

**Syllabus for INF350E (27890): Introduction to Scientific Data Informatics**  
(Crosslisted with INF382R (27950))  
(Spring 2017)

**Class meetings:** Thursdays 12-3pm, [UTA](#) 1.212

**Instructor:** William L. Anderson

**Office Hours:** By appointment.

**Email:** band at austin dot utexas dot edu

**See also:** [Course Schedule](#) - up to date weekly schedule of readings and assignments.

**Catalog Description:** Introduction to the characteristics of scientific data and the emerging practices applied toward their management and preservation.

The class will include (1) an introduction to the information properties of scientific data, and scientific data modeling; (2) an introduction to emerging semantic web approaches of describing these data sets; (3) hands-on assessments and evaluations of public scientific data sets; and (4) discussions of current issues in scientific data sharing, publication, and preservation.

**Course summary:**

To quote Godfray and Knapp (H.C.J. Godfray and S. Knapp (2004), Introduction to 'Taxonomy for the twenty-first century', <http://dx.doi.org/10.1098/rstb.2003.1457>):

"Any student of data needs to learn the tools and techniques needed to identify and describe the various elements and properties of data objects and data collections."

The goal of this class is to facilitate such learning. The focus of the class is on the information properties of scientific and technical data and collections of such data. In addition, we will devote time to learning about some methods and tools associated with data science. We will use existing libraries in the statistics system R to explore what these tools can provide for data informaticians. Those of us aspiring to be data informaticians need familiarity with emerging trends in research data and big-data domains.

A course synopsis: An introduction to scientific data informatics, in which we (1) examine the information properties of scientific data; (2) develop criteria to appraise the properties of scientific data sets, (3) learn about and use command-line data science tools and R libraries to answer data informatics questions; (4) take an expedition into the Semantic Web and the networks of Linked Data, and learn about their use in the sciences; (5) examine issues of long-term management of, and access to, scientific data from the perspectives of working scientists; and (6) summarize our own learnings from an inquiry-based project.

The course work includes weekly readings from information and computer science and systems literature, as well as current active science and science policy and practice media, and social media. Weekly news and views are collected on a class wiki. There are a set of assignments to

provide hands-on experience with data informatics tools. The data science and R sections also include hands-on assignments. There is a major project that results in a report and presentation. The entire scientific and research data area is in a period of growth, transformation, and development. I look forward to learning more about it with you.

**Prerequisites:** Other than comfort in using computers for common research and writing work, there are no prerequisites for this class.

**Textbook:** There is one required textbook: *A Vast Machine: Computer Models, Climate Data, and the Politics of Global Warming* (2010) by Paul N. Edwards (available in the Coop Bookstore and from online bookstores).

There is one recommended textbook from the National Academies Press: *Preserving Scientific Data on Our Physical Universe: A New Strategy for Archiving the Nation's Scientific Information Resources* (1995). A PDF version of this book can be downloaded from the National Academy Press website by following the download instructions at this link (you may need to create a free account).

**Final exam scheduling:** This class does not have a final exam.

---

## **Grading:**

### **Weekly News and Views (30% of final grade) Due each week.**

Each week we present to the class \*and\* post on the class wiki, short items related to S&T data description, management, preservation, policy, etc., that we have found in news, research, web, and social media sources. Each link is accompanied by a short summary in our own words of its content and relevance to scientific data informatics.

### **Exercises (20% of final grade) Occasional.**

There will several assigned exercises during the term on modeling, research data collection evaluation, data science and informatics tools, and the semantic web. These exercises include assessing usability properties of data sets, developing concept maps, using command-line tools to explore informatics questions, and applying RDF-based ontology markup to text content.

### **Term project and class presentation (40% of final grade) Due 7 May 2016, 23:59 CDT.**

Students are required to present to the class, the results of the agreed upon project work. A record of the project work is required to be deposited on designated Canvas pages. Any references cited during the project will be included on the appropriate Canvas page and marked up using a semantic citation ontology.

### **Discussion participation (10% of final grade).**

Students are expected to participate in class discussions of reading assignments.

Late assignments are not accepted. In case of university-excused absence (see below), work may be submitted electronically or in hard-copy by the start of the next class meeting.

See iSchool general guidelines on [Grades & Grading](#). iSchool students must receive B or higher to order to include this course in their program of work toward graduation.

