* Data flow diagrams.
* Textbook ‘documentation’ section.
* Follow structures from the textbook.
* Go over ISYS unit resources.
* Quantitative metric for data privacy; questionnaires, surveys, something that provides an answer.
* Ultimate question: is system safe? How can you measure and prove this?
  + Simulations, tests.
* Rewrite introduction:
  + Be more ‘research’-focused.
  + Focus on quantitative results from different options of system designs; “system A is better than system B because of these differences”.
* Design tests before building system; break system then build around these vulnerabilities.
* Testing that:
  + System behaves as it should given compliant user.
  + System behaves as it should given non-compliant user.
    - Lock them out
  + Malicious user, behaves as it should.
* Why not use another system?
  + Expenses
  + Doesn’t meet user requirements
  + System abilities:
    - Scalability
    - Flexibility
* ER diagrams, data flow diagrams.
* How to design a system like this? What steps to follow and then how to test? Best configuration for the databases?
* User requirements 🡪 design (3 different designs) 🡪 test plan 🡪 prelim work (e.g. using matlab to keep stuff in a database check r) 🡪 Thesis B plan: run other 2 designs with test plan 🡪 plan over the 15 weeks of next semester, how’s it gonna all work and be tested
* If the 3 designs don’t’ work or have flaws propose a 4th that combines the best of all
* HD: include preliminary test plan, show them a shitload of work.
* Implement something!!!
* 5-10 hours writing thesis, 5-10 hours practicing with matlab/practical work; create mock databases, fields, stuff you can test with. Tests: making sure data is going in the right places.
  + Test plan: these are the queries to retrieve data, metrics for safe vs unsafe.
* Break name into initials? Perhaps?
* SQL-lite.
* Second marker for this thesis? Ask around.
  + Security
  + Data privacy
  + Databases
  + Pure profiles for MQU staff; it’s their record of research. Researchers.edu.mq or something like that

Post meeting notes

* My ER diagram didn’t capture everything that we have discussed.
  + How does this lead to the final product? Need a clearer picture
* There’s a chapter on documentation in the textbook; how to actually document the design.
* For my thesis, I should have similar sections. Describe the system that I’m going to build on paper, so someone can look at it and imagine how it would all fit together.
* Research components:
  + “ive designed the system, how am I going to test it?”
    - Create a privacy metric.
  + Novel contributions to the field in this area
* Some kind of metric; something measurable where I can determine “yes, the patient data has remained private and useful research data is available”, or “no, the patient data has not remained private with the given research data”. Something **quantitative**.
* Ultimately the question is “is this system safe? And how do we measure the safety of it?”
  + Run simulations!
* Going to look at different structures and compare them in terms of efficiency, security, etc.
  + One central database (all eggs in one basket)
  + Split databases (split the risk)
  + Examine the trade-offs of each system, **discuss this shit**.
  + Alan discusses this around 14:30 of #4 recording.
* Design tests before the system; if you build the system first and then test it, you know what works and what doesn’t. What you want to do is break the system first, and then build the system that’s unbreakable.
* So testing a couple things:
  + That it behaves as it should given a compliant user.
  + It behaves as it should given an incompetent (find a better word) user.
    - System should give reasonable results, or lock the user out.
  + It behave as it should for a malicious user (hacker)
    - Be aware of hacks for this kind of system.
      * Buffer overflow.
      * Etc.
      * Talk about these certain hacks as well!
* Also talk about the rationale behind my system. Solutions exist, why’s this any different?
  + Free to use? Other options may require a license.
  + Doesn’t serve the needs of the user; too generic or too specific. Scalable system.
* ER diagrams, data-flow diagrams.
* Test plan; how the design will be tested.
* Design a system where I;m going to compare what happens if I keep everything in one database, or if I separate the databases in one way, or separate the databases in another way; 3 cases.
  + Keeping all the information in one database is already the worst case scenario; this is the control condition.
    - Vary in 2 other ways to see which way is better. Hopefully these 2 other ways are better than keeping all the information in one DB.
    - If the other 2 ways don’t then this contradicts the theory of separating the data.
      * The theory is that if pieces of information are broken up in different ways, it should be safer.
      * Test this with different test plans.
* User requirements 🡪 design (3 systems) 🡪 test plan (tests which of these systems is best; include a hypothesis) 🡪 preliminary work that has been completed (e.g. “I’ve started using matlab with sqlite and created the control condition where all info is kept in 1 database; here’s the code for my prelim work) 🡪 plan for thesis B (write the other 2 databases in code, run them with the test plan) [this is what I’m going to do with the 15 weeks of the second semester; have 3 different ways of running the databases and I’m going to test them to make a recommendation]. Also include some suggestions: “if I had more time, I’d make these changes due to flaws XYZ that I found”, suggested remedies. Due to time constraints, this is left for the next student 😊.
* Document and include all supervisor meeting documents into the thesis.
* To get a HD need to show a very substantial amount of work, very detailed test planning, maybe even have implemented that test planning into the preliminary work cause then I’d have some results.
* Prelim testing may include:
  + Ensuring that data is stored in the correct areas of the database.
  + Ensuring that the correct data is returned for a given query.
  + Ensuring that only the data requested by a query is returned.
* Justify your choices. Why are certain data fields given a certain privacy score? Use this to make a privacy metric.