



**A UCGIS Call to Action:
Bringing the Geospatial Perspective to Data Science Degrees and Curricula**

Summer 2018

As a long-established information science discipline, the Geographic Information Science & Technology (GIS&T) community has key contributions to make to evolving data science curricula. This statement articulates the University Consortium for Geographic Information Science's (UCGIS) position for the academic GIS&T community and provides recommendations and action items for the benefit of both internal and external audiences. On May 22-24, 2018, UCGIS held its annual Symposium under the theme of *Frontiers of Geospatial Data Science*, coordinated this year with the AutoCarto conference of the Cartography and Geographic Information Society (CaGIS). Drawing from discussions at that event, together with many months of internal exchanges, UCGIS offers these statements for the benefit of its member organizations as well as the broader geospatial community. The goals of this white paper and its recommendations are to 1) describe and clarify the value of incorporating geospatial knowledge, skills, and data for students, employees, and employers within the emerging field of data science; 2) highlight potential pathways and opportunities for academic geospatial scientists to establish connections with data science programs and personnel on their university campuses; and 3) initiate a national dialogue about the synergistic benefits of mutually enriching data science and geospatial science curricula.

Context

Virtually every sector of industry, business, government, and science is awash in data of great volume, variety, and velocity. In light of calls for fairness, accountability, transparency, and reproducibility, data accuracy and authority are also highly relevant. As an interdisciplinary field, there are high expectations for the capabilities of data science¹ to address myriad demands for innovative breakthroughs. "Data Scientist" has become an in-demand job title, though the nature of the positions varies widely. The most common skill sets required are analytical and quantitative in nature: to be able to manage and help others interpret large and diverse data sets.

Data that are geographically referenced or contain some type of location markers are both common and of high value (e.g., data subject to state-specific policies, laws and regulations; demographic data from the census; location traces of smartphones and vehicles; remotely sensed imagery from satellites, aircraft and small unmanned aerial vehicles; volunteered geographic information; geographically referenced social media postings). A 2011 McKinsey

¹F. Berman et al., [Realizing the Potential of Data Science](#), Communications of the ACM, 61(4):67-72, April 2018. DOI: 10.1145/3188721.

Global Institute report estimates a value of “about \$600 billion annually by 2020” from leveraging personal location data² to reduce fuel waste, improve health outcomes, and better match products to consumer needs. Spatial data are critical for societal priorities such as national security, public health & safety, food, energy, water, smart cities, transportation, climate, weather, and the environment. For example, remotely-sensed satellite imagery is used to monitor not only weather and climate but also global crops³ for early warnings and planning to avoid food shortages.

However, spatial data presents unique data science challenges. Recent court cases that address gerrymandering, the manipulation of geographic boundaries to favor a political party, offer a high-profile example. Instances of such exploitation of the modifiable areal unit problem (or dilemma) is not limited to elections since the MAUP affects almost all traditional data science methods in which results (e.g., correlations) change dramatically by varying geographic boundaries of spatial partitions. The fundamental geographic qualities of spatial autocorrelation, which assumes properties of geographically proximate places to be similar, and geographic heterogeneity, where no two places on Earth are exactly alike, violate assumptions of sample independence and randomness that underlie many conventional statistical methods. Other spatial challenges include how to choose between a plurality of projections and coordinate systems and how to deal with the imprecision, inaccuracy, and uncertainty of location measurements. To deal with such challenges, practitioners in many fields including agriculture, weather forecast, mining, and environmental science incorporate *geospatial data science*⁴ methods such as spatially-explicit models, spatial statistics⁵, geo-statistics, geographic data mining⁶, spatial databases⁷, etc.

