IST400/IST600 Science Data Management

Spring 2009
Course Time & Location:
3:45–5:05 p.m. MW
117 Hinds Hall
First Day of Class: January 12th

Instructor Information:
John D’Ignazio
Office: 114E Hinds Hall (in Student Services Suite)
Hours: Wednesdays, 1:00–3:00 p.m.
Phone: 443-5603
Email: jadignaz@syr.edu

Course Description
The Science Data Management course includes three modules:
1. fundamentals of science data and data management,
2. data management in aggregation, and
3. broader issues of science data including 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, and levels, data structures and models, data formats, and databases used to store, retrieve, and manage the data. The second module uses case studies regarding data collection, processing, transformation, and management to understand aggregations of data at three levels of organization: research, resource, and reference collections. In the last module, students will be introduced to methods and tools for evaluating data quality and working with data for use in various practice communities. Throughout the semester, students will work in an interdisciplinary team on a comprehensive science data management project under the instructor’s guidance. The performance of each student is based on exercises, quizzes, group reports, class participation, and the course project.

This course has been developed as part of the iSchool’s Science Data Literacy project with funding from the National Science Foundation’s Course, Curriculum and Laboratory Improvement program. For more information about the project, please visit the project website http://sdl.syr.edu.

Course Objectives
At the end of the course, students are expected to understand:
- concepts and characteristics of science data and practice
- 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 have hands-on experience in:
- identifying the needs for organizing, reporting, and managing science data
- describing data sets by using science metadata standards
- designing databases and systems for science data management and use
- evaluating data quality
Who Should Take the Course?

Undergraduate juniors/seniors or graduate students in any science, engineering, and information technology major may take this class. If you find you fit into the following description, this course would be for you:

- an interest in a career in science data management or science and engineering research in the corporate, academic, or government sector
- basic computer skills, including creating simple Web pages, using spreadsheets (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 via WebCT as electronic documents for reading and printing or distributed in print form in class. Students are responsible for familiarizing themselves with assigned materials before class to participate in discussions and apply the material in course assignments and projects.

Grading

The work for this class will involve a mixture of quizzes, individual assignments, group reports and a final project:

- **Quizzes** (3 x 5% = 15%) are designed to test your understanding of basic concepts in science data management.
- **Exercises** (4 x 7% = 28%) are designed for you to practice the necessary skills in carrying out data management project.
- **Group reports** (2 x 6% = 12%) are designed to maximize the usefulness of group discussion results.
- **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** (15%) includes your attendance and participation in class discussions and activities.

Course Management and Expectations

I try to make every class worth attending. Students will be responsible for all material 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 start of class, 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.

**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 = 94-100, A- = 90-93.9, B+ = 85-89.9, B =
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.

When a dispute over unequal group participation occurs, group-based assignments will 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 or attend the instructor’s office hours 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 5/15/2009.

**Attendance**

- Attendance in all class sessions (and throughout each session) is expected, exactly as it would be on a job. 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://supolicies.syr.edu/ethics/acad_integrity.htm](http://supolicies.syr.edu/ethics/acad_integrity.htm)

