Teaching Users to Work With Research Data:

Case Studies in Architecture, History and Social Work

by Aaron Addison[1] and Jennifer Moore[2]

Abstract
A tailored approach is ideal for teaching users to work with research data, which often varies significantly by domain and project depending on methodology, available data sources and intended outcomes. In this paper and presentation, three distinct contexts will be put forth, each using Geographic Information Systems (GIS) and focused problem-based learning (PBL) approaches to teach research data use: primary collection, digital data reuse and mined textual data. In each illustration, researchers are not only working to implement a functional methodology, but also to engage students in practices that equip them with theory, tools and skills to advance their own research trajectory. Further, these examples are from researchers in distinctly different disciplines: an architect working on climate change in the St. Louis region, three historians reconstructing history with data from texts and a professor of social work collecting data for villages in India. The Data and GIS Services (DGS) team at Washington University in St. Louis (WUSTL) has partnered with each project presented to support analyses, visualization, management, preservation and sharing of research data. Methods, challenges and opportunities are discussed..

Keywords
research data, problem-based learning, GIS, higher learning

Introduction

What is research data?

The definition of research data is nebulous because what data means to each knowledge domain differs. The federal government defines research data as 'the recorded factual material commonly accepted in the scientific community as necessary to validate research findings.' (White House, 2015). However, expanding the definition to include scholarly communities is more appropriate for an academic environment. It is widely agreed that data can be collected in a number of ways including by observation, data mining, modeling and from referential sources. While the form and function of data varies between disciplines, as research grows more collaborative and cross-disciplinary, interest in new and combined approaches to data and technology grow in tandem. Quantitative skills are increasingly sought in many professional fields (Uttl, 2013). While the call for 'data science' skills is resounding, learning outcomes are often difficult to define due to the noted variability (Wlodarczyk and Hacker, 2014).

Barriers to technical learning

Self-efficacy - The American Psychological Association defines self-efficacy as an individual's belief in his/her ability to produce specific performance attainments, which affect that individual's efforts and likelihood of success (American Psychological Association, 2015). Students who have not worked with data in the past have expressed insecurity in learning to use data and technical skills. Some students state 'I'm not so good with computers' and whether that assessment is true or not, it speaks to their self-efficacy. A pedagogical approach that focuses on students progressively mastering increasingly difficult tasks through a sequence of steps impacts efficacy. Social learning is also impactful as it allows students to observe peers struggling with materials and mastery of tools, as are they, which gives them a better sense of self-efficacy (Bandura, 1982).

Technical skills - Working with research data requires a number of skills that many students have not learned in previous educational experiences. These skills can include basic data finding and cleaning, data and file management, analyses and visualization. Skill levels vary, but even some very bright students have trouble with mundane tasks (e.g., downloading, moving and unzipping a folder), which are not intuitive.

Connection to real world scenarios - Skills taught in a vacuum are usually not easy for students to retain. Students receiving one-off instruction on data usage without application to a problem in their domain may have difficulty developing skills to work meaningfully with research data.

Blended approaches to pedagogy

Learning data science through GIS

Geographic information systems (GIS) are a combination of technologies, both software and hardware that allow users to describe, analyze, manage and visualize space using a myriad of data, including but not limited to spatial data. But this should not minimize the importance of space in GIS; in fact, Reed (2014, p. 280) calls location the 'great data integrator'. Location offers students who are new to data a familiar anchor for understanding. Throughout life, everyone uses data to make decisions and maps are commonly used to help audiences understand various data problems. Many individuals are unwittingly consuming data visualizations and analyses daily, but learning to make maps and see the data at the foundation allows students a better understanding (Drennon, 2005).

The foundation of problem solving with GIS is spatial thinking. Spatial thinking is the process by which students use the concept of space and the tools of representation to answer questions. The first function of spatial thinking is to define space and describe the objects or movement within it. Secondly, it functions to analyze the structure of objects in the space. These functions, facilitate making inferences, predictions and building arguments. By learning to apply spatial thinking to problems, students develop an understanding of the nature of data, how and where to find or create it, how to assess it and attribute it, how to build arguments and look critically at the arguments made in spatial representations (National Academies Press, 2005). Keenan and Fontaine (2012) describe combining methods of inquiry-based pedagogy, where students must be able to pose questions, find and assess reliable data and understand the interconnectedness within the data (e.g., objects in their environment) with student-centered instruction.

