Chapter 2
The Development of Medical Databases

Since the early 1900s physicians have followed the teachings of the famed clinician, W. Osler, to study and learn from their patients and from the medical records of their patients, in order to improve their knowledge of diseases. In the 2000s, as in the 1900s, physicians continue to initiate this learning process by taking a history of the patient's medical problems, performing a physical examination of the patient, and then recording the history and physical examination findings in the patient's medical record. To confirm a preliminary diagnosis and to rule-out other possible diagnoses, physicians refer the patients for selected tests and procedures that usually involve the clinical laboratory, radiology, and other clinical-support services. After reviewing the information received from these services, physicians usually arrive at a more certain diagnosis, and then prescribe appropriate treatment. For an unusual or a complex medical problem, physicians may refer the patient to appropriate clinical specialists, and may also review evidence-based reports of appropriate therapies by consulting relevant medical literature and bibliographic databases.

2.1 The Origins of Medical Databases

Lindberg (1979) described the degrees of difficulty in the development of medical innovations in the grades of their complexity: (1) the easiest was the automation of a simple function such as providing a patient's billing for services; (2) more difficult was the automation of a more complex function such as collecting and storing a patient's medical history; (3) very difficult was constructing a very complex function such as a medical database; and (4) the most difficult was developing the highly complex medical information and database-management system for a hospital, as Starr (1982) had aptly ranked the hospital to be the most complex organizational structure created by man.

Databases were defined by Frawley et al. (1992) as logically integrated collections of data in one or more computer files, and organized to facilitate the efficient storage, change, query, and retrieval of contained relevant information to meet the

M.F. Collen, *Computer Medical Databases*, Health Informatics,
DOI 10.1007/978-0-85729-962-8_2, © Springer Verlag London Limited 2012

needs of its users. Frawley estimated that the amount of information generated in the world doubled every 20 months, and that the size and number of computer databases increased even faster. *Clinical repositories* was the term proposed by Johnson (1996) as more accurately representing a shared resource of patient data that was collected for the purpose of supporting clinical care. Johnson advised that a large-scale, clinical repository required: (a) a data model to define its functional requirements and to produce a formal description, (b) a conceptual schema of all the data generated in the enterprise and how it was all related, and (c) a database structural design to define its technical requirements. Since a medical database usually operated within a medical database-management system, the database needed to be compatible with the information system of the enterprise of which it was a part; and it also needed to be operationally and structurally independent of all subsystems and applications programs. The evolution, design, implementation, and management of computer-stored databases were described in some detail by Connolly and Begg (1999), Collen (1986, 1990, 1994, 1995); and also by Coltri (2006) who considered computer-stored databases to be one of the most important developments in software engineering.

Database-management systems soon replaced the earlier file-based systems that often stored the same data in multiple files, and where it could be more difficult to retrieve and coordinate a patient's data. A database-management system was defined by Blum (1986a, b, c) as software consisting of a collection of procedures and programs with the requirements for: (1) entering, storing, retrieving, organizing, updating, and manipulating all of the data within its database; (2) managing the utilization and maintenance of the database; (3) including a metadatabase to define application-specific views of the database; (4) entering data only once, even though the same data might be stored in other subsystems; (5) retrieving, transferring, and communicating needed data in a usable format, and having the ability to create inverted files indexed by key terms; (6) maintaining the integrity, security, and required level of confidentiality of its patients' data; and (7) fulfilling all management, legal, accounting, and economic requirements.

In the 1950s with the development of computers, physicians began to bring their work in batches to a central computer to be processed. Patient-care data were initially collected, entered, and merged into computer files that were stored on magnetic tape, and a file-management system was designed to enter, store, and retrieve the data. In the 1960s time-shared, mainframe computers that communicated by telephone lines to remote data-entry terminals and printers, allowed many users to process their data concurrently, and also provided a relatively acceptable turnaround time for data services. Patients' data were initially stored in computer databases on magnetic tape; but were soon moved to storage on random-access, magnetic disc drives; and were then better organized in a manner more suitable for query and retrieval of the data. However, at that time the high costs for computer storage greatly limited database capacities. In the 1970s as clinical support subsystems evolved for the clinical laboratory, radiology, pharmacy, and for other clinical services, most developed their own separate databases. With the emergence of random-access disc storage, subsystem databases could be more readily merged into larger databases and then needed an integrating database-management system. The retrieval of subsets of selected data from various databases required some reorganization of the stored data, and also needed an index to the locations of the various data subsets. Attempts were made to design more efficient databases to make them independent of their applications and subsystems, so that a well-designed database could process almost any type of data presented to it. Terdiman (1982) credited the development of microcomputer technology in the 1970s with many of the subsequent advances in database-management systems.

In the 1980s microcomputers and minicomputers were increasingly used for small database systems. As storage technology continued to become more efficient, and larger and cheaper storage devices became available, then computer-based registries expanded their storage capacity for larger amounts of data and were then generally referred to as databases. When huge storage capacity became available at a relatively low-cost, very large collections of data were then often referred to as data warehouses. Bryan (1988) called 1988 the "year of the database"; and he reported that more than 20 new or improved database-management systems became available in that year. In 1989 the total number of computer-stored databases in the world was estimated to be about five-million; and although most of the databases were considered to be relatively small, some were huge as was the 1990 U.S. census database comprising a million-million bytes of data (Frawley et al. 1992). Prior to the 1990s most physicians documented their patient-care activities by handwriting in paper-based charts. Surgeons and pathologists usually dictated their reports describing their procedures and findings; and medical secretaries then transcribed their dictations. With the increasing access to larger computers in the 1990s, medical center-based physicians began entering a patient's data directly into the patient's electronic medical record (EMR) using keyboard terminals and clinical workstations. Dedicated computers became database servers to store and integrate multiple databases; and to be able to add new data without disrupting the rest of the system. In the 2000s EMRs became more common; and new advances in informatics technology resulted in more efficient data management of expanding, multi-media, patient-care databases (Coltri 2006). More details on the origins and the development of medical databases can be found in Blum (1983, 1986a), Blum and Duncan (1990), Collen (1986, 1994, 1995), Coltri (2006), Duke and Bowers (2006), Campell-Kelly (2009).

2.2 Requirements and Structural Designs for Medical Databases

Data-modeling designs to provide the conceptual schema that represented the information in clinical repositories were advocated by Johnson (1996) to be as important for large medical databases as were their structural designs. He defined the conceptual schema for patient care as a representation of all of the data types required to manage the health-care process, whether using a hierarchical, a relational, or an object-oriented structural database design, or a combination of database structural designs. He advised that the structural design of a database needed to be able to provide rapid retrieval of data for individual patients, and to have the capability to adapt to changing information needs of growth and new technology; yet he emphasized that the primary purpose of the database structural design was to implement the conceptual schema. To properly build a database, Johnson (1996) proposed that it was necessary to first develop a model of the database that defined its functional requirements, its technical requirements, and its structural design. The database model needed to produce a formal description, a conceptual schema of all the data generated in the enterprise, and how all of the data were related. Thus the users of a medical database needed to define its functional requirements as to exactly what they wanted the database and its database-management system to do. Since a medical database usually operated within a larger medical-information system, the functional requirements of the medical database needed to be compatible with those of the medical enterprise of which it was a part. Whether a medical database served as the primary electronic medical record (EMR), or served as a secondary medical database, such as a clinical research database with its data derived from the EMR, both had some similar basic functional requirements. Davis and Terdiman (1974) recommended that as a minimum, the major goals of a medical database should be: (1) to maintain readily accessible all of the relevant data for each patient served; and (2) to provide a resource for the systematic retrieval of all relevant data from all patients' records for any desired primary purpose (see Sect. 4.1), or for a secondary administrative or a research purpose (see Sect. 6.1).

