Vocabulary) for instance made a proposal on vocabulary for data base management systems. In the near future, another group might be established to comment the COADASYL proposal, at least this was the recommendation of WG3.

5.1.1. Members

Members of committees and sub-committees are national standard organisations and not individually named specialists.

There are two categories of members:

- P-members have committed themselves to take an active part in the work of the (sub-)committee.
- O-members just want to be kept informed of the progress of the work.

The countries represented in this project (Evaluation and Implementation of Database Systems) are all P-members of SC5. In contrast to committees and sub-committees the members of working groups act on their own, not as representives of their countries. WG3 consists currently of about thirty active members from 11 countries. Most of them has already been members of the study group. In addition, there are passive members who just receive minutes and technical reports of meetings.

So far no restrictions have been imposed on active or passive membership. Anyone interested in the work of the group should contact the convenor.

When the membership affiliation is examined, it can be seen that there are two dominating groups:

- participants coming from hardware vendors (for instance, there are several participants from IBM);
- participants belonging to universities or similar institutions.

Very few users or software houses are represented. There are strong links with national standards organisations. All German members for instance belong to a comparable working group establihed by DIN, the German Standards Organisation.

5.1.2 Working Program

The following working program was laid down by the SC5 resolution establishing WG3:

- define concepts for conceptual schema languages;
- define or monitor definition of conceptual schema languages;
- develop a methodology for assessing proposals for conceptual schema languages;
- assess candidate proposals for conceptual schema languages;
- define concepts for conceptual level end user facilities;
- define conceptual level end user facilities;
- get acquainted with and react to other data base developments as appropriate;
- develop vocabulary for database management systems.

The first six of these topics grew out of discussion within the study group on DBMS, the forerunner of WG3. Obviously, there is a strong interconnection between these topics as they all refer to the conceptual schema. This concept was introduced by the ANSI/X3/SPARC DBSG report in 1975 within the concext of a three schema framework for DBMS (see 3.1). Subsequently, this framework has been widely accepted and also assumed by the study group. There is, however, little consensus on the form and contents of a conceptual schema. Several approaches have been proposed using quite different concepts. The ANSI report itself is not very specific in this respect. In this situation, the study group thought it impracticable to pursue its ultimate goal which was to propose directly a conceptual schema language for standardization. The concepts to be used for such a language had to be first clarified. Thus, topic 1 has been set up. Topics 2 to 6 are clearly follow-up activities. They should be considered as possible tasks; which of them will be actually tackled is still an open question. Future work will obviously depend on the results gained from the first task. However, external events may also have some influence. Ongoing activities, especially within ANSI, should be mentioned here. Topic 7 has been included partly in response to the just mentioned ANSI activities. The above mentioned proposal on vocabulary for database management systems by SC1 led to topic 8, though the working group considers vocabulary more a by-product of their work than a separate task in itself.

5.2 Deutsches Institut fuer Normung - DIN -

DIN has a structure which deliberatly parallels that of ISO. The 'Normenausschuss Informationsvevarbeitung' (NI = Standards Committees Information Processing) corresponds to TC97. It is divided into 'Arbeitsausshuesse' (AA = Working Committee) following to ISO policy of division into sub-committees and using the same numbering. Thus, the committee responsible for programming, which includes responsibility for DBMS, is AA5. Arbeitsausschuesse may establish 'Unterausschuesse' (UA = Sub-committee) which are comparable to the ISO working groups. In January 1977, the UA5.10 sub-committee 'Datenbanken' was established . This group is the first to be taken into consideration when referring to DBMS standardisation in Germany.

The group which is headed by Dr. C. Heinrish, consists currently of about 15 members. Strict rules regarding membership have not been established.

Besides representatives from hardware vendors (IBM, SIEMENS) and scientists belonging to universities and comparable institutions, another group of members plays an important role: participants coming from data processing departments of various government administrative agencies. The members are in general more practically oriented than those of WG3.

5.2.1 Program and Work Status

The present task of UA5.10 is the standardization of a user interface, called KLDS or LINDA, which provides relatively simple file handling functions at about the level of indexed sequential access.

This interface is part of a system of interfaces called the 'compatible data base interface - KBDS' which has been propagated by a number of data processing departments of various government administrative agencies and is supported by the Federal Ministry of the Interior. The aim is to achieve portability of application programs while using different data base management systems.

