RESEARCH ARTICLE

Towards an information centric flood ontology for information management and communication

Yusuf Sermet[1] · Ibrahim Demir[1,2]

© Springer-Verlag GmbH Germany, part of Springer Nature 2019

Abstract

Advancements and new techniques in information technologies are making it possible to manage, analyze and present large-scale environmental modeling results and spatial data acquired from various sources. However, it is a major challenge to make this data accessible because of its unstructured, incomplete and varied nature. Extracting information and making accurate inferences from various data sources rapidly is critical for natural disaster preparedness and response. Critical information about disasters needs to be provided in a structured and easily accessible way in a context-specific manner. This paper introduces a group of information-centric ontologies that encompass the flood domain and describes how they can be benefited to access, analyze, and visualize flood-related data with natural language queries. The presented methodology enables the easy integration of domain knowledge into expert systems and voice-enabled intelligent applications that can be accessed through web-based information platforms, instant messaging apps, automated workflow systems, home automation devices, and augmented and virtual reality platforms. A case study is described to demonstrate the usage of presented ontologies in such intelligent systems.

Keywords Intelligent systems · Flood ontology · Information system ontology · Disaster communication · Disaster preparedness · Knowledge generation

Introduction

In a world with non-stop technological advancements, large streams of data are constantly being generated (Wu et al. 2014; Demir et al. 2015) and require effective tools for management, analysis, and communication (Demir and Szczepanek 2017). Information systems allow users to access, interpret and explore data and information quickly and effectively while offering a variety of capabilities for decision making (Peppard and Ward 2016; Demir and Beck 2009; Weber et al. 2018; Jones et al. 2018). Expert systems were used extensively for the purpose of replicating the decision-making capabilities of experts while possessing the knowledge that is essential to the

reasoning process (Jackson 1998; Liang et al. 2005). They consist of a knowledge base which stores factual and experimental information related to a particular subject, and an inference component which can apply human-like reasoning (Liao 2005). Inference components use the information available in the knowledge base to solve complex real-life problems. There are various examples of expert system applications in commercial environments, including Apple Siri, Google Now, and Wolfram Alpha, which are computational knowledge engines that answers factual queries directly by computing the answer from a knowledge base of externally sourced curated and structured data (Apacible et al. 2013).

Human-like interaction using natural language processing is vital for expert systems. As opposed to conventional keyword-based search methods, a natural language search engine has the ability to produce a direct answer to a given question. Natural language search systems aim to determine the intent of the question by analyzing ontological components.

Considering the needs of knowledge-based systems, structured and curated data should be provided in the form of an ontology. In the context of computer science, an ontology can be defined as "a formal, explicit specification of a shared

✉ Yusuf Sermet
   muhammedyusuf-sermet@uiowa.edu

1   Department of Electrical and Computer Engineering, University of Iowa, Iowa City, IA, USA

2   Department of Civil and Environmental Engineering, University of Iowa, Iowa City, IA, USA

conceptualization" (Studer et al. 1998). Data should be represented in a comprehensive domain model which accurately implements a consensual vocabulary along with conceptual relationships for shared understanding. As opposed to relational databases schemas, ontologies are easier to develop and maintain as the scope of the ontology widens. Ontologies have a strong focus on semantics; thus they allow the inference of new and implicit information. In the environmental domain where utilization of ever-growing sensor data to the fullest extent depends on easy accessibility and proper structuring (Demir and Krajewski 2013), ontologies are essential to effectively infer knowledge from data by allowing intelligent systems to comprehend the conceptuality and causality of real-world phenomena.

This paper presents a group of information-centric ontologies that encompass the flood domain while enabling the extension to other disasters. These ontologies are used as the semantic foundation of environmental information systems that allow users to access, analyze, and visualize flood-related data with natural language queries. In creating these ontologies, various resources were utilized, and domain experts were consulted to achieve a consensual representation of the domain. A web platform was used in the design of the ontology so that it is intuitive to domain associates with a little technical background. The platform has the ability to extract the data model to make it parsable by web-based systems. This mechanism enables easy integration of domain knowledge into any expert system and supports voice-enabled question answering systems. The knowledge system can be accessed through web-based information systems, instant messaging systems, automated workflow systems, home automation devices, and augmented and virtual reality platforms.

One of the major contributions of this work is the introduction of the information-centric ontologies which describes domain ontologies designed specifically for use in modern information systems and data retrieval approaches in the environmental domain. The development of domain-specific knowledge generation systems requires the integration of domain knowledge with the capabilities and resources of an information system. The large variance in the functionalities and expected outcomes of platforms designed to access information makes it challenging for a domain ontology to comply with such terms. Building upon previous works, presented ontologies encapsulate the knowledge of a domain expert in flooding specifically for use in natural language understanding and question answering systems. Another contribution of this work is the presented workflow for developing information-centric ontologies that takes the social aspect of ontology development into account by considering the backgrounds of various stakeholders. Furthermore, the development principles of the proposed ontologies enable extensions and their adaptation to different domains.

The remainder of this article is organized as follows. Section 1.1 provides the literature review on relevant work. Section 2 presents the details of the development and implementation of the ontologies. Section 3 describes the Flood and Information System ontologies and discusses a case study application. Section 4 discusses the future work and provides conclusions.

Background

There has been extensive work on ontologies in the environmental domain including earthquake and wildfires (Kalabokidis et al. 2011; Xu et al. 2014; Abburu and Golla 2017). More specifically, Sui and Maggio (1999) have discussed the importance of high-level common ontologies for hydrological modeling and how they can play a key role in the integration of hydrological modeling and geographical information systems. Bermudez and Piasecki (2003) created an OWL-based Hydrological Ontology for the Web (HOW) which aims to form a consensual vocabulary between the hydrologic communities. Tripathi and Babaie (2008) modified and extended the top-level SWEET ontologies to produce a domain-specific hydrogeology ontology. Hydroseek is an ontology-aided search engine (Beran and Piasecki 2009) that allows users to perform queries using keywords in the areas of water quality, meteorology, and hydrology. CUAHSI designed a hydrologic ontology to provide a common understanding of the shared domain and allow a keyword-based search feature in the Hydrologic Information System (CUAHSI-HIS).

