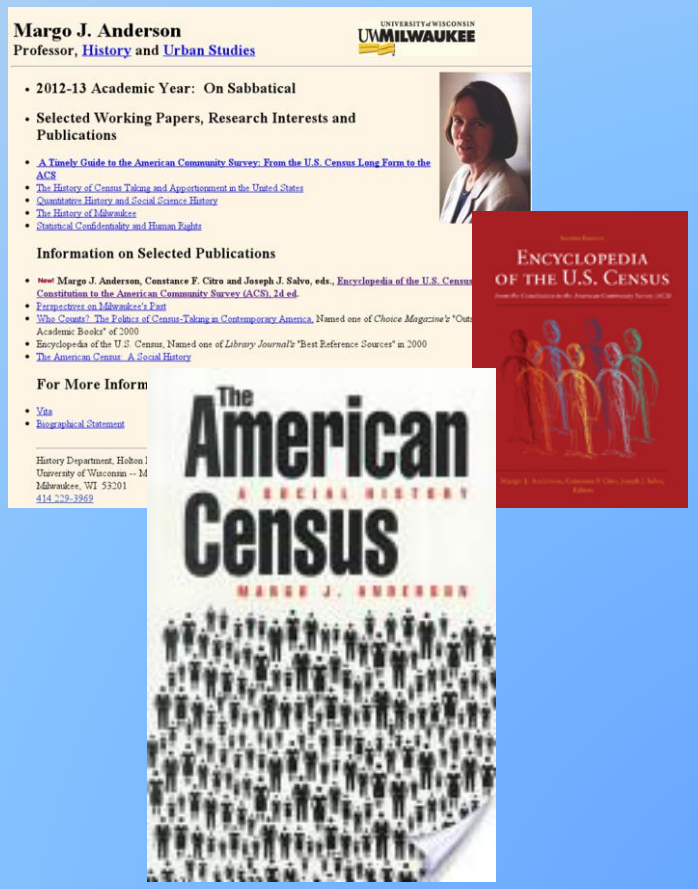
INFO 7470/ECON 7400/ILRLE 7400 Understanding Social and Economic Data

John M. Abowd and Lars Vilhuber

January 21, 2013

Session 1: History and Current State of the Federal Statistical System

- Overview of the (U.S.) federal statistical infrastructure, and how it came to be
- Guest lecture by Margo Anderson



Session 2: Technical Statistical Terminology and Tools

- Censuses, surveys, administrative records, contextual data, genomes, spatial records, web sources

- Populations
- Frames
- Sampling
- Coverage
- Bias
- Other errors

Session 3: Measuring People and Households

- Censuses of population
- Goals and methods
- Guest lecture by Warren Brown

- U.S. Decennial Census of Population and Housing
- American Community Survey
- Current Population Survey
- American Housing Survey

Session 4: Measuring Business and Economic Activity

- The fundamental concepts of the national income and product accounts
- Business entities
- Frame management
  - Census
  - BLS
- Births and deaths
- The Employer Business Register
- Economic Censuses
- Establishment Surveys

Session 5: Enhancing Traditional Methods

- Health surveys, Program participation surveys
  - SIPP
  - NHIS
  - Benefit recipient surveys
- Survey-driven linkages
  - Validation studies (CPS, PSID)
  - Two-stage sampling schemes (Canada's Workplace and Employee Survey, WES)

Session 5: Introduction to Integrated Data Systems

- Validation, augmentation using administrative data
  - SIPP linked to SSA
  - Retirement History Survey (RHS) linked to SSA
  - Health and Retirement Survey (HRS) linked to SSA
- Integrating data from multiple sources
  - What is it?
  - Key tools
  - Where is it applied

Session 6: 21st Century Statistical Systems

- Integrated administrative data systems
  - Longitudinal Employer-Household Dynamics (LEHD) data
  - IRS linked data
- Register-based Censuses
  - Variants in Europe
  - Canada: augmenting the Census with administrative data
  - US: planning 2020?

Session 6: 21st Century Statistical Systems

- Non-traditional data collection methods
  - Administrative
  - Electronic (web)
  - Non-traditional sources (Google, Twitter, etc.)
- Confidentiality and access methods
  - RDCs
  - Enclaves
  - Methods of gaining access
  - Justifications for gaining access
  - Learning about confidential data

Session 7: Replicable science

- How do you find literature?
  - (assumed to be known)
  - Review of how to cite literature

- Assessing quality of replicability
- Some tools

- How do you find data?
  - Less well developed, if at all
  - Referencing data used in articles
  - Developing standards on how to cite data
  - The conundrum of how to cite confidential data

Sessions 8-10: Statistical tools

- Edit and imputation
  - Why and when?
  - Methods
  - Do it yourself!
- Record linkage
  - Why, when, and what?
  - Methods and tools
  - Do it yourself!

