**Abstract.** Because little has been written about the design of integrated research systems, this article represents an initial effort to rectify this deficiency. It consists of four parts. An introduction to the topic and review of the literature is developed in part one, followed by the presentation of an integrated research system model in part two. An in-depth discussion of the interfaces between the system components is articulated in part three. And finally, the article concludes with some thoughts and best-practice recommendations for research system design.

**Introduction and Literature Review**

The use of educational stories is not new. Harvard University’s business and law schools first introduced case-studies into their curricula some 100 years ago. The case study can be thought of as a special kind of story, narrative designed to build reasoning skills while also imparting content. Herreid (1997) writes, “Cases are stories with a message. They are not simply narratives for entertainment. They are stories to educate” (p. 92). As such, the story has proven to be an effective pedagogical tool.

The practice of data science in the clinical sciences has some unique features that create an ideal environment for a story infused pedagogy. For example, data analysis in the clinical and translational sciences is concrete and embedded in the real world. This stands in contrast to statistics, math, and computer science where the focus is more abstract. In these fields, data science is largely about constructing and fitting models. Clinical data, on the other hand, is often messy and contradictory, abhorring generalizations and abstraction. While abstraction can help to formulate a problem, it is an appreciation for the complexity of clinical research and an appreciation that exceptions can be nearly as important as rules that are required of computational scientists working in this field. A data-driven approach to learning informed by messy data and interesting research questions – the way concepts are presented in data stories – aligns well with the realities of data analysis in precision medicine and public health.

Although data stories share many commonalities with case-studies – content in both, for example, is packaged in a story container or narrative arc – there are important differences as well. With data stories, the focus shifts to data analysis and/or the acquisition of the requisite technical skills to **do** data analysis. This is not the situation with case-studies where data may be present but is rarely the focus.

Interactivity is another feature of data stories that distinguishes them from the traditional case-study. The case-studies featured at the National Center for Case Study Teaching in Science website, for example, are offered in static containers – MS Word, PowerPoint, or Adobe. The data stories created as part of this project, on the other hand, will be developed with open source tools such as [RMarkdown](https://rmarkdown.rstudio.com/) or [Jupyter Notebooks](https://jupyter.org/). With these technologies, students can interactively run blocks of code and receive immediate feedback. Learners can also modify blocks of code to fit their needs and/or analyze similar kinds of data sets. In other words, these tools support dynamic, interactive learning experiences with feedback in real time, making them ideal learning tools for novice and experienced learners alike. Study participants will not need to have any prior programming experience. In addition to a mini-course, participants will be given access to existing introductory R, Python and statistics [MyDataStory](https://github.com/mydatastory) classes.

There is a substantive and growing body of literature that describes how to develop and use case-studies in science teaching. Clyde Herreid (2007, 2012) – the founder of the *National Center for Case Study Teaching in the Sciences* – has been especially active in this space. His articles and edited volumes are indispensable and provide a practical introduction to the art of case-study construction and teaching. The center’s website features some 778 case-studies available to registered faculty.

By contrast, studies that have examined the effectiveness of the case-study method in life science settings are minimal. Even fewer studies have attempted to understand when, how, and why case-studies work from a student perspective. About a dozen articles in science-related journals have reported positive learning outcomes related to the case-study method (Harman et al., 2014; Grunwald & Hartman, 2010; Rybarczyk et al., 2007; Chaplin, 2009; Nair et al., 2013; Wilcox, 1999; Bonney, 2015; Yadav & Beckerman, 2009; Bjorn et al. 2013; White et al., 2009). Only Yadav, Shaver, and Meckl (2010) reported “no significant differences between traditional lecture and case teaching method on students’ conceptual understanding” (p. 55). Even so, they still viewed case-studies in a positive light, given their ability to actively engage students in the learning process. Faculty also appreciate the benefits of case-studies as learning tools. Yadev et al. (2007) conducted a national survey of faculty perceptions of the case-study method and found that a majority reported positive outcomes when using this method. The research team anticipates that the data story instructional method will result in positive results, similar to those just reported.

