Gio Wiederhold, April 30, 1992

Database Use and Technology in Japan

JTEC Study, March 1991

Participants
David Beech, Charles Bourne, Nick Farmer, Sushil Jajodia,
David Kahaner, Toshi Minoura, Diane Smith,
John Miles Smith, Gio Wiederhold (chair)

As a draft, this document remains subject to revision. It was sponsored by the National Science Foundation (NSF) and the Department of Energy under NSF Grant ECS-8922947, awarded to the Japanese Technology Evaluation Center at Loyola College in Maryland. The views expressed herein are solely those of the authors and do not necessarily reflect those of the United States Government, the authors' parent institutions or Loyola College.

Table of Contents

Section	page
Foreword	i
Table of Contents	iii
Executive Summary	1
Qualitative Comparisons Between U.S. and Japan	2
1 Introduction	7
Coverage	7
Motivation	7
Overview	7
2 Database Research Infrastructure	9
Role of the Japanese Government	9
Funding Sources	15
International Collaboration	15
U.S. Influence and Faddism	15
Conclusions	15
3 Technology Topics	16
Object Databases	16
Knowledge Bases	23
Multimedia Database Systems	24
Interoperable Databases	33
Database Hardware	35
Security	40
Database Tools	41
Database Management Systems	42
4 DBMS Product Infrastructure	45
5 Database Use Industry	49
Database Search Services	49
Databases as Export Products	52
Database Technology	57
6 Summary	60
Acknowledgments	61
Appendices	62
Participants' Biographies	62
Sites Visited and Contacts	66
Trip Reports	71
Bibliography	108

Executive Summary

This report presents the findings of a group of database experts, sponsored by JTEC, based on an intensive study trip to Japan during March 1991. Academic, industrial, and governmental sites were visited. The primary findings are that Japan is supporting its academic research establishment poorly, that industry is making progress in key areas, and that both academic and industrial researchers are well aware of current domestic and foreign technology. Information sharing between industry and academia is effectively supported by governmental sponsorship of joint planning and review activities, and enhances technology transfer. In two key areas, multimedia and object-oriented databases, we can expect to see future export of Japanese database products, typically integrated into larger systems.

Support for academic research is relatively modest. Nevertheless, the senior faculty are well-known and respected, and communicate frequently and in depth with each other, with government agencies, and with industry. In 1988 there were a total of 1,717 Ph.D.'s in engineering and 881 in science. It appears that only about 30 of these were academic Ph.D.'s in the basic computer sciences.

Industrial research is well supported. Support services are marvellous in terms of hardware, and researchers are encouraged to publish. Industrial Ph.D.'s, given as recognition for publications (following the British model), are common. The number of industrially grounded Ph.D.'s is roughly equal to the number of academic ones.

The Japanese government, overall, seems to have less influence on database research directions than is perceived by outsiders. The funding supports some laboratories and projects, but academic researchers have considerable flexibility in choosing the directions for government-sponsored research. The level of government funding for industrial laboratories is relatively low and does not influence market-driven priorities. However, these projects do require regular meetings of academic, government, and industrial researchers, increasing mutual awareness and understanding. Similar information sharing results from review panels of NIH and, sometimes, NSF, where proposal reviewers learn much from each other. The presence, in Japan, of industry participants on these panels, broadens their relevance and enhances technology transfer.

Technology transfer is enhanced by governmental requirements and support of review boards joining academia and industry.

The database research we saw in industry is very much oriented toward support of development and new technology. It relies for its conceptual input greatly on publications from the U.S. and Europe. The researchers are well-read and often well-connected with foreign academic sources. They provide an important path for technology transfer.

An important driving mechanism in database development is the Japanese capability in the area of developing electronic products. High quality image acquisition, transmission, storage, display, and digitized voice data are emphasized. Database management systems are being expanded to provide support for such "multimedia." While fundamental database management systems are not being advanced, the incorporation of multimedia support will change their character greatly. We can expect that purchasers of systems with multimedia requirements will, with Japanese image-processing hardware, acquire Japanese database software. This field is likely to grow rapidly. Computer-aided-design, computer-aided-education, and other application areas that are critically dependent on graphics will be the initial users of this technology.

We expect that multimedia technology will provide an important path for the introduction of Japanese database software into European and U.S. markets.

Japanese hardware for computer systems is roughly equivalent to U.S. systems, except again in the areas of multimedia support and optical mass storage, where the Japanese have a substantial advantage. Parallel architecture and database accelerator schemes are of active interest.

Hardware support for database systems is provided equally well by Japanese and foreign companies. Sony is an important supplier of workstations, but U.S. companies such as SUN Microsystems are also well represented. Japanese mainframe-based database systems are similar to their U.S. counterparts, but this market shows less growth and is less fluid.

Relevant research on topics such as database accelerators is being pursued. This work can be seen as a specialization of research into parallel computation, which is pursued by computer researchers everywhere with equal intensity. The payoff is likely to come as demands on database computation increase.

The JTEC study also surveyed the industry that maintains databases and sells information retrieved from these databases. In this area Japanese databases provide useful services internally, but are not in a position to export their services. There is substantial use in Japan of Western databases, both via U.S. and European vendors and via Japanese resellers. Some internal developments are oriented towards providing image data as well. Providing such services on an international scale awaits high capacity communication lines and acceptance of standards. In this area the relative situation seems stable.

While Japan is not viewed today as a world-level player in the database area, the infrastructure is in place for Japan to make important contributions in areas where there is high growth potential and linkage with consumer hardware.

Qualitative Comparisons Between the U.S. and Japan

The panel has prepared a qualitative comparison of the present status and trends in database systems research in the U.S. and Japan. For the purposes of this comparison, the subject matter covered by the panel was divided into seven subtopics: mainframes, hardware-pc, workstations-servers, storage, database content, database management systems, and new database technologies. Figures 1 through 7 represent the panel's findings in each of these subareas.

1. Introduction

This report presents the findings of a study group composed of eleven database specialists who visited Japan from March 23 to March 29, 1991. The group included representatives from the database systems industries, database service industries, academia, and government. This was their first visit to Japan for several of the participants, while others had interacted with colleagues in Japan for many years. Short biographies of the participants are given in Appendix A.

Thirty-one sites were visited by teams with from two to six participants. The entire study group team was together for an initial visit to the Ministry of International Trade and Industry (MITI), and for planning and wrap-up sessions. The sites were diverse, and included major Japanese companies, some small specialized companies, government laboratories, and universities. They are listed with the names of the principal contacts in Appendix B.

Short site reports from the individual visits are in Appendix C. We had many pleasant social interactions as well, but are omitting them from this report. This omission should not diminish their importance, nor the appreciation we have towards our hosts.

1.1 Coverage

It is obvious that in the short timeframe of the study, complete coverage could not be achieved. We apologize to those Japanese colleagues we have missed, either because of scheduling problems or lack of awareness. We have tried to keep this report factual so that the effects of selection bias are minimized; generalities based on our observations are in the concluding section.

1.2 Motivation

It is important for U.S. researchers, technologists, marketing specialists, and negotiators to understand Japan. It is an important country that is playing a critical role in business and world affairs. This importance extends to the area of our expertise, computing and databases. We all rely on Japanese hardware in our computer systems, and it is likely that we will deal increasingly with Japanese data and software as well. We should not view this as a threat or, in a defeatist sense, as a loss. Instead, it appears that we can best envisage a world in which our systems are composed of a wide variety of components, organized and accessible in a manner that will serve all of our needs best.

Japan presents a social homogeneity which distinguishes it from the variety found in the U.S., Europe, or other regions of similar size in the world. But within this social structure, we find individuals and companies with a wide range of approaches, outlooks, methods, and aspirations. Specific, novel, and interesting examples cannot be viewed as a general prototype or a trend, but merely as a potential seed for new possibilities. It is important to note that this is not a report on databases for Japan, Inc., but rather a collection of reports that illustrate a range of research and application interests.

Readers of this report may be tempted to extrapolate our findings to other areas of computer science or even science in general. We do not believe that such extrapolations are necessarily valid.

1.3 Overview

The remainder of this report is organized into several sections. We begin with a short section on the Japanese database research and development establishment to place the remainder in context.

The bulk of this report consists of small summaries of individual topics stressing the

state of the art, but also indicating likely changes. The state of the industry that uses databases to disseminate information is the topic of Section III.

A subsequent section focuses on integration, since the capability of future database and information systems depends on the smooth meshing of its components.

We list the contributing authors in the various sections, but the responsibility for errors and omissions rests with the chair of the study, since many editorial changes were introduced.

2. Database Research Infrastructure

As is the case in other countries, the Japanese government uses databases to administer a variety of governmental activities. However, not all of the relevant databases are inside government offices. It does appear that the Japanese government has, through various measures, done more than most other governments to further database use and technology. The effort has not been massive, but has had subtle, yet important effects on the structure supporting technological progress in the database and in other related areas.

2.1 Role of the Japanese Government in Database Activities

Authors: Sushil Jajodia and David Kahaner

After an overview of the relevant overall structure we will cover in more detail the ministries and their offices involved in database activities.

2.1.1 The Japanese Government's Structure and Use of Databases

The Government of Japan (GOJ) is organized around the prime minister who heads the government. He is assisted by a cabinet composed of ministers (education, justice, foreign affairs, etc.). Each of the ministries has its own budget, and to some extent each provides support for database activities. In 1987, at an Inter-ministerial Council, a plan was formulated to promote GOJ databases; every ministry and agency is expected to improve their databases in accordance with this plan. Furthermore, in 1989 the Cabinet first decided to promote the open system interconnection (OSI) model in accordance with the international standard. As a result of these actions, databases have become an important aspect of many parts of the Japanese government.

Two ministries are significantly involved in database use, the Ministry of International Trade and Industry (MITI), and the Ministry of Education (MONBUSHO). Both MITI and MONBUSHO support significant database activities through grants for research and development, tax incentives, loans, and other mechanisms. The ministries have subunits that also provide support. Like NSF, MONBUSHO supports the national universities and provides research grants to academic researchers. For example, Japan's National Institute for Education Research, its National Education Center, and its National Women's Education Center are funded to produce various new types of databases. MONBUSHO also administers the National Center for Science Information System (NACSIS). MITI runs the highly automated Patent Information Organization (JAPIO). The Ministry of Post and Telecommunications (MPT) also provides support for database activities, by supervising the compilation of an online database directory and studying how the new digital communication capabilities provided by ISDN can be used to improve database distribution.

Another important part of the Japanese government is the Science and Technology Agency (STA), which reports directly to the prime minister's office. STA, in addition to having research programs, also supports the Japan Information Center of Science and Technology (JICST) with its extensive databases.

The GOJ also has a legislative arm, the Diet, which supports the National Diet Library (NDL), which is similar to the Library of Congress. At the NDL, all publications (including informal *grey* literature) published in Japan are deposited, except for classified military documents and other materials withheld from the public. The NDL has developed a database of its catalogue, and provides it on tapes and CD-ROMs, and via an online system to public libraries.

The discussion below is organized around the GOJ units. It is not meant to be a complete synopsis of GOJ database science activities, but only describes those aspects of significant interest to the JTEC database team.

2.1.2 MITI

The Ministry of International Trade and Industry supports several activities that affect database use and technology.

(a) Database Promotion Center. Japan is aware that it is behind the U.S. in the use and development of databases, particularly for science, and there is a strong effort being made to improve the position of the Japanese database industry. The Japan Database Promotion Center (DPC), established in 1984, has as its goals the promotion of database construction; research and development of the basic technologies related to databases; establishment of efficient clearing services; education propagation and training associated with databases; and international information exchange informing other countries about Japanese databases.

Japanese online databases in science and technology were first developed by JICST and JAPIO in the mid-1970s. Business databases were established a few years earlier. This sequence is about ten years behind corresponding efforts in the U.S. The delay was due in part to the lag in computer technology for handling the Japanese language and in part to the reluctance of the Japanese to sell intangibles, in this case information, as a product. The Japan Database Industry Association (DINA), established in 1979, now has over 100 members and is quite active. Furthermore, since the early 1980s both the information industry and the GOJ have attempted to promote database development and use. Since 1983, MITI has published an annual database directory that provides a comprehensive list of both Japanese and foreign databases that can be accessed in Japan. The 1989 directory, in Japanese, is about 3,000 pages long and lists over 3,000 databases. An English introductory version to Japanese original databases is about 100 pages long.

The DPC is a non-profit organization supervised by MITI dealing with databases. It is founded by private companies and keeps track of various database statistics such as the following: "MITI's survey reports that the Japanese database service industry's sales reached ¥106.3 billion in 1988 ... In their *Outlook of the Information Industry for the year 2000*, submitted in 1987, estimated sales were placed at ¥144.5 trillion for the entire information industry in the 2000's, and at ¥3.4 trillion for information provision services including database services. If we estimate the sales of database services in the early 2000's very roughly on the basis of ¥106.3 billion of 1988 with a tentative growth rate of 30%, it could be expected to reach ¥2.5 trillion." In addition to the database directory, DPC also publishes an annual *Database in Japan 19xx*, which should be referred to for additional statistical data.

Information is of growing importance to Japan's economy. The "Information Society" is a standard phrase used to describe the near term social environment. MITI has formulated the following sequence of policies to develop and improve databases:

- Promote important database production.
- Encourage international cooperation regarding databases.
- Adjust tax policy to encourage database-producing corporations.
- Support system development to increase database operational efficiency.
- Create a database directory.
- Support research into the production and organization of databases.
- Produce public databases and encourage movement of GOJ data to the private sector.

The DPC provides financial support for private organizations for database research and construction. For example, the Sharp Corporation has been funded to study a concept called Set Theory DataBase (STDB) for document retrieval.

MITI has the additional role of sponsoring certain national projects. The Institute for New Generation Computer Technology (ICOT) is a national project with an object oriented database as a subproject (see below). Another is the Interoperable Database System Project (INTAP), which endeavors to establish an open system architecture in Japan. It appeared to the JTEC visitors that the project has been mostly passive. It focuses on validation of conformance with OSI standards. Dr. Akio Tojo (Managing Director of the Information Technology Promotion Agency) considers this project to be one of the most successful MITI R&D projects in the IT area. MITI also supports the International Multimedia Association, which became a non-profit organization supervised by MITI in April 1991 and is expected to establish international relationships with relevant organizations. Finally we were told that MITI itself planned no new international projects in the database area.

(b) Japan Patent Information Organization. Data relating to patents has been processed by the Japan Patent Information Organization (JAPIO) since 1985. It currently has about 250 staff members. In addition to printed documents (or "Gazettes") there is a computerized information service built around large databases. Computer equipment includes six Hitachi mainframes (M682H, M680H, M680D x 2, M662K), 420 gigabytes (GB) of magnetic disks, and 3,100 GB of optical disks. The full texts of all Japanese patents are electronically filed and total entries in the database exceed 31 million. All the database data are in Japanese except for the foreign documents bibliographic database and English abstracts of Japanese unexamined patents. The latter have been available through the Orbit online service since 1976. JAPIO itself offers an online service called Patent OnLine Information System (PATOLIS) which has issued about 4,000 issued passwords. Three basic services are offered:

- Retrieval, using a special interactive terminal. Search keys corresponding to patent classification, applicant, keyword, and so forth, can be combined using the logical operators AND, OR, and NOT.
- File history in response to patent number specification.
- Correspondence search for foreign patents.

Western users can access the database via a leased line or through commercial services such as VENUS-P. Because most of the data is in Japanese, it is not surprising that all but a few of the passwords are issued to users in Japan, although in 1989 there were far more passwords issued in Europe than in the U.S. For more detailed patent information, JAPIO also offers a batch retrieval service which extends the keys allowed in PATOLIS. JAPIO also supplies the full text of patents and patent publications on optical disks.

(c) Sigma Project. The Sigma project was initiated by MITI in 1985 as a five-year project to improve software quality and productivity in a standard operating system environment. No new database management systems (DBMSs) were to be developed. The policy was to utilize commercially available RDBMSs on Unix workstations, with the manipulation of Kanji an important criterion. The project fostered the use of existing RDBMSs on Unix systems, and one such system (UNIFY) has sold more than 25,000 copies in Japan.

(d) ICOT. The Institute for New Generation Computer Technology was established in 1982 by MITI to provide a focus for original research in new technologies, specifically artifi-

cial intelligence (AI) and knowledge bases having persistent databases as a component. Its cornerstone was a ten-year project called "Fifth Generation," which focused on the construction of a logic-based parallel inference computer. Within this project ICOT developed two deductive databases, PHI [Hani:91] and CRL. Prof. Nishio and several other professors from the Obase consortium participate in an ICOT working group on deductive object-oriented database technology. Theoretical results have been obtained about query evaluation algorithms. A Deductive Object-Oriented Database (OODB) language, QUIXOTE, is also being developed.

One of the most important functions of ICOT is that it enables young database researchers to come together outside of their parent companies. Several working groups have been formed and meet periodically.

2.1.3 Science and Technology Agency

The Science and Technology Agency (STA) of the prime minister's office has a budget of approximately $3.8 billion. Two activities are worthy of note in the context of databases.

(a) JICST. The Japan Information Center of Science and Technology (JICST) is the country's central body for distributing science and technology information. (A related organization is NACSIS, described under MONBUSHO in Sect. 2.1.4.) JICST's databases store scientific and technological literature, factual and other data, that are made available to any user on a fee-for-service basis. The information is collected from more than 15,000 journals of which more than 8,000 are from outside of Japan. The database has almost 8,000,000 citations and abstracts, of which about half are domestic. Also included are technical reports, conference proceedings, research projects carried out at public research institutes, and government reports that are often difficult to obtain through other means. Financing is via government support and user fees; JICST's 1991 budget is about ¥15 billion. JICST has a main office in Tokyo, a library in Tsukuba Science City, and branch offices in a dozen Japanese cities, Paris, and Washington, D.C. JICST has a staff of about 325. Further information about JICST is provided in Section 5.

JICST is not a database research organization in the sense of developing new concepts. However, in addition to providing an online service, JICST does engage in applied research in machine translation, expert system technology for abstracting and indexing full text Japanese databases, fuzzy information retrieval techniques, friendly user interfaces, and development of a chemical substance molecular weight database for organic compounds, inorganic substances, metallic materials and high polymers.

