Health Informatics
Series Editors: Kathryn J. Hannah · Marion J. Ball

Morris F. Collen

Computer Medical Databases

The First Six Decades (1950–2010)

Springer

Health Informatics

Morris F. Collen

Kathryn J. Hannah • Marion J. Ball
(Series Editors)

Computer Medical Databases

The First Six Decades (1950–2010)

Springer

Morris F. Collen
Division of Research
Broadway 2000
94612 Oakland, California
USA

ISBN 978-0-85729-961-1 e-ISBN 978-0-85729-962-8
DOI 10.1007/978-0-85729-962-8
Springer London Dordrecht Heidelberg New York

British Library Cataloguing in Publication Data
A catalogue record for this book is available from the British Library

Library of Congress Control Number: 2011939795

Printed on acid-free paper

Springer is part of Springer Science+Business Media (www.springer.com)

Foreword I

This latest book by Dr Morris Collen, "**A HISTORY OF MEDICAL DATABASES"** is a delight. I have never asked Dr. Collen if he reads Rowling's Harry Potter books. Perhaps not; he may be too busy with his own writing. Nonetheless Doctor Collen is fully entitled to be known as a Wizard amongst the people of Medical Informatics. I note this distinction because diving into this book is very much like diving into the Wizard Dumbledore's magical vase the "Pensive". A delightful and surprising journey begins in which no one ever dies, the relationships between ideas are revealed, and the reader feels cleansed of any of his own mistakes and happy to be part of the story.

To be fair, Doctor Collen warns his readers that "this book is primarily a history of how people applied computers, so it is not a history about the people themselves". But, no matter, I recall that Harry's hero also took a similar wizardly "above the fray" tone. I find I myself have been treated all too kindly by our author; so please excuse this.

Doctor Collen also treats the National Library of Medicine rather well. I claim this is proper; it's a grand institution and a source of much stimulus and sustained support to research and training in bio-medical uses of computers and information systems. It is important to regard this as a real institutional commitment, separate from that of any individuals.

I hope he may consider another book to hail the other great American institutions, including our great universities and medical centers, that have also supported this important work. All these institutions - especially the Federal institutions - even great ones – are surprisingly fragile. They are very much subject to the whims and waves of societal hopes and of scientific "theories". Serious budget cuts could destroy them in only a few years.

On the positive side, the NLM can look back 175 years and be proud that Presidents and Congress have faithfully supported its mission through depressions, wars, and domestic and international hard times. NLM's proudest national collections and achievements resulted from abiding Congressional belief and financial support. I hope future readers' trips through the Pensive will again find Morris Collen at his keyboard and our American science institutions strong and faithful.

Donald Lindberg, M.D.

Foreword II

This new volume by Morris Collen, "Computer Medical Databases: The First Six Decades, 1950–2010," is sure to join his 1995 book, "A History of Medical Informatics in the United States," on the bookshelves of health informaticians in the U.S. and around the world as a trusted reference.

In this book in Chap. 2, Morrie credits Gio Wiederhold of Stanford University with the early definition and design of databases as collections of related data, organized so that usable data may be extracted. Morrie presents the history of medical databases in ten chapters. He traces their evolution, giving detailed exemplars of specialized clinical databases and secondary healthcare databases. He lays out illuminating examples of both knowledge and bibliographic databases and pays tribute to the National Library of Medicine – a remarkable institution that in 2011 celebrated its 175th year of providing the best-of-the-best medical knowledge to the worldwide healthcare community.

A mentor, teacher, and friend to many of us for 40 years, Morrie has made and continues to make invaluable contributions to Kaiser Permanente and to the global medical informatics community; contributions that have transformed many aspects of healthcare delivery, medical research, and clinical practice.

Every year the American College of Medical Informatics gives the coveted Morris Collen Lifetime Achievement Award to an individual whose work has advanced the field of health informatics. We are all blessed with the wisdom, friendship, and humanity Morrie has shared so generously. We owe him – who has truly earned the title of "father of medical informatics" – our thanks for his tireless and insightful work over the years. His contributions, including this latest volume, do honor to the field and to those of us who are privileged to have him as a colleague.

Marion J. Ball

Series Preface

This series is directed to healthcare professionals leading the transformation of healthcare by using information and knowledge. For over 20 years, Health Informatics has offered a broad range of titles: some address specific professions such as nursing, medicine, and health administration; others cover special areas of practice such as trauma and radiology; still other books in the series focus on interdisciplinary issues, such as the computer based patient record, electronic health records, and networked healthcare systems. Editors and authors, eminent experts in their fields, offer their accounts of innovations in health informatics. Increasingly, these accounts go beyond hardware and software to address the role of information in influencing the transformation of healthcare delivery systems around the world. The series also increasingly focuses on the users of the information and systems: the organizational, behavioral, and societal changes that accompany the diffusion of information technology in health services environments.

Developments in healthcare delivery are constant; in recent years, bioinformatics has emerged as a new field in health informatics to support emerging and ongoing developments in molecular biology. At the same time, further evolution of the field of health informatics is reflected in the introduction of concepts at the macro or health systems delivery level with major national initiatives related to electronic health records (EHR), data standards, and public health informatics.

These changes will continue to shape health services in the twenty-first century. By making full and creative use of the technology to tame data and to transform information, Health Informatics will foster the development and use of new knowledge in healthcare.

Kathryn J. Hannah
Marion J. Ball

Preface

I was privileged to have witnessed the evolution of medical informatics in the United States during its first six decades. Donald A. B. Lindberg, Director of the National Library of Medicine, advised me that documenting this history would be a worthy project since during this period the country moved into a new information era, and it was obvious that computers were having a major influence on all of medicine.

