Trends and Gaps in Ontology-Supported Environmental Health

Lucas Felipe Moreira Silva
Renato F. Bulcão-Neto
Instituto de Informática
Universidade Federal de Goiás
Goiânia, Goiás, Brazil
lucassilva@inf.ufg.br
rbulcao@ufg.br

ABSTRACT

Environmental Health (EH) refers to aspects of human health affected by factors in the environment, e.g., biological factors, and it is an essential part of any comprehensive public health system. Similar to other health-related fields, one observes an increasing movement in the adoption of IoT technologies into the EH domain. Regarding the data life cycle in IoT systems, data modeling and interpretation are crucial tasks in which ontologies are a feasible solution because of their expressiveness and reasoning support. In this paper, we structure the ontology-supported EH research theme through a systematic literature mapping. The identification and selection strategies of primary studies include the automatic search for studies published from 2010 to 2019 on five sources and the application of inclusion and exclusion criteria on an eight-hundred-eleven-distinct-paper group. The results of this original work provide an overview of the research theme with multiple classifications of thirty-four relevant studies remaining as well as the finding of trends and gaps for future work.

KEYWORDS

Ontology, Environmental Health, Systematic Literature Mapping, Internet of Things

1 INTRODUCTION

Environmental Health (EH) concerns aspects of human health affected by agents in the environment, e.g., biological ones. EH plays an essential role in any comprehensive public health system, and it is a topic of great interest in sustainability and smart-cities management [1]. Various physical, chemical, and biological factors affect the environment's quality, such as air quality, a source of many respiratory problems like asthma. Studies addressing EH face an inherent difficulty in the field: the massive number of interconnected elements and the effects on human health [2, 3].

Similar to other health-related fields, the literature observes a growing usage of IoT technologies in the EH domain. Most computer-supported EH research fits in the IoT paradigm, as they move towards the use of sensors, actuators, and other types of devices interconnected [4–6]. For instance, Reis et al. [4] highlight the importance of the integration of sensors and data models towards a Big Data scenario using human and environmental health challenges as an example. To prevent exposure risk to polluted internal spaces, Pitarma et al. [5] propose a low-cost, wireless sensor network system for indoor air quality monitoring composed of sensors of air temperature, humidity, CO, CO_2, luminosity, and specific pollutants. Finally, Bran et al. [6] designs a low-cost, remote sensor dot based on a simplified message-passing communication model as a means of addressing the negative environmental impacts of the buildings in their occupant's health.

Collection, modeling, reasoning, and distribution of sensor data constitute the data life cycle in IoT systems [7]. Due to their expressiveness and reasoning support, ontologies are a feasible solution for sensor data modeling and interpretation, which are critical tasks in the IoT data life cycle [8–11]. An ontology is a formal specification of shared conceptualizations of a specific domain [12]. This enables different IoT applications to represent and analyze EH entities, terms, and relationships in a consensual manner towards reaching their goals.

In this paper, we structure the ontology-supported EH research theme through a systematic literature mapping (SLM). The main goal of an SLM is to map the state of art of a specific topic and consequently point out trends and gaps for further research [13, 14]. Based on guidelines for performing SLMs [14], we elaborated on a protocol describing our research questions, the identification and selection strategies of primary studies, and the data analysis and synthesis tasks. We performed the automatic search for studies published from 2010 to 2019 on five sources, ACM DL, Engineering Village, IEEE Xplorer, PubMed, and Scopus, and applied inclusion and exclusion criteria on eight-hundred-eleven papers.

This original work chart thirty-four relevant studies on research maturity, the type of research contribution, and the types of environmental factors addressed in each paper. Moreover, we point out trends and gaps arisen from the SLM to guide future research.

The remainder of this paper is structured as follows: Section 2 details the SLM protocol and the results of the initial selection of studies; Section 3 describes the data extracted from the selected studies; Section 4 synthesizes our results; and Section 5 presents threats to the validity of the SLM as well as trends, gaps, and future work on ontology-supported EH.

2 THE SLM PLANNING AND CONDUCTION

Every SLM represents a systematic compilation of primary studies on a research topic; thus it shall follow a well-defined process to search, select, analyze, and synthesize the set of evidence available in a repeatable and non-biased way [14]. We defined a process composed of three phases, as depicted in Figure 1: protocol planning, conduction, and publishing of results.

The protocol planning includes the statement of the main objectives, the research questions, the search strategy, the search string definition process, the studies selection criteria, and a pilot test for protocol evaluation purposes. The conduction phase encompasses the activities of identification and initial selection of primary studies, the final selection of studies based on data extraction, and data synthesis. Finally, synthesis results are reported with textual, tabular, and graphical descriptions in the form of scientific papers or technical reports. We have been using the *Parsif.al* tool analyzed in [15] to support the protocol planning and the conduction phase of this SLM.

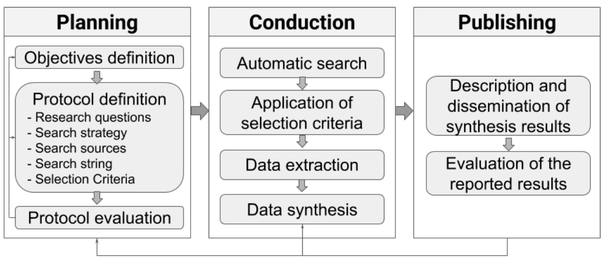

Figure 1: Phases and activities of this SLM.

2.1 Research Questions

The primary objective of this SLM is to investigate the state of the art of computing research focused on ontology-supported EH. That objective is compiled in the following research questions (**RQ**) and their respective justifications:

RQ1. *Which type of EH data is addressed?* – To recognize the heterogeneity of EH data.

