INDIANA UNIVERSITY
SCHOOL OF INFORMATICS AND COMPUTING
Bloomington

Management, Access, and Use of Big and Complex Data

On-line Course part of IU Data Science Program

Syllabus and Course Roadmap

V 5.0 – 07 Nov 2015

Instructor:

Professor Beth Plale

plale@indiana.edu

https://www.linkedin.com/in/bethplale

812 855 4373

@bplale

Associate Instructor:

Zong Peng, PhD student
zongpeng@indiana.edu
pengzong1111 (skype)

Course materials:

Course web page: http://datamanagementcourse.soic.indiana.edu/

No book required. All reading materials are accessible and downloadable either from Canvas (if distribution is limited or for convenience sake) or the web.

Schedule:

This is a 3 credit hour course that follows Indiana University's academic calendar for Fall 2015:

Begins	Mon, Aug 24
Labor Day	Mon, Sept 7
Fall Break	Oct 9-11
Thanksgiving	Nov 22-29
Ends	Fri, Dec 18

I. Course Overview:

Data is abundant and its abundance offers potential for new discovery, and economic and social gain. But data can be difficult to use. It can be noisy and inadequately contextualized. There can be too big a gap from data to knowledge, or due to limits in technology or policy not easily combined with other data. This course will examine the underlying principles and technologies needed to capture data, clean it, contextualize it, store it, access it, and trust it for a repurposed use. Specifically the course will cover the 1) distributed systems and database concepts underlying noSQL and graph databases, 2) best practices in data pipelines, 3) foundational concepts in metadata and provenance plus examples, and 4) developing theory in data trust and its role in reuse.

Evaluation: Competency in the course is evaluated on a student's engagement with and mastery of the content. This is through:

a) Exercises: allow you to put into practice what you learned in a lesson and through the readings and resources. Exercises are always done as an assignment that is submitted via Canvas.

b) Project: there are two projects: Lessons 6 and 9. A project is a larger effort that may pull together concepts from multiple lessons. Project 9 has two tracks, one for non-technology savvy graduate and undergraduate students, and a longer and more technical track for the technology savvy graduate students

c) Reflections: allow you to test yourself on what you have learned in a lesson. Reflections are done through a separate lesson feedback video.

d) Peer review: engagement in peer reviewing of classmate materials, and

e) Class interactions: engagement in class interactions (chat sessions, hangouts, engaging with the video content).

The breakdown used in determining a grade is 30% exercises (a), 30% projects (b), and 40% engagement (c, d, e). Graduate students can skip doing one reflection. Undergraduates may skip one reflection and one exercise (or for exercises, the lowest grade will be dropped.)

II. Mechanics:

An overview of the structure of the course is given in the introductory video posted to the course web site: http://datamanagementcourse.soic.indiana.edu. Students should consult the video.

The on-line course covers a semester of work. A student is expected to follow the timeline given on the course web site. It is expected that a student put in 6-7 hours a week every week for the course which includes time spent in readings, exercises, and engaging with instructional content.

Section	Cohort description	Engagement mechanism
34879	Graduate residential	In person topical discussion section every Fridays 1-2:15 pm, room 122 Informatics East; Canvas (canvas.iu.edu) for daily interactions and chat

34881	Graduate on-line	On-line discussion Fridays 5-6:15 pm; Canvas (canvas.iu.edu) for daily interactions and chat
34880	Undergraduate residential	On-line discussion Fridays 11:00 am - 12:15 pm; Office hours - 2 hours/week, location TBA; Canvas (canvas.iu.edu) for daily interactions and chat

Tools:

a) *On-line discussions:* we will use either Google Hangout or Big Blue Button for weekly discussions; the latter comes with Canvas.

b) *Chat:* using Canvas (canvas.iu.edu); talk to fellow classmates and instructors using chat

c) *Course web site:* go to place to get to next lesson video, http://datamanagementcourse.soic.indiana.edu/

d) *Reflections and exercises:* posted to Canvas

e) *Assignment submission:* all will be submitted through Canvas

f) *Lessons:* lessons are available in Office Mix cloud (accessible off of the course web page); they were created with Office Mix (mix.office.com) a plugin to Powerpoint.