In this book I address history as a chronological accounting of what I considered to be significant events. To attempt to preserve historical accuracy and minimize any personal biases, I have relied entirely on published documents; and since long-term memory can allow history to be mellowed or enhanced, and may blur fact with fantasy, I did not conduct any personal interviews. I recognize that innovators rarely publish accounts of their failures; but if they learn from their failures and publish their successes, then other innovators can build on their successes and advance the technology. This book is primarily a history of how people applied computers, so it is not a history about the people themselves. When people are mentioned, their associations and contributions are described, and they are usually referenced from their own publications.

Although the evolution of computer applications to medical care, to biomedical research, and to medical education are all related the rates of diffusion of medical informatics were different in each of these three fields. Since I was primarily involved in computer applications to patient care and to clinical research, the history of medical informatics for direct patient care in the hospital and in the medical office was presented in Book I, A History of Medical Informatics in the United States; 1959–1990 (M.Collen 1995). This present book describes the historical evolution of medical digital databases; and it omits the computer processing of digital images (for radiology), of photographs (for dermatology), and of analog signals (for electrocardiograms). The technical aspects of computer hardware, software, and communications are limited to what I judged to be necessary to explain how the technology was applied to the development and uses of medical databases. At the end of each chapter is a brief summary and commentary of my personal view on the chapter's contents.

The medical informatics literature in the United States for these six decades has been so voluminous that it was not possible for this historical review to be completely comprehensive. Undoubtedly I have overlooked some important contributions worthy of historical reference, especially of those never published. It is hoped that the sampling of the historical material herein presented will be considered by readers to be reasonably representative, and will serve as a useful bridge between medical informatics from the past into the future. The concurrent evolution of medical informatics in Canada, Europe, and Japan certainly influenced this field in the United States; however, the scope of this book is limited to the development of medical informatics in the United States.

Morris Frank Collen

Acknowledgements

This book is dedicated to Frances Bobbie Collen, my beloved wife and constant inspiration for 60 years; and who directed my career from engineering into medicine. I shall always be indebted to Sidney R. Garfield and to Cecil C. Cutting, who re-directed me from medicine into medical informatics and fashioned my entire professional career. I am very grateful to Donald A. B. Lindberg who inspired me to write about the history of medical informatics; and to Marion J. Ball who provided continuing encouragement and support during my years of writing in this exciting domain of medical informatics.

The first in this series of books on the "History of Medical Informatics" (M. Collen 1995) was initiated, and supported in part, by a contract with the National Library of Medicine arranged for me by Dr. Lindberg. In this book on the "History of Medical Databases", Betsy Humphreys contributed substantial editing of section 9.1 describing the National Library of Medicine's databases. While I was a resident scholar at the National Library of Medicine, the staff there was of inestimable help in facilitating the finding of many reference publications. While I was a fellow at the Center for Advanced Studies in the Behavioral Sciences on the Stanford University campus in 1986–87, this Center's staff arranged for my access to the Stanford University libraries where I found many of the earliest references used in this book.

In the Division of Research of the Northern California, Kaiser Permanente Medical Care Program, I collected many articles from its library with the great help and support of the always gracious and efficient librarians, Marlene R. Rogers and Brenda J. Cooke. Dr. Dana Ludwig provided many helpful suggestions for several sections of this book. Many Fellows of the American College of Medical Informatics generously contributed copies of their publications for my use. I have always found one of the best sources of publications on medical informatics to be the Proceedings of the Annual Symposia on Computer Applications in Medical Care (*Proc SCAMC*), first published in 1977, continued from 1994 as the Proceedings of the AMIA Annual Fall Symposia (*Proc AMIA*); and later as an annual Supplement to Journal of the American Medical Informatics Association (*JAMIA*). Also of great help were the Proceedings of the annual American Association for Medical Systems and Informatics (*Proc AAMSI*) Congresses, with its first volume in 1983; and the

Proceedings of the triennial International Congresses on Medical Informatics (*Proc MEDINFO*), with its first volume in 1974. To avoid undue repetitions in the references at the end of each chapter, these proceedings are referred to by their abbreviated titles, as shown on the following pages.

