SENG 5707: Data Modeling Homework

(Individual; <u>not</u> team) Homework 1.

DUE: by, at the latest [earlier is better – being nice to graders is always cool]:
<u>noon on</u> <u>THURSDAY 9/22</u> [Early submission is encouraged]

Submit <u>one</u> PDF (preferred) or MSWord file via Moodle. You can scan hand written material (returned on 9/19.)

A. [This is an easy one to get you started.]
On page 39 of Mastering Data Modeling, in the "outside-the-box bars" section the authors made two errors – incorrect reading of Fragment 3-14. [Look at the italicized text.] Find those errors, state what is wrong and state what we should have written.

B. [Data modelers are more precise than regular people.]
On page 41 we write "do not say *About a cow*". Explain what is wrong with that.

C. [Getting this wrong with your progress.]
Do exercise 3 on page 44, and answer this question: How many identifiers does registration event have.

D. [Even master data modelers count]
On page 62 do exercise 2 but only for Fragments 3-17 and 3-19.

E. [Reading correctly is necessary for even a novice level of **Logical Data Structure, LDS, skill.**]
For Fragment 11-24 on page 148, write the ten sentences that constitute reading the LDS.
[Advice: Before you write a sentence determine the counts. E.g., " About each Newspaper we can remember 2 descriptors.]

F. [Hint: Employ phrases like these: "As a user I say ..." and "As a modeler I ask ..."]
On page 85 do exercise 3, that is, make up a conversation.
Requirement: While Chapter 6 had two dozen fragments, you should aim for only about one dozen.

---

advice
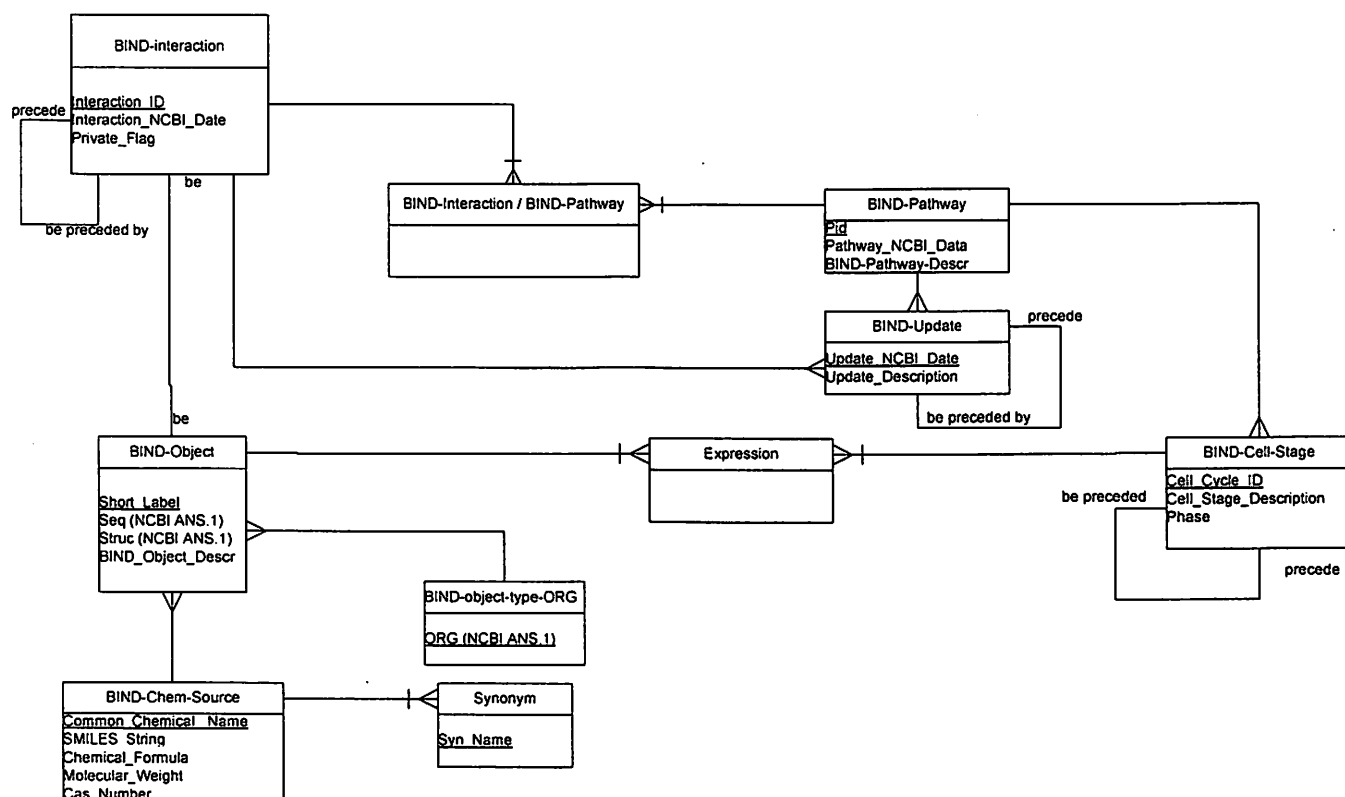You will find this software helpful: **www.erdraw.com**.
It was written by MSSE grad Allan Tokuda.

