CAN THERE BE A SPATIAL DATA SCIENCE ETHICS?

David DiBiase* and Anthony C. Robinson**

John A. Dutton e-Education Institute and Department of Geography, Penn State University, University Park PA, 16802, USA
* dibiase@psu.edu; ** arobinson@psu.edu

Keywords: ethics, spatial data science, education, research integrity

> What cartographers most earnestly seek is probably not so much a theoretical as a practical ethics, a set of principles that can be used to clarify moral disagreements or conflicts with the goal of resolving them.
> – J.B. Harley, 1991

Introduction

Luc Anselin, a Fellow of the University Consortium for Geographic Information Science, observed a few years ago that "GIScience [is] morphing into spatial data science" (Anselin 2017). In 2021, Alex Singleton and Daniel Arribas-Bel published "Geographic Data Science," in which they argue that Geographers have a lot to learn from Data Scientists, and vice versa. They go on to assert that an integration of Geography and Data Science would be beneficial in "sustaining the relevance of Geography and subdisciplinary approaches [including Geographic Information Science, or GISc] within a rapidly changing sociotechnological landscape" (p. 71). Evidence that higher education is paying attention to these arguments includes new undergraduate and graduate degrees in Spatial or Geographic Data Science or Analytics established at least eleven major universities in the U.S. and the U.K. by 2022.

Singleton and Arribas-Bel (2021) emphasize that "Geographic Data Science would closely align with core critical and ethical principles … that have been advanced within Geography and, in particular, the subdisciplinary field of GISc" (p. 70). The following year, Nelson, Goodchild and Wright (2022) issued a clarion call for "Accelerating Ethics, Empathy, and Equity in Geographic Information Science." Accordingly, it seems worthwhile to assess the extent to which ethical issues are included in Geographic Data Science curricula. This paper has a dual purpose: first, to share Penn State rationale and strategy for infusing ethics into its Spatial Data Science master's degree, and second, to explore a methodological path forward for evaluating spatial data science curricula vis-à-vis their support for ethics engagement.

Background

Formal training in research ethics has been mandatory at Penn State, and most other higher education institutions, since The America COMPETES Act was signed into law in 2007 (H.R.2272 - 110th Congress (2007-2008). Section 7009 of the legislation, which took effect in 2010, requires any student who receives NSF funds to obtain training in the Responsible Conduct of Research (RCR). Institutions are responsible for determining both the content and delivery method of their training. Many prescribe generic, university-provided online training materials to satisfy this requirement. At Penn State, each academic department and program proposes content and delivery of its training to the University's Scholarship and Research Integrity office, which reviews and approves trainings, and monitors compliance. Students pursuing the Master of GIS or Master of Science in Spatial Data Science degrees comply by successfully completing either a no-cost, 12-hour online workshop on "Responsible Scholarship and Professional Practice," or the 3-credit graduate seminar GEOG 581: Spatial Data Science Ethics. The objectives and content of both the workshop and the seminar were informed by our participation in the GIS Professional Ethics project (DiBiase 2017).

Both compliance and ethics are concerned with standards of behaviour, but compliance is concerned only with minimum standards. Since the rise of GIS' critics in the 1990s (Schuurman 2000), it's hard to imagine a university course, certificate, or degree program that doesn't consider the ethical implications of geospatial technologies and methods. There are numerous ways to include ethics in curricula: extracurricular events, independent studies, guest lectures, workshops and even entire dedicated courses. While many educators might idealize a "pervasive" approach in which ethical issues are infused in every course across the curriculum, "the pervasive approach has, however, proved hard to implement" (Davis 1999, p. 116). The persistent challenge of implementing pervasive approaches to ethics education is evident in Oliver and McNeil's (2019) analysis of undergraduate Data Science course descriptions.

Survey

Oliver and McNeil examined major requirements and corresponding course descriptions to assess the alignment of 18 undergraduate Data Science programs with two prominent frameworks for Data Science education. Their assessment led authors to conclude that "current data science undergraduate programs … may not deliver sufficient context in terms of … ethics considerations necessary for appropriate data science applications" (Oliver and McNeil, 2019). Their survey methods involved developing rubrics based on the two frameworks to score published course descriptions on a four-point scale using content analysis. One framework is the National Academies of Sciences, Engineering and Medicine's Data Science for Undergraduates (NASEM 2018). The NASEM identifies 10 "key concepts involved in developing data acumen," including Mathematical foundations, Computational foundations, Statistical foundations, Data management and curation, Data description and visualization, Data modeling and assessment, Workflow and reproducibility, Communication and teamwork, Domain-specific considerations, and Ethical problem solving. Oliver and McNeil evaluated probable learning outcomes of each of the ten key concepts using the following scoring rubric:

> 4: Student [graduate] is well-versed in the area, with at least one required course covering the area;
> 3: Student has some knowledge of the area, although likely only at rudimentary level;
> 2: Student may have been exposed to this area through an optional course satisfying a requirement;
> 1: No expectation that the student is familiar in this area.