JICST also has been active in promoting a National Information Policy in Japan. In the future, JICST will:

- Investigate the problems of using databases that result from overdiversification of dissemination modes.
- Enhance bibliographic database coverage to include more of the grey literature.
- Promote the construction and dissemination of factual databases.
- Encourage standardization.
- Develop full text databases.
- Promote networking between information processing and information dissemination centers.
- Extend dissemination systems to include multimedia and user friendly services.
- Promote the construction of English language databases.

(b) Other STA Projects. Two other projects were discussed with the JTEC team. The first project was the "Research on the Development of Knowledge Base System to aid Chemists in Designing Chemical Substances and Chemical Reactions." This project was funded at a level of approximately ¥100 million over five years, and included six universities, fourteen government research labs, and one industrial organization (NTT). The work involved software and systems developed at different places, with the core database system (SYNDES) from Tsukuba University. Three or four of these systems are to be demonstrated in October 1991 in Tokyo. We were told by Prof. S. Kito (Aichi Inst. Tech.) that the system developed in Kyoto was the largest. Since the total amount is quite modest (averaging less than $8,000 per year per institution), it is obvious that the main effect of STA funding was to support cooperation, integration, and technology transfer rather than fundamental research.

The other project was the "Development of a Self-Organizing Information Base to Aid Researchers in Creative Research." This project has the goal of creating an intelligent large hypertext database system. The first phase of this project began in April 1991 and will last three years; the second phase will last an additional two or three years. Project members include JICST; Ryukoku, Tsukuba, Kobe, and Hokkaido universities; and ETL, NTT, FujiXerox; and other industrial partners. STA provides funding at the level of $100K-150K per university per year. ETL administers the projects for each university.

2.1.4 MONBUSHO

The Ministry of Education, in addition to its general role of supporting the national universities, supports some specific database activities.

(a) NACSIS. The National Center for Science Information Systems (NACSIS) is a national inter-university research institute. It is the nucleus organization for scholarly information transfer, providing the services of shared cataloging, compilation, processing, and dissemination of information including (but not limited to) scientific and technical fields. The significant differences between NACSIS and JICST are:

- Most of JICST's users are from industry (see Section 5.2); NACSIS mainly serves the academic research community.
- JICST is limited to scientific and technical information; NACSISOB incorporates the social sciences and humanities.
- JICST does its own abstracting; NACSIS primarily collects, purchases or leases data.

NACSIS has a full-time staff of almost ninety, about one-third of the size of JICST. It was a branch laboratory of Tokyo University until 1986, but has been independent since then. NASCIS projects a very energetic and forward-looking image. In fact, while most of the staff are in the Operations or Administrative departments, almost two dozen members of the Research Department hold university titles (including seven professors) and some have joint appointments. More information on NACSIS, and its system SIS, can be found in Section 5.2.

(b) Grant-in-Aid. With the growth of large databases it will be necessary to learn how to mine them for nuggets of useful information or knowledge. Using the knowledge inherent in the data, very large knowledge bases can be built up. A new program has begun this year on Knowledge Discovery in Databases. This general area will probably become one of the most important new research topics in Japan.

(c) Academic Groups. University researchers supported by MONBUSHO are engaged in a variety of database projects. The projects are typically quite small.

Japanese university researchers face some serious difficulties. The number of graduate students interested in database research is declining. To begin with, the total number of graduate students in computer science is very small, and many of these are entering such "popular" areas as artificial intelligence.

About half of the graduate students are from outside Japan (mainly from the People's Republic of China), and almost all of these students leave Japan once they finish their education. This differs, of course, greatly from the U.S., which is better able to integrate productive individuals into its social fabric.

Japanese students are less motivated to continue on to academic graduate programs since industry relies on its own training programs. Moreover, capable industrial researchers can obtain a Ph.D. degree by showing that their publications are worthy of such a degree.

Although the Japanese government is committing more resources to build research infrastructure in Japanese universities, some fundamental problems are yet to be resolved.

(d) Professional Societies. There are two professional societies that promote database research: the Information Processing Society of Japan (IPSJ) established in 1970, and the Institute of Electronics, Information, and Communication Engineers (IEICE), established in 1971. There are two special interest groups within them that focus on databases: the Special Interest Group on Database Systems (SIGDBS) associated with IPSJ, and the Special Interest Group on Data Engineering (SIGDE) within IEICE. The IEICE is roughly equivalent to the IEEE in the U.S., while the IPSJ parallels the ACM. Each society has its own journals. Each holds annual conferences and workshops on various aspects of databases.

Two new professional societies, Japan Society for Software Science and Technology and Japan Society for Artificial Intelligence, were founded recently. Journals published by these societies also have a growing number of articles related to databases [Nish:91].

There is now a Japanese-printed English language publication: the monthly *IEICE Transactions*. It has non-Japanese associate editors. The journal covers

1 Fundamentals of electronics, communication and computer sciences,
2 Communications,
3 Electronics, and
4 Information and systems.

(Prior to 1992 these journals were issued bound together, but they will now be distinct.) This Japan-originated English publication is not well-known now outside of Japan. *IEICE Transactions* also contains English abstracts of Japanese-language papers published in the *IEICE Computer* and the *IEICE Information Processing* journals. These two journals cover the IEICE areas of Information and Systems. The volume of abstract material is modest and might be a candidate for reprinting in selected IEEE Computer Society newsletters, such as the *IEEE CS Data Engineering Bulletin*. *IEICE Transactions* also lists the English titles of the unrefereed workshop reports from the forty-seven technical groups of the IEICE.

2.2 Funding Sources

Tenured Japanese university researchers receive some support from the Japanese government directly. However, most research support for significant efforts comes from industry, which means that university researchers must show that technology transfer can take place effectively. Thus, there is a great deal of emphasis on prototyping.

2.3 International Collaboration

Japanese university researchers seem to collaborate freely with U.S. and European university researchers. Many Japanese industries provide study leaves, typically for one year, for their promising researchers. Several researchers we met have spent an extended period of time at U.S. universities. Most international conference series are now being held occasionally in Japan, increasing the potential for greater cooperation in the future.

The problems of language differences fall mainly on Japanese researchers. Their ability to read English is a significant benefit to them.

2.4 U.S. Influence and Faddism

While the basic research efforts are very interesting, we did not hear or see any novel research direction or idea in the database area that was developed by Japanese university researchers. Most of these efforts seem to follow in the footsteps of U.S. researchers.

When a new idea reaches a certain level of visibility it is easily picked up by multiple Japanese researchers at that time. Since their base research funding is not tied to commitments made in formal research proposals, they can redirect their attention rapidly. This flexibility can be advantageous, but can also lead to faddism and abandonment of long-term, basic research directions. In this sense academic and industrial research show some similarities, and probably some empathy as well.

2.5 Conclusions

There is no doubt that Japanese university researchers are very productive and maintain high levels of research activity. This is especially impressive since their overall number is small, and the amount of support they receive in the form of research funding and from graduate students is quite small. Although government is taking steps to improve the situation, it is likely that the shortage of graduate students will continue to exist. Unless this most fundamental problem is solved, the enormous potential for graduate research at Japanese universities cannot be fully realized.

3. Technology Topics

In this section we report on database technology as found in Japan, focusing on nine topics:

1 Object Databases
2 Knowledge Bases
3 Multimedia Databases
4 Interoperable Databases
5 Database Hardware
6 Database Applications
7 Database Security
8 Database Tools
9 Database Management Systems

We stress areas of great activity, and report only briefly on topics where activity is less.

3.1 Object Databases

Author: David Beech

There is considerable activity in Japan in the field of object databases. Indeed, there is a worldwide sense that objects are likely to be at the heart of the next generation of database technology. This may well be a self-fulfilling prophecy, since anything can be described as an object, and the technical sense of the term is still negotiable in the object database world. Even in Japan, the land of consensus, there is no commonly agreed-upon definition of what an object database is — consensus takes time.

It is therefore necessary for us first to identify the main contending emphases to establish a framework to which the Japanese work can be related. We will then survey ongoing work in industry, in the universities, and in a consortium that has both industrial and academic participation, before summarizing our findings.

3.1.1 What is an Object Database?

There are perhaps three major trends in the development of object database systems — to concentrate (a) on multimedia objects, (b) on complex, structured objects, or (c) on programming language objects. There is considerable overlap between these approaches conceptually, but the differences in emphasis can produce very different systems.

(a) Multimedia Objects. The simplest use of object concepts occurs in the storage and retrieval of multimedia information. The binary representation of a voice message, a color image, or a video recording is intuitively thought of as an *object* due to its individuality and its synthetic nature — it is as though it springs into life when created, and it is highly unlikely that it is the same as any other object unless there has been explicit copying. In everyday life, it is too unwieldy a model, except to a handful of mathematicians, to consider all such multimedia phenomena to be instances of a pre-existing domain of immutable *values*.

Yet in a database system, there is no conceptual dividing line between a short bit string and a long one. For example, a string of 2 gigabytes offers a domain of $2^{(2^{34})}$ possible values. Thus a relational database system could treat images as values, and its storage management could be adjusted to give efficient support to these vastly enlarged domains. Since there is some perceived marketing value in claiming object support, one is likely to encounter such claims made about relational or other systems capable of handling such "binary large objects" (BLOBs).

The positive side of this use of the term "object" for a BLOB is that it is a first step in the direction of a more general conceptual model that would fully justify the terminology.

If a relational system is extended, not just with a type `LONG`, but with types `VOICE`, `IMAGE`, and `VIDEO`, say, and these types have appropriately defined functions and can be given more specialized subtypes, then such a system would begin to share the rich model of other object systems, although perhaps offering only limited extensibility.

When this perspective on multimedia objects is allied to the market potential for their support by databases in the office — and the home — of the future, it should be clear that this is an important and ultimately reputable use of the terminology of object databases. Multimedia capability is already becoming an important emphasis in practical systems, as described in section 3.3 of this report.

(b) Complex Objects. Many applications still store their data in files rather than in databases, because their data structures often are not a good match with those supported effectively by database systems (such as rows in relational tables). The data may be in large aggregates composed of many interconnected simple structures. An aggregate of information often corresponds to an object in the user's mind, such as an aircraft wing, an architectural drawing, or a programming system, and these are usually referred to as *complex objects*.

There is considerable demand for more convenient and efficient database support for complex objects in such fields as computer-aided design, manufacturing, and software engineering, since a database system provides more functionality than a file system in controlling multi-user access and in automating recovery from system failures.

Programming languages are strong in providing complexity within a single structure, but usually rely on the programmer to *navigate* within and between structures rather than providing higher-level aggregation. This can lead to unacceptable performance in a database system if the navigation requires location of data in disk blocks instead of in internal memory. Systems that try to overcome this problem provide various forms of paging, indexing, physical clustering, and caching.

Although there is some work in complex objects underway in Japan, we did not find a strong emphasis on it.

(c) Programming Language Objects. Currently much programming and system design is said to be *object-oriented*. As applied to database systems, "oriented" is little more than a noise word, especially in the Orient where it is hard to know which way to turn to become oriented. In the programming language world, the term tends to be associated with the more interpretive object languages, such as SMALLTALK-80 and CLOS (the Common Lisp Object System). We shall not use the phrase unless we wish to give it this connotation. There is an older tradition, still very much alive, of compiled object languages that originated with SIMULA-67 and has led to such languages as C++ and EIFFEL. Most object database systems that emphasize programming language objects currently concentrate on a C++ interface.

The common characteristics of object languages are their type (or class) systems that classify objects according to their interfaces (the attributes of the objects; and the operations that may be performed on the objects). The types may be placed in a hierarchy where more specialized subtypes inherit the interfaces of their supertypes, adding their own extensions to the interface. The implementation of the type is encapsulated within the type definition, so that systems may be constructed in a modular fashion.

In object database systems emphasizing programming language objects, users are permitted to define new types, including the functions that operate on them, so that such systems can be extended and can embody user-defined semantics for their types. Such systems clearly have a general model that makes them suitable for handling the various multimedia

objects discussed above, provided that the storage management of large objects is suitably optimized. They are also capable of providing the basic underlying structures for complex objects, and of addressing such performance issues as physical clustering.

There are two main approaches to interfacing object languages and object databases. The first is driven from the programming language end, and aims to provide a "seamless" interface so that a minimal language extension will handle persistent objects similarly to transient ones. This usually means that the database system supports only a single language, at least initially. The other approach is driven from the database end, and aims to provide a language-neutral database system using a type or class system very similar to those of several programming languages. The language interfaces may then take various forms, such as embedded database language statements, function calls, or a seamless syntax that can map to these.

3.1.2 Industrial Projects in Object Database Systems

Although there are not yet any Japanese object database products, whereas in the U.S. there are several, we found a number of substantial prototype systems in Japan and extensive research activity that generally showed a good blend of theoretical and practical considerations. Japanese researchers are well-informed in this area and appear to find it congenial territory, which helps contribute to rising standards in both the content and presentation of their database literature.

Three industrial object database prototype systems were described to us: MANDRILL (Hitachi), JASMINE (Fujitsu), and ODIN (NEC).

Mandrill. The MANDRILL system [Yama:89] at Hitachi's Systems Development Laboratory at Kawasaki aims to improve database productivity, applicability, and performance. The wide range of applications includes not only Computer-Aided-Design (CAD), Manufacturing(CAM), and Software-Engineering (CASE), but also Office Information Systems (OIS), Management Information Systems (MIS) and Geographic Information Systems (GIS). Many of the points of the Object-oriented Database System Manifesto (presented by Atkinson, et al. in Kyoto, 1989) are being addressed. A preliminary version 1 of MANDRILL was built in 1988/89, with the version 2 prototype scheduled for completion in 1991, and application studies in C planned for 1992.

The approach at Hitachi has been to build an object layer over an existing DataBase Management System (DBMS) and its file system interfaces, with the intention of supporting multimedia information as well as conventional data. The central part, MANDRILL/CORE, first provides a logical media layer to establish system-independence from the underlying database and file systems, and then builds its object layer over it. Applications access the object layer via statements of the MANDRILL/QUEST database language embedded in programming languages.

The QUEST language is somewhat like SQL, the current standard database query language based on IBM's research, although no attempt has been made at a close integration or compatible extension. QUEST commands are treated as system-defined methods (functions) that can be applied to objects. Users may also define their own methods. The object model has the usual class hierarchy. In addition, attributes of a class may be tagged as inverses of other attributes or as referring to parts of an aggregate object. A complex object defined by means of attributes with `has_parts` tags may be processed as a whole, being retrievable, for example, by a single `complexselect` statement. Database integrity is supported by offering user-defined triggers, and authorization controls are applied at the granularity of the object

instance. MANDRILL version 2 is a completely different successor system not presented to us in detail. It is a pure object-oriented database system with a database programming language called MANDRILL/POP (persistent object-oriented programming language).

Jasmine. The JASMINE system [Aosh:90] is being developed by Fujitsu Laboratories Ltd., a Fujitsu subsidiary in Kawasaki. It is an object database system that supports set-oriented queries containing path expressions. Constraints and demons (triggers) may be defined, and the system may be extended by new nodes, attributes, and links defined procedurally in JASMINE/C. Both textual and image data are supported in the window interface, with hypermedia applications in mind.

The data model has classes with single inheritance, and attributes which may either be `property` (stored) or `procedural` (computed by methods written in JASMINE/C). Classes are either `immediate` (value) or `referential` (objects with identifiers). Inverse functions are supported for properties linking referential objects. Each member of a persistent object may also be annotated, like a slot in a frame-based AI language.

The prototype system described to us has been built on XDE, an extended relational storage manager supporting nested relations but lacking concurrency and recovery in its first version. Reservations were also expressed about its performance in navigating between objects. For knowledge-base applications especially, a facility called MEMORY KB provides an in-memory equivalent for storing and manipulating transient objects linked by memory pointers.

JASMINE may thus be characterized as an object database system offering a seamless C-based language interface, where the seamlessness applies to similar ways of referencing and using persistent and transient data. However, it is not seamless C — in both cases the objects are defined by the JASMINE/C class extensions where the language is in competition with similar extensions in C++. JASMINE/C also makes database-inspired extensions emphasizing associative retrieval by adding set-oriented operations with predicates much as in relational languages, and AI-inspired extensions to support demons. The OODB JASMINE project seems to be on a par with similar research prototypes produced in other parts of the world. It is used for multimedia projects as described in Section 3.3.3 and has been supported as part of the National Interoperable Database System Project (see Sections 3.3.2 and 3.4.2).

Odin. At the NEC Systems Research Laboratories, also in Kawasaki, the ODIN object database system [Hiro:91] is under development. This has emphasized the "seamless C++" approach, and the prototype has itself been written in C++ to run on an NEC Unix workstation. Applications written in ODIN/C++ are preprocessed by referencing and updating the class definitions in the database schema to produce pure C++ with ODIN method calls.

The class hierarchy permits multiple inheritance, and each class has a set of all its associated instances. The class concept is carried through the whole design, such that a database is a set of classes, and a set object is an instance of the set class.

View classes may be defined by selecting objects from a given class, by joining corresponding objects from two or more classes, or by grouping objects from a class into subsets. The original instance variables and methods may be restricted in the view, and users can overload the update operations with their own methods.

The implementation of the system has a physical structure virtualization layer, and pays particular attention to the storage of large variable-length multimedia objects, to object and page buffering, and to transaction control.

3.1.3 Multimedia Databases

In addition to the interest in multimedia applications of general-purpose object databases described above, we encountered two other projects focusing specifically on multimedia support. One, the IMAZONE image database system at Ricoh, is a product headed in an object direction. The other was an evaluation study of an object database management system for multimedia usage carried out by NTT Data.

IMAZONE. We mention IMAZONE [Kuni:90] [Shir:91] [Iiza:91] here because its developers see it as potentially capable of making effective use of an object database system. Presently the object database concepts in its design are implemented on G-BASE (exported as RICOHBASE to avoid confusion with a different G-BASE system in the U.S.), which is a relational system extended with structural links. It is based on the work carried out by the laboratory director who was at the University of Texas. It effectively supports multimedia data, as described in Section 3.3.3.

NTT Study. A different perspective is provided by an evaluation [Suzu:91] carried out by NTT Data Services, who are system integrators rather than DBMS vendors. In evaluating the GEMSTONE object database product from the U.S. for multimedia support, NTT was convinced of the value of a strong type system, a class hierarchy with multiple inheritance, and user-defined methods, for handling the ever-growing variety of media data types and subtypes, as well as the complexity of hypermedia objects. However, NTT saw the need for better concurrency and transaction management of such extended media as optical disks, and for network management, whether methods run on clients or servers. Their conclusion was that object database systems still need much practical improvement for effective multimedia application.