The structural design of medical databases was substantially developed by Wiederhold (1981, 1982, 1983, 1984), Wiederhold et al. (1975, 1987) at Stanford University. Wiederhold emphasized that the effectiveness of a database depended on its relevance to its organizational purposes; that it had to serve as a resource to the enterprise which had collected the data; and that a database-management system was needed to control, store, process, and retrieve the data. He advised that when using very large databases it was helpful to apply automated methods for the acquisition and retrieval of the desired information. Several database structural designs evolved as new medical and informatics technologies were developed to meet the various users' requirements.

Hierarchical tree-structured databases were considered by Coltri (2006) to be the simplest and earliest structural design used for medical databases. In a hierarchical designed database the data was organized in what was usually described as a "parent–child" relationship, where each "parent" could have many "children", but each "child" had only one "parent". Hierarchical data subclasses with inheritance of attributes could also appear in other designed databases, such as in relational and in object-oriented databases. A. Coltri reported that the best known early example of a hierarchical structured, medical database was the one developed in the 1960s by Barnett (1974), Barnett et al. (1981) and associates (Greenes et al. 1969; Grossman et al. 1973). Their Massachusetts General Hospital Utility Multi-Programming System (MUMPS) was designed for building and managing dynamic hierarchical databases with interactive computing applications and online transactional processing. MUMPS provided a good structure for medical databases with all their complexity, since its hierarchical structure functions as a fundamental persistent saved entity that enables a more complex design than does a simple relational table with rows and columns; and this greater complexity matches well with the needs of a medical record database. In the 1980s both the Department of Defense and the Veterans Hospitals began installing their MUMPS-based medical information systems; and in the 2000s the popular Epic medical information systems was also Mumps-based. Another example of an early hierarchical-structured medical database was that developed in the 1960s by Davis (1970, 1973), Davis et al. (1968), Davis and Terdiman (1974), Terdiman (1982) and associates at Kaiser Permanente (KP) in Oakland, California, to store patients' electronic medical records. The design of each KP patient's record included 12 levels of storage that allowed direct access by the patient's unique medical record number to each of the patient's computer-defined visits, which were subdivided into medical meaningful parts ("tree branches") such as laboratory data, diagnoses, and clinical services. The database was designed to store all patients' data received; and it also contained program-generated data related to the tree structure of the record that included data as to the level of the tree branch and of the length of the record, that provided a trail through the record.

Relational databases and their database-management systems were developed in the 1960s for large shared databases by Codd (1970, 1972, 1979), Codd et al. (1993) while at the IBM Research Center in San Jose. Codd required that all data in a relational database be expressed in the form of two-dimensional tables with uniquely labeled rows and columns. Every data element was logically accessible through the use of the names of its table and its column; and data transformations resulted from following defined logical rules. In a relational database the data were organized into files or tables of fixed-length records; each record was an ordered list of values, one value for each field. Information about each field's name and potential values was maintained in a separate metadatabase. Because of its simplicity, by the 1980s the relational database design had become dominant in industry and in medicine. Miller et al. (1983) at the University of Pittsburgh, Pennsylvania, described using a commercial relational database-management system, called System 1022, that provided its own programming language (1022 DPL) and permitted clinical data from large groups of patients to be entered, stored, queried, and analyzed for clinical studies. Friedman et al. (1990) and associates at Columbia University, noted that the typical relational design for a patient database could have a serious impact on query performance, because a patient's data was typically scattered over many different tables, so a query language needed to be added. Also noted by Deshpande et al. (2003) was that medical data parameters were often time-stamped, such as when representing the beginning and the end of a clinical event; and also when in a relational database special approaches were required to query various columns for desired temporal data. Structured Query Language (SQL) was developed in the 1970s by D. Chamberlin and R. Boyce at the International Business Machines (IBM) to construct, manage, and query relational databases (VanName and Catchings 1989); and SQL soon became the standard language used for programming relational databases.

In 1979 a commercial relational database named ORACLE became available from the ORACLE Corporation. In the 1980s Ashton-Tate developed dBASE for microcomputers (Connolly and Begg 1999). Johnson (1999) described an extension of SQL for data-warehouses that enabled analysts to designate groups of rows that could be manipulated and aggregated into large groups of data, and then be analyzed in a variety of ways to solve a number of analytic problems.

Marrs and Kahn (1995) and M. Kahn at Washington University, St. Louis, described developing a distributed, relational database-management system across multiple sites comprising a single enterprise, when they extended their clinical repository for Barnes Hospital to include data from Jewish Hospital in the BJC Health System that included 15 hospitals and other health care facilities. After considering alternative approaches, they chose to add the data from Jewish Hospital to their repository, and implemented required changes to accommodate mapping the data from other facilities into their database, and to adjust for differences in syntax and semantics in patient identifiers, medication formulary codes, diagnoses codes, and other information in their patients' records. As relational databases grew in size and developed multiple dimensions, some commercial search-and-query programs for very large relational databases became available, led by Online Analytic Processing (OLAP), that provided answers to analytic queries that were multi-dimensional and that used relational databases. OLAP generally stored data in a relational structured design, and used aggregations of data built from a fact-table according to specified dimensions. Relational database structures were considered to be multi-dimensional when they contained multiple attributes, such as time periods, locations, product codes, and other attributes that could be defined in advance and aggregated in hierarchies. The combinations of all possible aggregations in the database were expected to be able to provide answers to every query that could be anticipated of the stored data (Codd et al. 1993). Connolly and Begg (1999) described a way of visualizing a multi-dimensional database by beginning with a flat, two-dimensional table of data; then adding another dimension to form a three-dimensional cube of data called a "hypercube"; and then adding cubes of data within cubes of data, with each side of each cube being called a "dimension", with the result being a multi-dimensional database. Pendse (1998, 2008) described in some detail the history of OLAP, and credited the publication in 1962 by K. Iverson of A Programming Language (APL) as the first mathematically defined, multi-dimensional language for processing multi-dimensional variables. Multi-dimensional analyses then became the basis for several versions of OLAP that were developed in the 1970s and 1980s by IBM and others; and in 1999 the Analyst module was available in COGNOS that was subsequently acquired by IBM. By the year 2000 new OLAP derivatives were in use by IBM, Microsoft, Oracle, and others.

Object-oriented databases were developed in the 1970s at the Xerox Palo Alto Research Center (PARC), and used the programming language Smalltalk (Robson 1981). Object-oriented databases attempted to bring the database programming and the applications programming closer together; and treated the database as a modular collection of component data-items called *objects*. Objects were members of an "entity" that belonged to types or classes of data with their own data and programming codes; and objects incorporated not only data but also descriptions of their behavior and of their relationships to other objects. Whereas other database designs separately represented information and its manipulation, in an object-oriented system the object represented both. Objects used "concepts" such as entities, attributes, and relationships; and objects could be members of an entity that belonged to types or classes with their own data and programming codes. Objects had an independent existence; and could be persons, activities, or observations; and were sufficiently independent to be copied into other programs. Attributes were properties that described aspects of objects; and relationships described the association between objects (Dawson 1989). Connolly and Begg (1999) described some relational variances for an object-oriented database in order to use SQL.

Barsalou and Wiederhold (1989) described their PENGUIN project that applied, a three-layered architecture to an object-oriented database that defined the object-based data as a layer of data on top of a relational database-management system, with a hypertext interface between the object-oriented and the relational databases that provided conceptual integration without physical integration. Their workstations were Apple personal computers; and they used Apple's HyperCard program for their Macintosh computer that defined and manipulated "stacks" of data corresponding to a relational-database structure, with one field for each attribute, written in the Macintosh HyperTalk language that allowed querying visual images that moved through a hypertext document.