How to do a Peer Reviews

Peer review is an opportunity for a student to constructively offer feedback to another student. This enhances the learning experience for both the student receiving the feedback, and for the student providing the feedback. Feedback must always be constructive. The best feedback makes the submitter think more deeply about what they wrote; that is, it offers points for further thinking ("have you thought about this angle?") Feedback that only points out problems can be seen as unconstructive.

Students are asked to comment on no more than a couple points made in the submission, and make the feedback constructive by pointing out a counter argument, supporting an argument, taking a thought one step further, pointing out related literature, etc. Peer reviewing does not include assigning a grade. The peer reviewing process is evaluated on the quality of the peer review. The submitter of the assignment receives a grade for their original submission, the grade is independent of any peer review comments. The original submitter is encouraged to respond to the peer reviews through the Canvas system (not by resubmission of their assignment.)

III. Course Outline

The table below gives all 20 lessons in order as you will do them throughout the semester. The start and stop dates are given for each. Because there are 16 weeks and 20 lessons, some weeks you will do two lessons. Except for projects, the lesson feedback is due at the end of the week in which it is assigned.

The assignment completion dates are color coded to let you know what kind of assessment the lesson will use. Lessons in Yellow are Peer Reviewed; Green are Projects; Red use Reflections for feedback; Blue: have Exercises to turn in.

Lesson #	Lesson Title	Topic Area	Start by	Complete by
1	Big data in business	Intro	24 Aug	31 Aug midnight

2	Big data in scientific research	Intro	31 Aug	5 Sep midnight
3	Data processing pipelines in science; in 2 parts	Data pipelines	07 Sep	12 Sep midnight
4	Data processing pipelines in business	Data pipelines	07 Sep	12 Sep midnight
5	Data cleansing	Data pipelines	14 Sep	19 Sep midnight
6	Project: Data coding	Data pipelines	14 Sep	28 Sep midnight
7	Software systems design overview	Software systems	21 Sep	26 Sep midnight
8	Complexity in software systems	Software systems	21 Sep	26 Sep midnight
9	Project: Twitter dataset analysis and modeling	Project	28 Sep	17 Oct midnight
10	Distributed file system introduction	Distributed systems	05 Oct	10 Oct midnight
11	Role of caching in distributed computing	Distributed systems	12 Oct	17 Oct midnight
12	Role of fault tolerance in distributed computing	Distributed systems	19 Oct	24 Oct midnight
13	NoSQL data stores	Distributed data stores	26 Oct	31 Oct midnight
14	Consistency in distributed noSQL data stores	Distributed data stores	26 Oct	31 Oct midnight
15	Routing in noSQL data stores	Distributed data stores	07 Nov	14 Nov midnight
16	Comparison of data models through example	Distributed data stores	07 Nov	21 Nov midnight
17	Data provenance	Data provenance	07 Nov	28 Nov midnight
18	Linked Data	Linked data	07 Nov	05 Dec midnight
19	Science gateways and distributed computing: data in, data out	Open source open data	skip	skip
20	Social and Technical Barriers to Data Sharing	Data sharing	07 Nov	13 Dec midnight

IV. Course Roadmap

The Section IV portion of the syllabus is your roadmap for course completion. It is recommended that you keep a hardcopy of the roadmap and refer to it as you complete one lesson and move onto the next. Each lesson is described here in the syllabus with a purpose, and directions on what is asked of you. Each lesson has an accompanying reflection, exercise, or project. Reflections cover the content in the lesson video and any assigned readings. A reflection is carried out electronically through a separate video whose link is available on the course web site. Exercises cover the content in the video and any assigned readings. Exercises are submitted as a .pdf or .docx document through Canvas. Projects are submitted through Canvas in a format described in the project description. All links to lessons can be found on the course web page.

Staying on schedule: You should confirm that you are staying on schedule by consulting the start and due date for each lesson in the table above.

Peer review: Five lessons have peer review with them (lessons 1, 2, 3, 4, and 20). Peer review will be assigned at the end of the week upon which the assignment is due. Peer review will take place the week following the assignment due date. If you are unable to submit an assignment, you will be unable to participate in peer review and peer review is part of the engagement portion of your grade. Exceptions will be allowed on an as needed basis only.

0. Background Competency ----------

If you lack background in databases and file systems, you are advised to work through the following nicely done tutorial videos on MySQL and the Linux File System.