**Disability Statement**

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

<table>
<thead>
<tr>
<th>Semester</th>
<th>Course Date</th>
<th>Topics</th>
<th>Activities</th>
<th>Readings / Handout</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Module 1: Fundamentals of science data and data management</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Week 1</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Week 2</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>19-Jan</td>
<td>Martin Luther King, Jr. Day. No class</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>21-Jan</td>
<td>Intro to databases &amp; database programs &amp; data attributes</td>
<td>Rob, P. and C. Coronel (2004), chapter 1, pp. 6-25; chapter 2, pp. 30-65; Rob, P. and C. Coronel (2004), chapter 3, pp. 76-109 / on WebCT: Examples of data sets in relational databases,</td>
<td></td>
</tr>
<tr>
<td><strong>Week 3</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>28-Jan</td>
<td>Fundamentals about data: Forms, Scales, Types, Levels</td>
<td>In class presentations: Share your information analysis of data repository/database</td>
<td>NSF (2005), chapters 2&amp;3; Brooks, A. A. (1984) chapter 2</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Data structures and models: Physical data, Model data</td>
<td>Quiz 1: Data and database fundamentals (Th-Fri)</td>
<td>Lab: Designing a database</td>
</tr>
<tr>
<td>--------</td>
<td>-------</td>
<td>------------------------------------------------</td>
<td>-----------------------------------------------------------------</td>
<td></td>
</tr>
<tr>
<td></td>
<td>4-Feb</td>
<td>Describing datasets (1): introduction to Metadata: aboutness, types, elements, schemes, standards</td>
<td>In-class group work: investigate CSDGM profiles Due: Exercise 2 Database design</td>
<td></td>
</tr>
<tr>
<td></td>
<td>11-Feb</td>
<td>Describing datasets (3): metadata elements in schemes, tools. Case study</td>
<td>In-class group work: form use for data description</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Describing datasets (4): Relationship-Making and Relational Database lab</td>
<td>Hillmann, D. I. and E. L. Westbrooks (2004); Hill &amp; Janée Chapter 8; Arms &amp; Arms Chapter 14</td>
<td></td>
</tr>
<tr>
<td>Week 6</td>
<td>16-Feb</td>
<td>Managing data (1): Encoding, Storage, Import/Export, Cleaning, Transformation</td>
<td>Due: Exercise 1 &amp; 2 Revision Quiz 2: Data formats and description (Th-Fri)</td>
<td></td>
</tr>
<tr>
<td>Week 7</td>
<td>23-Feb</td>
<td>Managing data (3): Ownership and access, Data quality</td>
<td>Practice with SQL statements and data reports</td>
<td></td>
</tr>
<tr>
<td>-------</td>
<td>-------</td>
<td>-----------------------------------------------------</td>
<td>------------------------------------</td>
<td>--------------------------------------------------</td>
</tr>
</tbody>
</table>

**Spring Break**

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>18-Mar</td>
<td>Data aggregation scenario (3): resource collection</td>
<td>Guest speaker</td>
<td>Due: Exercise 4 XML for metadata and data records</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>25-Mar</td>
<td>Data and users (2): data set characteristic analysis, needs assessment</td>
<td>In-class group work</td>
<td>Case handouts to be delivered in class</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Week 11</th>
<th>30-Mar</th>
<th>Organizational planning (1): Goals and objectives, Procedures, Quality control</th>
<th></th>
<th>Muller, R. J. (1999). Chapter 7</th>
</tr>
</thead>
</table>

**Learning module 3: Broader issues in science data management**
<table>
<thead>
<tr>
<th>Week 12</th>
<th>6-Apr</th>
<th>Enabling technologies: organizing and managing data, storing and retrieving, using data</th>
<th>Group report 2 on interview result</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>8-Apr</td>
<td>Understanding Data Curation: metadata description, quality criteria, archival concepts</td>
<td></td>
</tr>
<tr>
<td></td>
<td>22-Apr</td>
<td>Sharing data: Ethics, Publishing, Citation</td>
<td>Uhlic, P. F. (2003)</td>
</tr>
<tr>
<td>Week 15</td>
<td>27-Apr</td>
<td>Project presentations and discussions, Wrap-up</td>
<td>Final Project Due; Post-course assessment survey</td>
</tr>
</tbody>
</table>

**Reading List**

**Book Chapters:**


- Chapter 2: A Syntax for Data. pp. 11–26
- Chapter 3: XML. pp. 27–50


Chapter 59: XML Path Language; pp. 844–871
Chapter 60: Extensible Stylesheet Language(XSL); pp. 844–893
Chapter 61: XML Pointer Language (XPointer); pp. 894–900
Chapter 62: XML Linking Language (XLink); pp. 902–917


Chapter 3: Gathering Requirements, pp. 55–74
Chapter 4: Modeling Requirements with Use Cases, pp. 75–98
Chapter 6: Building Entity-Relationship Models, pp. 105–125
Chapter 7: Building Class Models in UML, pp. 127–184


Chapter 3: Data Models, pp. 28–71
Chapter 3: The Relational Database Model pp. 74–121
Chapter 6: Structured query language, pp. 226–316


Chapter 1: Introduction pp. 3–13
Chapter 3: Schemas—Structure and Semantics pp. 87–129
Chapter 4: Schemas and Syntax pp. 131–147
Published Articles:


Reports (Institution & Government):


Chapters 2: The Elements of the Digital Data Collections Universe, pp. 17–23
Chapters 3: Roles and Responsibilities of Individuals and Institutions, pp. 25–30