Contents

Organization, Information, and Learning Sciences (OILS) 515 Goals, Objectives and Reading Assignments **3**

Week 1 - Introductions, Course Overview and Introduction to the Research and Data Lifecycles . . 4

Reading . . . 4

Week 2 - Introduction to Vector Data . . . 4

Background . . . 4

Expected Outcomes . . . 4

Key Concepts . . . 4

Readings . . . 5

Assignment . . . 5

Class Project . . . 5

Week 3 - Introduction to Raster Data and Other Related Geospatial Data Concepts . . . 5

Background . . . 5

Expected Outcomes . . . 5

Key Concepts . . . 6

Reading . . . 6

Assignment . . . 6

Class Project . . . 6

Data Links . . . 6

Week 4 - Database Design I . . . 6

Background . . . 6

Expected Outcomes and Key Concepts . . . 7

Reading . . . 7

Assignment . . . 7

Class Project . . . 8

Week 5 - Database Design II . . . 8

Background . . . 8

Expected Outcomes and Key Concepts . . . 8

Reading . . . 8

Assignment . . . 8

Week 6 - Geodatabase Design . . . 9

Background . . . 9

Expected Outcomes and Key Concepts . . . 9

Reading . . . 9

Assignment . . . 9

Week 7 - Managing Raster Data . . . 9
Background . . . 9
Expected Outcomes and Key Concepts . . . 9
Reading . . . 10
Assignment . . . 10
Week 8 - Data Formats for Analysis and Archiving . . . 10
Background . . . 10
Expected Outcomes and Key Concepts . . . 10
Reading . . . 10
Assignment . . . 10
Class Project . . . 10
Week 9 - Spring Break . . . 10
Week 10 - Documenting Data - The Interview . . . 10
Background . . . 10
Expected Outcomes and Key Concepts . . . 10
Reading . . . 10
Class Project . . . 11
Week 11 - XML Document Creation, editing and validation . . . 11
Background . . . 11
Expected Outcomes and Key Concepts . . . 11
Reading . . . 11
Week 12 - Metadata Standards - FGDC . . . 11
Background . . . 11
Expected Outcomes and Key Concepts . . . 11
Reading . . . 11
Class Project . . . 11
Week 13 - Metadata Standards - ISO and Dublin Core . . . 11
Background . . . 11
Expected Outcomes and Key Concepts . . . 11
Reading . . . 12
Assignment . . . 12
Week 14 - Data Management Planning . . . 12
Background . . . 12
Expected Outcomes and Key Concepts . . . 12
Reading . . . 12
Week 15 - Ethical, legal and privacy issues . . . 12

Background . . . 12
Expected Outcomes and Key Concepts . . . 12
Reading . . . 12
Assignment . . . 12
Class Project . . . 12
Week 16 - Emerging Concepts . . . 13
Background . . . 13
Expected Outcomes and Key Concepts . . . 13
Reading . . . 13
Class Project . . . 13

Organization, Information, and Learning Sciences (OILS) 515 Goals, Objectives and Reading Assignments

An understanding of core spatial data concepts and principles is increasingly important in the current world of collaborative, spatially enabled research and applications. We are no longer working in a vacuum as individual researchers that only need to understand and use the data that we create and use in our separate research projects. Successful research depends upon being able to integrate data generated by others with our own and by extension being able to share our data with others, both during our research projects and also for posterity (and to meet the requirements of funding agencies). This class will focus on the following aspects of spatial data management that relate to this need for effective integration, use, collaboration and sharing:

- The *Research* and *Data Lifecycles*
- Types of spatial data
- Spatial database design and management
- Working with and managing gridded data
- Spatial data documentation standards and practices
- Ethical, legal and privacy issues as they relate to spatial data
- Data management planning
- Emerging topics

Upon completion of the course students will have improved their knowledge and skills in the following areas:

- Locating and evaluating spatial data based upon knowledge of formats, content models and documentation standards
- Structuring data (both in terms of format selection and content) from a variety of sources to enable integrated research
- Evaluate data products to determine which elements of a dataset might raise ethical, legal or privacy issues if released or shared with others
- Documenting data as an ongoing process throughout the research cycle
- Producing machine- and human-readable documentation for data to support discovery, understanding, and use of data that they produce

Week 1 - Introductions, Course Overview and Introduction to the Research and Data Lifecycles

This week's required "collaboratory" session (Tuesday afternoon, 5:00-6:30 - method TBA) will allow us to share some background about each other, review the class structure and objectives and have a brief overview of the research and data lifecycles and the linkages beween them. While this week's lecture will be done "live" during our collaboratory session, future lectures will be pre-recorded and accessible at the beginning of each class week.

