



Stewart

EDGAR F. CODD

1923-2003

Elected in 1981

“For the origination of the relational approach to the organization of large data bases.”

BY C.J. DATE

SUBMITTED BY THE NAE HOME SECRETARY

BY NOW ALMOST EVERYONE in the database community is aware that Dr. E.F. Codd passed away on April 18, 2003, at the age of 79. Dr. Codd, known universally as Ted to his colleagues and friends—among whom I am proud to count myself—single-handedly put the field of database management on a solid scientific footing. The entire relational database industry, now worth many billions of dollars a year, owes its existence to Ted’s original work; the same is true of all relational database research and teaching programs in universities and similar organizations worldwide. Indeed, all of us who work in this field owe our careers and livelihoods to Ted’s contributions from the late 1960s to the early 1980s. This tribute to Ted and his achievements is offered in recognition of the great debt we all owe him.

Ted began his computing career in 1949 as a programming mathematician for IBM working on the selective-sequence electronic calculator. He subsequently participated in the development of several important IBM products, including the 701 (IBM’s first commercial electronic computer) and STRETCH, which led to IBM’s 7090 mainframe technology. Then, in the late 1960s, he turned his attention to the problem of database management—and over the next few years he created the relational model of data, with which his name will forever be associated.

The relational model is widely recognized as one of the great technical innovations of the twentieth century. Ted described it and explored its implications in a series of staggeringly original research papers published between 1969 and 1981. The effect of those papers was twofold: First, they changed for good the way the IT world perceived the database management problem; second, they laid the foundation for a whole new industry. In fact, they provided the basis for a technology that has had, and continues to have, a major impact on the very fabric of our society. It is no exaggeration to say that Ted is the intellectual father of the modern database field.

To give an idea of the extent of Ted's accomplishments, I will briefly survey some of his most significant contributions. The biggest of all was, of course, making database management into a science (thereby introducing clarity and rigor in the field). The relational model provided a theoretical framework within which a variety of important problems could be attacked scientifically. Ted first described his model in an IBM Research Report (RJ599) published on August 19, 1969, *Derivability, Redundancy, and Consistency of Relations Stored in Large Data Banks*. The following year he published a revised version of this paper, "A Relational Model of Data for Large Shared Data Banks" (*Communications of the ACM* 13(6): 377-387), which is usually credited with being the seminal paper in the field.

Most of the novel ideas described in outline in the following paragraphs, as well as numerous subsequent technical developments, were foreshadowed in these first two papers; some of these ideas have still not been fully explored. In my opinion, everyone professionally involved in database management should read, and reread, at least one of these papers every year.

Incidentally, it is not as widely known that Ted not only invented the relational model in particular, he invented the whole concept of a data model in general (cf., "Data Models in Database Management," *ACM SIGMOD Record* 11, No. 2 (February 1981)). For both the relational model and data models in general, he stressed the importance of the distinction between a data model and its physical implementation.

Ted recognized the potential of using predicate logic as a foundation for a database language. He discussed this possibility briefly in his 1969 and 1970 papers and then, using the predicate logic idea as a basis, went on to describe in detail what was probably the very first relational language to be defined, Data Sublanguage ALPHA, in “<http://www.informatik.uni-trier.de/~ley/db/conf/sigmod/Codd71.html> A Data Base Sublanguage Founded on the Relational Calculus” (Proceedings of 1971 ACM-SIGFIDET Workshop on Data Description, Access and Control, San Diego, Calif., November 11-12, 1971). Although ALPHA was never implemented, it was extremely influential on certain other languages, especially the Ingres language QUEL and to a lesser extent SQL.

Ted subsequently defined the relational calculus more formally, as well as the relational algebra, in “<http://www.informatik.uni-trier.de/~ley/db/labs/ibm/RJ987.html> Relational Completeness of Data Base Sublanguages,” in *Database Systems: Courant Computer Science Symposia Series 6*, edited by Randall J. Rustin (Prentice-Hall, 1972). As the title indicates, this paper also introduced the notion of relational completeness as a basic measure of the expressive power of a database language. It also described an algorithm—Codd’s reduction algorithm—for transforming an arbitrary expression of the calculus into an equivalent expression in the algebra, thereby proving the algebra was relationally complete (i.e., it was at least as powerful as the calculus) and providing a basis for implementing the calculus.