In the area of natural disasters, SemSorGrid4Env software architecture (Gray et al. 2010) describes its ontology network covering the fire and flood domains by reusing many of the ontologies in the SWEET suite to cover fire, forest and vegetation, weather, and geography domains. Kalabokidis et al. (2011) presented OntoFire, an ontology-based geo-portal on wildfires, which is the first geo-portal in the area of wildfires that relies on the semantic relationships between the resources utilizing an ontology. Babitski et al. (2009) developed a group of integrated ontologies founded on DOLCE ontology to cover the emergency management and risk assessment aspect of natural disasters. Kollarits et al. (2009) presented MONITOR risk management ontology to serve as a reference ontology for the management of natural hazards. Wang and Stewart (2014) presented an approach that automatically extracts disaster-related information from the web using NLP and geographical information retrieval methods. In their work, a hazard ontology was constructed on the NeOn (http://neon-toolkit.org/) platform using several resources including FEMA, The U.S. National Weather Service, and existing ontologies.

More specifically, in the area of flooding, there have been several studies that describe the development of flood

ontologies from a domain perspective (Garrido et al. 2012; Agresta et al. 2014; Kurte et al. 2017). Dolbear et al. (2005) presented an application ontology for flood risk management to be used by governmental agencies. Katuk et al. (2009) presented an architecture which demonstrated how the use of ontologies in web-based systems can be beneficial in flood response efforts in Malaysia. Mongula (2009) showed how ontologies in enterprise level can be useful in flood control with case studies of both The Netherlands and The United States. Scheuer et al. (2013) proposed a flood risk assessment ontology founded upon SWEET and MONITOR ontologies to put local and expert knowledge into operation.

Methods

An ontology should capture all concepts, attributes, and relationships for modeling an intended phenomenon, alongside with defining the necessary rules and constraints explicitly. To formalize the ontology, it should be implemented in a machine-readable format instead of using natural language. Because ontology aims at consensual knowledge, it should be the result of a collaboration rather than reflecting the perspective of an individual. To accomplish this goal, the best practice is to involve stakeholders from different backgrounds in the ontology development process.

Different approaches for designing and developing an ontology are available and discussed in the literature (Uschold and King 1995; De Nicola et al. 2009). However, there is no universally agreed-upon methodology for ontology development. The objective and scope of the ontology affect the development process substantially. Ontologies can broadly be categorized as Generic Ontologies, Upper-Level Ontologies, Domain Ontologies, Task and Method Ontologies, and Application Ontologies, based on their objective. Application ontologies are best described as a unification of domain and task ontologies (Guarino 1998). Along with possession of the knowledge and vocabulary for a certain field, it is also constructed in a way that allows problem-solving goals. A simple example of a problem-solving goal might be to analyze the CO_2 level in the Iowa River. To satisfy this analysis goal, an ontology should consist of the relevant knowledge, which is the carbon dioxide level in a river channel in this case. The ontology also should be able to interpret "the analysis of carbon dioxide levels" as the consensual understanding in the domain which agreed upon.

Purpose and scope definition

Several application scenarios for ontologies in flooding are listed in Table 1. These use cases facilitate the process to ensure accurate and appropriate encompassing of the flood domain. In some scenarios, ontologies are utilized to process natural language questions and map them to appropriate queries and resources, whereas in others, ontologies merely serve as a consensual data model between previously agreed stakeholders.

Competency questions (CQ) are used to capture the scope of an ontology by clarifying what questions ontology should be able to answer. From the software engineering perspective, competency questions act like acceptance tests. They provide a clear and concise way to evaluate and validate the ontology as it expands. The definition of a competency question is extended to comply with the scope of the ontology regarding its usage in information systems. In this work, a CQ can be used as the natural language input that the system in the corresponding application scenario should process and act on rather than being a question that is directly answered by the ontology. Table 2 lists example CQs for different scenarios along with their potential answers.

Domain knowledge acquisition

The initial collection of ontology concepts was elicited from the competency questions, previous work on flood and disaster ontologies, and various online resources including books, research articles, and educational platforms. These resources are made available by organizations such as The National Oceanic and Atmospheric Administration (NOAA), The Federal Emergency Management Agency (FEMA), The United States Geological Survey (USGS), IIHR - Hydroscience & Engineering, Iowa Flood Center (IFC), and online knowledge systems (e.g. Wikipedia). Once the lexicon was generated, concepts and their respective attributes were defined in a class hierarchy. The top-down development process was utilized to create classes which advises to start building the hierarchy with the most general classes and specialize afterward. Although other approaches are available, a top-down approach may facilitate the domain expert's efforts by allowing them to see the whole picture rather than focusing on particular areas. For example, in Flood Ontology, *Instrument* is one of the top-level classes in the hierarchy and can be specialized to *Radar* which further specializes to *NEXRAD Radars*.

After the concepts and attributes are captured and expressed in a class hierarchy, conceptual relationships are generated. Relationships between ontological entities are vital to accurately interpret natural language input. For example, a user might ask "What is the travel time for water in the basin to reach my community at the outlet?" *Travel time* is an attribute of the *Basin* class, however, the basin is not a sub-class or an attribute of the *Community* class. A relationship and a rule should be defined in the ontology to incorporate the knowledge that every community is part of a watershed (i.e. basin). Domain experts are consulted during the ontology development process.

Table 1 An overview of the application scenarios and their descriptions and expectations

Application scenario	Description
Knowledge engine	An artificial intelligence system for flood-related natural language questions that can be accessed via web-based systems, chatting applications, home automation devices, and AR/VR platforms. (Sermet and Demir 2018a, 2018b)
Tweet analysis	A real-time social media analysis system to support flood preparedness, response, and monitoring using a comprehensive information-centric flood ontology and natural language processing. (Sit et al. 2019)
Markup language	A standard data exchange language in flood domain for organizations to share different data products (flood watch, alerts, advisories, warnings, and forecasts) with each other. (Holz et al. 2006)

Synonym enrichment Most of the terms in the lexicon are prevalently used synonyms or alternative expressions. For each concept and attribute, synonyms are elicited and defined with the goal of achieving consensus and providing maximum reach to stakeholders.

Definition and explanation enrichment A dictionary is created with definitions for the whole lexicon. In addition, procedural and mathematical explanations are provided where appropriate (e.g. river discharge: $Q = A/\overline{u}$, where Q is the discharge in m^3/s or ft^3/s, A is the cross-sectional area of flow in m^2 or ft^2, and $\overline{u}$ is the average velocity in m/s or ft/s).

