EKATERINA EGOROVA
Swiss National Science Foundation
Email: ekaterina.egorova@geog.uzh.ch

Ekaterina Egorova gained her Ph.D. in GIScience from the Department of Geography, University of Zurich, Switzerland in 2018. She subsequently got a Postdoc Mobility Grant from the Swiss National Science Foundation for an 18-month research project that explores and maps aspects of human-environment interaction in the context of nature-based recreation activities. Conducting this project, she spent one year in New Zealand, working at the Geoinformatics Collaboratory, Massey University and the School of Environment, University of Auckland. She is currently finishing the project as a visiting scholar at the Centre for Spatial Studies, UCSB Santa Barbara.

Egorova's research lies at the intersection of GIScience, spatial cognition and cognitive linguistics, and combines methods from NLP, Geographic Information Retrieval, cognitive and computational discourse analysis. Her interests include the production and processing of geographic information across a variety of contexts, spatial semantics, and methods for the automated extraction of spatial concepts from multimodal data sources.

With her primary research interest focussing on the language of space, Egorova has served on the programme committee for the Spatial Language Understanding workshop (SpLU'2020), Geographic Information Retrieval workshop (GIR'19), Speaking of Location workshop (COSIT'19, COSIT'17), International Symposium on Platial Information Science (PLATIAL'19), Combined Workshop on Spatial Language Understanding (SpLU'2019) & Grounded Communication for Robotics (RoboNLP'2019), and has co-organized the Corpus in GIScience workshop (GIScience'2018).

Position Statement

The global society of today is facing a number of large-scale challenges, ranging from the changing climate to the loss of biodiversity. In parallel, the rapid development of technology, and the massive proliferation of geospatial data bring new opportunities for tackling the challenges by transforming the process of scientific discovery—resulting in what is known as the fourth paradigm of science (Hey, 2009)—but also require an understanding of the realia and the consequences of this transformation. Not only does the large volume and variety of the data have to be adequately addressed through new tools for their storage, retrieval and analysis. More importantly, given the data-driven nature of a large part of the analysis today, the data collection process itself should be treated with ultimate caution, not to introduce and reproduce data-inherent biases that can be further fossilized through knowledge production. While these and related issues are not completely

new to science, it appears to be high time to dwell and re-consider some of the aspects of how we treat the data and the scientific process in spatial data science in light of these developments. In what follows, I chose three areas related to data production and use that particularly resonate with my research, suggesting them as potential starting points for a discussion on advancing spatial data science.

Multi-use data. Given the amount of resources needed for collecting, pre-processing, and storing large geospatial datasets, it is high time for the community to collectively adhere to the FAIR principles by aiming at making datasets findable, accessible, interoperable and reusable (Wilkinson et al., 2016). One particularly important issue to be addressed is the multi-use data, which requires transparent data collection procedures and high quality metadata, as well as effective data-sharing methods and tools, including the design and development of *searchable metadatabases*. The international spatial data science community could serve as a driving force behind establishing collaborative networks for raising awareness on the value and potential uses of existing and newly collected geospatial datasets, thus fostering research and policy efforts that address large-scale issues.

Multimodal data sources. Data (in particular, VGI) containing geospatial components comes today in a variety of communication modes, such as image and text. Given that different modes schematize and convey space in different ways (Tversky and Lee, 1998), it is important to understand the nature of each mode, as well as the ways in which various modes of information relate to each other and can be used in a complimentary way. For example, the spatial distribution of images from social media is widely used today as a digital proxy of the general "worthiness" of places (Richards and Friess, 2015), but it fails to reflect more elusive, subjective and intangible aspects of human-environment interaction. To address that gap, we should aim at leveraging text narratives as data, applying more sophisticated linguistic models capable of capturing such subtle elements of discourse. Leveraging multimodal data sources will result in a more balanced, multiple-perspective, *holistic representation of human-environment interaction* that will include both qualitative and quantitative information. This appears especially important in light of biases that are inadvertently reinforced through machine learning in many spheres today, including spatial sciences (Wachter, 2018).

Public participation. While citizen science and other participatory approaches are now fairly established, spatial data science still appears to label and treat citizens as *sensors* (Goodchild, 2007), a term that implies a certain degree of passiveness. It seems that it is high time for the spatial data science community to design a framework for a systematic engagement of citizen scientists in the knowledge production process. This will, firstly, increase the scientific reasoning skills among the public, which is key for decision making processes focussing on environmental issues. Secondly, this will foster the public's literacy in a variety of areas, transforming their role of sensors into that of knowledgeable *actors*—and as I see it, educating public on aspects such as geoprivacy (Keßler & McKenzie, 2018) is the moral duty of spatial data scientists.

The issues outlined above are only a few out of many that require attention and (re)consideration. The multidisciplinary nature of many of these issues (e.g. ethics and AI) as well as the multidisciplinary

nature of "spatial" in general call for truly interdisciplinary researchers who are eager to join efforts for setting the frontiers of spatial data science that reflect the realities of today and tomorrow, and this symposium appears a timely step towards this goal.

References

Goodchild, M. F. (2007). Citizens as sensors: the world of volunteered geography. *GeoJournal 69*(4): 211–221.

Hey, A.J. ed., 2009. *The fourth paradigm: data-intensive scientific discovery* (Vol. 1). Redmond, WA: Microsoft research.

Keßler, C., & McKenzie, G. (2018). A geoprivacy manifesto. *Transactions in GIS 22*(1): 3–19.

Richards, D. R., & Friess, D. A. (2015). A rapid indicator of cultural ecosystem service usage at a fine spatial scale: Content analysis of social media photographs. *Ecological Indicators 53*: 187–195. https://doi.org/10.1016/j.ecolind.2015.01.034

Tversky, B., & Lee, P. U. (1998). How space structures language. In *Spatial cognition* (pp. 157–175). Springer, Berlin, Heidelberg.

Wachter, S. (2018). Normative challenges of identification in the Internet of Things: Privacy, profiling, discrimination, and the GDPR. *Computer law & security review 34*(3): 436–449.

Wilkinson, M. D. et al. (2016). The FAIR Guiding Principles for scientific data management and stewardship. *Sci. Data* 3(160018) doi: 10.1038/sdata.2016.18