IST687 Scientific Data Management

Spring 2012

Instructor: Jian Qin

Email: jqin@syr.edu
Office: 311 Hinds Hall
Phone: 315-443-5642

Time: any time
Location: anywhere

Course Description

The Scientific Data Management course includes three modules: 1) fundamentals about science data, 2) data management with case studies, and 3) use of science data and broader issues relating to tools for management and visualization, as well as quality and publication practices. The first module provides an overview of science data and data management, including data fundamentals such as forms, scales, types, levels, data structures and models, data formats, and databases used to store, retrieve, and manage data. The second module uses case studies regarding data collection, processing, transformation, management, and use to provide students with hands-on experience. In the last module, students will be introduced to methods and tools for evaluating data quality and working with data for use in various contexts. Students will work as an interdisciplinary team on a comprehensive science data management project throughout the semester under the guidance of the instructor. The performance of each student is based on exercises, quizzes, group reports, class participation, and the course project.

Learning Objectives

At the end of the course, students are expected to understand:

- Concepts and characteristics of science data and practices
- Principles and practices in data management and use
- Technologies used to manage and use data
- Procedures and methods of using data for inquiry

At the end of the course, students are expected to be able to:

- Identify the needs for organizing, reporting, and managing science data
- Represent datasets by using science metadata standards
- Design databases and systems for science data management and use
- Evaluate the quality of metadata describing science datasets

What does it take to succeed in the course?

- An interest and passion in a career in science data management or science and engineering research in corporate, academic, or government sector
- Basic computer skills, including creating simple web pages, using spreadsheet (e.g. Excel) and/or database (e.g. Access) software, and others
- Willingness to work in an interdisciplinary team

Required Readings

There is no required textbook, but required readings will be available in Blackboard as electronic documents for reading and printing. Students are responsible for complete reading the assigned materials for discussions and coursework. A complete reading list by week can be found at the end of this document.

Contributions to Grade

The work for this class will involve a mixture of quizzes, individual assignments, case study reports, and a final project.

- **Quizzes** (2 x 5% = 10%) are designed to test your understanding of basic concepts in science data management.
- **Exercises** (4 x 8% = 32%) are designed for you to practice the necessary skills in carrying out data management project.
- **Case study reports** (2 x 6% = 12%) are designed to maximize the usefulness of case study results and provide staged deliverables for the final project.
- **Final project** (30%): the project team will be formed early on in the semester and the team members will work together throughout the semester on many tasks.
- **Participation** (16%) includes your attendance and participation in class discussions and activities.

Expectations and Grading Policy

I try to make every class worth attending. Students will be responsible for all materials covered, handed-out, announced, etc. in class unless told otherwise. Attempts will be made, however, to place important announcements in class and/or on the class web page. Class will be held on all days as scheduled, unless notified otherwise.

Every attempt will be made to return assignments in a timely fashion. Assignments are due at the time indicated in the weekly schedule, unless specified otherwise, and will be annotated with the point count for individual components of the assignment. Late work will be accepted only for two days after the due date, with a 5% penalty per day. This is to facilitate the timely return of graded assignments with answers.

This syllabus (including course requirements, due dates, etc.) may be changed with sufficient notice.

If you have any type of disability that may require additional time or special consideration, please let me know at the beginning of the course.

Grading Policy

- Each assigned work will be graded on the scale as specified for the component (e.g. each exercise will receive a maximum points of 7), which will be summed at the end of the semester.
- Grade levels follow the scales below:
  A = 95-100, A- = 90-94.9, B+ = 85-89.9, B = 80-84.9, B- = 75-79.9, C+ = 70-74.9, C = 65-69.9, C- = 60-64.9, F = below 60

- An incomplete grade, **I,** can be given only if the circumstances preventing the on-time completion of all course requirements were clearly unforeseeable and uncontrollable. If an incomplete is required a written contract must be completed which specifies the nature of the missing work, the date it will be completed, and the default grade that will be given if that deadline is missed.
- It is unethical to allow some students additional opportunities, such as extra credit assignments, without allowing the same options to all students.
- Failure to complete any course requirement will result in a course grade of **C** or lower, regardless of the grades received in other components.
- Group-based assignments will usually have a component of the final grade based on each group member's assessment of the contribution made by the others in the group.
- To discuss a grade, arrange for a private meeting in which you identify the sources of your concern. It is important to bring with you to that meeting the relevant materials (e.g., marked papers). Except for extraordinary circumstances, no appeal for an individual assignment or project will be considered later than two weeks after the graded assignment was returned. For final grades, no appeal will be considered after one week of final project submission date.