Morris Frank Collen

Contents

Frequently Referenced Proceedings

Proc SAMS 1973	Proceedings of the Society for Advanced Medical Systems. Collen MF (ed). San Francisco: ORSA, 1973.
Proc SCAMC 1977	Proceedings of the First Annual Symposium on Computer Applications in Medical Care. Orthner FH, Hayman H (eds). New York: IEEE, 1977.
Proc SCAMC 1978	Proceedings of the Second Annual Symposium on Computer Applications in Medical Care. Orthner FH (ed). Silver Springs, MD: IEEE Computer Society Press, 1978.
Proc SCAMC 1979	Proceedings of the Third Annual Symposium on Computer Applications in Medical Care. Dunn RA (ed). New York: IEEE, 1979.
Proc SCAMC 1980	Proceedings of the Fourth Annual Symposium on Computer Applications in Medical Care. O'Neill JT (ed). New York: IEEE, 1980.
Proc SCAMC 1981	Proceedings of the Fifth Annual Symposium on Computer Applications in Medical Care. Hefferman HG (ed). New York: IEEE, 1981.
Proc SCAMC 1982	Proceedings of the Sixth Annual Symposium on Computer Applications in Medical Care. Blum BI (ed). New York: IEEE, 1982.
Proc SCAMC 1983	Proceedings of the Seventh Annual Symposium on Computer Applications in Medical Care. Dayhoff RE (ed). New York: IEEE, 1983.
Proc SCAMC 1984	Proceedings of the Eighth Annual Symposium on Computer Applications in Medical Care. Cohen GS (ed). New York: IEEE, 1984.
Proc SCAMC 1985	Proceedings of the Ninth Annual Symposium on Computer Applications in Medical Care. Ackerman MJ (ed). New York: IEEE, 1985.
Proc SCAMC 1986	Proceedings of the Tenth Annual Symposium on Computer Applications in Medical Care. Orthner HF (ed). New York: IEEE, 1986.
Proc SCAMC 1987	Proceedings of the Eleventh Annual Symposium on Computer Applications in Medical Care. Stead WW (ed). New York: IEEE, 1987.
Proc SCAMC 1988	Proceedings of the Twelfth Annual Symposium on Computer Applications in Medical Care. Greenes RA. (ed). New York: IEEE, 1988.
Proc SCAMC 1989	Proceedings of the Thirteenth Annual Symposium on Computer Applications in Medical Care. Kingsland L (ed). New York: IEEE, 1989.
Proc SCAMC 1990	Proceedings of the Fourteenth Annual Symposium on Computer Applications in Medical Care. Miller RA (ed). Los Alamitos, CA: IEEE , 1990.
Proc SCAMC 1991	Proceedings of the Fifteenth Annual Symposium on Computer Applications in Medical Care. Clayton PD (ed). New York: McGraw-Hill, Inc., 1991.
Proc SCAMC 1992	Proceedings of the Sixteenth Annual Symposium on Computer Applications in Medical Care. Frisse ME (ed). New York: McGraw-Hill, Inc., 1992.
Proc SCAMC 1993	Proceedings of the Seventeenth Annual Symposium on Computer Applications in Medical Care. Safran C (ed). New York: McGraw-Hill, Inc., 1993.

Proc AAMSI Conf 1982	Proceedings of the First Annual Conference, American Association for Medical Systems & Informatics, Bethesda, MD: AAMSI, 1982.
Proc AAMSI Conf 1983	Proceedings of the Second Annual Conference, American Association for Medical Systems & Informatics, Bethesda, MD: AAMSI, 1983.
Proc AAMSI 1983	Proceedings of the AAMSI Congress on Medical Informatics, AAMSI Cong 83. Lindberg DAB, Van Brunt EE, Jenkins MA (eds). Bethesda, MD: 1983.
Proc AAMSI 1984	Proceedings of the Congress on Medical Informatics, AAMSI Congress 84. Lindberg DAB, Collen MF (eds). Bethesda, MD: AAMSI, 1984.
Proc AAMSI 1985	Proceedings of the Congress on Medical Informatics, AAMSI Congress 85. Levy AH, Williams BT (eds). Washington, DC: AAMSI, 1985.
Proc AAMSI 1986	Proceedings of the Congress on Medical Informatics, AAMSI Congress 86. Levy AH, Williams BT (eds). Washington, DC: AAMSI, 1986.
Proc AAMSI 1987	Proceedings of the Congress on Medical Informatics, AAMSI Congress 87. Levy AH, Williams BT (eds). Washington, DC: AAMSI, 1987.
Proc AAMSI 1988	Proceedings of the Congress on Medical Informatics, AAMSI Congress 88. Hammond WE (ed). Washington, DC: AAMSI, 1988.
Proc AAMSI 1989	Proceedings of the Congress on Medical Informatics, AAMSI Congress 89. Hammond WE (ed). Washington, DC: AAMSI, 1989.
Proc AMIA 1982	Proceedings of the First AMIA Congress on Medical Informatics, AMIA Congress 82. Lindberg DAB, Collen MF, Van Brunt EE (eds) New York: Masson, 1982.
Proc AMIA 1994	Proceedings of the Eighteenth Annual Symposium on Computer Applications in Medical Care. Ozbolt JG (ed). JAMIA Symposium Supplement. Philadelphia: Hanley & Belfast, Inc., 1994.
Proc AMIA 1995	Proceedings of the Nineteenth Annual Symposium on Computer Applications in Medical Care. Gardner RM (ed). JAMIA Symposium Supplement. Philadelphia: Hanley & Belfast, Inc., 1995.
Proc AMIA 1996	Proceedings of the 1996 AMIA Annual Fall Symposium. Cimino JJ (ed). JAMIA Symposium Supplement. Philadelphia: Hanley & Belfast, Inc., 1996.
Proc AMIA 1997	Proceedings of the 1997 AMIA Fall Symposium. Masys DR (ed). JAMIA Symposium Supplement. Philadelphia: Hanley & Belfast, Inc., 1997.
Proc AMIA 1998	Proceedings of the 1998 AMIA Fall Symposium. Chute CG, (ed). JAMIA Symposium Supplement. Philadelphia. Hanley & Belfast, Inc., 1998.
Proc AMIA 1999	Proceedings of the 1999 AMIA Fall Symposium. Lorenzi NM, (ed). JAMIA Symposium Supplement. Philadelphia. Hanley & Belfast, Inc., 1999.
Proc AMIA 2000	Proceedings of the 2000 AMIA Fall Symposium. Overhage JM, (ed). JAMIA Symposium Supplement. Philadelphia. Hanley & Belfast, Inc., 2000.
Proc AMIA 2001	Proceedings of the 2001 AMIA Symposium. S. Bakken (ed). JAMIA Symposium Supplement. Philadelphia. Hanley & Belfast, Inc., 2001.
Proc AMIA 2002	Proceedings of the 2002 AMIA Fall Symposium. I. S. Kahane (ed). JAMIA Symposium Supplement. Philadelphia. Hanley & Belfast, Inc., 2002.

