Geospatial Health 2022; volume 17:1171

Does the Geohealth domain require a body of knowledge?

Ellen-Wien Augustijn-Beckers,[1] Sherif Amer,[1] Rob Lemmens,[1] Robert Bergquist[2]

[1]Faculty of Geo-Information Science and Earth Observation (ITC), University of Twente, The Netherlands; [2]Ingeröd, Brastad, Sweden

To answer the question if the Geohealth domain requires a body of knowledge (BoK), we need a general understanding of concepts associated with this field. Two years ago, the United Nation (UN) committee of experts on global geospatial information management identified "*semantic- and ontology-linked data"* as something that "*will become essential to support the next generation of autonomous systems*" (UN-GGIM, 2020). The term ontology is closely related to BoK. Unlike data models, ontologies are independent of application; they are generic, can be used in different ways and have clear advantages but they are challenging to create and even more difficult to maintain.

The brief description below summarizes what ontologies are, why they are needed to support linked data, what the role of the semantic web is, what is already going on within the Geohealth domain on ontologies and how a BoK can assist.

Ontologies and semantics and their use in GeoHealth

Ontologies consist of concepts and their relationships, where the former can have both a horizontal and a hierarchical relationship, like those of parent and child. A parent concept (class), *e.g.,* visualization, can be split into subclasses, such as disease diffusion maps, risk maps, etc. Concepts have properties ranging from name, definition and literature references to other information essential for ontology users, which need to be captured. Relationships, on the other hand, define how two such concepts are related (e.g., a patient is treated at a hospital). Ontologies are specific to a scientific field and create a common vocabulary leading to better information exchange.

Ontology formalizes semantics by explicating the meaning of the concepts and relationships that describe a scientific field. They can be used to overcome problems of semantic heterogeneity (Bittner *et al.*, 2009). Examples of semantic heterogeneity are dissimilar use of concepts due to changes in the name of a city over time, name differences depending on language and the use of different classifications of health-related attributes. These problems especially hinder longitudinal and global studies. Work on solving problems associated with ambiguous toponyms (placenames derived from topographical features), disease names and health services descriptions has been conducted by Somodevilla *et al.* (2013). The mapping of longitudinal study data in relation to risk factors of dementia is another example (Roantree *et al.*, 2016). In addition, One Health, a transdisciplinary approach that recognizes the connection between of people, animals and the environment as an indivisible element with regard to health, can face problems of different disciplinary and professional groups that use different and often incompatible terminologies to describe and structure relevant concepts, relationships and spatial data.

Linked data

Ontologies also facilitate data linkage that has to do with relationships or connections between *actual* data from different data sources, possibly from different organizations or even different scientific domains. By mapping data sources in this way, ontologies can express information at the conceptual level generating what is called *ontology-based data access* (Poggi *et al.*, 2008). Public health information systems are good examples of the use of linked data, which should ideally link patient data with information on the geospatial environment of patients, but this is often not the case. Ontologies can enhance the interoperability of data and help with respect to applications (Abburu, 2019). Effective healthcare systems require that all relevant data can be integrated (including spatial data), and ontologies play an important role here (Zimeras, 2012). This should not only relate to the exact locations but also include other spatial aspects, such as spatial relationships (Gao *et al.*, 2012). Integrating health and environmental components can lead to better data retrieval and analysis, overcome knowledge gaps (Rezaei & Vahidnia, 2022) and eventually improve public health policies.

Correspondence: Robert Bergquist, Ingeröd, Brastad, Sweden.
E-mail: editor@geospatialhealth.net

Key words: Body of knowledge; ontology; semantics.

Conflict of interest: The Authors declare no conflict of interest.

Received for publication: 14 November 2022.
Revision received: 14 November 2022.
Accepted for publication: 14 November 2022.

Geospatial Health 2022; 17:1171
doi:10.4081/gh.2022.1171

The semantic web

The semantic web is a knowledge network made up of linked data that allow computers and humans to process data from different sources and in this way produce answers to all kinds of questions. Semantic web technologies allow data to be processed by computers, which enables data sharing and reuse across data platforms (Gür *et al.*, 2012). An example of this is the public health emergency response system developed by Mao *et al.* (2009), in which the Internet is an information repository, both as data source and the place where information is published.

Body of knowledge (BoK)

A BoK is "*a collection of essential concepts, terms and activities within a profession or subject domain*" (Oliver, 2012), which makes it an overarching ontology defined by a scientific society or organisation. A BoK can also be called a domain ontology.

