



Working Session

January 19, 2017

Meeting Agenda

- 🧐 Open Data Migration Discussion
- 🧐 Brief bit on estimation and building timelines
- 🧐 Data modeling revisit

Negative points are a killer. If you got your dailies in and see negative points being awarded on Canvas, please let me know. There may be bugs in DERP and I don't want points to be lost on account of errors.

Where this is the last time we meet in class this week, I want to provide ample time for questions about the legacy data, the data model, and sql. We will start there for the discussion and run until there are no more questions. I am hopeful that this will take the majority of the class time today.

I have a few thoughts on estimation and timelines that I can fill a little bit of time with.

We can also try a data modeling exercise. I feel like I flubbed the lecture Tuesday and would like to try again because I think relational databases are really useful tools.

Migration Challenges

- 💡 Misspellings
- 💡 Model mismatches
- 💡 Non-machine readable sources
- 💡 Missing parts of the workflow
- 💡 Data in violation of the presented workflow
- 💡 Missing data



Misspellings - My data migration back east was for a title abstraction company that operated in Virginia. My system used a lookup for state names to make sure the names were valid; one of the records I had to read in was for Virginia

Model mismatches - In the source data the value may be represented one way (perhaps as arbitrary text) but needs to be represented by a different type or as a relation in the target.

Nonmachine readable data - The data exists in a PDF from another vendor or only exists in hard copy. On a security product we needed the list of controls from the NIST SP800-53. We could only find these in a PDF document I copied and pasted from the PDF into a text file and made a parser to go through the data. It worked ok but there was still some manual entry.

Data in violation of workflow - I haven't experienced this one directly but it is a huge problem on projects. For example, the process indicates that approvals must be done on a weekly basis, but they are not. Migrating the data fails because the existing process is not followed and the data is not in the source files.

Missing data - There is some data that is needed for one reason or another that doesn't exist or can't be provided. The customer list I mentioned in the first lecture is an example.

Solving these can sometimes be done by guessing and sometimes these involve discussions with the customer.