MySQL Database Tutorial - 1 - Introduction to Databases, Bucky Roberts, Jan, 2012.
https://www.youtube.com/watch?v=KgiCxe-ZW8o.

MySQL Database Tutorial - 2 - Getting a MySQL Server
https://www.youtube.com/watch?v=qgdKbmxR--w&index=2&list=PL32BC9C878BA7208 ()

MySQL Database Tutorial - 3 - Creating a Database
https://www.youtube.com/watch?v=O4SIpJMH7po&list=PL32BC9C878BA72085&index=3

MySQL Database Tutorial - 4 - SHOW and SELECT
https://www.youtube.com/watch?v=HQQ_hDCUUuI&index=4&list=PL32BC9C878BA72085

MySQL Database Tutorial - 5 - Basic Rules for SQL Statements
https://www.youtube.com/watch?v=evvg1h2ivDo&index=5&list=PL32BC9C878BA72085

MySQL Database Tutorial - 6 - Getting Multiple Columns
https://www.youtube.com/watch?v=TKbKAW0Fspc&index=6&list=PL32BC9C878BA72085

Learning the Linux File System, https://www.youtube.com/watch?v=HIXzJ3Rz9po, by Joe Collins, Jul 7, 2015

Lesson 1: Big Data in Business ----------
Key topic: business perspective on big data

Lessons 1 and 2 give you a perspective of how society is thinking and talking about big data today. We look the topic from the perspective of business (lesson 1) and science (lesson 2).

Watch the Lesson 1 video linked at the URL off the course web page. Read the two readings below. Respond to the assignment for the lesson, which you will find on canvas. Submit your completed assignment via Canvas. This lesson will utilize peer review.

Readings:

"A special report on managing information: Data, data everywhere," *The Economist*, February 25, 2010

Data Scientist, The Sexiest Job of the 21st Century, Thomas H. Davenport and D.J. Patil, Harvard Business Review, pp. 70-76, Oct 2012

Lesson 2: Big Data in Scientific Research ----------
Key topic: science perspective on big data

Lesson 2 gives a science perspective on big data. You will read a dozen or so short articles from Science magazine (Feb 2011). The articles are written by practitioners in a dozen or so fields including social science, medical, and scientific disciplines, each discussing their unique data

issues. Taken together, the collection highlights how differently one discipline sees data challenges from another discipline.

Watch the Lesson 2 video linked off the course web page and read the Lesson 2 readings. Canvas has a spreadsheet that helps guide you to the right articles to read. Respond to the assignment for the lesson, which you will find on canvas. Submit your completed assignment via Canvas. This lesson will utilize peer review

Readings:

"Dealing with Data", Special Online Collection, Science, 11 February 2011

In the 11 February 2011 issue, *Science* (Links to an external site.) joins with colleagues from *Science Signaling* (Links to an external site.), *Science Translational Medicine* (Links to an external site.), and *Science* Careers (Links to an external site.) to provide a broad look at the issues surrounding the increasingly huge influx of research data. This collection of articles highlights both the challenges posed by the data deluge and the opportunities that can be realized if we can better organize and access the data.
http://www.sciencemag.org/site/special/data/

Lesson 3: Data Processing Pipelines in Science ----------
Key topic: what is a data pipeline? How used in science?

What is a data pipeline? Data rarely instantly show up ready to use in whatever exploratory purpose you may have in mind. Data from creation to use undergoes numerous steps, some of which are end products in themselves. Lesson 3 is derived from a talk that database scholar Jim Gray gave in 2009 on the Laboratory Information Management System (LIMS), an early notion of the data processing pipeline for science.

Watch lesson 3 video (it is in two parts) at the link off the course web page. Read both readings and the three cited resources. Respond to the assignment for the lesson, which you will find on canvas. Submit your completed assignment via Canvas. This lesson will utilize peer review.

Readings:

Jim Gray on eScience: A Transformed Scientific Method, Edited by Tony Hey, Stewart Tansley, and Kirstin Tolle, in *The Fourth Paradigm: Data Intensive Scientific Discovery*, Tony Hey, Stewart Tansley, and Kritsin Tolle eds., Microsoft Research, 2009, pp. xix – xxxiii.

Jim Gray's Fourth Paradigm and the Construction of the Scientific Record, Clifford Lynch, in *The Fourth Paradigm: Data Intensive Scientific Discovery*, Tony Hey, Stewart Tansley, and Kritsin Tolle eds., Microsoft Research, 2009, pp. 177-183.