Entity-attribute-value (EAV) databases were developed to help manage the highly heterogeneous data within medical databases, where over several years of medical care a single patient could accumulate thousands of relevant descriptive parameters, some of which might need, from time-to-time, to be readily accessible from a large clinical database that contained multiple relational tables. Dinu and Nadkarni (2007), Nadkarni and Cheung (1995), Nadkarni et al. (1998), Nadkarni et al. (1999), Nadkarni et al. (2000), Nadkarni and Marenco (2001), Brandt et al. (2002) described an EAV database as an alternative to conventional relational-database modeling where diverse types of data from different medical domains were generated by different groups of users. The term, EAV database, was generally applied when a significant proportion of the data was modeled as EAV even though some tables could be traditional relational tables. Conceptually, an EAV design used a database table with three columns: (1) 'Entity', that contained data such as the patient identification, with a time-stamp of the date-and-time of the beginning and end of each clinical event; (2) 'Attribute', that identified the event, such a laboratory test, or showed a pointer to a separate attribute table; and (3) 'Value' column, that contained the value of the attribute (such as the result of a laboratory test). A metadatabase was usually added to help provide definitions of terms, keys to related tables, and logical connections for data presentation, interactive validation, data extraction, and for ad-hoc query. Tuck et al. (2002) described some alternate methods for mapping object-oriented software systems to relational databases by using an EAV approach. Chen et al. (2000) evaluated the performance of an EAV design; and concluded that the advantage of the EAV design was in supporting generic browsing among many tables of data, as when following changes in a clinical parameter over many periods of time; and that it also helped to provide schema stability as knowledge evolved and the metadata needed to change. However, attribute-centered queries were somewhat less efficient when using EAV designed databases because of the large numbers of data tables with many more rows than when using conventional relational databases. Some early users of variations of the EAV model were: McDonald et al. (1977a, b, 1982, 1988), McDonald and Hammond (1989) in the Regenstrief Medical Record (RMR) system; Warner et al. (1972, 1974), Warner (1990), Pryor et al. (1983) in the HELP system; Stead and Hammond (1988), Stead et al. (1992), Hammond et al. (1977), Pryor et al. (1982) in the TMR system; and Friedman et al. (1990), Hripcsak et al. (1996) at Columbia University, and the EAV model underlies the architecture of i2b2 (see Sect. 3.3).

Database-management systems in large medical centers began to evolve in the 1950s, and were designed as either: (a) clusters of computers tightly coupled to a central large mainframe computer, or (b) loosely-coupled in a distributed database system (London 1985). As information communication systems grew to service large medical centers, with all of their inpatient and outpatient clinical departments that included internal medicine, surgery, pediatrics, obstetrics, gynecology, pathology, clinical laboratory, radiology, and others, with their great variety of medical applications, all of these required a complex, computer-based, information system that communicated data to-and-from all of the various clinical subsystems. As databases grew larger and often contained redundant storage, Coltri (2006) noted that although a single, structural database model could initially allow for simpler coordination, operation, and reporting; yet as clinical databases enlarged and became more complex with many functional relationships and subsystem components, with frequent changes in their data content, then the ability to restructure a single large database in order to satisfy the important need for efficient querying of its data content became increasingly difficult.

Federated databases developed that could store large volumes of aggregated data in multiple partitions or as functional-oriented databases that were logically interconnected. They were directly accessible to-and-from multiple applications, and allowed multiple users to simultaneously access and query data in the various databases (Coltri 2006). *Data warehouses* was the term applied to large, extended, central databases that collected and managed data from several different databases; and they were capable of servicing the ever-increasing volume of patient data that were collected from the ever-changing and expanding medical technologies. As data warehouses further enlarged they often developed partitions and data-marts for specialized sub-sets of the data warehouse in order to better serve users with different functional needs (Connolly and Begg 1999). When data warehouses were found to satisfy the needs of different users and efficiently query large collections of data, this led to the development of online analytical processing (OLAP), and of translational data processing between multiple data warehouses.

Translational databases evolved in the late 1990s with more advanced designs of database-management systems to: (a) optimize the translation, transformation, linkage, exchange, and integration of the increasingly voluminous medical information that was becoming accessible from many large databases in multiple institutions that were located worldwide, by using wide-area-networks, the Internet, and the World Wide Web; (b) provide access to high-performance, super-computing resources; (c) facilitate the concurrent query, analyses, and applications of large amounts of data by multi-disciplinary teams; (d) encourage knowledge discovery and data mining, and support the transfer of new evidence-based knowledge into patient care; and (e) to advance the use of biomedical computational methods. Since most data warehouses had been developed with standard database-management system designs that often employed their own legacy and data-encoding standards, it usually required some reorganization and modification of their source data to be compatible with the data that was transferred from other different data warehouses and then be merged into a single database schema; so it became necessary to develop some translational informatics software.

2.3 Databases and Communication Networks

Distributed database systems evolved in the 1970s with the introduction of low-cost minicomputers and efficient communication networks that brought computers closer to the users. In a distributed database system with a cluster of specialized subsystem databases, each subsystem collected and stored in its separate database the data it generated; and a communications network provided linkages for data entry to, and retrieval from, an integrating central database, and also to other subsystem databases as needed. As each specialized clinical service developed its individual database to satisfy its own specific functional and technical requirements, this usually resulted in the need for an overall integrating database-management system that could better service the very complex organizational structure of a large hospital. This allowed physicians to use clinical workstations connected to client–server minicomputers connected in a local-area-network that linked the entire hospital. Patient data could be generated and used at the local sites, and collected from all of the distributed subsystem databases, and integrated in a central, computer-based patient record (Friedman et al. 1990; Collen 1995). However, since the computers were often made by different manufacturers that used different software, this introduced a major problem when interchanging data between differently designed computer-database systems. This stimulated the evolution of specialized communications computers and networks for the distribution of data. Computers began to be linked together, usually connected to a central mainframe computer from which data could be downloaded to the smaller computers; and this changed the requirements and the designs of database-management systems. Wess (1978) noted that the design and implementation of a distributed-database system was more complex and demanding than that for a simple networked, data-communication system. By the late 1970s a variety of forms of networks for distributed-database systems began to appear, either linked together or connected to a central mainframe computer from which data could be communicated to-and-from the distributed smaller computers. In 1979 Walters (1979) at the University of California, Davis, began to link microcomputers with their databases to a remote, large host computer using MUMPS-based software. Blois et al. (1971, 1974) advocated using a communications computer-processor that would perform code conversion, and provide a high-speed communicating link to each distributed computer. In 1971 Blois initiated the first distributed database system for the medical facilities at the University of California, San Francisco (UCSF) Hospital. He used a separate, dedicated, communications minicomputer to connect computers from several different vendors, and established the first local-area network (LAN) for medical data communications. Blois separated the functions for communications from those for data processing, since each subsystem had its own requirements for data input, data processing, and data communications. After developing modular subsystems that could stand alone, he linked them in a communications network using specific standards adopted at their onset. His distributed database-management system required a reliable high-bandwidth, communications computer to perform communications code conversion, and also required a high-speed link to each subsystem computer. Wasserman (1977, 1986) while associated with Blois, proposed that a distributed database system should be capable of functioning at all levels of data acquisition, data manipulation, data retrieval, and data communications for a variety of applications; and he advocated that distributed medical databases needed to support an interactive information system with advanced software design, and with clinical work-stations.