Another interface of this family, KSDS, is now being considered for standardization. KSDS, a modified version of KLDS, is a further candidate, but a decision has not yet been made. UA5.10 has met nine times since its establishment, but has not yet been able to present a description of KLDS in a form which permits standardization. The lack of a consistent, generally accepted terminology which could be used for the data description has proved to be the main obstacle. Most of the group's time has been spent on the development of such a system. An important further task of UA5.10 is to observe and eventually cooperate with other bodies involved in DBMS standardization. This applies especially to TC 97/SC5/WG3 and ANSI/X3. As has already been mentioned, several members of UA5.10 are also members of WG3. It is possible that this cooperation will be intensified.

5.3 British Standards Institution - BSI

In 1901, the Institution of Civil Engineers and other professional bodies set up the engineering standards committee. This later became the Engineering Standards Association. In 1929, it was granted a royal charter. A supplementary charter was granted in 1931 when the name was changed to the British Standards Institution.

In the United Kingdom, the BSI is the recognised body for the preparation and promotion of national standards.

It draws up voluntary standards by agreement among all of the interested parties and promotes their adoption. BSI plays a major role in the following international organisations:

ISO - The International Organisation for Standardization
IEC - The International Electrotechnical Commission
and within Europe:
CEN - The European Commitee for Standardization
CENELEC - The European Committee for Electrotechnical Standardization

In addition, BSI has links with many other standards bodies.

5.3.1 Database Standards

BSI has a number of active committees investigating data processing standards. The Committee responsible for programming language standards, which includes work on database standards, is DPS/13. This commitee has established a working group (DPS/13/WG1) which is assisting the work of the ISO Working

Group TC97/SC5/WG3. The chairman of the BSI group is T.W. Olle. The group currently has eight active members. Four of the members represent users, two represent computer manufacturers, one represents a software house and one is an independent consultant. In addition, the BSI COBOL Group (DPS/13/ WG3) has considered the CODASYL database proposals.

5.4 British Computer Society

The British Computer Society is a professional body which suitably qualified computer personnel may join. One facet of its many activities is the organisation of specialist groups. The advanced programming specialist group has three groups working in the database area:

- Data Dictionary Systems working party
- Data Base Administration working party
- Distributed Database working group

These groups cooperate with the national and international standards organisations.

5.5 Association Francaise de Normalisation - AFNOR -

AFNOR has been divided into technical committees, sub committees and working groups which fit the ISO organization exactly. The TC97/SC5/CE3 group was created in 1974.

This group is headed by H. Tardieu and is composed by about 30 members, of whom 10 currently participate at the meetings which are held quarterly. The composition of the group is as follows: computer manufacturers (IBM, CII-HB, SEMS, TEXAS Instruments), software houses (STERIA, SIS), Universities and user groups. Generally speaking, the activities of the group are more oriented towards practical aspects of standardization than in WG3.

5.5.1 Program and Work Status

The AFNOR group has devoted its entire acticity in the last three years to conceiving and writing the ISO report mentioned in 4.1. On the basis of the research work conducted in France in recent years on the dynamic aspects of databases, the group has mainly worked on the preparation of a first proposal on the dynamic modelling aspects.

The present objectives of AFNOR are to concentrate on the creation of a new group in order to examine new proposals for DBMS standardization. This evaluation should be conducted using the ISO/TC97/ SC5/WG 3 report as a reference. AFNOR will act as the secretary of the group and will organize the first international meeting.

5.6 The Conference on Data Systems Languages - CODASYL

CODASYL is the voluntary body that developed the Common Business Oriented Language (COBOL) and guided that language's evolutionary development. CODASYL has also developed language specifications for a FORTRAN Data Manipulation Language and a Subschema Data Description Language, a COBOL Data Manipulation Language and a Subschema Data Description Language, a host-language independent Data Description Language, and a draft Data Storage Description Language. The FORTRAN specifications were established in January 1977 and the others will be published in April 1978.

Though not a standards body, the importance of CODASYL's work on computing languages is well known, and its work in the DBMS areas has had significant impacts already. The various specification developing committees of CODASYL meet at intervals varying from six weeks to three or four months.

Approximately 25 different vendors and users are represented in the developmental committees. Proposals for improvements to the specifications arrive from all over the world, including the European Computer Manufacturing Association, the International Federation of Information Processing Societies, and several vendor user bodies. CODASYL specifications will continue to evolve and will be considered as standards candidate by ANSI. However, CODASYL is not a standards making body and CODASYL specifications can not become standards through CODASYL actions alone.

5.7 Americal National Standards Institute - ANSI

In autumn, 1972, the ANSI committee on Computers and Information Processing Committee (XS) through its Standards Planning and Requirements Committee (SPARC) established a Study Group on data Base Management Systems with a charter "to investigate the potential for Standards". The Study Group issued an interim report in 1975 and a final report in July 1977 (see 3.1).

