Final Assignment: Data Science 2014 (10% written of overall credit score)
Due: Tuesday December 2, 2014 (by 5pm ET, day of the class)
Submission method: email to pfox@cs.rpi.edu and dass4@rpi.edu. Please use the following file naming for electronic submission for any individual documents:
DataScience_2014_AFinal_YOURFIRSTNAME_YOURLASTNAME.xxx
Late submission policy: NOTE: assignments handed in after Dec 5, 5pm may not be graded in time for inclusion in your final grade.

Note: Your report for this assignment should be the result of individual work. Take care to avoid plagiarism ("copying"), including all web resources, texts, and class presentations. You may meet to discuss the tasks for this assignment with other members of the class but must produce the materials for the final assignment themselves, citing all outside sources.

General assignment: Stewardship: Workflow construction for preservation. Use the data workflow(s) from your group assignment (or for extra credit use the detail from another group and indicate this in your submission). The weighting score for each question is included below. Please use the question numbering (1-3) below for your written assignment.

1. Construct a scientific (data-oriented) workflow. Present the workflow in a diagram form with suitable annotations (documentation) for someone else to review. The workflow should be an accurate depiction of the data flow from the chosen project for your specific instance, i.e. it need not be generalized. The diagram should be included in what you submit. Provide a minimum of a 5-6-sentence description of what was required to carry out the construction of the workflow. (6%) [Hint: the diagram can be hand drawn, computer drawn, or developed in a workflow tool – in the latter case the workflow need not be ‘executable’]

2. Describe each major stage of the workflow and indicate how well (or poorly) data and information preservation is enabled or accommodated. You may include URLs to information sources, etc., minimum of 3-5 sentences (2%) 

3. Describe how the workflow and your assessment of existing documentation such as provenance and other contextual information enabled (or not) the data stewardship that was needed in your project. Minimum of 3-5 sentences (2%) [Hint: Review notes from class 2 as well as class 9]