The Changing of the Database Guard

Michael L. Brodie, research scientist, CSAIL, MIT

You can be right there at a historic moment and yet not see its significance for decades. I now recount one such time when the leadership of the database community and its core technology began to change fundamentally.

Dinner with the Database Cognoscenti

After spending the summer of 1972 with Ted Codd at IBM's San Jose Research Lab, Dennis Tsichritzis, a rising database star, returned to the University of Toronto to declare to Phil Bernstein and myself that we would start Ph.Ds. on relational databases under his supervision. What could possibly go wrong?

In May 1974, I went with Dennis to the ACM SIGFIDET conference in Ann Arbor, Michigan, my first international conference, for the *Great Relational-CODASYL Debate* where Dennis would fight for the good guys. After the short drive from Toronto, we went to a "strategy session" dinner for the next day's debate. Dinner, at the Cracker Barrel Restaurant in the conference hotel, included the current and future database cognoscenti and me (a database know-nothing). It started inauspiciously with Cracker Barrel's signature, neon orange cheese dip with grissini ('scuse me, breadsticks).

I was quiet in the presence of the cognoscenti—Ted Codd, Chris Date of IBM UK Lab, and Dennis—and this tall, enigmatic and wonderfully confident guy, Mike something, a new University of California, Berkeley assistant professor and recent University of Michigan Ph.D. According to him, he had just solved the database security problem with QUEL, his contrarian query language. During dinner, Mike sketched a few visionary ideas. Further evidence for me to be quiet since I could barely spell datakbase [sic].

The *Great* Relational-CODASYL Debate

The cognoscenti who lined up for the debate were, on the relational side, Ted Codd; Dennis Tsichritzis; and Kevin Whitney, from General Motors, who had implemented RDMS [1], one of

the first RDBMSs. On the CODASYL side were Charlie Bachman, who was awarded the 1973 Turing Award "for his outstanding contributions to database technology"; J. R. Lucking, International Computers Limited, England; and Ed Sibley, University of Maryland and National Bureau of Standards.

The much-ballyhooed debate was less than three years after Codd's landmark paper [2]; one year after Charlie's Turing Award; one year into Mike's and Eugene Wong's pioneering Ingres project[1] at UC Berkeley; coincident with the beginning of the System R project[2] at IBM Research, San Jose; five years before the release of Oracle, the first commercial RDBMS, in 1979, followed in 1983 by IBM's DB2[3]; and almost a decade before Ted was awarded the 1981 Turing Award "for his fundamental and continuing contributions to the theory and practice of database management systems," specifically relational databases.

SIGFIDET 1974 [3] focused largely on record-oriented hierarchical and network databases. Relational technology was just emerging. Most significantly, SEQUEL (now SQL) was introduced [4]. Three papers discussed concepts and six[4] RDBMS implementations: IBM Research's XRM-An Extended (N-ary) Relational Memory, The Peterlee IS/1 System, and Rendezvous; Whitney's RDMS; also, ADMINS and the MacAIMS Data Management System. Mike's paper [6] on a core relational concept, like those of Codd, Date, and Whitney, showed a succinct and deep understanding of the new relational concepts, in contrast to the debate.

The much-anticipated debate was highly energized yet, in hindsight, pretty ho-hum, more like a tutorial as people grappled with new relational ideas that were so different from those prevalent at the time. The 23-page debate transcript [8] should be fascinating to current database folks given the emergent state of database technology and the subsequent relational versus CODASYL history. Ted, some IBMers, Whitney, Mike, and about five others were the only people in the crowded room that had any RDBMS implementation experience. Of that number, only Ted and Kevin Whitney spoke in the debate. Everyone else was living in a different world. From the

---

[1] See Chapter XXX: The Ingres Years
[2] See Chapter YYY: Competitor, Collaborator, Friend: Don Haderle
[3] See Chapter ZZZ: IBM Relational Database Code Bases
[4] Amazingly, ~10 RDBMSs were implemented or under way within three years of Ted's landmark paper.

transcript, Mike seemed curiously quiet.[5] Truth was he had his hand up the whole time but was never called.[6]

In hindsight, most questions seem weird. "Why were ownerless sets better than navigating data?" "Why is the network model worse than the relational model as a target for English?" "I couldn't find many examples of the relational sublanguage compared to CODASYL subschemas." "I can think of many systems that I have had in which questions would come up so that it was almost impossible, and certainly impractical, to automate a way of coming up with the answer. To me, it boils down to a question of economics. Is it worth spending the money and taking the time to be able to provide this kind of availability to anybody?" In contrast, Ted's clear focus was on "applications programming, support of non-programmers (…), and implementation" and on the logical and physical data independence that remain the cornerstones of the relational model [2] [5], emphasized succinctly by Mike [6] and in sharp contrast to the network approach and most of what was said in the debate. The relational side was casting pearls [7].