Reading

Piwowar, Heather A., Roger S. Day, and Douglas B. Fridsma. 2007 "Sharing Detailed Research Data Is Associated with Increased Citation Rate." Plos One 2, no. 3: Science Citation Index, EBSCOhost (accessed January 20, 2014). link

Tenopir, Carol, Suzie Allard, Kimberly Douglass, Arsev Umur Aydinoglu, Lei Wu, Eleanor Read, Maribeth Manoff, and Mike Frame. 2011. "Data Sharing by Scientists: Practices and Perceptions." *Plos ON*E 6, no. 6: 1-21. Academic Search Complete, EBSCOhost (accessed January 19, 2014). link

Week 2 - Introduction to Vector Data

Background

This week we begin our consideration of the different classes of geospatial data that you are likely to encounter - both as you search for data to use in your research and as you produce data with a spatial component that you need to integrate into a spatial data management system or workflow. Our focus is on *vector* datasets this week, with a particular emphasis on the general types of vector data *features* you need to be aware of, and the key linkage between vector *geometries* and the *attributes* that are linked to those features. While the introductory lecture focuses on these aspects of vector datasets, the concepts of *map projections* or *coordinate reference systems* as introduced in the reading are also core concepts that you will need to understand in the context of all spatial data that you work with. Next week's lecture will include a more detailed overview of map projections (and focus on raster data), but for now see what you can also pick up from the reading as it blends both vector and raster data and map projections into a single chapter.

Expected Outcomes

By the end of this week's reading and work you should have an understanding of:

- The types of geometries that are likely to be encountered when working with vector data
- The concept of attributes and attribute types as they are linked to features
- The beginning of an understanding of how you can transform tabular data that may have implicit spatial content into explicitly spatial data
- The importance of map projections in clearly understanding the spatial context for coordinates and geometries based upon those coordinates.

Key Concepts

- Vector Feature Types: *point*, *line*, *polygon*
- Attributes and associated attribute data types
- Map projections

Readings

Westra, E. (2010). Python Geospatial Development : Build a Complete and Sophisticated Mapping Application From Scratch Using Python Tools for GIS Development. Olton, Birmingham: Packt Publishing. (link for access page for downloadable eBook - expires) (link for access page for online eBook). *Chapter 2* and skim *Chapter 1*.

Nikos Mamoulis (2012), Spatial Data Management. Synthesis Lectures on Data Management #21. Morgan & Claypool Publishers. DOI10.2200/S00394ED1V01Y201111DTM021. http://libproxy.unm.edu/login?url=http://search.ebscohost.com/login.aspx?direct=true&db=cat00503a&AN=unm.b7199537&site=eds-live&scope=site. Review the *Introduction*, keeping in mind that some of the database related issues will come up in a few weeks - so, don't be worried if some of the database terminology is confusing during this initial read through.

Galati, S. R. (2006). Geographic Information Systems Demystified. Boston: Artech House. Link for access to the online and downloadable eBook. Skim *Chapters 1 & 3*

Assignment

Link to this week's assignment

Class Project

Link to this week's task related to the class project

Week 3 - Introduction to Raster Data and Other Related Geospatial Data Concepts

Background

This week we will conclude our overview of geospatial data types and cover some additional core geospatial data concepts that you need to be familiar with as you work with geospatial data in your own research. Beginning with a discussion of *raster* data types, an introduction to the *geodatabase* concept (to be much expanded upon in the coming weeks) our overview of different classes of geospatial data is now complete. This week's lecture also includes a discussion of the key role of *coordinate reference systems* (AKA map projections) in understanding the map coordinates represented in geospatial data, the importance of the consideration of *accuracy and precision* in geospatial (and other) data, and the relationship between accuracy and precision and the concept of map scale.

Expected Outcomes

After reviewing this week's lecture and completing the reading, you should understand the following concepts:

- Raster, vector and tabular data and their respective characteristics in spatial data management
- The importance of coordinate reference systems, and the specific purposes for which they are defined, in characterizing locations encoded in spatial data
- The concepts of accuracy and precision in location data and some of the factors that can contribute to decreased accuracy (bias) in location data.

Key Concepts

- Spatial data types
- Map projects and the tradeoffs in defining them
- Accuracy and precision in location data

Reading

Review the readings from last week as they relate to the raster data and other data concepts that will be covered in this week's lecture and demonstration.

NationalAtlas.gov. *Map Projections: From Spherical Earth to Flat Map*. Online resource: http://nationalatlas.gov/articles/mapping/a_projections.html. Accessed on 2/2/2014.