3.1.4 University Projects

A brief review of object database research that the team encountered in universities follows. It illustrates our general finding that Japan has done considerable state of the art work in the conceptual modeling of object databases often allied to some prototyping or broad consideration of implementation issues.

Kobe and Kyoto Sangyo. The widest range of object database modeling and prototyping activities appeared to be at Kobe University and Kyoto Sangyo University. We will consider them together because of substantial amount of joint work conducted by the two universities. At the conceptual level, papers have been published during the past three years on schema design [Tana:89a], object views [Tana:88b], complex objects [Tana:88c], natural joins [Tana:90b], and versions ("alternative objects") [Chan:89].

Extensions to relational systems have included a BLOB approach to storing and retrieving POSTSCRIPT text and graphics objects in the database [Tana:88a], and an SQL-Navigator which dynamically attaches methods to objects to support a hypertext style of navigation [Sana:90]. Recent work has been on TEXTLINK which supports hypertext on an object database system.

This is an illustration of the trend we observed, to move from a good general understanding of the object database field to specific applications in the hypermedia world.

Kyushu. New work is being started at Kyushu University in the design and implementation of object-oriented persistent programming languages for multimedia databases. A planned series of languages take their names from the various stages in the life-cycle of the yellowtail tuna, for which in Japanese there are apparently as many words as eskimos

have for snow. The distinctive feature of this project [Maki:91] is that it has begun by examining performance aspects of object heap management using alternative virtual memory approaches on the MACH operating system.

Library and Information Science. The Library and Information Science University in the Tsukuba science city specializes in information management. Preliminary work over the past six years has led to a project to develop a system named OMEGA (Object-oriented Multimedia Database Environment for General Application) [Masu:90a]. Current emphasis is on the part-of relationship, and on referencing and synchronization issues that arise in dealing with complex objects.

Hokkaido. Although current activity at Hokkaido University is largely in the multimedia area, this was preceded by more general object database investigations.

In particular, their object data model ODM [Moza:89] brings together object ideas from SMALLTALK-80 and set-oriented ideas from the relational database model. The concept of a relation as a set of tuples is generalized to that of the "u-set" (uniform set) of objects. The uniformity of a u-set resides in the fact that all of its elements must be conformable to the base class of the u-set, i.e., must satisfy its interface. The u-set is itself defined as an object class with appropriate methods, both for the usual set operations and for the relational operations: project, Cartesian product, and join, on the instance variables of objects regarded as tuples.

3.1.5 Obase Consortium

A novel development is the formation of the Obase Consortium, which will operate from September 1990 through August 1993. Membership is as follows:

- Senri International Information Institute (Osaka)
- Universities
  - Himeji Dokkyo
  - Kobe
  - Kyoto Sangyo
  - Osaka
- Industries
  - Fujitsu
  - Fuji Xerox Information Systems
  - Hitachi
  - Kobe Computer Services
  - Kobelco Systems
  - Matsushita Electric
  - Toyo Information Systems

The objectives are to investigate current ODBMSs and applications, to develop a new Object-Database Management System (OBASE) together with ODB design tools, and to carry out related research. The emphasis in the current, preliminary applications study appears to be on hypertext, video, and all forms of document management.

Specifics about the Obase Consortium were provided by Prof. S. Nishio (Osaka University). Each industrial partner funds Obase at a level of about $150,000 over three years. The project began by investigating the advantages and disadvantages of conventional OODBMSs, and then moved on to the development of some specific applications. Currently the Obase researchers are using the VERSANT OODBMS to implement these applications, which include a fuzzy retrieval system for a movie database and a LATEX document OODB.

3.1.6 Summary

The main conclusions regarding object databases follow:

- There is not yet any commercially available Japanese object database management system.
- There is substantial current, well-informed object database research, especially in theory and concepts, as well as in relatively straightforward prototyping. Deeper investigation of efficient implementation is less in evidence, except in the treatment of binary large objects.
- It is unlikely that a Japanese general-purpose object database management system is imminent as a product, not only for technical reasons, but also because the strategic questions related to how to design such a system to become an industry standard rather than one among many proprietary systems have not yet been decided. Open object database standards, now in an early state of discussion, may change this expectation.
- There is a strong emphasis on the applicability of object database technology to the management of multi/hypermedia information, which is likely to lead to the early appearance of more specialized object database management systems for integrated media support.

This last point makes it likely that object databases, supporting multimedia on Japanese image and voice hardware, will be one of the early export items for the Japanese software industry.

3.2 Knowledge Bases

Author: Gio Wiederhold

We define knowledge bases as the combination of processable knowledge, stored as rules, logic, or frames, which can be applied to data, either in conventional databases, in in-memory databases, or as ground rules. It is a technology with great potential, since the application of knowledge to factual information permits effective use of even limited computer represented knowledge. As the factual data are updated to reflect changing situations, the knowledge can be reused.

3.2.1 Past Research

A major focus of the initial plans for the Fifth Generation Project was the development of knowledge-base technology. For the knowledge base a PROLOG-based schema was foreseen. The factual data would be processed, when needed, by powerful, parallel database machines. While the Fifth Generation Project initiated much research activity, we found no actual systems or prototypes that followed the original architecture. Its progress is well documented in ICOT publications. While knowledge-base technology is now well-understood, the lack of a database infrastructure, using common standards, limits its applicability.

3.2.2 Current Research

However, research on knowledge bases is progressing in academia, often based on in-memory storage models. This approach is similar to current directions in the U.S. While previously memory-based systems were not considered adequately scalable, it is obvious that the rapid growth of workstations (now they typically have more memory than the upper limit of the last generation of mainframe computers) makes this direction valid. In-memory databases do require continuous backup to assure persistence. Where knowledge changes slowly and the focus is on queries, simple backup techniques suffice in practice.

While Japanese researchers are well aware of knowledge representation technologies, we did not find any large-scale efforts to establish large collections of knowledge as, for instance, the MCC CYC project [Lenat]. This may be an omission on our part, since we focused only on knowledge bases that involve database storage. Databases are likely to be employed in such efforts, but are not an essential feature.

Neural Nets. Several recent research efforts consider neural nets. Here knowledge is not managed explicitly, but is captured automatically from example sets, for which answers, typically classification parameters, are given. The results are stored as discrimination values in internal nodes, and can then be used to classify new examples.

Neural net knowledge can then be exploitable to classify images, ranging from Kanji calligraphy to faces [Kuni:90]. This research area is relatively new and immature. Japanese researchers in this area are well aware of the state of the art, and appear to be experimenting widely with the current technology. If this technology can be usefully applied in future database systems, then the background knowledge available in Japan is adequate to implement these techniques. These methods could then complement multimedia technology.

3.3 Multimedia Database Systems

Authors: John Miles Smith and Diane Smith

Computer systems are now able to handle multimedia information such as audio, video, images and text. A huge commercial market is expected to develop in multimedia systems in the next decade. Success in this market requires the integration of consumer electronics and computer systems, and is likely to be an area of intense interest both in the U.S. and in Japan. Multimedia database systems, which manage repositories of multimedia information, are a central component of multimedia applications.

After a description of the background of multimedia database systems, the impact of the three major Japanese technology drivers (national projects, commercial products, and academic research) are reviewed. Subsequently, each key technical area is considered, and the results of Japanese projects visited by the team are described for each area. General comparisons are made with the state of the technology in the U.S. The final section draws some conclusions about Japanese research in multimedia databases.

It should be noted that the opinions and recommendations in this section are based on a sample of organizations visited by the study team. The study focused on organizations that are active in database research and development. However, due to the interdisciplinary nature of the multimedia area and due to emerging commercial competition, it is possible that some organizations doing important work were not visited, and that for some organizations visited, the study team was not informed about some proprietary work.

3.3.1 Background

Until recently, computer systems focused on the manipulation of alphanumeric data. However, much of the information processed by human beings is in the form of audio, video, images and text. Higher computer speeds and larger storage capacities are required to process these information media than are required to process alphanumeric data. Computer systems have now passed the threshold where these information media can be processed cost-effectively. As computer systems and communication networks continue to improve, higher volumes of media information can be supported and increasingly sophisticated processing operations can be performed at high speed.

The result is that new classes of computer applications can be developed for the home, education, entertainment and technical industries. For example, computers can be used to access encyclopedias of information containing text, images, video and audio recordings. Family photograph albums can be captured on computer storage and retrieved and displayed on the computer screen. Movies can be downloaded from commercial databases and played back through the computer. Videoconferencing can be held via computer interface, and sections of the conference (video, audio, and presentation material) can be stored for later review and editing. Geographic information, previously available in the form of paper maps, can be retrieved from computer storage and analyzed in powerful and flexible ways.

Multimedia computer systems can handle several information media. Typically, they support I/O devices such as microphones, scanners, videocameras, FAXmodems, and color monitors. For all of these devices, the multimedia computer systems can capture, process, and playback information. With conventional storage systems such as books and videotapes, the information is connected in a single linear sequence. However, when multimedia information is captured in computer storage, it can be linked together and retrieved in any and multiple meaningful sequences. This allows users to search for information in whatever way is most effective for the purpose at hand. Computer systems which allow multiple links for storing

and searching multimedia information are often referred to as *hypermedia* systems.

Multimedia systems are expected to be a huge growth market for the computer industry over the next decade. In addition, they will have a profound impact on existing markets for consumer electronics. Televisions, telephones, audio systems, VCRs, videocameras, still cameras, copiers, and FAX machines may all become extensions of integrated computer systems, and not stand-alone products. Japan's electronics industry is well aware of the opportunities provided by multimedia computer systems. Both the United States and Japan are positioned to acquire a share in this market. The United States is ahead but is losing ground in the computer industry, while Japan is well ahead in consumer electronics. In the 1990s the battle will be joined to successfully integrate these two technologies.

Multimedia database systems are a key component of multimedia computer systems. They will provide the services to store and retrieve multimedia and hypermedia information, and support the sharing of this information across user groups. There is a substantial technology gap that must be bridged to provide these services effectively for multimedia data. For example, an image requires approximately three orders of magnitude more space for storage than an alphanumeric string. At the same time, matching image segments against a retrieval pattern is a much more time-consuming operation for an image than a string. Finally, the index structures required for speeding access to information are quite different for certain kinds of multimedia data (e.g., maps). Solving these technical problems is critical to realizing the potential of multimedia computer systems.

As a result, the status and direction of Japanese multimedia database projects is an important factor in formulating U.S. strategy in multimedia computer systems.

3.3.2 Technology Drivers

From the study, three drivers emerged that are generating new technology development in the multimedia database area:

1 A Japanese national project called the "Interoperable Database System" (IN-DBS) project [Tojo:91, Tojo:85];

2 The commercialization of existing multimedia system technology in niche application areas; and

3 Some innovative academic research projects involving multimedia data.

The effects of these three drivers are different but synergistic. The first has drawn attention to the opportunities in multimedia databases, triggered investigation into various development problems, and increased the level of expertise in Japanese industry. The second has created practical extensions of existing database technology, and greater understanding of the performance and capacity problems to be solved. The third is creating more fundamental advances in multimedia database technology with the potential for use in future products.

The National Research and Development Project on Interoperable Database Systems started in 1986 and has just come to the end of its six-year program. The project was supported by the Agency of Industrial Science and Technology of MITI and executed by the Electrotechnical Laboratory (ETL) and sixteen participating companies. The project was started with the goal of developing technology in four areas:

1 The integration of distributed heterogeneous database systems;

2 The reliability of database systems;

3 The architecture of computer networks to support heterogeneous database systems;

4 The support of multimedia information in database systems.

The specific subgoals for the multimedia area include:

1 Network protocols for handling multimedia data;
2 Conversion algorithms for mapping between different information formats;
3 Data compression and feature extraction; and
4 Transformation of bit-mapped documents into structured objects.

Based on information obtained during the study, it appears that the multimedia work concentrated on the last two subgoals. A common project theme was to store maps (e.g., municipal, plant, and telephone service network) in a database, to build feature-recognition and feature-matching algorithms, and to provide real-time scrolling and zooming via a graphical user interface. This theme was apparent in projects discussed at ETL, NTT, Mitsubishi, and Toshiba, and [Tojo:91] indicates that Hitachi also has a similar project. All these projects appear to have been influenced directly, or indirectly, by the Interoperable Database System program.

From the point of view of commercial products, the most important multimedia data types are text, image and drawing. Virtually all office documents can be captured and processed using these three data types. Several Japanese companies offer document management products with a database system component. For example, Ricoh has the IMAZONE image database construction tool that runs on top of its RICOHBASE database system. Hitachi's "Document Filing System" uses a special-purpose database to support semantic matching. Toshiba has a product set called the "Total Office Productivity Support System" that includes a document management tool OA-FILING that operates in conjunction with its RDB/V database system.

From a database system perspective, Ricoh's RICOHBASE is the most advanced product. RICOHBASE is an extended-relational database that has been specially designed to manage images, and IMAZONE provides tools to help C programmers build image database applications.

Ricoh's manufacturing capability in optical and electronic products provides an important motivational focus to its research laboratory. We present their technological approach in Sect. 3.3.3.

Of the academic projects involved in the study, two stood out as making significant contributions in the multimedia database area:

- Dr. Tanaka's group at the Un. of Hokkaido [Tana:89, Tana:91],
- Dr. Fujiwara and Dr. Kitagawa's group at the Un. of Tsukuba [Jian:89,Luan:89].

The Hokkaido group is driven to create new multimedia technology. For example, storage of video information, content-based search for video segments, efficient representations and content-based searching for bit-mapped documents, and object-oriented graphical generators for multimedia applications. In contrast, the Tsukuba group is motivated by two information systems considerations. The first is to develop database systems for chemical/material substance information (the databases are actually published on CD-ROM). The second is to develop new database system architectures that can more effectively manage multimedia data. These efforts are investigating innovative approaches to fundamental problems in multimedia database systems.

To nurture multimedia activities in Japan and to foster cooperation with overseas activities, MITI is supporting the founding of an "International Multimedia Association." The goals are technology interchange, expanding the base of expertise, and popularizing the benefits of the technology. At this stage, national, multinational and foreign companies are

being sought as founding members of the association. According to the brochure describing admission to the association, U.S. companies IBM, Apple, Intel and Microsoft are expected to participate. It would appear that MITI is looking to the International Multimedia Association to help guide the next steps in Japan's commercialization of multimedia technology.

3.3.3 Technology Survey

To fully achieve the promise of multimedia database systems, technology must be developed and deployed in five technical areas:

1 Base technologies;
2 Database system architecture;
3 Content-based retrieval;
4 Hypermedia storage and retrieval; and
5 Application generation tools.

These five areas will be covered in this section. For each area, the technology will be briefly described and motivated. Then Japanese projects in the area will be summarized.

Base Technologies. The term *base technologies* refers to underlying capabilities in information capture, playback, storage, and transmission that are prerequisite to the commercial exploitation of multimedia systems. These nondatabase base technologies were not explicitly covered by this study and will be treated only briefly here.

Japan's technological capability in the audio, video, and imaging areas of consumer electronics is well-known. While basic research is probably on a par with that in the US, Japan is leading the commercialization of these technologies with high-definition TV, electronic still cameras, FAX and photocopiers, and digital audio systems. Due to the large information content of multimedia data, very high capacity storage devices are required. For example, a reasonable color image requires 1 megabyte (MB) or more of storage. Only forty such images can be stored on the 40 MB hard disk drive found in a typical PC. As a result, optical disk technology must be used, which provides about 5000 MB per optical platter with current products. Similarly, the transmission of multimedia data requires very high-speed communication networks. Japan has been commercially operating a limited ISDN (Integrated Services Digital Network) since the middle 1980s; this will soon expand across the country. Japan is a leader in both the base technologies and their commercial exploitation.

Database System Architecture. Database system architectures must be extended to handle multimedia data. One approach is to extend relational database systems (RDB) so that multimedia data can be stored and retrieved along with regular alphanumeric data. In general, extended relational database systems (ERDB) work well for multimedia data items that are always accessed as a whole (e.g., as a single image or as a single video segment). However, they do not work well for data items with a complex internal structure that must be accessed on a component-by-component basis (e.g., a structured document or a CAD model). For this purpose, object-oriented database systems (OODB) have been developed, and are becoming commercially available in the U.S. A disadvantage of commercial OODBs is that they do not provide the high performance we expect when querying data items with a uniform structure (e.g., business transaction records). It is an active research issue to combine the advantages of ERDB and OODB into a single database system.

Most of the commercial products and National IN-DBS Project efforts use a combination of an RDB and a file system to manage multimedia data. Multimedia data items are represented as files. Pointers to these files are stored in fields in the relational tables to connect

relational data with associated multimedia data. The file manager may use an optical disk jukebox for data storage. The advantage of this approach is that it requires no modification to the relational database. The disadvantage is that applications become responsible for ensuring that the relational and file data are mutually consistent during normal operation, and for ensuring the recoverability of mutually consistent states in the event of a failure.

Ricoh's commercial product IMAZONE [Rico:90] is an image database construction tool being built by Ricoh at their Software Division in Bunkyo-ku, Tokyo.

The emphasis is on support of complex graphical objects, variety of media, and high volume. IMAZONE uses an ERDB that can store multimedia data items directly in fields of the relation. However, the size of each such field was limited to 64 Kbytes at the time of the JTEC panel's visit. 2 Gbytes were supported from the end of 1991. In that case larger data items must be stored in a separate file system and then connected back to the ERDB by an ID number. Since IMAZONE supports linkage of records, that connection can be relatively transparent to the user. The file system can support an optical jukebox. Some vendors in the U.S. now allow very large field sizes in their ERDBs, and provide for mapping of these fields directly onto an optical jukebox.

From the discussion with the RICOHBASE group, it is clear that this group has faced and solved many of the practical performance and capacity problems associated with multimedia databases in the context of office documents.

Two research projects visited by the study team are developing ERDBs for multimedia data. At Hitachi Research, an experimental multimedia database system called MANDRILL [Yama:89] is being developed based on object-oriented extensions to SQL. At IBM Japan, the first version of an experimental database system called MODES [Kosa:87] has been implemented, and work is underway on a second version. The first version of MODES extends an RDB in three ways:

1 By introducing a field type that is a pointer to an external file,
2 By allowing the selection of storage devices (in particular, optical devices) for these files, and
3 By defining virtual input/output systems which are able to handle these file types.