Proc AMIA Annu Symp 2003	Proceedings of the 2003 AMIA Fall Symposium.
Proc AMIA Annu Symp 2004	Proceedings of the 2004 AMIA Fall Symposium.
Proc AMIA Annu Symp 2005	Proceedings of the 2005 AMIA Fall Symposium.
Proc AMIA Annu Symp 2006	Proceedings of the 2006 AMIA Fall Symposium.
Proc AMIA Annu Symp 2007	Proceedings of the 2007 AMIA Fall Symposium.
Proc AMIA Annu Symp 2008	Proceedings of the 2008 AMIA Fall Symposium.
Proc AMIA Annu Symp 2009	Proceedings of the 2009 AMIA Fall Symposium.
Proc AMIA Annu Symp 2010	Proceedings of the 2010 AMIA Fall Symposium.
Proc AMIA TBI 2010	Proceedings of the AMIA Summit on Translational Bioinformatics (TBI), P. Tarczy-Hornoch (ed). March 10 12, 2010, San Francisco.
Proc AMIA CRI 2010	Proceedings of the AMIA Summit on Clinical Research Informatics (CRI), P. Embi (ed). March 12 13, 2010, San Francisco.
Proc MEDINFO 1974	Proceedings of the First World Conference on Medical Informatics, MEDINFO 74, Stockholm. Anderson J, Forsythe JM (eds.). Stockholm: GOTAB, 1974.
Proc MEDINFO 1977	Proceedings of the Second World Conference on Medical Informatics, MEDINFO 77, Toronto. Shires DB, Wolf H (eds). Amsterdam: North-Holland Pub Co., 1977.
Proc MEDINFO 1980	Proceedings of the Third World Conference on Medical Informatics, MEDINFO 80, Tokyo. Lindberg DAB, Kaihara S (eds). Amsterdam: No-Holland Pub Co., 1980.
Proc MEDINFO 1983	Proceedings of the Fourth World Conference on Medical Informatics, MEDINFO 83, Amsterdam. Van Bemmel JH, Ball MJ, Wigertz O (eds). Amsterdam: North-Holland Pub Co., 1983.
Proc MEDINFO 1986	Proceedings of the Fifth World Conference on Medical Informatics, MEDINFO 86, Washington. Salamon R, Blum BI, Jorgensen M (eds). Amsterdam: North-Holland Pub Co., 1986.
Proc MEDINFO 1989	Proceedings of the Sixth World Conference on Medical Informatics, MEDINFO 89, Beijing. Barber B, Cao D, Qin D, Wagner G (eds). Amsterdam: North-Holland Pub Co., 1989.
Proc MEDINFO 92	Proceedings of the Seventh World Congress on Medical Informatics, MEDINFO 92, Geneva. Lun KC, Degoulet P, Piemme TE, Rienhoff O (eds). Amsterdam: North-Holland Pub Co., 1992.
Proc MEDINFO 95	Proceedings of the Eighth World Congress on Medical Informatics, MEDINFO 95, Vancouver, BC. Greenes RA, Peterson HE, Protti DJ (eds). Amsterdam: North-Holland Pub Co., 1995.
Proc MEDINFO 98	Proceedings of the Ninth World Congress on Medical Informatics, MEDINFO 98, Seoul, Korea. Cesnik B., McCray, AT, Scherrer JR (eds). Amsterdam: IOS Press, 1998.
Proc MEDINFO 2001	Proceedings of the Tenth World Congress on Medical Informatics, MEDINFO 2001,
Proc MEDINFO 2004.	Proceedings of the Eleventh World Congress on Medical Informatics, MEDINFO 2004,
Proc MEDINFO 2007	Proceedings of the Twelfth World Congress on Medical Informatics, MEDINFO 2007,
Proc MEDINFO 2010	Proceedings of the thirteenth World Congress on Medical Informatics, MEDINFO 2010,

Chapter 1
Prologue: The Evolution of Computer Databases

Databases have sometimes been called data banks, since like money banks that collect, store, use, exchange, and distribute money, data banks and databases collect, store, use, exchange, and distribute data. In this book the term, *data*, may include a single datum, like the number, 6; or the letter, a; or the symbol +; or combinations of these such as in a collection of facts or statistics; or information stored as textual natural-language data; or analog signals like phonocardiograms, electrocardiograms; or as visual images like x-rays. In this book a database can be more than a collection of data, since it can be an aggregate of information and knowledge, where information is a collection of data, and knowledge is a collection of information. Coltri (2006) wrote that the heart and the brain of a modern information system resides in its databases; and medical databases are especially complex because of the great diversity of medical information systems with their many different activities, their variety of medical services and clinical specialties with their computer-based subsystems; and with all of these actively changing and expanding in the ever-changing health-care environment. A database-management system is required to capture and process all of these data, and to implement all of the required functions of its database (Collen and Ball 1992).

To fully appreciate the historical development of computer-stored medical databases, it is helpful to have some knowledge of the evolution of informatics for databases, of the development of the computer hardware and software, and of the communications technology that are essential to fully exploit the remarkable capabilities of databases. This chapter very briefly describes some of the early important developments that led to modern computer-stored databases. The development of medical databases themselves is described in Chap. 2; and the great variety of medical databases that subsequently evolved in these six decades is described in later chapters.

