

# Edgar F. Codd

August 23rd, 1923 - April 18th, 2003

## a tribute

by Chris Date

By now there cannot be many in the database community who are unaware that, sadly, Dr. E. F. Codd passed away on April 18<sup>th</sup>, 2003. He was 79. Dr. Codd, known universally to his colleagues and friends—among whom I was proud to count myself—as Ted, was the man who, single-handed, 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 the fact of its existence to Ted's original work, and the same is true of all of the huge number of relational database research and teaching programs under way worldwide in universities and similar organizations. Indeed, all of us who work in this field owe our career and livelihood to the giant contributions Ted made during the period from the late 1960s to the early 1980s. We all owe him a huge debt. This tribute to Ted and his achievements is offered in recognition of that debt.

Ted began his computing career in 1949 as a programming mathematician for IBM 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 invention with which his name will forever be associated: the relational model of data.

The relational model is widely recognized as one of the great technical innovations of the 20<sup>th</sup> century. Ted described it and explored its implications in a series of research papers—staggering in their originality—that he published during the period from 1969 to 1981. The effect of those papers was twofold: First, they changed for good the way the IT world perceived the database management problem; second (as already mentioned), 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.

Let me remind you of the extent of Ted's accomplishments by briefly surveying some of the most significant of his contributions here. Of course, the biggest of all was, as already mentioned, to make database management into a science (and thereby to introduce a welcome and sorely needed note of clarity and rigor into the field): The relational model provided a theoretical framework within which a variety of important problems could be attacked in a scientific manner. Ted first described his model in 1969

in an IBM Research Report: “Derivability, Redundancy, and Consistency of Relations Stored in Large Data Banks,” IBM Research Report RJ599 (August 19<sup>th</sup>, 1969)

He also published a revised version of this paper the following year: “A Relational Model of Data for Large Shared Data Banks,” CACM 13, No. 6, June 1970 (This latter is usually credited with being the seminal paper in the field, though this characterization is a little unfair to its 1969 predecessor) and elsewhere (most of Ted’s papers were published in several places. Here I will just give the primary sources.)

Almost all of the novel ideas described in outline in the following paragraphs, as well as numerous subsequent technical developments, were foreshadowed or at least hinted at in these first two papers; what is more, some of them remain less than fully explored to this day. 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 as it should be that Ted not only invented the relational model in particular, he invented the whole concept of a data model in general. See his paper “Data Models in Database Management,” ACM SIGMOD Record 11, No. 2 (February 1981).

And in connection with both the relational model in particular and data models in general, he stressed the importance of the distinction—regrettably still widely underappreciated—between a data model and its physical implementation.

Ted also saw 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 “A Data Base Sublanguage Founded on the Relational Calculus,” Proc. 1971 ACM SIGFIDET Workshop on Data Description, Access and Control, San Diego, Calif. (November 1971). ALPHA as such was never implemented, but it was extremely influential on certain other languages that were, including in particular the Ingres language QUEL and (to a lesser extent) SQL as well.

Ted subsequently defined the relational calculus more formally, as well as the relational algebra, in “Relational Completeness of Data Base Sublanguages,” in Randall J. Rustin (ed.), DATA BASE SYSTEMS: COURANT COMPUTER SCIENCE SYMPOSIA SERIES 6 (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 (a) proving the algebra was relationally complete (i.e., it was at least as powerful as the calculus) and (b) 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). See the papers “Normalized Data Base Structure: A

Brief Tutorial,” Proc. 1971 ACM SIGFIDET Workshop on Data Description, Access, and Control, San Diego, Calif. (November 11<sup>th</sup>-12<sup>th</sup>, 1971), and “Further Normalization of the Data Base Relational Model,” in Randall J. Rustin (ed.), DATA BASE SYSTEMS: COURANT COMPUTER SCIENCE SYMPOSIA SERIES 6 (Prentice-Hall, 1972).

These papers laid the foundations for the entire field of what is now known as dependency theory, an important branch of database science in its own right (among other things, it established a basis for a truly scientific approach to the problem of logical database design).

Ted also defined the key notion of essentiality in “Interactive Support for Nonprogrammers: The Relational and Network Approaches,” Proc. ACM SIGMOD Workshop on Data Description, Access, and Control, Vol. II, Ann Arbor, Michigan (May 1974). This paper was Ted’s principal written contribution to “The Great Debate.” The Great Debate—the official title was Data Models: Data-Structure-Set vs. Relational—was a special event held at the 1974 SIGMOD Workshop; it was subsequently characterized in CACM by Robert L. Ashenurst 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 regarding the nature of data and DBMSs. In particular, The Information Principle (which I heard Ted refer to on 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 the research activities briefly sketched in the foregoing, Ted was professionally active in other areas as well. In particular, he founded the ACM Special Interest Committee on File Description and Translation (SICFIDET), which later became an ACM Special Interest Group (SIGFIDET) and subsequently 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 the level of acceptance for the relational model that he rightly believed it deserved—efforts that were, of course, eventually crowned with success.

Ted’s achievements with the relational model should not be allowed to eclipse the fact that he made major original contributions in several other important areas as well, including multiprogramming and natural language processing in particular. He led the team that developed IBM’s very first multiprogramming system and reported on that work in “Multiprogramming STRETCH: Feasibility Considerations” (with three coauthors), CACM 2, No. 11 (November 1959) and “Multiprogram Scheduling,” Parts 1 and 2, CACM 3, No. 6 (June 1960); Parts 3 and 4, CACM 3, No. 7 (July 1960)

As for his work on natural language processing, see among other publications the paper “Seven Steps to Rendezvous with the Casual User,” in J. W. Klimbie and K. L. Koffeman (eds.), DATA BASE MANAGEMENT, Proc. IFIP TC-2 Working Conference on Data Base Management (North-Holland, 1974)

The depth and breadth of Ted's contributions were recognized by the long list of honors that were 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 computing pioneer. He was an inspiration to all of us who had the fortune and honor to know him and work with him. It is a particular pleasure to be able to say that he was always scrupulous in giving credit to other people's contributions. Moreover—and despite his huge achievements—he was also careful never to overclaim; he would never claim, for example, that the relational model could solve all possible problems or that it would last forever. And yet those who truly understand that model do believe that the class of problems it can solve is extraordinarily large and that it will endure for a very long time. Systems will still be being built on the basis of Codd's relational model for as far out as anyone can see.

Ted was a native of England and a Royal Air Force veteran of World War II. He moved to the United States after the war and became a naturalized US citizen. He held MA degrees in mathematics and chemistry from Oxford University and MS and PhD degrees in communication sciences from the University of Michigan. He is survived by his wife Sharon; a daughter, Katherine; three sons, Ronald, Frank, and David; and six grandchildren. He also leaves other family members, friends, and colleagues all around the world. He is mourned and sorely missed by all.

A memorial event to remember and celebrate Ted's life and achievements will be held in Silicon Valley later this year.