The second version is developing a *card* user interface paradigm for the storage, retrieval and display of multimedia data.

Two other research projects are pursuing OODBs for multimedia data management. At ETL, an open database system called AXIS [Koji:88, Koji:91] is under investigation as part of the National IN-DBS Project. AXIS seeks to provide extensibility so that a system kernel can be extended to meet the needs of particular applications. The specifications of extensions to the system are stored in a system database that is accessible to application developers.

Multimedia systems being produced at ETL and private research laboratories may not be considered real innovations. Such products are already being used in some simple applications. Also, these projects have certainly awakened interest in multimedia technology.

At the University of Library and Information Science, OMEGA [Masu:91] is under development. The current effort is to prototype processors for the data definition language and the query language. The main emphasis is on integrating a number of semantic constructs and operations into an object oriented data model.

Content-Based Retrieval. No matter what database system is being used for multimedia data, a key challenge is to develop algorithms and data structures for data items so that they can be retrieved based on the semantic content of that data item. For example,

to retrieve all images that contain a human face, or to retrieve all documents that contain a certain phrase. Content-based retrieval requires efficient algorithms to match a given *pattern* against a similar *pattern* in the multimedia data item. Content-based retrieval allows retrieval conditions to be determined at the time of retrieval. In contrast, attribute-based retrieval requires that the attributes be known at the time the database is created, and that their values be given at the time an item is stored.

Several companies (ETL, NTT, Mitsubishi and Toshiba) are investigating feature extraction from a variety of map information including municipal maps, manufacturing plant layout, and utility networks. Mitsubishi, for example, is interested in finding areas in a plant layout where the topology is similar to another area in the layout. Maeda and his colleagues [Maed:88] have been working on algorithms for a simplified version of this problem involving the layout of rooms in a house. They abstract the room topology in a house to adjacency relationships. They then search a "realty" database to find houses with a layout similar to that of a given house by matching on their adjacency relationships. In another example, Toshiba has developed a geographic information system using optical disks. The database consists of image data for base maps with vector and alphanumeric data overlays. In [Okaz:90], techniques are described for obtaining a seamless base map from a conventional atlas and for distributing image data to multiple optical disks for concurrent retrieval.

A group at ETL has been pursuing content-based retrieval using a more abstract collection of images, namely trademark symbols. In [Kato:91], a *graphical feature vector* has been developed that abstracts a trademark symbol into five distinct analytic features. Trademarks are compared for similarity by differencing their vectors. The group has demonstrated that subjective similarity of symbols correlates well with small vector differences.

At Hokkaido University, Tanaka and his research group have been working on content-based retrieval of video information. Their initial work focused on the detection of *cuts* in a video segment. A cut is where one scene ends and another begins. They have developed algorithms to reliably detect cuts. Recently, they have extended their work to search a videotape for specific content. Prof. Tanaka showed a video of a car race, in which the same cars showed up from a variety of angles and in various contexts at different stages of the race. In a convincing demonstration, Prof. Tanaka was able to retrieve all of the video frames which included a specific selected car. Work of similar quality in the U.S. is largely sponsored by the intelligence communities. Although reported in the open literature, this work seems not to have had much commercial impact, perhaps because of the high cost of the hardware.

3.3.4 Character Recognition

In a project called the "Transmedia Machine," Tanaka [Tana:89] has developed techniques for implementing text processing on the content of scanned images of text documents. The usual approach today is to employ optical character recognition (OCR) to regenerate the original text from the image. However, OCR is often difficult, particularly when multiple fonts or different character sets are used. For many text editing operations, it is not necessary to recognize individual characters. It is sufficient to recognize just the line and character boundaries, and treat characters and lines as rectangular sub-images. Characters can then be made to flow as line length and spacing are changed, and as characters are inserted and deleted. Hitachi is developing a Kanji full-text search machine [Fuji:91].

Text-processing in Japanese is complicated by the absence of spaces to indicate word boundaries. The same Kanji symbol can initiate a one-, two-, or three-symbol word. The

hardware is further described in Section 3.5.1. Emphasis is on development of a very high-speed machine that can quickly scan very large text databases.

Hypermedia. As described before, hypermedia refers to the linking of multimedia data objects together so that information can be retrieved by following the links. For example, the service manual for repairing a car could be stored as a collection of information (text, diagrams, video) linked together based on symptoms and diagnostic results. The same piece of information may appear along several different links. Hypermedia systems must provide a graphical interface that allows the user to navigate (or browse) along the links to find the required data. In some cases the required data may be a composite of several linked items stored in the database. This may require automatic layout algorithms to compose the data items into an efficient form for human examination. For example, if a car repair step includes some text instructions with two diagrams, then all three items must be arranged on the user's monitor in such a way that important information is not obscured.

Ishikawa [Ishi:90] has used Fujitsu's experimental object-oriented database system JASMINE (see Section 3.1.2) to build a prototype hypermedia system that serves as a visitor's guide to Kyoto. JASMINE's object hierarchy is extended with special classes to contain multimedia information. Links between information are represented by object attributes. Users can directly navigate between objects, or use a query language for retrieving objects based on their classes and attributes. Key features of JASMINE are exploited for hypermedia information management. For example, the version control subsystem is used to keep track of different versions and configurations of visitor guidebooks, and triggers are used to notify users of changes to stored information.

Hitachi is starting a new project called "Personal Information Base." The general idea is to overlay a multimedia database with a semantic network of concepts. The user interface takes advantage of human memory characteristics to retrieve information by navigating the conceptual structures. A prototype browser and editor has been developed.

Dr. Kambayashi at Kyoto University has investigated an automatic layout problem associated with the presentation of map data. With a geographical information system, a user can select information from a database and request that it be displayed in the form of a map. Since the content and form of the map can be defined dynamically, the positioning of information on the map is determined at the time of map generation. A classical example is the positioning of names close to the map objects they describe without obscuring other critical information. Kambayashi describes practical methods for solving this latter problem in real-time.

Based on the rather fragmentary data gathered during the visit, the U.S. has research and commercial leadership in the hypermedia area. With the exception of Kambayashi's project, Japanese research of hypermedia is superficial. A number of U.S. projects have explored this area in much more depth.

However, work on hypertext and image-processing, as seen at NEC, provides an important infrastructure upon which hypermedia systems development can proceed rapidly.

Application Generation Tools. A multimedia (or hypermedia) database system only provides services for information storage and retrieval. To be useful for a particular purpose, an application program must be written that calls on those services on behalf of the user. For example, an application will typically allow a user to enter data from some input device (e.g., a keyboard or scanner), check the consistency of that data, retrieve data using certain predefined queries and parameters, organize and aggregate data into certain output

formats, and transmit or print the result. In many situations it is desirable for a user who is not a professional programmer to create or modify his or her own application. For this purpose, application generation tools are used. These tools normally provide a *skeleton* for a certain class of application. The user then *fleshes out* this skeleton to include the specific details required for the application on hand. Application generation tools are commercially available for relational database systems, and others have been developed for certain specialized multimedia applications, such as document processing. Application generation tools for general-purpose multimedia database systems are still in the research and development stages.

The multimedia application generation tool seen during the visit was the INTELLIGENTPAD system created by Tanaka [Tana:91]. INTELLIGENTPAD falls into the same software category as Apple's HYPERCARD and Xerox's NOTECARDS. They all provide an object-oriented programming system with a strong visual representation for objects. However, INTELLIGENTPAD goes beyond the other tools by allowing object construction via direct manipulation of the visual representation. Each object is represented as a *pad*, that is, a stack of sheets. Database properties are handled through the persistence of the underlying pad data. New pads are created by pasting one pad on top of another. This provides an elegant and surprisingly powerful programming paradigm. A prototype implementation of INTELLIGENTPAD has been built by Tanaka's research group at Hokkaido University.

While still in the experimental stage, INTELLIGENTPAD is a creative and promising research result, and stands on a par with other U.S. work in this area. INTELLIGENTPAD is an exception in that it presents novel advanced programming and system abstractions. The overall U.S. lead in software owes much to research into such novel paradigms.

3.3.5 Conclusions

The style of Japanese research in the multimedia database area contrasts strongly with that of U.S. research. Japanese research tends to be pragmatic. It is focused on creating a concrete target application that stresses the capabilities of current systems. Research results tend to be incremental improvements to current system software, together with insights into the target application. U.S. research tends to be more profound. It is focused on creating new system concepts that can be applied to many target applications. Research results include more fundamental advances in system software technology.

At the database system level, Japanese researchers rely on the U.S. to provide the major software technologies. There are academic projects aimed at architectural advances, but in reality they are largely just variations on U.S. research themes. At the content-based retrieval level, there is good stand-alone research in progress, but the results are generally not integrated back into an overall database system architecture. There is a noticeable lack of research in Japan in multimedia database tools and application generators. The Japanese researchers are more concerned with building applications directly than with understanding the general methodologies involved.

From the examples mentioned above, it is clear that Japanese groups are covering a wide spectrum of media types in their research on content-based retrieval. Their techniques are being investigated in isolated but important applications. It will be some years before these techniques are fully integrated into the architecture of a database system that can handle the full spectrum of media types. The U.S. is ahead in most aspects of research into content-based retrieval, as well as in some aspects of commercialization – particularly in the areas of geographical information systems and CAD/CAM database systems. As might be

expected, Kanji text retrieval is an area where Japan is ahead. The most innovative work is underway at the universities (e.g., Tanaka's group). However, this work is under-funded and it will be difficult for researchers to keep pace with relatively well-funded groups such as MIT's "Media Lab."

The principal Japanese advantage in the multimedia database area is at the hardware component level — their capability to design and manufacture high quality displays, storage devices, and audio/video systems. This capability is supported by a strong vertical integration going all the way up to application products. The lack of standards in new technology is overcome by joint product planning teams. This setting, combined with the Japanese focus on concrete applications, positions them to commercialize packaged hardware solutions based on available system software technology in the multimedia area. At some point in the future, software technology may become the roadblock to continued progress. At that point Japan is likely to invest more research effort in creating the fundamental software technology needed. In the interim, Japan can rely on the rest of the world.

Japanese industrial research takes a conservative approach in the multimedia database area. On the whole, it is following the direction of the "Interoperable Database System" National Project. The more widely publicized "Fifth Generation Project" has had little or no influence on Japanese industry in this area. Innovative research is underway at the universities (e.g., Hokkaido and Tsukuba). However, this research has significantly less funding than key multimedia research centers in the U.S. (e.g., MIT's Media Lab), and the researchers will find it difficult to keep pace except in specialized niche areas.

The Japanese choice of target application areas throws some light on where they see early commercial opportunities for multimedia databases. The applications seen by the study team include:

Industrial Plan Maps (Mitsubishi)
Utility Service Maps (NTT)
Urban Maps (Toshiba)
Museum Holdings Database (National Museum of Ethnology)
Fashion Industry Database (NTT Data Systems)
Japanese Document Reader (Toshiba)
Chemical Substances (University of Tsukuba)

In some areas, such as chemical substances data management and Kanji document processing, the Japanese work is on the leading edge. In other areas such as mapping applications, Western companies have leadership capabilities and the Japanese are in catch-up mode. Since the field is not yet mature, the eventual results of the research, development, and marketing investment are unclear. Even less clear is the eventual allocation of market-share, since the market is in its infancy.

3.4 Interoperable Databases

Author: Toshi Minoura

We expect that most database activity of the future will involve multiple computers. When computing activities are distributed, issues of joint operation must be addressed.

3.4.1 Current State

Most of the current distributed applications in Japan, such as banking systems, are processed by centralized computer systems. Only a few distributed database systems are deployed. However, the Japanese understand the importance of interoperation among computers of different sizes and from different vendors in the "information-based society" they envision.

3.4.2 Interoperability and Standards

The most notable move toward the vision of an information-based society is the National Research and Development Program on Interoperable Database Systems (Interoperability Program), managed through INTAP (Interoperability Technology Association for Information Processing). INTAP is an industry consortium created to perform research and development focused on open network systems architecture.

The Interoperability Program was started by MITI in 1986 with a total projected budget of $100 million. It is expected to end in March 1992. The official goal of the program was "to establish fundamental technology indispensable for realizing multimedia reliable distributed database systems on interoperable computer network systems."

The projects supported by the program were divided into four major areas:

1 Database systems architecture (discussed earlier under the acronym IN-DBS),
2 Multimedia technology,
3 Distributed systems technology, and
4 Open network systems architecture.

Open systems permit multiple participants to use and contribute software. Without that concept there is much dependence by a user on a single vendor. As such, the concept changes drastically the entry cost for new, innovative companies in the software and hardware arena.

The major participants in this MITI program are ETL, Fujitsu, Hitachi, Matsushita, Mitsubishi, NEC, Oki, Sharp, Sumitomo and Toshiba. Furthermore, NTT, IBM Japan, and Nihon Unisys participated in the interoperability experiments among heterogeneous computer systems. Fuji Xerox and DEC Japan are also members of INTAP.

Tojo summarized the results of this project as follows [Tojo91]: "*Although it is true that there may still exist a number of problems not resolved completely, it is also true that most of the key issues of interoperability are being attacked in the project. It is expected that the result of the project will broadly accelerate the development and use of real open interoperable information systems* ".

In other words, the project has not yet produced a coherent set of technologies as aimed in the official goal. Real-time interoperation is not yet a practice. However, there are some important results.

Implementation and Documentation. Based on the ISO OSI reference model [ISO:84], the INTAP participants have produced detailed implementation standards for the protocols from layer 1 to layer 7. There are 3,830 pages of these standards. The networks supported include ISDN, FDDI, and CSMA/CD LAN. The application level (layer 7) protocols include FTAM (File Transfer, Access and Management), MOTIS (Message Ori-

ented Text Interchange Systems), ODA/ODIF (Open Document Architecture/Interchange Format), distributed transaction processing, and RDA (Remote Database Access).

The implementation protocols are tested at Interoperable Networking Events (INEs) '88, '90, and '91. The successful experiments of FTAM involved BULL (France), DEC (Great Britain), ICL (Great Britain), and Olivetti (Italy) in addition to the member companies of INTAP. They have established the INTAP Conformance Test Center (ICTC) to certify OSI products implemented by various vendors. About twenty OSI products from various vendors have been certified so far. To provide the environment for interoperability tests of OSI products, INTAPnet was established in 1989.

The Japanese are actively participating in international standard setting in this field. Their primary stage is Asia-Oceania Workshop (AOW), with participation by China, Australia, Korea, and Japan. They intend to cooperate fully with EWOS and NIST OIW. A key player of the Japanese standardization activity, Fujitsu, is participating also in the RDA standardization activity on the U.S. side, the SQL Access Group.

This program can be regarded as another example of Japanese government projects whose actual goals are to disseminate information (education) and to provide some seed money for research activities at private companies.

Distributed Databases. The Japanese have not yet built homogeneous distributed database systems that perform distributed query processing. (This omission may be due to the observation that, from a commercial point of view, such systems are generally not viable.) We encountered no efforts to build federated systems. A project on federated heterogeneous database systems started at ETL (Interim Report, 1989) seems to have fizzled out.

Hitachi has installed its relational systems at about 170 sites. Among them are about thirty distributed systems with an average of two to three nodes. Mostly ISO standards are followed, and OSI RDA support is being developed. However, some customers want compatibility with DB2, the principal IBM product for relational databases.

Fujitsu was the first company in Japan that developed a relational database system. Fujitsu seems to move directly to heterogeneous systems based on the OSI RDA standard.

Currently, Oki does not sell distributed database systems. Its product group is extending its relational system REAM, incorporating the RDA standard for remote data access. The system may be ported to non-Oki platforms. ORACLE has been ported to Oki minicomputers.

3.4.3 Conclusion

Distribution of databases, while less applied in Japan than in the U.S., is being built on a solid foundation of standards, especially the RDA standard favored in Europe, rather than the direct use of SQL, as used for current implementations. Since RDA is based on SQL, the transition will not be difficult, but any incompatibility can hurt. For autonomous federated systems, an important compound of broad-based information systems, the RDA standard is especially relevant. Unless RDA is provided and supported in U.S. systems, as promoted by the RDA-SQL-ACCESS consortium, this style of distribution will assure early system compatibility of Japanese software with the European market. Such a direction could cause isolation of more narrowly-SQL-based U.S. vendors.

3.5 Database Hardware

Author: Nick Farmer

In Japan, as in other parts of the world, most databases are accessed using conventional computers. While there are a large number of small databases on personal computers, most significant databases have, in the past, been loaded on a large central mainframe, using a variety of different database management systems.

There appears to be considerable dissatisfaction with the current environment. In a survey by Hayashi in 1985, users noted significant dissatisfaction with the performance of all types of database management systems running on conventional computers.

	Problems	No.	Rank
1.	Poor flexibility with changes to the data structure	65	1
2.	Inadequate efficiency for processing	63	2
3.	Poor capability to manage a large quantity of data	36	3
4.	Limitation of program development	29	4
5.	Poor fittable functions to each application	27	5
6.	Too expensive to install the DBMS	27	5
7.	Too expensive to operate the DBMS	19	7
8.	Others	12	8
9.	Poor expandability of the system	6	9
10.	Bugs in the DBMS	4	10

Some of these reasons motivate research and development activities for special purpose hardware devices, with the goal of improving the performance of database management systems. Use of hardware support also provides a more stable interface than more flexible software and operating systems conventions, especially where standards are absent.

3.5.1 Research and Development

These research and development activities are being carried out by a number of Japanese universities and by most Japanese computer manufacturers. Of special note are the activities at the following institutions and corporations:

Organizations Involved in Research on Database Hardware Technology	
Universities	Computer Manufacturers
University of Hiroshima	Fujitsu
Hokkaido University	Hitachi
Tohoku University	Matsushita
Tsukuba University	Mitsubishi
	NTT
	Ricoh
	Toshiba

Most of the research and development activities discussed in this chapter were carried out between 1984 and 1990. Since the most active area of database interest during this time

was associated with the relational model, this is also where most of the hardware-related activities were focused. The simplicity of the relational model favors hardware oriented approaches.

In addition to the relational model, there has been some interest in full-text searching of both Japanese and English language material. No activities were mentioned relative to special purpose hardware for the object database model probably because this model is not yet sufficiently stable enough to support special-purpose hardware implementations.