1.1 The Evolution of Digital Computing

In the 1890s John S. Billings, a physician and the Director of the Army Surgeon General's Library (later to become the National Library of Medicine) initiated a series of events that led to the conceptual foundation for the development of medical

M.F. Collen, *Computer Medical Databases*, Health Informatics,
DOI 10.1007/978-0-85729-962-8_1, © Springer-Verlag London Limited 2012

informatics in the United States (Augarten 1984). As an advisor to the Census Bureau for the 1880 and the 1890 census, Billings advised Herman Hollerith, an engineer, that there should be a machine for tabulating statistics; and he suggested using punched paper cards. In 1882 Hollerith prepared paper cards (the size of dollar bills so he could store them in U.S. Treasury filing cabinets) with 288 locations for holes. He built machines for electrically punching holes in the cards in appropriate locations for numbers, letters, and symbols; and he invented machines for automatically reading the punched cards and tabulating the data. Hollerith generated the first computerized database for the 1890 census. T. Watson Sr. joined Hollerith's Automated Tabulating Machines Company; and in 1924 Watson took over the company and changed its name to the International Business Machines (IBM) Corporation; and initiated the development of computer hardware and software. *Informatics* was the term developed to satisfy the need for a single overall word to represent the domain of computing, information science and technology, and data communications. In 1968 A. Mikhailov, in the Scientific Information Department of the Moscow State University, published a book with the word *informatika* in its title (Mikhailov et al. 1976). In 1968 an article was published in the French literature with the word 'informatique' in its title (Pardon 1968). The derived English word, *informatics*, first appeared in print in the *Proceedings of MEDINFO 1974* (Anderson and Forsythe 1974). Variations of the term evolved, such as *bioinformatics* (Altman 1998).

Electronic digital computers began to be described in the scientific literature in the 1950s. Blum (1986a, b) noted that in the 1940s the word *computer* was a job title for a person who used calculators that usually had gears with ten teeth so calculations could be carried out to the base-ten. In the 1950s the term began to be applied to an *electronic digital computer*. In 1942 the first electronic digital computer was reported to be built in the United States by J. Atanasoff, a physicist at Iowa State University (Burks and Burks 1988; Mackintosh 1988). In 1943 the Electronic Numerical Integrator and Calculator (ENIAC) was built by J. Mauchly, J. Eckert, and associates at the University of Pennsylvania; and it is also considered by some to be the first electronic digital computer built in the United States (Rosen 1969). ENIAC performed sequences of calculations by rewiring its circuits for each sequence; and gunners in World War II used it to calculate trajectories of shells. In 1945 J. vonNeumann, at the Princeton Institute for Advanced Study, devised a method for storing the operating instructions as well as the data to be used in calculations (Brazier 1973). In the late 1940s J. Mauchly and associates used von Neumann's stored-program technology, that made possible high-speed computer processing, to build the Universal Automatic Computer (UNIVAC) that used 5,000 vacuum tubes as on-off switches so that calculations were then carried out to the base-2. In 1949 the Electronic Discrete Variable Automatic Computer (EDVAC) was built by the Moore School of Electrical Engineering; it used the internally stored-programs that had been developed by J.von Neuman; and it was an improvement over the UNIVAC (Campbell-Kelly 2009). The UNIVAC and the EDVAC were the first commercially available computers in the United States. In 1951 UNIVAC was transferred to Remington Rand, and was used by the U. S. Census Bureau to complete the 1950 census. In 1948 IBM began to market its first

commercial computer, the IBM 604, with 1,400 vacuum tubes and a plug board for wiring instructions. In 1952 IBM built its 701 computer, with 4,000 vacuum tubes, that was used in the Korean War; and in 1954 IBM built the 704 computer using FORTRAN programming (Blum 1983).

Magnetic core memory was invented in 1949 by A. Wang at the Harvard Computation Laboratory. In 1953 J. Forrester at the Massachusetts Institute of Technology, fabricated the magnetic cores from ferrite mixtures and strung them on three-dimensional grids; and magnetic core was the basic element of computer primary memory until the invention of the microchip in the 1960s (Augarten 1984). In 1956 IBM developed its IBM 704 computer with magnetic core memory, FORTRAN programming, a cathode-ray monitor, and some graphics capability; and it was one of the earliest computers used for biomedical research (Reid-Green 1979). *Random-access memory* (RAM) chips became commonly used for primary main memory in a computer because of their high speed and low cost. The earliest secondary storage devices for computer data used drives of reels of magnetic tape to sequentially record and store digital data. In the late 1940s magnetic disc drives became available that made possible direct random access to store and retrieve indexed data. The earliest small digital compact-disc (CD) was developed to store primarily audio material; but in the 1990s the magnetic compact disc, read-only memory (CD-ROM) became popular because of its high-density storage capacity. Laser-reflective, optical-storage discs were developed (Schipma et al. 1987), that by 2010 were used in wireless high-definition (Wi-Fi) blu-ray compact disc players. In the 2000s *flash* (*thumb*) *drives* for storage and for memory were developed that consisted of small printed circuit boards. Low-cost storage could be easily added to a computer by plugging in a flash drive with a Universal Serial Bus (USB).

Transistors were invented by W. Shockley and associates at Bell Laboratories in 1959, and initiated the second generation of electronic digital computers when IBM began marketing its first transistorized computer, the IBM 7090 (Blum 1983). In 1959 J. Kilby at Texas Instruments and R. Noyce at Fairchild Semiconductors independently made the silicon crystal in a transistor serve as its own circuit board, and thereby created the first integrated circuit on a chip (Noyce 1977; Boraiko 1982). In 1961 Fairchild at Texas Instruments introduced logic chips that in addition to the arithmetic *AND* function could also perform Boolean *OR* and *NOT*.

Minicomputers were first developed in 1962 by W. Clark and C. Molnar at the Lincoln Laboratory of the Massachusetts Institute of Technology (MIT); and it was a small special-purpose computer called the "Laboratory Instrument Computer" (LINC) (Clark and Molnar 1964). In 1964 the Digital Equipment Company (DEC) began the commercial production of the LINC (Hassig 1987). C. Bell designed DEC's first Programmed Data Processor (PDP); and by 1965 DEC's PDP-8 led in the use of minicomputers for many medical applications since it could outperform large mainframe computers for certain input/output processing tasks, and at a lower cost (Hammond and Lloyd 1972).