RQ2. *Which type of environmental data source is exploited?* – To verify if data come through sensors, datasets, or simulation.

RQ3. *If applicable, which well-established, medical terminologies and coding systems (MTCS) support EH research, and for what purpose?* – Exploiting these expert-curated terminologies and coding systems makes knowledge machine-understandable, easy to exchange, and their analysis becomes more reliable [16]. /

RQ4. *What is the primary type of contribution?* – To characterize the main contribution of each primary study; algorithms, methods, models, metrics, and others.

RQ5. *How can the research be classified?* – To determine the type of research developed in each primary study, according both to the applied research methods and a classification scheme proposed in [14, 17].

2.2 Search Strategy, Sources, and Search String

For this SLM, we chose the *automatic search* method over studies' sources available through the "Portal de Periódicos CAPES"[1] website. In essence, the sources should meet the following criteria: relevant content indexed according to the research topic, including the Health area; and a Web-based search mechanism over studies' abstracts (at least), with support of boolean operators and temporal filtering, and multiple formats of searches' results exporting.

[1] Available at http://www.periodicos.capes.gov.br/.

As a result, the sources chosen for this SLM include the following search engines and digital libraries: *ACM Digital Library*, *Engineering Village*, *IEEE Xplorer*, *PubMed*, and *Scopus*. Yet related to the search strategy, we decided to select *English*-written papers only, published in the last decade, i.e., *from 2010 to 2019*.

To support the definition of standardized terms about ontology-supported EH, the search terms are borrowed from specialized literature in EH [18]. The following is the set of keywords and synonyms candidate for the definition of the search string: *environment*, *environmental*, *air*, *water*, *noise*, *sound*, *soil*, *pollution*, *sanitation*, *chemical*, *biological*, *radiation*, *climate*, *health*, *ontology*, *ontologies* and *ontological*. The goal is to identify studies addressing ontologies and any type of pollution (e.g., air, noise, and water) or other environmental factors that affect human health.

A pilot search with those candidate terms should allow us to achieve a search string balancing comprehensiveness and accuracy. After several attempts searching on the Scopus (due to its broad scope of knowledge fields), we decided not to exclude generic terms (e.g., *environment* and *health*) that led us to obtain many false positives. The reason is that the exclusion of these terms also discarded relevant papers after reading their abstracts during the pilot test. After all these observations, we elaborated on the following final search string:

(*environment* OR *environmental* OR *air* OR *water* OR *noise* OR *sound* OR *soil* OR *pollution* OR *sanitation* OR *chemical* OR *biological* OR *radiation* OR *climate*) AND *health* AND (*ontology* OR *ontologies* OR *ontological*)

When necessary, the search string was tailored in response to the search capabilities of each study source. In some sources, for example, variations of the term *ontology* and *environment* are necessary. Owing to indexing a broader collection of papers, the search scope of the ACM Digital Library must be configured to *The ACM Guide to Computing Literature* option.

2.3 Selection Criteria

The initial selection of studies based on selection criteria firstly relies on the reading of papers' metadata (title, abstract, and keywords). Whenever a paper is excluded by one exclusion criterion (EC) at least, it is removed from this SLM. Otherwise, the paper should be included by one inclusion criterion (IC). The following is the list of exclusion criteria we defined for this SLM:

EC1. The study is a preliminary or simplified version of another published elsewhere.

EC2. The study does not address Environmental Health.

EC3. The study does not describe computing research.

EC4. The study does not address ontologies in Environmental Health.

EC5. The study is not a journal article or a conference paper.

EC6. The full-text of the paper is not available.

EC7. The full-text of the paper is not English-written.

EC8. The publication year of the paper is before 2010 or after 2019.

We also defined eight inclusion criteria (IC) primarily for studies classification purposes. Those papers we could not classify into a specific criterion (IC1-IC7) based on the reading of metadata, we associate them with a general inclusion option (IC8). Afterward, the reading of the full-text of those papers (i.e., in the extraction activity) included by the IC8 criterion can reallocate them to one of the other seven inclusion criteria. The following is the list of inclusion criteria we defined for this SLM:

IC1. A primary study on ontologies and air pollution.
IC2. A primary study on ontologies and sound pollution.
IC3. A primary study on ontologies and water pollution.
IC4. A primary study on ontologies and soil pollution.
IC5. A primary study on ontologies and radioactive pollution.
IC6. A primary study on ontologies and chemical pollution.
IC7. A primary study on ontologies and biological pollution.
IC8. A primary study on ontologies and Environmental Health.

2.4 Conduction Phase

In this section, we detail the results from the automatic search, the duplicate elimination process and initial and final selections, by reading metadata and full-texts, respectively. We also overview the whole conduction phase with the respective number of primary studies in each activity.

Table 1 describes in detail the number of studies returned per source, including duplicates[2] as well as the corresponding number of non-duplicate studies[3]. As a result, 811 non-duplicate studies were retrieved and submitted to the screening through the selection criteria we describe next. The high number of duplicate papers found through automatic search (993 of 1804 = 55%) is mainly due to the non-predictable overlap among the indexed bibliographic databases regarding the ontology-supported EH research theme.

Table 1: The total number of studies returned per source.

Source	Identified	Non-duplicate
ACM DL	119	32
Engineering Village	887	243
IEEE Xplorer	235	127
PubMed	136	114
Scopus	427	295
Total	**1804**	**811**

As the initial selection activity relies on the reading and interpretation of papers' metadata only, we selected 42 candidate papers as relevant studies for the final selection based on the full-text analysis. In the data extraction activity, we read the full-text of 42 studies from which we excluded four papers more by the EC2 criterion and four others by EC4. We describe the process of data extraction of the 34 studies remaining in Section 3. In general, the main reasons for the removal of papers are as follows:

- many studies focus on different types of environments (e.g., clinical), but not on EH (111 by EC2);
- others address ontologies from the perspective of other knowledge areas[4] (e.g., Genetics) (96 by EC3);
- 63% of studies excluded (491 of 777) represent proposals, usage reports, or assessments of ontologies from a computing point of view, but not related to EH (EC4).

