Sustainability Science – manuscript accepted September 13, 2019. In press.

Bridging Sustainability Science, Earth Science, and Data Science through interdisciplinary education

Deana Pennington · Imme Ebert-Uphoff · Natalie Freed · Jo Martin and Suzanne A. Pierce

Abstract Given the rapid emergence of data science techniques in the sustainability sciences and the societal importance of many of these applications, there is an urgent need to prepare future scientists to be knowledgeable in both their chosen science domain and in data science. This article provides an overview of required competencies, educational programs and courses that are beginning to emerge, the challenges these pioneering programs face, and lessons learned by participating instructors, in the broader context of sustainability science competencies. In addition to data science competencies, competencies collaborating across disciplines are essential to enable sustainability scientists to work with data scientists. Programs and courses that target both sets of competencies – data science and interdisciplinary collaboration - will improve our workforce capacity to apply innovative new approaches to yield solutions to complex sustainability problems. Yet developing these competencies is difficult and most instructors are choosing instructional approaches through intuition or trial and error. Research is needed to develop effective pedagogies for these specific competencies.

Keywords: Education · Interdisciplinary studies · Competencies · Data science applications

D. Pennington - Geological Sciences, University of Texas at El Paso, El Paso, TX 79901

I. Ebert-Uphoff – Cooperative Institute for Research in the Atmosphere, Colorado State University, Fort Collins, CO 80523

N. Freed - Texas Advanced Computing Center, University of Texas at Austin, Austin, Texas 78758

J. Martin - Mathematics and Statistics, University of Vermont, Burlington, Vermont 05405

S.A. Pierce - Texas Advanced Computing Center, University of Texas at Austin, Austin, Texas 78758

Corresponding author: Pennington, E-mail: ddpennington@utep.edu; Phone: (915)747-5867; fax: (915)747-5073

Introduction

The past decade has seen the rise of data-intensive science (Hey et al., 2009), driven by advances in sensor technologies, expanded access to a wide variety of data sources, and a deluge of data generated by simulations. The term data science (DS) is widely used to denote the collection of scientific methods to extract meaningful insights or knowledge from data, and include emerging and rapidly changing methods in artificial intelligence, statistics, machine learning and data mining. As outlined in a recent National Academy of Sciences report on DS education (2018, p. 6),*"data science is inherently concerned with understanding and addressing real-world problems"*. Progress in application of DS to real world problems requires *integration* of general methodological knowledge along with knowledge of the particular data being analyzed (Fox and Hendler, 2014). Repko (2011) defined *interdisciplinary integration* as a process by which ideas, data and information, methods, tools, concepts, and/or theories are synthesized, connected, or blended. In this article we address the challenges of integrating DS methods with earth and sustainability science (ESS) content. While knowledge of both general DS methods and the ESS context of the data being analyzed can be held by a single individual, in practice this most often requires interdisciplinary (ID) collaboration between two or more individuals who work across their different disciplines (Pennington 2011b).

As the amount of data relevant to ESS is increasing dramatically every year (Plale et al., 2013), new methods from DS yield new scientific insights at record rates. For example, improved access to data and new technologies has driven innovations in sustainable cities, including the use of big visual data for managing construction sites (Tibaut and Zazula, 2018) and smart energy grids (Caputo et al., 2018). Yarime (2017) highlighted the importance of data science for achieving and monitoring the United Nations' Sustainability Development Goals. Seele (2016) and Kitchin (2014) envisioned big data providing a pathway for rigorous observation of sustainability performance. Barile et al. (2018) stated that the challenge of sustainability requires a transformation in research and education to move towards approaches that combine people, technology, and governance, fostering “smartness” through application of DS. Data-intensive methods are increasingly being employed to answer environmental questions relevant to sustainability, such as prediction and forecasting of natural processes (e.g., for weather, climate, environmental conditions, sea level, sea ice, availability of ground water and food) and pattern recognition and event detection (e.g., identifying conditions leading to extreme weather events, earthquakes and volcanic eruptions). Great strides have been made in all of these areas in recent years (Gibert et al., 2018; The World Economic Forum, 2018; Xie et al., 2017; Sellars, S.L. and others, 2017; Gil, Y. and others, 2015; Monteleoni et al., 2013). The World Economic Forum (2018) has recently identified artificial intelligence and other emerging technologies as key disciplines to help address environmental issues and redesign how we manage our shared global environment. The potential of data-intensive approaches in sustainability science has also been recognized by the DS community, which has called for a pooling of *“talents and knowledge to help find efficient and effective ways of managing and allocating natural resources”* (Gomes, 2009, p. 6). Likewise, the education section of the American Geophysical Union (AGU) concluded that *“The changing landscape of information technology (e.g., big data, emerging technologies, access to a wide variety of tools, rich multimedia) also affects the kinds and quantities of resources that are available for problem solving. Students must learn to navigate this rapidly changing space, identifying and harnessing resources (e.g., tools, data, models, experts, collaborators [...]) that can be brought to bear on the convergent problems”* (St. John et al., 2019, online - no page number).

The research community and industry are quickly embracing new DS methods, and it is important that curricula in ESS keep up with these rapid developments to give future scientists working on sustainability-related problems the competencies they need to be effective and competitive in an increasingly information-driven environment. Professional competencies are frequently categorized as required knowledge, skill, and attributes (attitudes and behaviors) to be successful in a particular career, that can be addressed as learning outcomes implemented in specific class or training activities, as exemplified in Wiek et al. (2015). Yet there are no existing articles that assess the state of formal education in DS and ESS. This article seeks to fill that gap in knowledge. The authors are part of a collaboration between geoscientists (mostly focused on water resource and climate issues) and data scientists. We set out to identify existing formal educational opportunities (as opposed to informal training workshops) focused on applying DS techniques to earth and environmental data. Our objective was to learn what skills are being targeted, what pedagogies are being used and how well they are perceived to work by those faculty. In the process we discovered substantial DS education efforts under the umbrella of computational sustainability. Hence, this article focuses on existing opportunities for advanced DS training in the environmental pillar of sustainability science and across sustainability science more broadly, yet undoubtedly similar educational efforts are emerging in other specific fields relevant to sustainability science.

The article consists of three major parts. The first part discusses key competencies needed to accelerate data-driven discovery in ESS, especially the need for collaboration competencies; the second part addresses programs and courses we identified and strategies being used in the design of individual courses, and the third part places our findings into the broader context of sustainability education competencies. While the second part integrates empirical information gathered from surveys and interviews of educators at the forefront of these emerging fields, the first and third parts synthesize theory and concepts from existing literature.

Challenges of Interdisciplinary Collaboration to Advance Data-Driven Discovery in Sustainability Science

The necessity of collaboration in sustainability data science

It is becoming clear that attacking ESS problems with DS requires significant knowledge in both (Sellars, S.L. and others, 2017; Ebert- Uphoff and Deng, 2017; Gil, Y. and others, 2015; Monteleoni et al., 2013). DS typically requires training in statistics, computer science, or a similar discipline, and choosing appropriate techniques requires a deep understanding of the available data analysis methods, their underlying model assumptions, computational effort, required sample size and pitfalls. While it is possible for an ESS expert to independently learn to apply DS techniques to problems of interest if they have substantial training in statistics, many ESS programs generally do not require statistics (Haider et al., 2018) or other DS courses, and it is well-recognized that earth and environmental scientists, specifically, need additional training to manage and use complex data (Hou, 2015). The Belmont Forum performed a skill gap analysis for data-driven research in the environmental sciences, identifying a lack of competencies in programming, visualization, data management, and data exchange - all core DS competencies (Belmont Forum, 2017).

Conversely, while some data scientists focus on particular kinds of ESS data and develop knowledge in that domain, defining the science question to be considered requires deep ESS knowledge, such as awareness of important open science questions, understanding of their underlying physical, economic, and social processes, including spatial and temporal scales, availability of suitable datasets, and selection of data preprocessing strategies to maximize the strength of the signal or patterns to be detected and minimize noise (Ebert-Uphoff and Deng, 2017). In addition, ESS data display spatial autocorrelation and geographic heterogeneity, requiring special statistical techniques (Berman et al., 2018). Combined knowledge is required throughout all phases of the research, in particular when it comes to interpreting the results and explaining what they mean. As economist and Nobel laureate Ronald Coase once said, *if you torture the data long enough it will confess* (Tullock, 2001, p. 205), implying that algorithms that look for patterns long enough will eventually identify *some kind* of pattern, but it might not actually be meaningful or new. Thus, it is important to carefully evaluate the results obtained from such analyses. Do the results represent a real physical phenomenon, or are they merely an unforeseen by-product of the data collection or analysis method? Both ESS and DS competencies are required to make this assessment. Thus, working independently neither a data scientist, nor an ESS expert, as trained in today's educational environment, can be effective in this area.

