Pedagogical and Future Implications for the Training of Data Translators

Boba Samuels, Donna Kotsopoulos, and Douglas G. Woolford

Abstract Sharing data stories is an exciting opportunity for integrating data into the workplace. The examples provided in the chapters of this book reveal three imperatives for training data translators: 1) interdisciplinarity, 2) a knowledge exchange framework (KEF), and 3) language calibration. We propose that a focus on developing students' skills in these areas must be coupled with data science training if we are to prepare effective data translators. Instructors must look to adjust not just their methods, but the philosophy behind their instruction to train students for the next generation of data translation jobs.

The sharing of data science stories within and beyond single disciplines presents an exciting opportunity that is increasingly recognized by data users within academia, workplaces of all types, and the public realm. In the chapters of this book, we have gathered a variety of examples demonstrating how such data stories can be created and taught to new generations of data scientists and, most especially, those who aspire to using the wealth of data available yet who lack as strong a background in data science and analytics. It is this group—those we call data translators—who navigate the shifting boundaries between disciplinary expertise and data science, interpreting data in meaningful and accurate ways for others.

How can we teach new generations of data translators to navigate these boundaries? The case studies in this book suggest three factors need attention. The first is that to develop effective data translators we must acknowledge interdisciplinarity as

B. Samuels (✉)
Faculty of Kinesiology and Physical Education, University of Toronto, Toronto, Ontario, Canada
e-mail: boba.samuels@utoronto.ca

D. Kotsopoulos
Faculty of Education, University of Western Ontario, London, Ontario, Canada
e-mail: dkotsopo@uwo.ca

D. G. Woolford
Department of Statistical and Actuarial Sciences, University of Western Ontario, London, Ontario, Canada
e-mail: dwoolfor@uwo.ca

D. G. Woolford et al. (eds.), *Applied Data Science*, Studies in Big Data 125,
https://doi.org/10.1007/978-3-031-29937-7_13

a guiding force in data science. In other words, data science is not just for data scientists. As this book demonstrates, scholars and stakeholders across many different fields are teaching and learning how to tell data science stories for their audiences. However, this ubiquity in the use of data science and analytics methods also runs the risk of incorrect analyses leading to incorrect conclusions—just because someone has data that can be run through a data modelling method, doesn't mean that it is being done correctly. This was highlighted by Li in Chap. 10 through an analysis of binary (correct/incorrect) response data that demonstrated the need to understand the underlying assumptions of data modelling methods and to select an appropriate technique accordingly. Hence, more attention to the ways in which such teaching occurs across disciplines, and particularly when such attempts are multidisciplinary, is needed. Drawing on the expertise and experiences of different stakeholders in this endeavour is good practice. The expertise required to generate big data, for instance, does not lead directly to the expertise needed to use this data. Even with so much data available, only 18% of companies can use the data they generate effectively [1]. More is needed, therefore, than technical skills in data science to effectively use data to address real discipline-specific or workplace issues.

In Schanzer et al.'s Chap. 3 description of public health policy work, multidisciplinary committees are used to shape the creation of policy guidelines, demonstrating that approaches that include multiple different perspectives can be effective. They suggest that teaching using class discussion to identify multiple interpretations of guidelines can raise students' familiarity with such guidelines as well as the critical review process. Such knowledge and skills are necessary in understanding how data analysis is used to support creation of the guidelines. Creating opportunities for interaction between disciplinary experts and creating classroom models that replicate the sharing of different perspectives are promising approaches for encouraging the critical review of data analysis necessary for ensuring high quality data interpretations.

The second factor necessary for teaching people to become effective data translators is to identify a framework for knowledge exchange (KE) between stakeholders. In discussion of KE in the wildland fire management context, McFayden et al.'s Chap. 12 describes the need for intentional engagement in the sharing of knowledge and proposes that a KE framework can facilitate such iterative engagement between researchers and practitioners. In their words, "Effective KE in fire management helps ensure that real-world problems are understood by researchers, the research is relevant, and the results are integrated into fire management practices." They describe how an active learning approach in the classroom can facilitate use of such a KE framework by students who may go on to become data translators. In this approach, which is facilitated by knowledge brokers, students practice the behaviours and skills relevant to address a typical real-world problem by actively engaging with multidisciplinary researchers and practitioners rather than focusing on the technical requirements of data analysis.