For the sake of transparency, Table 2 presents in detail the number of studies removed per exclusion criteria throughout the conducting phase, including initial and final selections activities.

Table 2: The total number of studies removed per exclusion criteria in the conduction phase.

Criterion	Initial selection	Final selection	Total
EC1	3	0	**3**
EC2	107	4	**111**
EC3	96	0	**96**
EC4	487	4	**491**
EC5	72	0	**72**
EC6	1	0	**1**
EC7	1	0	**1**
EC8	2	0	**2**
Total	**769**	**8**	**777**

3 DATA EXTRACTION

This section describes the data extraction process from the full-text-reading of the 34 relevant studies (S1 to S34) of this SLM shown in Table 3. We gathered the following data from each study: (i) the types of EH data addressed; (ii) the type of environmental data source; (iii) the type of MTCS support, if applicable; (iv) the primary type of contribution; (v) the research method and classification; and (vi) the respective venue and year of publication.

With information needed to answer the research questions RQ1 and RQ2, Table 4 shows the type of pollution related to EH data handled in each study as well as the type of environmental data source involved. We inserted a column labeled "Other" because the studies S13 and S16 follow a different approach.

In S13, the authors develop a semantic framework that aligns ontologies of different domains (e.g., environmental and human health) through a bridge ontology. As a general solution for enabling cross-domain search, S13 does not explicit associations to a particular EH data type, and it also assumes that EH data are accessible through the Web, regardless of the source of them.

In turn, S16 proposes a human health risk assessment ontology called *RsO*, which links the existing relevant health and environmental ontologies. As a general ontology, *RsO* is not associated with specific EH data types, but the utility of its concepts was validated with curated data on radiation and water pollution.

To answer the research question RQ3, we found six studies with MTCS support within 2010-2019, as shown in Table 5. Observe that two of them (i.e., S28 and S32) exploit more than one MTCS.

Regarding the research question RQ4, Table 6 highlights the main contribution of each relevant study. The contribution types we found are applications, frameworks, system or software architectures, and specific- or general-purpose ontologies. Observe that some studies have a dual contribution (e.g., S7–13).

[2]Search carried out on March, 2020.
[3]The *Parsif.al* tool controls the duplicate studies finding process; thus, there is no reason to assess the prevalence of a source over another based on these numbers only.
[4]Most of these papers came from automatic search over PubMed.
[5]International Program in Chemical Safety [53].

Table 3: The final list of studies that contributed to the data synthesis results.

ID	Paper title	Ref.
S1	A context-awareness architecture for managing thermal energy in an nZEB building	[19]
S2	AllergyLESS. An intelligent recommender system to reduce exposition time to allergens in smart-cities	[20]
S3	A methodology for indoor human comfort analysis based on BIM and ontology	[21]
S4	An effective inference method using sensor data for symbiotic healthcare support system	[22]
S5	An intelligent system to improve T-H-C parameters at the workplace	[23]
S6	An ontology based personal exposure history	[24]
S7	An ontology-based decision support framework for personalized quality of life recommendations	[25]
S8	An ontology-driven approach for integrating intelligence to manage human and ecological health risks in the ...	[26]
S9	An ontology for proactive indoor environmental quality monitoring and control	[27]
S10	A pervasive healthcare system for COPD patients	[28]
S11	A proposal for a computer-based framework of support for public health in the management of biological ...	[29]
S12	A smart space application to dynamically relate medical and environmental information	[30]
S13	Approaching cross-domain search in environmental applications - Towards Linked Data	[31]
S14	Arduino based system for indoor and outdoor ECG monitoring: Functions and extended user model ontology	[32]
S15	Broad, interdisciplinary science in tela: An exposure and child health ontology	[33]
S16	Building a human health risk assessment ontology (RsO): A proposed framework	[34]
S17	Enhancement of a body area network to support smart health monitoring at the digital home	[35]
S18	Environmental GIS to identify municipalities with high potential of biogas production in Mexico	[36]
S19	Health and environment monitoring service for solitary seniors	[37]
S20	Intelligent healthcare service based on context inference using smart device	[38]
S21	Leveraging ontology to enable indoor comfort customization in the smart home	[39]
S22	Linked data for air pollution monitoring	[40]
S23	Modeling of test specifications of raw materials in seafood ontology using Semantic Web Rule Language (SWRL)	[41]
S24	Ontology driven cross-linked domain data integration and spatial semantic multi criteria query system for ...	[42]
S25	Ontology-based approach to the discovery of human health and environmental risks assessment	[43]
S26	Ontology-based context aware for ubiquitous home care for elderly people	[44]
S27	Ontology-based model to support ubiquitous healthcare systems for COPD patients	[45]
S28	Personalized health knowledge graph	[46]
S29	Semantically enabling the SEMAT project: Extending marine sensor networks for decision support and hypothesis ...	[47]
S30	Semi automated transformation to owl formatted files as an approach to data integration: A feasibility study using ...	[48]
S31	The Apollo structured vocabulary: An OWL2 ontology of phenomena in infectious disease epidemiology and ...	[49]
S32	The Translational Medicine ontology and knowledge base: Driving personalized medicine by bridging the gap ...	[50]
S33	Towards semantically-enabled exploration and analysis of environmental ecosystems	[51]
S34	Using ontologies to relate resource management actions to environmental monitoring data in South East Queensland	[52]

Table 4: Types of environmental pollution and input data acquisition methods.