Students from other departments should refer to their program's appropriate guidelines. The UT Graduate School requires a minimum grade of C or higher to count a course for credit. For detailed information please refer to UT's [Graduate Catalog](#).

---

### **Weekly readings**

Several articles are assigned each week for reading and review. Some will be original published research, some will be literature surveys on particular topics, and in a few cases we'll read presentation slides.

Links to readings can be found on the Course Schedule document (see above). **If you encounter a broken link when looking for an assigned reading, try searching for the paper online**, and please notify the instructor so an updated link can be posted. Readings without links contain sufficient citation information to be found by searching online. Notify the instructor of any discovery and/or access issues you encounter.

**Some papers are available only through university subscription (restricted access)** requiring you to (a) download while you are on campus, or (b) use UT's [electronic proxy](#) or [VPN connection](#) from off-campus (if you haven't used the proxy before, it's worth setting up to access UT libraries and restricted materials when off-campus). Plan ahead.

Readings assigned for a given class are discussed in that class meeting, and are expected to be read before class. The discussion in class is intended to support understanding of the course material. Everyone is expected to participate in every class.

---

## Standard UT Austin Course Information & Policies

### Academic Honor Code

You are encouraged to discuss assignments with classmates, but anything submitted must reflect your own, original work. If in doubt, ask the instructor. **Plagiarism and similar conduct represents a serious violation of UT's Honor Code and standards of conduct:**

- <http://deanofstudents.utexas.edu/conduct/standardsofconduct.php>

**Students who violate University rules on academic dishonesty are subject to severe disciplinary penalties, such as automatically failing the course and potentially being dismissed from the University. *\*\*PLEASE\*\** do not take the risk.**

**Assignments submitted to Canvas are automatically checked for plagiarism.** We are REQUIRED to report any suspected case to central administration for investigation and disciplinary hearings. Honor code violations ultimately harm yourself as well as other students, and the integrity of the University. Academic honesty is strictly enforced. For more information, see the Student Judicial Services site: <http://deanofstudents.utexas.edu/sjs>.

### Notice about students with disabilities

The University of Texas at Austin provides appropriate accommodations for qualified students with disabilities. To determine if you qualify, please contact the Dean of Students at 512-471-6529 or UT Services for Students with Disabilities. If they certify your needs, we will work with you to make appropriate arrangements. UT SSD Website: <http://diversity.utexas.edu/disability/>

### Emergency Preparedness

Any students requiring assistance in evacuation must inform the instructor in writing of their needs during the first week of classes. This must then be provided to the Fire Prevention Services office by fax (512-232-2759), with "Attn. Mr. Roosevelt Easley" in the subject line.

### Coping with stress and personal hardships

The [Counseling and Mental Health Center](#) offers a variety of services for students, including both individual counselling and [groups and classes](#), to provide support and assistance for anyone coping with difficult issues in their personal lives. As mentioned above, life brings unexpected surprises to all of us. If you are facing any personal difficulties in coping with challenges facing you, definitely consider the various services offered and do not be shy to take advantage of them if they might help. These services exist to be used.

### Notice about missed work due to religious holy days

A student who misses an examination, work assignment, or other project due to the observance of a religious holy day will be given an opportunity to complete the work missed within a reasonable time after the absence, provided that he or she has properly notified the instructor. It is the policy of the University of Texas at Austin that the student must notify the instructor at

least fourteen days prior to the classes scheduled on dates he or she will be absent to observe a religious holy day. For religious holy days that fall within the first two weeks of the semester, the notice should be given on the first day of the semester. The student will not be penalized for these excused absences, but the instructor may appropriately respond if the student fails to complete satisfactorily the missed assignment or examination within a reasonable time after the excused absence.

### **Electronic mail Notification Policy**

In this course the [Canvas](#) Announcements are used as used as a means of communication with students. You are responsible for checking your Canvas e-mail regularly for class work and announcements. If you are an employee of the University, your e-mail address in Canvas is your employee address.

The University has an official e-mail student notification policy. It is the student's responsibility to keep the University informed as to changes in his or her e-mail address. Students are expected to check e-mail on a frequent and regular basis in order to stay current with University-related communications, recognizing that certain communications may be time-critical. The policy is found here: <http://www.utexas.edu/cio/policies/university-electronic-mail-student-notification-policy>.

You can find and change your official email address of record at [https://utdirect.utexas.edu/apps/utd/all\\_my\\_addresses](https://utdirect.utexas.edu/apps/utd/all_my_addresses)