NATHAN PIEKIELEK
Geospatial Services Librarian
The Pennsylvania State University Library
Email: nbp104@psu.edu

Nathan Piekielek promotes spatial research and teaching in all departments and units at Penn State. Recent activities include establishing a maps and geospatial information support center from within the library. His research interests are varied and have included scientific workflow software and targeting non-geography academic disciplines for GIS outreach. He is currently working to develop fully- and semi-automated methods to convert print geographic resources (e.g., maps and historic aerial photographs), to research-quality digital geospatial datasets to be posted to centralized data portals. He is a taskforce member of the Big Ten Academic Alliance (BTAA) geoportal project and serves on their collection development subcommittee (https://geo.btaa.org/). Piekielek holds degrees in ecology, conservation ecology and sustainable development, and geography as well as postdoctoral experience working on climate change.

Perspective Statement

• How can spatially mediated discovery provide single-point access to research data, across distributed repositories and catalogs?
• How can the discovery of research objects in general be spatially supported? and □
• How can spatial discovery be applied to topic spaces, not just geographic ones?

Spatial discovery across repositories/catalogs is often mediated by metadata, specifically metadata elements that are populated and structured in common (i.e. compatible schemas). The metadata records and/or the data themselves are then aggregated (and cross-walked if necessary), and discovery tools built on top of the aggregated metadata records. This has been our approach in the BTAA geoportal project with the caveat being that we have only aggregated the metadata records—the data themselves remain in each individual institution’s repository. Issues with this approach are that it requires a substantial amount of effort (i.e., mediated deposit), to get an item into the collection and the collection itself requires constant update and maintenance to remain relevant. When considering applying this approach to research data one might realize that not every research project interacts with geographic space or frame of reference in the same way—for example in lab-based experiments, does the geographic location of the lab where the experiment took place have meaning? Was there a specialized instrument used in the experiment with a unique frame of reference that is meaningful to the results and for which “spatial” search should be enabled? Inconsistencies in representation of space and frame of reference make it difficult to define and enforce a common metadata element on which spatial discovery can be performed. For example, searching Penn State’s mediated deposit research data repository (DataCommons), by geographic search criteria will not produce every dataset in the result because a few (~30) escape geographic
description. Contrast this by searching Penn State’s self-deposit repository (Scholarsphere), where you cannot search by geography and where many datasets are not documented well enough for reuse.

In general, my perspective is that supporting interdisciplinary research data discovery is the far more tractable problem when compared to trying to get researchers to document their data well enough to support reuse. Rather than “metadata,” I might call this barebones characterization of data “documentation” including some kind of assessment of quality so that others may know for what other research questions the data are appropriate? What good is discovery if the data are not reusable by anyone except for (or including), the original researcher?