Problem-based learning and GIS

The approach of problem-based learning (PBL) directs students to work through complex, real life problems and attempt to solve them, often collaboratively. As Barrows (1986) explains, there is not one single PBL methodology, but the common thread in all of them is using problems as the basis of instruction. Often, PBL requires students to self direct, digest, reflect and identify knowledge and tools they may need. Instructors are tasked with creating problems with specific, yet implicit, learning outcomes that students can reach through the problem solving activity. Instructors may help students develop necessary skills to solve the problem, but don't include a step-by-step guide. This approach requires instructors to be ready to let students fall off-course and find their way back. Research suggests that PBL may facilitate a better conceptual understanding of their discipline as well as the development of soft skills (Allen et. al., 2012). Groups of students working collaboratively on a problem may increase their understanding and belief in their abilities to work through it. Inherently, PBL allows for students to take problems in any direction; it's noted that this can make a classroom chaotic (White, 1996).

Table - Research data learning outcomes using GIS

Data information literacy -	sourcing, understanding and using research data
Data cleaning -	correcting or removing records in a dataset; making the dataset operable
Data and file management -	organizing data in a structure that makes it easy to access and share
Combining datasets -	adding or joining relevant data to an extant, primary dataset
Querying datasets -	creating expressions to draw out particular data from a dataset
Analysis skills -	showing relationships, regression, clustering, etc. through analytical processes
Data modeling -	developing a structure that reflects how data and datasets will relate to others
Data collection -	gathering and inputting data
Data creation -	inputting or drawing new datasets in a workspace
Deriving data -	creating or modifying a dataset from a selection of a larger dataset
Visualization -	displaying selected data in a way that communicates an intended message
Assessment of arguments -	critically evaluating the products of others using data based on acquired knowledge and skills
Data attribution -	identifying and expressing data sources

Combining spatial thinking and PBL allows students from a number of disciplines to work with and understand data and analysis required to solve problems. Drennon (2005, p.397) found that by integrating PBL into a spatial project students became competent with data and GIS analyses 'almost by accident'. Because GIS is a complex tool on its own, integrating it into PBL requires scaffolding to give students footing in basic GIS skills and tasks before introducing the problem. Scaffolding, first articulated by Wood, Bruner and Ross (1976), introduces students to tasks just out of their capability and assists them in completing the tasks; as tasks are sequentially mastered, assistance for that task fades. Howarth (2011) suggests that GIS learning requires introduction to core, sequential skills working through guided problems and solutions first. Following the introduction to basic tools, the instructor may hand students a problem, which can be solved by building on those skills and with minimal guidance.

This approach to research data learning through problem-based, scaffolded GIS instruction can bridge the gap between novice students and research data. GIS learning encompasses many of the needed outcomes and through the problem based approach students connect to tangible problem solving and therefore better retain the technical skills.

Problems and praxis – case studies

History and mined textual data - close reading combined with GIS

Building on constructivist and active learning theory, Calandra (2005) described a model of learning using digital history resources and methods to help students develop their own understanding of history. To that end, technology employed in the classroom must be authentic, flexible, scaffolded and foster creative, independent thinking. Three case studies are presented in this history section, all of which are based on a problem that students were introduced to and investigated in a history course; one uses historical legislation, another uses a Holocaust memoir and another is based on research about a territory in dispute.

History Problem 1 - Turning historical legislation into digital data

The first case study is based on a project developed by a faculty member to visualize the establishment and growth of the early U.S. Federal government. The project has huge amounts of data to be digitized, attributed and mapped. In the first phase of the project a graduate student built a working model from mined legislative documents and existing legislative data. This student had been involved in a summer internship in the Humanities Digital Workshop (HDW), which supports long-term projects in the digital humanities on campus. The student had some basic experience with digital data, but not GIS. With assistance from the Data and GIS Services (DGS) team, which operates out of the Libraries, this student built a proof of concept mapping project. Working from that proof of concept, a team formed between the faculty, the HDW and DGS to develop a framework that would drive the project vision forward and also to introduce a pedagogical experience for undergraduates. The students would be introduced to research data, GIS skills and methods through the lens of history. Meeting these students in a familiar domain of study and working on content they understand creates an opportunity for conceptual understanding of how this data fit into the historical equation.