The final report contained neither specifications for a recommended standard nor recommendations for any action for standardization of any existing products or specifications. The report does contain a "framework" which can be used to consider future standards actions. After accepting the report, SPARC initiated three pertinent actions: referred actions for subschema data description language specifications and data manipulation language specifications to the COBOL committee, referred actions for a subschema data description language specifications and data manipulation language specifications to the FORTRAN committee, and initiated a committee for data description language specifications.

5.8 U.S. National Bureau of Standards - NBS

The role of the NBS of the US Department of Commerce in the Federal Government differs from the role of ANSI and ISO. While voluntary standards processes depend on the cooperation and contributions of essentially independent organisations which make up the membership, the Federal standards process is supported as a management function and its standards are mandatory. The U.S. Department of Commerce which has the specific responsibility for this task, in turn, delegated this responsibility to the Institute for Computer Sciences and Technology (ICST) within NBS.

ICST exercises this responsibility through its own research efforts and through a Federal Information Processing Standards (FIPS) program. The organisation of the FIPS program included a FIPS Coordinating and Advisory Committee (FIPSCAC), an interagency body that provided a two-way communication channel with all the Federal agencies and several Task Groups.

Thus the ICST role resembles a standards staff position in a major corporation. Because FIPS are mandatory rather than voluntary, they should first be subjected to a process of validation, testing, and reporting.

5.8.1 Work Program

Within the FIPS program, three Task Groups have direct relevance to DBMS.

- FIPS TG 9: COBOL Standards

The COBOL Task Group has not yet considered the CODASYL approved DBMS extensions to COBOL. At present, because ANSI appears to have initiated work by beginning a study of the possible standardization of CODASYL DML and the DDL, FIPS TG9 intends to parallel this study.

- FIPS TG 17: Data Element Directories

The Task Group characterizes its activities as examining concepts, principles, and applications of data resource management and data resource directories. Its task is to develop guidelines for constructing data element dictionaries and directories and to identify the relevant performance characteristics of the automated processes designed to use and maintain DDD.

- FIPS TG 24: DBMS Standards

This Task Group has a charter to determine the need for Federal data base management system standards and any appropriate activities necessary to meet such needs. The Task Group first met in March, 1977. It will asssess Federal usage of DBMS, survey existing DBMS specifications, collect information about implementations, and make formal recommendations for Federal action.

Sub-tasks include: Federal users survey, data conversion needs, time--sharing DBMS, DDD, DDL, DML, and Distributed Data Bases. In addition, specific subject areas include the impact of standards on DBMS technology progress, and the role of proprietary pakages in a DBMS standardization process.

6. CONCLUDING REMARKS

In conclusion, we can make some general comments:
All the work being conducted on the standardization of data base systems should be based on the contribution of the following groups of organizations:

- Professional computer and information system organizations (IFIP societies, etc.);
- Professional associations with a strong data processing aspect;
- Development and User groups (CODASYL, SHARE, GUIDE, etc.);
- Standards organizations (ISO, ANSIX3 DIN, AFNOR, BSI, etc.);
- National governmental bodies (NBS, CCA, etc.).

Although the advantages of standardization should be widely propagated in the DBMS area, no attempt should be made to create a new body for standardization on a European basis. The only feasible way to influence the future seems to be a cooperation between the appropriate ISO-bodies. The CODASYL proposal, for practical reasons, is now the only suitable candidate for a standardized DBMS.

While DBMS standards should be encouraged, it must be clear that they are non exclusive; the possibility for other ideas to coexist must always be open. For host languages such as COBOL or PL/1, it is clear that the first thing to standardize should be the user/application program interfaces. Indeed, the biggest obstacle to any DBMS changes are the application programs. Several different interfaces would be necessary to cope with different data models. If the TP-interface is to be taken into account the situation becomes even more difficult.

Furthermore, in order to ensure a coordinated development of standardization activities two preliminary working groups would be necessary:

- Terminology - to allow for better communication between developers and users of database systems
- Criteria - to show when a particular area was ready for standardization. These criteria may be assumed to refer to a sufficiently large class of

users, as well as economic considerations.

REFERENCES

1) "Experience of DBMS usage in Europe.

2) Berg, J.L., Editor "Data Base Directiions: The Next Steps", Proceedings of the Workshop of NBS and ACM held at Ft. Lauderdale, Fl. Oct. 29-31, 1975. National Bureau of Standards Special Publication = 451.