Zeichner et al. (1979) at Mitre Corporation and Tolchin et al. (1980) at Johns Hopkins University described their distributed database system that contained a variety of different independent minicomputers. They used microcomputer-based, interface-units between each network minicomputer-processor and the communications bus. Data exchange used a standard set of protocols between network units, so each new or modified application or device could interact with its communications bus. In 1980 the Johns Hopkins group implemented a fiber-optic, local-area-network to integrate several subsystems built by three different manufacturers, each with a different operating system. They used microprocessor, network-integrating units to perform the conversions of communications codes needed to exchange data (Tolchin et al. 1981a; Tolchin and Stewart 1981). In 1985 they expanded their distributed clinical-information systems, all linked by Ethernet technology that supported 10-megabit-per-second data rates on a shared coaxial-cable medium which was logically a broadcast bus (Tolchin et al. 1985a, b). Kuzmak et al. (1987) described their addition of a central, clinical-results database to contain all of the reports for the clinical laboratory, radiology, and surgical pathology; and it was networked to permit the viewing of patients' reports from any terminal, any personal computer, or workstation in their hospital. Tolchin et al. (1982) at the Johns Hopkins University and D. Simborg (1984) at the University of California in San Francisco (UCSF), made a significant contribution to networking technology that reduced the problem of interfacing multiple incompatible computers, when they implemented at the UCSF medical center a fiber-optic, local-area network that integrated four different minicomputers, by using a fifth host-computer that was interfaced to the network to provide a monitoring service for performance analysis. Hammond et al. (1985) and associates at Duke University reported implementing an

Ethernet local-area-network for three types of computers connecting their clinical laboratory system to their central "The Medical Record" (TMR) database.

Network models were one of the earliest organizational structures used for clusters of computers with distributed databases; and they displayed pointers to link various data sets. Since the same data could reside in more than one data base, it required a communications network to link such data. This led in 1971 to a Conference on Data Systems Languages (CODASL) that advocated a variance of the network model in a hierarchical form of database with a tree-like branching structure that, at the start of the database, defined connections between files (Taylor and Frank 1976).

Communications standards for both the communications networks and for their transmission of data became essential requirements for the exchange of data between different computer systems. In the late 1970s the International Standards Organization (ISO) developed an important model and reference base for network systems that specified seven layers for the exchange of data between computers, with each layer corresponding to the same layer in the other computers. ISO layer one, the physical layer, included interface hardware devices, modems, and communication lines, and the software driver for each communication device that activated and deactivated the electrical and mechanical transmission channels to various equipment. Layer two, the data-link layer, provided for transfer of blocks of data between data-terminal equipment connected to a physical link, and included data sequencing, flow control, and error detection to assure error-free communication. Layer three, the network control layer, provided routing and switching of messages between adjacent nodes in the network. Layer four, the transport layer, provided an end-to-end control of the transmission channel once the path was established. Layer five, the session-control layer, opened communications, established a dialogue, and maintained the connection including the control and synchronization for the transfer of messages between two computers. Layer six, the presentation layer, insured the message was transferred in a coded form that the receiving computer could interpret. Layer seven, the application-user layer, the only part of the system apparent to the user, provided services that facilitated data exchange between application processes on different computers (Blaine 1983; Huff 1998). Thus each of the seven ISO layers had a defined set of functions and a layer protocol that established the rules for exchange with the corresponding layer in another computer. Orthner (1998) noted that network protocols required standardization of a variety of processes involved in data communications; and this led the International Organization to foster the development of the Open System Interconnection (OSI) Reference Model. To permit the connection and integration of local-area networks (LANs) with other LANs required the development of: (a) *bridges* that operated at level two of the ISO/OSI seven-level architecture to connect one LAN to another; (b) *routers* that operated at layer three and routed packets of data between dissimilar networks; and (c) *gateways* that operated at level seven, providing high-speed communications from a host computer to the network.

Medical data standards for data transmission began to be developed in 1983; and its early history was reviewed by McDonald (1990, 1983), McDonald and

Hripsak (1992). The proposed standards addressed what items of information should be included in defining an observation, what data structure should be employed to record an observation, how individual items should be encoded and formatted, and what transmission media should be supported. Formal attempts to improve the standardization of medical data were carried out by collaborating committees, such as the subcommittees on Computerized Systems of the American Standards for Testing Materials (ASTM), the oldest of the nonprofit standards-setting societies, and a standards-producing member of the American National Standards Institute (Rothrock 1989). The ASTM technical subcommittee E31.12 on Medical Informatics considered nomenclatures and medical records (Gabrieli 1985). In 1988 ASTM's subcommittee E31.11 on Data Exchange Standards for Clinical Laboratory Results published its specifications, E1238, for clinical data interchange, and set standards for the two-way digital transmission of clinical data between different computers for laboratory, for office, and for hospital systems; so that, as a simple example, all dates for years, months and days should be recorded as an eight-character string, YYYYMMDD. Thus the date, January 12, 1998, should always be transmitted as 19980112 (ASTM 1988a, b, 1989). The Medical Data Interchange (MEDIX) P1157 committee of the Institute of Electrical and Electronics Engineers (IEEE), formed at the Symposium on Computer Applications in Medical Care (SCAMC) in 1987, was also developing a set of standards based on the ISO application-level standards for the transfer of clinical data over large networks from mixed sources, such as from both a clinical laboratory and a pharmacy, for both intra- and inter-hospital data exchange. Linkages of data within a hospital were considered to be "tight, synchronous", and between hospitals were assumed to be "loose, asynchronous" (Rutt 1989). McDonald (1990), McDonald and Hripsak (1992) emphasized the need for clinical-data interchange standards that became essential when electronic medical records (EMRs) became technically feasible, and needed to integrate all of the various formats and structures of clinical data from the computer-based, clinical laboratory system, the radiology system, pharmacy system, and from all of the medical specialty subsystems such as the intensive-care unit, the emergency department, and others. Orthner (1992) described several important advances for digital communication systems that evolved in the 1990s, including: (1) time division multiplexed (TDM) systems that allowed several lower-speed digital communication channels to interleave onto a higher-speed channel; (2) the evolution of Integrated Services Digital Network (ISDN) that developed international standards to satisfy the needs for medical database systems; and to provide users with universal, digital inter-connectivity regardless of modality, including natural-language text, voice, and three-dimensional images; (3) the increasing use of broadband fiber-optics for digital data communication; and (4) the evolving global use of wireless communications.

Health Level 7 (HL7), an international organization made up of computer vendors, hospital users, and healthcare consultants, was formed in 1987 to develop interface standards for transmitting data between medical applications that used different computers within hospital information systems, with the goal of creating a common language to share clinical data (Simborg 1987). HL7 communicates data as a sequence of defined ASCII characters, which are hierarchically organized into segments, fields, and components. The message content of HL7 conforms to the International Standards Organization (ISO) standards for the applications level seven of the Open Systems Interconnection (OSI) model. The HL7 standards use the same message syntax, the same data types, and some of the same segment definitions as ASTM 1238 (McDonald and Hammond 1989; McDonald and Siu 1991; McDonald and Hripsak 1992). HL7 expanded its activities in the 1990s, and became one of the accredited Standards Developing Organizations (SDOs) in the American Standards Institute (ANSI) to collaborate with other SDOs to develop standards, specifications and protocols for the interoperability of hospitals clinical and administrative functions. HL7 version-3 published in 1995 its Reference Information Model (HL7 RIM) with the goal of providing improved standard vocabulary specifications for the interoperability of healthcare information systems including electronic medical records; and to improve representation of semantical, syntactical and lexical aspects of HL7 messages (Smith and Ceusters 2006). Bakken et al. (2000) described some activities of the HL7 Vocabulary Activity Committee related to vocabulary domain specifications for HL7-coded data elements, and for its guidance in developing and registering terminology and vocabulary domain specifications including those for HL7 RIM. In 2004 HL7 released its draft standards for the electronic medical record that included: (1) direct care functions, including care management and clinical decision support, (2) supportive care functions, including clinical support, research, administrative and financial functions; and (3) information infrastructure functions of data security and records management (Fischetti et al. 2006).