Areas of potential interest relative to special-purpose hardware include:

- Special purpose VLSI components
- Vector processors
- Intelligent disks
- Parallel processing
- Pipelined operations

While a specific project may be focused on one of these areas, projects often span multiple areas. For instance, intelligent disks may also use parallel logic, and vector processing may take advantage of pipelined operations. Nevertheless, the discussion that follows will be structured along the lines of the five principal areas listed above.

3.5.2 Special Purpose VLSI Components

Most of the interest in special purpose VLSI components for database access has focused on three areas:

- Key search in the relational model
- Character match for full-text searching
- Sorting

Professor Tanaka of Hokkaido University has done extensive work on special-purpose VLSI devices for database machines. He has proposed a special-purpose sort module that can handle sorting of variable length character strings in a pipelined fashion, that can begin to output the sorted list immediately after the last input character has been received.

Fujisawa and his colleagues at Hitachi have worked on a system to search Japanese text files [Fuji:91]. They use a two-stage search process, with a bit-map surrogate search followed by a condensed text search. They can scan text at approximately 100 million bits per second if the surrogate hit ratio is less than or equal to 15 percent. Their goal is to be able to process 25,000 documents per hardware unit, for document sizes up to 20,000 bytes per document, with retrieval times of approximately five seconds, for queries that have 1,000 or fewer terms.

3.5.3 Vector Processors

Most examples of the use of vector processors have been restricted to numerically intensive computing. Vectorizing algorithms are better understood in this area, and the applications requiring vector processors are more prevalent. However, some of the same basic techniques used for numerically intensive computing can be carried over to special-purpose character-oriented vector processors.

In areas where vector processors have been used to support database processing, they have typically been connected to conventional processors running a relational database management system. This allows the vendor of the database management system to be able to selectively use the vector processor in those areas where the vector processor has the greatest advantage. It does not require the application system to be aware of the vectorizing

algorithms. It also takes advantage of the file updating capability of the native database management system.

Torii and his colleagues at Hitachi have developed an integrated database processor (IDP) that exploits the parallelism inherent in a relational system by dynamically rearranging the pointer structures into a vectorized format. The integrated database processor is connected to a conventional computer running the Hitachi RDB1 relational database management system. The query analyzer in RDB1 determines whether or not each query is suitable for the vector processor. Performance improvements of between two and thirty have been observed when combining the specialized vector processor with a standard RDB1 configuration.

3.5.4 Intelligent Disks and Accelerators

One of the characteristics of applications based on a database management system is that it is often necessary to transfer large amounts of data from secondary storage to primary storage, and perform a significant level of computation on that data, with the objective of selecting a very small percentage of the retrieved data to pass along to the application. This tends to stress all of the major architectural components of conventional computers: input and output devices, channels, memory, and processors. In order to reduce the load on the other system components, special-purpose "intelligent" disks have been constructed that allow some of the processing to be moved much closer to the data on secondary storage.

Various implementations have placed special-purpose processor components at different levels. Some have been placed at the disk controller level, some at the disk device itself, and some at the individual track on the disk. Implementations have been oriented towards both full-text retrieval and relational database applications. Full-text retrieval applications have used special-purpose character string-matching devices. Relational database applications have focused primarily on the select and join operators.

Inoue and his colleagues at NTT Communications have developed a special relational database processor called RINDA that performs key database operations such as search and sort at very high speeds. RINDA consists of one or more content search processors (CSP) and relational operational accelerator processors (ROP) connected between the channel of the host computer and the disk controller.

RINDA is controlled by the conventional database management systems DEIMS-5, running on a conventional processor. Studies have shown that RINDA dramatically reduces retrieval time through the use of parallel retrieval from multiple disks and multiple track read mechanisms. For certain queries RINDA performance has been found to be 10 to 100 times better than a conventional database management system alone.

Kitajima and his colleagues at Hitachi have emphasized disk caching techniques to increase online transaction processing throughput. They have experienced improvements ranging from 1.6 at a 50% cache hit ratio, to 3.0 at an 80% cache hit ratio. They have also investigated a data filter within a disk controller to reduce the channel load. This device has shown a throughput improvement factor of between 2 and 100 in an early prototype.

3.5.5 Parallel Processing

One obvious way to improve the performance of database applications, especially when the database is large, is to partition the database function into several independent operations and perform the operations in parallel. Equally obvious is the fact that the success of this approach is heavily dependent on the effectiveness of the partitioning algorithm. Most past research into partitioning algorithms has been directed towards numerically intensive computing applications, and not much has been done on database-related applications.

Most research into partitioning algorithms for database applications has focused on the join operator for the relational model. Some database applications, text searching of document database for example, allow a simple partitioning by document or range of documents.

Most research in the past has been oriented towards coarse-grain parallel systems with only a few processors. However, there is some current interest in massively parallel systems. Professor Tanaka at Hokkaido University has proposed a multiport disk cache system that will support several thousand processors.

3.5.6 Pipelined Operations

In addition to the performance gains from using special purpose hardware such as VLSI components, vector processors, and intelligent disks, additional gains can be realized by segmenting the problem into separate operations and running the operations in parallel, as described in the previous section. Within each parallel system it is possible to gain even more performance by structuring the system in a way that allows multiple operations to be processed in an overlapped fashion. That is, the total processing can be treated as a "pipeline," with information flowing in one end in "raw" form, with the information processed in several distinct steps so that a final fully-processed data object can flow out one end of a pipe while raw data is still flowing in the other end. The components of a database system that have been treated in this fashion include character string matching, join operations, and sorting.

3.5.7 Jukebox

Sony Corporation and Ricoh are building jukeboxes to hold sets of their optical platters and increase total system storage capacities. The availability of consumer products that handle multiple platters enables rapid development of reliable devices in this arena. The availability of jukeboxes reduces pressure to increase the capacity of individual platters, and permits the establishment of standards that represent state-of-the-art technology.

3.5.8 Conclusion

The preceding sections have covered the specific areas where special-purpose hardware has been used to address the performance-related aspects of database management systems. In general, the areas of activity in Japan over the last five years are similar to those pursued in the U.S. and in Europe. The results of the research and development activities, while of a substantial nature, have not yielded any breakthroughs.

The area of hardware support for text processing is a promising innovation. As discussed in the section on object databases, many modern DBMSs provide large fields (Binary Large Objects or BLOBs), but lack extensions for operating on them. BLOBs are useful for storing text fields, but with that facility comes the need to process text. Either extended indexing or fast text searching must be provided before textual BLOBs can be used effectively.

There are very few commercial database products built around special purpose processing hardware, either in Japan or elsewhere. The improving price-performance ratio of conventional computer systems, especially the newer Reduced Instruction Set Computers (RISC), reduce the potential of special-purpose hardware and make it more difficult to recover the significant investment required to develop and produce special-purpose hardware systems.

Japanese industry is in a good position to take advantage of these options, if such hardware developments can be effectively integrated. Japan is a major vendor of storage hardware and large capacity optical units. Together with Japanese capability in image and sound products, there is an excellent structure in place to build vertically integrated multimedia systems.

3.6 Security

Author: Sushil Jajodia

During our visit, we investigated the information security efforts at various sites in Japan. We also received a copy of the Security Guidelines (in English) from Mr. Hideo Setoya, who is the manager of the Information Processing Systems Development Division at MITI. (Since our visit Mr. Setoya moved to another bureau. His successor is Mr. Satoru Ishida.)

Our overall conclusion is that the Japanese are in the very early stages of considering the security aspects of database systems. Their Security Guidelines mainly involve physical and procedural security. Little security work is being supported by the Japanese government. Some work is being done by the Japanese Electronics Industry Association (JEIA), which has a security committee headed by Mr. Koichi Mori.

Lack of interest in security appears to be influenced by two factors:

1. Japanese culture appears to be such that product vendors are trusted to develop correct products.
2. Computers in Japan are not as highly networked as in the West. As a result, the Japanese have not, to our knowledge, suffered major virus attacks or intrusions from unauthorized users.

However, Japanese companies are keenly aware of the information security evaluation and requirements activities in the U.S. and in Europe. For example, Japanese companies such as Hitachi and Fujitsu are working with the United Kingdom's Communications-Electronics Security Group (CESG) so that their companies' products can be evaluated for use in the United Kingdom. Japanese companies also provided detailed comments on the Information Technology Security Evaluation Criteria (ITSEC) published by the European Community (EC).

In the U.S. the major driver for database security is the Department of Defense, which mandates requirements. Industry adapts slowly, but cannot ignore the requirements. The Self-Defense Forces of Japan do not seem to issue such mandates or purchase systems that are vertically integrated, so that these concerns become internal to the suppliers.

Concerns for personal privacy seem negligible in Japan, especially in comparison with Western Europe. Some public databases (see Section 5) seem to provide information which is close to that planned for the Lotus MARKETPLACE product, which was cancelled due to public uproar as a perceived invasion of privacy.

3.7 Database Tools

Author: Toshi Minoura

In addition to the basic DBMS, most users need software that effectively utilizes the information obtained from the databases. High-level tools of this type are referred to as fourth-generation languages, or 4GLs.

In Japan, most software is custom-made, and the use of software tools is not yet common. However, leading computer companies like Fujitsu and Hitachi provide a set of state-of-the-art software tools for their own machines. Package software that can be used on platforms from different vendors has mostly U.S. origins. For example, the most popular 4GL used in Japan is the Japanese version of ACCELL, which runs on Unix-based workstations provided by several Japanese vendors.

To give an idea of the software packages and tools available for the machines produced by Japanese computer manufacturers, we list those provided by Fujitsu for their mainframes.

1 EPG II (Executive Planning Guide).
2 C-NAP II/NA (Needs Analysis), SA (Systems Analysis), DA (Data Analysis).
3 YAC II (Yet Another Control chart), YPS (YAC II Programming System), YPS/APG (YPS Application Program Generator). These are COBOL program generators.
4 ADAM/IRD (Application Development And Management / Information Resource Dictionary).
5 CASET (Computer Aided Software Engineering Tools). An extensive set of CASE tools, including data dictionary support, data structure and form generators, and prototype test support.
6 BAGLES (Business Application Generator's Library for Extensive Support).

Hitachi also provides a similar line-up within its Software Engineering Workbench (SEWB). The PAD (Program Analysis Diagram) system automatically generates COBOL programs from structure diagrams.

Fujitsu provided the following list of application development tools for its workstations.

1 (Knowledge system HCI Server, interface generator).
2 KR (Knowledge Representation and inference).
3 NOAH (New Object-oriented And knowledge media management system for intelligent applications with Human interface). A commercial version of JASMINE.
4 IP (Intelligent Pad). A simple hypermedia system.

As standards for databases and their interfaces become established, the market for generic database tools will increase. Here is an opportunity for established U.S. vendors to serve a growing market. It will of course be necessary to invest in adaptation of these tools to the language and system requirements of the Japanese community.

3.8 Database Management Systems (DBMSs)

Authors: Sushil Jajodia, Toshi Minoura, Gio Wiederhold

As indicated in earlier sections, most databases in current use in Japan operate on U.S. (DB2, IMS, ORACLE, etc.) or German (ADABAS) database management system platforms. Very similar database systems are available on Japanese computers which do not provide compatible operating systems, such as the Hitachi (RDB1), Oki (REAM), and NTT (DEIMS-5) relational database management systems.

These systems are not perceived as being very satisfactory, as shown in [Hayashi:85], summarized in the table introducing Section 3.5. These problems strongly mirror dissatisfaction experienced everywhere among database users. Japanese database system software does not seem to differ inherently in quality from that seen elsewhere. However, comparisons are difficult since systems are in different stages of maturity.

3.8.1 Database Management Systems of Japanese Origin

In this section we describe the current state of affairs with respect to development of database management systems in Japan. Since Japanese industry is very strong in the area of electronic products, it is not surprising that hardware plays an important role in the development of DBMSs. Performance is an overriding consideration for the users of DBMSs. For this reason the Japanese have put their strength to work and are building database accelerators to enhance overall database performance, as described in Section 3.5.4. Projects representative of this enhanced performance follow.

Two notable efforts, one by Hitachi Ltd. and the other by NTT also are described. Hitachi is currently working on XDM/DF (eXtensible Data Manager), a distributed database management system. It provides many advanced distributed database features, including location transparency, unified access interface, and a two-phase commitment protocol.

We also describe the system integration efforts of NTT Data Systems, an independent service company that focuses on consulting and system integration. Every major organization today has systems purchased from several different vendors. These systems must be linked into a unified system, although components may be loosely coupled.

RDB1/DP (Hitachi Ltd.) . RDB1/DP is Hitachi's commercial relational DBMS. To improve overall performance, RDB1 (Hitachi's commercial relational DBMS), uses a new pipelined vector processor called IDP [Tori:87].

RINDA (NTT). RINDA is a relational database processor for speeding up two common relational operations – search and sort – efficiently. This makes RINDA especially suitable for performing nonindexed queries. In conventional database management systems, nonindexed queries consume much CPU time and I/O time and are often not feasible. The two common approaches create bottlenecks. If a search is performed on all rows of a relation, the amount of data that must be transferred becomes large and much CPU time is consumed. If on the other hand tuples are sorted on selected columns, the processing cost for sorting is high – on the order of $(n \log n)$ – and a large amount of memory is required.

RINDA removes both bottlenecks by providing specialized hardware that can perform an order n time sort on a large internal memory and on-the-fly search on disk storage.

A key property of RINDA is that it can be easily introduced into the existing systems without any modification of existing applications [Inou:89].

SDC (Kitsuragawa, University of Tokyo). Kitsuregawa and members of his group at the University of Tokyo are working on a high-performance, parallel relational database

server called the Super Database Computer (SDC), for a join-intensive environment. The basic architecture of SDC is shared-nothing providing high performance where search results are independent [Kits:90].

XDM (Hitachi Ltd.). The XDM (eXtensible Data Manager) is Hitachi's large mainframe DB/DC system which supports both relational and structured databases. It provides many advanced distributed database features including location transparency, unified access interface, and two phase commitment protocol.

3.8.2 Other DBMSs Mentioned in This Report

A Deductive Object-Oriented Database language, QUIXOTE
The core database system (SYNDES) from Tsukuba, which uses QUIXOTE
The JASMINE system being developed by Fujitsu Laboratories Ltd.
The IMAZONE image database construction tools for RICOHBASE
RICOHBASE, an extended relational DBMS supporting linked connections
OMEGA, Object-oriented Multimedia DB developed at the UNLib Tsukuba
The Hitachi RDB1 relational and XDM/DF distributable database management system
REAM, by Oki, an RDBMS being enhanced with RDA

3.8.3 Adaptation to the Japanese Market

To reach the Japanese market, foreign products have to be adapted at least to handle Japanese character sets. In earlier foreign products little consideration was given to handling non-ASCII or non-EBCDIC character representations. The absence of a single standard for Japanese characters hinders foreign developers as well. The simpler conventions do not provide for the following complications found in practical Japanese text:

- integration of Roman numbers and terms
- vertical and horizontal orientation
- font changes concomitant with changes in orientation
- complications induced by the combination of the first two points
- layout problems induced by mixed orientations

Some Japanese companies have invested greatly in making these adaptations.

Oracle and NTT Data. NTT Data uses various database products in building large, vertically integrated applications. They have a copy of the source code of the U.S.developed ORACLE system, and have modified it to meet the needs of the turn-key systems that they build. Recently Oracle Corporation has established an office in Japan to serve their Japanese customers directly with their own, now internationalized products.

ADABAS and Software a.g., Far East. A similar level of adaptation was made by an entrepreneur to the product of a German company, Software a.g.: ADABAS. This product has achieved an important niche in the Japanese market. It has focused on the IBM and compatible mainframe market, where many other offerings were not suited to adaptation. Ongoing cooperation with the source company and its U.S. offshoot have motivated further internationalization.

For several important areas, such as Japanese textual retrieval for JICST, adaptations to ADABAS have been made available as well. This has created new market opportunities not covered by other DBMS products. Modules, such as report generators specifically oriented toward the Far East market, are also available.

3.9 Summary

The DBMS market is yet far from settled. While RDBMS approaches are well-known, many applications use older and less general software. Where this software has been adapted to the Japanese market it will be at least as hard to replace as in the U.S., where the application codes are yet often older.

However, the demands for new applications and advanced facilities are high, so that there remain many opportunities in this market, which seems to be equally open to all participants.

4. DBMS Product Infrastructure

Author: Diane Smith and John Smith

This section presents the market context within which database technology is deployed, as products and services. It describes the channels used by American companies and the roles these channels play in the Japanese marketplace and in Japanese global marketing strategy. The market factors that have shaped database products in both the U.S. and Japan are identified. The section concludes with a discussion of Japan's response to these drivers.

4.1 DBMS Technology Channels

To facilitate comparison between U.S. and Japanese products and services, we characterize the DBMS marketplace by three channels: off-the-shelf (OTS), original equipment manufacturer (OEM), and systems integration (SI). While this is not the only possible characterization, it does permit us to highlight the differences between the two countries.

"Off-the-shelf" products are sold through the most familiar channel in the U.S. marketplace. Database products distributed through this channel include DBMSs and database tools. Both hardware vendors and independent software vendors sell a wide variety of such products: DBMSs based on the CODASYL, relational, and object data models; DBMSs that operate on mainframes, mini-computers, workstations, personal computers, and database servers.

In the OEM mode, products are marketed to third party, value-added resellers (VARs). VARs typically build applications on top of DBMS products and resell them as "turn-key" systems. While this mode of of marketing a product is frequently thought of as part of the OTS channel, it is distinguished here because this channel is just emerging within Japan, and is worth separate comment.

In systems integration, existing hardware, software and communication products are integrated, frequently with custom software, to produce large customized systems that automate business, engineering, and manufacturing processes. In SI typically, services in addition to products are sold: integration, documentation, installation, and maintenance services as well as the production of custom software.

4.2 DBMS Market Drivers

The Japanese market appears to differ markedly from the U.S. market with respect to DBMS products. In the U.S., there is a strong demand for both OTS products and for open systems. This is particularly true in the commercial marketplace. In the 1970s, this led to the emergence of both hardware and third party independent software vendor-supported DBMS products. In the 1980s, with the widespread acceptance of the relational model as a de facto standard for DBMS interfaces, there was an explosion of DBMS products. In the 1990s, these products are expected to be available on a wide variety of hardware platforms and to be compliant with OS and communications standards.