Figure 1 summarizes Oliver and McNeil's evaluation of the 18 data science degree programs across ten key NASEM concept areas. Notably, mean scores among programs were lowest on the "Ethical problem solving" area.

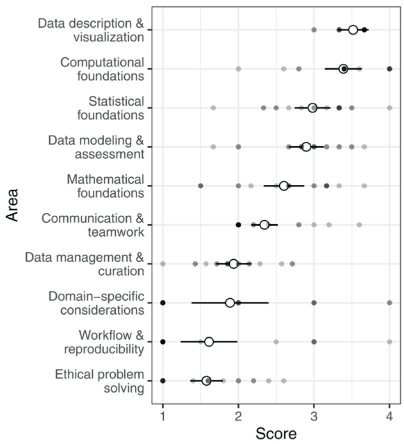

Figure 1: Undergraduate Data Science programs' scores on the NASEM framework. Open circles show the average score for each area across all programs. Filled circles show mean scores in each area of the framework for individual programs. Oliver & McNeil (2019), DOI: 10.7717/peerj-cs.441/fig-1

Oliver and McNeil's scoring rubric for the "Ethical problem solving" area appears in Figure 2. We propose to apply this rubric to assess the possible outcomes of ethics education in at least a dozen Spatial/Geographic Data Science masters programs. As a first step, we used part of Oliver & McNeil's rubric to evaluate the likelihood that graduates of Penn State's MGIS or MS-SDS programs would be knowledgeable about ethics as proposed in the NASEM framework.

10. Ethical Problem Solving

- **Ethical precepts for data science and codes of conduct**
  4: Required course(s) covering ethical precepts for data science and codes of conduct.
  3: Required course(s) introducing ethical precepts for data science and codes of conduct, but not at the level necessary to receive a 4.
  2: Optional course covering ethical precepts for data science and codes of conduct.
  1: No course covering material.
- **Privacy and confidentiality**
  4: Required course(s) covering privacy, confidentiality, or ethics of data.
  3: Required course(s) introduced privacy, confidentiality, or ethics of data, but not at the level necessary to receive a 4.
  2: Optional course covering privacy, confidentiality, or ethics of data.
  1: No course covering material.
- **Responsible conduct of research**
  4: Required course description specifically mentioning responsible conduct of research or something similar.
  3: Required course description mentioning responsible conduct of research or something similar, but not at the level necessary to receive a 4.
  2: Optional course description specifically mentioning responsible conduct of research or something similar.
  1: No course covering material.
- **Ability to identify "junk" science**
  4: Required course description specifically mentioning junk science or pseudoscience.
  3: Required course description including language like "critically evaluating the quality of prior published work."
  2: Required course description explicitly mentioning the scientific method; or optional course description including language like "critically evaluating the quality of prior published work."
  1: No course covering material.
- **Ability to detect algorithmic bias**
  4: Required course description explicitly mentioning algorithmic bias, ethics of algorithm development/application, or something of that ilk.
  2: Optional course description explicitly mentioning algorithmic bias, ethics of algorithm development/application, or something of that ilk.
  1: No course covering material.

Figure 2: Portion of rubric used by Oliver & MacNeil (2019) to evaluate the alignment of data science curricula with the NAESM's (2018) recommended areas of emphasis for undergraduate Data Science programs in higher education. Shown here are the criteria for evaluating academic programs' emphasis on Ethical Problem Solving.

We examined 41 course descriptions published at the Online Geospatial Education programs website (https://geospatial.psu.edu/the-experience/course-calendar). Descriptions of six graduate courses and one workshop mentioned Ethics topics explicitly. The ethics workshop or GEOG 581 seminar that are required for all MGIS and MS-SDS students address Ethical Precepts, Privacy and Confidentiality, and Responsible Conduct of Research specifically. GEOG 583: Geospatial System Analysis and Design – which is also required for all MGIS and MS-SDS students – addresses "the ability to identify 'junk' science" at least to the extent of "including language like 'critically evaluating the quality of prior published work.'" None of the 41 course syllabi mentioned "the ability to detect algorithmic bias." Our self-assessment of ethics in Penn State's MGIS and MS-SDS curricula are as follows:

Ethical precepts for [spatial] data science and codes of conduct: 3 (not 4, because the Ethics workshop most students complete is non-credit, and just 12 hours in length.

Privacy and confidentiality: 3 (not 4, because the Ethics workshop most students complete is non-credit, and just 12 hours in length.

Responsible conduct of research: 3 (not 4, because the Ethics workshop most students complete is non-credit, and just 12 hours in length.