Pollution type	Sensor	Dataset	Simulation	Other
Air pollution	S1 S2 S4 S5 S8–S10 S12 S14 S17 S19–S21 S26–S28	S2 S4 S20 S22 S24 S27 S28 S30	S3 S7	–
Biological pollution	–	S23	S31	–
Chemical pollution	–	S25 S32	–	–
Radiation pollution	–	–	–	S16
Sound pollution	–	–	–	–
Soil pollution	–	S18	–	–
Water pollution	S29 S34	S33	–	S16
Unspecified	–	S11 S15	S6	S13

Given the research question RQ5, the thirty-four-study group was classified by the type of research developed, using the definition of Petersen et al. [14]. The classification by research type of a study depends on a number of conditions. Whenever an empirical evaluation is missing, the study is a solution proposal. On the other hand, whether the solution validated or evaluated is novel is not a key criterion. Both have to be empirically evaluated; however, validation is not used in practice, while evaluation studies take place in a real-world industrial context [14].

Table 5: Association between studies and MTCS.

MTCS	Studies identification
ICD-10	S24 S28
IPCS[5]	S16
SNOMED-CT	S27 S28 S30 S32
UMLS	S32

Table 6: Main contribution of the studies.

Contribution	Studies identification
Architecture	S1 S17
Framework	S13 S18 S26
Ontology	S7–S13 S15 S16 S20 S22 S24–S27 S31–S34
Application	S1–S12 S14 S17 S19–S23 S25 S28–S30 S32–S34

Table 7: Types of research and validation method.

Research	Method	Studies identification
Proposal	Proof of concept	S2 S7 S12 S13 S16 S28–S31 S33 S34
	No validation	S1 S14 S19 S23 S25–S27 S32
Validation	Case study	S8 S17 S18 S21 S22
	Experiment	S4 S5 S6 S9 S20
	Survey	S15 S24
	Simulation	S3 S10
	Prototyping	S11
Evaluation	Case study	–
	Experiment	–
	Survey	–

Wohlin et al.'s [17] classification of research methods has also been frequently reported in mapping studies. The method selection has to be consistent with the designation of the research type as of evaluation and validation. For instance, experiments, surveys, and case studies in academia are classified as validation research, while those are performed with industry practitioners would be classified as evaluation research. Finally, Table 7 characterizes the thirty-four-study group according to the criteria proposed by [14, 17].

4 DATA SYNTHESIS

This section presents a synthesis of the data extracted from the thirty-four relevant studies highlighted in Table 3. The goal here is to answer the research questions of this SLM.

4.1 Question 1

To answer the research question *"which type of EH data is addressed?"*, most of the studies (23 of 34) treat air pollution-related data, as shown in Table 4. Moreover, we identify there is no report of research on sound pollution with ontologies support.

We also analyzed the distribution of the thirty-four studies taking into account the environmental data type. We observed that at least one study is published about air pollution-based EH with ontologies support since 2013.

4.2 Question 2

Regarding the research question *"which type of environmental data source is exploited?"*, we observed that roughly 80% of the studies collect EH data from sensors and datasets, after analysis of Table 4. There is also a somewhat predominance (44.5%) of studies capturing EH data through physical or virtual sensors in comparison with through datasets, and more than 80% of those focus on the air pollution effects on human health.

Even with the widespread of sensors in several fields of study, as in the development of EH applications, we believe that other data sources will continue for specific purposes. Public data sets cured by specialists and modeled by ontologies can support methods of associating probable diseases according to the readings of air pollution data regardless of physical sensors. These same data sets can serve as a basis for the development of simulation models that allow, for example, to anticipate, in experiments, the performance of different architectures of an EH system, still in the design phase, which can reduce risks, time and costs.

The following word cloud in Figure 2 summarizes **RQ1** and **RQ2**. It shows the predominance of Semantics, Sensors, Health and several other related terms, but it is interesting to notice that the only EH data type (related to **RQ1**) present is Air, and that the studies have indoor approaches, contributing with smart solutions in this purpose. The cloud is build upon the abstracts of all the 34 accepted studies.

Figure 2: Environmental data type and sources.

4.3 Question 3

The research question *"Which well-established MTCS supports EH research, and for what purpose?"* is one of the most important questions in this SLM because of the relevance of MTCS for the Health field. By providing a uniform and consensual interpretation of healthcare information in general, MTCS addresses the issues of miscommunication, erroneous understanding, and serious adverse effects in patients.

Table 5 reveals a small number of studies (6 of 34) making use of MTCS in ontology-supported EH research. Except for the case of S16, we found well-established MTCS, such as SNOMED-CT and ICD-10. The *RsO* ontology proposed in S16 borrows terms from IPCS [53], a WHO initiative for the standardization of a generic terminology for risk and exposure assessment in chemical hazards. S16 demonstrates the utility of RsO concepts with curations of risk assessments related to radiation and water pollution.

In S24, S27, and S28, ICD-10 and SNOMED-CT help with the semantic enrichment of diseases caused by air pollution and the integration of these EH data with specific-domain ontology proposals (S27 and S28) or with healthcare systems initiatives (S24). In turn, S32 presents a unifying semantic web ontology to integrate patient and biomedical data with diseases, treatments, and textual electronic health records. The UMLS thesaurus lends its generic terminology to the ontology in question, whereas SNOMED-CT adds more semantics to the clinical terms.

Despite the small number of studies exploiting MTCS's benefits, five ones date in the last five years (S16, S24, S27, S28, and S30). We believe that this may suggest a trend on the MTCS usage in EH research due to the intrinsic relation between MTCS and ontology knowledge representation. At the same time, more empirical investigation is required because only the S24 study describes validation research using MTCS (i.e., ICD-10).

