

# Syllabus

## Scientific Data-Collection for Business (BMGT883)

Fall 2009, Instructor: Prof. Galit Shmueli

**Office:** 4361 Van Munching Hall

**Contact:** [gshmueli@rhsmith.umd.edu](mailto:gshmueli@rhsmith.umd.edu)

**Office hours:** Wed 11:00-12:00, by appointment

**Class hours:** Wed 13:00-15:40

**Class location:** VMH 3509 (eMarkets Research Lab)

### A. DESCRIPTION

This course focuses on collecting data for scientific research in the Internet age. In particular, it covers three popular data collection methods: experiments, surveys, and observational data. Although these have been the most common data collection methods in the offline world, the uniqueness of this course is that it discusses them in an online context where the data and/or the collection mechanism are online. In particular, we will discuss Internet experiments, online and email surveys, and web crawling and other tools for collecting Internet data.

Emphases of the course will be on designing efficient and ethical data collection plans and executing them. We will discuss statistical, ethical, technological and practical aspects of collecting data in an online environment. We will also cover the academic approval process necessary for conducting human subjects research in the USA.

The course is intended for any PhD student in the business school who will be performing empirical research.

### B. ORGANIZATION

This is a lecture-lab course in which topics are presented by the instructor, and students learn by hands-on homework and projects. Several readings will be assigned, and students are expected to be prepared to discuss them in class.

### C. OBJECTIVES

1. To introduce students to different data collection tools and the suitability of the generated data for different research purposes.
2. To introduce students to data collection in an online environment, and its ethical, technological, and statistical aspects.
3. To educate students about the proper use and evaluation of online data collection tools.
4. To educate students about collecting data in an academic environment.
5. To provide a hands-on experience for learning about the various pitfalls of data collection.
6. To provide students with an opportunity to collect data in practice, before they pursue collecting their own data for their dissertation work.

### D. TOPICS

1. Introduction: Data collection methods; Types of data and their research value
2. Collecting Web data
3. Online and email surveys
4. Online experiments

## E. READINGS

Mandatory: A set of articles will be given through Blackboard.

Suggested:

- Hill and Lewicki, "Statistics: Methods and Applications", Chapter on Experimental Design (available online at <http://www.statsoft.com/textbook/stexdes.html>)
- Best SJ and Krueger BS, *Internet Data Collection* (ISBN: 0761927107)
- Fowler JF *Survey Research Methods*, 3rd edition (ISBN: 0761921907)
- Lohr S *Sampling Design and Analysis* (ISBN: 0534353614)

## F. RESOURCES

### Course Management System

The course will be hosted on Blackboard, and all materials will be posted there (assignments, readings, grades, etc.)

### Software

We will use several software programs and packages in the course:

- 1) Google Spreadsheet Surveys and/or Survey Monkey (available online)
- 2) Minitab (available on the Smith Portal)
- 3) PHP (open source)
- 4) TIBCO Spotfire (available in the lab and for personal download)

**Computer Facilities:** The course meetings will take place in the eMarkets Research Lab (VMH3509). Students will have access to the lab and will be able to use its computers and software for working on projects and homework assignments. Alternatively, students can use their own computers and use the required software either by installing it on their computer, or by using it via the Smith Portal.

## G. COURSE POLICIES

### Attendance

Students are expected to attend all class meetings. A student who misses class for a good reason is responsible for delivering any homework or other deliverable on time. Attendance on the last Presentations Session is mandatory. Absence from this session will reduce your final letter grade.

### Homework Assignments

Students are expected to complete all homework assignments and to submit each homework by the delivery date. Unless specified otherwise, students should complete the homework assignments on their own, without consulting with others.

### Final Project

The final project in the course combines the various aspects taught during the class. Each student will work separately on a project, and will present it during the final Presentation Session. In addition to the presentation, students will write a professional project report, to be submitted within two days from the presentation.

## H. GRADING

Final project presentation	15%
Final project report	15%
Wiki project page	10%
Homework assignments	5% each (total 60%)

**COURSE SCHEDULE (TENTATIVE): timing of topics will likely change**

<b>Week</b>	<b>Topic</b>	<b>Deliverable</b>
2-Sep-09 Week 1	<b>Intro to data collection:</b> data types, collection methods, research goals <b>Course overview</b> and the Internet context <b>Collecting web data:</b> Intro; Why? What? How? HTML basics	
9-Sep-09 Week 2	<b>Web crawling:</b> Goals, rules and ethics PHP: installation, writing programs, saving webpages, parsing	Assignment 1 (web) Read: Allen et al. (2006) Wiki: research Q
16-Sep-09 Week 3	<b>Web crawling:</b> PHP (regExp, writing data to a file, grabbing multiple pages); Hidden web; Web crawling vs. other technologies: <b>RSS and Web services</b> (API); XML format	Assignment 2 (web) Web crawling Read: Bapna et al. (2006), Kenett & Shmueli (2009) Wiki: Information quality
23-Sep-09 Week 4	<b>Web services / API</b> <i>In class:</i> <b>Surveys:</b> Intro	Assignment 3 (web) Web crawling & RSS Wiki: Population of interest
30-Sep-09 Week 5	<b>Surveys:</b> Intro, concepts, sampling and non-sampling errors; Ethics and IRB	Assignment 4 (web) API Read: Shmueli et al. (2005) Wiki: Sampling concepts
7-Oct-09 Week 6	<b>Questionnaire design:</b> validity & reliability; pre-testing; pilot study; focus groups Online survey tools; email vs. Internet survey	Assignment 5 (surveys) Sampling concepts Read: FTC report Project: Create frame, IRB form
14-Oct-09 Week 7	<b>Surveys:</b> Improving response rate; solicitation email <b>Sample design:</b> Sampling from finite populations	Assignment 6 (surveys) Questionnaire design Read: FTC report Project: create questionnaire
21-Oct-09 Week 8	<b>Sample design:</b> SRS; finite population; bias & variance	Assignment 7 (surveys) Solicitation email and pretesting IRB form – cont.
28-Oct-09 Week 9	<b>Sample design:</b> Stratified sampling (sample size allocation, post-stratification); cluster sampling	Assignment 8 (surveys), Sample design Wiki: sample design
4-Nov-09 Week 10	<b>DOE:</b> Intro to DOE (goals, principles) Complete randomized and block designs Using Minitab	Assignment 9 (surveys) Sample design Read: Angst (2007), FTC report
11-Nov-09 Week 11	<b>DOE:</b> Full factorial designs, Blocking and confounding	Assignment 10 (DOE) Read: Souza et al. (2006) Wiki: Integrate DOE into project
18-Nov-09 Week 12	<b>DOE:</b> Random effects, Nested designs, Confounding and Aliasing; Fractional factorial designs	Assignment 11 (DOE) Read: Bakos et al. (2006) Deploy survey!
25-Nov-09 Week 13	<b>DOE:</b> 3-level designs, Response surface methodology	Assignment 12 Response rate & data analysis Project: response analysis and data presentation
2-Dec-09 Week 14		
9-Dec-09 Week 15	Project Presentations	Project report due 11-Dec-09