TOWARDS A DATABASE BODY OF KNOWLEDGE: A STUDY FROM SPAIN

Coral Calero, Mario Piattini and Francisco Ruiz
ALARCOS Research Group
University of Castilla-La Mancha - Spain
{Coral.Calero, Mario.Piattini, Francisco. RuizG}@uclm.es

ABSTRACT
Databases, the center of today's information systems, are becoming more and more important judging by the huge volume of business they generate. In fact, database related material is included in a variety of curricula proposed by international organizations and prestigious universities. However, a systemized database body of knowledge (DBBOK), analogous to other works in Software Engineering (SWEBOK) or in Project Management (PMBOK) is needed. In this paper, we propose a first draft for this DBBOK based on degree programs from a variety of universities, the most relevant international curricula and the contents of the latest editions of principle books on databases.

Keywords: databases, curriculum

1. INTRODUCTION

Databases are extremely important within information systems since they are the basic nucleus in most cases (especially in management systems). This importance is demonstrated in a number of ways. First of all, the database market works with an enormous budget (over 14 billion US dollars) that has an annual growth of approximately 20% (Gartner group http://www3.gartner.com, IDC http://www.idcresearch.com, Meta Group http://www.metagroup.com).
A second determinant factor is that databases are included in all international curricula.
However, in our opinion no one has carried out a systemization of fundamental knowledge on databases such as those in the fields of project management, e.g. PMBOK (PMI, 1997) or in Software Engineering, e.g. SWEBOK (IEEE, 2001). In fact, it is surprising to note that databases are barely included in this second work. The cited Entity/Relationship (E/R) model is included as a notation on structural design within sub-section 5 (Notations for Software Design) in the Design Section, but the methods and strategies of database design are overlooked in sub-section 6 of the same section.
This lack of systemization may be in large part due to the fact that databases are associated to a few different disciplines: Software Engineering, Computer Science, Information Science and Information Systems (Khazanchi and Munkvold, 2000).
In general, all of the universities offering Computer Studies (be it Computer Science, Software Engineering or Information Systems) include some subjects on databases in their degree programs. In a national conference in Spain in 1997, a group of university Database professors from a variety of Spanish universities formed a team that sought to define a database curriculum to be used as a guide for Spanish universities to determine their degree programs. This curriculum was published in 1999 (Blesa et al., 1999).
However, it is well known that computer science, and therefore databases, undergo continual changes. What was a state-of-the art discipline a few years ago may now be obsolete or what was only beginning to appear in research laboratories then is fundamental nowadays.
In order to propose a DBBOK (DataBase Body of Knowledge) we decided to carry out a study from three different perspectives in order to determine the current situation of the world of databases:

- The contents of the database courses from a few Spanish universities in order to determine what is being taught in our universities and how the content compares to the Blesa et al (1999) curriculum.
- The contents of some of the most important international curricula.
- The latest editions of the main international books on databases.

In the next section we will briefly present the curriculum proposed by Blesa et al (1999), which we can use as the framework for studying the different proposals. The third section includes an analysis of the degree programs from seven Spanish universities, a variety of international curricula and some of the most important texts in the area of databases. In the fourth section we present our proposed DBBOK and our conclusions appear in the fifth section.

2. BLESA ET AL (1999) DATABASE CURRICULUM

At the *II Jornadas de Investigación y Docencia en Bases de Datos* (2nd Conference on Database Research and Education- JIDBD'97) held in July 1997, a team of Spanish University professors was formed to define the contents of database courses in the University.

The definitive version of the proposal drawn up by this team was published in *Novática*[1] over three years ago (Blesa et al., 1999). It was agreed that the subjects appearing in Table 1 were necessary.

Course	Hours	
Databases 1 (DB1)	90	Required
Databases 2 (DB2)	60	Required
Database Design (DBD)	45/60	Elective
Database Administration (DBAd)	45/60	Elective
Database Applications (DBAp)	45/60	Elective
Advanced Databases (ADVDB)	45/60	Elective

Table 1. Blesa et al. (1999) proposal

The DB1 course entails an introduction to databases and their architecture, the relational model, SQL language and an introduction to the E/R model and the normalization theory; while DB2 reviews the evolution of DBMSs, introduces concurrence, recovery and optimization, explains distributed databases and gives an introduction to object oriented databases and their future.