Understanding execution time complexity: the Selection Sort versus the Heap Sort
http://en.wikipedia.org/wiki/Selection_sort
http://nova.umuc.edu/~jarc/idsv/lesson3.html

Selecting the Right LIMS, Keith O'Leary, Scientific Computing, Aug 2008
http://www.scientificcomputing.com/articles/2008/08/selecting-right-lims

Lesson 4: Data Processing Pipelines; Role of Cloud Computing ----------

Key topic: how does business view a data pipeline? Role of cloud computing in the business view of data pipelines.

Lesson 4 introduces the business perspective of data pipelines. It draws inspiration from a 2011 talk by Wernert Vogels *"Data Without Limits".* Vogels is CTO of Amazon, and in this nice 2011 talk discusses data pipelines in context of business computing. He argues that cloud computing is core to a business model "without limits". The pipeline he proposes is: collect | store | organize | analyze | share.

Watch Lesson 4 video; there is no accompanying reading. The resources for this lesson point to a primer for MapReduce. If you are not familiar with MapReduce, you should look at the primer. Respond to the assignment for the lesson, which you will find on canvas. Submit your completed assignment via Canvas. This lesson will utilize peer review

Readings:

No readings for this lesson.

Resources:

Vogels talks about mapreduce extensively during his discussion of analysis. If you're not familiar with MapReduce, a decent primer on MapReduce (Hadoop really; MapReduce is built into the open source Hadoop tool) can be found here:

http://readwrite.com/2013/05/23/hadoop-what-it-is-and-how-it-works

Lesson 5: Data Cleansing ----------

Keywords: data cleaning, missing data, quantitative coding.

Lesson 5 gives use cases on data cleansing from environmental science and social science. It includes a link to a nice video “Dealing with Missing Data and Data Cleansing”, Part 3 of 3 on Quantitative Coding and Data Entry, Graham R Gibbs, Research Methods in Social Sciences, University of Huddersfield.

Watch lesson 5 video. There are no readings or resources for this lesson. Do the exercise; turn in your answers via Canvas.

Readings:

No readings for this lesson.

Lesson 6: Data Coding (project) ----------

Keywords: tagging data, categorizing data, coding data, feature extraction

In this lesson the student will gain basic knowledge about coding data, or the tagging or categorizing of it on important features of the data so that it can be clustered, etc.
The student will get a chance to try their hand at coding a dataset.

One possible project uses 278 media mentions from Pervasive Technology Institute over the year 2013-2014. The categorization that the student does will be illustrated through visualizing the results as a simple pie chart. Other possible projects will be introduced later. Watch the lesson video, then complete the project and turn it in via Canvas.

Lesson 7: Software Systems Design Overview ----------

Keywords: distributed systems, emergent behavior, tradeoffs in software system design

Lesson 7 is an introduction to the general concepts of software systems. These concepts are used during design of the large software systems needed to handle any of today's large applications (social media, cloud services, shopping carts, video rentals …).

Watch lesson 7 video. There are no readings for this lesson. Complete the video for the reflection for this lesson.

Readings:

No readings for this lesson.

Lesson 8: Complexity in Software Systems ----------

Keywords: complexity, layering, abstraction, modularity, hierarchy.

A key problem in large-scale applications is complexity. We will examine the sources of that complexity, and design principles that are used to overcome the complexity.

Watch lesson 8 video. There are no readings for this lesson. Complete the exercise for this lesson; the exercise can be found on Canvas.

Readings: No readings for this lesson.

Lesson 9: Project: Twitter Dataset Analysis and Modeling ----------

Lesson 9 is the course project. An explanation of the project is the subject of the lesson 9 video. You will have 3 weeks to complete the project. The project description is in Canvas, and will be turned in through Canvas. Expect to give a demo of the project.

Lesson 10: Distributed File System Intro ----------

Keywords: transparencies, session semantics, fault tolerance, naming

We looked in Lessons 7 and 8 at systems design and some of the complexities therein. The focus of this course is for the student to understand the distributed systems concepts that are inherent in today's noSQL stores. The next step leading up to a study of noSQL stores themselves is to look at distributed file systems, where key concepts like transparencies, session semantics, fault tolerance, and naming all have a form that can be easily understood from the perspective of files and directories, which we all work with.