Meanwhile, in response to the aggressive demand for skilled data scientists, higher education has been establishing data science degrees and programs, typically involving departments of computer science, statistics, and mathematics. As articulated in statements by the Computing Research Association⁸ and the American Statistical Association⁹, robust computational and statistical skills are necessary to achieve the types of analytical results that decision makers have been promised from data science, and have come to expect. At the same time, only spatial skills¹⁰ allow for the holistic analysis of the local and neighboring environmental context (e.g., climate, infrastructure, hazards), and social data (e.g., laws, demographics, culture) at multiple scales. Integrating this geographic contextual perspective significantly advances students’ skill sets to

² J. Manyika et al., Personal Location Data (Chapter 3e), [Big data: The next frontier for innovation, competition and productivity](#). McKinsey Global Institute, May, 2011.

³ Global Crop Monitoring (www.geoglam.org), Group on Earth Observations.

⁴ Y. Xie et al., [Transdisciplinary Foundations of Geospatial Data Science](#), *ISPRS Intl. Jr. of Geo-Informatics*, 6(12):395-418, 2017. DOI: [10.3390/ijgi6120395](https://doi.org/10.3390/ijgi6120395).

⁵ N. Cressie, [Statistics for Spatial Data](#), Wiley, 1993 (1st ed.), 2015 (Revised ed.).

⁶ H. Miller and J. Han, [Geographic Data Mining and Knowledge Discovery](#), CRC Press, 2009 (2nd Ed.).

⁷ S. Shekhar and S. Chawla, [Spatial Databases: A Tour](#), Prentice Hall, 2003.

⁸ [“Computing Research and the Emerging Field of Data Science,”](#) Computing Research Association October 2016.

⁹ [ASA Statement on the Role of Statistics in Data Science](#), Am. Statistical Association, October 2015.

¹⁰ S. Shekhar et al., [Spatial Computing](#), *Communications of the ACM*, 59(1):72-81, Jan 2016.

address the data science questions within public health, public safety, agriculture, water, energy, and the environment. As examples, spatially-informed data science enables understanding of gene-environment interactions affected by local impacts of climate change, or sophisticated analyses of the geographic variations in data protection regulations (e.g., EU GDPR¹¹). The bottom line: spatial is key within the “variety” dimension of big data and data science, and represents a vital variety of methods beyond the initial “one-size-fits-all” toolkit.

Whether an academic data science program is new or simply rebranded, the curricula necessary to support the emerging expected outcomes must be flexible to accommodate the necessarily dynamic skill sets for graduates. The National Academy of Sciences recently concluded a 16-month project¹² for [Envisioning the Data Science Discipline for Undergraduate Students](#), and one of its recommendations is that “academic institutions should be prepared to evolve programs over time” (Recommendation 5.1, p. 3), reflecting the newness of the academic field and the unusually dynamic and rapidly shifting landscape of data sources, computational platforms, analytical approaches, and algorithms. Accordingly, the data science curriculum has rich opportunities to incorporate spatial aspects via courses (required or elective) and specializations (e.g., a minor or a major). Additionally, the report suggests that “ethics be woven into the data science curriculum from the beginning and throughout” (Recommendation 2.4, p. 25). Spatial data, methods, and use cases may be leveraged to integrate ethically-based decision-making in an engaging manner via current societal debates such as those in election gerrymandering court cases or the negative impact of social feedback loops¹³ in crime hotspot analyses.

UCGIS is committed to providing channels for individuals from its member organizations to share their opinions around issues that matter to the community. At the May 2018 Madison, Wisconsin event, approximately 110 participants heard from a panel of geospatial scientists and professionals about their observations on data science trends and opportunities for geospatial science integration and then, working in small breakout groups, participants discussed and recorded their responses to a series of related questions. The resulting collection of insights and suggestions have informed the UCGIS set of recommendations and action items, listed below. For reference, the discussion notes have been organized into two sets of collective responses: 1) *What types of data science activity are taking place at your institution, and how is geospatial being coordinated, if at all?* (Appendix A), and 2) *What could UCGIS be doing with regards to data science?* (Appendix B).

¹¹ The [General Data Protection Regulation](#) (GDPR) 2016/679 is a regulation in the European Union (EU).

¹² [Envisioning the Data Science Discipline: The Undergraduate Perspective](#), National Academies of Science, Engineering and Medicine, National Academy Press, 2018.