The demand for standardization of software product interfaces increased in the 1980s and has continued into the 1990s. The relational query language SQL was released as an ANSI standard in 1984. Major efforts are being made to develop application data model standards and interchange standards such as RDA to support computing in networked environments.

In Japan, the emphasis in large corporate software acquisitions has been on buying total system solutions. These large hardware/software systems are typically custom systems developed in partnership by the customer and the vendor, with the vendor not only supplying the system, but installing it, operating it, and maintaining it.

In summary, one can characterize the major U.S. drivers of database technology as being the development of off-the-shelf products and open systems, while in Japan, the drivers appear to be the requirement for total, custom systems. This difference in market drivers in the U.S. and Japan has produced differing patterns as to how the various distribution channels are perceived and used.

4.3 Japanese Positioning

To better understand how database products are positioned with respect to the OTS, OEM, and SI channels in Japan, we first distinguish the different classes of companies that produce these products within Japan. The class of companies most visible outside of Japan is that of the full-product line hardware vendors. These are typically part of very large Japanese corporations. Members of this class that we visited are Hitachi, Mitsubishi, Toshiba, and NTT. IBM Japan also falls into this class although it is not, per se, a Japanese company.

A second class consists of manufacturers of office products and hardware peripherals such as faxes, scanners and printers. Ricoh, visited by the JTEC team, is typical of this class. The final class is that of independent software vendors. The only member of this class visited by JTEC was the Japanese associate of a German company, Software a.g.

4.3.1 Japanese OTS Database Products

Japanese computer hardware companies produce a variety of off-the-shelf database products. Their software offerings are typically products that run on large mainframes and are based on proprietary operating systems. For example, Toshiba produces both a CODASYL-based DBMS and a relational DBMS for its mainframe computer. These products are well supported — they are upgraded on a regular basis to include the same types of features as U.S. DBMS software products. For example, the Hitachi XDM/RD (Relational Database) now supports distributed data.

The Japanese computer companies also offer an interesting range of database hardware products. Hitachi, for example, produces intelligent disks, search engines, and hardware accelerators. Toshiba offers a similar set of products, a number of which use specialized chips produced by Toshiba.

In discussions with Toshiba we explored the rationale for developing specialized hardware, since U.S. companies have not found it profitable to develop a wide range of similar offerings. A number of these products are bundled with either mainframe or peripheral systems. Toshiba views these products as adding to the quality of the overall system in terms of either reliability or performance and therefore are not considered optional. Furthermore, Toshiba is unconcerned about the need to upgrade the specialized chips as technology advances, since this type of consideration is built into its normal engineering and manufacturing processes.

The products discussed at the site visits and described in the brochures made available to us are summarized in Section 3.9.

4.3.2 The OEM Channel in Japan

The OEM channel is used in two different ways in Japan. Hardware vendors use independent software vendor DBMS products to fill out the software product lines for their hardware offerings. For example, Toshiba uses ORACLE on its UNIX-based workstation offering. It also uses a tool package developed by McDonnell-Douglas as part of its software product line.

The second use of the OEM channel is to produce turn-key systems. This is atypical in Japan, although it is standard practice in the U.S., where they represent a major marketplace

opportunity. The particular example we saw of a company producing a product for the OEM market was Ricoh. Ricoh has developed a software integration platform called IMAZONE. It integrates the use of Ricoh's scanners, faxes, optical storage devices, and printers for use in building imaging and document management systems. It provides a set of tools (e.g., an image editor; image storage and retrieval software; and a relational DBMS (RICOHBASE)) specially tuned to handle images and graphical data.

Several Japanese value-added resellers (VARs) are using IMAZONE to build turn-key systems. One is building a technical manual management system. A second is building a grading system to support the Japanese "cram" schools. This system will direct assignments faxed in by students to the instructors who will grade them. The instructors will view the assignments in electronic form, use an editor to mark them with grade, corrections and comments, and return them to the students via fax. Electronic copies of the transactions will be stored and managed by the system.

Although it was not on the official visit list, Murata, a small Japanese company that has been succeeding as a VAR, was visited informally by Toshi Minoura. Murata has built a CIM system using the ORACLE DBMS product. This is one of the few examples we saw of this channel being used as it is in the U.S.

4.3.3 System Integration

Because the emphasis in Japan has been on the acquisition of total and custom solutions, systems integration is the most important channel in Japan. To meet this demand, the large Japanese hardware vendors have developed strong, vertically integrated hardware, software and service offerings. These include mainframes and peripherals, operating systems, DBMSs and tools, text and image subsystems, and a service organization that can design, implement, document, test, install and maintain large systems.

Typically, a company like Mitsubishi will specialize in selected vertical markets (e.g., utilities and railroads). To support these markets, the company will develop as part of its product offerings whatever technology is needed. At Mitsubishi, we saw the prototyping of a scanner-based system for the input of utility floor plans. An image recognition system specific to the symbols used on floor plans was being tested. The system being developed links real-time sensors on a plant floor to a display showing the floor plan, and to data representing the interrelationships of equipment on the floor so that problems can be recognized and diagnosed.

The research activities supported by these companies frequently tie in directly to their targeted markets.

4.4 The Japanese Response to the Global Database Market

As the Japanese anticipate expanding their markets, they are responding to the factors they see driving the U.S. and global marketplaces. The U.S. market drivers, open systems and OTS software, also are strong factors in the European and global marketplaces.

The response from the Japanese government to the global market drivers has been twofold. MITI, through INTAP, has initiated a major program on interoperability and is providing significant support of international standards.

Japanese companies have also made a visible and significant commitment to standards support. Each company we visited supports representation on the international standards committees. In addition, they keep in close touch with technology developments worldwide by sending staff members abroad and by interacting with Japanese university faculty that maintain close contact with the international database research community.

Japanese university database faculty are particularly aggressive in interacting with the international research community. This is in part because many of them feel that there is not enough interest in their research by Japanese companies. However, this works to the advantage of the Japanese companies because faculty members become a very efficient means of culling external results and introducing them into Japan, through student training and research publications made available at Japanese language conferences. This and the fact that many Japanese professionals can read English provide efficiency in propagating international technology.

4.5 Observations

While the information we gathered with respect to the Japanese database technology marketplace is purely anecdotal, it set the context for understanding differences between Japan and the U.S. with respect to developing and deploying database technology.

In the U.S., the OEM market is entrepreneurial. The two examples we found in Japan follow this pattern. While the group we visited at Ricoh is part of a relatively large company, it differed from the groups we visited at other large companies in several ways. To begin with, the group is headed by a woman, Dr. Kunii. At no other site did we meet with a woman in a professional leadership role. Not only was the group headed by a woman, but one of the chief technical leads was a female Ph.D. We saw at least one other woman on the staff, a programmer from Ireland.

Dr. Kunii received her Ph.D. from the University of Texas. Members of her staff, both junior and senior, described their work environment as being open, in the American style. They were proud of their creativity and in what they had accomplished. They credited Dr. Kunii with creating their environment, and with the success of the group. There is a shortage of trained scientific workers foreseen in Japan. Since Japan does not easily accept foreigners, we can expect that Dr. Kunii's example presages changes for the future. How these changes will affect the fabric of Japanese society is a question outside the scope of our report.

The requirement for total and customized solutions established in Japan is emerging as a global market driver. U.S. companies are repositioning to meet this demand. Defense contractors have decades of experience in performing systems integration. Typically, they put together teams of companies to deliver the components and services needed to pursue a given opportunity. However, the U.S. defense marketplace is very different from the commercial marketplace. The procurement process is standardized and slow; ideally, selection is made on the basis of technical adequacy and low bid. Continuity is often broken. Requirements are often obsolete and inflexible.

This contrasts with the Japanese commercial marketplace, where procurement is based on long-term strategic relationships and involves deep application knowledge restricted to a limited number of application areas. As discussed in Section 4.3.3, Japanese firms are strongly vertically integrated, and hence are able to supply total solutions without negotiating complex teaming arrangements. They have built up foreign experience and reputation in the Middle East and Asia within their target markets.

Large U.S. firms such as IBM, DEC, and UNISYS are now emphasizing this marketplace. Smaller software companies selling specific applications are offering integration services. We can expect intense competition between the U.S. and Japan over systems integration business in the near term.

5. Database Use Industry

Authors: Charles P. Bourne and Nick Farmer

In addition to database technology we consider the industry that sells information from databases. Funding sources for the government-sponsored services were identified in Section 2 of this report, and include STA (JICST) and MONBUSHO (NACSIS). This industry is mature, and typically uses software that differs from the software provided by generalized DBMS providers, although there are important exceptions, as noted earlier for ADABAS, as adapted by Software a.g., Far East. This industry focuses on effective search services, and on acquisition and maintenance of large information files.

5.1 Database Search Services

In contrast with the computer database searching in the 1960s and 1970s, which was predominantly a mainframe batch searching activity, most of the computer-based search systems in operation throughout the world today are operating now in an interactive online mode. This panel concentrated its attention on interactive mainframe systems, which is the focus of this chapter. New computer architectures, as networked workstations, have not yet impacted the suppliers, whose focus is the maintenance and storage of large databases.

5.1.1 Points of Comparison

In reviewing the Japanese database search services we considered the following attributes as the basis for comparison.

Size of the Service. Thousands of online search systems are in operation today throughout the world. Most of them are strictly inhouse systems that are in operation to serve the needs of individual organizations (e.g., a corporate management information system). Others have been developed to serve larger but limited constituencies (e.g., regional or national law enforcement networks, airline reservations systems). A relatively small number of online systems have been established to operate on a worldwide basis to provide information (often for a fee) to almost any interested user (e.g., the NLM Medline service). Some services have been developed for end-user or consumer markets to serve customer populations with over 100,000 subscribers. We have reviewed the Japanese online services in this context, and have not found any that have yet operated on the scope and scale of activities that are seen with several U.S. services. They would have to go through a development and learning process before they could do that — but they are able to do so.

Content. Online search systems are in operation throughout the world today that collectively address a great variety of types of information to be stored and searched. Online search service is provided throughout the U.S. on a regular production basis with at least the types of file content shown in Table 5.1 (although not usually all provided from the same search service). North American and European online services are extremely competitive with each other, and because of that, have been fairly proactive in developing new system features as enhancements or improvements to their search systems (e.g., multi-file search software, word proximity searching, online search aids, improved document delivery service, improved customer service). The online services in Japan, which are oriented toward Japanese-language databases, generally do not compete directly with North American and European online services. Furthermore, since many of the Japanese systems are government supported and directed towards a particular market segment, e.g., industry or academia, there is little competition within Japan. This lack of competition has resulted in a somewhat slower pace relative to adoption of new system features.

Types of File Content	Description (Examples)
bibliographic files	indexing and abstracting records (Biological Abstracts, Chemical Abstracts); catalog records (Catalog of the Library of Congress)
fulltext	full character representation of news publications, journals, newsletters, and other material (New York Times, Washington Post, New England J. of Medicine, Associated Press newswire)
biographic information	information about people (Amer. Men and Women of Science, Who's Who)
directories	organization descriptions (Moody's Corporate Profiles, Encyclopedia of Associations)
financial data	company financial reports (SEC Online), stock and commodity price, financial time series
demographic information	information about populations (1990 U.S. Census data)
images	scanned bit-representations of image data (U.S. trademarks and patent drawings)
chemical structure	representations of chemical structures (Chemical Abstracts Services Registry, Beilstein)
scientific numerical data	physical properties, measurements of various substances, spectrographic or other measured data about various substances

Table 5.1 File content of searchable databases

Operational Support. The level and extent of computer system support in place at some of the major online services in North America and Europe seems to be far above that provided by almost all of the online services in Japan. Examples of such activities are shown in Table 5.2.

Type of Support	Description (Examples)
search service for multiple files per search	There is concurrent availability of more than 100 files for users to choose from on several U.S. search services. Most Japanese search services provide only a few databases.
current file updates	Many U.S. public search services provide daily or even hourly updates to their search files (Official Airline Guide, some news files). Few Japanese online services have updates as frequently as daily.
large files	No public online search service in Japan operates with the file sizes and storage requirements experienced by several U.S. online search services.
extensive file maintenance	Several U.S. online search services perform massive file maintenance activities on a regular production basis (many file updates and reloads for multiple files on a single system, with a combined storage size of over 500 Gbytes). No large-scale support activities such as these were seen with any of the Japanese online search services.
full-time availability	A 7-day/week, 24-hour/day (public utility equivalent) service with over 99% real availability is a service objective that many U.S. search services approach. Few Japanese online services approach this level of computer system support.

Table 5.2 Support levels for database searching

None of the Japanese online systems seemed to be operating with the power and sophistication of computer support activity provided by the U.S. search services.

Database Search Service as an Export Industry. A small number of the Japanese online services make their online services available to users outside of Japan. Examples are listed in Table 5.3. The actual use of the services outside of Japan is rather small.

Service	Acronym	Host Organization
JICST Online Information System Service	JOIS	JICST of STA
NACSIS Information Retrieval Service	NACSIS-IR	NACSIS of MONBUSHO
Japan Economic Journal	NIKKEI	Nihon Keizai Shimbum, Inc.
COMLINE Business/Industry Monitors	COMLINE	COMLINE Int'l Corp.

Table 5.3 Japanese services available internationally

The STN Service, an international cooperative of online search services that provides service worldwide, includes a Japanese participant (JICST), but this special case, is not an example of a Japanese search service trying to independently reach and serve a global market. Japanese online search services are not a significant export activity at this time.

5.2 Databases as Export Products

This segment of the report will consider databases as a product for distribution inside and outside of Japan.

As pointed out in reports by the Database Promotion Center in Japan [DPC:90], a large and growing number of databases are produced in Japan. However, in contrast with hundreds of North American and European organizations that produce English-language databases and expect to realize some revenue from the use of those databases, few Japanese organizations have prepared databases as a proprietary product to be sold or licensed (particularly outside Japan) as a revenue-generating product for use on computer systems other than those of the originator. A key reason for this is undoubtedly that databases in Japanese have a very limited international market, and few English-language databases are produced in Japan. Specific examples of the few English-language databases produced in Japan and made available to search services outside of Japan are shown in Table 5.4. These are in addition to those listed in Table 5.3. These files are not used extensively outside of Japan.

Supplier Initial Year	Database	Volume Content
Asahi-Shimbum 1986	Asahi-Shimbum Online News fulltext records	> 10 000
COMLINE International Corp. 1986	COMLINE Business Analysis fulltext records	2 000
1986	COMLINE Industrial Monitor daily updates	40 000
	COMLINE Japan Daily daily update Areas: Biotechnology, Chemicals, Computers, Electronics, Industry Automation, Telecommunications, Transportation	
1986	Tokyo Financial Wire citations, daily updates	20 000
Japan Info. Center of Science & Technology (JICST) 1985	JICST File on Science, Technology and Medicine in Japan citations and abstracts	1 000 000
	Japanese Government & Public Research in Progress descriptions of research projects covering 600 Japanese institutions	30 000
Japan National Diet Library 1977	Japan MARC catalog records	700 000
Japan Patent Info. Org. 1955	JAPIO Japanese patents abstracts and drawings	31 000 000
Japanese Standards Assoc.	Kikaku Net citations to Japanese and foreign standards	213 000
JIJI Press Ltd. 1980	JIJI Press Ticker Service JIJI Press wire service fulltext records	
	JIJI Securities Data Service (JSD) Tokyo Stock Exchange Stocks current information	≈ 1 600
Kyodo News Int'l Inc. 1984	Japan Economic Newswire (Tokyo) English-language fulltext newswire stories reported by Kyodo News Service	> 140 000

Table 5.4 to be continued

Supplier Initial Year	Database	Volume Content
Nihon Keizai Shimbum Inc.	Japan Economic Journal	
(NIKKEI) 1980	fulltext records	
	NEEDS (separate time-series files)	
	Areas: demographics, commodity prices, business/finance, retail point-of-sale data	
	NIKKEI Financial File	1,200
1964	financial time series	
	NIKKEI Stock and Bond Price File	
	Stocks traded in Japan	
	twice-daily update	
	NIKKEI Telecom II Japan	
	Financial News and Data	
	fulltext coverage of Japanese	
	business/financial community	
Nomura Research Inst.	NRI/E Japan Economic and Business DB	5 600
1965	time series for the Japanese economy	
QUICK Corp.	QUICK Information System	
1971	economic information	
Sumika Technical Info.	Japan High Technology Monitor	30 000
Service 1983	citations	
Teikoku Databank, Ltd.	Teikoku Databank	45 000
	profiles of Japanese companies	

Table 5.4 English language accessible databases

JICST, supported by the prime minister's office through STA, provides various online services such as bibliographic databases and factual databases through the JICST Online Information System, JOIS and JOIS-F respectively, in Japanese. JOIS and JOIS-F run on Hitachi mainframes.

Services are provided to branch offices through a network of 14.4, 19.2, and 64 Kbps lines and then to users through 300-2400 bps lines. Overseas users can also use these systems to access some of the databases but must do so via the Japanese language. In addition, JICST participates in the Scientific and Technical Information Network (STN), which is a worldwide integrated computer network system for scientific and technical information.

Table 5.5 gives details of the various services and their usage.

	JOIS		JOIS-F	STN
Type of Service	Bibliographic DB		Factual DB	Bib. and Fac. DB
When Started	1976		1988	1987 (in Japan)
Number DB	15		7	96
Number Items	23M		0.6M	67M
Used Hours/Month	12 000		30	3 000 (Japan)
Number Queries/Month	88 000		250	26 000 (Japan)
Number of users	('90)	('86)	('90)	('90)
	8 700	4 600	900	2 500 (Japan)
Company	74%	76%	85%	79%
University	9	11	5	11
Government	5	5	4	5
Institute	3	3	2	3
Hospital	3	3	-	-
Other	6	2	4	2

JOIS: JICST Online Information System (Japanese)
JOIS-F: JICST-Factual Database System (Japanese)
STN: Scientific and Technological Network International:
USA-Germany-Japan (English)

Table 5.5 JICST database parameters

The Science Information System (SIS) of MONBUSHO has NACSIS at its center. SIS electronically links university libraries, computer centers, and national research institutes (such as the National Lab for High Energy Physics), and provides various online database services. The communication hardware part of this service is the Science Information Network (SIN), which is a privately operated packet switching network employing packet multiplexers installed throughout Japan at major research areas. This network will serve as the main communication facility for the Inter-University Computer Network of Japan (an N-1 network). Plans are to permit fulltext, graphics, voice and image communications. NACSIS's computer system includes Hitachi mainframes (M-684H, M-682H) with special database accelerator hardware, and large magnetic, optical and semiconductor disks. A 10 Mbps Ethernet is used internally for R&D, and to provide workstation access to the main system. To provide electronic mail service, an NEC ACOS 1000 mainframe and disks are provided, along with two 48 Kbps lines to SIN and some lower speed lines to public telephone networks. From SIN a 9.6 Kbps line connects to the British Library, and a 14.4 Kbps line connects to NSF and the Library of Congress, both in Washington, D.C.