The question of the motivational efficacy of the data story lies at the heart of the research reported in this article. Educators have long recognized that motivation and engagement are key drivers of student learning. A motivated and engaged student is one who will eventually master the concepts and techniques, absent any significant extenuating circumstances. Achieving either, however, entails some level of emotional involvement on the part of the student. Indeed, the importance of emotion in the learning process has received increasing attention in the literature (Herreid et al. 2014).

The article by Young and Anderson (2010) is especially relevant in this context.

*Significance*

The long-term importance of this proposal lies in two dimensions; a) the development of a clinical data story community, and b) the data story as a new and innovative way to impart scientific concepts and issues to clinical translational science professionals. Writing for *Nature*, Vivien Marx (2013) details the big data challenges facing the life sciences, biology in particular. She writes, “Data mountains and analysis are altering the way science progresses, and breeding biologists who get neither their feet nor their hands wet” (p. 260). Hey, Tansley, and Tolle (2009) argue that science is undergoing a *paradigm shift*. If so, then the larger question of how to educate a generation of data-savvy public health professionals capable of navigating this shift is truly significant.

*Innovation/Potential Impact*

As noted earlier, the data story is a pedagogical innovation that shares many similarities with the case-study method of instruction. The data story, however, has unique properties which distinguish it from similar approaches. These unique properties – a high level of interactivity, for example – creates new ways to engage students and impact learning outcomes.

A data story approach to precision public health education is likely to improve student technical and communication skills. A pedagogical approach grounded in stories also allows CTS professionals to develop and practice emerging skills in the context of topics that are both more engaging and relatable than the abstract examples typically employed in technical education settings. And finally – rather than present technical skills in isolation – the “data story” approach enables learners to understand how these skills can be jointly used to solve a particular problem while also, due to the availability of the auxiliary material, enabling skills to be more thoroughly learned on demand. As well, data stories will lend themselves to being translational vectors. That is, the inherent accessibility of data stories makes them an ideal choice for delivering multimedia prevention and intervention messages to a broader audience, including patients, journalists, and science communication professionals in related fields.

A defining feature of a story-driven approach to data science education is the innovative use of visual and video learning modalities in combination with narrative, data, and code. This is a new and unexplored line of pedagogical inquiry. The College of Pharmacy’s Video Production Services Department will be an ideal partner, assisting the research team in its exploration of innovative ways to communicate complex data science concepts and ideas. The department is a proven innovator, having pioneered the development and use of explanatory and interactive videos across campus.

*Preliminary Studies*

A second research team, led by the PI, has already begun to assess the impact of the data story method of instruction. The team has been meeting bi-weekly for the past six months to advance this initiative and discuss findings. The findings so far have been positive, though anecdotal. For example, the PI on this proposal taught an online class (MCB4934 -Data Storytelling in R) for the Microbiology and Cell Science Department. Student response was overwhelmingly positive. Additionally, data story modules have been tested in the graduate-level course STA6093 (Introduction to Applied Statistics for Agriculture and Life Sciences), taught by Denis Valle, and the undergraduate-level course BSC2891 (Python Programming for Biology), taught by Bryan Kolaczkowski. Both STA6093 and BSC2891 are fully online, and enrollment often exceeds 100 and 200 students per semester for STA6093 and BSC2891, respectively. Both are foundational courses for life-science students at UF that introduce basic data science concepts through in-depth interaction with a variety of datasets.

*Design and Methodology*

The foundational question of this research study asks, “Does the data story method of instruction engage and motivate students to a greater extent than traditional methods of instruction?” The typical data science learning experience today, either in-person or online, adheres to the traditional lecture format. A knowledgeable instructor lectures from the front, often demonstrating the features of a software program written in R or Python. The Carpentries advocate the use of “live coding” as it demystifies the art of programming, frequently giving students the chance to view a knowledgeable instructor making coding mistakes and recovering from them. Data Camp – a popular online data science learning environment – presents content in a similar fashion, with instructional lectures discussing a particular technique or concept, interspersed with interactive coding exercises.