Formalization and implementation

Ontology development is a collective process that should involve developers and stakeholders (e.g. domain experts) to maximize the integrity of the model. Domain experts in environmental issues may lack an advanced level of expertise in computer science. To facilitate expert interaction with the ontology and reduce time spent on the development, abstraction should be increased to free the experts from the technicality of ontology development. Furthermore, a group of domain experts may benefit from working together and consulting with each other. Allowing different parties to work on a single ontological model brings spatial challenges as some of

the interested parties may be located in another physical location. Offering an intermediary environment that supports simultaneous online access and editing of an ontology overcomes these challenges. Finally, the ontology should be capable of being converted to a format that software can make use of without any loss of knowledge. In Table 3, the requirements for an information-centric ontology development intermediate are listed.

Initially, we have evaluated many ontology development tools, including Protégé (i.e. Web and Desktop), a widely used ontology editor, by checking if they satisfy the requirements that were described in Table 3. As of today, most of the ontology development tools do not offer simultaneously accessed workplace and online development including Protégé Desktop (Stanford University 2019). Though various illustration and visualization options are available, visual editing is still limited. In contrast, Protégé Web provides an online and collaborative environment for ontology development with simultaneous access capabilities. However, it does not offer any illustration and visual editing options. Both versions of Protégé may be highly technical to some stakeholders with a little ontological background.

Considering the requirements in Table 3, an online modeling platform that features Unified Modelling Language (UML) to create class diagrams (i.e. GenMyModel) was utilized. UML is a standard of Object Management Group

Table 2 Example competency questions for different application scenarios

Application scenario	Competency question	Expected action
Knowledge engine	Which areas will be inundated in my community by a flood with a 10-year return period?	Map Output (i.e. System shows requested areas on the map)
	What is the definition of rating curve?	Textual Output (i.e. System returns the definition retrieved by the ontology or external resources)
	What will be the stream level forecast for the next 5 days?	Graphical Output (i.e. System displays a time series figure based on forecast model)
Tweet analysis	"Wow heavy rain is causing heavy traffic and roads to flood #drivesafe"	Classification (i.e. System infers that the tweet is reporting a natural event with potential to cause a flood.)
Markup language	What is the specification of flood forecasting model in your organization?	Structured Output (i.e. Parameters, inputs, assumptions about flood forecasting model)

Table 3 Requirements for ontology development tool selection

Requirement	Description
Illustration and visual editing	Ontology should be illustrated with graphical components to make it easy to understand. Editing feature should be enabled in the visualized model to decrease the fidelity to the implementation language.
Simultaneous access	Since ontology development involves many people in different fields, the ability to simultaneously work together is a necessity to maximize efficiency.
Online and intuitive development	Stakeholders from various domains are likely to be present in different places. Furthermore, the ontology should be a shared conceptualization which only can happen if it is committed by many experts in the domain. To minimize the meeting costs and make it possible to involve as many stakeholders as possible, online development is vital.
Machine readable representation	One of the most fundamental requirements for an ontology is that it should be implemented in a machine-readable format. The medium chosen to accomplish visual representations with editing features for the development process should allow translation of the ontology to a formal language.

(OMG) and widely used in software design. Unlike major programming languages (e.g. Java, C++), UML is created to visually model a software system via behavioral (e.g. activity diagram, sequence diagram, use case diagram) and structural diagrams (e.g. class diagram, package diagram). A class diagram is a static structural representation of a system. It consists of classes with respective attributes. Relationships between classes can be represented by associations. Multiplicity can be specified for each association. A special case of association is the generalization (a.k.a. is-a relationship) which represents inheritance. Furthermore, constraints can be specified to assert restrictions that must be satisfied for any realization of the model. Adapting UML to describe ontologies had been studied thoroughly and research (Kogut et al. 2002; Djurić et al. 2005) shows that complete conversions are possible despite some difficulties (e.g. mapping of ontology individuals or properties to UML). In this work, ontology representation in UML was approached from a more practical perspective and a complete translation was not intended.

Benefiting from an online collaborative UML tool satisfies most requirements. In order to achieve the goal of powering information systems, the ontology should be represented in a parsable intermediate language. XML Metadata Interchange (XMI) is also a standard of OMG which can be used to represent UML class diagrams in terms of Extensible Markup Language (XML) elements and attributes (OMG 2015). Parsing XML files has a wide support in major programming languages which eliminates the need for a translation module that converts the visual model of the ontology to a structured representation in the relevant language. Open-source ready-to-use libraries for parsing XML files are available in client-side languages (e.g. JavaScript), server-side scripting languages (e.g. PHP), and many prevalent languages (e.g. C++, Python, C#, Java). For example, server-side libraries (e.g. SimpleXML) can parse the textual content of an XML document and outputs an iterable collection of arrays and objects consist of the ontological elements.

Considering the scope of the ontology defined by the application scenarios, ontology can be represented in terms of relational databases to facilitate access by online software systems. Five relational database tables were created using the data structures produced by an XML parser for demonstration. Concept table consists of *concept id*, *name*, and *parent id* fields. In the case of multiple inheritance, *parent id* field accepts an array of ids. Attribute table defines each attribute's id, name, and id of the concept whose state represented by the attribute. Relationship table encompasses the ids of two related concepts along with the name, id, and multiplicity of the association. Having similar structures, synonym and description tables record synonyms, definitions, and explanations for ontological concepts, attributes, and relationships. Three fields keep track of the ontological element type, the id of the element, and respective information. Figure 1 summarizes the transitions and processes in the implementation of the ontology.

Revision of the ontology

Designing a domain-specific ontology requires the active participation of domain experts to keep up with changing standards. As the number of parties that are committed to the ontology development increases, different requirements and relationships can arise. Thus, periodic revision of the ontology is needed to assure completeness and integrity of the conceptualization. Researchers and professionals in the environmental field may not have time to learn nuances of an ontology in the artificial intelligence context. Benefiting from an online and collaborative UML design tool makes the technicalities of the implementation trivial to domain experts. The mechanism will allow experts and professionals to make suggestions by modifying the ontological elements, naming preferences and synonyms, description and method definitions. These suggestions will be reviewed by an advisory committee and applied in case of approval.