4.4 Question 4

Regarding the research question *"what is the primary type of contribution?"*, applications and ontologies represent the most common result of 29 of 34 studies presented in Table 6. A small number of studies propose frameworks (S13, S18, and S26) or system and software architectures (S1 and S17).

Software frameworks hide low-level details of designers' and programmers' tasks and automate part of these tasks. Software architectures, in turn, are abstractions of design decisions that can be useful before the software is implemented. In both cases, a software team can concentrate more on how to meet software functional and non-functional requirements, with reduction of risks, costs, and the overall software development time.

In brief, the significant number of studies proposing ontology-based EH applications can guide the proposal of new research on software frameworks and architectures on this topic. For instance, common functionalities, data and service design decisions, security and privacy concerns, among others.

4.5 Question 5

The research question *"how can the research be classified?"* is closely related to the study's maturity. Table 7 shows that 56% of studies (19 of 34) are solution proposals, and a significant part of these does not acquire EH data from physical sensors, even if the main contribution is a software application. Despite the fact of 44% of the studies remaining are of validation type, only five of them (S4–S6, S9, and S20) are academic-based experiments, i.e., not in real-world scenarios.

Moreover, the absence of empirical results based on real-world settings (evaluation research) makes us conclude that ontology-supported EH research is not a mature research area yet.

The following Figure 3 presents the amount of studies by year, stacking the publication venues. This shows the area usually rather publish in conferences proceedings.

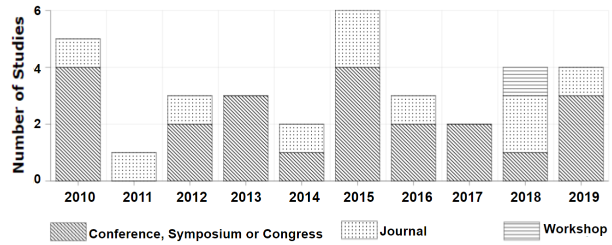

Figure 3: Publications and Venues used per Year.

Finally, the bubble chart in Figure 4 presents a mapping for the contributions and environmental data present in the studies. It combines the information in Table 4, Table 6 and Table 7 to make explicit the number, and thus preference, of studies that are solution proposals or validation research in air pollution, proposing or validating applications and ontologies, mostly.

5 TRENDS, GAPS, AND FUTURE WORK

The process of finding relevant research on a topic is a significant issue in systematic studies. We carried out some procedures to mitigate the potential threats to the validity of this SLM.

We made a search string with terms of specialized literature in EH and performed an automatic search over five sources embracing research on both Computing and Health fields. Planning and conduction followed the *one reviewer - one evaluator* approach, as discussed in [54]. A research leader performed the planning and verified the results of searching, selection, extraction, and synthesis carried out by the postgraduate student, as a reviewer. This approach could be improved if a third person was available for decision makings. Before the development of this SLM, we also elaborated on an extensive search for systematic literature studies, but none has been found.

Therefore, we contribute to the state-of-art with the protocol[6] and the results of this first systematic literature mapping on environmental health with ontological support. Synthesis results allow us to point out the following trends and gaps for further research:

- air pollution has been of great relevance for the research community, mainly within 2015-2019, but we did not find research on sound pollution with ontologies support;
- sensor-based EH data acquisition has been widely used, but the increasing adoption of sensors should also promote the use of public datasets to support EH research;
- a small, but an increasing number of studies taking advantage of the EH-related information integration benefits provided by MTCS within the last five years;
- and EH research should surpass the solution proposal barrier and present more mature empirical results using ontology-oriented frameworks and architectures.

[6]The entire protocol documentation and the data extracted from the thirty-four studies is available for public usage at https://bit.ly/3rC8n6w.

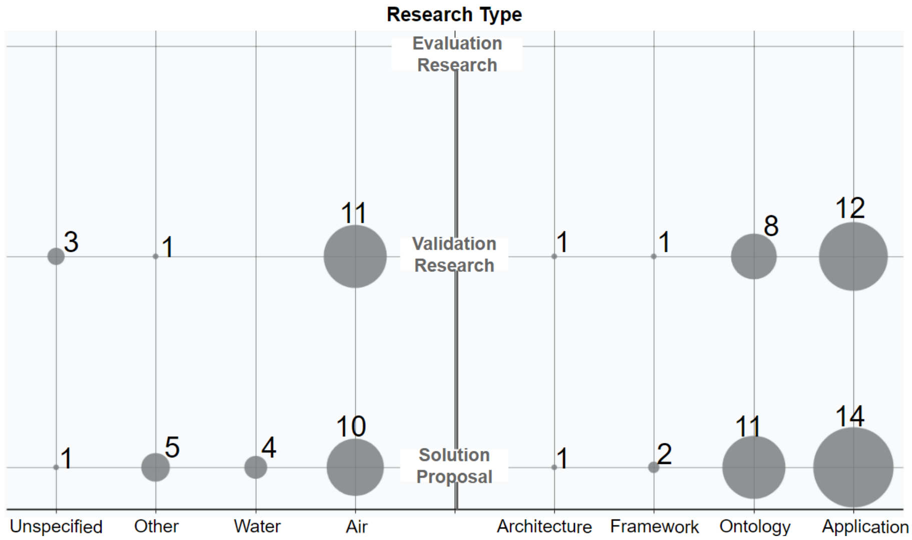

Figure 4: Mapping of contributions and acquisition means.

These results have geared our current and future work. We are investigating the effects of internal air pollution on human health by developing physical sensors and software architecture based on primary requirements healthcare information systems should consider. These include sensor-based data acquisition, ontology-based data representation, and information enrichment and integration supported by the valuable vocabulary of MTCS.

6 ACKNOWLEDGMENT

This study was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Brasil (CAPES) – Finance Code 001.