Virapongse et al. (2018) identified the need for education and training activities that improve data scientists' understanding of domain topics and improve ESS's knowledge of data management. There is a need to train current and next generation ESS professionals to extract usable information from the vast supply of data, train DS professionals to be competent in specific ESS areas, and train both to collaborate effectively to develop appropriate new methods for identifying meaningful, interpretable patterns across heterogeneous datasets (Blei and Smyth, 2017; Pankratius et al., 2016; Sellars et al., 2013; Szalay and Gray, 2006). Workforce development in ESS and DS and interdisciplinary collaboration are closely intertwined (Hampton et al., 2017; Kempler and Mathews, 2017).

Barriers to collaborating across disciplines

There is a large body of literature on ID collaboration (see Stokols et al. (2008); Bozeman et al. (2013); Hall et al. (2018) for summaries). While research on organizational teams dates back many decades, research on science teams, also known as the science of team science, is relatively recent (Falk-Krzesinski et al., 2011; Hall et al., 2008) and important challenges have become apparent in these teams compared with better studied teams in organizations (Hall et al., 2018).

ID research is known to be fraught with difficulties, taking considerably more time and effort compared with disciplinary research. Describing all important teamwork competencies is far beyond the scope of this article; here we only summarize a few basic concepts. The next section will address one of these concepts in detail.

Klein and Newell (1998, p. 3) provide the following widely used definition of interdisciplinary studies (IDS): *"interdisciplinary studies may be defined as a process of answering a question, solving a problem, or addressing a topic that is too broad or complex to be dealt with adequately by a single discipline or profession. […] IDS draws on disciplinary perspectives and integrates their insights through construction of a more comprehensive perspective."* The concept of *integrating* is central in this definition and is key to any interdisciplinary effort (Repko, 2011).

Interdisciplinary collaboration can be categorized by the level of integration, as outlined by Klein (2010) and illustrated in Fig. 1. For an extensive review of the many ways these terms have been used in literature, see (Choi and Pak, 2006). Although terminology has varied substantially, most have converged on the following definitions:

- Multidisciplinary: Combines separate perspectives under a theme.
- Interdisciplinary: Combines separate perspectives through the development of connections between them.
- Transdisciplinary (meaning 1): Combines separate perspectives across academic and other sectors (e.g., government agencies, industry, organizations, etc.) through the development of connections between them to generate research that is informed by stakeholders.
- Transdisciplinary (meaning 2): Generates a new area of knowledge that deeply combines originally separate perspectives and may lead to new scientific frameworks, paradigms, or disciplines (for example, bioinformatics combined genomics and data mining).

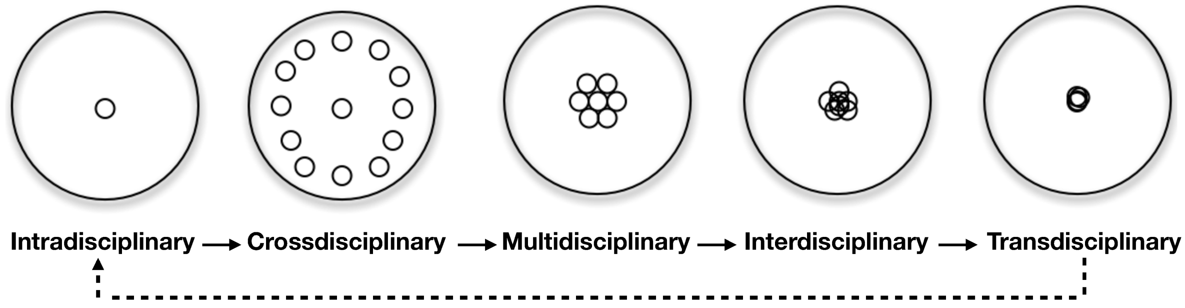

Figure 1. Overview of different types of collaborations, where level of integration increases from left to right. Once transdisciplinarity (meaning 2) is reached a new discipline is generated (e.g., the field of bioinformatics) and the cycle can start over. Transdisciplinary (meaning 1) extends interdisciplinary integration to sectors outside of academia. *Image credit: Blog post by Alexander R. Jensenius (Jensenius, 2012). Image modified to switch the terms Crossdisciplinary and Multidisciplinary. Permission to use and modify image granted by the blog author.*

In practice, it is exceedingly difficult to evaluate where any given team is along the continuum from multi- to inter- to transdisciplinary (Masse et al., 2008). The remainder of this article will use the term interdisciplinary (ID) to refer to the entire continuum from multi- to transdisciplinary.

It is known that disciplines each have their own culture, vocabulary, epistemology, and ways of thinking that are barriers to integration across disciplines (O'Rourke et al., 2016; Stone, 2013) and that integration across disciplines can be challenging (National Research Council and others, 2015; Kliskey et al., 2017). Researchers who successfully engage in ID efforts consistently report that despite the challenges, they enjoy ID work because it causes them to think about their own research in new ways. Hence, there is a cost/benefit trade off that must be assessed individually.

ESS and DS professionals have little formal training in common beyond a core curriculum. They have very different approaches to research, and little understanding of what constitutes significant research for the other discipline (Pennington, 2011a). For example, we have observed that research in ESS *revolves around a specific science question to be answered,* thus datasets and algorithms are only seen as *tools* to answer that question. In contrast, research in DS is often driven by a specific algorithm or dataset. In addition, ESS has traditionally relied on physics-based approaches, namely developing models from governing equations and principles that were discovered over the span of centuries. In contrast, today's data-driven approaches often do away with all of the rich domain knowledge and develop models entirely based on a single dataset - an approach that has some well-known short-comings and may yield questionable results (Karpatne et al., 2018). A great deal of learning across disciplines and close collaboration must occur before connections can be made that result in integrated research for both (Pennington, 2008; Pennington 2011b). ESS students, especially when first exposed to DS methods, often get the impression that data-driven approaches may disregard their domain knowledge. In reality, success using purely data-driven approaches without domain expertise is exceedingly rare. While data-driven approaches have yielded good results in image processing or identification of gene interactions, their application in other science domains is not as straight forward (Karpatne et al., 2018). Hence, both DS and ESS science competencies must be synergistically applied. In fact, the growing field of theory-guided data science is based on the idea that new DS approaches for ESS applications are best developed in collaboration with scientists and that the most effective DS approaches are achieved by incorporating domain knowledge (such as physical constraints or relationships) in their algorithms (Faghmous and Kumar, 2014; Faghmous et al., 2014; Karpatne et al., 2017).

Even when DS and ESS scientists are working well as a team, this type of work still requires each one to have significant knowledge in the other area to achieve true synthesis. However, to date there is very little consensus on how to teach these collaboration competencies. The report of the 2017 workshop on *Big Data and the Earth Sciences: Grand Challenges* (Sellars, S.L. and others, 2017, p. 9) summarizes a discussion on education as follows: "In the end, the recognition that there is a dire need for people with skills in both camps was unanimous, but there was no clear answer on how best to integrate or coordinate their knowledge, or what should be expected from students and researchers who participate in this interface." Thus, it is fair to say that even among experts at the forefront of research in the sustainability data science field there is little consensus on how to overcome these obstacles themselves and to best train the future workforce.

ID competencies can only be developed in conjunction with certain attitudes (Haider et al., 2018). A willingness to learn difficult topics from another discipline and appreciation of different perspectives are critical (Pennington 2008). Also critical is the ability to work with others whose personalities and work styles may be quite different from one's own (Gosselin et al., unpublished data). Methods are being developed for overcoming ID issues, focused on a variety of challenges, including generic teamwork issues (Fiore, 2008); differing disciplinary values, epistemologies, and philosophies (Eigenbrode et al., 2007; O'Rourke et al., 2013); and the challenge of integrating knowledge across disciplines (Pennington, 2016; Pennington et al., 2016; Pennington et al., in review). Research on ID education suggests that these skills are best taught in the context of solving particular problems (Bosque-Prez et al., 2016; Derrick et al., 2013).

Models of interdisciplinary programs

There are many different models for developing ID educational programs, each of which has been successful in some settings. The most common approach is often referred to as T-shaped education (Oskam, 2009) (Fig. 2(a)), with students acquiring specialized knowledge in one field (the vertical leg of the T), along with broad elementary knowledge in adjacent fields and soft skills (the horizontal bar of the T). Another proposed model is pi-shaped education (Fig. 2(b)), which has been described in different ways. It may represent deep expertise by two separate people in two disciplines (the two legs of the pi) connected either by a person who understands something about both, or through development of improved cross-disciplinary skills by experts from each discipline. It has also been used to represent a single person obtaining deep expertise in two separate disciplines. There are examples of both of these in the team science literature. Recently, Bosque-Perez et al. (Bosque-Prez et al., 2016) have proposed a shield-shaped model of ID education (Fig. 2(c)), with deep expertise in one discipline combined with practical understanding of one or two other disciplines, sufficient to enable collaboration with researchers from those disciplines.