Fig. 1 Ontology development for use in web-based information systems

Visual Model (UML)   Processing Module (PHP)   Database Representation (PostgreSQL)
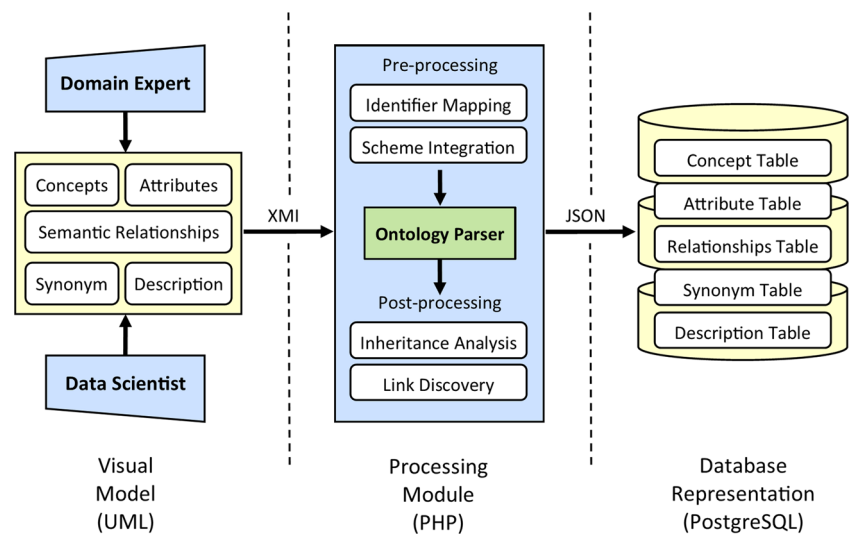
Results

Flood ontology

Flood ontology was developed under three major ontological branches (i.e. Natural Hazard, Instrument, and Environmental Phenomena) which are detailed in three paragraphs in this section, respectively. Definition and scope of natural hazards should be specified for accurate classification. A naturally occurring phenomenon that affects a human population and/or causes fatalities and economic damage was categorized as a natural hazard (OAS 1990). Given the broad scope of hazardous events and upon inspection of various reports, books, and resources (e.g. FEMA, USGS, EM-DAT), natural hazards were categorized initially into five main groups to enable the extension of the Flood Ontology to other disasters while keeping the focus on floods (Fig. 2). These groups are Geological Hazards (e.g. earthquake), Meteorological Hazards (e.g. tornado), Diseases, Wildfires, and Floods. Flood concept in the ontology can be grouped by their formation types including river floods, coastal floods, storm surge, inland flooding, and flash floods. The importance of the structure that allows extension lies in ontology's ability to provide a complete picture in terms of the semantic relationships of natural phenomenon. For example, areas that experienced wildfires may be prone to flash floods as the reduction in vegetation causes increased water runoff (NOAA 2018).

Technological devices that enable the measurement of quantities useful for flood preparedness, management, and response were described under the *Instrument* branch (Fig. 3). Gathering environmental data for a point or region is a critical task that requires the assessment of the different type of resources in order to have a better understanding of the environmental processes and develop models. Several environmental monitoring device types (i.e. *Stream Sensor*, *Reservoir Gauge*, *Radar*, *Rain Gauge*, *Groundwater Well Gauge*, and *Soil Moisture Gauge*) were grouped under the *Instrument* branch.

Selected environmental concepts that play a role in the lifecycle of a flooding event were defined under *Environmental Phenomena* branch. Although it can be extended, this branch was grouped under three subcategories (i.e. *Hydrological*, *Meteorological*, and *Social*) that are directly related to flooding while satisfying the requirements for the application scenarios described in this work. The three branches of the Flood Ontology are tightly interconnected with each other through semantic relationships to assure completeness. This structure allows the ontology to be capable of comprehending and resolving complex environmental queries such as "Which sensors were used to support forecast models in the Iowa flood of 2008?" In response, the ontology is capable to deduce the rivers and locations which are part of that flood, find the stream sensors that are installed on those locations, filter the stream sensors by their ability to measure stage and discharge, and finally determine the exact response to the given query. The Flood Ontology is available in XMI file format on GitHub (https://github.com/uihilab/floodontology).

Information system ontology

Information System (IS) Ontology defines ontological entities for the tools and actions of an IS along with the attributes and semantic relationships of these entities. The IS Ontology was created to encompass the requirements and capabilities of the software for each application scenario. It was created and maintained by software engineers, data scientists, and system owners. IS Ontology was categorized under three main branches (i.e. Data, Analysis, and Action). *Data* branch encompasses all retrieved or calculated data channels for the targeted information system. *Analysis* branch gathers all
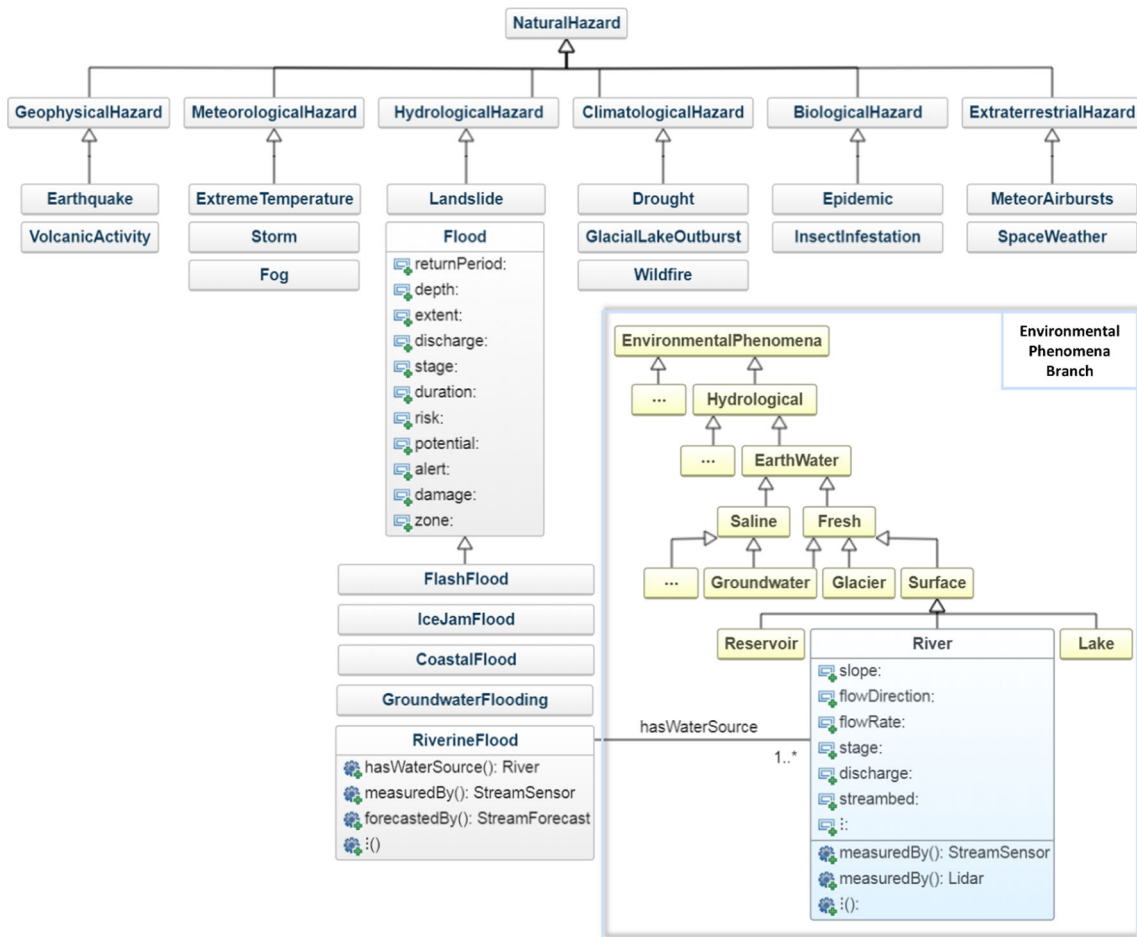
Fig. 2 Natural Hazard and Environmental Phenomena branches of the Flood Ontology with emphasis on riverine floods while highlighting its semantic relationship with rivers. Grey boxes represent the concepts of Natural Hazard branch, while yellow boxes represent the partial depiction of the Environmental Phenomena branch. The ontological concept *River* is highlighted with blue color