3) "Recommendations for Database Management System Standards", National Bureau of Standards Special Publication 500-51, Aug. 1979.

4) Sibley, H.E. "Standardization and database Systems", Proceedings of the Third International Conference on Very Large Data Bases, Oct., 1977.

5) ANSI/X3/SPARC DBMS Study Group, "Interim Report on Data Base Management Systems", Feb, 1975.

6) Tsichritzis, D. & Klug, A, "ANSI/X3/SPARC DBMS Framework", University of Toronto, Computer Systems Research Group, Technical Note R, July, 1977.

7) Berg, L.J., "A DBMS Architecture for Prudent Managers", Information & Management 1 (1978).

8) CODASYL Data Description Language Committee Journal of Development, 1978.

9) ISO TC 97/SC5/WG3, Interim Report on "Concepts and Terminology for the Conceptual Schema", Nov. 1979.

10) ANSI/X3H2, "Data Description Language" Working Paper, Oct. 1979.

11) CODASYL FORTRAN Data Base Facility Journal of Development, 1977.

12) British Computer Society "Data Dictionary Systems", Working Party Report, March 1977.

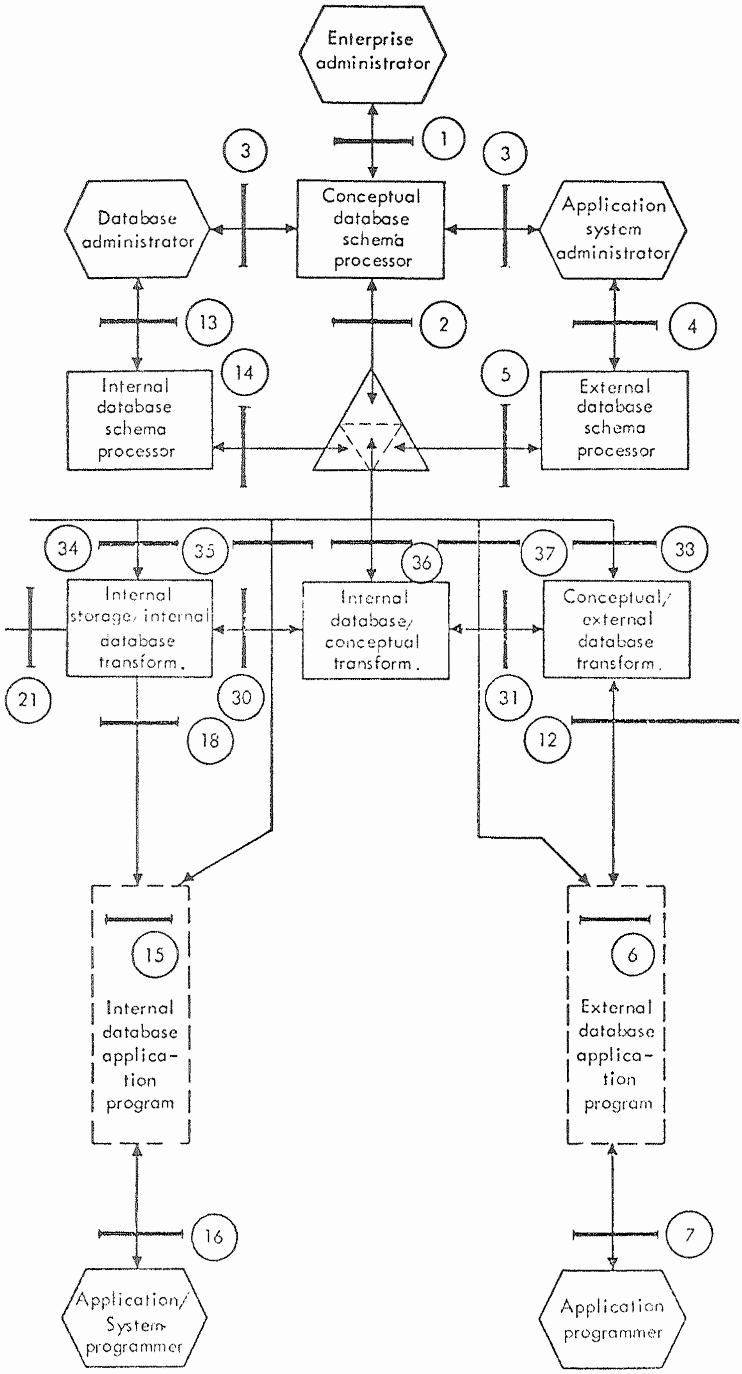

Fig. 1 — Partial schematic