The initial set of research questions to be investigated are listed below. (These questions are directly linked to objective one in the *Specific Aims/Objectives* section – Research the Efficacy of the Data Story as a Precision Public Health Learning Tool.)

1. What features of a data story approach elicit the necessary thinking processes to understand precision public health issues and concepts?
   1. How do we create data stories that appeal to wide and diverse audiences?
   2. Are students motivated to a greater degree when content is delivered in a story?
   3. How do we create data stories that activate foundational reasoning processes?
   4. What defines the experiences of students with the data story approach and how does a student’s academic and cultural background impact that experience?
   5. What features of the learning environment support positive student experiences with the data story instructional method and elicit the necessary thinking processes for understanding data science?
   6. How usable is the data story method of instruction at imparting clinical translational to the research workforce?
   7. What usability factors ought to inform the creation of data story learning experiences?

The research team plans to employ a mixed methods approach to address these questions and others which arise during this preliminary investigation (Creswell, 2014). The information gleaned from this work will also be supplemented by a pre-and-post survey. Research methods inspired by phenomenology will be used to investigate the cognitive psychological dimensions of the data story instructional method (questions 1a – 1d). The research executed in response to these questions will be informed by the methods of classical phenomenology (Van Manen, 1990; 2016), phenomenography (Marton, 1986), and protocol analysis (Ericsson & Simon, 1993; Fonteyn, Kuipers, & Grobe, 1993).

[Usability testing](https://en.wikipedia.org/wiki/Usability_testing) data will be collected in order to evaluate the Data Story Method of Instruction from the perspective of study participants (questions 1f & 1g). Data collection will be informed by a mental model theoretical framework (Jonassen, 2001) and implemented using the DECIDE framework (Rogers, Sharp & Preece, 2011). The DECIDE framework involves a six-step process to evaluate a prototype's capacity to address its design and development goals. The goals for usability testing will include: a) participant’s perspective on usefulness, satisfaction and ease-of use, and b) usability issues (e.g., problematic behaviors, frustration, misinterpretations). We will use pre and post surveys and interviews to analyze study participant expectations and how these goals and expectations are met as a result of the data story way of instruction.

*Research Protocol*

|  |
| --- |
| *Recruitment of Participant(s)*  A sample of ~40 student participants with cohorts at UF, FSU, and UF Jacksonville will be recruited from a target population of clinical research professionals and graduate students at all three locations. The sample will be participants interested in gaining competencies in fundamental Data Sciences as applied to managing, coordinating or conducting translational research in data rich areas of precision health.  Participants will be recruited from the population of research professionals identified through CTSI networks and training programs. Recruitment messages will be disseminated via relevant list servers as well as direct approaches from key personnel. |
| *Informed Consent*  Respondents will be sent an email with a request to participate in the study. Details of the study with potential benefits and risks will be provided in the recruitment email. Follow-up consent will be completed online via a Qualtrics survey. |
| *Research Methodology*  A small sample (~40) of individuals will be recruited to participate in a CTSI translational data science mini-course which features a PPH data story as the initial learning experience. Participants will be asked to fill out pre-and-post surveys to acquire demographic background data and assess the usability of the data story module and selected features of the mini-course. Additionally, a small subset (~5) participants will be invited to an interview, either individually or as part of a focus group. Interviews will take between 45 minutes and one hour. |
| *Data Collection*  The range of questions pursued in the interviews and surveys will be limited to those of demography and usability. Data will be collected from the pre-and-post surveys, interviews, and focus groups. The focus groups and individual interviews will be audio recorded with the consent of each participant. A professional transcriptionist – not associated with the study – will transcribe each interview. Data will be scanned and stored electronically in applications such as Microsoft OneNote or Dropbox. |
| *Data Analysis*  Descriptive statistics will be used to analyze the study’s survey data. The transcribed interviews and focus group conversations will be analyzed using appropriate qualitative and phenomenological methods. Open science best-practices will be employed in the sharing of the analysis and findings. |
| *Research Locations*  The mini-course and surveys will be conducted online. In-person interviews may be supplemented with Zoom conference calls. The primary research location will be UF’s Gainesville campus. |
| *Benefits and Risks*  The CTSI translational data science online mini-course will be offered as a “CTSI Certificate of Completion” at no charge to study participants. UF participants who complete the training will have this recorded in their HR training records. It is anticipated the time commitment will be approximately 15 hours over the course of 5 weeks. There are no anticipated risks to participating in this research. |