REFERENCES

[1] Maged Kamel Boulos and Najeeb Al-Shorbaji. On the internet of things, smart cities and the who healthy cities. *International journal of health geographics*, 13: 10, 03 2014.
[2] World Health Organization. World health organization who. action plan for environmental health services in central and eastern europe and the newly independent states. http://www.euro.who.int/__data/assets/pdf_file/0003/136227/EUR_ICP_CEH_212A.pdf, oct 1993.
[3] World Health Organization. Environmental health. http://origin.searo.who.int/topics/environmental_health/en/, 2015.
[4] Stefan Reis, Edmund Seto, Amanda Northcross, Nigel W.T. Quinn, Matteo Convertino, Rod L. Jones, Holger R. Maier, Uwe Schlink, Susanne Steinle, Massimo Vieno, and Michael C. Wimberly. Integrating modelling and smart sensors for environmental and human health. *Environmental Modelling and Software*, 74:238 – 246, 2015. ISSN 1364-8152.
[5] Rui Pitarma, Gonçalo Marques, and Bárbara Roque Ferreira. Monitoring indoor air quality for enhanced occupational health. *J. Med. Syst.*, 41(2):1–8, February 2017. ISSN 0148-5598.
[6] C. Bran, C. Hernández, and A. Gómez. System based on sensor dots for monitoring smart buildings. In *2019 IEEE 39th Central America and Panama Convention (CONCAPAN XXXIX)*, pages 1–6, Guatemala City, Guatemala, Guatemala, Nov 2019. IEEE.
[7] Charith Perera, Arkady B. Zaslavsky, Peter Christen, and Dimitrios Georgakopoulos. Context aware computing for the internet of things: A survey. *IEEE Commun. Surv. Tutorials*, 16(1):414–454, 2014.
[8] M. Bermudez-Edo, T. Elsaleh, P. Barnaghi, and K. Taylor. Iot-lite: a lightweight semantic model for the internet of things and its use with dynamic semantics. *Personal and Ubiquitous Computing*, 21:475–487, 2017.
[9] Ernesto Fonseca Veiga, Lucas Balbino de Melo Ferreira, Mayke Ferreira Arruda, Alessandra Alaniz Macedo, and Renato de Freitas Bulcão-Neto. A lightweight mobile service for context representation through an iot-oriented ontology. In *Anais Principais do XXIV Simpósio Brasileiro de Multimídia e Web*, pages 299–306, Porto Alegre, RS, Brasil, 2018. SBC.
[10] V. M. Tayur and R. Suchithra. A comprehensive ontology for internet of things (coiot). In *2019 Second International Conference on Advanced Computational and Communication Paradigms (ICACCP)*, pages 1–6, Gangtok, Sikkim, India, 2019. IEEE.
[11] Tarek Elsaleh, Shirin Enshaeifar, Roonak Rezvani, Sahr Acton, Valentinas Janeiko, and María Bermúdez-Edo. Iot-stream: A lightweight ontology for internet of things data streams and its use with data analytics and event detection services. *Sensors*, 20:953, 02 2020.
[12] Thomas R. Gruber. Toward principles for the design of ontologies used for knowledge sharing? *International Journal of Human-Computer Studies*, 43(5):907 – 928, 1995. ISSN 1071-5819.
[13] Barbara A. Kitchenham, David Budgen, and O. Pearl Brereton. Using mapping studies as the basis for further research - A participant-observer case study. *Inf. Softw. Technol.*, 53(6):638–651, 2011.
[14] Kai Petersen, Sairam Vakkalanka, and Ludwik Kuzniarz. Guidelines for conducting systematic mapping studies in software engineering: An update. *Information and Software Technology*, 64:1 – 18, 2015. ISSN 0950-5849.
[15] Hannah Harrison, Simon J. Griffin, Isla Kuhn, and Juliet A. Usher-Smith. Software tools to support title and abstract screening for systematic reviews in healthcare: an evaluation. *BMC Med Res Methodol*, 20(7):1–12, 2020.
[16] David Gefen, Jake Miller, Johnathon Kyle Armstrong, Frances H. Cornelius, Noreen Robertson, Aaron Smith-McLallen, and Jennifer A. Taylor. Identifying patterns in medical records through latent semantic analysis. *Communications of the ACM*, 61(6):72–77, June 2018.
[17] Claes Wohlin, Per Runeson, Martin Hst, Magnus C. Ohlsson, Bjrn Regnell, and Anders Wessln. *Experimentation in Software Engineering*. Springer, Berlin, Heidelberg, 2012. ISBN 3642290434.
[18] World Health Assembly. Health, environment and climate change: draft who global strategy on health, environment and climate change: the transformation needed to improve lives and well-being sustainably through healthy environments. Technical Report A72/15, World Health Organization (WHO), Geneva, Switzerland, apr 2019.
[19] O. H. Uribe, M. Santos, M. C. Garcia-Alegre, and D. Guinea. A context-awareness architecture for managing thermal energy in an nzeb building. In *2015 IEEE First International Smart Cities Conference (ISC2)*, pages 1–6, Guadalajara, Mexico, Oct 2015. IEEE.
[20] José Antonio García-Díaz, José Ángel Noguera-Arnaldos, María Luisa Hernández-Alcaraz, Isabel María Robles-Marín, Francisco García-Sánchez, and Rafael Valencia-García. Allergyless. an intelligent recommender system to reduce exposition time to allergens in smart-cities. In Fernando De La Prieta, Sigeru Omatu, and Antonio Fernández-Caballero, editors, *Distributed Computing and Artificial Intelligence, 15th International Conference*, pages 61–68, Cham, 2019. Springer International Publishing. ISBN 978-3-319-94649-8.