For all the fireworks projected for the debate, it was bland. So, my mind wandered to J.R. Lucking, who smoked a cigarette throughout. It was, after all, 1974. Why pay attention? It was distracting. Smoke never came out. We imagined that J.R. periodically left the room to empty an otherwise hollow leg of smoke and ash.

The debate had little impact outside the room. The real debate was resolved in the marketplace in the mid-1980s after the much-doubted adoption of Oracle and DB2[7] and as SQL became, as Mike called it, "intergalactic data speak." The elegance of Codd's model would never have succeeded had it not been for RDBMS performance due to Pat Selinger's query optimization, enabled by Ted's logical and physical data independence,[8] oh ya, plus tens of thousands of development hours spent on query and performance optimization.

---

[5] Holding back like that didn't last long for Mike.
[6] The database cognoscenti who were running the debate may not have foreseen that in 40 years the tall, new guy with the unanswered hand would receive the Turing Award for the very issues being debated.
[7] DB2 was IBM's #2 DBMS product after its #1 DBMS product, IMS.
[8] Mike was already at the heart of the performance issue [6] described so eloquently by Chris and Ted [5] in the same conference and missed by debate questioners. Mike has generalized this as the key requirement of any new data model and its data manager.

The debate and conference had a huge impact … on me. Ted Codd became a mentor and friend, calling me almost daily throughout the Falklands War to review the day's efforts of Britain's Royal Air Force (RAF).[9] Charlie, who lived up the street from me in Lexington, Mass., in a large, elegant, white house, later offered me a CTO job. I declined but gained sartorial knowledge about buttons that I didn't know I had. Ed Sibley, my first academic boss at the University of Maryland, got me appointed chair of the ANSI/SPARC Relational Standards Committee, where I proposed, with other academics, to standardize the relational calculus and algebra, to allow multiple syntaxes, e.g., SQL, QUEL, QBE. I lost that job to an IBMer who came with a 200-page SQL specification. (Who knew that standards were a business and not a technical thing? Nobody tells me anything.)

While the debate had little impact on the community at the time, it marked the changing of the guard from the leaders of the hierarchal and network period of database research and product development. In the debate, they had posed the odd questions presumably trying to understand the new ideas relative to what they knew best. The torch was being passed to those who would lead the relational period that is still going strong almost half a century later. As the 1981, 1998, and 2014 Turing Awards attest, the new leaders were Ted Codd, Jim Gray, and Michael Stonebraker. With more than ten relational DBMSs built at the time of the debate and the three most significant relational DBMSs in the works, the database technology shift to relational databases was under way.

Mike: More Memorable than the Debate, and Even the Cheese

Apart from the neon orange cheese, SQL, and being awed by database cognoscenti, there was little memorable about SIGFIDET 1974, except meeting Mike Stonebraker. Mike Something became a colleague and friend for life. Although a stranger and the most junior academic at the strategy dinner (I don't count), Mike was unforgettable, more so as time went on. Me: "Hey, Mike, remember that dinner before the Relational-CODASYL Debate?" Mike: "Sorry, I don't remember." Maybe it's like a fan meeting Paul McCartney: only one of the two remembers. For

[9] In WWII, Ted trained as a pilot in Canada with the British Royal Air Force. I am Canadian, and my mother, an Englishwoman, had been in the British Women's Royal Air Force.

this chapter, I asked Dennis Tsichritzis and other database cognoscenti for memorable moments at this event, to a uniform response of "not really." Don Chamberlain and Ray Boyce, SQL inventors, were there [4]. But most future relational cognoscenti had not even joined the database party. Bruce Lindsay and Jim Gray were at UC Berkeley and would move that year to the System R project at IBM. The instrumental Pat Selinger was at Harvard (Ph.D. 1975) and wouldn't join System R until after her Ph.D. SIGFIDET 1974 was a milestone that marked the end of the old guard and emergence of the relational era with most of the relational cognoscenti still whet behind the relational model, and Mike Stonebraker, unwittingly, taking the lead.

