

INFX 551

Course Introduction & Overview
01.01.2018

- Me + An (your TA)
- Course Objectives + Expectations
- Course Format
- Assignments
- Grades

About Me

- Phd in Information Science from the University of Illinois.
- My research includes Open Data, Open Science, and Information Infrastructures for Transparency

About An Yan

- PhD Student in the iSchool + DataLab
- Background in climate science (but we converted her to Information Science!)
- Research: Open Government Data + Open Science Data

INFX 551

- Survey course - meant to **introduce** you the the ideas, practices, and tools of data curation.
- Data Curation (a simple definition): The active and on-going management of data throughout a lifecycle of use, and reuse in an anticipated contexts.
- The challenge of data curation is simple in form, but complex in application: How can we systematically improve the presentation and preservation of data... such that the data required for scholarship, governance, and business is both relevant for intended uses, and viable for future reuse?

Types of Knowledge

Declarative

I know **that** something is the case...

I know that longitude is the geographic coordinate that specifies the east-west position of a point on the Earth's surface.

Often precedes and informs procedural knowledge

I know that XML is an encoding standard for structuring data...

Procedural

I know **how** to do something...

I know how to locate a point on a map given the Longitude and Latitude coordinates...

Often follows from declarative knowledge (but not always)

I know how (and when it is appropriate) to create well-formed XML encodings of free text data.

Course Outline

This course consists of three modules.
Each module consists of three topics.

| Data | Data Systems | Policy, Privacy, & Ethics |
|--------------------------|--------------|---------------------------|
| <u>Types & Roles</u> | Repositories | Policy |
| <u>Documentation</u> | Preservation | Privacy |
| <u>Standardization</u> | Cost Models | Ethics |

Course Engagement

- Lecture (Monday)
- Lab / Demo / Tech Lecture (Wednesday)
- Chat / Office Hours (Friday)
- Forum posts (any day all day)

You can certainly work ahead in each module; but you will get much more out of the class if you follow our weekly schedule.

Module Schedule

| Monday | Tuesday | Wednesday | Thursday | Friday | Saturday | Sunday |
|---------|---------|-----------|----------|--------|----------|-----------------|
| Lecture | | Lab | | Chat | | Homework Due |

Module Schedule

| 1 | Monday | Tuesday | Wednesday | Thursday | Friday | Saturday | Sunday |
|----------|---------|---------|-----------|----------|--------------|----------|-----------------|
| | Lecture | | Lab | | Chat | | Homework Due |
| 2 | Monday | Tuesday | Wednesday | Thursday | Friday | Saturday | Sunday |
| | Lecture | | Lab | | Chat | | Homework Due |
| 3 | Monday | Tuesday | Wednesday | Thursday | Friday | Saturday | Sunday |
| | Lecture | | Lab | | Quiz Open | | Quiz Due |

- **Homework** (6 total) - Mix of written statements (declarative) and exercises (procedural). Submit file to canvas. [1 hour]
- **Quiz** (3 total) 10 short-answer questions. [1 hour]
- **Data Profile** - Structured description of a dataset in the wild. This is meant to help your evaluative skills in curating data. (declarative)
- **Curation Protocol** - You will be given a data source to curate. The curation will require you to transform, clean, and create metadata about the source. You will record these steps in a “protocol” that would allow anyone to reproduce your work. (procedural)

- **Homework** - 30% of your grade. (6 total x 5% each). 0 = no credit; 2.5 = partial credit; 5 = full credit.
- **Quiz** - 30% of your grade (3 total x 10% each).
- **Data Profile** - 20% of your grade. Jan 17th (forum post); Jan 31st profile due.
- **Curation Protocol** - 20% of your grade. Due March 7.

Getting in Touch with your Instructors

General Questions (Will anyone else have this question?) : Forums

Specific personal questions (This is a *me* question) : Email both instructors

Grading question: Email both instructors.