2.4 Classification of Medical Databases

Medical databases are classified in this book in accordance with their objectives, which can be to support clinical patient care, or to support medical research, or support administrative functions, or public health objectives. Medical databases collect, integrate, and store data from various sources; and they are usually considered to be *primary* databases if the data were initially collected and used to serve the direct purposes of the user; and are considered to be *secondary* databases when data derived from primary databases were stored in other databases and used for other objectives (Glichlich et al. 2007).

Clinical databases include a variety of primary and secondary databases that are used primarily by physicians to support their clinical patient care by helping in making decisions for the diagnosis and treatment of patients. The great utility of clinical databases resides in their capacity for storing huge volumes of information collected from large numbers of patients and from other clinical sources; and for their ability to help users to search, retrieve, and analyze information relevant to their clinical needs. Michalski et al. (1982) described clinical databases as constructed to collect patient data and to learn more about the phenomena which produced the data; and he divided techniques for using clinical databases into: (a) descriptive analyses to extract summaries of important features of a database, such as grouping patients with similar syndromes and identifying important characteristics of each syndrome; and (b) predictive analyses to derive classification rules, such as developing rules which predict the course of a disease. Clinical databases were differentiated by Hlatky (1991) as either primary medical databases that are intended to assist and support decision making in direct patient care; or as secondary medical databases that are the repositories of data derived from medical primary databases, and these include medical specialized databases (see Chap. 5) and medical research databases (see Chap. 6)

Primary medical record databases, also more commonly referred to as patient record databases, as electronic medical records (EMRs) or as electronic health records (EHRs), are the data repositories used by physicians, nurses, and other health-care providers to enter, store, and retrieve patients' data during the process of providing patient care. The National Library of Medicine's MESH terms defines an electronic medical record (EMR) as a computer-based system for input, storage, display, retrieval, and printing of information contained in a patient's medical record (Moorman et al. 2009). Primary clinical databases also include the separate repositories for storing data collected from clinical specialties, such as from surgery, pediatrics, obstetrics, and other clinical services; and from the clinical support services, such as from laboratory, radiology, pharmacy, and others. Patient record databases may contain data collected over long periods of time, sometimes for a patient's life-time; they are accessed by a variety of users for different patient-care purposes; and they need to satisfy legal requirements for maintaining the security, privacy and confidentiality of all of their patients' data (see also Sect. 4.1.1). When computer-based patients' records replaced paper-based patients' charts, the hospital record room was replaced by a computer center that initially stored the patient-record databases on magnetic tapes or discs. The rapidly increasing volume of computer-based information stimulated the development of larger storage devices and more efficient database-management systems. For most medical applications, Blum (1986a) emphasized that the primary utility of a clinical information system depended on its database-management system. It soon became apparent that the complex requirements of patient-record databases required combined hierarchical, relational, and object-oriented structural approaches. After a review of the patient-record database structures employed in the 1990s, Stead et al. (1992) et al. reported that the major problem for a patient-record database-management system was the difficulty of mapping complex logical structures into a physical media; and concluded that patient-record databases were much more complicated than were databases used for other purposes, that none of the existing database structural designs was adequate for developing, as an example, a common national patient-record database, and some combination of database designs would needed to be employed. Dick and Steen (1992) and E. Steen also reviewed the essential technologies needed for a patient-record database system, and agreed that in the 1990s there was not yet one medical database system available that could serve as a model for computer-based patient record systems, and that could satisfy all the continual changing requirements for timely processing of all the information commerce in a comprehensive patient-care system, with all of its different information modalities and changing patient-care technologies, and with its strict legal requirements for assuring patients' data security and confidentiality.

Camp et al. (1983) described some of the complexities of primary clinical databases, namely: (1) at the time when patient-care information was being obtained it was not always known what data might be needed in the future,so this tended to enlarge a database with some data that was never used; (2) the database had to store information that could be differently structured and formatted, and was often unstandardized; (3) it needed to allow exploring complex data relationships in (frequently) a minimal access time, and not unduly interfere with the productivity of busy health care providers who were not computer programmers; and (4) a common deficiency of primary clinical databases was that they tended to lack patients' data for events that occurred between recorded visits to their health care providers. Connolly and Begg (1999) noted that since most clinical data were "time-stamped", it was necessary that data transactions be recorded and retrieved in their correct time sequence. Graves (1986) added that another requirement for a medical database was to provide a natural language processing (NLP) program that had the capability to query textual information such as were obtained by patient interviews, and that could include relevant expressed feelings and experiential information. The availability of online access to clinical databases greatly facilitated the process of searching and retrieving information when needed in a timely way by physicians for clinical decision-making. The factors that influenced the rate of diffusion of medical databases and other computer applications in medical practice were studied by Anderson and Jay (1984) at Purdue and Indiana Universities; and they concluded that physicians had the major role in their diffusion.

Specialized clinical databases can be disease-specific (as for heart disease or cancer), or device- or procedure-specific (as for coronary artery bypass surgery), or therapy-specific (as for anti-viral drugs), or population-specific (as for a geriatric or a racial group). Safran and Chute (1995) observed that a clinical database could be used to query for information on an individual patient, or to find data on patients with similarities to the one being cared for, or to describe a group of patients with some common attributes, or to analyze data patterns in terms of trends or relationships. Fries (1984) noted that some of the most important medical problems were the chronic diseases, such as arthritis, cancer, and heart disease; and a study of the management of these disorders could be benefited by chronic diseases databases (see also Sect. 5.3). A large medical center often had many specialized clinical databases for its various inpatient and outpatient clinical services, and for its clinical support subsystems (laboratory, radiology, pharmacy, and others). As a result it usually needed a distributed database-management system to service them all. Each clinical subsystem's database needed extract-load-transfer (ETL) programs to move data to-and-from its subsystem database and the central, integrated, clinical database.

Clinical research databases may be primary databases when the clinical patient data was collected for the primary purpose of supporting clinical research, such as for clinical trials; but they are usually secondary research databases that contain selected data extracted from primary medical databases, such as when they contain clinical data extracted from primary medical records for groups of patients with the same problem. This differs from primary patient care databases where the medical record of each patient needs to contain all of the information collected for all of the medical problems of that individual patient. In a secondary research database it is usually necessary to extract and transfer the selected data from the primary patient-record database into the secondary research database; and all patient data transferred from a primary patient-record database has additional special legal requirements for assuring the data validity, data security, and the strict privacy and confidentiality of each patient's data. Garfolo and Keltner (1983) emphasized the importance of the need to de-identify patient data when a clinical database is also used for research purposes (see Sects. 4.1.1 and 6.1.1).

Biosurveillance databases were developed by the FDA for the surveillance of adverse drug events (see Sect. 7.1); and by the CDC for the surveillance of epidemics of infectious diseases (see Sect. 7.2). *Claims databases* were established by Medicare, Medicaid, and commercial health care insurers for collecting from health care providers their relevant sub-sets of primary medical record data for the purpose of arranging payments for claims of provided clinical services (see Sect. 7.3). *Medical knowledge databases* are comprehensive collections of information from a variety of sources, including clinical and research databases, textbooks, and publications by experts in specific medical issues. They are used to communicate medical information in order to support the clinical decision-making process (see Sect. 8.1); and large knowledge bases with other medical databases have been combined and used for data mining to discover new knowledge (see Sect.8.2). *Medical bibliographic databases* are collections of medical literature developed as fact-and-information locators in libraries and other collections of relevant medical publications, and are used to provide and communicate medical information. The National Library of Medicine (NLM) is the primary resource in the world for a variety of bibliographic databases (see Sect. 9.1).