The DBD course covers the whole process of creating a relational database and the methodology used for its design, from the conceptual model to the physical design, including logical design and normalization and distributed database design.

The administration of databases is the objective of DBAd, which includes an introduction to the work environment and to the role of the administration, all of the aspects related to security, integrity, retrieval and evolution, as well as database optimization. It also includes the administration of distributed databases.

The DBAp course has two main blocks, the first on architecture of database applications: Client/Server architectures and programming techniques, embedded static and dynamic SQL and the implementation in multimedia databases; and the second on development environments for database applications: 4GL and RAD, generation of graphic user interfaces through virtual class libraries, development of Web interfaces for database applications and integration of database applications in office environments.

In Advanced Databases (ADVDB) the limitations of the relational model are shown and the RMT and relational V2 models, the OO databases (pure –ODMG-as well as object relational–SQL: 1999-), the active, deductive, distributed and heterogeneous databases, data warehouses and other database models (temporal, Web, document, multimedia, workflows, mobile, parallel, etc.) are explained.

[1] Novática is the *ATI (Asociación de Técnicos en Informática)* journal, one of the main Spanish computer associations.

3. ANALYSIS OF MAIN CURRICULA AND BOOKS

In this section we will present the study we have carried out on the degree programs of a few Spanish universities, the main international curricula and the most important books on databases. We have used the subjects described in the previous section as the framework for this section of the study.

3.1. Degree program contents in Spanish universities

The database courses from some of the most important universities were include in the study: the Polytechnic University of Valencia (UPV), the University of the Basque Country (EHU), the Polytechnic University of Madrid (UPM) – including the degree programs from the university, as well as from the School of Computer Science-, the Polytechnic University of Cataluña (UPC), the Carlos III University (UC3M), the University of Seville (US) and the University of Castilla-La Mancha (UCLM).

The study entailed determining whether or not the courses from the Blesa et al (1999) curriculum were included in the degree program of the university (Table 2 is a summary where T indicates that the content is treated totally and P that the content is treated partially).

	UPV	EHU	UPM	UPC	UC3M	US	UCLM
DB1	T	T	T	T	T	T	T
DB2	T	P	P	P	P		T
DBD	T	P	T	T	T	T	T
DBA	T	T	T	T		P	T
DBA		P					P
ADV	T	P	T	T	P		T

Table 2. The database courses in the Spanish Universities

Almost all of the subjects appearing in the DB1 course are covered in almost all of the universities. The DB2 course is a deeper look at the material from DB1 and the contents are not covered totally in these universities. The Database Design (DBD) course contents are included in most of the universities' degree programs; however, Database Administration (DBAd) appears less frequently. The Database Applications course is the least frequent of all. On the other hand, diverse aspects of Advanced Databases (ADVDB) are present in quite a few of these universities. It is interesting to note that UPV has a course specifically dedicated to documentary databases. This course is not highlighted in the subjects from the Blesa et al. (1999) curriculum.

3.2 International curricula

We will briefly present the most current versions of the curricula that we appraise to be the most interesting and prestigious in the field of computer science: ACM/IEEE CS 2001, IRMA/DAMA 2000, ACM/AIS MSIS 2000, ISCC´99, IFIP/UNESCO ICF-2000.

Although others such as ACM/IEEE CS 1991, ACM/AIS/AITP IS'97, UNESCO-IFIP of 1995, etc., have traditionally been studied, we did not include them because we believe that they have become outdated. However, we are conscious of the fact that many of their ideas persist in current proposals.

Due to space restrictions, we will only present the final summary. Table 3 also shows, using our personal judgment, whether the DB courses are covered totally (T) or partially (P).

	CC2001 ACM /IEEE	IRMA/ DAMA 2000	MSIS 2000 ACM/ AIS	ICF-2000 IFIP/ UNESCO	ISCC'99
DB1	T	T	T	T	T
DB2	P	T	T	P	P
DBD	T	P	P	P	T
DBAd	T	P	T	P	P
DBAp	T			P	
ADVDB	P		P	P	P

Table 3.Database courses in international curricula

It is important to point out that:

- A few curricula already present object oriented database models and even multidimensional (datawarehouses) models as basic knowledge for their students to acquire.
- There are some subjects, deductive databases for example, that are not specifically considered in any of the curricula.
- Distributed databases are also not considered within the basic knowledge material.