(Team) Data Modeling Homework 2.
DUE: by the end of Thursday 10/13 ;Returned on Saturday 10/15
Each team [not each person] should submit one PDF or MSWord file via Moodle.

1. For the following fragment,

Write out how you would read this LDS. [You do not need to understand BIND to accomplish this.]
   a.   Make and fill out a table like this:

	How many does each entity have?		Kind of Entity
Entity name	Identifier	Descriptor	Refer to Chapter 8

In "Mastering Data Modeling ... " by Carlis and Maguire:
2.   Using Fragment 5-3 on page 73
   a.   Do exercise 3.c on page 73.
   b.   Pat Down, who was promoted from the security force, informs you that the stadium management plans to add tiers to the stadium. Document at least four questions you would want to ask Pat about adding tiers. With your user hat on, become Pat and make up document answers. *(Hint: allow each team member to practice being the user, while the other team members practice asking questions as the data modeler.)*
   c.   Based on Pat's responses, evolve the LDS in Fragment 5-3 on page 73 to accommodate the changes.

3.   Do exercise 1 on page 360 **or**   Do exercise 5 on page 363.
   (We grade only the first one if you submit both.)

(Team) Data Modeling Homework 3.

DUE Saturday 11/12 (in class, <u>paper submission only</u>)
Goal: Improve your skills as a content neutral data modeler.
Task: Reverse engineer the data model for an application's data, using documentation (and, if available, real instances too).

The application should be something that you nothing about (or only a bit – you really are never entirely ignorant). The idea is that you become comfortable with never putting on a user hat. Reverse engineering is fun and satisfying. This task can give you large amounts of well-deserved confidence.

For this task, it is *not* a good idea to reverse engineer an application where you already understand it, that something at your work or elsewhere, e.g., IMDB.com, where you will be tempted to put on a user hat.

You get to pick the application. When you have decided, send email to the TA, copying Carlis. Begin searching soon. [We càn also help if you have no idea of what to work on.] You can ask the TA or me to judge whether the system you choose to model is too large or too small. How big? Well, the real criteria is that you have a variety of LDS experiences (roughly, shapes), and are not overloaded. Historically, most teams having success have picked models with 20-60 entities. 5-10 entities is too small and >100 is too large – so use some judgment. For example, IMDB used to be a good size, but they have been successful and added lots of new content and now reverse engineering all of IMDB is too large for this assignment – besides being too easily understood --"Damn Yankees".

For this task, it is good to work from forms or formats, in addition to other documentation. For example, the protein data bank, PDB, has this file format [http://www.wwpdb.org/docs.html]. It is not so good to work from a relational schema – the reverse engineering generally is too simple for mastery learning. (If a work db schema is really poor it might provide enough data modeling experiences.)

Here are some sites with a biomedical flavor – likely to fit the content neutral guideline. [Many of these are about graphs – like the pirate example.] These are only a few of hundreds available – do some scouting around. Pick something you find interesting.

> http://nar.oxfordjournals.org/content/41/D1.toc [~100 articles]
> http://www.oxfordjournals.org/nar/database/a/ [1512 databases]
> http://www.ncbi.nlm.nih.gov/Database/
> http://bbid.grc.nia.nih.gov/bbid499.html
> http://www.genome.jp/kegg/
> http://metacyc.org/

Printed Deliverables:
- A well-formed data model [no sandbox!]. (Bring 2 extra copies for other students to look at and whatever else might help you present)
- A brief description of the application. [Some teams show instances via a laptop.]
- A description of the easy and hard parts of your reverse engineering task.
- A list of questions that you want to ask the users about their data.

Alternate task
You can use an application where you know something about the data – only if you have users available to work with you as data modeler. If you are considering this alternative, talk to Carlis. The danger is that real projects are generally much too big for a class project, but that risk can be offset by having real users to work with

Pep Talk:
Being a content-neutral modeler is quite empowering. I have lessened my ignorance about lots of subjects by performing this reverse engineering of a "how" up into "what" plus questions. You can dive into something new, see some shapes, ask some questions, and learn a lot in a short amount of time. This skill is consistent with a constructionist model of learning.

In-class Presentation Instructions for Reverse Engineering Assignment

Phase 1.   15 minutes for each presentation

Get into gaggles of three teams; decide who presents and who listens
Each presenting team should divide up the presentation (nobody gets to just sit)

For one team of presenters:
- Presenters: give a quick overview of the application
  and then a tour of your model
        Anchor with instances

- Listeners: look for "sandbox"
        Examine names
        Examine goodness of identifiers
        Ask about degree 1 link
        Find shapes: "chicken feet in out & across" plus "plus"

- Presenters: describe your hardest modeling
- Presenters: state several of the questions you want to ask users

- Thank each other for listening & presenting

Switch roles repeating phase I until everyone has presented

Phase 2.   5 minutes for each presentation

Take an LDS that you were listening about.
Re-gaggle into 4 person gaggles;
You present another's model