CRM 318: Introduction to Data & Information Management in Health Science

Instructors: Haresh Bhatia, PhD & David Mulvihill

Contact: vrbhatias@gmail.com or dmulvihill@wustl.edu

Class Time: Wednesdays, 5:30-8:00 PM (CRM 318)
Location: FLTC Room 203

Dates: January 17 – May 2, 2018

Requirements:
1. Students are expected to use their own laptops during the class sessions, for their practical work, and assignments.
2. Working knowledge of MS-Excel.

Description
This introductory course will cover education and training in data and information management, including: what data are, best practices in data collection, data preparation, and working with data management plans with respect to healthcare application. Students will be given a dataset for the class assignments in order to clean and visualize data, and they will be required to find, critique, and present health data as found in the relevant media (e.g. scientific literature and research data repositories). The lectures and activities are designed for those who work or will work with healthcare data but may have had no formal training in managing data and information.

Goal
The goal of the course is to instruct and train researchers in best practices in data collection, maintenance, and manipulation as well as how to make data discoverable and shareable.
Class Structure
This course is structured as an undergraduate course. Students are expected to attend and participate in class, thus learning from each other by utilizing the breadth and expertise of the group. The first portion of the class will be a lecture of pertinent topics with the last portion of the class being a training session to learn software tools and techniques in data management. Assignments will be available through WUSTL Blackboard (bb.wustl.edu).

Attendance
It is expected that students will attend and actively participate in class. In-class exams and projects will be given according to the schedule on the syllabus. In case of inclement weather, the course will adhere to Washington University policy and any associated scheduled changes. Any weather associated modifications other than those set by WUSM will be sent to the class by e-mail.

Academic Integrity
From the UC policy on academic integrity (see website below): “In all academic work, the ideas and contributions of others must be appropriately acknowledged, and work that is presented as original must be, in fact, original.” It is expected that students will submit their own original work for all course assignments and projects, and properly cite others’ work/ideas when appropriate. For more information on UC policy, please go to: http://www.wustl.edu/policies/undergraduate-academic-integrity.html

Disability Resources
Students can seek assistance and class accommodations for learning and other disabilities through WU’s Cornerstone: The Center for Advanced Learning. Please go to http://cornerstone.wustl.edu/Home.aspx for further information.

Text
1. Biomedical Informatics – Computer Applications in Healthcare and Biomedicine, By Edward H. Shortliffe and James J. Cimino
   Only some parts of this text book are suggested for reading – as indicated in the class-session details below.
2. Research papers as indicated in the class-session details below.
Readings (done by class, before the respective session)

Class Sessions

Week 1 – 17th Jan [DM]
1. Introductions
2. Module 1: Overview of Research Data Management (RDM)
   a. Chapter 26.1, 26.2 from the Biomedical Informatics text

Week 2 – 24th Jan [DM & HB]
1. Module 1: Research Data Management (RDM)
   a. Chapter 26.3 from the Biomedical Informatics text
2. Module 2: Types, Formats, and Stages

Week 3 – 31st Jan [DM & HB]
1. Intro to Clinical Data Management (CDM): Hands-on work with REDCap, OnCore, OpenSpecimen.
   a. Chapter 2 from the Biomedical Informatics text
2. Quiz 1 – Module 1

Week 4 – 7th Feb [HB & DM]
1. Merging Data and Data Cleaning. [HB]
2. Data Cleaning: Hands-on work [HB]

Week 5 – 14th Feb [HB & DM]
1. Module 3: Metadata

Week 6 – 21st Feb [DM & HB]
1. Importing & Merging Data: Hands-on work [DM]
2. Importing data into RDM / CDM tool (REDCap / OnCore) [DM]

Week 7 – 28th Feb [DM & HB]
1. Module 5: Legal and Ethical Considerations
2. Module 6: Sharing and Reuse
   a. Chapter 26.3.3.3.2 and 26.3.3.3.3 from the Biomedical Informatics text

Week 8 – 7th Mar [DM & HB]
   1. Module 7: Repositories, Archiving, & Preservation
   2. Data visualization

Week 9 – 14th Mar [DM & HB]
   Spring Break - No Class

Week 10 – 21st Mar [HB]
   1. In-class DMP Review
   2. HANDS-ON Data visualization & cleaning