statistical, hydrological, map-based, and terrain analysis and calculations. *Action* branch provides all functions and interactions of the information system and should be revised

according to the application scenario. Considering the capabilities of the information system, the output of the action can be a report, table, or a visualization. Visualization and

Fig. 3 Partially-depicted instrument branch of the flood ontology in UML representation isolated from the rest of the ontology
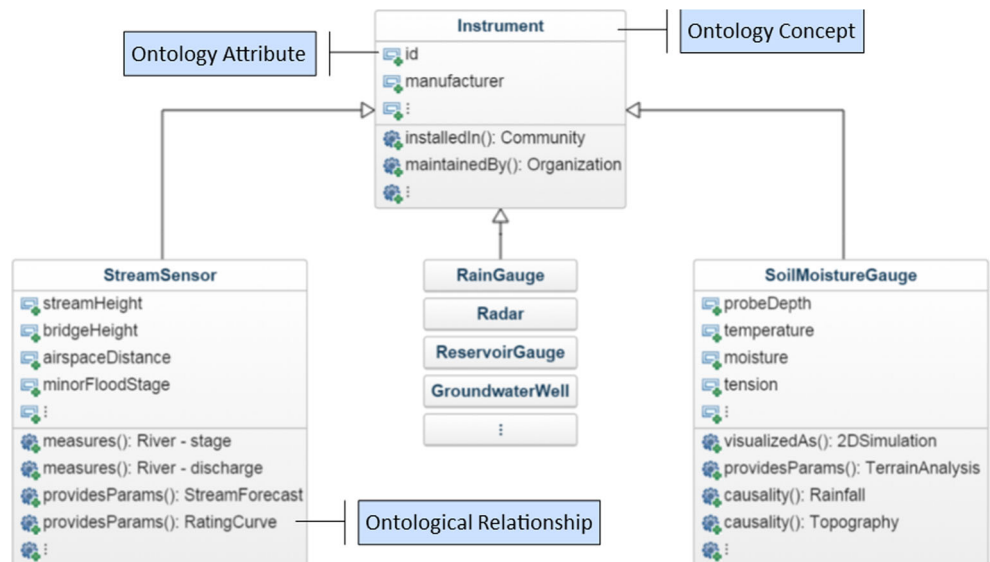
communication of environmental datasets can be in various forms (Demir et al. 2009) including graphs (e.g. rainfall accumulation), map layers (e.g. flood maps), base map (e.g. terrain, streams), and games (e.g. serious gaming; Demir et al. 2018). Some map layers require different data representation such as raster, grid, and vector (i.e. point, line, and polygon).

The IS ontology's main purpose is to integrate the domain knowledge of the Flood Ontology with capabilities of a software application (e.g. web, mobile, and desktop). Resources and functionality of the targeted application are connected to the appropriate ontological elements in the Flood Ontology. For instance, river discharge calculation (i.e. $Q = A/\overline{u}$) which is a function definition for the *discharge* attribute of the *river* concept, is represented under the *Hydrological Analysis* branch in the IS ontology. This connection assures that river discharge is an actual method that expects certain variables to be computed rather than just a textual value as a description of an attribute.

To embody the integration of Flood and IS ontologies, consider the query "Show me the flood map for Cedar Rapids by 100-year return period" for the Knowledge Engine application scenario. Cedar Rapids is a river community, thus have a drainage basin and data resources such as stream gauges. Stream gauges provide certain parameters (e.g. stream height, bridge elevation, flood stages) for rivers that they are installed to measure. Rivers are defined by their slope, flow direction, flow rate, and stage, in addition to other parameters. Furthermore, fluvial floods inherit the properties of the flood event attributed by parameters such as return period, depth, extent, discharge, stage, duration, damage, and so on. Domain-specific conceptual relationships had been discovered by the Flood Ontology, however, there is a need for an intermediary that comprehend the requested action along with the required parameters to achieve it. In the IS ontology, *Flood Maps* is a concept under *vector-based* (i.e. polygon) *visualization* concept that is categorized into three main groups (i.e. custom stage, real-time stage, return period). Each flood map is specific to a river with the retrieved or calculated flood level. Additionally, flood maps have an association with certain 2D or 3D graphics and simulators (e.g. flight simulator), thus allowing the system to present compatible features and functionalities to enhance the user interaction. Figure 4 provides the conceptual schema of the IS ontology separated by the attributes, semantic relationships, descriptions, and synonyms.

