Assignment 2: Data Science CSCI/ERTH/ITWS (15% written and 5% presentation of overall credit score)
Due: October 6, 2015 (by 12pm ET)
Submission method: email to hughet2@rpi.edu and koiraa@rpi.edu, including your class presentation (after you present it). Use your rpi.edu email address to send.
Please use the following file naming for electronic submission:
DataScience_2015_Assignment2_YOUR_NAME.ext
Late submission policy: first time with valid reason – no penalty, otherwise 20% of score deducted each late day.
Office hours: Monday 3:00-4:00, JEC 5018 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 performed as a pre-requisite to this assignment, document and present the data that you collected. NOTE: data analysis is NOT required for this assignment but may be performed.

The weighting score for each question is included below. Please use the question numbering (1, 2, 3 and a, b, etc.) below for your written assignment response.

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) (7%)
   a. The goal (what it was driven by), and the mode(s) of collection
   b. How data was really acquired, any problems, lessons learned, etc., converted
   c. Physical and logical organization of data/ metadata, success, failures
   d. Metadata and documentation collected/ stored, what provenance was identified
   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. 6% (graduate students)/ 8% (undergraduate students)
   a. Highlight what worked and what did not and why (minimum 1-2 sentences for 4-5 categories of items in your plan (max). You do not need to cover all 9 areas)
   b. What you would do differently next time and why?
   c. Were there effects on the data or metadata collected, its quality, etc.?

3. Graduate student question: related to Q 1 and Q 2. 2%
   a. What standards (data format, metadata, etc.) did you use (whether planned or not) and/or what ones couldn’t you use and why? Minimum 1-2 sentences.
   b. What standards/ conventions would have been helpful (‘none’ is not an acceptable answer), i.e. what area? Minimum 1 sentence.
4. Oral presentation of the data collection and management in practice (5%). Your grade will be assessed by giving the presentation to demonstrate that you collected the data and how you followed (or modified) your data management plan. Please submit your presentation (ppt, pdf or similar using the same naming scheme noted above) after your presentation. Guidelines:
   a. ~ 8 mins. each - allow time for at least 1-2 questions (2 mins.)
   b. Presentations will be split into four groups and held in parallel (sections) during class time thus 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 ‘data management’ aspects that you deem important
      v. Metadata and documentation collected/stored
      vi. Present some of the data in some form (i.e. curation)

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