Metadatabases are developed to store *metadata,* that are data that describe the data contained in a database for the purposes of providing a dictionary with definitions of terms; and a list of coded data in the database with their codes; and to serve as a thesaurus to recognize different terms that have similar meanings; and to provide a lexicon of standard, accepted, defined, and correctly spelled terms. A metadatabase needs to contain associated relevant information to aid: in the storage and retrieval of data in the database; in providing linkages to other data items and files; in providing keys to related tables; in providing logical connections for data presentation, interactive validation, data extraction, permitting ad-hoc query; and also providing users with interfaces for any metadata additions or corrections. A data dictionary was usually initiated as a part of a metadatabase by selecting commonly used terms from a standard medical dictionary and from related medical literature; and needed to be capable of adding new terms from the database itself; so the design of the data dictionary had to allow for incorporating new data items when they were introduced, such as for new procedures. As these lexicons became the basis for automated natural-language processing, they also usually included: (1) syntactical information as to whether a word was a noun, a verb, or other; and (2) the word's semantical information as to its meaning in the language of medicine (McCray et al. 1987).

For a primary patient-record database the metadatabase needed: to provide any special instructions for conducting clinical procedures; needed to describe all processes and procedures such as clinical laboratory tests; and needed to specify the normal and the "alert" boundary limits for each clinical test and procedure. Warner (1979) emphasized that the purpose of a metadatabase was to minimize the chance of ambiguity in data representation between the point of data entry and the point at which the data was used. Anderson (1986) credited the Veterans Administration (VA) with publishing the first data dictionary as a part of the VA's computer-based medical record (see Sect. 4.2). Hammond et al. (1977, 1980, 1985) and W. Stead described in some detail the metadatabase developed for Duke University's TMR (The Medical Record) (see Sect. 4.2). Their metadatabase included patients' identification data; it defined and coded all clinical variables including patients' medical problems, diagnostic studies, and therapies. They used their metadatabase as a dictionary to define the codes for their computer-based, clinical laboratory system that was linked to their TMR system. Their metadatabase permitted modifying and updating specific clinical functions, and allowed for differences between various medical specialties and clinics. Sections of the metadatabase were devoted to system specifications; to medical problems, procedures, therapies; and to health-care providers' information. It contained patients' demographic and examination data, clinical reports and messages, and also professional fees and accounting data. An alphabetically arranged thesaurus provided definitions of synonyms. Where appropriate for free-text input, all codes and their text equivalents were defined in the metadatabase. The user could enter a code directly; or could type in the textual data and then let the program do an alphabetic search in the metadatabase and convert the text-string into the appropriate code. With the advent of the World Wide Web, Munoz and Hersh (1998) reported using a Java-based program for generating a Web-based metadatabase.

2.5 Summary and Commentary

In the 1950s patients' medical records were paper-based and were stored in stacks of charts on shelves in a medical record room. In the early 1960s the development of computers allowed patient-care data to be entered into a computer by using punched paper cards; and the data were stored and accessed sequentially in computer flat files that had little structured relationships; and they were aggregated in file-management systems. In the late 1960s structured computer databases began to evolve with associated database-management systems. In the 1970s distributed database systems began to be developed; and in the following decades of the 1980s, 1990s, and the 2000s the development of increasingly large and ebhanced medical databases was truly phenomenal.

In the 1960s hospital information systems began to be developed that used large mainframe computers with integrated databases that serviced all clinical departments.

It was soon found that although a single, large, mainframe computer could readily integrate patient data into a single database, it could not adequately support the information processing requirements for all of the clinical specialty and ancillary services in a large medical center. In the 1970s the advent of minicomputers permitted many hospital services to have their subsystems databases directly linked to a central mainframe computer that integrated all patients' data into the patients' clinical records that were stored in the mainframe computer's database (Ball and Hammon 1975a, b). Some patient data were manually encoded before being entered into the database to facilitate billing for payments of claims, and for the retrieval of data for management and clinical research purposes. In the 1980s a diffusion of minicomputers and microcomputers were incorporated into a variety of medical applications. Micro-computer-based subsystems that had evolved independently for specialized clinical and ancillary services usually became subsystems of larger medical information systems with an integrating central database-management system. Storage technology improved, storage devices became cheaper and larger; registries grew in size to become databases; databases became data warehouses; and a great variety of secondary clinical databases evolved.

In the 1990s international communications used computers and local-area networks; and the use the Internet and the World Wide Web became commonplace. As patient-care data expanded in both volume and complexity, frequent innovations in informatics technology provided more efficient computer-based, clinical-information systems in hospitals and in medical offices. In the 2000s distributed information systems allowed physicians to enter orders and retrieve test results using clinical workstations connected to client–server computers in local-area-networks that linked multiple medical center databases. By the end of the 2000s there had evolved global wireless communications with translational networks that linked data warehouses in collaborating medical centers in the nation.

References

Anderson J. Data dictionaries – a way forward to write meaning and terminology into medical information systems. Methods Inf Med. 1986;25:137–8.

Anderson JB, Jay SJ. The diffusion of computer applications in medicine: network location and innovation adoption. Proc SCAMC. 1984:549–52.

ASTM (American Society for Testing and Materials) E 1238–88. Standard specifications for transferring clinical laboratory data messages between independent computer systems. Philadelphia: ASTM; 1988.

ASTM (American Society for Testing and Materials) E 1239–88. Standard guide for description of reservation/registration-admission, discharge, transfer (R-ADT) systems for automated patient care information systems. Philadelphia: ASTM; 1988.

ASTM (American Society for Testing and Materials) E 1238–88. Standard specifications for transferring clinical observations between independent computer systems. Philadelphia: ASTM; 1989. Revision November 30.

Bakken S, Campbell KE, Cimino JJ, et al. Toward vocabulary domain specifications for health level 7-coded data elements. J Am Med Inform Assoc. 2000;7:333–42.