Upwards of thirty hours were dedicated to scaling the project goals into a classroom learning experience, including designing a flexible, data model that could fit into the scope of the class project, data preparation, looking for synergy with an existing database and working through example legislations to understand the parameters of the problem. The data/GIS portion of this class took place over six sessions. Basic skills were introduced in session one, and sessions two through four were designed to equip students with the specific skills needed to work through a piece of legislation in sessions five and six. The importance of sequence and scaffolding became apparent during the classroom experience. In some cases students moved into tasks before they were able to tackle the technology, but the students did have a very clear idea of the conceptual purpose of the exercise and what the data could do. By the final session, students became more comfortable manipulating the data. While using GIS to input and visualize this data made sense to the students, they also became aware of the hiccups that come along with that. Students were particularly engaged by the organization of the data model. They had questions regarding how to delineate between data that can be represented through a spatial feature and data as an attribute of a feature (e.g., when is a jurisdiction a boundary and when is it a level in the bureaucracy). Learning outcomes included: data creation, data and file management, data editing, data literacy and data modeling.

History Problem 2 - Making an argument using modern digital data and historical maps

The second case study was built on a historical problem that has bled into modern-day disputes over the Senkaku/Diaoyu Islands off the coast of east Asia. Currently Taiwan, Japan and China have claims to these uninhabited islands. The professor teaching this class wanted to introduce the dispute to freshmen and ask them to make arguments for different sides of the dispute. Part of that argument had to be expressed in a map they created that included both modern data and historical maps.

This course took place over eight sessions after intense study of texts on the conflict. Session one through three focused on data and GIS skills, which were presented sequentially and based on the functions of spatial thinking, to describe, analyze and make inferences in space. While moving through these skills students were reminded what functions the skills relate to. Sessions one through three focused purely on description, placing data, georeferencing, data editing and creation. Sessions four and five focused on understanding relationships through spatial and attribute data queries, basic analyses (e.g., buffering features), data finding and metadata. In session six, skills were reviewed. Session seven and eight served as guided practice so students could use the data and skills acquired to work in teams on their problem, to build an argument for a particular side of the dispute.

File management was a challenge to these students. Concepts of file naming, hierarchy and versioning were introduced in session one, but were difficult in practice. Working with zipped folders also presented challenges. While instructors focused on scaffolding the data representation and analyses methods, more attention to basic practices of file management and versioning was needed and will be emphasized in future sessions. Learning outcomes included: data creation, data literacy, data and file management, data editing, data analysis, digitization, and making an argument using data.

History Problem 3 - Using a memoir and digital data to recreate a journey & human experience

In this case study students began working from a text written by a Holocaust survivor chronicling her experience moving from her home, between camps and finally home again. Students were tasked with visualizing this experience using something more than dots on a map; the goal was to make a map that depicted something students deemed impactful about the journey, for example emotions (e.g. hope or fear), languages spoken, separation from families. etc. Students used methods like applying number values that described levels of emotion felt at each location based on their assessment of the text and then applied meaningful symbology to express it. Instruction took place over three sessions followed by three open guided-practice sessions. In session one, students learned to organize their data into features with attributes that they placed in a table; between session one and two they populated that table with whatever they deemed impactful. In session two they combined their data with established geodata and described the area. In session three students learned techniques for visualizing a map that tells a story. In this case, with only three short sessions to learn skills, students were left alone to tackle their problem. Some students made use of group open editing sessions designed for guided practice, but many needed individual appointments as well. In this case greater attention to the pace of scaffolding was needed, but in the process students did become more familiar with data and produced a unique map based on the attributes they selected to visualize. Learning outcomes included data literacy, data and file management, combining datasets, data creation, editing and visualizations.

Social Work – Teaching through primary collection - in the field
The field of social work has continued its move towards evidence-based research over the past several years. This shift has required students to develop proficiencies in primary data collection and data analysis. Although these skills are important, there continues to be debate about what research skills social work students need. For students in the UK, requirements encompass capabilities to interpret information through collection and analysis and an ability to assess materials and data (MacIntyer and Paul 2011). Data in the social work domain originates in numerous forms and much of it is applicable for GIS analysis. Felke (2014) makes an argument for the importance of GIS to the social work student's dossier, and he embedded within his own undergraduate course data literacy, data creation, aggregating data, data management and visualization outcomes. For some students such skills led to employment, and for others the skills were utilized for further study.