Ability to identify "junk" science or pseudoscience: 3 (GEOG 583)

Ability to detect algorithmic bias: 1

Based on course descriptions and titles alone, we do not find evidence that ethics education is "pervasive" in Penn State's Online Geospatial Education programs. However, these preliminary findings do suggest that MGIS and MS-SDS graduates are likely to have some working knowledge of key ethical issues, including ethical precepts and codes of conduct, privacy and confidentiality, the Responsible Conduct of Research, and the ability to identify "junk" science or pseudoscience.

Discussion and Conclusion

One limitation of Oliver and McNeil's approach is that the absence of explicit mentions of ethical issues in course descriptions does not mean that such issues are absent from course content. Indeed, many course authors and instructors are likely to argue that ethical issues are considered in courses on geospatial data acquisition and positioning, data analysis, modelling, and visualization, and software and app development – despite what official course descriptions may say or not say.

Replicating Oliver and McNeil's assessment in the emerging spatial data science domain will be worthwhile, we believe, if only to incentivize Spatial Data Science instructors and programs to foreground their contributions to ethics education. Furthermore, following Singleton and Arribas-Bel's argument, encouraging Geographic/Spatial Data Science educators of all specializations to prioritize ethical issues strengthens the value proposition of Geographic Information Science for Data Science.

More important than pervasiveness is the *effectiveness* of Spatial Data Science Ethics education. Individual instructors are primarily responsible for evaluating their students' achievement of their courses' educational objectives. At the programmatic level, however, effectiveness is reflected in student outcomes that have lasting positive impacts for graduates and the organizations in which they practice their professions.

May and Luth's (2012) "quasi-experimental field study" shows that formal ethics education – whether embedded with domain subject courses or delivered in "standalone" ethics courses – can positively affect students' empathy (i.e., ability to take different perspectives), moral efficacy (confidence in one's ethical abilities), and moral courage ("unwavering commitment to carry out tasks in an ethical manner despite potential negative consequences"). In this paper we will discuss strategies for evaluating the impact of ethics education in Spatial Data Science Ethics programs.

Coda

At the outset of this paper, we quoted J.B. Harley's to acknowledge his seminal thinking about ethics in mapping. Regarding his own question – Can there be a cartographic ethics? – Harley wrote "Ethics cannot be divorced from questions of social justice" (p. 15). We hold that Harley's concerns are even more urgent in our era of Spatial Data Science (or whatever you wish to call it) than they were in the context of cartography, a generation ago. We imagine that Brian would argue, but not disagree.

Acknowledgements

Thanks to the John A. Dutton e-Education Institute and the College of Earth and Mineral Sciences in the Pennsylvania State University, much of the courseware developed for Spatial Data Science Ethics instruction is licensed for reuse under CC-BY-NC-SA. Video lectures are available at the GEOG 581 YouTube playlist https://www.youtube.com/watch?v=3Opv5j5oJmE&list=PLnoT6jlTOgwDpytC5GpfOC5StC_IB3JgH. Educators are welcome to request further free access to ethics courseware by contacting the authors.

References

Anselin, L. (2017) Space Skepticism. University Consortium for Geographic Information Science blog, October 24, https://bit.ly/3Fx0mHI

Davis, M, (1999). Ethics Across the Curriculum. Chapter 7 in *Ethics in the University,* Routledge.

DiBiase, D. (2017). Professional and Practical Ethics of GIS&T. *The Geographic Information Science & Technology Body of Knowledge* (2nd Quarter 2017 Edition), John P. Wilson (ed.). doi: 10.22224/gistbok/2017.2.2

Harley, J.B. (1991), Can there be a cartographic ethics? *Cartographic Perspectives* 10, doi: 10.14714/CP10.1053

H.R.2272 - 110th Congress (2007-2008): America Creating Opportunities to Meaningfully Promote Excellence in Technology, Education, and Science Act. (2007, August 9). http://www.congress.gov/

May, D.R. and M.T. Luth (2013). The Effectiveness of Ethics Education: A Quasi-Experimental Field Study. *Science and Engineering Ethics*, 19. doi: 10.1007/s11948-011-9349-0

Nelson, T.A., Goodchild, M.F., and Wright, D.J. (2022) Accelerating Ethics, Empathy, and Equity in Geographic Information Science. *PNAS Environmental Sciences Computer Sciences,* 199, doi: 10.1073/pnas.2119967119

Oliver, J.C. and McNeil, T. (2021). (2010). Undergraduate Data Science Degrees Emphasize Computer Science and Statistics but Fall Short in Ethics Training and

Domain-Specific Context. *PeerJ Computer Science*, 7:e441, doi: 10.7717/peerj-cs.441

Schuurman, Nadine (2000). Trouble in the Heartland: GIS and its Critics in the 1990s. *Progress in Human Geography* 24:4, pp. 569-590.

Singleton, A. and Arribas-Bel, D. (2021). Geographic Data Science. *Geographical Analysis* 53, doi: 10.1111/gean.12194