*Quotes*

We love the tangible, the confirmation, the palpable, the real, the visible, the concrete, the known, the seen, the vivid, the visual, the social, the embedded, the emotionally laden … Most of all we favor *the narrated.*

Alas, we are not manufactured, in our current edition of the human race, to understand abstract matters – we need context (Taleb, 2007, p. 132).

If I have a set of 200 random variables, completely unrelated to each other, then it would be near impossible not to find in it a high correlation of sorts, say 30 percent, but that is entirely spurious (Taleb, 2012, p. 418).

**References**

Bonney, K. M. (2015). Case study teaching method improves student performance and perceptions of learning gains. *Journal of Microbiology & Biology Education 16*(1), *21 – 28.* doi: 10.1128/jmbe.v16il.846

Bjorn, H. K., … White, C. D. (2013). Student performance in a multimedia case-study environment. *Journal of Science Education Technology 22,* 215 – 225. doi: 10.1007/s10956-012-9387-7

Causey, M. (2019). Clinical trial industry demands increasingly nimble workforce. Retrieved from <https://acrpnet.org/2019/04/15/clinical-trial-industry-demands-increasingly-nimble-workforce/?utm_campaign=News&utm_medium=email&utm_source=internal&utm_content=CRBeat-04182019&utm_term=text-ReadMore&_zs=6HvCX&_zl=Ee4b1>

Chaplin, S. (2009). Assessment of the impact of case studies on student learning gains in an introductory biology course. *Journal of College Science Teaching 39*(1), 72 – 79.

Creswell, J. W. (2014). *Research Design: Qualitative, quantitative, and mixed methods approaches.* Thousand Oaks, CA: Sage.

De Veaux, R. D., … Ye, P. (2017). Curriculum guidelines for undergraduate programs in data science. *Annual Review of Statistics and Its Applications.* doi: 10.1146/annurev-statistics-060116-053930

Donovan, M. S., Bransford, J. D., & Pellegrino, J. W. (Eds.). (1999). *How people learn: Bridging research and practice.* Washington, DC: National Academy of Sciences Press. Retrieved from <http://nap.edu/9457>

Ericsson, K. A., & Simon, H. A. (1993). *Protocol analysis: Verbal reports as data*. Cambridge, MA: MIT Press.

Fonteyn, M. E., Kuipers, B, & Grobe, S. J. (1993). A description of think-aloud method and protocol analysis. *Qualitative Health Research 3*(4), 430 – 441.

Grunwald, S, & Hartman, A. (2010). A case-based approach improves science students’ experimental variable identification skills. *Journal of College Science Teaching 39*(3),28 – 33.

Harmon, T., … Babatunde, O. T. (2014). Case-based learning facilitates critical thinking in undergraduate nutrition education: Students describe the big picture. *Journal of the Academy of Nutrition and Dietetics 115*(3), 378 – 388.

Hey, A. J., Tansley, S., & Tolle, K. (2009). *The fourth paradigm: Data-intensive scientific discovery*. Redmond, WA: Microsoft Research.