Previous sections of this editorial have given examples of ontologies being developed and used within the GeoHealth domain. However, they are driven by specific applications and systems requiring data linkage and do not provide a complete overview of the GeoHealth domain. Ontologies have a wider use than just data linkage; they define what a scientific community considers to be the main subjects they are working on. The BoK defining GeoHealth could also assist in specifying how this domain links to other scientific fields, especially the connected fields of Geoinformation science and Health that already have BoKs of their own. Two important BoKs for Geoinformation science are the EO4GEO BoK (Lemmens *et al.*, 2022) and the GIS&TBoK (DiBiase, 2007). In the health domain, there are standardized health ontologies, such as the Systematized Nomenclature of Medicine – Clinical Terms (SNOMED-CT) and the Unified Medical Language System (UMLS).

BoK creation

Domain ontologies are typically created by a group of experts within the domain, in which discussion helps formalize the domain further (Abdelghany *et al.*, 2019). The development of ontologies is iterative and includes cycles of improvement. The design of the ontology is defined by the envisioned use/users of the ontology. Several methodologies to generate ontologies exist, including expert-defined ontologies and literature data-mining, and there are several ways to reuse existing ontologies.

Different data-mining techniques exist for identifying concepts and relationships. For concepts, a keyword extraction method for journal articles can be used (Rospocher *et al.*, 2012), e.g., based on an ordered list of keywords. The frequencies of occurrence can be included. To determine relationships, the co-occurrence of terms can be used. A co-occurrence is the appearance of two concepts close together, e.g., within four words (de Boer & Verhoosel, 2020). Relationships can be mined by finding two co-occurring concepts and identifying the verb that binds these two together. Review papers related to the Geohealth domain such as, for example, Utzinger *et al.* (2011) and Krauskopf (2018), are good sources that could be used to identify concepts and relationships.

In the next step of ontology creation, concepts from other existing ontologies are imported/linked to the new ontology. The GeoHealth domain is positioned between domains, for which ontologies already exist: the Geoinformation science domain and the Health domain. Identifying overlaps can help to avoid inconsistency between ontologies by linking directly to the concepts in these other domains; descriptions of the concepts do not have to be recreated and consistency between ontologies can be ensured. However, certain tools are required to visualize and use an ontology. The EO4GEO BoK is available in The Living Textbook tool (LTB), which is available from an open-source link[1] and creating the GeoHealth BoK by the same tool would ease this integration.

How can scientific experts contribute and benefit from ontologies?

BoKs are developed via a broad involvement based on open principles (Ong & He, 2016). Involvement of many experts ensures that all subdomains are covered and that the BoK developed provides a balanced overview of the complete scientific domain. Experts should also write a description of the concepts. When made an integrated part of ontology development, experts are more likely to use the ontology and promote it with their students and colleagues and via their projects. This means that a great effort is needed to start a BoK, which might require an international project to ensure enough financial means and commitment. A nice spin-off is that building a BoK can lead to stronger bonding within the community, strengthening the network and lead to new collaborations.

Maintaining a BoK is also important as a scientific domain is dynamic, and the BoK needs to represent new insights and developments. Within expert groups, different opinions can make it hard to decide on the boundaries of the BoK, its granularity and its concept descriptions. BoKs should be publicly accessible but also require a board of editors that keeps them up-to-date. This means that dedicated experts are needed as board members or editors to define this ontology's scope and application domain and make it publicly available.

References

Abburu S, 2019. Ontology Driven Cross-Linked Domain Data Integration and Spatial Semantic Multi Criteria Query System for Geospatial Public Health: Concepts, Methodologies, Tools, and Applications. In I. Management Association (Eds.), Geospatial Intelligence: Concepts, Methodologies, Tools, and Applications IGI Global. pp. 714-743.

Abdelghany A, Ramadan N, Hefni H, 2019. An agile methodology for ontology development. Int J Intell EngSyst 12:170-181.