Third-generation computers appeared in 1963 using solid-state integrated circuits that employed large-scale integration (LSI) consisting of hundreds of transistors, diodes, and resistors that were embedded on one or more tiny silicon chips

(Blum 1986a). In 1964 IBM introduced its system 360-series that allowed data processing operations to grow from a smaller machine in its 360-series to a larger one in its 370-series without the need to rewrite essential programs. By the late 1960s the fourth-generation of computers employed very-large-scale integration (VLSI) that contained thousands of components on very tiny silicon chips (Boraiko 1982). Soon magnetic primary-core memory was replaced with semiconductor, random-access memory (RAM) chips; and by the early 1970s IBM's system/370 series used only integrated circuit chips. In 1965 S. Cray at Control Data Corporation (CDC), designed its CDC 6600 computer that contained six computer processors working in parallel. It was the most powerful computer at the time and was considered to be the first super-computer (Runyan 1987).

Microprocesssors were developed in 1968 when R. Noyce left Fairchild Semiconductors to begin a new company called Intel; and produced the Intel 2008 that was an 8-bit microprocessor which sold at a price of $120 each, and required 50–60 additional integrated circuits to configure it into a minimum system. In 1973 Intel's 8080 microprocessor was introduced and it required only five additional circuit devices to configure a minimum system. The next Intel 8748 was also an 8-bit microprocessor; but it was considered to be a microcomputer since it incorporated some read-only memory (ROM) chips (Titus 1977). In 1969 M. Hoff fabricated at Intel the first central processing unit on a single silicon chip. Intel then developed a series of microprocessor chips that revolutionized the personal computer industry. In 1970 G. Hyatt filed a patent application for a prototype microprocessor using integrated circuits. In 1971 J. Blankenbaker assembled what is generally credited as being the first personal computer (Bulkeley 1986). Further development in the 1970s led to large-scale integration with tens-of-thousands of transistors on each chip. In 1975 Intel's 8080 microprocessor was the basis for the Altair 8800 that became the first commercial personal computer. In 1976 S. Jobs and S. Wozniak founded the Apple Computer Company, and designed the first Apple computer that used the Motorola 6502 chip. In 1984 the Apple Macintosh computer contained a Motorola 68000 central processor chip and used a Smalltalk-like operating system; and employed some of the features that had been developed at the Xerox Palo Alto Research Center (PARC) that included: a mouse pointing device, the ability to display symbols and icons representing files and documents; and provided a graphical-user-interface (GUI); and it could support applications with multiple windows of displays-within-displays (Miller 1984; Crecine 1986). The power of a microprocessor is greatly influenced by the number of transistors it contains on a chip, and whether they are connected to function in series or in parallel. The earliest chips functioning as a central-processing unit (CPU) had a small number of cores of transistors, with each core performing a task in series in assembly-line style; and they were used for running operating systems, browsers, and operations requiring numerous decisions. In 1980 Intel's 8080 chips contained 2,300 transistors. In 1981 IBM introduced its Personal Computer (IBM PC) that used the Microsoft DOS operating system and the Intel 8088 chip it had introduced in 1979 that was a 16-bit processor containing 29,000 transistors and performed 60-thousand operations-per-second (KOPS), the equivalent of 0.06-millions of instructions-per-second (MIPS). In 1986

Intel's 80386 contained 750,000 transistors; in 1989 its 80486 was a 32-bit processor that contained 1.2 million transistors; in 1992 its Pentium chip contained 3.1 million transistors; in 2002 its Pentium 4 had 55 million transistors; and in 2006 Intel's dual-core chip contained 291 million transistors.

Parallel processing units were developed in the late 1990s as multi-core processor chips became available. As the number of cores-per-chip increased, then transactional memory techniques evolved that allowed programmers to mark code segments as transactions, and a transactional memory system then automatically managed the required synchronization issues. Minh et al. (2008) and associates at Stanford University developed the Stanford Transactional Applications for Multi-Processing (STAMP) to evaluate parallel processing with transactional memory systems by measuring the transaction length, the sizes of the read-and-write sets, the amount of time spent in transactions, and the number of retries per transaction. With increasing computer memory and data storage capabilities, databases rapidly evolved to store collections of data that were indexed to permit adding, querying, and retrieving from multiple, large, selected data sets. In 1999 a single chip processor that functioned as a graphics-processing unit (gpu) with numerous cores that simultaneously processed data in parallel, was marketed by NVIDIA as GeForce 256; and it was capable of processing 10-million polygons per second. By 2010 NVIDIA had a product line called TESLA, with a software framework for parallel processing called CUDA; and NVIDIA marketed its NV35 graphics-processing unit with a transistor count of about 135-million that could process very large calculations in 2 min that had previously taken up to 2 h. Graphics-processing units are much better at processing very large amounts of data, so they are increasingly used for high-definition video and for 3-dimensional graphics for games. *Computer graphics* was defined by Fung and Mann (2004) as image synthesis that takes a mathematical description of a scene and produces a 2-dimensional array of numbers which is an image; and Fung differentiated it from *computer vision* that is a form of image analysis that takes a 2-dimensional image and converts it into a mathematical description. In 2010 the Advanced Micro Devices (AMD) Opteron 6100 processor, a core package of two integrated circuits, contained a total of more than 1.8-billion transistors. Traditionally, a central-processing unit processed data sequentially; whereas a parallel-processing unit divided large amounts of similar data into hundreds or thousands of smaller collections of data that were processed simultaneously. In 2010 a graphics-processing unit could have about 3-billion transistors, as compared to about 1-billion for a central-processing unit. Further advances were occurring in the development of multi-core, parallel processing, transactional-memory chips for creating general-purpose, high-speed, parallel-processing computers. The evolving hybrid combinations of central-processing units and embedded graphics-processing units that were called integrated-graphics processors, or high-performance units, or even called personal desk-top supercomputers, were expected to greatly increase computational efficiency and at a much lower cost (Toong and Gupta 1982; Wikipedia 2010b; Villasenor 2010).