Ceri (2018) has proposed that DS should be taught using a pi-shaped model, with students from any discipline also acquiring deep knowledge of DS, arguing that these are skills needed by students from all disciplines. However, his description more closely resembles the shield-shaped model. Hampton et al. (2017), discussing DS education for environmental scientists, argue that these students should all obtain foundational knowledge and skills in 1) data management, 2) analysis, 3) software for science, 4) visualization, and 5) communication methods for collaboration and dissemination, although their description more closely resembles a T-shaped model with the horizontal bar including DS basics. We think that perhaps these models represent a continuum similar to the discussion above on multi-, inter-, and transdisciplinarity, and that the model to choose depends on many factors at the host institution, including existing focal points, resources and expertise available for new course development, and intended scope and goals. For example, is the program's goal to prepare graduates: (1) to become *productive team members* of such ID efforts in industry or academia; (2) to *lead* such ID efforts; or (3) to be able to conduct ID efforts on their own, and thus be fully trained in both ESS and DS? These roles require different training and thus different types of educational components.

We propose that all ESS students should be exposed to basic DS concepts and should develop collaboration skills as part of an institutional core curriculum (T-shaped). In addition, ESS students who desire to work on ESS/DS ID teams should be encouraged to take basic statistics and DS courses as upper division electives to gain a practical understanding of DS (shield-shaped). Statistics is not required for many ESS programs. Applied DS and statistics courses should be developed for ESS graduate students who intend to conduct data-intensive research on their own (pi-shaped, one person with depth in two disciplines). Finally, educational paths should be developed for people who specialize in being the "connector" in a pi-shaped model, critical to the success of ID research teams. This is especially relevant to role (2) above, leading ID efforts. Graduate certificate programs could be developed exposing PhD level students and/or post-graduates to practical results from the team science literature.

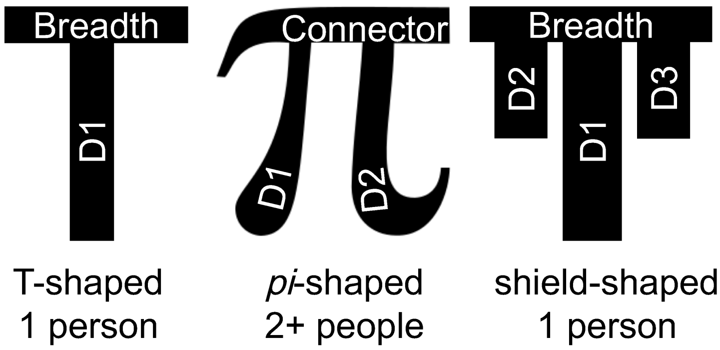

Figure 2. Models of interdisciplinary education, using (a) T-, (b) pi-, and (c) shield-shapes as metaphors for describing the desired relationship between knowledge in a primary and secondary disciplines (D1, D2, and D3).

Social Learning in the Context of Interdisciplinary Collaboration

The National Research Council report (2015) on enhancing the effectiveness of team science identified seven features that challenge such ID efforts: 1) high membership diversity, 2) deep knowledge integration, 3) large team size, 4) goal misalignment, 5) permeable boundaries, 6) geographic dispersion, and 7) high task interdependence.

Among these, three are especially relevant in the context of ESS in general, and ESS data science in particular: the high degree of disciplinary diversity between ESS and data scientists; the need for deep knowledge integration across these fields; and high task interdependence.

These are particularly challenging because integrating deep knowledge depends on understanding enough about the other discipline to be able to find interesting, synergistic connections between the two that provide research opportunities for both.

Spelt et al. (2009, p. 365) define interdisciplinary thinking as: "*the capacity to integrate knowledge of two or more disciplines to produce a cognitive advancement in ways that would have been impossible or unlikely through single disciplinary means"* and include the ability to change perspectives, to synthesize knowledge of different disciplines, and to cope with complexity as specific skills that are needed. Newell and Luckie (2013) suggest students should develop ID *habits of the mind* including 1) drawing insights from diverse perspectives; 2) evaluating insights; 3) modifying insights; and 4) integrating insights, with specific suggestions for actions that accomplish each of these. Yet in very diverse disciplines, there may be few basic concepts understood by both, impeding progress integrating deeper knowledge. It requires significant time spent learning each other's vocabulary, methods, and concepts, usually in real time during collaboration (Pennington 2008, 2011a, 2011b, 2013). This experiential, social learning includes developing an understanding of the vocabulary, methods, and concepts around specific research topics of interest to participants and also the more basic concepts needed to understand those research topics. It is important to realize that no research team begins its work at the inter- or transdisciplinary level; rather, teams begin as multidisciplinary groups and develop into inter- or transdisciplinary teams as they learn each other's perspectives and find relevant connections (Pennington, 2011b). Teams that successfully navigate this social learning space end up with deeply integrated knowledge across disciplines that can ultimately develop into inter- and transdisciplinary outcomes. ESS data science and computational sustainability, if they are to become more widespread and mainstream, depend on finding more effective ways of supporting the difficult learning process involved in collaborating across ESS and DS disciplines.

The Employing Model-Based Reasoning in Socio-Environmental Synthesis (EMBeRS) initiative is a collaborative, social learning research effort initially formed through the U.S. National Center for Socio-Environmental Synthesis (SESYNC) followed by funding from the U.S. National Science Foundation (NSF). The goal of the initiative is to synthesize learning, social, and organizational science theories relevant to social learning across disciplines to design and test collaborative knowledge integration activities in the context of ESS, and develop new understanding of how social learning occurs in this context (Pennington 2016; Pennington et al. 2016; Pennington et al., in review). "Model-based reasoning" is a theory derived from the cognitive science community, who have found that during complex problem solving, progress may be enabled by offloading information in the form of visual (or other) representations of internal mental models, and that the creation of new external representations not only supports conceptual change but actually *invokes* such change (Nersessian, 1999). Related theories have come out of the social sciences, particularly the notion of "boundary objects" that support exchange of information across disparate social groups (Star & Griesemer, 1989). Certain boundary objects, boundary *negotiating* objects, have been found to facilitate negotiation of new conceptualizations that cross boundaries between perspectives (Lee, 2007; D. Pennington, 2010). Hence, model-based reasoning is a knowledge exchange process through which boundary negotiating objects can be purposefully (co-)constructed to facilitate knowledge integration, and is the basis for the EMBeRS method (Pennington, 2016; Pennington et al., 2016).

The EMBeRS method purposefully integrates boundary object negotiation with participatory processes, with a specific focus on learning issues, applying transformative (Mezirow, 1997) and experiential learning (Kolb, 1984) theories in addition to model-based reasoning theory. The method balances individual mental model representation with group co-created representations, enabling individuals to organize their own thinking before contributing to the group. In addition, participants are guided to interact in certain ways during the process, by attending to fundamental learning concepts such as active listening, jargon reduction and/or lay descriptions, individual and group reflection on the process, among others. A key concept that must be understood is that the group itself needs to develop into a working, distributed cognitive system (Hutchins, 1995), and that takes time (Pennington, 2011b; Pennington 2016). Synergistic research ideas will eventually emerge from the distributed cognitive system as it evolves and connections are identified. Synergistic ideas are an emergent property of the system (Pennington 2011b, 2016). Until the multidisciplinary group truly becomes an inter- or transdisciplinary distributed cognitive system, formulation of the collaborative research to be undertaken will be vague and ambiguous (Pennington et al. 2013).

The EMBeRS method has been tested in two workshops for doctoral students held at the University of Texas at El Paso (UTEP). A total of twenty-five students from seventeen different U.S. institutions participated. Participants were recruited from eighty-six NSF-funded research projects related to water sustainability. Five were doctoral students at UTEP. Participants had a wide range of research interests and represented a variety of disciplines in the natural sciences, social sciences, data science and engineering. The participants were 58% women, 23% Hispanic, 19% Asian, and 8% Black. A third (31%) were international students from Brazil, Ghana, India, Libya, Nepal, Nigeria and Vietnam.