Case study

As a use case of the presented ontologies, an intelligent system, namely Flood AI, on knowledge generation and communication for flooding is developed using Flood and Information System ontologies (Fig. 5), and reported at Sermet and Demir (2018a). Flood AI presents an intelligent

system for processing linguistic inputs and providing answers on flood-related data and information. It utilizes information-centric ontologies, natural language processing and artificial intelligence methods, and voice recognition to provide factual and definitive responses to flood-related inquiries. Flood AI demonstrates how the presented ontologies can be used in a contemporary knowledge generation system to power natural language understanding and to connect user input to relevant knowledge discovery channels and resources on flooding. Interaction with the user is achieved through various communication channels including voice and text input via web-based systems, agent-based chatbots (e.g. Microsoft Skype, Facebook Messenger), smartphone and augmented and virtual reality applications (e.g. Google Assistant, Microsoft HoloLens; Sermet and Demir 2018b), automated web workflows (e.g. IFTTT), and smart home devices (e.g. Google Home, Amazon Echo). Flood AI's comprehensive communication channels show the viability of developing an information-centric ontology by opening the knowledge discovery for flooding to thousands of use cases and applications.

The Flood AI's modular structure is comprised of ontology management, natural language processing (NLP), query mapping and execution (QME) modules. Ontology Management module accepts and processes the given ontology to provide engine the domain-specific knowledge, and stores it in a relational database format for efficient use. NLP module utilizes ontology to extract useful features from the question including location, date and time, ontological entities and relationships, and intent. Query Mapping and Execution module analyses the extracted features of the user question to determine the relevant internal and external information resources and environmental and numerical analyses, and returns the definitive answer in an appropriate format.

The Flood AI utilizes Service Oriented Architecture (SOA) where each module has an Application Programming Interface (API) interacting with each other. The advantage of this approach is to offer each module of Flood AI as a service for different applications to benefit from the system according to their needs. For example, if an ontology in another domain was given to the ontology management module, the same process will be performed independently, assuming ontology is formed with the same specifications described in the paper. NLP module can proceed to analyze the natural language question in the relevant domain without any modifications. Thus, usage of the information-centric ontologies allows the expansion of the knowledge engine to other natural disasters, and any science or engineering domains with minimal effort. Another benefit of this architecture is that separate modules can be used for applications with different goals. For instance, an application that retrieves flood-related tweets from Twitter can analyze them by using only the Ontology Management and NLP modules.
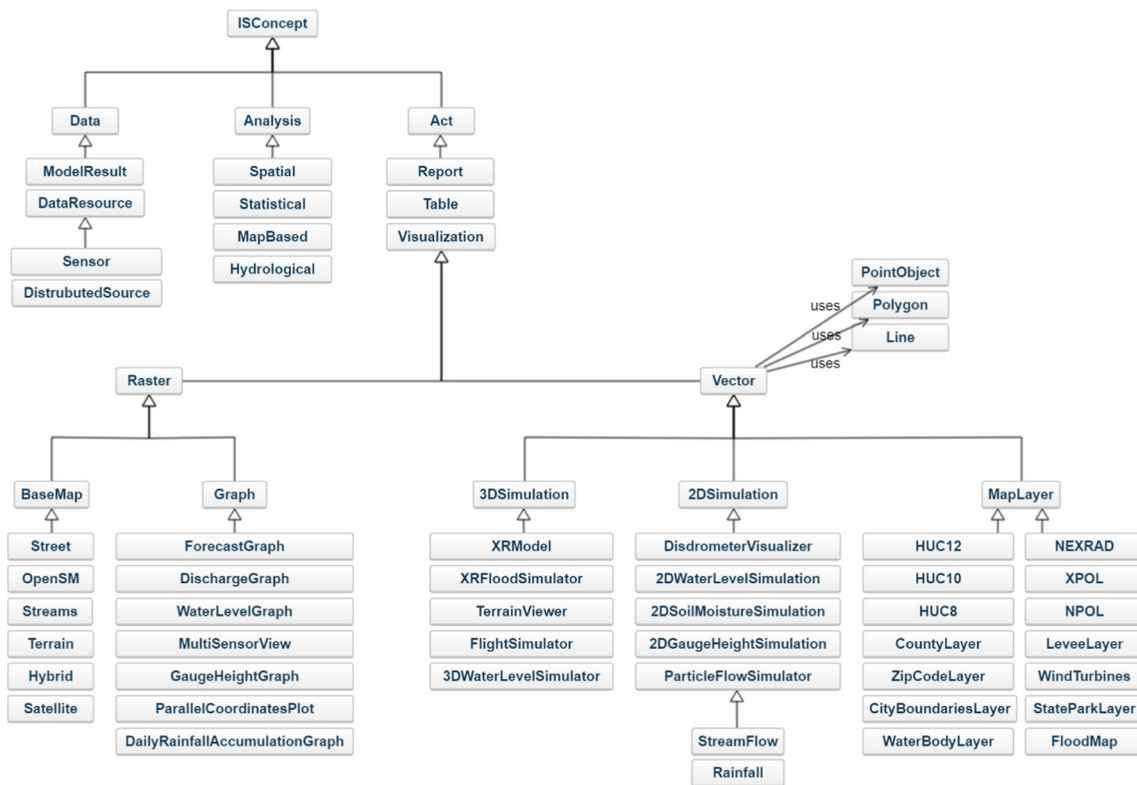
Fig. 4 Overview of the Information System Ontology's conceptual schema with an emphasis on the Visualization concept

Evaluation

In order to build confidence for the proposed ontologies and methodically assess which use cases are suitable for an application, a validation and evaluation strategy needs to be utilized. There are several systematic approaches and techniques to employ when evaluating an ontology as summarized by Brank et al. (2005). These strategies can be summarized into four categories: comparing to a 'golden standard', application-based evaluation, comparison to a collection of documents (i.e. data-driven), and assessment by humans. Golden standard is a consensual and comprehensive representation of the related domain in the forms of a set of strings or an ontology, which acts as a reference to evaluate the quality of other ontologies. As the notion of an information-centric ontology for flooding is novel and no comparable ontology exists, the 'golden standard' approach may not be suitable for evaluation. The fourth approach (i.e. evaluation by domain experts) was