Attendance

Attendance in all class sessions is expected, exactly as it would be on the job. Attendance in online classroom means that you will log into the course site regularly and read/view the lecture notes/recordings, and participate in discussion. If an emergency or illness occurs, have someone notify your team and the course instructor as soon as possible--even if you are out of town. If you are going to be absent from class or from team meetings you need to arrange to catch-up on what you missed and to make sure your part of the workload is covered. Too many absences are sufficient cause to lower the final course grade. Exceptions will be made for emergencies and other extenuating circumstances provided they are verified by appropriate documentation that is received no later than 1 week after the absence(s).

Academic Integrity

The academic community of Syracuse University and of the School of Information Studies requires the highest standards of professional ethics and personal integrity from all members of the community. Violations of these standards are violations of a mutual obligation characterized by trust, honesty, and personal honor. As a community, we commit ourselves to standards of academic conduct, impose sanctions against those who violate these standards, and keep appropriate records of violations. The academic integrity statement can be found at: http://www.ist.syr.edu/courses/advising/integrity.asp

Student with Disabilities

In compliance with section 504 of the Americans with Disabilities Act (ADA), Syracuse University is committed to ensure that "no otherwise qualified individual with a disability…shall, solely by reason of disability, be excluded from participation in, be denied the benefits of, or be subjected to discrimination under any program or activity…" If you feel that you are a student

who may need academic accommodations due to a disability, you should immediately register with the Office of Disability Services (ODS) at 804 University Avenue, Room 308 3rd Floor, 315.443.4498 or 315.443.1371 (TTD only). ODS is the Syracuse University office that authorizes special accommodations for students with disabilities.

Schedule

Date	Topics	Readings	Activities and Dues
Module 1: Overview of science data and data management			
1/17 Wk 1	Introduction to the course Science data life cycle and basic concepts	See readings for week 1	Sharing your experience in data handling
1/23 Wk 2	Fundamentals about data • Forms • Scales • Types • Levels Data structures and models • Physical data • Model data Data formats • Data format standards Representation of data Communication of data	See readings for week 2	Due (1/29 11 pm): Exercise 1 Identifying datasets
1/30 Wk 3	Describing datasets (1) • Metadata as solution to science data management • Metadata types • Metadata standards for scientific datasets	See readings for week 3	
2/6 Wk 4	Describing datasets (2) • Examples of metadata records for science datasets • Principles and requirements • Adoption of metadata standards	See readings for week 4	Due (2/12 11 pm): Exercise 2 Analyze and model metadata for datasets
2/13 Wk 5	Data provenance • Understanding science workflows • What is data provenance? • Provenance metadata • Data provenance and curation	See readings for week 5	Identifying a case for case study reports and final project (due any time this week)
2/20 Wk 6	Relational databases • Data attributes • Data relationships • Databases • Example data sets in relational databases	See readings for week 6	Due (2/26 11 pm): Exercise 3 Creating metadata record for datasets Quiz 1: Data management fundamentals

2/27 Wk 7	Managing data with repositories • Data management tasks • Data curation tasks • User requirements for data repositories • Technical requirements for data repositories	See readings for week 7	
3/5 Wk 8	Challenging issues in data repository services • Interoperability • Discovery and search • Ownership and access • Security • Evaluation	See readings for week 8	Due (3/11 11 pm): Exercise 4: Database structure and querying
3/12 Wk 9	Spring Break. No class.		
Module 2. Case studies: step-by-step guides			
3/19 Wk 10	Data use scenario analysis Guest speaker: Neil Sarkar	See readings for week 10	Start interviewing faculty for identifying data management needs
3/26 Wk 11	Data management scenario analysis Guest speaker: to be decided	See readings for week 11	Report 1 for data use based on interview results
4/2 Wk 12	Developing data management project (1): • Data set characteristic analysis • Needs assessment • User roles (researchers, lab staff, etc.) • Planning • Goals and objectives • Procedures • Policy development • Quality control	See readings for week 12	

4/9
Wk 13
	Developing data management project (2): Data organization and preservation
	• Metadata issues
	• Long-term preservation of data
	• Enabling technologies:
	  ○ for organizing and managing data
	  ○ for storing and retrieving data
	  ○ for using data
		See readings for week 13	Report 2 on data management based on interview results

Learning module 3: Using science data			
4/16 Wk 14	Data quality, discovery, and publishing • Read metadata description • Quality criteria • Data repositories and discovery • Directory services • Data publishing and sharing	See readings for week 14	Quiz 2: Using data

4/23 Wk 15	Data analysis • Data mining • Data meshing-up • Data presentation Visualization • Tools Formatting for Publication Using data • Ethics • Citing datasets	See readings for week 15	
4/30 Wk 16	Project presentations and discussions Wrap-up Post-course assessment survey		Due 5/6 11 pm: final group project