Assignment 2: Data Science CSCI/ERTH/ITWS (10% written and 10% presentation of overall credit score)
Due: September 27, 2011 (by 1200m ET, day of the class**)
Submission method: email to pfox@cs.rpi.edu and chenw8@rpi.edu including your class presentation. Please use the following file naming for electronic submission: DataScience_2011_Assignment2_YOUR_NAME.ext
Late submission policy: first time with valid reason – no penalty, otherwise 20% of score deducted each late day
Office hours: normally Tuesday 2:30-3:30, Winslow 2120 or by appointment or email.

Note: Your report and presentation for this assignment should be the result of your own individual work. Take care to avoid plagiarism (“copying”), including all web resources, texts, and class presentations. You may discuss the problems with other students, but do not take written notes during these discussions, and do not share your written solutions.

General assignment: Based on the practical experience from your data collection exercise, document and present the data that you collected.

The weighting score for each question is included below. Please use the question numbering (1 and 2 and a, b, etc.) below for your written assignment.

1. Report on your data management practical experience (not the plan but the practice) from the data collection exercise using a minimum of 2-3 sentences (for each item) (5%)
   a. The goal (what it was driven by), and the mode of collection
   b. How data was acquired, any problems, lessons learned, etc.
   c. Physical and logical organization of data/ metadata, success, failures
   d. Metadata and documentation collected/ stored
   e. A link to the data/ metadata (or a copy of it)
2. Describe your experience in how well (or not) the data management plan turned out in practice. 5%
   a. Highlight what worked and what did not and why (minimum 1-2 sentences for each, 4-5 items max, do not need to cover all 9 areas)
   b. What you would do differently next time?
   c. Were there effects on the data or metadata collected, its quality, etc?
3. Oral presentation of the data collection and management in practice (10%). Your grade will be assessed giving the presentation and demonstrating that you collected the data and followed (or modified) your data management plan. Please submit your presentation (ppt, pdf or similar using the same naming scheme as for questions 1-2) after your presentation. Guidelines
   a. ~ 8 mins. each, allow time for at least one question (2 mins.)
   b. Presentations will be split over two weeks but everyone needs to be prepared to present in week 5 (**)
   c. 5-6 slides MAXIMUM (slides are optional as is the use of a computer)
d. What you should present:
   i. The goal (what it was driven by), and the mode of collection
   ii. How data was acquired, any problems, lessons learned, etc.
   iii. Physical and logical organization of data/ metadata, success, failures
   iv. Other ‘management’ aspects that you deem important
   v. Metadata and documentation collected/ stored
   vi. Some of the data in some form

4. Graduate question addition to Q 1 and Q 2. a) what standards (data format, metadata, etc.) did you use (whether planned or not)? b) What standards would have been helpful (‘none’ is not an acceptable answer), i.e. what area? c) Were you aware of a best practice for the type of collection you carried out? Minimum 3 sentences (1 for each sub-question).

Include all references, citations and attributions of external information sources used.