Relational Database Normalization In-Class Exercise
"An Organization Metadata Database"

Organizations are increasingly facing what I call the “file cabinet and distributed PC problem.” This describes a situation where organizational employees collect data for various projects or business needs but do not keep good track of where the data reside, where they are backed up, who is the designated owner of the data, etc. The organizational data resources are scattered everywhere and are not very well maintained.

You have been asked to design the relational database structure for a metadata database for your organization. To begin design of this database, you need to gather requirements. Specifically, you interviewed various people in the organization who deal with organizational data to identify the data that needs to be in this database.

From these discussions you learned several things about data management in your organization.

- Datasets are assigned to specific departments
- There sometimes can be multiple people assigned to maintain a particular dataset
- Sometimes data managers maintain multiple datasets

These meetings also provided a list of possible fields for the database tables:

- Dataset name
- Type of dataset
  - Geographic (e.g., GIS coverage)
  - Nongeographic (e.g., a spreadsheet with client information)
  - Publications/Reports

Version #
- Citation for the dataset
- Beginning date of content (if a time series dataset)
- Ending date of content (if a time series dataset)
- Department who published the dataset
- Name of person or persons who own and maintain the dataset (Last name, First name, Middle Initial)
- Dataset progress (Completed, in progress, planned)
- URL of the dataset (if information about it is on an existing web page)
- Dataset owner’s email address
- Appropriate citation for the dataset
- Abstract of the dataset
- Internal mailing address and phone of dataset owner person
- Rules for dataset usage
- Number of datasets maintained by a department
- Name of the manager of the department
- Status of dataset owner (full time, part time)
- Methods used to create the dataset
- Any other comments about the dataset

Your task: Develop an entity-relationship diagram for this metadata database.
An Environmental Remediation Project  
Entity-Relationship Exercise

You have just taken a job with a major government agency chartered with the cleanup of many environmental restoration projects throughout the United States. One of the primary responsibilities of this position is to keep track of current project status and periodically reprioritize clean-up projects as new ones are discovered. Lately, several projects have fallen behind schedule and have resulted in bad public relations for your agency. Your boss, the head of the environmental remediation branch of this agency, has asked you to develop a method to improve project management and reporting at the headquarters level. Since management has traditionally been done primarily through paper reports you decide it is time to develop a clean-up project management database that could be accessible through the organization’s LAN.

To gather requirements of this database, you held a requirements gathering session with several project administrators. In these meetings, you discussed how projects are currently managed and then brainstormed about the type of information needed for effective project management. You learned several things about the way projects are currently managed in the agency.

- A “cleanup project” is a cleanup initiative at some location somewhere in the U.S.
- Several “cleanup sites” may exist at any given cleanup project location.
- “Cleanup sites” are areas that may be contaminated with one or possibly more contaminants.

In this meeting you also were able to define the types of information needed for the high-level management of these projects:
- Cleanup project name -- usually named after the city/town they exist near
- Number of cleanup sites in cleanup project
- Total cleanup project area contaminated (a value in acres).
- Project status (NS-not started, A-analysis phase, R-remediation phase, C-completed)
- Contaminant names (these can be up to 50 characters long)
- Contaminant-id
- This is a code that corresponds to a list of particular contaminants the agency is concerned about (e.g., 1-mercury, 2-PCB, 3-oil, etc...). You do not have to worry about what these codes stand for--just let this be an autonumber field.
- Estimated contaminant amount at a cleanup site
- Contaminant unit of measurement
- Contaminant toxicity level (1-10, where 10 is the highest)
- Best remediation method(s) (a description)
- Cleanup site name
- General cleanup site location longitude (in decimal degrees. E.g., -86.3434)
- General cleanup site location latitude (in decimal degrees. E.g., 35.2323)
- Cleanup site status (NS-not yet started, A-analysis phase, R-remediation phase, C-completed)
- Cleanup project priority level (1-10, 10 being highest priority)
- Cleanup project percent complete (percentage)
- Other comments related to cleanup project as whole
- Other comments related to specific cleanup site
- Estimated cleanup project end date
- Cleanup project start date
- Cleanup project estimated cost
- Current level of expenditures for the cleanup project

Develop the E-R diagram for this database.