The workshop focused on training students to integrate knowledge across disciplines using water sustainability in the Middle Rio Grande Basin in the southwestern U.S. as a case study. This region was selected because almost any ESS researcher can relate their research in some way to water; research has shown that ID projects are more likely to be successful if they are place-based (Bosque-Prez et al., 2016; Derrick et al., 2013); UTEP is located in the region, and the workshop leader (Pennington) is familiar with its water sustainability issues. That enabled the workshop design to incorporate a wide variety of perspectives on the problem; diverse methods and data; guest lectures; and fieldtrips to meet with stakeholders of different types (including farmers, water managers, environmental restoration specialists, and civil engineers). Since the students recruited were already associated with water sustainability projects, they already understood parts of the problem and did not need to struggle with completely new problem content while also struggling to acquire new skills working across disciplines. Therefore, while the goal of the workshop was to provide training on ID collaboration and especially knowledge integration across disciplines, these skills were embedded in a real problem so that students could experience how these skills are applied in a concrete rather than abstract way. Experiential learning theory, under development and tested rigorously for decades, has indicated that concrete experiences lead to abstraction of content and transfer to other applications (Kolb 1984 and thousands of subsequent references). This same approach has been used by Pennington in a semester-long graduate course that has been taught four semesters and is now a required course for Masters and Doctoral programs at UTEP.

Data were collected before, during, and six to nine months after each workshop. Detailed design of the workshops and research agenda are described in Thompson et al. (2017) and Pennington (in review); participant outcomes are reported in Pennington et al. (in review) and findings from studies of the process are in (Thompson, 2009; Thompson et al., 2013; Thompson, et al., 2016).

We have continued to hear from many of these students through time. One measure of the effectiveness of the EMBeRS method is the enthusiastic and frequent use of the ID skills they learned in all aspects of their professional lives. Several have reported the value of the EMBeRS approach in planning and presenting their own research and in a diverse array of ID research projects and proposals. The 2017 cohort highlighted two aspects they found extraordinary: 1) how they all came to respect and trust each other and how that allowed them to effectively integrate individual personalities and strengths, perspectives and disciplines; and 2) how the experience gave them the confidence to better understand and articulate their own ideas. They also noted that the workshop broadened their perspectives and helped them develop understanding of how to effectively negotiate and compromise in developing shared understanding and consensus.

The two EMBeRS workshops did not include participants from DS by design, although some participants had strong, targeted DS skills. However, early work on the method successfully brought together environmental scientists with computer scientists (Pennington, 2008, 2010, 2011b, 2011a). As the method continues to develop, it may be an effective way to train students from any discipline to more effective integrate their knowledge with other disciplines. Such skills are a critical need in the future ESS workforce.

Methods for collecting information on existing programs and courses

We now turn our attention to how all of the above education and competency issues are playing out in specific ESS and DS education contexts. We contacted an array of relevant research communities and individual educators to get a broad selection of existing programs and courses integrating earth science and data science. The first set of activities included reaching out to mailing lists, conducting web searches and querying the NSF award database. Although our search was not constrained to the U.S., it resulted in a U.S. bias because we searched English language sources, personally contacted participants in our own, U.S. based network, and searched the database of a U.S. funding agency.

1. **Mailing lists:** We reached out to the following three mailing lists.
   (a) *Intelligent Systems in the Geosciences (IS-GEO)* email list: serves our NSF-funded network that partners earth, computer, data, and intelligent systems scientists;

(b) *American Geophysical Union (AGU) Earth and Space Science Informatics (ESSI)* Section email list: reaches a large international community of earth, computer, and data scientists who are very active in research and development of sustainability data science-related techniques; and

(c) *Machine Learning-news* email list: ML-news is an international Google group that covers all topics related to machine learning, data mining and other data analysis techniques and reaches a large community of computer scientists working in this area.

2. **Web searches:** We performed extensive web searches using the following keywords alone and in many combinations, e.g., combining one term each from Groups 1, 2 and 3:
   - Group 1: data science, data mining, machine learning, intelligent systems, artificial intelligence.
   - Group 2: computational sustainability, geo, geosciences, earth sciences, geology, climate, environment, earth.
   - Group 3: course, curriculum, education.
3. **ICS resources:** In addition to the above web searches we perused the extensive resources of Cornell University's Institute for Computational Sustainability (ICS) focused on "developing computational methods for balancing environmental, economic, and societal needs for a sustainable future", see https://computational-sustainability.cis.cornell.edu/.
4. **NSF award searches:** The above activities yielded only two ID programs funded by the NSF through Research Traineeship (NRT) grants. We then searched through all current and past NSF NRT and IGERT grants (IGERT was the predecessor program to NRT) using NSF award searches. This yielded one more ID program, bringing the total to only three, all in the U.S.
5. **Survey and follow-up interviews**: We conducted a survey of people who responded to the above contacts using a Google form that requested information about any programs, courses, course modules, or training workshops in which they participated. The survey requested information regarding the type of course, level taught, focus of course, and other information detailed in a later section on Course Design. The survey was open from January of 2017 through February of 2018. This returned 11 different responses to the survey from 9 unique respondents some of whom filled out the survey for more than one course, for a total of 14 courses. Seven respondents were from the U.S. and 2 from non-U.S. institutions. At the beginning of April 2018, survey respondents were contacted with a more detailed survey,

and 2 more detailed responses were received. Using a grounded theory approach (Corbin and Strauss, 1990) keywords related to subject matter and pedagogy were extracted from the survey responses, and courses were classified based on whether or not the responses mentioned these keywords.

From the surveys, 2 people were selected for follow up video conference interviews. These interviews were conducted on April 24, 2018 and May 16, 2018 and were 28 and 18 minutes long, respectively. Participants for video interviews were selected based on survey responses, particularly the amount of detail, specificity and reflection provided. The video interviews were semi-structured, with questions informed by the interviewees' prior survey responses. Semi-structured interviews were chosen to allow both survey participants the opportunity to speak on the same topics while allowing the interviewer to pursue things brought up in the interview with more depth. Participants were asked to give more detail about their responses to the questions about overall course design, challenges encountered, and successful student projects. The interview notes were then summarized into key points, and keywords from the interviews were used to classify which courses from the survey responses addressed points brought up by interview participants.

The survey and interview research on human subjects were overseen and determined to be exempt by the University of Texas at Austin Institutional Review Board under protocol 2018-05-0069.

6. **Authors' experience:** We synthesized our own knowledge and experiences working in these contexts with our findings from the surveys and interviews. The authors are all part of the afore-mentioned IS-GEO program and some of the authors have long term (many years) of experience working across science and information technology disciplinary boundaries on other research and education projects.

Results

First, we report on the programs and courses identified through our searches. Then we present the results from surveys and interviews with instructors.

Existing interdisciplinary education programs

We identified the following three ID data science education programs, all of which implement some of the principles discussed above. Two are directly relevant to sustainability science; the third is earth science focused:

- Data Science for Energy and Environmental Research – University of Chicago (USA),
- Environment and Society: Data Sciences for the 21st Century – UC Berkeley (USA),
- Integrated Data-Driven Discovery in Earth and Astrophysical Sciences (IDEAs) – Northwestern University (USA),

All three are NSF NRT programs. They primarily supplement traditional graduate work, and they most closely resemble the shield shape model of ID education, with a student's PhD discipline serving as the field in which they have deep expertise. The NRT then exposes the students to a breadth of disciplines through a variety of means, including bootcamps, workshops, and traditional courses. Each NRT builds up practical understanding in DS through coursework and projects. All three programs emphasize ID teamwork, either through coursework or, in the case of the IDEAs program, a citizen science project that partners students with non-scientists in the community to train citizens to collect and contribute relevant and valid scientific data. Communication is also emphasized in all programs, with the University of Chicago and Northwestern programs having courses to train students in communication, and the Berkeley program including a communication bootcamp. The Berkeley program explicitly addresses other aspects of ID education, including interdepartmental immersion programs and an

Interdisciplinary Research Design and Methods class. The Appendix contains links to all three programs.

In addition, the University of Colorado Boulder offers an Earth Data Analytics Certificate that is obtained by completing three courses (9 credits) within three years. Available courses are 1) Earth Analytics Data Science Bootcamp, 2) Earth Analytics Python, and 3) Earth Analytics Applications. The only prerequisite is an undergraduate degree in *any* field. As a 3-course program with a target audience that might have neither an earth science nor DS background, this is a good example of an entry-level program. It does not specifically address topics such as ID collaboration but provides many detailed lesson plans; see the list of related resources in the appendix.

Existing courses

A major driving force for courses in Computational Sustainability is Cornell University's Institute for Computational Sustainability (ICS), created through a large U.S. NSF Expeditions in Computing grant awarded in 2008, that focuses on "developing computational methods for balancing environmental, economic, and societal needs for a sustainable future". In addition to engaging in research activities, ICS runs CompSustNet, a multi-institutional research network that includes Cornell, 11 other U.S. academic institutions, and international collaborators. CompSustNet runs an annual doctoral consortium on Computational Sustainability, and a Journal of Computational Sustainability is planned for the near future, to serve the growing needs of this emerging field. Table 1 provides a list of courses in computational sustainability identified by Fisher et al. (2016), most of which are described in more detail in the corresponding CompSustNet blog post. Fisher et al. (2016) point out that most of these courses are seminar-based, and only a few include projects.