Week 11 – 28th Mar [DM & HB]
   1. Data cleaning continued

Week 14 – 4th Apr [HB]
   1. Statistics

Week 13 – 11th Apr [HB]
   1. Concepts in epidemiology and biostatistics part I
   2. HANDS-ON – Crosstabs

Week 14 – 18th Apr [HB & DM]
   1. Concepts in epidemiology and biostatistics part II
2. HANDS-ON – Crosstabs & Analyses

**Week 15 – 25th Apr [DM & HB]**
1. Project preparation

**Week 16 – 2nd May [DM & HB]**
1. Project presentation

**Grading**
- A 90-100%
- B 80-89%
- C 70-79%
- D 60-69%
- F <60%

**Points per Assignment**

<table>
<thead>
<tr>
<th>Points</th>
<th>Assignment</th>
</tr>
</thead>
<tbody>
<tr>
<td>20</td>
<td>Quiz 1 – Module 1</td>
</tr>
<tr>
<td>30</td>
<td>Data Cleaning Assignment</td>
</tr>
<tr>
<td>20</td>
<td>Quiz 2 – Modules 2 &amp; 3</td>
</tr>
<tr>
<td>30</td>
<td>Quiz 3 – Modules 4-6</td>
</tr>
<tr>
<td>30</td>
<td>Data Management Plan</td>
</tr>
<tr>
<td>50</td>
<td>Project</td>
</tr>
<tr>
<td>20</td>
<td>Class Participation</td>
</tr>
</tbody>
</table>

200 Total Points Possible
<table>
<thead>
<tr>
<th>Week</th>
<th>Date</th>
<th>Lecture</th>
<th>Notes</th>
<th>Due Dates</th>
<th>Points</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>17 Jan</td>
<td>Introductions Module 1: Overview of Research Data Management</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>24 Jan</td>
<td>Module 1: Overview of Research Data Management Module 2: Types Formats and Stages</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>31 Jan</td>
<td>HANDS-ON Intro to Clinical Data Management</td>
<td>REDCap, OnCore, OpenSpecimen</td>
<td>Quiz 1 – Module 1</td>
<td>20</td>
</tr>
<tr>
<td>4</td>
<td>7 Feb</td>
<td>HANDS-ON Importing &amp; Merging data; Data cleaning</td>
<td>Institutional Resources</td>
<td></td>
<td>30</td>
</tr>
<tr>
<td>5</td>
<td>14 Feb</td>
<td>Module 3: Metadata Module 4: Data Storage, Backup and Security</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>21 Feb</td>
<td>HANDS-ON Data conversion; reformatting data; syntax</td>
<td></td>
<td>Quiz 2 – Modules 2 &amp; 3</td>
<td>20</td>
</tr>
<tr>
<td>7</td>
<td>28 Feb</td>
<td>Module 5: Legal and Ethical Considerations Module 6: Sharing and Reuse</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>7 Mar</td>
<td>Module 7: Repositories, Archiving, &amp; Preservation Data visualization</td>
<td></td>
<td>Quiz 3 – Modules 4, 5 &amp; 6</td>
<td>30</td>
</tr>
<tr>
<td>9</td>
<td>14 Mar</td>
<td>Spring Break – No class</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>21 Mar</td>
<td>In-class DMP Review HANDS-ON Data visualization &amp; cleaning</td>
<td></td>
<td>Data Management plans</td>
<td>30</td>
</tr>
<tr>
<td>11</td>
<td>28 Mar</td>
<td>Data cleaning continued</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>4 Apr</td>
<td>Concepts in epidemiology and biostatistics part I HANDS-ON – Crosstabs</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>13</td>
<td>11 Apr</td>
<td>Concepts in epidemiology and biostatistics part II HANDS-ON – Crosstabs &amp; Analyses</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Date</td>
<td>Activity</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>---</td>
<td>--------</td>
<td>---------------------</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>14</td>
<td>18 Apr</td>
<td>Project preparation</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>15</td>
<td>25 Apr</td>
<td>Project preparation</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>16</td>
<td>2 May</td>
<td>Project presentation</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>