Sessions 8-10: Statistical tools

- Disclosure limitation methods
  - Why and when?
  - Methods
  - Do it yourself!

- Session 13
  - Releasing synthetic data combines many of these tools:
    - Extreme case of imputation
    - Use as a disclosure limitation method
    - Record linkage as a way to prove that protections are valid

Sessions 11-13: More Tools

- Geographic Information Systems
- Guest lecture by Nicholas Nagle
- Basic Geocoding
- Spatial data analysis methods

Sessions 11-13

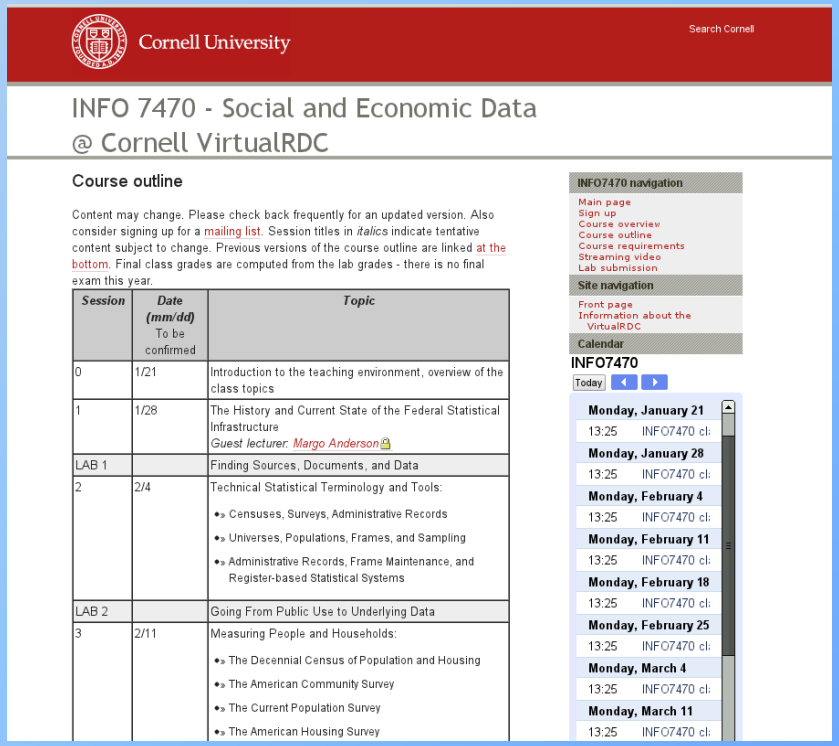
- Modeling integrated data
- The relational database model
- Alternative representations: graphs, networks
- Bayesian methods for edit, imputation, estimation
- Synthetic data used to represent the simulation outputs

TECHNICAL SETUP

Technical setup

Primary website:

http://www.vrdc.cornell.edu/info7470/



Cornell University

Search Cornell

INFO 7470 - Social and Economic Data
@ Cornell VirtualRDC

Course outline

Content may change. Please check back frequently for an updated version. Also consider signing up for a mailing list. Session titles in *italics* indicate tentative content subject to change. Previous versions of the course outline are linked at the bottom. Final class grades are computed from the lab grades - there is no final exam this year.

Session	Date (mm/dd) To be confirmed	Topic
0	1/21	Introduction to the teaching environment, overview of the class topics
1	1/28	The History and Current State of the Federal Statistical Infrastructure *Guest lecturer: Margo Anderson*
LAB 1		Finding Sources, Documents, and Data
2	2/4	Technical Statistical Terminology and Tools: •» Censuses, Surveys, Administrative Records •» Universes, Populations, Frames, and Sampling •» Administrative Records, Frame Maintenance, and Register-based Statistical Systems
LAB 2		Going From Public Use to Underlying Data
3	2/11	Measuring People and Households: •» The Decennial Census of Population and Housing •» The American Community Survey •» The Current Population Survey •» The American Housing Survey