In their example, McFayden et al. describe a scenario in which students work in teams with an external "client" with whom they meet to address the client's problem. The class instructor acts to facilitate communication between the students and client by brokering knowledge between the stakeholders, working with the client to guide students through the KE framework. Students take on different roles in their team depending on their background and technical skill level. Various modes of communication are practiced including email exchanges, phone/video meetings, progress reports, presentations, and a final report; peer feedback is also used to encourage the sharing of multiple perspectives and develop the ability to provide constructive criticism. Such an active learning approach provides students with an opportunity to not only practice essential skills necessary for data analysis, but also to experience how a KE framework can guide them to integrate such skills with behaviours conducive to developing mutually beneficial relationships with clients and peers. Identifying and applying such a KE framework is thus likely to be productive for data translators.

Teimourzadeh and Kakavand in Chap. 7 present a staged process for students to learn how to define and address a business problem using data. The "framework" for this staged process is the data analytics workflow with students using two Tableau apps to develop data literacy and data translation skills. The authors describe an "experiential" learning approach in which students practice—as in the active learning approach described by McFayden et al. –the skills necessary for real world use. Instructors are provided with a step-by-step plan that focuses on helping students use freely available software apps to move through six stages of a data analytics workflow, from identification of a business problem, through data collection, processing, analysis, and visualization, to data translation. This approach may be particularly helpful for students with little background in data analysis because it presents them with a framework for their activities that they can apply across a range of possible contexts as well as experience in applying this framework.

Finally, the third factor identified in these chapters' case studies is language calibration, namely the attention to language in flexible ways to carry out effective data translation between participants. Zwiers and Zhang's Chap. 1 discusses the calibration of language related to climate change by an intergovernmental organization, and shows how such development of consistency in use of calibrated terms enables confidence in users and an ability to demonstrate progress over time. They describe, for instance, how specifying the probability related to various terms (e.g., *virtually certain*, *likely*, *very unlikely*) has enabled stronger claims to be stated about scientific conclusions. Most significantly, such attention to clear communication through development of consistent, calibrated terminology provides policy makers and scientists a tool to share information on climate change with confidence that uncertainties in meaning are minimized. The pedagogical implications of this work include drawing students' attention to the value that data translators must place on precise language and the internal values of a discipline or organization. In this case, the need for climate scientists and organizations to avoid confusion and emphasize consensus is necessary to avoid attacks on their credibility in the contentious context of climate change debates.

Falconer takes a wider view on language in Chap. 11, focusing on identifying the rhetorical strategies used when data created in one context is taken up and transformed for use in another domain. By distinguishing between recontextualizing knowledge and data translation, with the former describing a process of transforming knowledge or data from one form to another, while the latter refers to making knowledge accessible (i.e., translated) for different audiences for different purposes, Falconer describes how multidisciplinary and multisectoral panels transform scientific data for use by governmental departments for the purposes of policy development. Falconer identifies strategies including the visual representation of people, numerical representations, and written explanations as some of the means by which data is transformed for later use. In the classroom, these strategies can be identified, analysed, and taught to students who envision themselves in roles doing such "boundary work."

These three factors—interdisciplinarity, a knowledge exchange framework, and language calibration—necessary for effectively teaching people to become data translators reflect issues familiar to writing instructors working to teach students how to write in the disciplines. A well explored concern, for instance, is how to enculturate novice students to a discipline so they can increasingly participate and be seen as legitimate actors within that discipline. For example, how can we teach students to think and write like a scientist? For students new to a discipline, the conventions, genres, language, and values of that discipline are largely unknown and invisible. Making such disciplinary elements visible and enabling practice to use these elements incrementally and appropriately allows students to begin identifying as disciplinary insiders, a process known as legitimate peripheral participation [8]. This process is relevant not only to novice students, but to those who are established within disciplines and who now must interact with and adjust to other disciplines to take advantage of the expanded opportunities available through interdisciplinarity.

This discussion of conventions, genres, values, and participation in a disciplinary community is central to the issue of how to become a data translator. Being a data translator requires that individuals participate and are perceived as legitimate actors within that discipline, not as outsiders. Providing students with experience conducting data analyses using authentic data and then translating that data for use within a specific discipline to address a discipline-specific question teaches students about the conventions that must be followed, the genres that are possible, and the values that require acknowledgment. Students thus learn what it means to be a disciplinary insider and are therefore able to communicate appropriately with disciplinary experts. It is this ability to communicate in discipline-appropriate ways that makes them translators.