Ball MJ, Hammon GL. Overview of computer applications in a variety of health care areas. CRC Crit Rev Bioeng. 1975a;2(2):183–203.
Ball MJ, Hammon GL. Maybe a network of mini-computers can fill your data systems needs. Hosp Financ Manage. 1975b;29(4):48–51.
Barnett GO. Massachusetts general hospital computer system. In: Collen MF, editor. Hospital computer systems. New York: Wiley; 1974. p. 517–45.
Barnett GO, Souder D, Beaman P, Hupp J. MUMPS – an evolutionary commentary. Comput Biomed Res. 1981;14:112–8.
Barsalou T, Wiederhold G. A cooperative hypertext interface to relational databases. Proc SCAMC. 1989:383–7.
Blaine GI. Networks and distributed systems. A primer. Proc MEDINFO. 1983:1118–21.
Blois MS, Henley RR. Strategies in the planning of hospital information systems. In: Journee D'Informatique Medicale. Toulouse: Institut de Recherche d'Informatique et d'Automatique; 1971. p. 89–98.
Blois MS, Wasserman AI. The integration of hospital information systems. Journee D'Informatique Medicale, Toulouse; also: Tech Report #4, Office of Med Inform Systems. San Francisco: University of California; 1974.
Blum RL. Machine representation of clinical causal relationships. Proc MEDINFO. 1983b:652–6.
Blum BI. A history of computers, chap 1. In: Blum BI, editor. Clinical information systems. New York: Springer; 1986a. p. 1–32.
Blum BI. Programming languages, chap 4. In: Blum BI, editor. Clinical information systems. New York: Springer; 1986b. p. 112–49.
Blum BI. Design methods for clinical systems. Proc SCAMC. 1986c:309–15.
Blum BI, Duncan K, editors. A history of medical informatics. New York: Addison-Wesley Pub. Co; 1990. p. 1–450.
Brandt CA, Morse R, Mathews K, et al. Metadata-driven creation of data marts from an EAV-modeled clinical research database. Int J Med Inform. 2002;65:225–41.
Bryan M. 1988: the year of the data base. Personal Comput. 1988;12(1):100–9.
Camp HN, Ridley ML, Walker HK. THERESA: a computerized medical consultant based on the patient record. Proc MEDINFO. 1983:612–4.
Campell-Kelly M. Computing. Sci Am. 2009;301:63–9.
Chen RS, Nadkarni P, Marenco L, et al. Exploring performance issues for a clinical database using an entity-attribute-value representation. J Am Med Inform Assoc. 2000;5:475–87.
Codd EF. A relational model of data for large shared data banks. Comm ACM. 1970;13:377–87.
Codd EF. Further normalization of the data base relational model. In: Rustin R, editor. Database systems. Englewood Cliffs: Prentice-Hall; 1972. p. 33–64.
Codd EF. Extending the data base relational model to capture more meaning. ACM Trans Database Syst. 1979;4:397–434.
Codd EF, Codd SB, Salley CT. Providing OLAP (On-line Analytical Processing) to User-Analysts: an IT mandate. San Jose: Codd and Date, Inc; 1993.
Collen MF. Origins of medical informatics. West J Med. 1986;145:778–85.
Collen M. Clinical research databases – a historical review. J Med Syst. 1990;14:323–44.
Collen MF. The origins of informatics. J Am Med Inform Assoc. 1994;1:91–107.
Collen MF. A history of medical informatics in the United States. Bethesda/Indianapolis: American Medical Informatics Assn/Bookscraft; 1995.
Coltri A. Databases in health care, chap 11. In: Lehman HP, Abbott PA, Roderer NK, et al., editors. Aspects of electronic health record systems. 2nd ed. New York: Springer; 2006. p. 225–51.
Connolly TM, Begg CE. Database management systems: a practical approach to design, implementation, and management. 2nd ed. New York: Addison-Wesley; 1999.
Davis LS. Prototype for future computer medical records. Comput Biomed Res. 1970;3:539–54.
Davis LS. A system approach to medical information. Methods Inf Med. 1973;12:1–6.
Davis LS, Terdiman J. The medical data base, chap 4. In: Collen MF, editor. Hospital computer systems. New York: Wiley; 1974. p. 52–79.

Davis LS, Collen MF, Rubin L, Van Brunt EE. Computer-stored medical record. Comput Biomed Res. 1968;1:452–69.
Dawson J. A family of models. Byte. 1989;4:277–86.
Deshpande AM, Brandt C, Nadkarni PM. Temporal query of attribute-value-patient data: utilizing the constraints of clinical studies. Int J Med Inform. 2003;70:59–77.
Dick RS, Steen EB. Essential technologies for computer-based patient records: a summary. In: Ball MJ, Collen MF, editors. Aspects of the computer-based patient record. New York: Springer; 1992. p. 229–61.
Dinu V, Nadkarni P. Guidelines for the effective use of entity-attribute-value modeling for biomedical databases. Int J Med Inform. 2007;76:769–79.
Duke JR, Bowers GH. Scope and sites of electronic health record systems. In: Lehman HP, Abbott PA, Roderer NK, et al., editors. Aspects of electronic health record systems. New York: Springer; 2006. p. 89–114.
Fischetti L, Schloeffel P, Blair JS, Henderson ML. Standards. In: Lehmann HP, Abbott PA, Roderer NK, et al., editors. Aspects of Electronic Health Record Systems. New York: Springer; 2006. p. 252–82.
Frawley WJ, Piatetsky-Shapito G, Matheus CJ. Knowledge discovery in databases: an overview. AI Magazine. 1992;13:57–70.
Friedman C, Hripcsak G, Johnson SB, et al. A generalized relational schema for an integrated clinical patient database. Proc SCAMC. 1990:335–9.
Fries JF. The chronic disease data bank: first principles to future directions. J Med Philos. 1984;9:161–89.
Gabrieli ER. Standardization of medical informatics (special issue). J Clin Comput. 1985;14: 62–104.
Garfolo BT, Keltner L. A computerized disease register. Proc MEDINFO. 1983:909–12.
Glichlich RE, Dreyer NA, eds. Registries for evaluating patient outcomes: a user's guide. AHRQ Pub. # 07-EHC001-1. Rockville: Agency for Healthcare Research and Quality; 2007(Apr). p. 1–233.
Graves J. Design of a database to support intervention modeling in nursing. Proc MEDINFO. 1986:240–2.
Greenes RA, Papillardo AN, Marble CW, Barnett GO. Design and implementation of a clinical data management system. Comput Biomed Res. 1969;2:469–85.
Grossman JH, Barnett GO, Koepsell TD, et al. An automated medical record system. JAMA. 1973;224:l6l6–1621.
Hammond WE, Stead WW, Feagin SJ, et al. Data base management system for ambulatory care. Proc SCAMC. 1977:173–87.
Hammond WE, Stead WW, Straube MJ, Jelovsek FR. Functional characteristics of a computerized medical record. Methods Inf Med. 1980;19:157–62.
Hammond WE, Stead WW, Straube MJ. Planned networking for medical information systems. Proc SCAMC. 1985:727–31.
Hlatky M. Using databases to evaluate therapy. Stat Med. 1991;10:647–52.
Hripcsak G, Allen B, Cimino JJ, Lee R. Access to data: comparing AcessMed with Query by Review. J Am Med Inform Assoc. 1996;3:288–99.
Huff SM. Clinical data exchange standards and vocabularies for messages. Proc AMIA. 1998:62–7.
Johnson SB. Generic data modeling for clinical repositories. J Am Med Inform Assoc. 1996;3:328–39.
Johnson SB. Extended SQL for manipulating clinical warehouse data. Proc AMIA. 1999:819–23.
Kuznak PM, Kahane SN, Arsenlev M, et al. The role and design of an integrated clinical result database within a client-server networked hospital information system architecture. Proc SCAMC. 1987:789–95.
Lindberg DAB. The growth of medical information systems in the United States. Lexington: Lexington Books; 1979.

London JW. A computer solution to clinical and research computing needs. Proc SCAMC. 1985:722–26.

Marrs KA, Kahn MG. Extending a clinical repository to include multiple sites. Proc AMIA. 1995:387–91.

McCray AT, Sponsler JL, Brylawski B, Browne AC. The role of lexical knowledge in biomedical text understanding. Proc SCAMC. 1987:103–7.

McDonald CJ. Standards for the transmission of diagnostic results from laboratory computers to office practice computers – an initiative. Proc SCAMC. 1983:123–4.

McDonald CJ. Standards for the electronic transfer of clinical data: programs, promises, and the conductor's wand. Proc SCAMC. 1990:09–14.

McDonald CJ, Hammond WE. Standard formats for electronic transfer of clinical data. Editorial. Ann Intern Med. 1989;110:333–5.

McDonald CJ, Hripsak GH. Data exchange standards for computer-based patient records. In: Ball MF, Collen MF, editors. Aspects of the computer-based patient record. New York: Springer; 1992. p. 157–64.