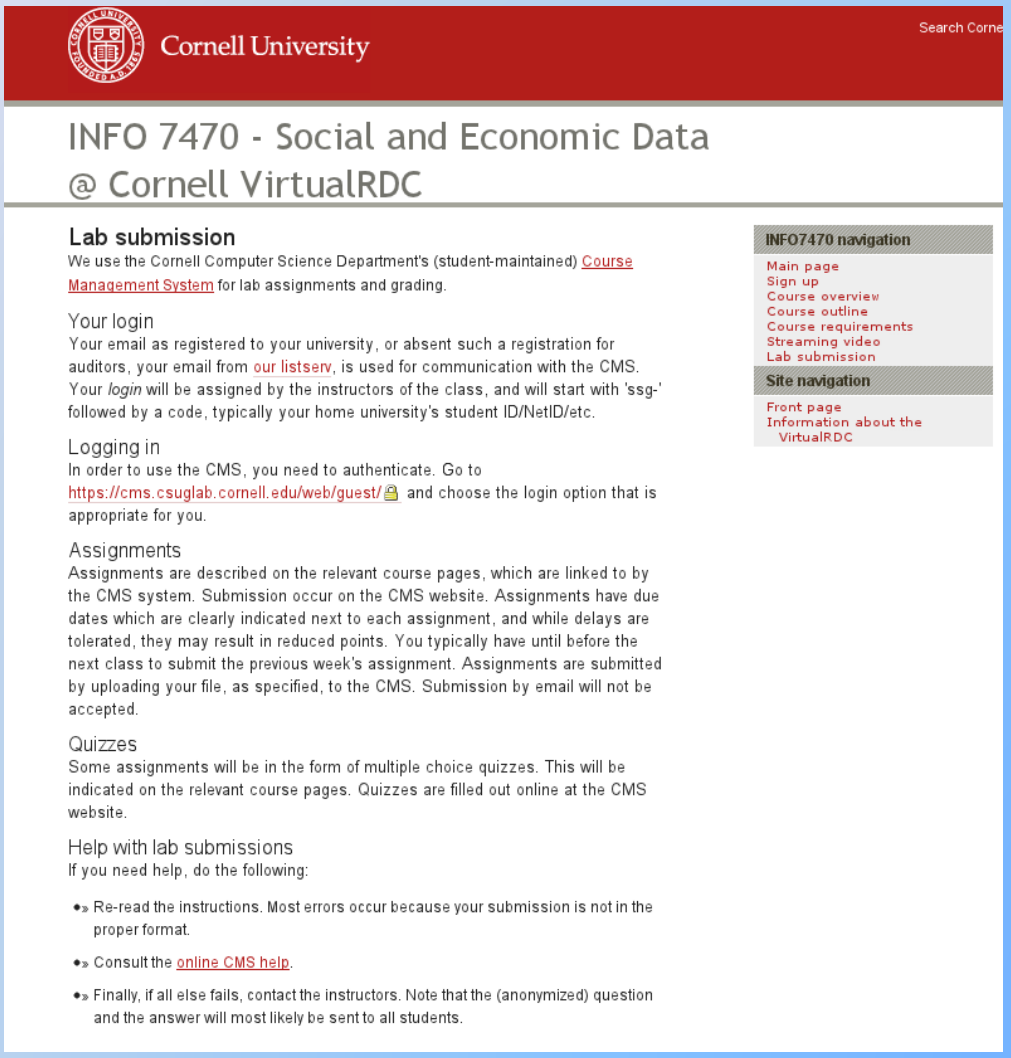
INFO7470 navigation
Main page
Sign up
Course overview
Course outline
Course requirements
Streaming video
Lab submission

Site navigation
Front page
Information about the VirtualRDC

Calendar
INFO7470
Today

Monday, January 21
13:25 INFO7470 cl:
Monday, January 28
13:25 INFO7470 cl:
Monday, February 4
13:25 INFO7470 cl:
Monday, February 11
13:25 INFO7470 cl:
Monday, February 18
13:25 INFO7470 cl:
Monday, February 25
13:25 INFO7470 cl:
Monday, March 4
13:25 INFO7470 cl:
Monday, March 11
13:25 INFO7470 cl:

Submitting labs and quizzes



Cornell University

Search Corne

INFO 7470 - Social and Economic Data @ Cornell VirtualRDC

Lab submission

We use the Cornell Computer Science Department's (student-maintained) Course Management System for lab assignments and grading.

Your login

Your email as registered to your university, or absent such a registration for auditors, your email from our listserv, is used for communication with the CMS. Your *login* will be assigned by the instructors of the class, and will start with 'ssg-' followed by a code, typically your home university's student ID/NetID/etc.

