Assignment 1: Data Science 2013 - ITEC/CSCI/ERTH (10% of overall credit score)
Due: September 10, 2013 (by 0900 ET) – before class!
Submission method: email to pfox@cs.rpi.edu **and** sharms3@rpi.edu
Document naming: DataScience_2013_Assignment1_YOUR_NAME.ext (e.g. txt, pdf, doc, zip).
Late submission policy: first time with valid reason – no penalty, otherwise 20% of score deducted each late day
Office hours: Monday 3:00-4:00pm, Winslow 2120 or by arrangement

Note: Your report 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. Use the numbering below when completing you responses to this assignment.

General assignment: propose **two** data collection exercises (label them A and B) and perform a survey of data formats, metadata and application support for data management suitable for the data you will collect in ~ week 4. Note the overall modes of data collection:

- Observation
- Measurement
- Generation

Driven by

- Questions
- Research idea
- Exploration

1. Data collection – propose two data collection options - 4%
    a. State the details of the mode of collection and what the data collection need is being driven by. Suggested minimum response is 3-4 sentences.
    b. Describe a management plan for the data and metadata acquisition and initial curation using the **9** headings under Management lecture slides from week 2. Minimum 1-2 sentences per category.
2. Survey of data storage/ formats - 3%
    a. Based on Q1 (for both proposed collections), research and describe existing suitable data formats that could be used. If no suitable format is available, describe the data format needs and your choice including any structure choices for the data. Minimum 4-5 sentences.
3. Survey of metadata conventions, standards - 3% (undergraduate), 2% (graduate)
    a. Based on Q1 (for both proposed collections), research and describe existing metadata conventions/ standards in use. If there are no suitable ones, describe the metadata needs including any structure choices. Minimum 3-4 sentences.
4. Graduate student question - 1%
    a. Describe the provenance information you plan to collect (for both proposed collections) and how they may support the overall data collection and goals of the investigation. Min 2-3 sentences.

Include **all** citations and sources of information you use especially for Q 2, 3 and 4.