Geospatial Metadata for Discovery
in Scholarly Publishing

Tom Niers ⦿ and Daniel Nüst ⦿

Institute for Geoinformatics, University of Münster
tom.niers@wwu.de, daniel.nuest@wwu.de

Almost every scientific article that refers to existing regions of the earth, contains "[...] a narrative description of the study area" (Karl, 2019). At the same time, Shapiro and Báldi (2012) found that more than a quarter of articles omit maps and coordinates and only use vague descriptions albeit the relevance of location for said articles' content. Geospatial metadata can help to detect biases in research coverage (Karl et al., 2013; Young & Lutters, 2017), to filter search results for scientific articles (Howell et al., 2019; Karl, 2019; Karl et al., 2013), and to enhance the understanding of relations within a study area (Margulies, Magliocca, Schmill, & Ellis, 2016). In more than half of the scientific articles that refer to locations, coordinates are used to determine the location (Karl et al., 2013; Shapiro & Báldi, 2012). However, geospatial information on scientific articles is not yet exploited in scholarly publishing platforms. Coordinates can be included in articles in different formats (Karl, 2019; Kmoch, Uuemaa, Klug, & Cameron, 2018) and therefore are prone to errors such as improper formatting, incompleteness, and ambiguity (Karl, 2019; Margulies et al., 2016), so that demand for standardization increases (Karl, 2019; Karl et al., 2013; Kmoch et al., 2018; Margulies et al., 2016; Young & Lutters, 2017). In this work, we report on a novel approach to integrate well-defined geospatial metadata in a scholar publishing platform so it can enhance discovery of scientific articles.

geoOJS offers a novel way for authors to provide spatial properties of research works when submitting an article to a journal based on the open source software Open Journal Systems (OJS, https://pkp.sfu.ca/ojs/). Previous work applied text recognition (Kmoch et al., 2018) or pattern matching algorithms (Karl, 2019) to derive geospatial metadata from papers, but such fully automated workflows are not without shortcomings. Instead, we decided to streamline the user interaction to create geospatial metadata so that user's understanding of the most suitable connections with location(s) or area(s) is unambiguously recorded. Authors can either search for a location and accept the suggested bounding box or manually create one or more suitable geometric shape(s) on a map. If authors enter geometries, a gazetteer is used to suggest a matching administrative unit's name to the author. This allows *geoOJS* to store geospatial data in two forms: as text, using the above administrative unit or standardised geographical norm data, and as geospatial coordinates in GeoJSON format. Thereby the coordinates are stored accurately, while at the same time a textual description is accessible and flexible for non-map-related usage. In addition to displaying geospatial information on maps, it is also added to the HTML source

code of articles' landing pages in a semantically meaningful way, e.g., using Schema.org vocabulary. This facilitates indexing by search engines and can improve accessibility by supporting screen readers better than a regular map. To evaluate these goals, we implement a prototype for *geoOJS* and demonstrate the metadata input by the authors, the storage as precise coordinates and standardized texts, and the integration of location information in article views (see Figure 1).
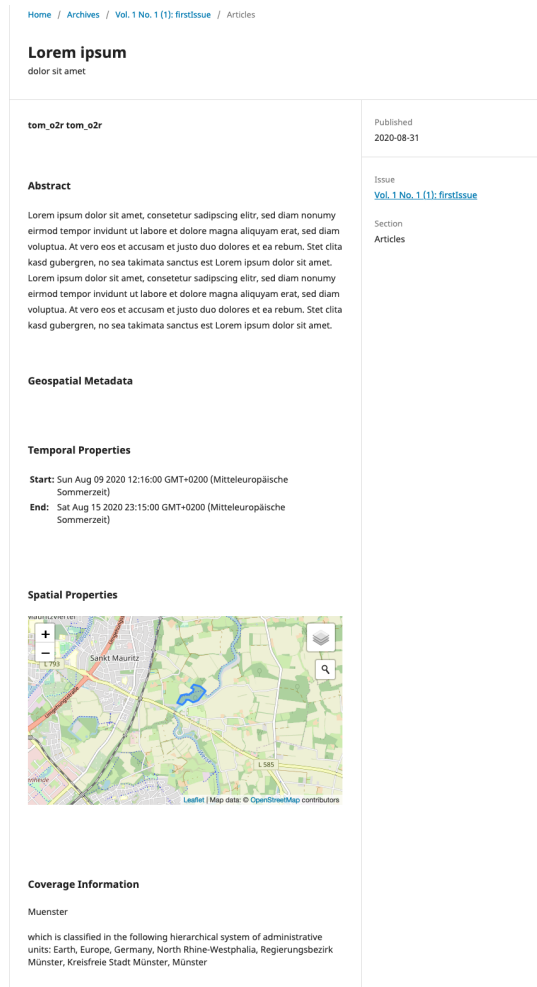

Figure 1: Screenshot of *geoOJS*: geospatial properties in the OJS article view

Future work includes research into usability and usefulness of geospatial metadata for discovery of articles, a search engine across OJS instances, geospatial and temporal filters in article search, and validation of geospatial metadata as part of the review process. We plan to release *geoOJS* in the OJS plugin gallery so that the increasing number of independent and Open Access journals may benefit from it.

References

Howell, R. G., Petersen, S. L., Balzotti, C. S., Rogers, P. C., Jackson, M. W., & Hedrich, A. E. (2019). Using webgis to develop a spatial bibliography for organizing, mapping, and disseminating research information: a case study of quaking aspen. *Rangelands*, *41*(6), 244–247. doi: 10.1016/j.rala.2019.10.001

Karl, J. W. (2019). Mining location information from life-and earth-sciences studies to facilitate knowledge discovery. *Journal of Librarianship and Information Science*, *51*(4), 1007–1021. doi: 10.1177/0961000618759413

Karl, J. W., Herrick, J. E., Unnasch, R. S., Gillan, J. K., Ellis, E. C., Lutters, W. G., & Martin, L. J. (2013). Discovering ecologically relevant knowledge from published studies through geosemantic searching. *BioScience*, *63*(8), 674–682. doi: 10.1525/bio.2013.63.8.10

Kmoch, A., Uuemaa, E., Klug, H., & Cameron, S. G. (2018). Enhancing location-related hydrogeological knowledge. *ISPRS International Journal of Geo-Information*, *7*(4), 132. doi: 10.3390/ijgi7040132

Margulies, J. D., Magliocca, N. R., Schmill, M. D., & Ellis, E. C. (2016). Ambiguous geographies: connecting case study knowledge with global change science. *Annals of the American Association of Geographers*, *106*(3), 572–596. doi: 10.1080/24694452.2016.1142857

Shapiro, J. T., & Báldi, A. (2012). Lost locations and the (ir) repeatability of ecological studies. *Frontiers in Ecology and the Environment*, *10*(5), 235–236. doi: 10.1890/12.WB.015

Young, A. L., & Lutters, W. G. (2017). Infrastructuring for cross-disciplinary synthetic science: Meta-study research in land system science. *Computer Supported Cooperative Work (CSCW)*, *26*(1-2), 165–203. doi: 10.1007/s10606-017-9267-z