

Young Prof. Stonebraker amongst the Database Illuminati

Michael L. Brodie

Computer Science and Artificial Intelligence Lab, Database Group

mlbrodie@mit.edu

Dinner with the Database Illuminati

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 me that we would start PhDs 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 long 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 me, a database know-nothing, and the current and future database illuminati. It started inauspiciously with Cracker Barrel's signature, neon orange cheese dip with grissini ('scuse me, bread sticks).

I was quiet in the presence of the illuminati Ted Codd; Chris Date, IBM UK Lab; Dennis, and this tall, quiet, but wonderfully confident guy, Mike something, a new UC Berkeley assistant professor and recent U Michigan grad. 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 databbase.

The *Great* Relational-CODASYL Debate

The illuminati lined up for the debate were, on the relational side, Ted Codd; Dennis Tsichritzis; and Kevin Whitney, who had implemented, at General Motors, RDMS [1]— one of the first RDBMSs. On the CODASYL side, 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 and Eugene Wong's pioneering Ingres project at UC Berkeley; coincident with the beginning of the System R project at IBM San Jose; five years before the release of Oracle, the first commercial RDBMS, in 1979, followed in 1983 by IBM's DB2; 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, esp. 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 conceptual and six¹ 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[5] on a core relational concept, like those of Codd, Date, and Whitney, shows a succinct and deep understanding [see abstract] 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

¹ Amazingly, ~10 RDBMSs were implemented or underway within 3 years of Ted's landmark paper.

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 transcript Mike was curiously quiet. Truth was, he had his hand up the whole time but was not called on.

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 on “applications programming, support of non-programmers [...], and implementation”, and on logical and physical data independence that remain the cornerstones of the relational model[5], emphasized succinctly by Mike[6] and in sharp contrast with the network approach and most of what was said in the debate. The relational side was casting pearls.

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? Because he didn’t exhale the smoke. I imaged J.R. periodically visiting the men’s room during the day 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-1980’s after the much-doubted adoption of Oracle and DB2, 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 independence², oh ya, plus tens of thousands of hours developing optimization and implementation.

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). In WWII, Ted trained in Canada as an RAF pilot and I was Canadian! Charlie, who lived up the street from me in Lexington, MA in a large, white, elegant house, later offered me a CTO job. I declined but gained sartorial knowledge about buttons 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, and 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.

Mike more memorable than the debate, and even the cheese

Apart from the neon orange cheese, SQL, and being awed by database illuminati, there was little memorable about SIGFIDET 1974, except meeting Mike. 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 this note I asked Dennis and other illuminati for memorable moments at this event to a uniform “not really”. Even the instrumental Pat Selinger wouldn’t join System R for two more years. Jim Gray and Bruce Lindsay were also not present. SIGFIDET may have been a milestone that marked the end of old guard and emergence of the relational era with most of the relational illuminati yet to appear.

To this day, Mike is quiet, 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

² 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 data manager.

he was gaining his sea legs. He had been an assistant professor for about two years. 40 years later, at his Festschrift, Mike recalled those tenure-grind years as the worst of his career. At SIGFIDET he was also new to databases and already friends with the relational illuminati, the then future of databases. Today at conferences, people universally wait to hear Mike's opinions. On issues critical to him, he speaks out with succinct questions that get right to the heart of the matter, e.g., "What use case and workload do you envisage?" Answer: Rhubarb, rhubarb, rhubarb. Mike: "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 PC 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 to some as confrontational. I have never seen such an intent, rather he is getting to the heart of the matter quickly. That only enriches such discussions.

My first meeting with Mike 43 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 and Postgres, 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] C. J. Date and E. F. Codd. 1975. The relational and network approaches: Comparison of the application programming interfaces. In [3]
- [6] Michael Stonebraker. 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] R.P. van de Riet, 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](https://doi.org/10.1016/0167-739X(86)90015-4).