Ted also introduced the concept of functional dependence and defined the first three normal forms (1NF, 2NF, 3NF) in “Normalized Data Base Structure: A Brief Tutorial” (Proceedings of 1971 ACM-SIGFIDET Workshop on Data Description, Access and Control, San Diego, Calif., (November 11-12, 1971) and “Further Normalization of the Data Base Relational Model,” in *Data Base Systems: Courant Computer Science Symposia Series 6*, edited by Randall J. Rustin (Prentice-Hall, 1972). These papers laid the foundations for the field of what is now known as dependency theory, a branch of database science. Among other

things, it established a basis for a truly scientific approach to the problem of logical database design.

Ted defined the key notion of essentiality in “Interactive Support for Nonprogrammers: The Relational and Network Approaches,” *Proceedings of the ACM SIGMOD Workshop on Data Description, Access, and Control, Vol. II*, Ann Arbor, Mich. (May 1974). This paper was Ted’s principal written contribution to “The Great Debate”—the official title was “Data Models: Data-Structure-Set vs. Relational”—a special event at the 1974 SIGMOD Workshop subsequently characterized by Robert L. Ashenhurst as “a milestone event of the kind too seldom witnessed in our field.”

The concept of essentiality introduced by Ted in this debate is a great aid to clear thinking in discussions on the nature of data and database management systems. The Information Principle (which I heard Ted refer to one occasion as the fundamental principle underlying the relational model) relies on it, albeit not very explicitly: “The entire information content of a relational database is represented in one and only one way: namely, as attribute values within tuples within relations.”

In addition to all of his research activities, Ted was active professionally in other areas. He founded the ACM Special Interest Committee on File Description and Translation (SICFIDET), which later became an ACM Special Interest Group (SIGFIDET) and changed its name to the Special Interest Group on Management of Data (SIGMOD). He was also tireless in his efforts, both inside and outside IBM, to obtain acceptance for the relational model.

Ted’s achievements with the relational model should not eclipse his original contributions in several other important areas, such as multiprogramming. He led the team that developed IBM’s first multiprogramming system and reported on that work in: “<http://www.informatik.uni-trier.de/~ley/db/journals/cacm/Codd59.html> Multiprogramming STRETCH: Feasibility Considerations” (with “http://www.informatik.uni-trier.de/~ley/db/indices/a-tree/l/Lowry:E=_S=.html” E.S. Lowry, “<http://www.informatik.uni-trier.de/~ley/db/indices/a-tree/m/McDonough:E=.html>” E. McDonough, and

["http://www.informatik.uni-trier.de/~ley/db/indices/a-tree/s/Scalzi:Casper_A=.html"](http://www.informatik.uni-trier.de/~ley/db/indices/a-tree/s/Scalzi:Casper_A=.html) C.A. Scalzi), *Communications of the ACM* 2(11): 13-17 (November 1959), and "Multiprogram Scheduling," Parts 1 and 2, *Communications of the ACM* 3(6) (June 1960); Parts 3 and 4, *Communications of the ACM* 3(7) (July 1960). In addition, his work on natural language processing was described in several publications, including "Seven Steps to Rendezvous with the Casual User," in *Data Base Management, Proceedings of the IFIP TC-2 Working Conference on Data Base Management 1974*, edited by J.W. Klimbie and K.K. Koffeman (North-Holland, 1974).

The depth and breadth of Ted's contributions are reflected in the long list of honors conferred on him during his lifetime. He was an IBM fellow, an ACM fellow, and a fellow of the British Computer Society. He was also an elected member of both the National Academy of Engineering and the American Academy of Arts and Sciences. And, in 1981, he received the ACM Turing Award, the most prestigious award in the field of computer science. He also received numerous other professional awards.

Ted Codd was a genuine pioneer and an inspiration to everyone who had the good fortune and honor to know him and work with him. He was always scrupulous about crediting other people's contributions, and, despite his huge achievements, he was careful never to make extravagant claims. For example, he would never claim that the relational model could solve all possible problems or that it would last forever. Yet those who truly understand that model believe that the class of problems it can solve is extraordinarily large and that it will endure for a very long time. Systems will be built on the basis of Codd's relational model as far out as anyone can see.

A native of England, Ted served in the Royal Air Force during World War II. He moved to the United States after the war and became a naturalized U.S. citizen. He held M.A. degrees in mathematics and chemistry from Oxford University and an M.S. and a Ph.D. in communication sciences from the University of Michigan. He is survived by his wife, Sharon; a daughter, Katherine; three sons, Ronald, Frank, and David;

six grandchildren; other family members; and friends and colleagues around the world. He is mourned and sorely missed by all.

