Spatial Intelligence by Design: Embedding GIS Expertise in AI-Driven Data Infrastructure

***Stephanie Deitrick, Chief Data & Analytics Officer**

Center for Integrated Data Innovation, City of Tempe, AZ
*Stephanie_deitrick@tempe.gov

Keywords: data infrastructure, data quality, GeoAI, spatial data governance

Whether geographic meaning is preserved throughout the data pipeline is shaped by whether spatial expertise is present when organizational decisions about data are made.

The Problem

Recent work in data-centric AI has identified domain knowledge integration and real-world contextualization as among the least developed dimensions of data understanding in AI pipelines, with existing methods providing limited support for connecting data to its real-world meaning before it reaches a model (Holstein et al. 2026). GIScience has recognized the need for disciplinary engagement in data production as a field priority, noting that spatial data quality depends on expertise being present when decisions about how data are structured and what it represents are being made (Nelson et al. 2025). What neither conversation has examined is whether spatial expertise reaches those decisions in the organizations where spatial data are produced and managed, and what happens to data quality when it does not. That is the quality condition this paper examines: whether spatial expertise reaches the organizational decisions where data quality is established

The problem is not primarily at the model layer. Spatial meaning is lost long before AI systems engage with the data. Addresses are stored as strings. Jurisdictional identifiers are codes without geometry. Records with location attributes often live in business systems with no connection to spatial data infrastructure. By the time data reaches an AI system, the geographic context that would make outputs geographically reliable has already been stripped away, and better models do not recover it.

Where the Work Actually Happens

What AI models can do with spatial data is determined by decisions made well before any downstream access layer comes into play. Those decisions include how data are collected, how location is represented in source systems, what geometry gets attached to records, and whether the people making those decisions understand the spatial implications. In most organizations those decisions get made without spatial thinking as a first consideration, not because location is seen as unimportant, but because the people shaping data infrastructure seldom include someone with spatial training. By the time data are available for use, most of those decisions are already locked in. The access layer and AI models inherit those decisions, and the resulting outputs reflect them.

The judgment that cartographers and GIS professionals develop through their work, understanding scale, fitness for use, what a dataset can and cannot support, is exactly what needs to be present when data systems are being designed, not when outputs are being reviewed. That expertise does not require a large GIS team. In smaller organizations it may live in one person who bridges GIS, data governance, and IT. What matters is that it enters the process before infrastructure decisions are made, not after they are locked in.

A Different Organizational Model

The City of Tempe's Center for Integrated Data Innovation embeds spatial expertise across all enterprise data work, including AI development, rather than treating GIS as a service function. The intent is to have spatial thinking present early in decisions about data infrastructure and data design, not applied as a review process after systems are built. That organizational choice is still being worked out in practice, but it changes what questions get asked at the beginning of the process and who is in the room when data decisions get made.

The work has focused not just on integrating spatial data but also on identifying where spatial context is missing from datasets that contain location information in non-spatial forms. It also considers when that context should have been present and what its absence means for AI use. That work requires someone who can see across the enterprise, recognizing where spatial context matters to AI use even in data that was never designed with location in mind. That broader view does not emerge naturally from how most organizations are structured. Working through these questions in practice suggests that treating spatial integrity as a data quality dimension in enterprise AI environments is a research direction worth pursuing, and one the field is well positioned to take on. As AI systems gain more direct and automated access to institutional data, what upstream decisions built either holds up or falls short at scale, with no mechanism for detection on the AI side.

An Agenda for Research and Practice

The research direction this paper proposes centers on spatial integrity as the aspect of spatial data quality concerned with preserving geographic meaning throughout the data pipeline. It considers how spatial integrity is established at different stages of a data pipeline, how those decisions shape AI outputs, and how organizations can assess whether spatial context has been preserved. GIScience is well positioned to develop those frameworks, and doing so would connect established spatial data quality thinking to the organizational and infrastructure conditions where GeoAI is deployed.

The Tempe work is offered here not as a proven model but as one documented illustration of what those conditions look like in practice, and as an invitation to document similar efforts elsewhere. Organizations working through these questions are making consequential decisions about spatial data quality in AI environments in territory that GIScience research has not yet mapped. That is where the field has the most to contribute, because defining what spatial expertise contributes at each stage of a data pipeline, and what the pipeline loses without it, requires disciplinary knowledge that GIScience is best positioned to apply.

Conclusion

Most organizations will encounter GeoAI after AI tools are already in use. At that point, they are under pressure to retrofit spatial infrastructure. Having spatial expertise present when data systems are being designed, at each stage where geographic meaning can be preserved or lost, is a more promising approach than trying to restore spatial integrity after the fact. That is, at its core, a data quality argument, and it is one the GIScience community is well positioned to make and build frameworks around.

References

Holstein, J. et al. (2026) Data understanding for data centric AI, Business Information Systems Engineering, 68(4): 897-971.

Nelson, T. et al (2025) A research agenda for GIScience in a time of disruptions, International Journal of Geographical Information Science, 39:1, 1-24.