It is also worthwhile to note some of the subjects from the Blesa et al (1999) curriculum that do not appear in the international curricula. Among these:

- CS`2001 includes information storage and retrieval and digital libraries, and on the other hand, also considers hypertexts/hypermedia and data mining.
- IRMA/DAMA 2000 emphasizes the role of databases as a support for decision-making, database strategic planning, and information systems for executives.

Other interesting proposals for the field of databases are presented by Surynt and Augustine (2001). They propose to meet the demands of electronic commerce professionals by offering a variety of courses (in Business Schools), such as "Development of Databases for Electronic Commerce". On the other hand, there are growing numbers of experts in the academic world, as well as in the professional world, such as Mathieu and Khalil (1998) or Kahn and Strong (2001), who insist on the need to instruct students on Data Quality.

3.3. The texts

We selected the latest editions of the main international texts on databases. They are used in the most important international universities as a basis for the database programs. First we will present details on the books we worked with and then, a comparison of the contents to the Blesa et al (1999) curriculum.

The selected books were:

- Atzeni, P., Ceri, S., Paraboschi, S. and Torlone, R. **Database Systems. Concepts, languages and architectures.** McGraw-Hill. ISBN 0-07-709500-6. 1999. 612 p
- Connolly, T. and Begg, C. **Database Systems**. 3rd edition. Addison-Wesley. ISBN 0-201-70857-4. 2002. 1236 p.
- Date, C.J. **An introduction of database systems**. 7ª ed. Prentice-Hall. ISBN 968-444-419-2. 2001. 936 p.
- Elmasri, R. and Navathe, S. **Fundamentals of Database Systems**. 3rd ed. Prentice-Hall. ISBN 0201542633. 1999. 873 p
- García-Molina, H., Ullman, J.D. and Widom, J. **Database Systems: The complete book**. Prentice Hall. ISBN 0-13031-995-3. 2002. 1100 p
- Piattini, M. and Díaz, O. **Advanced database technology and design**. Artech House Publishers. ISBN 0-89006-395-8. 2000. 535 p
- Silberschatz, A., Korth, H.F. and Sudarshan, S. **Database Systems Concepts**. 4rd ed. McGraw-Hill Higher Education. ISBN 0-077-228363-7. 2002. 1064 p.

Again owing to space restrictions, we will only present the final summary, seen in Table 4. Using our own judgment, we indicate whether the subjects proposed by the texts totally (T) or partially (P) cover the database courses.

	Atzeni et al. (1999)	Piattini and Díaz (2000)	Connolly et al (2001) 3 ed.	Date (2001) 7 ed.	Elmasri and Navathe (2002) 3 ed.	García-Molina et al (2002)	Silberschatz et al. (2002) 4 ed.
Databases 1 (DB1)	T		T	T	T	T	T
Databases 2 (DBf2)	P	T	T	P	T	P	T
Database Design (DBD)	T		T		P	P	P
Database Administration (DBAd)	T		T	T	T	P	T
Database Applications (DBAp)							
Advanced Databases (ADVDB)	T	T	T	T	T	P	T

Table 4. Coverage of DB subjects in the main books

With respect to the DB1 course, files are not considered in the database texts, although all of the texts comment on the use of files for information storage prior to databases. The remaining points in this introductory course are covered by all of the studied texts, except for one of them that is, in fact, geared towards advanced databases.

In general, the DB2 course is also more or less covered in the texts studied. However, it can be observed that not all of the texts deal with evolution, the future of databases, and distributed databases (which can sometimes be out of the scope of a general text on databases).

The case of the DBD course is disparate in that four of the texts cover DBD totally or almost totally (except for some points that can be considered secondary in reference to database design), whereas other texts do not consider it at all. This difference may be due to the specific objectives of each text.

Only two books cover the Database Administration (DBAd) course satisfactorily and the remaining texts only deal with the most basic aspects with little depth. This may be due to the fact that general concepts on databases are often kept separate from DBAd concepts, which are often dealt with in specific texts or manuals on DBAd products (e.g., Loney and Theriault, 2002).