¹³ Cathy O’ Neil, [Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy](#), Crown, 2016.

UCGIS Recommendations for Actions around Geospatial Data Science

Recommendation #1: Revise and update your curriculum. Integrate geospatial data science in your data science program, and vice versa. To do so:

- Examine and evaluate geospatial science curricula as core and optional components of the data science curriculum. Use the [GIS&T Body of Knowledge](#) to help you identify topics of relevance. Make sure that theoretical principles constitute the basis of your curriculum, with techniques and methods to complement the theory
- Familiarize yourself with the way that peer institutions are integrating data science principles and methods into existing undergraduate and graduate programs. Examples include the University of Oregon's BS in [Spatial Data Science & Technology Major](#), the University of Southern California's MS in [Spatial Data Science](#), North Carolina State's new [Geospatial Analytics](#) PhD, the University of Pennsylvania [MS in Urban Spatial Analytics](#), and others.
- If feasible, consider creating a new track (e.g., a minor or a specialization) within a data science program that blends geospatial science in a holistic and integrated manner.
- Integrate geospatial scientists and faculty members into your research teams, teaching groups, and advisory boards. Look around you on your campus, starting with Geography and GIScience and Technologies programs.

Recommendation #2: Establish a line of communication with alumni and potential employers.

Seek feedback on your existing GIS and GIScience and Technologies undergraduate and graduate programs to determine whether your graduates have acquired the quantitative, analytical, and/or programming skills (collectively, data science skills) needed to become successful professionals. Design and implement a process to self-evaluate your program on a regular basis and with clearly defined methods.

- Consult with your program's alumni and employers of alumni to determine if your program's graduates are qualified to take a position as geospatially-focused "data scientist." Consider annual "open house" event for students, alumni, and potential employer
- Implement a protocol for your program's self-evaluation, identifying methods, goals, and personnel dedicated to the task. Learn from the process to improve your program
- Educate your student to explain to potential employers how their knowledge and skills in the geospatial field are relevant to seeking employment in data science. This could be done in a classroom setting, via student organizations, special events, etc.

Recommendation #3: Engage with your campus community. Connect with faculty and students in your department, at the college level, and across campus.

- Be prepared to engage in geospatial data science conversations with a mix of faculty and administrators. Highlight how a focus on geospatial data science may benefit students' education and collaborative research (see also below). Be prepared to concede that not all audiences will be receptive, especially as you establish connections outside your

college. Once connections are established, offer to make presentations on the topic and make sure to follow up.

- Share spatially-based research with other data science people in the department and on campus, via meetings, events (including student-led), and formal and informal networks. For example:
 - Dedicate one track or session or set of posters prepared during the annual GIS Day to topics that highlight the data science dimensions of the research.
 - Invite a data science person (preferably one who regularly employs geospatial approaches in his/her work) to give a department's colloquium.
 - Invite an alumnus or alumna to give a presentation to students and faculty and ask him or her to highlight how data science is relevant to their profession.
- Identify and promote manageable-scale data-science research projects that could involve students and faculty from multiple departments.
 - Pursue funding from internal sources at the department, college, Provost's office, or research office in your campus, as applicable.

Appendix A: [Group Responses for "What is happening on your campus with regards to Data Science?" \(pdf\)](#)

Appendix B: [Group Responses for "What Should UCGIS Do About Data Science" \(pdf\)](#)

UCGIS is a non-profit scientific and educational organization comprised of 60+ member and affiliate institutions. Established in 1995, the mission of UCGIS is to advance research in the field of Geographic Information Science, expand and strengthen multidisciplinary Geographic Information Science education, and advocate policies for the promotion of the ethical use of and access to geographic information and technologies. We achieve these goals by building and supporting scholarly communities and networks. UCGIS serves as the ongoing steward of the [Geographic Information Science & Technology Body of Knowledge](#), hosts an annual Symposium, and supports an NSF-funded initiative of professional development for academic women in the geospatial sciences. UCGIS is a hub for the GIS research and education community in higher education.