Logging in

In order to use the CMS, you need to authenticate. Go to https://cms.csuglab.cornell.edu/web/guest/ and choose the login option that is appropriate for you.

Assignments

Assignments are described on the relevant course pages, which are linked to by the CMS system. Submission occur on the CMS website. Assignments have due dates which are clearly indicated next to each assignment, and while delays are tolerated, they may result in reduced points. You typically have until before the next class to submit the previous week's assignment. Assignments are submitted by uploading your file, as specified, to the CMS. Submission by email will not be accepted.

Quizzes

Some assignments will be in the form of multiple choice quizzes. This will be indicated on the relevant course pages. Quizzes are filled out online at the CMS website.

Help with lab submissions

If you need help, do the following:

- Re-read the instructions. Most errors occur because your submission is not in the proper format.
- Consult the online CMS help.
- Finally, if all else fails, contact the instructors. Note that the (anonymized) question and the answer will most likely be sent to all students.

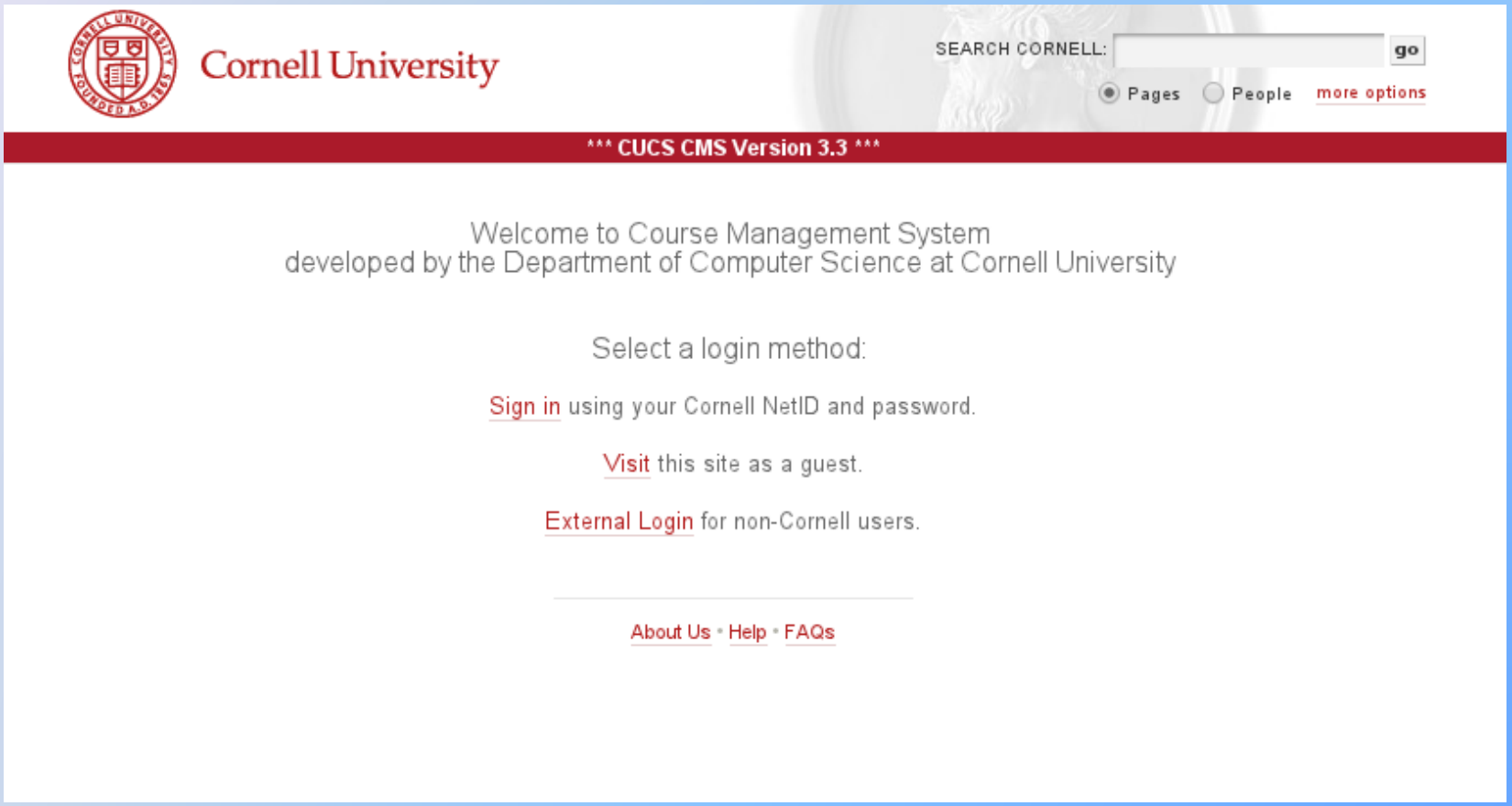
INFO7470 navigation

Main page
Sign up
Course overview
Course outline
Course requirements
Streaming video
Lab submission