Watch lesson 10 video. Read just sections 1-3 of the selected reading. Complete the video for the reflection for this lesson.

Readings:

Distributed File Systems: Concepts and Examples, E. Levy, A. Silberschatz, ACM Computing Surveys, Vol 22(4), Dec 1990, pp. 321-374

Lesson 11: Role of Caching in Distributed Computing ----------

Keywords: caching, locality of reference, cache replacement strategy, cache coherency

Caching of data is a key concept to efficient performance of a large distributed application. We turn back to the Levy and Silberschatz paper to see what they have to say about caching.

Watch lesson 11. Read just section 4 of the selected reading, skipping sections 4.2 and 4.3. Complete the exercise for this lesson; the exercise can be found on Canvas.

Readings:

Distributed File Systems: Concepts and Examples, E. Levy, A. Silberschatz, ACM Computing Surveys, Vol 22(4), Dec 1990, pp. 321-374. Section 4 (skip 4.2 and 4.3)

Lesson 12: Role of Fault Tolerance in Distributed Computing ----------

Keywords: stateful and stateless servers, idempotence, transactions

When distributed systems span multiple locations or computers, the incidence of failure increases substantially. In this lesson we'll turn back to the Levy and Silberschatz paper to learn about fault tolerance.

Watch lesson 12. Read just section 5.1 of the selected reading. Complete the exercise for this lesson; the exercise can be found on Canvas.

Readings:

Distributed File Systems: Concepts and Examples, E. Levy, A. Silberschatz, ACM Computing Surveys, Vol 22(4), Dec 1990, pp. 321-374

Lesson 13: NoSQL Data Stores ----------
Keywords: document, key-value, graph, and column stores; ACID, Consistency, replication

This lesson is the introduction to large-scale data stores. The lesson video gives an overview of noSQL data stores using a 2012 talk by Martin Fowler at the GoTo Aarhus Conference 2012. The reading covers the concept of data services. As the reading states, "while data services were initially conceived to solve problems in the enterprise world, the cloud is now making data services accessible to a much broader range of consumers."

Watch lesson 13 video and read the readings. If you are familiar with MongoDB data model and CRUD operations, skip the second and third references in the readings list. Complete the video for the reflection for this lesson.

Readings:

Thinking in Documents, Part 1, https://www.mongodb.com/blog/post/thinking-documents-part-1?jmp=docs&_ga=1.180119460.1286782280.1445028759. Blog post Apr 2015.

MongoDB CRUD Operations, http://docs.mongodb.org/manual/core/crud-introduction/.

MongoDB Data Models, http://docs.mongodb.org/manual/core/data-modeling-introduction/

Data Services, Michael J. Carey, Nicola Onose, Michalis Petropoulos, Communications of the ACM, Vol 55, no. 6, pp. 86-97, Jun 2012 DOI:10.1145/2184319.2184340

Lesson 14: Consistency in Distributed NoSQL Data Stores ----------
Keywords: eventual consistency, CAP Theorem, Quorum protocol, Vector clocks

When storage in a data store is distributed across multiple storage devices, consistency of reads and writes becomes a paramount issue.

Watch lesson 14 video. Read the nice introductory reading by W. Vogels from 2009. Complete the exercise for this lesson; the exercise can be found on Canvas.

Reading: *Eventually Consistent*, W. Vogels, Communications of the ACM, 52, 1, Jan 2009

Lesson 15: Routing in NoSQL Data Stores ----------
Keywords: routing, distributed hash tables, Chord, peer-to-peer, local versus global knowledge

When data are stored across multiple computers in a single noSQL data store, and the data store can be accessed any of the multiple servers that support the noSQL data store, how does the noSQL data store ensure that a request for a data object is routed to the right location where the data are stored? This is a routing problem in noSQL data stores. This lesson discusses ways of keeping track of the information needed to route requests to the right server.

Watch lesson 15. There are no readings. Complete the exercise for this lesson; the exercise can be found on Canvas.

Readings: No readings for this lesson.

Lesson 16: Comparison of data models through example ----------
Keywords: schema, data model

While efficient locating and retrieval of data from a distributed noSQL data stores is important, equally important is the flexibility that noSQL stores give you in storing data objects. Relational databases give structured and normalized tables for rapid and precise querying. noSQL stores support less structured data and less precise querying. This lesson walks you through a real life comparison using ecological data and storing it in a relational, document store, key-value pair, and column store data model.