A substantial set of services is associated with cataloging, aiming eventually at having only one catalog of all scholarly material in Japanese universities. Basic software is XDM-RDB enhanced to allow multi-valued entries. Currently almost 500 concurrent terminals can be supported, with an average interval of about 45 seconds between commands per user (thinking time). Most Japanese workstations can be used as cataloging terminals as well as IBM 5500-series terminals. A special full-screen interface has been provided to allow the input and display of Japanese characters.

For retrieval, about thirty databases can be queried including standard databases such as MathSci, COMPENDEX Plus, and the Harvard Business Review. The NACSIS databases include reported synopses or research projects conducted through research grants-in-aid by the Ministry of Education (Japanese and English), dissertation index (Japanese), academic conference papers (Japanese and English), a union catalog of Japanese and foreign books and serials, and a database directory (Japanese). The NACSIS databases include about three million pieces of data from Japanese university library catalogs. Hitachi provides an enhanced version of the U.S. system ORION, on which many applications have been written to adapt the user interface. Typically, several dozen users will be concurrently on the system.

NACSIS focuses on service, but it has a number of research activities, as well. There is substantial interest in advanced hardware such as optical media, and in parallel systems such as an N-cube; new database models such as nested-relation; quality control methods for database construction; human interfaces; and formal description languages for communication protocols based on temporal logic. However, the staff is concerned that performance not be degraded by the use of new techniques.

While we were at NACSIS, we were given a demonstration of the query system, which generated about two dozen references to object databases from Japanese authors. Although several members of the JTEC database team had queried other databases available in the U.S., accessing the NACSIS system would have been very helpful in deciding which sites to visit. We encourage future teams to use this service, which is available via the NSF. At the same time the team had the opportunity to discuss impediments to its use in the U.S. Although the NSF staff is exceptionally cooperative and helpful, because the facility uses specialized hardware, it is difficult for remote researchers to use. This is an important item for potential improvement.

Some files about Japanese activity are produced by organizations outside of Japan. Articles in Japanese scientific and technical journals have been routinely covered for years by the major international indexing and abstracting services outside of Japan (e.g., Biological Abstracts, Chemical Abstracts, Engineering Index, Index Medicus, and Mathematical Reviews). Furthermore, several databases have recently been produced by Western organizations to specifically cover Japanese activities for Western users (e.g., Japan Technology, JAPI).

One of the major reasons for the low number of databases produced by the Japanese for export is simply that the international information industry operates largely with English-language information, and English is not the primary language of the Japanese database producers. In comparison with many other countries where the database is originally produced in English, any Japanese database must undergo a Japanese-to-English translation effort before it can be made available as a significant export product. North American and European database producers would be similarly disadvantaged if the international information industry happened to serve primarily a Japanese-speaking customer base.

Given the extra effort and cost that is required for Japanese organizations to produce English-language databases, it does not seem likely that the Japanese information industry will produce English-language databases that would be direct competitors to existing U.S. databases. This would be particularly true for any databases that provide worldwide coverage of a particular subject area (e.g., Medline, Biological Abstracts, Mathematical Reviews), but it might be less true for niche databases with an audience that is accustomed to paying higher prices for information (e.g., Pacific Rim business information). There are several efforts to develop an effective Japanese-to-English machine translation capability, but current systems do not seem to be operating yet with a cost and performance level that is likely to have a significant impact in the immediate future.

Japanese database producers have little experience to date in providing databases and their updates to multiple computer services worldwide, and in providing the accompanying marketing, training and customer support services to the recipients of those databases. In contrast with some U.S. database producers (e.g., NTIS, NLM) who provide their databases and current updates to more than twenty organizations worldwide, very few Japanese database producers have the experience of providing their databases, updates, and associated support to more than two or three other organizations.

In summary, the Japanese database industry is in a position to develop and improve the information products that are relevant to activity in Japan (e.g., Japanese newswires, Japanese organization directories, Japanese stock prices and time series data, and other source publications), but does not appear to be in a position to provide significant direct competition with the U.S. database industry in topical areas.

5.3 Database Technology

Currently available Japanese databases represent a range of technical complexity with respect to record structure and file organization that is quite familiar to U.S. database producers and search services. Files and record structures associated with library catalog files (Japan MARC vs. LC MARC), bibliographic files (JICST-E vs. Engineering Index), company directory files (Teikoku vs. Dun's), and newswire services (Kyodo vs. UPI) are of the same degree of technical complexity for databases produced in both countries.

However, the U.S. producers and services have, on a case-by-case basis, pushed a relatively plain database technology to establish more complex structures and capabilities in the file and record formats to increase utility and advantage for the user. These organizations have, in a sense, pushed the operational envelopes of the search system capabilities by changing file and record structures, not by changing the computer hardware requirements. Examples of expanded file features in operation today in the U.S. include:

- expansion of the number of separately searchable fields in a record (say, over 200)
- expansion of the number of records that can be searched at one time in a single file (say, over 20 million records per file)
- expansion of the number of separate files that can be searched at one time by the user (say, over 200)
- searching and consolidation of master/subordinate records (say, sections within a handbook, questions within a questionnaire, Congressional Hearings records subordinate to a specific Bill).

In no case did we see Japanese databases that matched the large and more complex U.S. databases of the type listed above. We have no reason to believe that they could not handle these problems. We mention it only to note that U.S. institutions have already faced and

passed a number of technical and application obstacles that the Japanese probably have not faced yet. In that regard, the U.S. could be considered to be further ahead with database technology.

Image Management. There is one area of database technology where the Japanese should receive special recognition, namely high quality image storage and retrieval. Team members visited the National Ethnology Museum to observe a system used there to store high-quality three-dimensional color images of museum artifacts, along with associated descriptive information. A file of over 50,000 such images had already been assembled (at 7 MB/image, this represented a total storage volume of 350,000 MB), and a retrieval capability implemented for an inhouse system to search, retrieve, and display specific color images on the basis of their indexed attributes. The system was built by the IBM Tokyo Research Laboratory. A dial-up capability was planned for researchers at remote locations.

Another color-image storage and retrieval system was demonstrated by NTT Data Communication Systems Corp. in Tokyo for operational use by the Textile Rationalization Agency. Both of these image retrieval systems were developed by Japanese R&D organizations, but neither of them have received any extensive operational use to date.

Additional Japanese image storage and retrieval systems were seen at Ricoh, IBM Japan, and other institutions as part of office information systems in development. Several other systems were mentioned by our hosts during our visits to various Japanese organizations.

Several U.S. search systems also provide operational online search, retrieval, and display of scanned images (e.g., U.S. trademarks) or generated images (e.g., chemical structures), but this aspect is still not seen on a large scale in the U.S.

Multimedia storage and retrieval systems appear to be an area where the Japanese organizations are as advanced as those in the U.S.

5.4 Summary Assessment

For database search services, the Japanese organizations at this time are judged to be:

- behind the U.S. in terms of size and extent of offerings, technical features, and performance
- not a significant challenge to the U.S. online search service industry
- not catching up in database services.

For database content as an export product, the Japanese organizations at this time are judged to be:

- far behind the U.S. in producing databases as an export product
- short on experience in producing/supporting databases for use by large numbers of search services
- facing an inherent and continuing disadvantage in competing to produce English-language products
- not a significant challenge to the U.S. database industry
- not catching up in database content production.

Japanese online database service organizations at this time are judged to be:

- familiar and experienced with mainstream technology
- less familiar and experienced with large and complex files and retrieval situations that tend to stretch the bounds of normal database practice
- probably as current as U.S. organizations with regard to image or multimedia retrieval systems
- catching up with technology

6. Summary

In this section we integrate the findings of the preceding chapters and focus on areas where Japanese efforts have been intense, and are likely to have effects that transcend Japan in a significant way.

6.1 Multimedia Databases

In terms of infrastructure technology, the most significant advances we saw were in multimedia technology. By building database systems that exploit the Japanese strength in electronic devices, such as high-capacity optical and magneto-optical storage, document and image scanners, and image presentation on standard video, high-definition video, and computer-driven monochrome and colorfax equipment, unique advances are being made. These capabilities are not yet well integrated into networks and standards are lagging. The Japanese ISDN network is poised for a major expansion in bandwidth (from 256 KB to 4 GB). The availability of such links will further motivate this direction and cause pressure for integration of advanced multimedia, especially image technology, into databases.

Today there is significant dependence in Japan on foreign database management system technology. Most of the vendors of these database management systems are planning to provide support for the management of large, variable-sized data elements, as needed for multimedia database management. It will depend on the effectiveness of these extensions whether established DBMSs will be used for the multimedia services of the future. Otherwise it will be necessary for the developers of multimedia systems to develop their own DBMSs. The availability of standards such as SQL and ADA, makes entry of new DBMSs that satisfy these standards feasible. Even if they are less mature, having multimedia capability can be a decisive factor in the market.

Intermediary solutions do exist. Conventional DBMSs can reference images in distinct files for images and the large objects, and these can be accessed indirectly. However, such solutions are more complex to manage and are likely to be intermediate solutions. Furthermore, if pattern matching or associative access to image and voice data becomes a reality, then the indirect approach will no longer be feasible.

It is becoming understood that eventually access to multimedia databases will be required. Associative access means finding an image that 'looks like this image' or that contains features 'like these'. Speech files can be interpreted for voice print identification as well as contents. Research into this problem is in the early stages, both in the U.S. and in Japan, so that its relative success cannot now be assessed. Japanese efforts have focused on neural-net technology, which is likely to be quite effective for the simpler matching problems, but may not deal well with feature-based searching. The availability of excellent technology in Japanese laboratories reduces their entry cost for researchers interested in this field. If this research direction either catches the interest of Japanese industrial research, or if academic research in this field finds support, rapid progress is possible.

6.2 Government and Database Research and Development

The influence of governmental initiatives seems to be pervasive, but much less deep than other reports on Japanese technology indicate. It may also be less deep than governmental officials hope and surmise.

As indicated in Section 2, the Japanese academic establishment receives relatively little research project and student support from the Japanese government; support for industrial research laboratories is also minor. There are of course substantial government-funded research laboratories, for which governmental support is critical.

But overall, the direct government influence on technology directions seems minor. However, there is a very strong indirect effect. The participants meet regularly in committees and workshops. This communication benefits rapid dissemination of Japanese research, and even more importantly, of evaluated foreign research.

The dependence on foreign research does cause some faddism, since novel and superficial ideas are easier to disseminate in discussion groups than deep results. However, weak ideas are rapidly recognized when put to the test in industrial research laboratories. When the ideas are government sponsored, they may live longer, but are unlikely to confuse the primary direction of Japanese industry.

Acknowledgments

This study was supported by the National Science Foundation, the Department of Energy, and the Department of Commerce. Individuals were also partly supported by their institutions. Operational management of this study rested with the JTEC office at Loyola College in Maryland. Duane Shelton accompanied us in Japan. Important contributions were made by Geoff Holdridge, specifically the graphics for the qualitative comparison (see Executive Summary). Without the assistance of all supporters, institutional and personal, this report could not have been created. Most of the arrangements for our visit were made by Cecil Uyehara under contract with Loyola College. His knowledge, insight, and participation in the study visit considerably enhanced this report.

We also must convey a warm appreciation to our Japanese hosts. They handled our questions cooperatively and with concern, often providing needed background information. At many sites, informative formal presentations had been prepared at considerable effort. There were also occasions with wonderful food and beautiful scenery. Even though the press of time was always upon us, we also were able to appreciate the country and its inhabitants.

We thank our Japanese hosts for their useful comments during the review process of this report.

Appendices

A. Participants' Biographies

David Beech

Affiliation: Oracle Corporation, 500 Oracle Parkway, Redwood Shores, CA 94065
Telephone: 415-506-6420 Fax: 415-506-7200
Email: dbeech@oracle.com

David Beech is Senior Architect and Manager of Object Systems for Oracle Corporation. He joined Oracle in 1988 after eight years with Hewlett-Packard Laboratories, where he initiated work in object databases in 1982, which evolved into the Iris project, for which he designed and implemented the Object SQL language. Prior to this, he was with IBM from 1963 to 1980 and was a member of the original design team for the DB2 relational database system from 1976 to 1979. Earlier work was in programming languages, where he was PL/I Language Manager for IBM and was a major contributor to the design, implementation, formal specification and standardization of the language. He has also pursued interests in user interfaces, and was chairman of the IFIP Working Group 2.7 on Computer System User Interfaces from 1982-1989. He has published numerous technical papers, and has chaired and organized international conferences. He received his M.A. in Mathematics from the University of Cambridge, and initially taught mathematics for five years.

Charles Bourne

Affiliation: DIALOG Information Services, 3460 Hillview Avenue, Palo Alto, CA 94304
Telephone: 415-858-3775 Fax: 415-858-3847

Charles Bourne is Vice President of DIALOG Information Services, where he has been in a product development/product manager capacity since 1976. From 1971-76 he was simultaneously Professor-in-Residence, School of Library and Information Studies, and Director, Institute of Library Research, both at the University of California-Berkeley. He has worked on the development of a great variety of computer-based information systems. He has served on advisory boards to several professional publications. He was a consulting correspondent to the National Academy of Science Committee on Scientific and Technical Information, a UNESCO consultant to Indonesia and Tanzania, a U.S. National Academy of Sciences consultant to Ghana, and a member of the U.S.-Egyptian Task Force on Technical Information. He is presently a member of the Library of Congress Network Advisory Committee. He received his B.S.E.E. from the University of California-Berkeley, and his M.S.I.E. from Stanford University. He has been President of the American Society for Information Science and a member of the Board of Directors of the National Information Standards Organization.

Nick Farmer

Affiliation: Chemical Abstracts Service, 2540 Olentangy River Road, Columbus, Ohio 43210
Telephone: 614-447-3714 Fax: 614-447-3713
Email: naf20@cas.org

Nick A. Farmer is Director of Information Systems at the American Chemical Society's Chemical Abstracts Service. He is responsible for all aspects of computer systems, including research, technology, development and operations. Farmer joined Chemical Abstracts Service in 1970 as a senior programmer and subsequently served as advanced development manager, development projects manager, and senior research and development engineer before being named Assistant Director of Research and Development in 1982, and Director of Research

and Development in 1984. He was appointed Director of Information Systems in 1986. In addition to his responsibilities at Chemical Abstracts Service, Farmer is a member of the Board of Directors of Hampden Data Services, a software development company located in the United Kingdom. Farmer received the B.S. and M.S. degrees in electrical engineering from the University of Maryland.

Sushil Jajodia

Affiliation: George Mason University, ISSE Department, Science and Technology II, Room 459, Fairfax, VA 22030-4444

Telephone: 703-993-1653 Fax: 703-993-1521

Email: jajodia@gmuvax2.gmu.edu

Sushil Jajodia is Professor, Information Systems and Systems Engineering at the George Mason University. He joined GMU after serving as Director of the Database and Expert Systems Program within the Division of Information, Robotics, and Intelligent Systems at the National Science Foundation. Previously he was the head of the Database and Distributed Systems Section in the Computer Science and Systems Branch at the Naval Research Laboratory, Washington, D.C., and Associate Professor of Computer Science and Director of Graduate Studies at the University of Missouri, Columbia. His research interests include information systems security, database management and distributed systems, and parallel computing. He has published more than sixty technical papers in refereed journals and conference proceedings and has co-edited three books. He has served in different capacities for various journals and conferences. He is on the editorial board of the IEEE Transactions on Knowledge and Data Engineering. He is a senior member of IEEE Computer Society and a member of the Association for Computing Machinery. Jajodia received his Ph.D. from the University of Oregon, Eugene.

David Kahaner

Affiliation: Office of Naval Research, Asia, APO San Francisco, 96503-0007

Telephone: 81 3 3401-8978 Fax: 81 3 3403-9670

Email: kahaner@xroads.cc.u-tokyo.ac.jp

David K. Kahaner joined the Office of Naval Research, Asia in November 1989. He obtained his Ph.D. in applied mathematics from Stevens Institute of Technology in 1968. From 1978 until 1989 Dr. Kahaner was a group leader in the Center for Computing and Applied Mathematics at the National Institute of Standards and Technology, formerly the National Bureau of Standards. He is currently on leave of absence from that position. He was responsible for scientific software and library development on both large and small computers. From 1968 until 1979 he was in the Computing Division at Los Alamos National Laboratory. Dr. Kahaner is the author of two books and more than fifty research papers. He also edits a column on scientific applications of computers for the Society of Industrial and Applied Mathematics. He has had sabbatical appointments at the University of Michigan, Vienna, Turin, and the ETH in Zurich. His major research interests are in the development of algorithms and associated software. His programs for solution of differential equations, evaluation of integrals, random numbers, and others are used in many scientific computing laboratories. Since joining ONR's Tokyo office, Dr. Kahaner has written more than thirty reports on computing activities in Japan, which are circulated to hundreds of readers worldwide.

Toshi Minoura

Affiliation: Oregon State University, Dept. of Computer Science, Corvallis, Oregon 97331.

Telephone: 503-737-5574 Fax: 503-737-3014
Email: minoura@cs.orst.edu
Toshimi Minoura is Associate Professor of Computer Science at Oregon State University. He has conducted research on deadlock problems, true-copy token schemes, multi-version concurrency control, and reliable storage subsystems. He also supervised an application development using a 4GL system for the U.S. Environmental Protection Agency. He is currently working on an action-base system for manufacturing control, which is an active object system using transition rules. He was co-recipient of the 1988 Asahi Newspaper Invention Award for his contribution to a rule-based process control system. This technology is widely used to fully automate large power plants in Japan. Dr. Minoura received the B.S. and M.S. degrees from Tokyo University in 1968 and 1979 and the Ph.D. degree from Stanford University in 1980, all in Electrical Engineering.

