

Assignment 3: CSCI/ERTH/ITWS-6961 (20% written of overall credit score)

Due: Tuesday November 2, 2010 (by 12pm ET, day of the class)

Submission method: email to pfox@cs.rpi.edu. Please use the following file naming for electronic submission for any individual documents:

6961-2010_A3_YOURFIRSTNAME_YOURLASTNAME_<something extra>.xxx

Late submission policy: first time with valid reason – no penalty, otherwise 20% of score deducted each late day

Office hours: Tues, Oct 19 2010 2:30-3:30, otherwise by appointment (email)

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 are allowed discuss the tasks for this assignment with other students in your group.

General assignment: You have been asked to submit two datasets to an institutional repository for preservation after you graduate. The repository uses the netCDF format exclusively. The repository will accept both standard and user-defined metadata conventions in creating the netCDF files. Using knowledge on data management, convert TWO sets of data (from Assignment 2) to netCDF format. One dataset will be yours and the other will be from someone else (you can choose the person) from the class. You may need to consult with the other person in completing this assignment. Label the datasets: Dataset I and Dataset II. The weighting score for each question and part are included below. Please use the question numbering (1-2, a, b, etc.) below for your written assignment.

1. For BOTH datasets answer the following questions (total 10%)
 - a. Describe how the logical organization and physical organization may need to change in the transfer to netCDF. This includes a indication of whether all metadata will be encoded in the netCDF file or not, i.e. externally and why, choices of file names, etc. Describe whether you retained an existing metadata standard or convention or converted to another one, with details of your choice. Minimum 3-4 sentences (4%)
 - b. Describe what additional metadata and/ or information you would include for cataloging and preservation purposes. Min. 2-3 sentences (4%)
 - c. Describe any difficulties you encountered and the solutions you developed in the conversion process. Min. 1 sentence (2%)
2. Implementation for BOTH datasets (total 10%)
 - a. Convert the data from the original formats (ones that were handed in for Assignment 2) to netCDF. (4%)
 - b. Document the implementation and include a code/ method to read (example sufficient for someone else to use) the data/ metadata. (3%)
 - c. Create and submit a ‘package archive’ (bonus points for an OAI Archival Information Package) for each set of data that could be delivered to the repository. This includes any codes or documentation on methods you used. (3%)