[21] W. Chen, K. Chena, V. J. L. Gan, and J. C. P. Cheng. A methodology for indoor human comfort analysis based on bim and ontology, 2019. Copyright - Copyright IAARC Publications 2019; Última atualização em - 2019-08-06.

[22] Satoru Izumi, Yusuke Kobayashi, Hideyuki Takahashi, Takuo Suganuma, Tetsuo Kinoshita, and Norio Shiratori. An effective inference method using sensor data for symbiotic healthcare support system. In *Proceedings of the 2010 International Conference on Computational Science and Its Applications - Volume Part IV*, ICCSA'10, page 152–163, Berlin, Heidelberg, 2010. Springer-Verlag. ISBN 3642121888.

[23] Martin Frešer, Anton Gradišek, Božidara Cvetkoviundefined, and Mitja Luštrek. An intelligent system to improve t-h-c parameters at the workplace. In *Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing: Adjunct*, UbiComp '16, page 61–64, New York, NY, USA, 2016. Association for Computing Machinery. ISBN 9781450344623.

[24] Stacy A. Doore, Kate Beard, and Carol Bult. An ontology based personal exposure history. In *Proceedings of the 1st ACM International Health Informatics Symposium*, IHI '10, page 674–683, New York, NY, USA, 2010. Association for Computing Machinery. ISBN 9781450300308.

[25] Marina Riga, Efstratios Kontopoulos, Kostas Karatzas, Stefanos Vrochidis, and Ioannis Kompatsiaris. An ontology-based decision support framework for personalized quality of life recommendations. In *4th International Conference on Decision Support System Technology*, pages 38–51, Heraklion, Greece, 01 2018. Sringer-Verlag.

[26] Xiaoliang Meng, Feng Wang, Yichun Xie, Guoqiang Song, Shifa Ma, Shiyuan Hu, Junming Bai, and Yiming Yang. An ontology-driven approach for integrating intelligence to manage human and ecological health risks in the geospatial sensor web. *Sensors*, 18(11):3619, 2018.

[27] Jude A. Adeleke and Deshendran Moodley. An ontology for proactive indoor environmental quality monitoring and control. In *Proceedings of the 2015 Annual Research Conference on South African Institute of Computer Scientists and Information Technologists*, SAICSIT '15, New York, NY, USA, 2015. Association for Computing Machinery. ISBN 9781450336833.

[28] Hicham Ajami, Hamid Mcheick, and Karam Mustapha. A pervasive healthcare system for copd patients. *Diagnostics*, 9(135):1–35, 10 2019.

[29] Vladimír Bureš, Tereza Otčenášková, Pavel Čech, and Karel Antoš. A proposal for a computer-based framework of support for public health in the management of biological incidents: the czech republic experience. *Perspectives in Public Health*, 132(6):292–298, 2012.

[30] F. Vergari, S. Bartolini, F. Spadini, A. D'Elia, G. Zamagni, L. Roffia, and T. Salmon Cinotti. A smart space application to dynamically relate medical and environmental information. In *2010 Design, Automation Test in Europe Conference Exhibition (DATE 2010)*, pages 1542–1547, Dresden, Germany, March 2010. IEEE.

[31] Gerald Schimak, Luca Petronzio, and Tomás Pariente Lobo. Approaching cross-domain search in environmental applications - towards linked data. In *Environmental Software Systems. Fostering Information Sharing. ISESS 2013.*, volume 413, pages 71–84, Neusiedl am See, Austria, 10 2013. Springer-Verlag.

[32] Carmelo Pino and Alfio Costanzo. Arduino based system for indoor and outdoor ECG monitoring - functions and extended user model ontology. In Andreas Holzinger, Stephen H. Fairclough, Dennis Majoe, and Hugo Plácido da Silva, editors, *PhyCS 2014 - Proceedings of the International Conference on Physiological Computing Systems*, pages 331–335, Lisbon, Portugal, 2014. SciTePress.

[33] James P. McCusker, Sabbir M. Rashid, Zhicheng Liang, Yue Liu, Katherine Chastain, Paulo Pinheiro, Jeanette A. Stingone, and Deborah L. McGuinness. Broad, interdisciplinary science in tela: An exposure and child health ontology. In *Proceedings of the 2017 ACM on Web Science Conference*, WebSci '17, pages 349–357, New York, NY, USA, 2017. ACM. ISBN 978-1-4503-4896-6.

[34] Thomas E. McKone and Lydia Feng. Building a human health risk assessment ontology (rso): A proposed framework. *Risk Analysis*, 35(11):2087–2101, 2015.

[35] Laura Vadillo, Miguel A. Valero, and Gema Gil. Enhancement of a body area network to support smart health monitoring at the digital home. In *Proceedings of the 8th International Conference on Body Area Networks*, BodyNets '13, page 213–216, Brussels, BEL, 2013. ICST (Institute for Computer Sciences, Social-Informatics and Telecommunications Engineering). ISBN 9781936968893.

[36] C. Ambriz, N. Sánchez, F. Mata, M. Torres, and D. Castro-Frontana. Environmental gis to identify municipalities with high potential of biogas production in mexico. In *2016 IEEE 1er Congreso Nacional de Ciencias Geoespaciales (CNCG)*, pages 1–4, Mexico City, Mexico, 2016. IEEE.

[37] K. Kim, E. Lee, S. Kwon, D.-H Park, and S.-I Jin. Health and environment monitoring service for solitary seniors. In *Proceedings of the ISWC 2014 Posters and Demonstrations Track a track within the 13th International Semantic Web Conference (ISWC 2014)*, volume 1383, pages 1–2, Riva Del Garcia, Trentino, Italy, 01 2014. CEUR.