Diane Smith

Affiliation: Xerox Advance Information Technology, 4 Cambridge Center, Cambridge, MA 02142.
Telephone: 617-499-4404 Fax: 617-499-4409
Email: dsmith@xait.xerox.com,
Dr. Diane Smith is the Director of Technology for Xerox Advanced Information Technology (XAIT). She is responsible for setting technology directions, the execution of projects, and technology transfer activities. She has held this position since 1979 when XAIT was a division of Computer Corporation of America (CCA). In 1988, she transitioned the CCA Advanced Information Technology (both staff and technology) into the corporate structure of Xerox. She has participated in the development of technology in the areas of object management, database design tools and methodologies. Her group developed an object-oriented data management program incorporating advances in spatial data handling, inferencing, and interoperability. She led the development of a database design workbench supporting requirements analysis, and conceptual, logical and physical design. Her product group is currently developing a workflow manager which supports the definition and execution of long-term processes using and developing complex objects. Prior to joining CCA, she taught in the Department of Computer Science at the University of Utah.

John Smith

Affiliation: Digital Equipment Corporation, 55 Northeastern Blvd., Nashua, NH 03062
Email: jmsmith@databs.dec.com,
Dr. John Miles Smith joined Digital Equipment in 1988 and is manager of the Database Tools group. He is responsible for defining the end-user and application development toolset layered on Digital's Database Management Systems. His technology interests include multimedia document management, engineering data management and the user environment. He was previously at Computer Corporation of America as Vice President of the Research and Systems Division. After joining CCA in 1979, he was responsible for the acquisition and execution of contract R&D projects in the areas of Distributed Database Systems, Heterogeneous Database Integration, CAD/CAM Data Management, Highly-Available Database Systems, and Expert Database Systems. He was on the faculty of the University of Utah from 1972 to 1978. He and his wife, Diane, built a research program in database technology with notable contributions in query processing, database design methodologies, data models, and database machines.

Gio Wiederhold

Affiliation: Stanford University, Dept. of Computer Science, Stanford, CA 94305-2140.
Telephone: 415-725-8363 Fax: 415-725-7411
Email: gio@cs.stanford.edu

Gio Wiederhold is a professor of Computer Science and Medicine (Research) at Stanford University. He worked in industry from 1958 until 1976, when he left industry to obtain a Ph.D. at the University of California in San Francisco. Since that time he has been on the Stanford faculty. He is active in the application and development of knowledge-based techniques to database management. He is the Editor-in-Chief for ACM's Transactions of Database Systems and Associate Editor of Springer-Verlag's M.D. Computing magazine. Wiederhold has written more than 200 publications in computing and medicine, including a widely used McGraw-Hill textbook on Database Design, and a 1987 book, File Organization for Database Design. A textbook on Medical Informatics, with Ted Shortliffe, was published in 1991. He has been chairman and program chairman of several conferences. He consults for many governmental and commercial enterprises, including the United Nations Development Program, the U.S. Department of Health and Human Services, various U.S. defense agencies, Silicon Valley innovators and companies in Japan and Europe. From 1991 to 1993 he is on research leave at DARPA, managing their programs in knowledge and databases.

B. Sites Visited and Contacts

6.2.1 Government

Ministry of International Trade and Industry (MITI)

Information Systems Development Division
Machinery and Information Industries Bureau
Address: 1-3-1 Kasumigaseki, Chiyoda-ku, Tokyo 100
Telephone: 3-3580-3922
People Visited: Mr. Hideo Setoya

Institute for New Generation Computer Technology (ICOT)

Address: Mita Kokusai Bldg, 1-1-28 Mita, Minato-ku, Tokyo 108
Telephone: 3-3456-3195
People Visited:
Dr. Kazuhiro Fuchi, Director, Research Center
Takashi Kurozumi, Deputy Director, Research Center
Kazumasa Yokota, Chief, Third Research Laboratory
Dr. Shumichi Uchida, Manager, Research Department

Information Technology Promotion Agency

Address: Shuwa-Shibakoen 3 - chome Bldg., 3-1-38, Shibakoen, Minato-ku, Tokyo 105
Telephone: +81-3-3433-2350 Fax: +81-3-3437-5386
People Visited: Dr. Akio Tojo, Managing Director

Japan Information Center of Science and Technology (JICST)

Address: 2-5-2 Nagatocho, Chiyoda-ku, Tokyo 100
Telephone: 3-581-6411
People Visited: Ms. Hisako Uchida

6.2.2 Industry

Fujitsu at INTAP

Address: 24 Daikyocho, Sjinjuku-ku, Tokyo 160
Telephone: 3-3358-2721
People Visited:
Mr. Shozo Tanaka, Senior Manager
Mr. Fumiaki Tsuboi, Manager, 1st Engineering Department
Mr. Yoshio Izumida, Section Manager, Artificial Intelligence Laboratory
Mr. Hiroshi Ishikawa, Researcher, Knowledge-Based Engineering Section

Hitachi

Address: 1099 Ohzenji, Asu-ku, Kawasaki 215
Telephone: 044-966-9111
People Visited:
Takashige Kubo, Deputy General Manager, Systems Development Laboratory
Seiichi Yoshizumi, Department Manager, 3rd Research Department, SDL

Hiromichi Fujisawa, Senior Researcher, Central Research Lab
Shunichi Torii, Senior Researcher, Advanced Processor Research Department, CRL
Kazuhiro Satoh, Senior Researcher, 3rd Research Department, SDL
Shigeru Yoneda, Senior Researcher, R/D Planning Office, SDL
Hiroyuki Kitajima, Senior Researcher, 3rd Research Department, SDL
Hidefumi Kondo, Senior Researcher, 5th Department, SDL
Kazuo Masai, Senior Engineer, Software Development Center

IBM Japan
Address: Tokyo Research Laboratory 5-19, Sanbancho, Chiyoda-ku, Tokyo 102
Telephone: +81-3-3288-8280
People Visited:
Dr. Jung-Kook Hong, Manager, Applied Image Processing Systems
Mr. Kohichi Kajitani, Researcher
Dr. Norihisa Suzuki, Director

Kozo Keikaku Engineering
Address: 4-38-13, Honcho Nakano-ku, Tokyo
Telephone: 3382-6581
People Visited:
Hiroshi Ando, Senior Researcher
Terumi Hanmyo, Chief of Research and Development Department

Mitsubishi Electric Corp.
Address: 5-1-1 Ofuna, Kamakura, Kanagawa 247
People Visited:
Mr. Junichi Shibayama
Mr. Satoshi Tanaka
Mr. Koji Wakimoto

NEC
Address: Ugarashi Bldg, 2-11-5 Shibaura, Minato-ku, Tokyo 108
Telephone: 3-5467-1080
People Visited:
Kiichi Fujino, Vice President, C & C Software Development Group
Masao Matsumoto, Associate Chief Engineer, C & C Software Development Group
Kunihiro Sugawara, Assistant General Manager, EDP Systems Engineering Division
Isao Kamoi, Engineering Manager, Database Development Department
Ryuichi Ogawa, Supervisor, Applied Information Technology Research Laboratory
Kyoji Kawagoe, Research Manager, Terminal Systems Research Laboratory
Yoichi Miyashita, Engineering Manager, Software Architecture Department
Shuji Nakata, Systems Manager, 2nd Systems Department
Kunitoshi Tsuruoka, Research Manager, Basic Technologies Research Laboratory

NTT Processing Laboratories
Address: 1-1-7 Uchisaiwaicho, Chiyoda-ku, Tokyo 100

Telephone: 0468-59-2694
People Visited:
Dr. Masato Haihara, Director
Mr. Masaharu Araki, R&D Public Relations

NTT Data Services
Address: 1-26-5 Toranomon, Minato-ku, Tokyo 105
Telephone: 3-3580-4535
People Visited:
Yutoki Sasaki, International Affairs
Tsutomu Shibata
Akikazu Ida
Kouichi Suzuki
Dr. Y. Tachibana, Senior Vice president

Oki Electric Industry Co. Ltd.
Address: 4-11-22 Shibaura, Minato-ku, Tokyo 108
Telephone: 3-3456-3219
People Visited:
Mr. Yunosuke Haga, General Manager, Systems Laboratory
Mr. Suguru Kawakami, Computer Systems R & D Division
Dr. Atsushi Ohori, Senior Researcher, Kansai Laboratory
Dr. Nobuyoshi Miyazaki, Senior Researcher, Manager, Systems Laboratory

Ricoh
Address: Tomin-Nissei-Kasugacho Bldg., 1-1-17 Koishikawa, Bunkyo-ku, Tokyo 112
Telephone: 3-3815-7261
People Visited: Dr. Hideko Kunii, General Manager, Software Division

Software AG, Far East
Address: 1-6-1 Nishi Shinjuku, Shinjuku-ku, Tokyo 160
Telephone: 3-3340-2467
People Visited: Mr. Y. Ishii, President

Toshiba
Address: 1 Komukai Toshibacho, Kawasaki 215
Telephone: 44-549-2078
People Visited: Dr. Kazuo Narita

Universities

Hokkaido University
Address: 8 Nishi, 13 Kita-Ku, Sapporo 060
Telephone: 011-716-2111
People Visited: Professor Yuzuru Tanaka

Kobe University
Address: Rokkodai Nada, Kobe 657
Telephone: 78-881-1212
People Visited:
Professor Katsumi Tanaka and Professor Shojiro Nishio

Kyoto University
Address: Sakyoku, Kyoto
People Visited: Professor Yahiko Kambayashi

Ryukoku University
Address: Seta, Otsu 520-21
Telephone: 0775-43-5111
People Visited: Professor Toshiyuki Sakai

Tokyo University
Address: 7-3-1 Hongo, Bunkyo-ku, Tokyo 113
Telephone: 3-3816-1783
People Visited: Professor T.L. Kunii

Aichi Institute of Technology, Toyota, Nagoya
Address: 1247 Yachigusa, Yagusa-cho, Toyota, Aichi
Telephone: 0565-48-8121
People Visited: Professor S. Kito

Tsukuba University
Address: Tennodai, Tsukuba, Ibaraki 305
Telephone: 0298-53-5532
People Visited: Professor H. Kitagawa and Professor Y. Fujiwara

University of Library and Information Science
Address: 1-2 Kasuga, Tsukuba, Ibaraki 305
Telephone: 0298-52-0511
People Visited: Prof. Yoshifumi Masunaga

Kyushu University
Address: 6-10-1 Hakozaki, Fukuoka 812
Telephone: 92-641-1101
People Visited:
Dr. Toshihisa Takagi, Associate Professor
Mr. Susumu Goto, Ph.D. Student
Dr. Hirofumi Amano, Research Associate
Dr. Mohammed E. El-Sharkawi, Research Associate

6.2.3 Research Laboratories

Interactive Interface Systems Laboratory at ETL

Address: 1-1-4 Umezono, Tsukuba Science City, Ibaraki 305
Telephone: 0298-54-5413
People Visited:
Dr. Akio Tojo
Dr. Toshikazu Kato

Research Center for Advanced Science and Technology (RCAST)

Address: 4-6-1 Komaba, Meguro-ku, Tokyo 153
Telephone: 3-481-4483
People Visited:
Setsuo Osuga, Professor
Koichi Hori, Associate Professor
Hirayuki Yamanouchi, Research Assistant

6.2.4 Inter-University Research Institutes

National Museum of Ethnology

Address: 10-1 Banpaku, Senri, Suita, Osaka 565
People Visited: Dr. Shigeharu Sugita

National Center for Science Information Systems (NACSIS)

Address 3-29-1, Otsuka, Bunkyo-ku, Tokyo
Telephone: 3-3942-2351
People Visited:
Prof. Akira Miyazawa
Prof. Hisao Yamada, Director, R&D Department
Prof. Teruo Koyama
Prof. Masamitsu Negishi
Prof. Kimio Ohno, General Coordinator
Prof. Hiromichi Hashizume

Topics for Recommendations

For U.S. Dept. of Commerce.

1. Japanese firms that use U.S. DBMS software should cooperate with the original U.S. DB vendors, specifically as the Japanese resellers provide enhancements to

1 support Japanese and other Asian (Chinese, Korean) charactersets

2 facilitate image storage and image transmission.

3 develop other multimedia support features

The objective is to prevent U.S. firms from being shut out of this growing market.

2. Participation of U.S. scientists in Japan-based standards-setting activities should be supported. Now many potential participants are excluded because of high travel and lodging costs. The price for industry is high, and the payoff hard to foresee. For academics participation even the U.S. is difficult, to participate in Japan is completely out of bounds. We recommend focusing on areas where Japanese products are likely to set standards, as in the areas of multimedia storage and transmission, and in extended charactersets.

3. We need to enhance the use of commercial database services to disseminate information gathered in Japan from Japanese sources. There is a current impedance mismatch:

a. The U.S. government does not want to spend its resources on making this a governmental function.

b. the Japanese model makes this function governmental or quasi-governmental. Its institutions then want to deal with NSF etc.

If need be, a subcontractor relationship of NSF with such vendors might be established.

- To reduce the user cost for distinct access to U.S. and Japanese information we could have a federated database project where users can access in parallel U.S. and Japanese files. A possible topic would be Japanese patent files. There may be others?

  Such work would technically be based on the client/server model for information retrieval following the Z39.50 information retrieval standard (or the ISO counterpart, or some other suitable standard) as the communication protocol between the client and the server. It would give both parties useful experience with the standard.
- Allow a Japanese language searcher to access an English language server. Even if the database content is not translated, at least the interface could be treated, and perhaps part of the database content could be handled, and most Japanese searchers can deal with technical English text.
- Allow an English language searchers to access a Japanese and vice-versa. Here it is necessary that English commands be understood and English messages be produced. Some of the database content should be available in translation, so part of the database content (titles, abstracts, citations) could be handled.
- Such an enterprise would benefit from more general multi-lingual access capability, or, at least, provide a basis to stimulate joint research in this area.

4. An alternative to distributed access is to at least have a central repository in the where all Japanese publications are available to researchers.

5. Increase the number of English-language papers from Japanese sources. This could be accomplished by

- Increasing the rate of translation of papers from Japanese to English.
- Encouraging also more Japanese publication in English language media.

There is now already a Japanese printed English language publication: the *IEICE Transactions*. It is published by the Institute of Electronics, Information and Communication Engineers (IEICE). It has non-Japanese associate editors. The IEICE is roughly equivalent to the IEEE in U.S. The monthly *IEICE Transactions* cover

1 Fundamentals of Electronics, Communication and Computer Sciences,
2 Communications,
3 Electronics, as well as
4 Information and Systems.

This Japan originated English publication is not well known now. To encourage its use, and indirectly Japanese authorship in it

- Have computing topic papers regularly reviewed in the *ACM Computing Reviews* (I have never seen one reviewed there).
- Have papers on other topics reviewed in appropriate speciality journals.
- The *IEICE Transactions* also contains English abstracts of Japanese-language papers published in the *IEICE Computer* and the the *IEICE Information Processing* journals. These two journals cover the IEICE areas of Information and Systems. This volume of material is modest and might be a candidate for reprinting in selected IEEE Computer Society newsletters, as the *IEEE CS Data Engineering Bulletin*. The *IEICE Transactions* also lists the titles in English of the unrefereed workshop reports from the 47 technical groups of the IEICE.

6. Not enough takers have been found for postdoctoral positions in Japan, even though we feel that participation could help in mutual understanding, establishing long-term collegial relationships, and, in the end, improve the tracking in depth of Japanese research and practice.
We believe that the potential still exists for having postdocs in Japan, partially because there is now a supply, due to the slowdown in hiring. Another motivation would be access to Japanese multimedia technology.

- To avoid the disincentive of participants losing contact and inertia here, we recommend to augment such postdocs with a funded period for readjustment in the U.S., say 50% time or 1 year for a two year stay in Japan.

7. Set up or expand a branch of a U.S. university (or University Consortium), for one year/semester stays by U.S. students (including undergraduates) focusing on Japanese Technology and staffed by both U.S. and and Japanese scientists. Many of existing efforts focus on cultural aspects, and are not attractive to U.S. technical students, although Stanford does have a fledgling effort in Kyoto, oriented though to graduate students.

Use these facilities to hold joint workshops (say, every six months) where results are presented. That would avoid the need U.S. computer scientists go to Japan for an extended stay, but could provide better continuity than ad hoc visits.

An open question: What about Japanese Labs in the U.S., staffed by U.S. scientists. Should these cooperate in some way?

8. We should define some joint projects in the light of the 6th-generation project goals. The objectives are vague, and a call-for-joint-participation, best mediated through the Department of Commerce or a designated contractor (NRI?), may help.
A specific pre-commercialization research project might have the goal of exploiting any cross-fertilization possible between word recognition in Kanji text and continuous speech recognition where the U.S. could contribute. There might be a similarity in the way syntactic and semantic clues need to be used to optimize the pattern recognition, given that Kanji is continuous text. Blackboard technology, as developed at CMU and Stanford could provide a foundation.
Such work could support intelligent text retrieval, and fall into several categories of work considered for the 6th generation.

These following recommendations are for changes in U.S. practices, based on our observations:

A1. We found the MITI-sponsored interactions on joint stable, committees involving industry, researchers, and government very effective for rapid technology transfer. We recommend that such committees, with industry participants, be established here in concert with major projects. Specifically, we consider that the use of study sections (the NIH model) for new initiatives for new major and novel proposals submitted to NSF, so that pros and cons can be discussed, and not rejected for lack of understanding by some reviewers or program managers. Industrial managers should participate. This will indirectly enhance U.S. dissemination of new technology.

A2. Support the establishment of a technology lobbying group in the U.S., addressing the area of DB technology, similar to MITI's Database Promotion Center. For instance

- An organizations focused/interested in DB technology working with the government trying to influence how research money is spent?
- A consortium, possibly government funded, focusing on advancing the state of the art in DB systems?

Robert Kahn's (formerly of DARPA ISTO) institute for National Research Initiatives (NRI) can be seen as a effort in that direction, oriented towards electronic networking.

A corporate entry is the Computer Systems Policy Project, a group of companies interested in maintaining a U.S. leadership position in key technical areas. That project decides if some technical areas need special attention and effort to either maintain or achieve technical leadership.