Bittner T, Donnelly M, Smith B, 2009. A spatio-temporal ontology for geographic information integration. Int J Geogr Inf Sci 23;765-98.

de Boer M, Verhoosel JPC, 2020. Towards Data-driven Ontologies: a Filtering Approach using Keywords and Natural Language Constructs. Proceedings of the Twelfth Conference

[1] *https://www.itc.nl/about-itc/organization/resources-facilities/living-textbook/*

on Language Resources and Evaluation (11–16 May 2020).European Language Resources Association, Marseille, France. Accessed on 11.12.2022. Available from: https://aclanthology.org/2020.lrec-1.278
Dibase D, 2007. Is GIS a wamperter? Transactions in GIS 11:1-8.
Gao S, Anton F, Mioc D, Boley H, 2012. Non-spatial and geospatial semantic query of health information. Int Arc Photogramm, Remote Sens Spat Inf Sci XXXIX-B2:167-72.
Gür N, Sanchez LD, Kauppinen T, 2012. GI systems for Public Health with an Ontology Based Approach. Proceedings of the AGILE'2012 International Conference on Geographic Information Science, Avignon, April, 24-27, 2012.In Multidisciplinary Research on Geographical Information in Europe and Beyond.Eds: Jérôme Gensel, Didier Josselin and Danny Vandenbroucke. Agile, Avignon, France. Accessed on 11.12.2022. Available from: https://agile-online.org/images/conferences/2012/documents/proceedings/papers/paper_guer_gi_systems_for_public_health_with_an_ontology_based_approach_2012.pdf
Krauskopf E, 2018. A bibiliometric analysis of the Journal of Infection and Public Health: 2008–2016. J Infect Public Health 11:224-9.
Lemmens R, Lang S, Albrecht F, Augustijn E, Granell C, Olijslagers M, Pathe C, Dubois C, Stelmaszczuk-Górska M, 2022. Integrating concepts of artificial intelligence in the eo4geo body of knowledge. Int. Arch. Photogramm. Remote Sens Spat Inf Sci XLIII-B4-2022:53-59.
Mao X, Li Q, Zhang Z, Zhu Q, 2009. Application of Spatial Information Search Engine Based on Ontology in Public Health Emergence. Proceedings of the Third International Conference on Bioinformatics and Biomedical Engineering, Beijing, China 11-13 June 2009. IEEE Engineering in Medicine and Biology Society 2009. Accessed on 11.12.2022.Available from: https://ieeexplore.ieee.org/document/5162396
Oliver GR, 2012. Foundations of the assumed business operations and strategy body of knowledge (bosbok) : An outline of shareable knowledge. Sydney University Press.
Ong E, He Y, 2016. Community-based Ontology Development, Annotation and Discussion with MediaWiki extension Ontokiwi and Ontokiwi-based Ontobedia. AMIA Jt Summits Transl Sci Proc 65-74.
Poggi A, Lembo D, Calvanese D, De Giacomo G, Lenzerini M, Rosati R, 2008. Linking data to ontologies. In Lecture Notes in Computer Science, Spaccapietra, S. (ed) vol 4900. Springer, Berlin, Heidelberg. J Data Semant X. Doi: 10.1007/978-3-540-77688-8_5
Rezaei Z, Vahidnia MH, 2022. Effective medical center finding during COVID-19 pandemic using a spatial DSS centered on ontology engineering. GeoJournal, 11 pp. Doi: 10.1007/s10708-022-10777-3
Roantree M, O'Donoghue J, O'Kelly N, Pierce M, Irving K, Van Boxtel M, Köhler S, 2016. Mapping longitudinal studies to risk factors in an ontology for dementia. Health Informatics J 22;414-426.
Rospocher M, Tonelli S, Serafini L, Pianta E2012. Key-Concept Extraction for Ontology Engineering. Knowledge Engineering and Knowledge Management. EKAW 2012. Lecture Notes in Computer Science, vol 7603. Springer, Berlin, Heidelberg. Doi: 10.1007/978-3-642-33876-2_36
Somodevilla MJ, Perez de Celis Herrero C, Hernández JA, Peneda Torres, I. H, 2013. Ontology Based Framework to Represent Relationships between Biomedical Spatial Data. The Twelfth Mexican International Conference on Artificial Intelligence, pp. 257-261.
UN-GGIM, 2020. p. 36. in Future trends in geospatial information management: the five to ten year vision. United Nations,Department of Economic and Social AffairsStatistics Division24-30 Nov. 2013. Accessed on 11.12.2022. Available from: https://ggim.un.org/
Utzinger J, Rinaldi L, Malone JB, Krauth SJ, Kristensen TK, Cringoli G, Bergquist R, 2011. Geospatial Health: the first five years. Geospat Health 6:137–54.
Zimeras S, 2012. Bayesian Ontologies in Spatial Integrating Medical Information Systems. pp. 220-232 in Quality Assurance in Healthcare Service Delivery, Nursing and Personalized Medicine: Technologies and Processes. A. Lazakidou & A. Daskalaki Eds. IGI Global.