To this day, Mike is succinct in the extreme, intense, visionary, and superbly confident. Yet at the debate, he was strangely quiet (not called upon) especially as he was in the 1% who understood Ted's model and had implementation experience. Perhaps he was gaining his sea legs. He had been an assistant professor for about three years. Forty years later, at his Festschrift, Mike recalled those tenure-grinding years as the worst of his career due to the pressures of an academic life – teaching and tenure, in a new domain, developing one of the most significant database systems from scratch, while, as Don Haderle says in Chapter XXX, having to learn "how to create and operate a business". At SIGFIDET he was new to databases, having learned what a database was two years earlier when, while Mike was wondering what to do at UC Berkeley Gene Wong had suggested that he read Codd's paper [2]. Mike's first Ph.D. student, Jerry Held, had already implemented a DBMS. By May 1974, Mike had already impressed the relational cognoscenti, the then-future of databases. Today at conferences, people universally wait to hear Mike's opinions. Or in his absence, as at VLDB 2017, Mike's opinions tend to be quoted in every keynote speech. On issues critical to him, he speaks out with succinct observations and questions that get right to the heart of the matter, e.g., he might ask "What use case and workload do you envisage?" Answer: Rhubarb, rhubarb, rhubarb. Mike replies: "Interesting. VoltDB is in that space but in seven years has never encountered a single customer asking for those features."

A Decade Later: Friend or Foe?

At the First International Conference on Expert Database Systems, Kiawah Island, South Carolina [9], I debated with Mike on the topic "Are Data Models Dead?" I do not recall the content nor the tone, which must have appeared confrontational because I do recall a look of utter surprise from Larry Kerschberg, the program committee chair, as Mike and I hugged off stage. Mike had arrived just before the debate, so we had not yet greeted each other. When it matters, Mike speaks his mind pithier than most. His directness and honesty may seem confrontational to some. I have never seen such an intent, rather he is getting to the heart of the matter quickly. That enriches the discussion for some and can end it for others.

My first meeting with Mike over 40 years ago was memorable. There were others at the strategy dinner, but I do not recall them. Mike was quiet, calm, succinct, scary smart, and contrarian. He was a Turing laureate in the making. My impression was that he was the smartest man in the room. My impression, like data in Ingres, Postgres, and his many other DBMSs, persists.

References

[1] V Kevin M. Whitney, Relational data management implementation techniques. In [3]

[2] Codd, E. F. (1970). "A relational model of data for large shared data banks." *Communications of the ACM. 13 (6): 377.* doi:10.1145/362384.362685.

[3] Proceedings of the 1974 ACM SIGFIDET *(Now SIGMOD) Workshop on Data Description, Access and Control: Data Models: Data-Structure-Set Versus Relational (SIGFIDET '74)* ACM, New York, NY, USA, 121-144. DOI=http://dx.doi.org/10.1145/800297.811534

[4] Donald D. Chamberlin and Raymond F. Boyce, SEQUEL: A structured English query language. In [3]

[5] Date, C.J. & Codd, E.F. (1975). The relational and network approaches: Comparison of the application programming interfaces. In [3]

[6] Stonebraker, M. (1974). A functional view of data independence. In [3]

Abstract: Many researchers have used the term "data independence" without indicating a precise meaning. One common, definition is—the isolation of a program from considerations of the data which it processes. Another is—the ability of an applications program to execute correctly regardless of the actual storage of its data. Although these suggest the general concept, a precise framework is clearly needed. The current paper provides such a framework and explores its ramifications.

[7] Mathew 7:6.

[8] Discussion—panel and audience. In [3]

[9] van de Riet, R.P. (1986). Expert database systems. In Future Generation Computer Systems, Volume 2, Issue 3, 1986, Pages 191-199, ISSN 0167-739X, https://doi.org/10.1016/0167-739X(86)90015-4.