Computer software development methodology was advocated by Wasserman (1982) to cover the entire software development cycle, and support transitions between phases of the development cycle; and to support validation of the system's

correctness throughout the development cycle to its fulfilling system specifications and meeting its user needs. Although advances in computer hardware were the basis for many innovations, the software made the hardware usable for computer applications. *Computer programming languages* were defined by Greenes (1983) as formal languages used by humans to facilitate the description of a procedure for solving a problem or a task, and which must be translated into a form understandable by the computer itself before it could be executed. *Algorithms* are commonly used in computer programming as a method for providing a solution to a particular problem or a set of problems; and consist of a set of precisely stated procedures that can be applied in the same way to all instances of a problem. For complex problems, such as data mining (see Sect. 8.2), algorithms are indispensable because only those procedures that can be stated in the explicit and unambiguous form of an algorithm can be presented to a computer (Lewis and Papadimitriou 1978). For ENIAC, the first electronic digital computer, the programs of instructions were wiring diagrams that showed how to set the machine's plug boards and switches. In 1945 J. von Neumann demonstrated that instructions for the computer could be stored in the computer's electronic memory and treated in the same manner as data (Brazier 1973). The first machine language was a series of binary numbers that addressed memory cells for storing data; and used accumulators for adding and subtracting numbers; and then storing them in registers. To reduce the tedium of writing in machine code, programmers soon invented an assembly language so that commonly used English words, such as *add* or *load*, would be translated automatically into the appropriate machine code instructions. In subsequent higher-level languages, one English statement could give rise to many machine instructions; and programs tended to be shorter, quicker to write; were less prone to error, and had the ability to run on different computers (Davis 1977).

FORTRAN (FORmula TRANslator) was developed in 1957 by J. Backus and associates at International Business Machines (IBM), and it soon became the standard language for scientific and engineering applications. COBOL (COmmon Business-Oriented Language) was created in 1960 by a joint committee of computer manufacturers and users, government and academic representatives interested in developing a high-level language that would use ordinary English statements for business data processing. By the 1980s COBOL was one of the most commonly used programming languages. BASIC (Beginners All-purpose Symbolic Instruction Code) was developed in 1964 by J. Kemeny and T. Kurtz at Dartmouth, as a language modeled after FORTRAN, to be used for teaching computer programming. BASIC was used in 1975 by W. Gates and P. Allen to program the ALTAIR computer that led to the founding of Microsoft (Gates 1989).

MUMPS (Massachusetts General Hospital Utility Multi-Programming System) was developed in 1966 by N. Pappalardo and associates in G. Barnett's Laboratory of Computer Science at the Massachusetts General Hospital. MUMPS provided an operating system, a database-management system for handling large volumes of information, and an easy interactive mode for programmer-computer communication (Barnett et al. 1981). In 1969 MUMPS became commercially available by Pappalardo's Medical Information Technology, Inc. (Meditech); and MUMPS was

soon the most commonly used programming language in the United States for medical computing applications. MUMPS provided an excellent structure for medical databases with all their complexity; and in the 1980s both the Department of Defense and the Veterans Hospitals began installing MUMPS-based medical information systems; and in the 2000s the popular Epicare medical information systems was also Mumps-based (see also Sect. 4.2). Pascal programming language was developed in the early 1970s by N. Wirth using structured programming. Versions of Pascal were used in the 1980s by Apple computers, and also for the IBM 370 system; and in the 1990s it was the basis for Oracle's language PL/SQL. The Smalltalk language was developed in the 1970s by A. Kay and associates at Xerox's Palo Alto Research Center (PARC) for their Alto computer. Smalltalk provided a graphical-user interface (GUI) that could move displayed text and images by using a mouse pointer (Kay 1984).

Structured Query Language (SQL) was developed in the early 1970s by Chamberlin and Boyce (1974) and R. Boyce at IBM, as a language designed for the query, retrieval, and management of data in a relational database-management system, such as had been introduced by Codd (1970). In the 1980s the relational database design became dominant in industry; and versions of SQL were generally used to construct, manage, and query relational databases (VanName and Catchings 1989) (See also Sect. 2.2). C-language was developed in the mid-1970s by D. Ritchie and K.Thompson at Bell Laboratories, as a structured-programming language that used block structures of statements and could be used for object-oriented programming (Kernighan and Ritchie 1988). In the mid-1980s a new version of C-language called C++ began to be used for large-scale software development; and by the 2000s it was one of the most common languages used for commercial-grade software; and soon other specialized third-generation languages were developed (Blum 1986b). dBASE was developed by Ashton-Tate as a database-management system for microcomputers; and dBase II was used by the Apple computer, and by the IBM personal computer under Microsoft's DOS; and dBase III was used by UNIX. The language PERL was developed in 1987 with some of the features of C-language; and it was widely used for building Web-based applications, for interfacing and accessing database modules, for generating SQL queries; and also was used for text processing (Chute et al. 1995). Java language was developed in the 1990s by Sun Microsystems as an object-oriented, high-level programming language and, was used for a variety of operating systems including Apple Macintosh, Linux, Microsoft Windows, and Sun Solaris.