Database Applications (DBAp) is, beyond doubt, the course that is most ignored by these texts. Hardly any of the general books deal with the subject of DBAp.

There are recently published books, for example Melton and Eisenberg (2001), where contents of the type proposed for this course can be found.

As can be seen, almost all of the texts consider advanced databases to be important however, there is no consensus on which ones (with exception of the object oriented and distributed databases). One subject that is beginning to emerge with strength (data warehouses) has taken root over the past few years in specific texts, for example Inmon (1997). It is interesting to note that none of the studied texts deals with the subject of documentary databases, although again we can find specific books on this subject, for example, Ribeiro-Neto and Baeza-Yates (1999).

Other types of databases that are covered by some of the studied texts but not covered by the Advanced Databases (ADVDB) course are temporal databases, multimedia databases, mobile databases, component databases, secure databases, database quality, parallel databases, databases and the web and XML and databases. These DB types are also part of the "future" of databases and therefore, it could be advisable for them to be included in future degree programs.

4. OUR PROPOSAL

After studying the degree programs of the Spanish universities, the international curricula and the contents of the main books on databases, we propose a DBBOK that can be seen in Table 5.

We believe that the diffusion of object orientation, datawarehouses, the use of databases in Web environments, and XML language are substantially changing the outlook on databases and that a Computer Science Engineering student (from both 3-year and 5-year programs) cannot complete his/her degree without having "heard of" these subjects. Since the DB course may well be the only required course, we feel that the student should be presented with the various database models that s/he will encounter in the work environment. Following an introduction to DBMS (concepts, objectives, etc.), different architectures of DB (centralized, client/server, Web, parallel, distributed, etc.) could be presented, to be followed by the concept of data models.

The following step would be to briefly present network models, and then spend most of the time -approximately 60% of the total course hours- on relational models (including normalization and SQL language), followed by object-relational systems (including active rules and object extensions and the SQL: 1999 version), as well as pure object oriented model (including ODMG 3). The multidimensional model and data warehouse would also be presented, along with semi-structured models and XML. The course would be tied up with a general look at database administration and the development of applications with databases.

Distributed databases are not included because we believe that the subjects included in this proposal are currently more important.

However, distributed databases do receive some coverage in architectures and in the Advanced Databases (ADVDB) course.

COURSE	CONTENT
DATABASES (DB) **90/120 hours** **Required**	1. Introduction to DB 2. Architecture of DBMS 3. Concept of data model 4. Network models (elective) 5. Relational model 6. Object models 7. Multidimensional model 8. Semi-structured models (XML) 9. Introduction to database administration 10. Introduction to the development of database applications 11. The future of databases Practicum: Using SQL with relational, object-relational and multidimensional databases
DATABASE DESIGN (DBD) **60 hours** **Elective**	1. The database creation process 2. Methodology for database design 3. Conceptual modeling 4. Extended E/R model/UML model 5. Logical design 6. Physical design 7. Distributed database design Practicum: Complex cases of design using CASE tools and implementing the database on an existing product
DATABASE ADMINISTRATION (DBAd) **60 hours** **Elective**	1. Introduction 2. Confidentiality 3. Integrity 4. Concurrency 5. Recovery 6. Optimization of databases Practicum: Set up, creation of users, privileges, etc. Analysis of the optimizer and access pathways, refinements, etc. Lock and concurrence
DATABASE APPLICATIONS (DBAp) **60 hours** **Elective**	1. Client/server programming techniques 2. Static embedded and dynamic SQL. Using cursors. Comparison with CLL. 3. Working in 4GL and RAD environments. 4. APIs: ODBC, SQLJ, JDBC, etc. 5. Development of Web interfaces for database applications (CGI and others). 6. Integration of database applications for office environments Practicum: Complete development of a client/server application using an API including Web access.
ADVANCED DATABASES (ADVDB) **60 hours** **Elective**	1. Introduction 2. In-depth object oriented databases 3. In-depth multidimensional databases 4. Distributed and heterogeneous databases 5. Document databases 6. Mobile databases 7. Parallel databases 8. Geographic/spatial databases 9. Other database models: deductive, multimedia, secure. Practicum: Student presentation on subjects related to the latest advances in DB Design of an OODB using ODMG 3 Design an Object relational DB using SQL: 1999 Use a deductive DB prototype

Table 5. Our proposal

The design course content is very classical, especially when considering object orientation. In our opinion, it is becoming less and less justifiable to propose different courses on data design and Software Engineering Analysis and Design. We propose to have both aspects integrated in a Software Engineering/Information Systems design course, thereby offering the

student the possibility of understanding the crossed validation between data and processes, as well as using a single language (UML) for the design of both IS aspects. We feel that 60 hours would be needed for this course as well as for the others.