Quantum GIS (QGIS) Documentation:

- A gentle introduction to GIS
- User's Guide (skim for reference and familiarity with content and organization)

Assignment

Continue your work on the literature review that you began last week.

Class Project

Continue your work on defining your data management focus for the term

Data Links

Here are the links to the datasets that were used in this week's QGIS demonstration:

- http://kkb-classes.s3.amazonaws.com/2014/OILS515/data/bernalillo_tm2011_img.zip
- http://kkb-classes.s3.amazonaws.com/2014/OILS515/data/gnis_nm_all09.zip
- http://kkb-classes.s3.amazonaws.com/2014/OILS515/data/hyp2shp_shp.zip
- http://kkb-classes.s3.amazonaws.com/2014/OILS515/data/NM_Features_20130811.zip
- http://kkb-classes.s3.amazonaws.com/2014/OILS515/data/nm60mdem.zip
- http://kkb-classes.s3.amazonaws.com/2014/OILS515/data/tl_2010_35_state10_shp.zip
- http://kkb-classes.s3.amazonaws.com/2014/OILS515/data/tl_2010_35_zcta510_shp.zip

Week 4 - Database Design I

Background

This week we begin our consideration of database design, beginning with the basic terminology, concepts and types of databases that have been developed to date. We will have a short in-class presentation during this week's collaboratory session relating to the considerations you want to keep in mind when starting the process of *modeling* your data to meet a specific analytic and management goal.

Expected Outcomes and Key Concepts

After completing this week's reading and participating in the collaboratory you should understand the following concepts:

- What the *relational database model* is
- What *Structured Query Language (SQL)* is
- What the *Objectives* and *Benefits* of good database design are
- What the distinction is between *data* and *information*
- What *null* values are and why consideration of nulls is important in your database design
- What *tables*,*rows*, *fields* and*views* are when designing a database
- What *Primary* and *Foreign* keys are in a database
- What *one-to_one*, *one-to-many* and *many-to-many* relationships are within a database
- What *data integrity* is and why you must design your database to maintain it

Reading

This week's readings in *Database Design for Mere Mortals* cover the basic concepts of database terminology, leading into the design process as you consider the "model" for your data. Both editions are on 1-day reserve at the Centennial Science and Engineering Library.

Michael J. Hernandez (2003). Database Design for Mere Mortals: a Hands-on Guide to Relational Database Design. 2nd ed. Addison-Wesley. *Chapters 1-3*

or

Michael J. Hernandez (2013). Database Design for Mere Mortals: a Hands-on Guide to Relational Database Design. 3rd ed. Addison-Wesley. *Chapters 1-3*

As some technical background for the database platform that underlies the geodatabase (SpatiaLite) we will be working with as part of this class, skimming the following materials will be helpful.

Grant Allen and Mike Owens (2010). *The Definitive Guide to SQLite*. 2nd ed. Apress. http://www.books24x7.com.libproxy.unm.edu/marc.asp?bookid=37960 (you will need to create an account with Books 24x7 to access this digital resource). *Chapters 1-3* - skim 2, the installation process of SQLite is taken care of through the SpatiaLite installation process you perform.

and, as a beginning of our Python thread for the class, I recommend the following two iPython notebooks (an emerging technology that enables the development, execution and sharing of Python code through a web interface - http://ipython.org/notebook.html - both run on your local computer, but also potentially hosted on other systems, such as Wakari) developed by J.R. Johansson (robert@riken.jp) http://dml.riken.jp/~rob/:

Lecture 0 - Scientific Computing with Python

Lecture 1 - Introduction to Python Programming

Assignment

Please be prepared to give an ~10-minute presentation to the class over our Google Hangout during the Collaboratory Session this week on Tuesday from 5:00-6:30 pm. The presentation should address the requested information in the assignment for each of the items you addressed in your literature review:

Complete Citation A complete citation that would allow another researcher to locate the publication that you used. If available, provide the DOI (Digital Object Identifier) and online link for the publication.

Summary A brief summary of the publication's focus and conclusions.

Relevance A discussion of the relevance of the publication's findings to your interest in spatial data management in your research area/project.

Please post your writeup as a new "Thread" in the class's "Literature Review Results" discussion forum in Learn.

Class Project

Please continue your work in defining your data management focus for the term. Please schedule a time for us to "meet" to discuss your thoughts on this before 2/19 so we can make sure you are on track to be able to begin defining the specific datasets that you will be producing and documenting as part of your term project.

Week 5 - Database Design II

Background

This week you will be digging deeper into the concepts that you began working with last week in the context of relational database design.

Expected Outcomes and Key Concepts

Similar to last week, just in more detail as the readings for this week go into much more depth relating to these concepts.