A social work case study introduced students to primary data collection fieldwork for a large-scale project in India to map villages with GPS devices and using colloquial understanding of the village and locals' sense of place. The field experience required students to contemplate how real world phenomenon will be represented in GIS and what will be valuable in their research. Once back in the classroom, all data points were mapped using Google Earth and GIS files generated for use in ArcGIS. Faculty worked closely with DGS on design and delivery of workshops to develop methodology for teaching students and locals to research, collect, process manage and archive data, including de-identification of data. A key aspect of this project was for students to learn the value of place in the context of analysis and decision-making. GIS was not the focus of their research, but rather a data visualization and analysis tool to support evidence based research. Learning outcomes include data literacy, data and file management, data collection and cleaning.

Architecture - Digital data reuse - aggregation for collaborative workshops
Monsur and Islam (2014, p.49) argue that architects and landscape architects are in a position to make better and more informed designs based on available digital data. In this case, data is not limited to features in the landscape, but also includes socio-cultural, socio-economic, behavioral or demographic information and the contextual relationships across a regional area. However, while using data, as a part of architectural decision-making is not necessarily common practice, Monsur argues that data and GIS methods should be an integral part of the planning process both to design efficiently and to decide whether it makes sense to design at all.

In the Sam Fox School of Design and Visual Arts at WUSTL numerous faculty have embraced spatial data and GIS approaches in their pedagogy. While several instructors have invited the DGS team in for talks and one time instruction on using GIS methods and data, some have dedicated large portions of instruction to data analysis using GIS. One example where this approach was successful involved two faculty members collaborating to deliver a multidisciplinary workshop focused on flooding and climate change in the St. Louis region. The workshop also investigated the regional relationship to the changing environment, both culturally and economically. Students utilized sourced and extant landscape feature data as well community data they collected. Source data used was primarily publicly available hydrologic, levee, boundary, agricultural and soil data. Faculty created base maps and used georeferenced photographs to record data from members of the community and other stakeholders. Students participated in various scenarios throughout the workshop and used the data to work through problems, including proposing future architectural development and addressing agricultural, ecological and navigational challenges. The workshop outcomes were models of potential future consequences based on different data-driven decisions.

Workshops like this one not only incorporate the idea that data can be applied to structural planning, but they illustrate that useful data comes in a variety of forms, from river levels to micro-data from local residents. It introduced the concept that data can be sourced or organically generated and that it requires management. Lastly, it presented the idea that factoring new data variables could change the parameters and outcomes of student design. This problem-based experience provided students tools and skills required to explore potential directions of architectural and landscape development, which created a rich learning experience for everyone involved. Learning outcomes include data creation, sourcing data, data and file management, data editing, data literacy, data analysis, digitization, and making an argument using data.

Conclusion
As shown through these case studies, research data learning outcomes were achieved through problem-based GIS learning. This evidence demonstrates how this approach can be an effective method to introduce students to skills and tools needed to work with research data. However, each case presented is an example of the first cycle of instruction. At least four of the five cases will be repeated, offering an opportunity for enhancement. Instructors have a better sense of the students' starting point and, following further assessment, can tailor materials in collaboration with faculty to better match learning outcomes. Determining the lasting effects of teaching the use of research data through GIS problem based learning will require more local assessment, but the techniques outlined here are working toward a higher rate of long-term retention. GIS can be applied to all sorts of problems; therefore it is a very suitable medium to deliver this material in various domains. Drennon (2005) asserted that GIS in research data management and modeling furthers scientific enquiry. With the growing call for skills around research data use, visualization, analysis, management and sharing from every corner of professional and academic life, the types of learning experience described can launch students forward to meet that call in various disciplines.

Acknowledgements
We would like to acknowledge the Faculty Members involved, Dr. Derek Hoeferlin, Dr. Peter Kastor, Dr. Tabea Linhard, Dr. Lori Watt, Dr. Gautam Yadama as well as Doug Knox and Bill Winston. We would also like to thank our colleagues who reviewed our manuscript, Ruth Lewis, Cynthia Hudson-Vitale and Bill Winston.