Herreid, C. F. (1997). What is a case? Bringing to science education the established teaching tool of law and medicine. *Journal of College Science Teaching 27*(2), 92 – 94.

Herreid, C. F. (2007). *Start with a story: The case study method of teaching college science.* Arlington, VA: NSTA Press.

Herreid, C. F. (2012). *Science stories: Using case studies to teach critical thinking.* Arlington, VA: NSTA Press.

Herreid, C. F., … Ribbens, E. (2014). Emotion, engagement, and case studies. *Journal of College Science Teaching 44*(1), 86 – 95.

Jonassen, D. H. (2001). Operationalizing mental models: Strategies for assessing mental models

to support meaningful learning and design- Supportive learning environments. Paper presented at the CSCL '95 The First International Conference on Computer Support for Collaborative Learning, Bloomington, IN. Retrieved from <http://www.cica.indiana.edu/cscl95/jonassen.html>

Marton, F. (1986). Phenomenography: A research approach to investigating different

understandings of reality. Journal of Thought 21(3). Retrieved from

<https://www.jstor.org/stable/42589189>

Marx, V. (2013). The big challenges of big data. *Nature 498*(7453), 255 – 260.

Nair, S. P., … Shah, G. V. (2013). Case based learning: A method for better understanding of biochemistry in medical students. *Journal of Clinical and Diagnostic Research 7*(8), 1576 – 1578. doi: 10.7860/JCDR/2013/5795.3212

National Academies of Sciences, Engineering, and Medicine. (2018). *Data science for undergraduates: Opportunities and options.* Washington, DC: Committee on envisioning the data science discipline.

Olson, R. (2015). *Houston, we have a narrative.* Chicago, IL: University of Chicago

Press.

Rogers, Y., Sharp, H., & Preece, J. (2011). *Interaction Design: Beyond Human – Computer*

*Interaction*: Wiley.

Rybarczyk, B. J., … Wilkins, H. (2007). A case-based approach increases student learning

outcomes and comprehension of cellular respiration concepts. *Biochemistry and*

*Molecular Biology Education 35*(3), 181 – 186.

Schoenfeld, A. (1985). Making sense of ‘Out Loud’ problem-solving protocols. The

Journal of Mathematical Behavior 4, 171-191.

Taleb, N. N. (2007). *The black swan: The impact of the highly improbable.* New York, NY: Random House.

Taleb, N. N. (2012). *Antifragile: Things that gain from disorder.* New York, NY: Random House.

Van Manen, M. (1990).  *Researching lived experience: Human science for an action*

*sensitive pedagogy*. New York: State University of New York Press.

Van Manen, M. (2016).  *Phenomenology of practice: Meaning-giving methods in*

*phenomenological research and writing*. Abingdon: Routledge.

Wilcox, K. J. (1999). The case method in introductory anatomy & physiology: Using the

news. *The American Biology Teacher 61*(9), 668 – 671.

White, T. K., … Hayes, E. (2009). The use of interrupted case studies to enhance critical

thinking skills in biology. *Journal of Microbiology & Biology Education 10*(1),25

– 31.

Yadav, A. & Beckerman, J. L. (2009). Implementing case studies in a plant pathology

course: Impact of student learning and engagement. *Journal of Natural Resource*

*Life Science Education 38,* 50 – 55.

Yadav, A., … Herreid, C. F. (2007). Teaching science with case studies: A national survey

of faculty perceptions of the benefits and challenges of using cases. *Journal of*

*College Science Teaching 37*(1), 34 – 38.

Yadav, A., Shaver, G. M., & Meckl, P. (2010). Lessons learned: Implementing the case

teaching method in a mechanical engineering course. *Journal of Engineering*

*Education Research 99*(1), 55 – 69.

Young, L., & Anderson, R (2010). The use of personal narrative in classroom case study

analysis to improve long-term knowledge retention and cultivate professional

qualities in allied health students. *Journal of Microbiology and Biology Education,*

*11,* 107 – 112.