McDonald CJ, Siu SL. The analysis of humongous databases: problems and promises. Stat Med. 1991;10:511–8.

McDonald CJ, Wilson G, Blevins L, et al. The Regenstrief medical record system. Proc SCAMC. 1977a:168–9.

McDonald CJ, Murray M, Jeris D, et al. A computer-based record and clinical monitoring system for ambulatory care. Am J Public Health. 1977b;67:240–5.

McDonald CJ, Blevins L, Glazener T, et al. Data base management, feedback control and the Regenstrief medical record. Proc SCAMC. 1982:52–60.

McDonald CJ, Blevens L, Tierney WM, Martin DK. The Regenstrief medical records. MD Comput. 1988:34–47.

Michalski RS, Baskin AB, Spackman KA. A logic-based approach to conceptual database analysis. Proc SCAMC. 1982:792–6.

Miller RA, Kapoor WN, Peterson J. The use of relational databases as a tool for conducting clinical studies. Proc SCAMC. 1983:705–8.

Moorman PW, Schuemie MJ, van der Lei J. An inventory of publications on electronic medical records revisited. Methods Inf Med. 2009;48:454–8.

Munoz F., Hersh W. MCM Generastors: a Java-based tool for generating medical metadata. Proc AMIA. 1998:648–52.

Nadkarni PM, Cheung K. SQLGEN: a framework for rapid client-server database application development. Comput Biomed Res. 1995;28:479–99.

Nadkarni P, Marenco L. Easing the transition between attribute-value databases and conventional databases for scientific data. Proc AMIA. 2001:483–7.

Nadkarni PM, Brandt C, Frawley S, et al. Managing attribute-value clinical trials data using ACT/DB client-server database system. J Am Med Inform Assoc. 1998;5:139–51.

Nadkarni PM, Marenco L, Chen R, et al. Organization of heterogeneous scientific data using the EAV/CR representation. J Am Med Inform Assoc. 1999;6:478–93.

Nadkarni PM, Brandt CM, Marenco L. WebEAV: automatic meta-driven generation of web interfaces to entity-attribute-value-databases. J Am Med Inform Assoc. 2000;7:343–56.

Orthner HF. New communication technologies for integrating hospital information systems and their computer-based patient records, chap 11. In: Ball MJ, Collen MF, editors. Aspects of the computer-based patient record. New York: Springer; 1992. p. 176–200.

Orthner HF. New communication technologies for hospital information systems. In: Bakker AR, Ball MJ, Scherrer JR, Willems JL, editors. Towards new hospital information systems. Amsterdam: North-Holland; 1998. p. 203–12.

Pendse N. OLAP Omnipresent. Byte. 1998;111:751–6.

Pendse N. Online analytical processing. Wikipedia. Retrieved in 2008. http://en.wikipedia:org/wiki/Online_analytical_processing.

Pryor DB, Stead WW, Hammond WE, et al. Features of TMR for a successful clinical and research database. Proc SCAMC. 1982:79–83.

Pryor TA, Gardner RM, Clayton PD, Warner HR. The HELP system. J Med Syst. 1983;7:87–102.

Robson D. Object-oriented software system. Byte. 1981;6:74–86.

Rothrock JJ. ASTM: the standards make the pieces fit. Proc AAMSI Congress. 1989:327–35.

Rutt TE. Work of IEEE P1157 medical interchange committee. Proc AAMSI Congress. 1989:403–22.

Safran C, Chute CG. Exploration and exploitation of clinical databases. Int J Biomed Comput. 1995;39:151–6.

Simborg DW. Local area networks: why? what? what if? MD Comput. 1984;1:10–20.

Simborg DW. An emerging standard for health communications: the HL7 standard. Healthc Commun (HC&C). 1987;3:58–60.

Smith B, Ceusters W. HL7 RIM: an incoherent standard. Stud Health Technol Inform. 2006;124:133–8.

Starr P. The social transformation of American medicine. New York: Basic Books; 1982.

Stead WW, Hammond WE. Computer-based medical records: the centerpiece of TMR. MD Comput. 1988;5:48–61.

Stead WW, Wiederhold G, Gardner R, et al. Database systems for computer-based patient records. In: Ball MJ, Collen MF, editors. Aspects of the computer-based patient record. New York: Springer; 1992. p. 83–98.

Taylor RW, Frank RL. CODASYL data base management systems. Comput Surv. 1976;8:67–103.

Terdiman J. Ambulatory care computer systems in office practice: a tutorial. Proc AMIA. 1982:195–201.

Tolchin SG, Stewart RL. The distributed processing approach to hospital information processing. J Med Syst. 1981;5:345–60.

Tolchin SG, Blum BI, Butterfield MA. A system analysis method for a decentralized health care information system. Proc SCAMC. 1980:1479–84.

Tolchin SG, Simborg DW, Stewart RL, et al. Implementation of a prototype generalized network technology for hospitals. Proc SCAMC. 1981a:942–8.

Tolchin SG, Stewart RL, Kahn SA, et al. A prototype generalized network technology for hospitals. J Med Syst. 1982;6:359–75.

Tolchin SG, Barta W, Harkness K. The Johns Hopkins Hospital network. Proc SCAMC. 1985a:732–7.

Tolchin SG, Arsenlev M, Barta WL, et al. Integrating heterogeneous systems using local network technologies and remote procedure call protocols. Proc SCAMC. 1985b:748–9.

Tuck D, O'Connell R, Gershkovitch P, Cowan J. An approach to object-relational mapping in bioscience domains. Proc AMIA Symp. 2002:820–4.

VanName ML, Catchings B. SQL: a database language sequel to dBase. Byte. 1989;14:175–82.

Walters RF. Microprocessors as intelligent front-end devices for medical information systems. Med Inform. 1979;4:139–50.

Warner HR. Patient data file, chap 3. In: Computer-assisted medical decision-making. New York: Academic; 1979. p. 102–23.

Warner HR. History of medical informatics at Utah. In: Blum BI, Duncan K, editors. A history of medical informatics. New York: Addison-Wesley Pub. Co; 1990. p. 357–66.

Warner HR, Olmsted CM, Rutherford BD. HELP – a program for medical decision-making. Comput Biomed Res. 1972;5:65–74.

Warner HR, Morgan JD, Pryor TA, et al. HELP – a self-improving system for medical decision-making. Proc MEDINFO. 1974:989–93.

Wasserman AI. Minicomputers may maximize data processing. Hospitals. 1977;51:119–28.

Wasserman AI. Interactive development environments for information systems. Proc SCAMC. 1986:316–25.

Wess BP. Distributed computer networks in support of complex group practices. Proc SCAMC. 1978:469–77.

Wiederhold G. Database technology in health care. J Med Syst. 1981;5:175–96.

Wiederhold G. Databases for ambulatory care. Proc AMIA Symp. 1982:79–85.
Wiederhold G. Modeling databases. Inf Sci. 1983;29:115–26.
Wiederhold G. Databases, A tutorial. Proc AAMSI. 1984:423–30.
Wiederhold G, Fries GF, Weye S. Structural organization of clinical data bases. Proc AFIPS Conf. 1975;44:479–85.
Wiederhold G, Walker MG, Blum RL, et al. Acquisition of medical knowledge from medical records. Proc Benutzer-gruppenseminar Med Syst. 1987:213–4.
Zeichner ML, Brusil OJ, Tolchin SG. Distributed processing architecture for a hospital information system. *Proc SCAMC*. 1979:859–65.

http://www.springer.com/978-0-85729-961-1

Computer Medical Databases
The First Six Decades (1950–2010)
Collen, M.F.
2012, XX, 288 p., Hardcover
ISBN: 978-0-85729-961-1