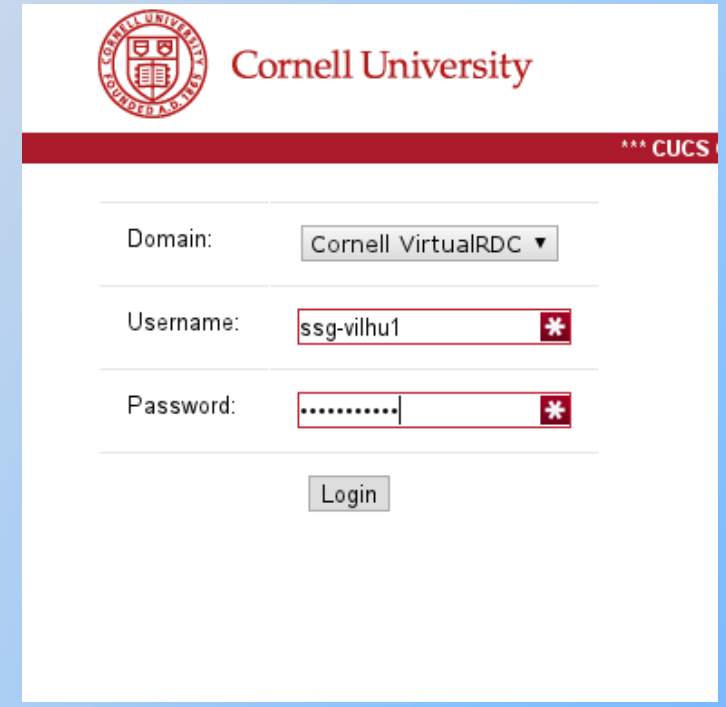
Site navigation

Front page
Information about the VirtualRDC

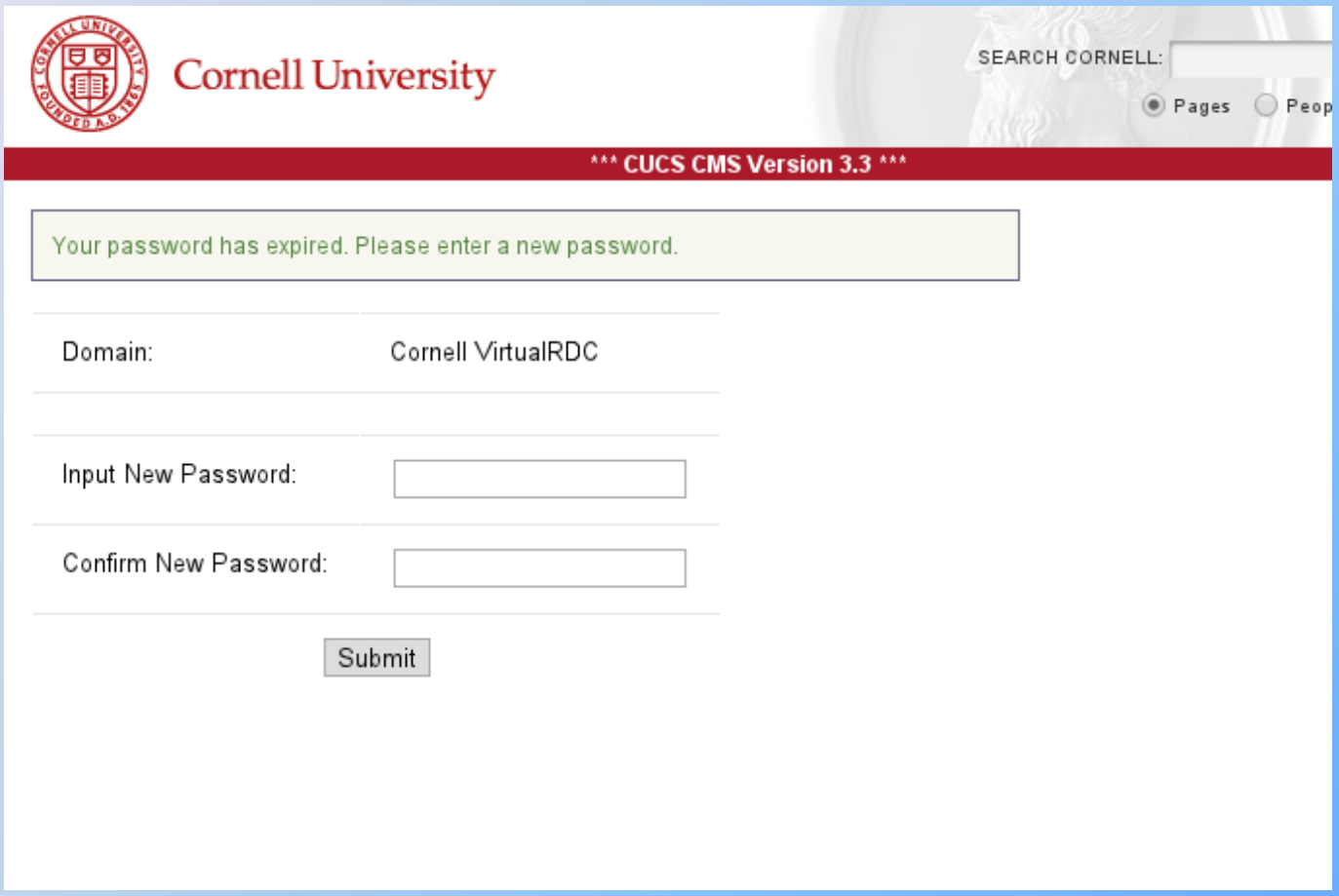
Course Management System



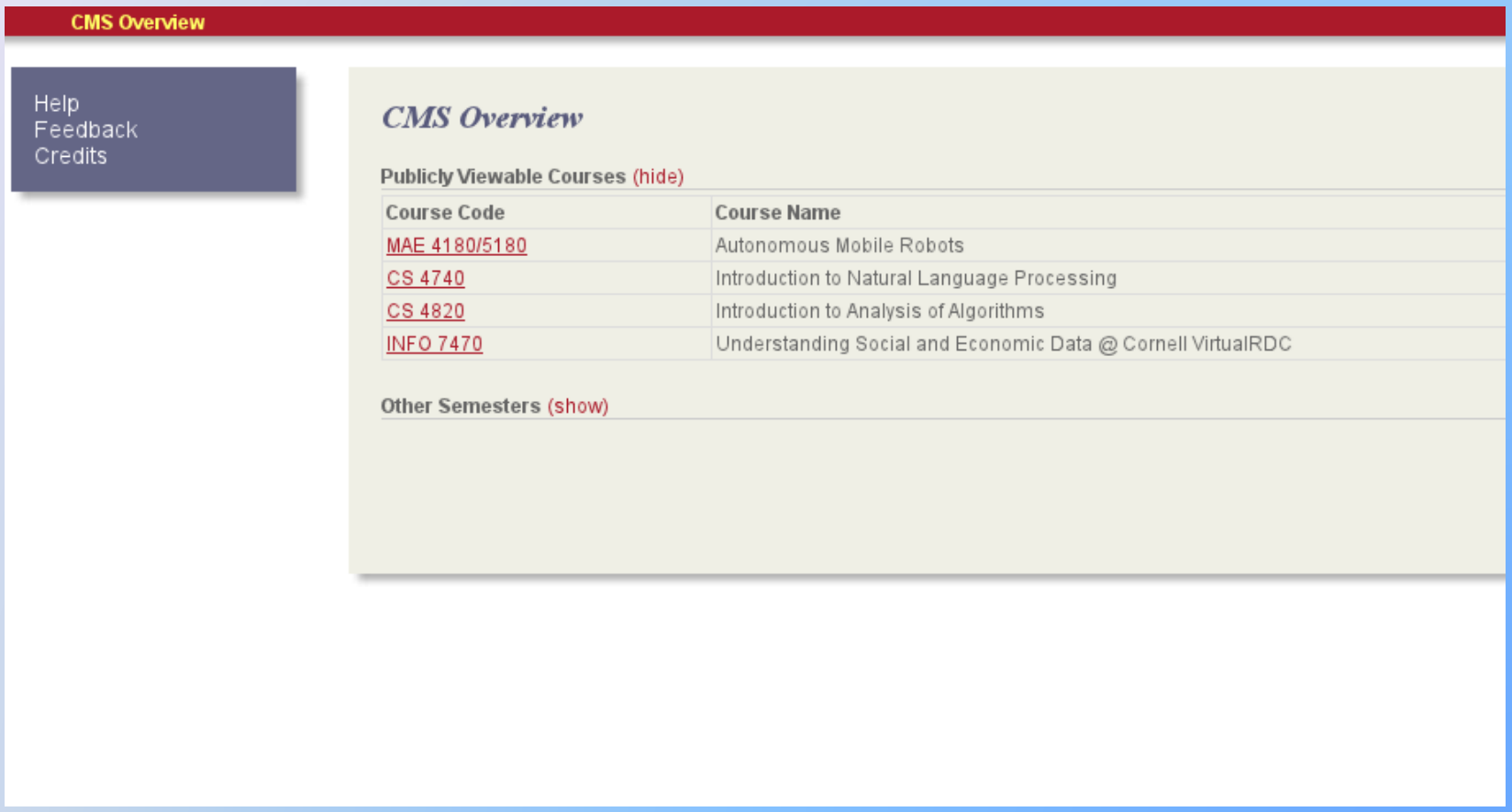
Entering the CMS



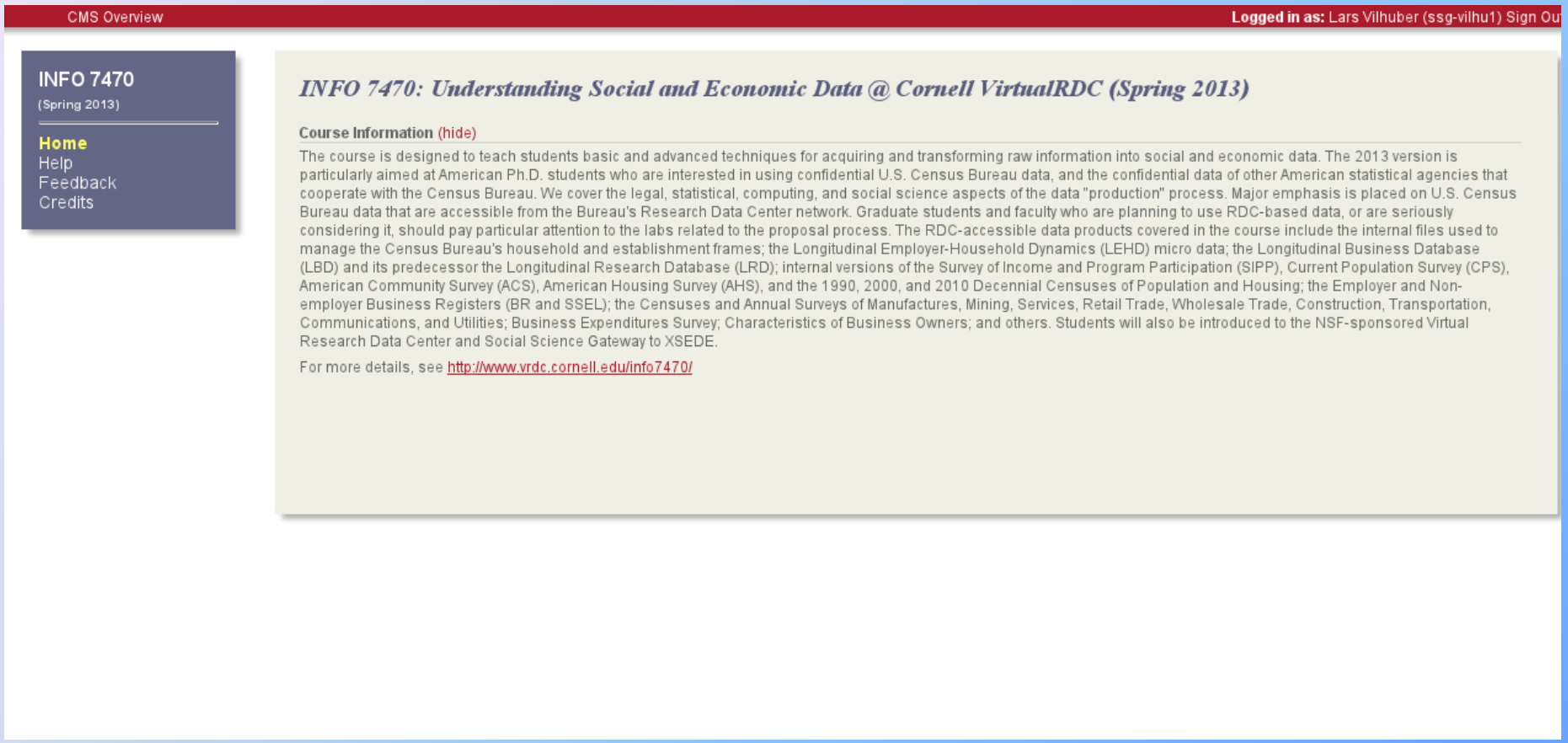
First-time CMS password reset



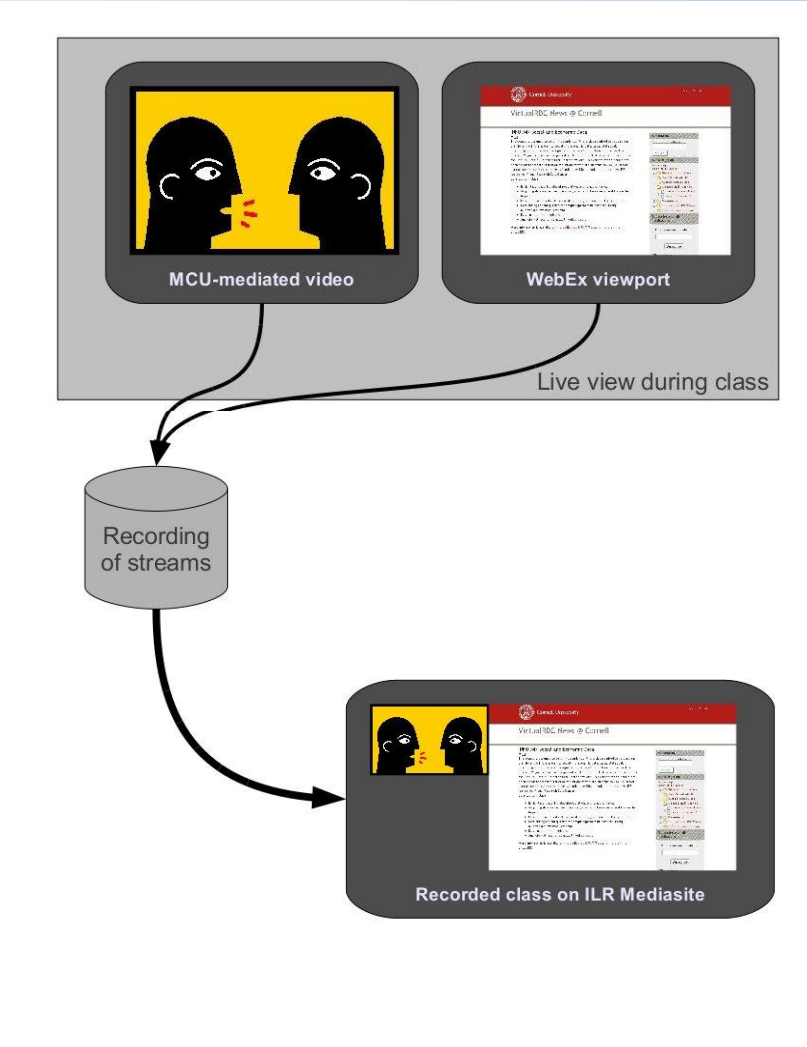
In the CMS


CMS Overview
Help
Feedback
Credits
CMS Overview
Publicly Viewable Courses (hide)
Course Code
Course Name
MAE 4180/5180
Autonomous Mobile Robots
CS 4740
Introduction to Natural Language Processing
CS 4820
Introduction to Analysis of Algorithms
INFO 7470
Understanding Social and Economic Data @ Cornell VirtualRDC
Other Semesters (show)

In the CMS course page



Video setup



- Two-screen setup
  - Screen 1 will have people
  - Screen 2 will have slides/live demos/etc
  - Is this what you are seeing now?
- Recording will merge both streams

Video etiquette

- Please mute your mike!

(Source: http://www.flickr.com/photos/raquelcamargo/3296054642/CC BY-NC 2.0)

Video etiquette

- If asking questions, try and get close to the mike
- If speaking, try and get the camera focussed on the person asking the question

Recording

- We will be recording all sessions
- Recording will focus primarily on the main camera, but all pictures and sounds are liable to be recorded and made available in the recorded classes
- We will edit the recordings after the class ends, with the goal of making a MOOC-like experience possible in 2014.
- All presenters will be asked for permission; other participants will not appear in those recordings