Table 1. Courses related to Computational Sustainability as identified in Fisher et al. (2016) and discussed in more detail in the corresponding CompSustNet blog entry (see https://blog.computational-sustainability.org/2016/04/11/university-courses-in-computational-sustainability/).

Course Title	**Institution**	**Source**
Sustainability and Assistive Computing (Fall 2010)	Bryn Mawr College	Fisher et al. (2016) & CompSustNet blog
Computing and the Environment (Spring 2011)	Vanderbilt University	Fisher et al. (2016) & CompSustNet blog
Computing, Energy, and the Environment (Fall 2016)	Vanderbilt University	Fisher et al. (2016)
Topics in Computational Sustainability (Spring 2011)	Cornell University	Fisher et al. (2016)
Computational Methods in Sustainable Energy (Fall 2012)	Carnegie Mellon University	Fisher et al. (2016) & CompSustNet blog
Computational Sustainability (Winter 2013-2014)	University of British Columbia	Fisher et al. (2016) & CompSustNet blog
Computational Sustainability (Spring 2014)	Georgia Tech	Fisher et al. (2016) & CompSustNet blog
Seminar on Computational Sustainability: Algorithms for Ecology and Conservation (Spring 2014)	University of Massachusetts Amherst	Fisher et al. (2016) & CompSustNet blog
Topics in Computational Sustainability (Spring 2016)	Stanford University	Fisher et al. (2016) & CompSustNet blog

Furthermore, the body of literature focusing on *Computing Education for Sustainability* (Mann et al., 2008; Mann et al. 2009; Mann 2016) seeks to integrate sustainability topics into computing education. Mann et al. (2009) provide important guidelines for material selection and material collections, which are discussed in the subsection on material selection below, while Mann (2016) discusses challenges for this area and how they might be resolved.

We also identified a broad selection of courses that combine ESS and DS (Table 2) spanning undergraduate and graduate courses, taught primarily by DS instructors (e.g., [C14]), by ESS instructors (e.g., [C3,C7]), or by a combination of the two (e.g., [C1-C2,C4-C6]).

Table 2. List of additional courses we identified that bridge ESS and DS. Contact person and email addresses are included with permission from each instructor, since many of these are available online at course websites.

Course title	Institution	Contact Person
[C1] Spatial Computing	U Minnesota	Shashi Shekhar (shekhar@umn.edu)
[C2] Spatial Data Science Research	U Minnesota	Shashi Shekhar (shekhar@umn.edu)
[C3] Computer Applications in the Geosciences	UTEP	Deana Pennington (ddpennington@utep.edu)
[C4] Geoinformatics for Natural Hazards Monitoring	MIT	Victor Pankratius (pankrat@mit.edu)
[C5] Decision Pathways for Earth Resources	UT Austin	Suzanne Pierce (spierce@tacc.utexas.edu)
[C6] Big Data	Wageningen U	Ioannis N. Athanasiadis (ioannis@athanasiadis.info)
[C7] Food, Energy, Water, the Environment, and Public Policy	U. of Kansas	Mary C. Hill (mchill@ku.edu)
[C8] Climate Science and Engineering	Northeastern U	Auroop Ganguly (a.ganguly@northeastern.edu)
[C9] Temporal and Spatial Data Science	Northeastern U	Auroop Ganguly (a.ganguly@northeastern.edu)
[C10] Critical Infrastructure Resilience	Northeastern U	Auroop Ganguly (a.ganguly@northeastern.edu)
[C11] Machine Learning	U Oklahoma	Amy McGovern (amcgovern@ou.edu)
[C12] Artificial Intelligence	U Oklahoma	Amy McGovern (amcgovern@ou.edu)
[C13] Advanced Machine Learning	U Oklahoma	Amy McGovern (amcgovern@ou.edu)
[C14] Data camp / RAMP (https://www.ramp.studio)	U Paris Saclay	Balazs Kegl (balazs.kegl@gmail.com)

Strategies for course design

This section presents quotes by the instructors of the above courses that highlight their motivation in course design, challenges encountered, and the lessons they learned while teaching these classes. Quotes are attributed to specific people with permission. These quotes often refer to the terms *IS* (for intelligent systems) and *GEO*, because those terms are commonly used by the community from which many of the surveys and interviews were derived. In the context of this article, one can think of replacing IS with DS and GEO with ESS in these quotes.

Course prerequisites

A challenge in any ID course is the need to establish common ground for students from different disciplines. For example, Sellars et al. (2017) state that students with interests in DS from disciplines outside of the traditional Statistics and Computer Science "can't just pop in" on courses that focus on statistical and computational methods. Ritu Arora (course [C5]) observes that *"A preliminary programming course should be made a prerequisite so that the IS-GEO class can be made more about the application of CS/IS techniques and has more depth on selected topics"*. Thus, a sequence of special introductory courses or bootcamps is essential, to bring both DS and ESS students up to speed with the other discipline.

Group projects

Group projects are an extremely powerful means to develop ID skills, such as communication and working in teams. Of the courses surveyed (Table 2), every course that included both DS and ESS students featured group projects. Pankratius (video interview) observes that *"Bringing together PhD students from different departments (earth science, computer science, physics, etc.) and working on case study projects together enables cross-fertilization and interdisciplinary learning experience for everyone at early career stages."* McGovern (video interview) pointed out that although desirable, bringing together students from different disciplines is not always easy. For group projects, given the option, students may be inclined to pair up with peers from the same discipline because of shared background knowledge, approaches, and/or interests. Instructors may explicitly require ID teams. Leaving room for more ID teamwork to emerge from teams that do not start out aligned, reinforces that even teams that do not start out functioning in a perfectly ID way can grow into it, as pointed out by Pierce ([C5]): *"Leaving ambiguity in the project description results in the best outcomes and learning by students; while the students are uncomfortable at first they usually commit to the project as a team to define it (e.g. active co-design) and I receive messages from past students in the workforce stating that the course was the most valuable in their graduate careers."*

Material selection

In this section we discuss general frameworks and considerations for material selection, including a summary of feedback obtained from our surveys and follow-up interviews. Please see the appendix for a list of additional educational resources we identified from other sources. Mann et al. (2009) provide an interesting framework for the evaluation and selection of resources for computer science educators to help them integrate sustainability topics in computing courses. The evaluation framework was developed by a working group and consists of a list of five categories, and within each category of several questions that guide the evaluation of any considered resource. They also provide a list of 14 sample resources they identified, namely 8 papers or news stories, 1 set of lecture notes, and 5 related books.

Considering the question of breadth vs. depth, we observed a variety of approaches among the courses we identified, ranging from specifically targeting an ESS domain for *deep* study ([C4]) to providing a *wide* set of example data sets to support a variety of student interests and backgrounds ([C11]). Furthermore, many instructors emphasize the importance of grounding courses in concrete examples and real science questions, rather than disconnected sample data, in order to keep students' interest and to encourage synthesis[11]. Case study data and real open problems seem to work best. Athanasiadis ([C6]) suggests to *"Teach with good examples. Even if they come from a different application area, students can follow it as long it is well explained*

[1] In fact, Mann (2016) even goes one step further and proposes to make a computing student's entire curriculum project-based, namely selecting the student's courses for the degree based on a self-selected sustainability project. Learning would be on-demand as needed by the project.

and documented." McGovern ([C11]) keeps the DS students interested as follows: *"I use a lot of GEO examples when I teach the methods. The biggest challenge is keeping the non-GEO student's interest. Since we live in an area with high severe weather, I use a lot of severe weather examples because everyone is familiar with that around here*." Pennington ([C3]) motivates the ESS students as follows: *"Geo students have difficulty understanding abstract techniques unless they are shown applications, and purposefully reflect on how a technique might be applied in their own work."* Hill ([C7]) gets all students excited through her "policy conference": "*The whole point is to make the students realize they can be part of societal decision making. It worked really well to focus on producing a poster and related research paper, and the two-class policy conference where they role play currently active politicians, business people, scientists, and NGO leaders worked really well."*

Pankratius (video interview) described his strategy for helping students select tools and techniques that match their science goals. In this course, students present their project ideas early in the semester. The instructors then use the list of projects to curate the selection of data science concepts and techniques to be addressed in subsequent classes in order to align with student project interests. Narrowing in on a specific geoscience problem of interest in combination with guidance from instructors in choosing appropriate techniques can help provide a good synthesis of geoscience and DS strategies. Without this guidance, students may be inclined to overly rely on the first few DS techniques they master, regardless of their match to a particular problem or dataset.

Shekhar ([C2]) emphasizes that - in order to learn about different disciplinary values, epistemologies, and philosophies - it is important to "*compare research cultures across IS (conference publication) and GEO (e.g., journal publication). Include papers from both IS and GEO in the reading list"*.