Fig. 5 System architecture for information-centric ontology-based knowledge engine
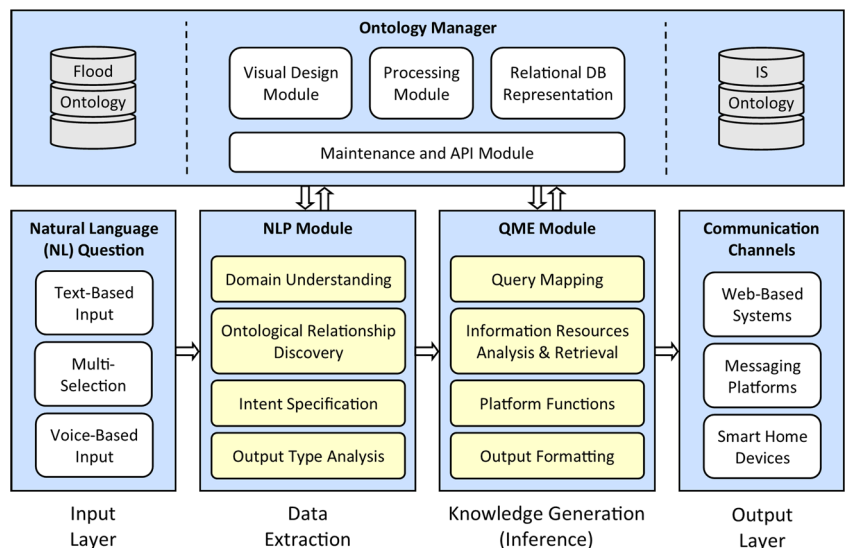
partially implemented as the assessment had been done throughout the development of the ontologies by domain experts and data scientists who were involved in the project. However, a formal evaluation by objective third parties had not been done. Thus, we have employed both data-driven and application-based evaluation strategies.

The data-driven approach is used to validate the Flood Ontology by eliciting and confirming the lexical vocabulary, concepts, hierarchy, and other semantic relationships using existing validated data. For instance, the Natural Hazard branch of the Flood Ontology (Fig. 2) can be observed to be compatible with the scientific report on disaster classification by The International Disaster Database (Guha-Sapir et al. 2012). The application-based strategy measures the quality of the ontology by evaluating the application in which the ontology is used. The correctness of the results produced by the application that depends on the ontologies indicates the quality of the ontologies. This approach is particularly suitable for the Flood and IS ontologies since their main objective is to facilitate knowledge generation in modern information systems. As discussed in the previous section, the Flood and IS ontologies were used to support Flood AI intelligent system. The verification and validation of the results were discussed in detail in Sermet and Demir (2018a).

Additionally, the ontologies developed in this project were evaluated against the five core principles for ontology development as described by Gruber (1995). Although the Flood Ontology is designed specifically for use in modern information systems, all concepts are defined in regard to their objective meanings to avoid any dependency on computational context. The conceptual relationships are carefully elicited by consulting to domain experts in order to avoid potential implicit inferences that are not consistent with the domain. As discussed in the previous sections, both ontologies are designed intentionally to allow extensions to other disasters and relevant domains without having to change the existing conceptualism. Encoding bias is minimized by allowing the ontology developers to freely browse and modify the ontology using visual editors and natural language. During, the Flood Ontology's design, various resources from different countries, organizations, and fields were benefited in order to represent the flood domain requiring as less commitment as possible. Thus, it is shown that the flood ontology abides by the quality standards, which were introduced by Gruber, including clarity, coherence, extendibility, minimal encoding bias, and minimal ontological commitment.

Conclusion

In this project, the development of an information-centric flood ontology is presented with generation process, best practices and challenges, and potential use cases. The ontology can be utilized in cyberinfrastructure systems for natural hazard preparedness, monitoring, response, and recovery. It is described how the information-centric approach allows the mapping of the capabilities and resources of a knowledge generation system to the flood ontology. Ontology was created to allow visual, intuitive, and collaborative development for domain experts who may have a limited technical background. Conversion of the ontology from UML to XMI makes it possible for software systems to easily integrate the domain knowledge. The ontological model provides a common framework for scientists and professionals in other fields to adapt and make further development.

A web-based knowledge engine for emergency preparedness and response was presented as a use case of the flood ontology. Knowledge engine allows the stakeholders to ask questions in natural language and instantly get the computed factual answer. The engine was developed as a software-as-a-service (SaaS) application and integrated into web-based systems, agent-based bots, smart assistants, and automated workflows. Alongside assisting the decision makers and authorities to take precautions in time and to respond to an emergency, another goal of the system is to raise public awareness of natural hazards.

The successful development of information-centric ontologies depends on experts and interested parties to examine and suggest revisions through an online platform to allow expansion, assure completeness and improve clarity. Scientists in other fields can utilize the framework to implement domain knowledge for use in cyberinfrastructure systems. The results from this project will facilitate the development of the next generation information generation and sharing systems for disaster-related applications as well as other domains.

References

Abburu S, Golla SB (2017) Ontology and NLP support for building disaster knowledge base. In: 2017 2nd International Conference on Communication and Electronics Systems (ICCES). IEEE, pp 98–103

Agresta A, Fattoruso G, Pollino M, Pasanisi F, Tebano C, De Vito S, Di Francia G (2014) An ontology framework for flooding forecasting. In: International conference on computational science and its applications. Springer, pp 417–428

Apacible JT, Encarnación MJ, Nareddy KR (2013) U.S. patent no. 8,606,739. Washington, DC: U.S. Patent and Trademark Office

Babitski G, Probst F, Hoffmann J, Oberle D (2009) Ontology design for information integration in disaster management

Beran B, Piasecki M (2009) Engineering new paths to water data. Comput Geosci 35(4):753–760

Bermudez LE, Piasecki M (2003) Hydrologic ontology for the web. In: AGU Fall Meeting Abstracts

Brank J, Grobelnik M, Mladenic D (2005) A survey of ontology evaluation techniques. In Proceedings of the conference on data mining and data warehouses (SiKDD 2005). Citeseer Ljubljana, Slovenia, pp 166–170

De Nicola A, Missikoff M, Navigli R (2009) A software engineering approach to ontology building. Inf Syst 34(2):258–275

Demir I, Beck MB, (2009) GWIS: a prototype information system for Georgia watersheds. In Georgia water resources conference: regional water management opportunities, UGA, Athens, GA

Demir I, Krajewski WF (2013) Towards an integrated flood information system: centralized data access, analysis, and visualization. Environ Model Softw 50:77–84

Demir I, Szczepanek R (2017) Optimization of river network representation data models for web-based systems. Earth and Space Science 4(6):336–347

Demir I, Jiang F, Walker RV, Parker AK, Beck MB (2009) Information systems and social legitimacy scientific visualization of water quality. In: EEEE International conference on systems, man and cybernetics, 2009. SMC 2009. IEEE, pp 1067–1072

Demir I, Conover H, Krajewski WF, Seo BC, Goska R, He Y, McEniry MF, Graves SJ, Petersen W (2015) Data-enabled field experiment planning, management, and research using cyberinfrastructure. J Hydrometeorol 16(3):1155–1170