[38] Ilsun You, Junho Choi, Chang Choi, and Pankoo Kim. Intelligent healthcare service based on context inference using smart device. *Soft Computing*, 18(12): 2577–2586, Dec 2014. ISSN 1433-7479.

[39] Daniele Spoladore, Atieh Mahroo, and Marco Sacco. Leveraging ontology to enable indoor comfort customization in the smart home. In Alfredo Cuzzocrea, Sergio Greco, Henrik Legind Larsen, Domenico Saccà, Troels Andreasen, and Henning Christiansen, editors, *Flexible Query Answering Systems*, pages 63–74, Cham, 2019. Springer International Publishing. ISBN 978-3-030-27629-4.

[40] M. U. H. A. Rasyid, I. Syarif, and I. A. H. Putra. Linked data for air pollution monitoring. In *2017 International Electronics Symposium on Knowledge Creation and Intelligent Computing (IES-KCIC)*, pages 65–70, Surabaya, Indonesia, Sep. 2017. IEEE.

[41] P. V. Vinu, P. C. Sherimon, and Reshmy Krishnan. Modeling of test specifications of raw materials in seafood ontology using semantic web rule language (swrl). In *Proceedings of the 2015 International Conference on Advanced Research in Computer Science Engineering and Technology (ICARCSET 2015)*, ICARCSET '15, New York, NY, USA, 2015. Association for Computing Machinery. ISBN 9781450334419.

[42] Sunitha Abburu. Ontology driven cross-linked domain data integration and spatial semantic multi criteria query system for geospatial public health. *Int. J. Semant. Web Inf. Syst.*, 14(3):1–30, July 2018. ISSN 1552-6283.

[43] Jiri Hrebicek, Ladislav Duvsek, Jiři Kalina, Miroslav Kubasek, Ivan Holoubek, and Jana Klanova. Ontology-based approach to the discovery of human health and environmental risks assessment. In *6th International Congress on Environmental Modelling and Software (iEMSs)*, pages 1–8, Leipzig, Germany, 2012. UFZ.

[44] Kurnianingsih, L. E. Nugroho, Widyawan, L. Lazuardi, and K. Non-alinsavath. Ontology-based context aware for ubiquitous home care for elderly people. In *2015 2nd International Conference on Information Technology, Computer, and Electrical Engineering (ICITACEE)*, pages 454–459, Semarang, Indonesia, 2015. IEEE.

[45] Hicham Ajami and Hamid Mcheick. Ontology-based model to support ubiquitous healthcare systems for copd patients. *Electronics*, 7(12):371–399, 2018.

[46] Amelie Gyrard, Manas Gaur, Krishnaprasad Thirunarayan, Amit Sheth, and Saeedeh Shekarpour. Personalized health knowledge graph. In *Joint Proceedings of the International Workshops on Contextualized Knowledge Graphs, and Semantic Statistics (CKGSemStats)*, number 2317 in CEUR Workshop Proceedings, pages 1–6, Monterey, USA, 2018. CEUR.

[47] T. Myers, I. Atkinson, and R. Johnstone. Semantically enabling the semat project: Extending marine sensor networks for decision support and hypothesis testing. In *2010 International Conference on Complex, Intelligent and Software Intensive Systems*, pages 974–979, Krakow, Poland, 2010. IEEE.

[48] Shao Liang, Adel Taweel, Simon Miles, Yevgeniya Kovalchuk, Anastassia Spiridou, Ben Barratt, Uyen Hoang, S Crichton, Brendan Delaney, and Charles Wolfe. Semi automated transformation to owl formatted files as an approach to data integration a feasibility study using environmental, disease register and primary care clinical data. *Methods of information in medicine*, 53, 06 2014.

[49] William Hogan, Michael Wagner, Mathias Brochhausen, John Levander, Shawn Brown, Nicholas Millett, Jay DePasse, and Josh Hanna. The apollo structured vocabulary: an owl2 ontology of phenomena in infectious disease epidemiology and population biology for use in epidemic simulation. *Journal of Biomedical Semantics*, 7:50, 08 2016.

[50] Joanne Luciano, Bosse Andersson, Colin Batchelor, Olivier Bodenreider, Tim Clark, Christine Denney, Christopher Domarew, Thomas Gambet, Lee Harland, Anja Jentzsch, Vipul Kashyap, Peter Kos, Julia Kozlovsky, Timothy Lebo, Scott Marshall, James McCusker, Deborah Mcguinness, Chimezie Ogbuji, Elgar Pichler, and Michel Dumontier. The translational medicine ontology and knowledge base: Driving personalized medicine by bridging the gap between bench and bedside. *Journal of biomedical semantics*, 2 Suppl 2:S1, 05 2011.

[51] P. Wang, L. Fu, E. W. Patton, D. L. McGuinness, F. J. Dein, and R. S. Bristol. Towards semantically-enabled exploration and analysis of environmental ecosystems. In *2012 IEEE 8th International Conference on E-Science*, pages 1–8, Chicago, IL, USA, 2012. IEEE.

[52] Jane Hunter, Peter Becker, Abdulmonem Alabri, Catharine Ingen, and Eva Abal. Using ontologies to relate resource management actions to environmental monitoring data in south east queensland. *International Journal of Agricultural and Environmental Information Systems*, 2:1–18, 01 2011.

[53] World Health Organization. International program in chemical safety (ipcs). risk assessment terminology; part1: Ipcs/oecd key generic terms used in chemical hazard/risk assessment; part 2: Ipcs glossary of key exposure assessment terminology. Technical report, World Health Organization (WHO), Geneva, Switzerland, 2004.

[54] Barbara Kitchenham and Stuart Charters. Guidelines for performing systematic literature reviews in software engineering. Technical Report EBSE 2007-001, Keele University and Durham University Joint Report, 2007.