You may want to review Lesson 13 again to refresh your memory on the basic data models used in key-value, document, and column noSQL stores. Watch lesson 16 and complete the video for the reflection for this lesson.

Readings: No readings for this lesson

Lesson 17: Data Provenance ----------

Keywords: data provenance, causality graph, Open Provenance Model

As data sharing increasingly moves from a friendly exchange between two parties that know each other to a transaction on an open data sharing market, the need also grows for data to carry with it sufficient information that the recipient can use to establish whether or not they trust and can use the data. Data provenance lies at the heart of the descriptive data needed to discern data trustworthiness. This lesson will introduce data provenance and give you a sense of what provenance data is and how it is represented.

Watch lesson 17 and read the survey paper. Complete the exercise for this lesson; the exercise can be found on Canvas.

Reading: *A Survey of Data Provenance in e-Science*, Yogesh Simmhan, Beth Plale, Dennis Gannon, Association of Computing Machinery SIGMOD Record, vol 34, no. 3, pp. 31-36, Sep 2005

Lesson 18: Linked Data ----------

Keywords: linked data, JSON-LD, RDFa, semantic architecture

The Internet contains vast numbers of pages linked together. Linked data enables more detailed information to be linked together in a loose way that doesn't require standardization on a small set of names (a vocabulary). Linked data is key to data sharing because it allows concepts about data objects to be linked together loosely (again, without seeking complete agreement.) This lesson introduces linked data through a video by Manu Sporny "Intro to Linked Data" 2012, and discusses some of the issues around linked data through the readings.

Watch lesson 18 video and read the readings. Complete the video for the reflection for this lesson.

Readings:

The semantic web. Berners-Lee, T., Hendler J., & Lassila, O. Scientific American, May 2001.

Semantic web in action, Feigenbaum, L., Herman, I., Hongsermeier, T., Neumann, E., & Stephens, S., Scientific American, Dec 2007.

Lesson 19: Science Gateways, Scientific Workflows and Distributed Computing: Data In, Data Out ----------

Keywords: supporting large-scale science, open source, open governance

Lesson 19 steps back from the details and gives a broad discussion of data cyberinfrastructure in science: how it is structure in the United States and how it is sustained. Discusses science gateways, their use in several scientific disciplines, and the open source and open governance philosophy of one project in support of science.

Watch lesson 19 video. Complete the video for the reflection for this lesson.

Readings: No readings for this lesson.

Lesson 20: Social and Technial Barriers to Data Sharing ----------
Keywords: data trustworthiness, data sharing, economies of data sharing

This last lesson ends the semester with several short readings on the social and economic barriers to sharing data.

Watch the video for lesson 20. Read the readings. Respond to the assignment for the lesson, which you will find on canvas. Submit your completed assignment via Canvas. This lesson will utilize peer review.

Reading:

Who Will Pay for Public Access to Research Data? Fran Berman and Vint Cerf, Science 9, August 2013, Vol. 341 no. 6146 pp. 616-617 DOI: 10.1126/science.1241625 (www.sciencemag.org). Preprint version available at: http://www.cs.rpi.edu/~bermaf/Berman%20+%20Cerf%20Public%20Access%20--%20Author%20Version.pdf

If Data Sharing is the Answer, What is the Question? Christine L. Borgman, http://ercim-news.ercim.eu/en100/special/if-data-sharing-is-the-answer-what-is-the-question

Data sharing: empty archives, Bryn Nelson, Nature 461, published online 9 Sep 2009, pp. 160-163. Doi: 10.1038/461160a http://www.nature.com/news/2009/090909/full/461160a.html

V. Final remarks

Students are expected to conduct themselves in a manner befitting their status as a student of a respected and distinguished institution of higher education. In college courses, we are continually engaged with other people's ideas: we read them in texts, hear them in lecture, discuss them in class, and incorporate them into our own writing. As a result, it is very important that we give credit where it is due. Plagiarism is using others' ideas and words without clearly acknowledging the source of that information.

See http://www.indiana.edu/~wts/pamphlets/plagiarism.shtml for help in addressing plagiarism in your own work.

Articles and data posted to Canvas for this course are for the convenience of students only and should not be shared outside the course.