

Engineering Ethics Cases with Numerical Problems

from an NSF & Bovay Fund sponsored workshop

August 14-18, 1995

Texas A&M University

Electrical Engineering Case 5

Database Discredited

Authors:

Wally Roth

(wlroth@tayloru.edu)Suggested Courses:

Computer Science

Level:

All

I. Narrative

As a consultant industrial engineer hired by a credit bureau named RWT (and pronounced "Right"), you have been asked to analyze problems which have occurred with their 20-million-record credit file. (Such large files do exist and one large credit agency claims to have records on 67 million U.S. households). RWT management became concerned when the following situation came to their attention.^{1,2}

A couple moving to a retirement community has an eye on their "dream home". They have a good credit history, so they assume they will have no trouble getting a mortgage to purchase their dream home through a local bank in their new community. A routine credit check through RWT uncovers the "fact" that they are a bad credit risk. When Igor Mendes Qurius (I. M. Qurius) from the local bank pursues the case, he discovers the couple has been mis-identified in the RWT databank and had been confused with another party having a very bad credit history. In making amends, the local bank approves the loan, but by now the home has already been sold to someone else. The couple is heartbroken and, worse yet, continues to have credit problems for some time.

II. Numerical and Design Problems

1. RWT has called you in as a consultant to make recommendations. Where do you begin?
2. What design flaws in the database have allowed this problem to occur?
3. Management at RWT claims they have only 1 error per 100,000 records in their databank. How would you

develop an experimental design to credit (or discredit) this statement?

4. After assuming or validating whether RWT is right in statement #3, how many bad records do they have in their major file at the present time? What implications does this have, if any?

5. At what cost per record do you decide they need to rework the database? What other data or assumptions do you need to make before recommending a solution?

[Assume you need (1) a cost per record to update, (2) the number of errors per year (which can be estimated), and (3) the cost per error found by users.]

[You also need to know the rate of updates or new records per month or year. Assume there are 300,000 accesses or updates per month to the database.]

Also assume it costs \$10/record to clean up the database plus \$50,000 in fixed costs. Finally, assume the cost for insurance, lawyers. etc. for each bad record found is \$100,000.]

6. Compare the two costs, draw conclusions, and recommend a course of action in a one page memo to RWT management. Alternatively, write out a dialogue for a discussion of the matter with RWT.

7. Estimate the time required to clean up the database. Can you design a solution which would not take the database off-line for _____ (your estimate) in time (mos/yrs)?

III. Questions on Ethics and Professionalism [As suggested by Kallman and Grillo¹, p. 61.]

1. List the "stakeholders" (those with something to lose or win in this case).

2. Should someone have done (or not done) something earlier?

3. Who benefits here? Who is harmed here? (There could/should be multiple answers.)

4. Here are three important ethical tests:

The Golden Rule Test asks whether you would be willing to accept the consequence of your action if you were the one affected.

The Rights Test asks whether your rights, such as the right to free and informed consent and the right to equal treatment, are being violated by a course of action.

The Utilitarian Test asks whether a course of action produces the greatest overall good for the greatest number of people, regardless of what it does to a few individuals.

Evaluate a decision to clean up the database from the perspective of each of these tests.

5. How does one go about preventing this situation from occurring again?

IV. Solutions to the Numerical Problems

1. You would first want to find out how often such errors occur and what the source of the typical error is in the system--data entry, updates, software, indices, etc.
2. This assumes there is a design flaw. There may not be any at all.
3. You would like something like 50 random records from the 20,000,000 records they claim to have (That number should be verified also). Then, every 20 meg/X = 50 records should be chosen. Hence, selecting "every 400,000th" record (X = 400K) in some random fashion would yield 50 records. Later, if a pattern emerges as to what type of records are in error, a subset of those could be randomly tested.
4. This would mean they have 200 bad records. That is really rather impressive if it turns out to be correct. Also, the type of error would be significant. A small address error may be trivial as compared to a pointer to the wrong person's record. If all 200 errors are bad pointers, RWT needs a major software rework.
5. If they have 300,000 transactions per month and that is where the errors occur, then there will be 3 x 12 or 36 errors per year entered into the system. However, the far tougher problem will be in finding these errors!
6. Using the assumptions made earlier one can conclude it will take about 60 years to check and verify the entire *static* file and justify the cost of clean-up. If one assumes \$1 million in cost per loss, it will still take 6 years to "break-even". One can ignore the \$50,000 fixed cost as irrelevant in the calculation.
7. A software solution might be implemented by taking the system down over a weekend. Any other pseudo-manual system would leave employees demoralized by its "snail's pace" of cleaning up the bad data.

V. Possible answers to the ethical and professional questions.

1. Everyone listed in the problem and the public as well.
 1. Probably not. These mistakes will occur in any file system.
3. No one benefits.
4. The first two tests would probably require reworking the database. If I were in that position of the retired couple, I would want the data base corrected. The rights of people to free and informed consent and to good treatment are violated by the dissemination of false information. Nobody would consent to false charges of financial irresponsibility being made against them, and such people are certainly treated unequally. Whether the third test would require it is more problematic. The cost of correcting the problem might outweigh the violation of the rights of individuals.
5. One probably can't!

VI. Endnotes

¹ This case is based on an idea from Ethical Decision Making and Information Technology, Kallman and Grillo, McGraw-Hill, 1994. The case is called "Credit Woes", p. 59.

² There is some similarity in this case to the famous Ford Pinto case. See Harris, Pritchard, and Rabins,

Engineering Ethics: Concepts and Cases, Wadsworth, CA, p. 205.

³ There is now legal recourse for the couple in such a case, but the law focuses on the responsibilities of the credit supplier and how the data can be corrected, not on how individuals can resolve follow-up errors. The author had a similar experience in his home state last year.