Translation is not simply about participants speaking different "languages"—one speaking with technical terms from data science and analytics and one speaking with discipline-specific terminology. Rather, when complex problems intersect with large and complex data structures, it is likely that neither the background knowledge, skills, epistemological values, nor vocabulary of the various stakeholders are aligned, with the result that communication even in a common language would be challenging. Data translators therefore bring together an immense variety of skills crucial to the success of an organization using data, including their data analytics skills, technical fluency, communication skills, project management, and an intimate knowledge of how the data can address organizational barriers to advance the companies' or organizations' interests [6].

Just as expert translators of poetry or novels from one language to another are highly valued for their insights into both languages and both cultures, so too are data translators increasingly valued for their ability to bridge differences between data science and other disciplines or institutions. The specific contributions of data translators have been noted, for instance, in the workplace [7]. For example, Marr (2018), writing for Forbes, went so far as to say that due to their ability to distill complex information to make it understandable to decision makers, employers should hire data translators rather than data scientists. Data translators, in other words, are those who can bridge the technical-stakeholder gap [9]. The hiring of those who can tell stories with data may thus make up the next big wave of hiring [5].

While it is too soon to tell if such predictions prove accurate in the workplace, in the academy there are increasing signs that the need to disseminate information to non-experts is well established and growing. Students and scholars are, for instance, increasingly required to write plain language abstracts for research grant applications and publications, with such translation work valued as an integral component of the research process. Though such tasks may not involve complex data structures or data science, they demonstrate the need to learn how to tell accurate and engaging stories to those without expert knowledge, making knowledge interpretable. In the workplace, data translators help to bridge the "interpretation gap", providing the insights from large data sets to make this information actionable to less-technical workers [4]. While the utility of data translators has been noted in STEM fields such as sustainable development [2] and biomedicine, much of the focus on data translation in academic disciplines is on translating data in underlying models rather than on the training of individuals to be data translators [3].

In this collection, we have shared a number of data science stories to describe early approaches to the pedagogy of data translation. Many other stories exist across a multitude of disciplines and fields, stories of instruction undertaken perhaps tentatively at first, but then with increasing confidence. These stories may provide models for future instruction. By bringing together experts who can demonstrate how such instruction is possible, we hope to encourage other instructors to share their approaches, with the goal of creating a robust and flexible pedagogy for future data translators.

References

1. Ariker, M., Breuer, P., McGuire, T.: How to get the most from big data. McKinsey Inst. (2014). https://www.mckinsey.com/capabilities/mckinsey-digital/our-insights/how-to-get-the-most-from-big-data
2. Auad, G., Fath, B.D.: Towards a flourishing blue economy: Identifying obstacles and pathways for its sustainable development. Curr. Res. Environ. Sustain. **4**, 100193 (2022)
3. Biomedical Data Translator Consortium: The Biomedical Data Translator program: conception, culture, and community. Clin. Transl. Sci. **12**(2), 91 (2019)

4. Brady, C. Forde, M., Chadwick, S.: Why your company needs data translators. MIT Sloan Manag. Review. (2016). https://sloanreview.mit.edu/article/why-your-company-needs-data-translators/
5. Fisher, A.: Now hiring: People who can translate data into stories and actions. Fortune. (2019). https://fortune.com/2019/10/12/human-intelligence-data-translation/
6. Henke, N., Levine, J., McInerney, P.: You don't have to be a data scientist to fill this must-have analytics role. Harv. Bus. Review. (2018). https://hbr.org/2018/02/you-dont-have-to-be-a-data-scientist-to-fill-this-must-have-analytics-role
7. Kenney, A.: 6 steps to become a finance data 'translator'. Financ. Management. (2021). https://www.fm-magazine.com/news/2021/oct/6-steps-become-finance-data-translator.html
8. Lave, J., Wenger, E.: Situated learning: legitimate peripheral participation. Camb. Univ. Press., Cambridge, UK (1991)
9. Maynard-Atem, L., Ludford, B.: The rise of the data translator. Impact **1**, 12–14 (2020)