Reading

Michael J. Hernandez (2003). Database Design for Mere Mortals: a Hands-on Guide to Relational Database Design. 2nd ed. Addison-Wesley. *Chapters 4, 5 (skim as you are most likely both the interviewer and interviewee in the context data that you are creating yourself), 6 & 7*

or

Michael J. Hernandez (2013). Database Design for Mere Mortals: a Hands-on Guide to Relational Database Design. 3rd ed. Addison-Wesley. *Chapters 4, 5 (skim as you are most likely both the interviewer and interviewee in the context data that you are creating yourself), 6 & 7*

Continuing the technical background for creating and adding data to SQLite (the database upon which SpatiaLite is based)

Grant Allen and Mike Owens (2010). *The Definitive Guide to SQLite.* 2nd ed. Apress. http://www.books24x7.com.libproxy.unm.edu/marc.asp?bookid=37960 (you will need to create an account with Books 24x7 to access this digital resource). *Chapter 4*

Also, related to data resources,

Westra, E. (2010). Python Geospatial Development : Build a Complete and Sophisticated Mapping Application From Scratch Using Python Tools for GIS Development. Olton, Birmingham: Packt Publishing. (link for access page for downloadable eBook - expires) (link for access page for online eBook). *Chapter 4.*

Assignment

Start working on the assignment to locate and describe data and review for documentation, usability and understanding.

Week 6 - Geodatabase Design

Background

This week we introduce the concepts related to the integration of geospatial data into relational database systems. This integration includes both the introduction of new data types into the database (i.e. the `geometry` data type), specialized indices that are optimized for working with those data, and additional functions that extend the SQL foundation of these databases to support additional query types that are explicitly spatial.

Expected Outcomes and Key Concepts

At the end of this section of the course, you should understand the following:

- The model for expanding standard relational databases to include support for geometries and their associated attributes
- The range of specialized spatial queries that can be used within a geospatially enabled database

Reading

For a more "cookbook" style presentation of the processes of working with SpatiaLite, check out:

The SpatiaLite Cookbook

For an overview of geospatially enabled databases and a description of the capabilities of three geodatabases, including SpatiaLite.

Westra, E. (2010). Python Geospatial Development : Build a Complete and Sophisticated Mapping Application From Scratch Using Python Tools for GIS Development. Olton, Birmingham: Packt Publishing. (link for access page for downloadable eBook - expires) (link for access page for online eBook). *Chapter 6.*

Finally, for a higher-level overview of some issues related to the development and performance of geodatabases:

Nikos Mamoulis (2012), Spatial Data Management. Synthesis Lectures on Data Management #21. Morgan & Claypool Publishers. DOI10.2200/S00394ED1V01Y201111DTM021. http://libproxy.unm.edu/login?url=http://search.ebscohost.com/login.aspx?direct=true&db=cat00503a&AN=unm.b7199537&site=eds-live&scope=site. Review the *Chapter 2.*

Assignment

Continue working on data review assignment

Week 7 - Managing Raster Data

Background

TBA

Expected Outcomes and Key Concepts

TBA

Reading

TBA

Assignment

Continue working on data review assignment

Week 8 - Data Formats for Analysis and Archiving

Background

TBA

Expected Outcomes and Key Concepts

TBA

Reading

TBA

Assignment

Present results of data review assignment in required collaboratory session

Class Project

Enumerate the specific datasets that you will create and document as part of your class project

Week 9 - Spring Break

Week 10 - Documenting Data - The Interview

Background

TBA

Expected Outcomes and Key Concepts

TBA

Reading

TBA

Class Project

Begin creating class project datasets

Week 11 - XML Document Creation, editing and validation

Background

TBA

Expected Outcomes and Key Concepts

TBA

Reading

TBA

Week 12 - Metadata Standards - FGDC

Background

TBA

Expected Outcomes and Key Concepts

TBA

Reading

TBA

Class Project

Begin creating documentation for class project datasets

Week 13 - Metadata Standards - ISO and Dublin Core

Background

TBA

Expected Outcomes and Key Concepts

TBA

Reading

TBA

Assignment

Create a data management plan

Week 14 - Data Management Planning

Background

TBA

Expected Outcomes and Key Concepts

TBA

Reading

TBA

Week 15 - Ethical, legal and privacy issues

Background

TBA

Expected Outcomes and Key Concepts

TBA

Reading

TBA

Assignment

Data management plan and Class project dataset and documentation peer review

Class Project

Share class project datasets and documentation with peers for review

Week 16 - Emerging Concepts

Background

TBA

Expected Outcomes and Key Concepts

TBA

Reading

TBA

Class Project

Present class project results and peer review outcome