Markup languages had begun to evolve in the 1960s when Generalized Markup Language (GML) was developed by IBM to enable the sharing of machine-readable, large-project documents used in industry, law, and in government. In 1986 Standard Generalized Markup Language (SGML) was developed as an International Standards Organization (ISO) version of GML, and was used by industry and the Armed Services. In 1996 SGML began to be used for Web applications; and in 1998 it was modified as Extensible Markup Language (XML) that was designed to provide a standard set of rules for encoding documents in machine-readable form, and to help simplify and support the usability of Web services. Hypertext Markup

Language (HTML), with some features derived from SGML, was developed in 1990 by T. Berners-Lee, while at CERN. HTML could be used by Web browsers to dynamically format text and images; it became the predominant markup language for describing Web pages; and in 2000 HTML became an international standard (Wikipedia 2010a).

Computer operating systems were initially sets of routines for data input and output; such as consisting of a few-hundred machine instructions for storing binary codes from punched paper tape into successive memory locations. In the 1950s operating systems ran the programs submitted by users in batch modes. In the 1960s time-sharing programs were developed that switched rapidly among several user programs; and could give the impression that the programs were being executed simultaneously. In 1969 K. Thompson and associates at AT&T Bell Laboratories developed the UNIX operating system; a powerful time-sharing operating system that was multi-user (it could serve more than one user at a time), multi-tasking (it could run several applications at the same time), and with open-architecture (useable by different vendor's computers). By 1983 about 80-percent of colleges that granted computer science degrees had adopted UNIX; and several versions of UNIX had evolved (Lockwood 1990). In early 1987 SUN Microsystems joined with AT&T to create a new version of UNIX with a graphical-user interface and used Internet protocols. Since UNIX could run on a large number of different computers, singly or in a network, including IBM compatibles and Apple Macintoshes, UNIX became a major competitor for the operating systems of networks of desktop computers and workstations. In 1974 the first microcomputer operating system, Control Program for Microcomputers (CP/M), was developed for 8-bit microprocessors by G. Kildall, the founder of Digital Research. CP/M contained an important module called BIOS (Basic Input/Output Subsystem) that applications programs and operating systems have continued to use to interface with their hardware components. By the end of the 1970s, CP/M was used world-wide (Kildall 1981).

In the early 1980s IBM needed an operating system for its new 16-bit microprocessor, its Personal Computer (IBM-PC); and IBM contracted with W. Gates to develop the Microsoft Disk-Operating System (MS-DOS) (Cringely 1992). In the 1980s MS-DOS became the most widely used operating system in the nation for IBM- compatible personal computers. In the late 1980s W. Gates independently developed an operating system called MS-Windows; and Gates separated from IBM that continued the development of its IBM-OS/2. The Apple Macintosh appeared in 1984 using a Smalltalk operating system with a graphical-user interface that permitted the use of displayed menus (lists of options available for selection), and icons (symbols representing options) from which the user could select items by pointing and clicking a mouse pointer. In May 1990 Microsoft announced its MS-Windows 3.0 operating system, that employed a graphical-user interface and a mouse-pointer selector such as was used by the Apple Macintosh; and it also provided some networking capabilities. By the mid-1990s Microsoft's Windows 95 outsold IBM's OS/2; and MS-Windows became the operating system most commonly used in personal computers. In 1989 L.Torvalds, a student at the University of Helsinki, and R. Stallman released an operating system that supported the functionality of UNIX

called LINUX; and it was made freely available to the public on the condition that its users would make public all of their changes as well. LINUX Online (http://www.linux.org/) provided a central location from which users could download source code, submit code fixes, and add new features. In 1999 Version 2.2 of the LINUX kernel was released and was shared by all LINUX distributors; and this core component of its operating system supported multiple users, multitasking, networking and Internet services, and some 64-bit platforms. In 1999 it already had more than 1,000 contributors and about 7-million LINUX users; and it became a competitor to MS Windows (Seltzer 1999).

1.2 The Evolution of Data Input and Output Devices

Data acquisition, data input, data retrieval, and data output are all challenging basic functions of a medical database; and the various devices used to carry out these functions changed greatly through these decades with innovations in technology. *Punched paper cards* were the earliest mode for entering data into to a computer. Punched cards were invented by H. Hollerith in 1882 (Warner 1979; Augarten 1984); and he invented a machine that punched a hole into a paper card in a specific location that corresponded to a digital code for each alphabet letter, for each number, and for each symbol that was selected. The punched paper cards were then passed through card readers that sensed the holes by wires that brushed over the cards, and thereby made electrical connections to the metal plate under which the cards were passed. Schenthal (1960, 1963) also used mark-sense paper cards, on which a mark was made by a graphite pencil instead of generating a punched hole, and the mark could be electrically sensed as data input to a computer. Schenthal also used portable-punch cards by using prescored cards, and instead of machine punching holes, or marking the desired response with a pencil, one could punch-out with a stylus the appropriate prescored hole and thereby produced a directly readable punched card. Prepunched cards, prepared with specific data items for computer input, were often used for requisitioning clinical laboratory tests, and also used for entering patients' responses to a questionnaire (Collen 1978). Soon data-entry devices that used electronic readers for punched-paper tape followed the use of punched cards. Punched paper cards and paper tape became unnecessary for data input when keyboard devices, structured like a typewriter but with additional special-function keys, were directly connected to computers; and when a key was pressed then an electric circuit was closed and sent a corresponding specific digital code to the computer.

Optical character readers (OCR) were developed in the 1970s that could scan documents, and read and input alphanumeric characters that were printed in standard fonts of type. Optical character scanners contained light sensors that converted light into an electrical voltage that could be sensed by an electronic circuit. OCR recognized the shape of the characters by the contrast of light and dark areas created when light was reflected from the surface of the document, and converted to a bit-map of