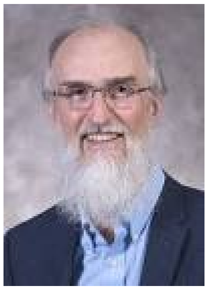
ALAN M. MACEACHREN
Department of Geography
Penn State University
Email: maceachren@psu.edu



Alan M. MacEachren is Professor of Geography and Information Sciences & Technology, and Director of the GeoVISTA Center (www.GeoVISTA.psu.edu) at the Pennsylvania State University. MacEachren's research foci include: geovisual analytics, cartography, geovisualization, visual semiotics, spatial cognition, user-centered design, place and big data, and geographical information retrieval. He is author of How Maps Work: Representation, Visualization and Design and Some Truth with Maps; and co-editor of additional books (including *Exploring Geovisualization*) as well as of nine journal special issues. He chaired the International Cartographic Association (ICA) Commission on Visualization (1995–1999) and the Commission on Visualization & Virtual Environments (1999–2005). MacEachren has served on the U.S. NRC, Rediscovering Geography Committee, (1993–1997), the NRC Computer Science and Telecommunications Board Committee on the Intersections Between Geospatial Information and Information Technology (2001–2002) and as a member of the National Visualization and Analytics Center R&D Agenda panel (2004–2005). He was also Associate Editor of IEEE Transactions on Visualization and Computer Graphics (2007–2011) and is an Associate Editor of Information Visualization (2001–present). Most recently, he was an invited contributor at a set of three AAAS-organized workshops focused on ethics and data (Developing Ethical Guidelines and Best Practices for the Use of Volunteered Geographic Information and Remotely Sensed Imagery in Crisis Situations, 2017–2019). MacEachren was named as an Honorary Fellow of the International Cartographic Association in 2005; elected as a Fellow of the American Association for Advancement of Science in 2014, and the given the Association of American Geographers Wilbanks Award for Transformational Research in Geography in 2019.

A Few Perspectives on Human-Centered Spatial Data Science—How can we "Do No Harm" while still doing some good?

Conceptualizing *Spatial Data Science* as distinct from Data Science more generally continues the long tradition in GIScience (and its precursors) of considering *spatial* to be *special*. There have been many arguments put forward over time about why this contention is true. But, it may prove to be most obvious with the rapid developments in data science for which location often provides the "glue" that makes it possible to connect fragments of information from big and diverse data sources. This role as a connector for a vast array of other information is the feature that will make spatial data science special and powerful. But, it is also the feature that raises two sets of challenges (at least): (a) conceptual and technological challenges about how to extract meaningful information from big, dynamic, heterogeneous, unconfirmed, hyper-connected data; and (b) equally important ethical challenges about applying data science methods to hyper-connected, location-linked data. In an era of fake news, erosion of privacy, and technologies that enable misinformation at light-speed, these latter challenges are ones that the community must address head-on from the beginning, not as an after-thought.

The core argument underlying this brief note is that success of spatial data science will depend on valuing humans over data. Thus, I contend that for spatial data science to make a positive impact (on science and society), it needs to be human-centered. Here, I sketch four research opportunities/challenges that together can lead to a *Human-Centered Spatial Data Science (HCSDS),* a data science that addresses place-related scientific and societal needs while continually reflecting on human/societal implications of the research we undertake and the practices it enables.

The *first* human-related opportunity and challenge to address, in developing HCSDS, is the dramatic increase over the past decade of *humans as sensors*. At least in the developed world, the typical human moves through the world carrying or wearing multiple location-based sensors and/or using vehicles with such sensors. These include our smart-phones, smart-watches, cars, e-Scooters, health monitoring devices, and increasingly even our clothes. Interfaces designed (or not) to collect location-based data from the "crowd," and those intended for the crowd to specifically collect data for us, impact the data volume, representativeness, quality, and relevance of resulting data. There are tremendous opportunities to leverage advances in location-aware technologies and the related miniaturization of sensors in order to utilize the crowd for collecting data at a scale and resolution never before possible, as well as data about topics for which data has previously been virtually non-existent. Doing some good with this endeavor while avoiding harm will require research on systems that enable everyone to make choices about what location-based data to collect when and for whom, as well as on visual interfaces that help users process the messy heterogeneous data to generate useful information.

The *second* human-related opportunity and challenge to address is the role of (and necessity for) *human expertise* in guiding computational tools—thus how do we achieve human-in-the-loop computing that is "better" than either computational or human solutions alone. A guiding premise here is that computational models benefit from human input on the front-end as well as throughout what is often an iterative process of model refinement. This is not a new idea, but one that has remained (mostly) on the back-burner since the early 1990s when visualization-based, computational model steering strategies were first introduced. Recent efforts to link advances in visual analytics with those in deep active learning offer an exemplar of the potential.

The *third* human-related opportunity / challenge is to develop methods that enable *human understanding* of computationally generated results—thus methods that provide support for *explainable spatial data science*. As with approaches to integrating human expertise into the process of developing and guiding computational models, we will not be pursing this goal alone; major efforts are underway in AI/machine learning to address the "explainability" challenge. A contention made with increasing frequency by visual analytics researchers is that interactive visual interfaces can enhance explainability of computationally-generated results. But, spatial is special. Thus, to enhance spatial data science explainability, those interactive visual interfaces need explicit support for spatial representation. Beyond that, however, my contention here is that *place* may matter more than *space* for spatial data science explainability. Human knowledge about the world is grounded in real-world places, and place rather than space is the basis for contextualizing analytical results. We have decades of research in cartography and geovisualization to draw upon as a starting point for developing innovative strategies that enable human understanding of computational results. But, most of that work takes a spatial perspective. Advances in *platial representation* will be essential for *explainable "spatial" data science*.

The *fourth* human-related opportunity and challenge is ethics. A key question is, "to what extent is ethical spatial data science possible"? Assuming the answer is yes (however qualified), how do we recognize and take into account the ethical implications of spatial data science—for *humans* and for the *environment* more generally? Spatial data science has unique ethical challenges that are essential to address as a core aspect of the agenda, particularly challenges associated with location as the 'glue' that connects data from diverse sources. Here, a useful starting point for developing an *ethics for spatial data science* is a recent AAAS series of workshops (on "Developing Ethical Guidelines and Best Practices for the Use of Volunteered Geographic Information and Remotely Sensed Imagery in Crisis Situations") and the subsequent report, *Location-Based Data in Crisis Situations* (https://www.aaas.org/resources/location-based-data-crisis-situations). That report outlines five principles that spatial data science can draw upon: do no harm, define your purpose, do good science, collaborate and consult, and give access to your data.

Spatial Data Science has clear potential to enable scientific advances across a wide range of fields and to support solutions for diverse societal challenges. Just like data science generally, however, it brings at least an equal risk to do harm; this is the case because links to location inherent in spatial data escalate the potential risk to individuals and place-based communities in many ways. Thus, as we work to adapt, extend, and invent new data science methods and technologies for applications with spatial data, a key question for spatial data science is *How can we "Do No Harm" while still doing some good?*