Demir I, Yildirim E, Sermet Y, Sit MA (2018) FLOODSS: Iowa flood information system as a generalized flood cyberinfrastructure. International Journal of River Basin Management 16(3):393–400

Djurić D, Gašević D, Devedžić V, Damjanović V (2005) A UML profile for OWL ontologies. In: Model driven architecture. Springer, Berlin, Heidelberg, pp 204–219

Dolbear C, Goodwin J, Mizen H, Ritchie J (2005) Semantic interoperability between topographic data and a flood defence ontology. Ordnance Survey Research & Innovation, Southampton

Garrido J, Requena I, Mambretti S (2012) Semantic model for flood management. J Hydroinf 14(4):918–936

Gray AJG, Galpin I, Fernandes AA, Paton NW, Page K, Sadler J, Kyzirakos K, Koubarakis M, Calbimonte JP, García Castro R, Corcho O (2010) Semsorgrid4env architecture

Gruber TR (1995) Toward principles for the design of ontologies used for knowledge sharing? Int J Hum Comput Stud 43(5–6):907–928

Guarino N (ed) (1998) Formal ontology in information systems: proceedings of the first international conference (FOIS'98), June 6–8, Trento, Italy, vol. 46. IOS press

Guha-Sapir D, Vos F, Below R, Ponserre S (2012) Annual disaster statistical review 2011: the numbers and trends. Centre for Research on the Epidemiology of Disasters (CRED)

Holz KP, Hildebrandt G, Weber L (2006) Concept for a web-based information system for flood management. Nat Hazards 38(1–2):121–140

Jackson P (1998) Introduction to expert systems. Addison-Wesley Longman Publishing Co., Inc

Jones CS, Davis CA, Drake CW, Schilling KE, Debionne SH, Gilles DW, Demir I, Weber LJ (2018) Iowa statewide stream nitrate load calculated using in situ sensor network. JAWRA J Am Water Resour Assoc 54(2):471–486

Kalabokidis K, Athanasis N, Vaitis M (2011) OntoFire: an ontology-based geo-portal for wildfires. Nat Hazards Earth Syst Sci 11(12):3157–3170

Katuk N, Ruhana Ku-Mahamud K, Norwawi N, Deris S (2009) Web-based support system for flood response operation in Malaysia. Disaster Prevention and Management: An International Journal 18(3):327–337

Kogut P, Cranefield S, Hart L, Dutra M, Baclawski K, Kokar M, Smith J (2002) UML for ontology development. Knowl Eng Rev 17(1):61–64

Kollarits S, Wergles N, Siegel H, Liehr C, Kreuzer S, Torsoni D, Sulzenbacher U, Papez J, Mayer R, Plank C, Maurer L (2009) ONITOR–an ontological basis for risk management. Tech. rep., MONITOR (January 2009)

Kurte KR, Durbha SS, King RL, Younan NH, Potnis AV (2017) A spatio-temporal ontological model for flood disaster monitoring. In: 2017 IEEE International Geoscience and Remote Sensing Symposium (IGARSS), Fort Worth, TX, pp 5213–5216

Liang TP, Turban E, Aronson JE (2005) Decision support systems and intelligent systems. Penerbit Andi, Yogyakarta

Liao SH (2005) Expert system methodologies and applications—a decade review from 1995 to 2004. Expert Syst Appl 28(1):93–103

Object Management Group (OMG) (2015) XML Metadata Interchange (XMI) Specification, Version 2.5.1. https://www.omg.org/spec/XMI/2.5.1. Accessed 1 Aug 2019

Organization of American States (OAS) (1990). Disaster, planning and development: managing natural hazards to reduce loss. https://www.oas.org/dsd/publications/unit/oea54e/begin.htm#Contents

Peppard J, Ward J (2016) The strategic management of information systems: building a digital strategy. Wiley

Scheuer S, Haase D, Meyer V (2013) Towards a flood risk assessment ontology–knowledge integration into a multi-criteria risk assessment approach. Comput Environ Urban Syst 37:82–94

Sermet Y, Demir I (2018a) An intelligent system on knowledge generation and communication about flooding. Environ Model Softw 108:51–60

Sermet Y, Demir I (2018b) Flood action VR: a virtual reality framework for disaster awareness and emergency response training. In: Proceedings of the international conference on modeling, simulation and visualization methods (MSV), pp. 65–68

Sit MA, Koylu C, Demir I (2019) Identifying disaster-related tweets and their semantic, spatial and temporal context using deep learning, natural language processing and spatial analysis: a case study of hurricane Irma. Int J Digital Earth 1–25

Stanford University (2019) Protégé – Products. https://web.archive.org/web/20190430114157/https://protege.stanford.edu/products.php. Accessed 30 April 2019

Studer R, Benjamins VR, Fensel D (1998) Knowledge engineering: principles and methods. Data Knowl Eng 25(1):161–198

Sui DZ, Maggio RC (1999) Integrating GIS with hydrological modeling: practices, problems, and prospects. Comput Environ Urban Syst 23(1):33–51

The NOAA National Severe Storms Laboratory (2018) Severe Weather 101 – Floods. https://www.nssl.noaa.gov/education/svrwx101/floods/types/. Accessed 1 Aug 2019

Tripathi A, Babaie HA (2008) Developing a modular hydrogeology ontology by extending the SWEET upper-level ontologies. Comput Geosci 34(9):1022–1033

Uschold M, King M (1995) Towards a methodology for building ontologies. Artificial Intelligence Applications Institute, University of Edinburgh, Edinburgh, pp 15–30

Wang W, Stewart K (2015) Spatiotemporal and semantic information extraction from web news reports about natural hazards. Comput Environ Urban Syst 50:30–40

Weber LJ, Muste M, Bradley AA, Amado AA, Demir I, Drake CW, Krajewski WF, Loeser TJ, Politano MS, Shea BR, Thomas NW (2018) The Iowa watersheds project: Iowa's prototype for engaging communities and professionals in watershed hazard mitigation. International Journal of River Basin Management 16(3):315–328

Wu X, Zhu X, Wu GQ, Ding W (2014) Data mining with big data. IEEE Trans Knowl Data Eng 26(1):97–107

Xu J, Nyerges TL, Nie G (2014) Modeling and representation for earthquake emergency response knowledge: perspective for working with geo-ontology. Int J Geogr Inf Sci 28(1):185–205