When asked about their wish list for future ID education resources, several researchers described the need for curated datasets for students to work with, particularly for open problems in ESS. There is a need to develop such datasets (Ebert-Uphoff et al., 2017). The appendix contains a list of data sets and other educational material we identified that are already available for immediate use.

Discussion: The Role of Data Science Competencies in Sustainability Science

Wiek et al. (2011) identified systems, anticipatory, strategic, and normative thinking as key competencies in sustainability science, all dependent on interpersonal (collaboration and teamwork) competencies. They provided examples of concepts and methods relevant to each. Expanding on that initial work, Wiek et al. (2015) considered how to operationalize these skills in the context of formal education and expanded the list of concepts and methods related to each. The methods they identified are data analysis and modeling methods, all of which depend on data inputs, and generate data outputs. DS competencies are implicitly required by the methods. DS competencies, then, are foundational in the same way that interpersonal & collaboration competencies are foundational (Figure 3). For example, some of the methods require simulation modeling (e.g. scenario analysis, forecasting, backcasting). Simulation models require data inputs that can be quite complex, and they generate voluminous data outputs that must be analyzed and/or visualized to be understood. In addition, the data inputs and outputs must be managed effectively. The model itself depends on programming and software development. All of these are DS competencies (Hampton et al, 2015). Similarly, methods that require statistical analysis depend on solid knowledge of statistics and the most important data analysis algorithms, e.g. methods for prediction, classification, pattern recognition and causal discovery (Kempler & Mathews, 2017). Equally important is an understanding of the limitations of these methods, in

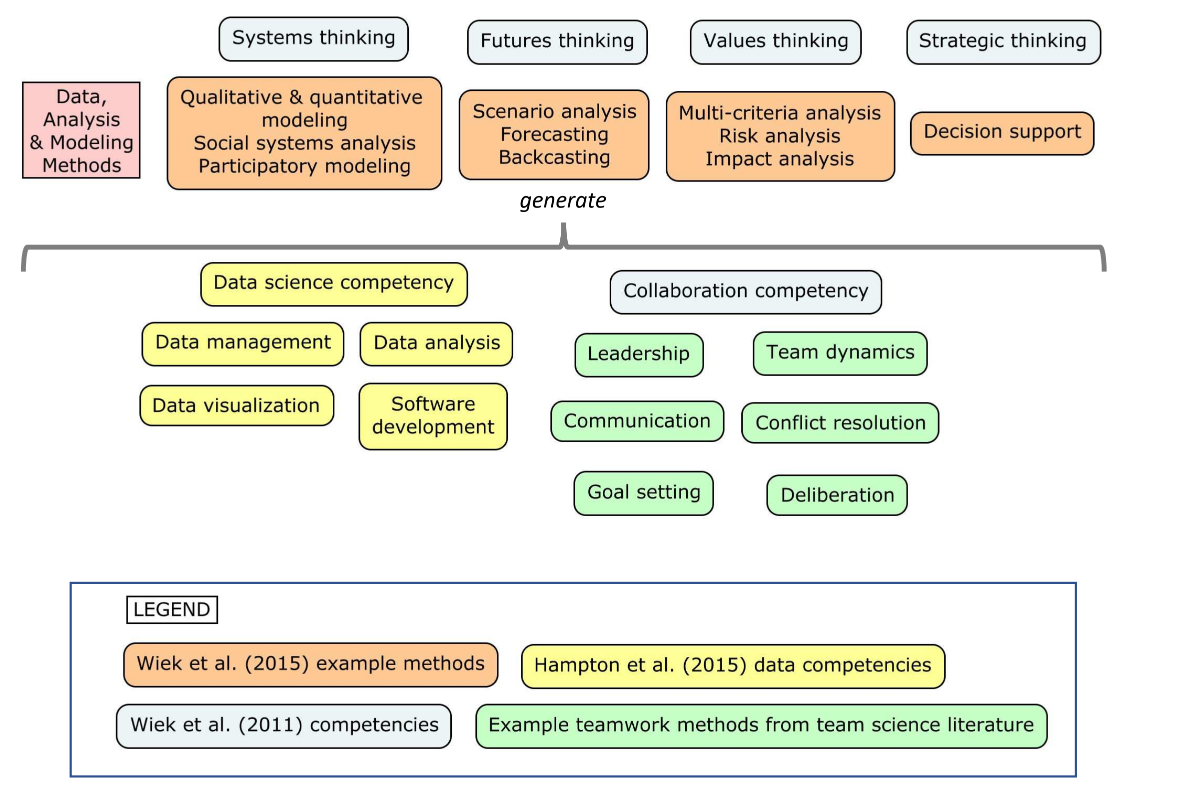

Figure 3 A framework for incorporating data science competencies proposed by Hampton et al. (2015) into sustainability competencies proposed by Wiek et al (2011).

particular how to interpret results from these algorithms and when to question their validity. In practice this has typically been achieved by a person with training in both the content area and data/programming skills. Yet people with both sets of knowledge and skills are few, and as shown above there are few formal programs that target both. As DS progresses it is increasingly difficult to maintain awareness of how the field is changing (Huppenkothen et al., 2018). In addition, making data and models findable, accessible, interoperable, and reusable (FAIR) requires substantial time, effort, and expertise (Wilkinson et al., 2016). Yet this is essential to address the complex sustainability issues confronting society (Creutzig et al., 2019). Not everyone applying data analysis and modeling methods needs to have a full set of DS competencies but they should have enough background in DS to enable collaboration with someone who does. Similarly, those with DS competencies must collaborate with someone who has deep ESS knowledge in order to generate valid models.

Limitations of the research

Although we used a variety of mechanisms to identify relevant programs and courses that could inform our effort, we did not attempt a rigorous, comprehensive search. There exists a large body of literature on ESS and DS education that could provide additional insights. However, our goal was primarily to initiate discussion within the ESS and DS communities regarding the challenges of workforce development in this area and provide examples for consideration of how some faculty are navigating this space.

Conclusion

Despite the fact that data science is rapidly changing virtually all areas of knowledge, and that data science has many important applications in the earth and sustainability sciences, there are very few efforts in formal educational programs attempting to incorporate data science training for earth and sustainability students. New courses and programs are urgently needed at the intersection of sustainability science and data science to prepare the future workforce. We identified several pioneering programs and courses that can serve as guides and inspiration for instructors and institutions. Nevertheless, much more work remains to make this type of education easy to implement, and thus to enable it to become mainstream. Curated datasets and accompanying education resources that support such efforts were identified as one of the major gaps to be addressed.

These efforts must be accompanied by purposeful training in interdisciplinary collaboration, because many sustainability scientists will need to collaborate with data science experts in order to generate and/or apply advanced techniques that are appropriate for the problem being addressed. Interdisciplinary collaborations are known to be fraught with difficulties; these are compounded in collaborations between sustainability and data scientists because of the extreme diversity of these disciplines and the lack of shared background concepts. Hence, progress developing curricula to systematically incorporate data science competencies into sustainability science education must be accompanied by development of more systematic training in interdisciplinary collaboration. New methods for purposeful development of interdisciplinary collaboration skills are beginning to emerge.

There is currently active discussion in the sustainability science education community regarding competencies needed by the future workforce to work effectively in this arena. Because sustainability science problems are complex, they can rarely be addressed effectively by one discipline or one type of knowledge. For this reason, collaboration has been previously identified as a fundamental competency that underlies every other required competency. In the same way, basic competencies for managing, integrating, analyzing, and/or visualizing data, as well as developing models and software, are data science competencies that underlie most (if not all) other competencies. Both sets competencies – collaboration and data science – require additional research to determine the most effective pedagogies for instilling these in students.

Acknowledgements This work was supported in part through NSF grant #1632211 *EarthCube RCN IS-GEO: Intelligent Systems Research to Support Geosciences* (S. Pierce). Support was also provided by NSF awards DGE-1545404 and DBI-1356707 (Pennington), and AGS-3891445978 (Ebert-Uphoff).

References

Barile, S., Orecchini, F., Saviano, M., and Farioli, F. (2018). People, technology, and governance for sustainability: the contribution of systems and cyber-systemic thinking. *Sustainability Science*, pages 1–12.

Belmont Forum (2017). Data skills curricula framework. Technical report. Available at http://www.bfe-inf.org/resource/data-skills-curricula-framework-full-recommendations-report.

Berman, F., Stodden, V., Szalay, A. S., Rutenbar, R., Hailpern, B., Christensen, H., Davidson, S., Estrin, D., Franklin, M., Martonosi, M., and Raghavan, P. (2018). Realizing the potential of data science. *Communications of the ACM*, 61(4):67–72.

Blei, D. M. and Smyth, P. (2017). Science and data science. *Proceedings of the National Academy of Sciences*, 114(33):8689–8692.