Bibliography
Allen, D. E., R.S. Donham and S. A. Bernhardt. (2012) Problem-Based Learning. New Directions for Teaching and Learning, 2011 (128) p. 21–29. Available from: http://onlinelibrary.wiley.com/doi/10.1002/tl.465/abstract

American Psychological Association. (2015) Self-Efficacy Teaching Tip Sheet. Available from: http://www.apa.org/pi/aids/resources/education/self-efficacy.aspx

Bandura, A. (1982) Self-Efficacy Mechanism in Human Agency. American Psychologist 37 (2): 122–47. Available from: http://dx.doi.org/10.1037/0003-066X.37.2.122

Barrows, H. S. (1986) A Taxonomy of Problem-Based Learning Methods. Medical Education 20 (6): 481–86. Available from: http://onlinelibrary.wiley.com/doi/10.1111/j.1365-2923.1986.tb01386.x/abstract

Drennon, C. (2005) Teaching Geographic Information Systems in a Problem-Based Learning Environment. Journal of Geography in Higher Education 29 (3): 385–402. Available from: www.tandfonline.com/doi/pdf/10.1080/03098260500290934

Felke, T.P. (2014) Building Capacity for the Use of Geographic Information Systems (GIS) in Social Work Planning, Practice, and Research. Journal of Technology in Human Services 32 (1-2): 81–92. Available from: http://dx.doi.org/10.1080/15228835.2013.860365

Howarth, J. T. and D.Sinton. (2011) Sequencing Spatial Concepts in Problem-Based GIS Instruction. Procedia - Social and Behavioral Sciences, International Conference: Spatial Thinking and Geographic Information Sciences 21 (2011): 253–59. Available from: http://www.sciencedirect.com/science/article/pii/S1877042811013656

Keenan, K. and D. Fontaine (2012) Listening to Our Students: Understanding How They Learn Research Methods in Geography. Journal of Geography 111 (6): 224–35. Available from: http://dx.doi.org/10.1080/00221341.2011.653651

Monsur, M. and Z. Islam. (2014) GIS for Architects: Exploring the Potentials of Incorporating GIS in Architecture Curriculum. ARCC Conference Repository. Available from: http://www.arcc-journal.org/index.php/repository/article/view/250

MacIntyre, G. and S. Paul. (2013) Teaching Research in Social Work: Capacity and Challenge. British Journal of Social Work 43 (4): 685–702. Available from: http://bjsw.oxfordjournals.org/content/early/2012/03/22/bjsw.bcs010

National Academies Press, ed. (2005) Learning to Think Spatially. Washington: National Academies Press. Available from: http://www.worldcat.org/oclc/63674143

Rickles, P. and C. Ellul, (2014) A Preliminary Investigation into the Challenges of Learning GIS in Interdisciplinary Research. Journal of Geography in Higher Education 0 (0): 1–11. Available from: http://dx.doi.org/10.1080/03098265.2014.956297

Reed, Carl. (2014) OGC Standards and Geospatial Big Data. In: Karimi, H. (ed.) Big Data: Techniques and Technologies in Geoinformatics. Boca Rotan: CRC Press. Available from: www.crcnetbase.com/doi/abs/10.1201/b16524-15

Uttl, Bob, Carmela A. White, and Alain Morin. (2013) The Numbers Tell It All: Students Don't like Numbers! PloS One 8 (12). Available from: http://journals.plos.org/plosone/article?id=10.1371/journal.pone.0083443

White, H. B. (1996) Dan Tries Problem-Based Learning: A Case Study. In: Richlin, L. (ed.), To Improve the Academy Vol. 15 (pp. 75 - 91). Stillwater, OK: New Forums Press and the Professional and Organizational Network in Higher Education.

The White House. (2015) Federal Register Notice Re OMB Circular A-110. Available from: https://www.whitehouse.gov/omb/fedreg_a110-finalnotice

Wlodarczyk, T. W. and T.J. Hacker. (2014) Problem-Based Learning Approach to a Course in Data Intensive Systems. 2014 IEEE 6th International Conference on Cloud Computing Singapore 2014. IEEE. Available from: http://ieeexplore.ieee.org/xpl/articleDetails.jsp?reload=true&arnumber=7037788

Wood, D., J.S. Bruner and G. Ross. (1976) The Role of Tutoring in Problem Solving. Journal of Child Psychology and Psychiatry 17 (2): 89–100. Available from: http://onlinelibrary.wiley.com/doi/10.1111/j.1469-7610.1976.tb00381.x/abstract

Notes

1. Director of Data and GIS Services at Washington University in St. Louis, 1 Brookings Dr., Campus Box 1169, St. Louis, MO 63130; aaddison@wustl.edu
2. GIS & Data Projects Manager & Anthropology Librarian at Washington University in St. Louis, 1 Brookings Dr., Campus Box 1061, St. Louis, MO 63130; j.moore@wustl.edu