Database administration also has traditional content. It should be pointed out that in this subject there is a lack of quality texts on the subject and the ones that can be found deal with a concrete version or product and they do not offer an overall view of the subject.

We think that the DBAp course is interesting and that new standards, like SQLJ, JDBC, etc. could be incorporated into the course. The contents of this course could be included in others that deal with all of the aspects of application development (not only those related to databases).

Concerning advanced databases, the concentration would be on the object oriented and multidimensional databases. Furthermore, distributed, document, mobile, geographic and parallel databases would be stressed because of their significance. Deductive (which we feel has lost some interest nowadays), secure and multimedia models would appear in the "other models" section.

5. CONCLUSIONS

Databases are a core area in computer curricula. We think that is very important that we make the subjects taught at universities in line with technological advances and market needs in the world of databases.

In this article we present our proposal for a Spanish DBBOK. This study was centered on Spanish peculiarities and took into account some of the most important curricula as well as the most important texts in the field.

It would be very interesting to run similar studies in other countries in order to define a general and international Database Body of Knowledge.

REFERENCES

- Atzeni, P., Ceri, S., Paraboschi, S. and Torlone, R (1999). Database Systems. McGraw-Hill.
- Blesa, P., Brisaboa, N., Canivell, V., Garbajosa, J., Maudes, J. and Piattini, M. (1999), "Propuesta de contenidos en bases de datos de los planes de estudio de Informática". NOVÁTICA (Revista de la Asociación Técnicos de Informática), Nr. 137. 60-63.
- Cohen, E. and Kozminski, L. (2001). "Rationale for the IRMA/DAMA 2000 Model Currículum". Proc. of the 2001 IRMA (Information Resource Management Association) International Conference, 612-616.
- Connolly, T. and Begg, C. (2002). Database Systems. Addison Wesley. 3rd ed.
- Date, C.J. (2001). An introduction of Database Systems. Prentice Hall. 7th ed.
- Elmasri, R. and Navathe, S. (2002). Fundamentals of Database Systems. Prentice-Hall. 3rd ed.
- García-Molina, H., Ullman, J. and Widom, J. (2002). Database Systems: the complete book. Prentice Hall.
- IEEE (2001). Software Engineering Body of Knowledge (SWEBOK). Trial version, May. IEEE
- Inmon, W.H. (1997). Building the Data Warehouse, second edition, John Wiley and Sons, USA.
- Khazanchi, D. and Munkvold, B.E. (2000). "Is Information Systems a Science? An Inquiry into the Nature of the Information Systems Discipline". The DATA BASE for Advances in Information Systems, 31 (3), 24-42.
- Loney, K. and Theriault, M. (2002). Oracle9i DBA Handbook. Oracle Pc press.
- Mathieu, R.G. and Khalil, O. (1998). Data Quality in the Database Systems Course. Data Quality Journal, 4 (1), 1-12.
- Melton, J. and Eisenberg, A. (2000). Understanding SQL and Java together. Morgan Kaufmann.
- Piattini, M. and Díaz, O (2000). Advanced database technology and design. Artech House Publishers.
- PMI (1997). Project Management Body of Knowledge (PMBOK). Project Management Institute.
- Ribeiro-Neto, B. and Baeza-Yates, R. (1999). Modern Information retrieval. Addison-Wesley and ACM Press.
- Silberschatz, A., Korth, H.F. and Sudarshan, S. (2002). Database systems concepts. McGraw-Hill Higher Education. 4rd ed.
- Surynt, T. and Augustine, F. (2001). "E-Business Technology: A Component-Based Curriculum for the Future". Proc. of the IRMA International Conference, 686-689.