Bosque-Prez, N. A., Klos, P. Z., Force, J. E., Waits, L. P., Cleary, K., Rhoades, P., Galbraith, S. M., Brymer, A. L. B., ORourke, M., Eigenbrode, S. D., and others (2016). A pedagogical model for team-based, problem-focused interdisciplinary doctoral education. *BioScience*, page biw042.

Bozeman, B., Fay, D., and Slade, C. P. (2013). Research collaboration in universities and academic entrepreneurship: the-state-of-the-art. *The Journal of Technology Transfer*, 38(1):1–67.

Caputo, F., Buhnova, B., and Walletzk, L. (2018). Investigating the role of smartness for sustain- ability: insights from the smart grid domain. *Sustainability Science*, 13:1299–1309.

Ceri, S. (2018). On the role of statistics in the era of big data: A computer science perspective. Statistics & Probability Letters.

Choi, B. C. and Pak, A. W. (2006). Multidisciplinarity, interdisciplinarity and transdisciplinarity in health research, services, education and policy: 1. definitions, objectives, and evidence of effective- ness. *Clinical and investigative medicine*, 29(6):351.

Corbin, A., & Strauss, A. (1990). Grounded theory research: Procedures, canons, and evaluative criteria. *Qualitative Sociology*, *13*(1), 3–21.

Creutzig, F., Lohrey, S., Bai, X., Baklanov, A., Dawson, R., Dhakal, S., … Walsh, B. (2019). Upscaling urban data science for global climate solutions. *Global Sustainability*, *2*, e2. https://doi.org/10.1017/sus.2018.16

Derrick, E.G., Falk-Krzesinski, H.J., and Roberts, M.R. (2013). Facilitating interdisciplinary research and education: A practical guide. Available at https://www.aaas.org/report/facilitating-interdisciplinary-research-and-education-practical-guide.

Ebert-Uphoff, I. and Deng, Y. (2017). Three steps to successful collaboration with data scientists. EOS, Transactions American Geophysical Union.

Ebert-Uphoff, I., Thompson, D., Demir, I., Gel, Y., Hill, M., Karpatne, A., Guereque, M., Kumar, V., E., C.-C., and Smyth, P. (2017). A vision for the development of benchmarks to bridge geoscience and data science. In *Proceedings of the Seventh International Workshop on Climate Informatics (CI 2017)*.

Eigenbrode, S. D., O'Rourke, M., Wulfhorst, J. D., Althoff, D. M., Goldberg, C. S., Merrill, K., Morse, W., Nielsen-Pincus, M., Stephens, J., Winowiecki, L., and Bosque-Perez, N. A. (2007). Employing philosophical dialogue in collaborative science. *BioScience*, 57(1):55–64.

Faghmous, J. H., Banerjee, A., Shekhar, S., Steinbach, M., Kumar, V., Ganguly, A. R., and Samatova, N. (2014). Theory-guided data science for climate change. *Computer*, 47(11):74–78.

Faghmous, J. H. and Kumar, V. (2014). A big data guide to understanding climate change: The case for theory-guided data science. *Big data*, 2(3):155–163.

Falk-Krzesinski, H. J., Contractor, N., Fiore, S. M., Hall, K. L., Kane, C., Keyton, J., Klein, J. T., Spring, B., Stokols, D., and Trochim, W. (2011). Mapping a research agenda for the science of team science. *Research Evaluation*, 20(2):145–158.

Fiore, S. (2008). Interdisciplinarity as teamwork - how the science of teams can inform team science. Small Group Research, 39(3):251–277.

Fisher, D. H., Bian, Z., and Chen, S. (2016). Incorporating sustainability into computing education. IEEE Intelligent Systems, (5):93–96.

Fox, P. and Hendler, J. (2014). The science of data science. *Big Data*, 2(2):68–70.

Franks, D., Dale, P., Hindmarsh, R., Fellows, C., Buckridge, M., and Cybinski, P. (2007). Interdisciplinary foundations: reflecting on interdisciplinarity and three decades of teaching and research at Griffith University, Australia. *Studies in Higher Education*, 32(2):167–185.

Gibert, K., Horsburgh, J. S., Athanasiadis, I. N., and Holmes, G. (2018). Environmental Data Science. Environmental Modelling & Software, 106:4–12.

Gil, Y. and others (2015). Final workshop report. Workshop on Intelligent and Information Systems for Geosciences.

Gomes, C. P. (2009). Computational sustainability: Computational methods for a sustainable environment, economy, and society. *The Bridge*, 39(4):5–13.

Haider, L. J., Hentati-Sundberg, J., Giusti, M., Goodness, J., Hamann, M., Masterson, V. A., Meacham, M., Merrie, A., Ospina, D., Schill, C., et al. (2018). The undisciplinary journey: early-career perspectives in sustainability science. *Sustainability Science*, 13(1):191–204.

Hall, K.L., Feng, A.X., Moser, R.P., Stokols, D., and Taylor, B.K. (2008). Moving the science of team science forward. *American Journal of Preventive Medicine*, 35(2): S243–S249.

Hall, K. L., Vogel, A. L., Huang, G. C., Serrano, K. J., Rice, E. L., Tsakraklides, S. P., & Fiore, S. M. (2018). The science of team science: A review of the empirical evidence and research gaps on collaboration in science. *American Psychologist*, *73*(4), 532–548. https://doi.org/10.1037/amp0000319

Hampton, S. E., Jones, M. B., Wasser, L. A., Schildhauer, M. P., Supp, S. R., Brun, J., Hernandez, R., Boettiger, C., Collins, S. L., Gross, L. J., et al. (2017). Skills and knowledge for data-intensive environmental research. *Bioscience*, 67(6):546–557.

Hey, T., Tansley, S., Tolle, K. M., et al. (2009). *The fourth paradigm: data-intensive scientific discovery*, volume 1. Microsoft research Redmond, WA.

Hou, C.-Y. (2015). Meeting the needs of data management training: The federation of earth science information partners (esip) data management for scientists short course. *Issues in Science and Technology Librarianship*.

Huppenkothen, D., Arendt, A., Hogg, D. W., Ram, K., VanderPlas, J. T., & Rokem, A. (2018). Hack weeks as a model for data science education and collaboration. *Proceedings of the National Academy of Sciences*, *115*(36), 8872–8877. https://doi.org/10.1073/pnas.1717196115

Hutchins, E. (1995). *Cognition in the Wild*. Cambridge, MA: MIT Press.

Jensenius, A. R. (2012). Disciplinarities: intra, cross, multi, inter, trans. Available at http://www. arj.no/2012/03/12/disciplinarities-2/. Blog.

Karpatne, A., Atluri, G., Faghmous, J. H., Steinbach, M., Banerjee, A., Ganguly, A., Shekhar, S., Samatova, N., and Kumar, V. (2017). Theory-guided data science: A new paradigm for scientific discovery from data. *IEEE Transactions on Knowledge and Data Engineering*, 29(10):2318–2331.

Karpatne, A., Ebert-Uphoff, I., Ravela, S., Babaie, H. A., and Kumar, V. (2018). Machine learning for the geosciences: Challenges and opportunities. *IEEE Transactions on Knowledge and Data Engineering (TKDE)*.

Kempler, S. and Mathews, T. (2017). Earth science data analytics: Definitions, techniques and skills. Data Science Journal, 16(0).

Kitchin, R. (2014). The real-time city? big data and smart urbanism. *GeoJournal*, 79(1):1–14.

Klein, J. T. (2010). A taxonomy of interdisciplinarity. In Frodeman, R., Klein, J. T., and Mitcham, C., editors, *The Oxford Handbook of Interdisciplinarity*, pages 15–30. Oxford University Press.

Klein, J. & Newell, W. (1998). Advancing Interdisciplinary Studies. In W. Newell (Ed.), Interdisciplinarity: Essays from the literature (pp. 3-22). New York: College Board.

Kliskey, A., Alessa, L., Wandersee, S., Williams, P., Trammell, J., Powell, J., Grunblatt, J., and Wipfli, M. (2017). A science of integration: frameworks, processes, and products in a place-based, integrative study. *Sustainability Science*, 12(2):293–303.

Kolb, D. A. (1984). *Experiential Learning: Experience as the Source of Learning and Development*. Englewood Cliffs, NJ: Prentice-Hall.

Lee, C. (2007). Boundary negotiating artifacts: Unbinding the routine of boundary objects and embracing chaos in collaborative work. *Computer Supported Cooperative Work*, *16*, 307–339.

Mann, S. (2016). A rethink for computing education for sustainability. *International Association for Development of the Information Society*.

Mann, S., Muller, L., Davis, J., Roda, C., and Young, A. (2009). Computing and sustainability: evaluating resources for educators. *ACM SIGCSE Bulletin*, 41(4):144–155.

