

Scale as an agent of transformation in geographic knowledge discovery and production

May Yuan, the University of Texas at Dallas

Innovations in observations and measurements are major drivers to scientific breakthroughs. The invention of the telescope pushed a huge leap forward in astronomy since the 1600s. Hubble Space Telescope captured high-resolution imagery to derive measurements that bolster the need for new physics to explain discrepancies in the Universe's Expansion rate (NASA¹). The invention of the microscope made possible histology, cell biology, and bacteriology. A new 3D microscope enables observations of fast biological processes, like blood cells move through two heart chambers and potentially the activity and dynamics of neural cell populations (AAAS²). Data are the foundation for science. Among the three broadly accepted definitions of scale, the minimum unit of observations (and therefore analysis) has the most potential to new knowledge discovery and production. Observations at a higher spatial resolution uncover the unknown, challenge the state of theories, and stimulate new algorithms and analytical methods.

GIS experience many advances in data acquisition technologies. Satellite imagery marks major improvements in spatial resolution, spectral resolution, and temporal resolution. As a result, updates are at more details, increasingly semantics, and higher frequencies on geographic knowledge about the environment, natural processes, and human activities. Federal, state, and city governments commonly build partnerships to fly manned areal photographs once or twice a year to estimate crop production, land-use land-cover change, or urban development. Now, unmanned aviation systems (UAS) – CubeSat, drone, and balloon-sat can capture fine resolution imagery on-demand and in short notice. Furthermore, increasingly extensive networks of google street views, surveillance cameras, parking cameras, weather cameras, traffic cameras, and cameras and sensors on manned or unmanned vehicles record observations of micro geography. UAS imagery, for example, can generate 1000s of aerial photographs in an hour over several acres of land. These UAS aerial photographs can be at the sub-mm resolution, uncovering previously unknown micromorphology and microstructures.

Besides high-resolution imagery, digital footprints of human activities are ubiquitous. Volunteered geographic information, georeferenced call detail records, geotagged social media posts, responses to location-based services, geo-enabled apps, and smart-card transit data are few but common producers of high-resolution data at high volumes. In addition, there are spatially implicit data producers of human activities at high resolution and high volumes, such as business transactions, social events, news reports, travel postings, and online reviews. These sources are spatially implicit because the location information is not given as coordinates but is with addresses, directions, place names, or spatial features. These data are at high resolution and inherently high temporal and semantic resolutions that can lead to detailed accounts for spatial events, activities, behaviors, and functions.

Have these innovations in geographic observations and measurements been driving geographic knowledge discovery or production? To date, strong evidence is yet to come. Many GIS publications utilize these new geospatial data and apply machine learning or artificial intelligence algorithms to categorize data, identify spatial clusters, uncover hotspots or cold-spots, and resolve patterns of relationships (such as association rules) for example. Many studies use these data to elicit point of

¹ <https://www.nasa.gov/feature/goddard/2019/mystery-of-the-universe-s-expansion-rate-widens-with-new-hubble-data>

² https://www.eurekaalert.org/pub_releases/2019-04/embl-n3m042619.php

interest (POI), human movement patterns, or social communities, for example. However, most studies suffer two drawbacks: (1) most findings are uninspiring. For example, social media clusters influenced by the underlying population distributions; regardless of what keywords are used, tweet clusters are often correlated with major cities; and (2) research lacks model validation and data are not shareable due to contractual agreements with data owners or concerns for privacy or confidentiality. As a result, most studies are not reproducible nor replicable. Also, there are various issues with biases and uncertainty in these data. As such, most of these studies have limited or even no contributions to geographic knowledge despite the fancy data and trendy algorithms adopted in the studies.

The three approaches below set forward frameworks of reference to promote transformation in geographic knowledge discovery with data at a high spatial resolution:

- (1) Rethink spatial problems. For example, traditional aerial photogrammetry is based on parallelism of overlapping photograph pairs to derive 3D measurements. High-resolution UAS images with extensive overlaps ($\geq 80\%$) have evoked a new framework of reference in which the Structure from Motion (SfM) algorithm meshes point clouds from these photographs to build a 3D model of buildings, trees, and the landscape. Because micro-scale data allow us to take advantage of minute features and structures with high fidelity across multiple observations, we can recognize something previously unknown or use them as anchors to fuse geospatial data across observations. Is spatial autocorrelation still meaningful for micro-scale properties? Will spatial similarity or spatial agglomerates of micro features (processes) to meso features (processes) be more interesting, instead?
- (2) Reinvent representation and methods to capture the duality of objects/fields, patterns/processes or forms/functions and how their relationships vary across scales. Most of the current GIS methods apply to objects or fields (but not both), identify patterns and compare patterns, simulate processes, and describe forms and relationships. Very few studies consider changes in object vs. field-view in a problem, relate patterns and processes, and connect forms and functions. A seemingly homogeneous unit at a larger scale would be spatially heterogeneous with high-intensity data, switching from a field-view (attentive to transition and gradation) to an object-view (attentive to segmentation and identity) of the world. Patterns emerge at one scale, but the processes responsible for the patterns may operate at one scale up or down. High-resolution data not only capture features in sub-mm units but also allow many levels of generalization or aggregation to cross-reference duality among these things to produce geographic knowledge from bottom-up (the finest data).
- (3) Relate site, situation, and event. The current GIS data and methods mainly concern about what we have (geographic features). What we have in a city characterizes the city's environmental resources, socio-economic conditions (e.g., demographic, social, economic variables), and community infrastructure (such as libraries, roads, etc.). Our GIS databases have little data about natural, historical, archaeological, or contemporary events. In particular, what people do reflect the characters and outlooks of a place. A city filled with arts or music festivals and another with rodeos or carnivals reflect their different characters beyond what we can draw from demographic or socioeconomic data. Are certain event assemblies reflect the social priority of the city or promote a sense of community, for example?

The three possible frameworks of reference are examples with aims to promote thinking about new geographic problems and approaches with geographic data at the microscale. We also need new methods, not for the often-mentioned reasons regarding big data, but for utilizing the unique advantages of geographic data the microscale to discover and produce new geographic knowledge.