Mann, S., Smith, L., and Muller, L. (2008). Computing education for sustainability. *ACM SIGCSE Bulletin*, 40(4):183–193.

Masse, L. C., Moser, R.P., Stokols, D., Taylor, B.K., Marcus, S.E., Morgan, G.D., Hall, K.L., Croyle, R.T., and Trochim, W.M. (2008). Measuring collaboration and transdisciplinary integration in team science. *American Journal of Preventive Medicine*, 35(2): S151–S160.

Mezirow, J. (1997). Transformative learning: Theory to practice. *New Directions for Adult and Continuing Education*, *1997*(74), 5–12.

Monteleoni, C., Schmidt, G. A., and McQuade, S. (2013). Climate informatics: accelerating discovering in climate science with machine learning. *Computing in Science & Engineering*, 15(5):32–40.

National Academies of Sciences, Engineering, and Medicine (2018). *Envisioning the* Data Science Discipline: The Undergraduate Perspective: Interim Report. The National Academies Press, Washington, DC. Available at https://www.nap.edu/catalog/24886/envisioning-the-data-science-discipline-the-undergraduate-perspective-interim-report.

National Research Council and others (2015). *Enhancing the effectiveness of team science*. National Academies Press. Available at https://www.nap.edu/catalog/19007/enhancing-the-effectiveness-of-team-science.

Nersessian, N. J. (1999). Model-Based Reasoning in Conceptual Change. In L.

Magnani, N. J. Nersessian, & P. Thagard (Eds.), *Model-Based Reasoning in Scientific Discovery* (pp. 5–22). Retrieved from http://link.springer.com/chapter/10.1007/978-1-4615-4813-3_1

Newell, W. H. and Luckie, D. B. (2013). Pedagogy for interdisciplinary habits of the mind. In McCright, A. M. and Eaton, W., editors, *The May 2012 Conference on Interdisciplinary Teaching and Learning, White paper*. East Lansing, MI: Michigan State University. Available at http://lbc. msu.edu/faculty_staff/CITL%20White%20Paper.pdf.

O'Rourke, M., Crowley, S., Eigenbrode, S. D., and Wulfhorst, J. (2013). *Enhancing communication & collaboration in interdisciplinary research*. Sage Publications.

O'Rourke, M., Crowley, S., and Gonnerman, C. (2016). On the nature of cross-disciplinary integration: A philosophical framework. *Studies in History and Philosophy of Science Part C: Studies in History and Philosophy of Biological and Biomedical Sciences*, 56:62–70.

Oskam, I. (2009). T-shaped engineers for interdisciplinary innovation: an attractive perspective for young people as well as a must for innovative organisations. In *37th Annual Conference–Attracting students in Engineering, Rotterdam, The Netherlands*, volume 14.

Pankratius, V., Li, J., Gowanlock, M., Blair, D. M., Rude, C., Herring, T., Lind, F., Erickson, P. J., and Lonsdale, C. (2016). Computer-aided discovery: Toward scientific insight generation with machine support. *IEEE Intelligent Systems*, 31(4):3–10.

Pennington, D. (2008). Cross-disciplinary collaboration and learning. *Ecology and Society*, 13(2):8.

Pennington, D. (2010). The dynamics of material artifacts in collaborative research teams. *Computer Supported Cooperative Work*, *19*(2), 175–199.

Pennington, D. (2011a). Bridging the disciplinary divide: Co-creating research ideas in eScience teams. *Computer Supported Cooperative Work, Special Issue on Embedding eResearch Applications: Project Management and Usability*, 20(3):165–196.

Pennington, D. (2011b). Collaborative, cross-disciplinary learning and co-emergent innovation in informatics teams. *International Journal of Earth System Informatics*, 4(2):55–68.

Pennington, D. (2016). A conceptual model for knowledge integration in interdisciplinary teams: orchestrating individual learning and group processes. *Journal of Environmental Studies and Sciences*, 6(2):300–312.

Pennington, D., Bammer, G., Danielson, A., Gosselin, D., Gouvea, J., Habron, G., Hawthorne, D., Parnell, R., Thompson, K., Vincent, S., and Wei, C. (2016). The EMBeRS project: employing model-based reasoning in socio-environmental synthesis. *Journal of Environmental Studies and Sciences*, 6(2):278–286.

Plale, B., McDonald, R. H., Chandrasekar, K., Kouper, I., Konkiel, S., Hedstrom, M. L., Myers, J., and Kumar, P. (2013). Sead virtual archive: Building a federation of institutional repositories for long-term data preservation in sustainability science. *International Journal of Digital Curation*, 8:172–180.

Repko, A. F. (2011). Interdisciplinary research: Process and theory. Sage, 2nd edition.

Seele, P. (2016). Envisioning the digital sustainability panopticon: a thought experiment of how big data may help advancing sustainability in the digital age. *Sustainability Science*, 11(5):845–854.

Sellars, S., Nguyen, P., Chu, W., Gao, X., Hsu, K.-l., and Sorooshian, S. (2013). Computational earth science: Big data transformed into insight. *Eos, Transactions American Geophysical Union*, 94(32):277–278.

Sellars, S.L. and others (2017). Big Data and the Earth Sciences: Grand Challenges Workshop. Technical report. Available at

http://pacificresearchplatform.org/images/reports/BigDataWorkshop2017_Report_FINAL_082417.pdf

Spelt, E. J., Biemans, H. J., Tobi, H., Luning, P. A., and Mulder, M. (2009). Teaching and learning in interdisciplinary higher education: A systematic review. *Educational Psychology Review*, 21(4):365.

St. John, K., Bitting, K., Cervato, C., Kastens, K., Macdonald, H., McDaris, J., … Teasdale, R. (2019, July 8). An Evolutionary Leap in How We Teach Geosciences. Retrieved July 31, 2019, from Eos website: https://eos.org/opinions/an-evolutionary-leap-in-how-we-teach-geosciences

Star, S., & Griesemer, L. (1989). Institutional ecology, translations and boundary objectsâ”Amateurs and professionals in Berkeleys Museum of Vertebrate Zoology, 1907-39. *Social Studies of Science*, *19*(3), 387–420.

Stokols, D., Misra, S., Moser, R., Hall, K., and Taylor, B. (2008). The ecology of team science: Under- standing contextual influences on transdisciplinary collaboration. *American Journal of Preventative Medicine*, 35(2): S96–S115.

Stone, D. A. (2013). The experience of the tacit in multi- and interdisciplinary collaboration. *Phenomenology and the Cognitive Sciences*, 12(2):289–308.

Szalay, A. and Gray, J. (2006). 2020 computing: Science in an exponential world. *Nature*, 440(7083):413–414.

The World Economic Forum (2018). Harnessing artificial intelligence for the earth. Technical report. Available at http://www3.weforum.org/docs/Harnessing_Artificial_Intelligence_for_the_Earth_report_2018.pdf.

Thompson, J. L. (2009). Building collective communication competence in interdisciplinary research teams. *Journal of Applied Communication Research*, 37(3):278–297.

Tibaut, A. and Zazula, D. (2018). Sustainable management of construction site big visual data. Sustainability Science, 13:1311–1322.

Tullock, G. (2001). A Comment on Daniel Klein's" A Plea to Economists Who Favor Liberty". Eastern Economic Journal, 27(2):203–207.

Virapongse, A., Duerr, R., and Metcalf, E. C. (2018). Knowledge mobilization for community re- silience: perspectives from data, informatics, and information science. *Sustainability Science*, (online first).

Wiek, A., Bernstein, M. J., Foley, R. W., Cohen, M., Forrest, N., Kuzdas, C., … Keeler, L. W. (2015). Operationalising competencies in higher education for sustainable development. In M. Barth, G. Michelsen, M. Rieckmann, & I. Thomas (Eds.), *Handbook of Higher Education for Sustainable Development* (pp. 241–260). London: Routledge.

Wiek, A., Withycombe, L., and Redman, C. L. (2011). Key competencies in sustainability: a reference framework for academic program development. *Sustainability science*, 6(2):203–218.

Wilkinson, M. D., Dumontier, M., Aalbersberg, Ij. J., Appleton, G., Axton, M., Baak, A., … Mons, B. (2016). The FAIR Guiding Principles for scientific data management and stewardship. *Scientific Data*, *3*(1), 160018. https://doi.org/10.1038/sdata.2016.18

Xie, Y., Eftelioglu, E., Ali, R. Y., Tang, X., Li, Y., Doshi, R., and Shekhar, S. (2017). Transdisciplinary Foundations of Geospatial Data Science. *ISPRS International Journal of Geo-Information*, 6(12):395. Yarime, M. (2017). Facilitating data-intensive approaches to innovation for sustainability: opportunities and challenges in building smart cities. *Sustainability Science*, 12(6):881–885.