# Report on the Second International Workshop on Semantic Web Meets Health Data Management (SWH 2019)

Haridimos Kondylakis
ICS-FORTH
Heraklion, Greece
kondylak@ics.forth.gr

Kostas Stefanidis
Tampere University
Tampere, Finland
kostas.stefanidis@uta.fi

Praveen Rao
Univ. of Missouri-Columbia
Columbia, USA
praveen.rao@missouri.edu

Dave Parry
Auckland University of Technology
New Zealand
dave.parry@aut.ac.nz

## ABSTRACT

The advancements in health-care have brought to the foreground the need for flexible access to health-related information and created an ever-growing demand for efficient data management infrastructures. To this direction, many challenges must be first overcome, enabling seamless, effective and efficient access to several health data sets and novel methods for exploiting the existing information. The second international workshop on semantic Web technologies for health data management aimed at putting together an interdisciplinary audience that is interested in the fields of semantic web, data management and health informatics to discuss the challenges in health-care data management and to propose new solutions for the next generation data-driven health-care systems. In this article, we summarize the outcomes of the workshop, and we present a number of key observations and research directions that emerge.

## 1. INTRODUCTION

Precision medicine is the next frontier for research and innovation in healthcare. It deals with treatment and prevention of diseases by taking into account the genetic makeup, environmental and lifestyle factors of an individual [2]. As a result, medical professionals can precisely prevent and treat diseases rather than using a "one-size fits all" approach.

Key in achieving the vision of precision medicine as well as affordable, less intrusive and more personalized care, is to efficiently and effectively harness the value of healthcare data to gain meaningful insights. Ultimately this has the potential to improve patient outcomes, increase the quality of life of patients, and lower mortality. Another important benefit is the potential to lower healthcare costs and reduce medical errors. Electronic health records (EHR) of patients are rich and complex and contain hundreds of attributes [1]. An EHR contains data about a patient's medical history, demographics, diagnosis, medications, allergies, radiology images, lab test results, and other pertinent information. In addition, healthcare data exists in many different formats, from textual documents and web tables to well-defined relational data and APIs. Furthermore, they pertain to ambiguous semantics and quality standards resulted from different collection processes across sites. Data pertaining to healthcare can also be found on social media through healthcare conversations, in wearables and monitoring devices that continuously stream information about a person's fitness and health.

Much effort has been spent in developing interoperability standards for healthcare systems over the last few decades. HL7's Fast Healthcare Interoperability Resources (FHIR) [10] is emerging as a popular standard for healthcare data exchange and developing new applications. In fact, FHIR supports Semantic Web technologies such as the Resource Description Framework (RDF) and SPARQL. Thus, Semantic Web technologies can provide effective solutions for enabling interoperability and common language among healthcare systems, and can lead to the disambiguation of the information through the adoption of various terminologies and ontologies available. In addition, artificial intelligence (AI) and machine learning can enable data-driven decision making and extracting meaningful insights from complex healthcare datasets. Thus, knowledge representation and reasoning on healthcare data become even more important. Semantic Web technologies have matured over the years and can provide these capabilities by design.

The goal of International Workshop on Semantic Web Meets Health Data Management (SWH) is to bring together researchers cross-cutting the fields of Semantic Web, data science, data management, and health informatics to discuss the challenges in healthcare data management and to propose novel and practical solutions for the next generation of data-driven healthcare systems. Developing optimal frameworks for integrating, curating and sharing large volumes of EHR data has the potential for a tremendous impact on healthcare, enabling better outcomes at a lower and affordable cost. The ultimate goal is to enable new innovations in Semantic Web, knowledge management, and data management for healthcare systems to move the needle to achieve the vision of precision medicine.

Next, we summarize the outcomes of the second workshop instance held in conjunction with the $18^{th}$ International Semantic Web Conference (ISWC 2019) in Auckland, New Zealand.[1]

## 2. INVITED TALKS

### 2.1 Semantic AI for Healthcare

Explainable Artificial Intelligence (XAI) aims at explaining the algorithmic decisions of AI solutions with non-technical terms in order to make these decision trusted and easily understandable by humans [3]. HORUS.AI [6] adopts XAI within the healthcare domain based on logical reasoning that supports the monitoring of users' behaviors and persuades them to follow healthy lifestyles.

Specifically, HORUS.AI is an AI-based system built upon the integration of semantic web technologies and persuasive techniques for motivating people to adopt healthy lifestyle or for supporting them to cope with the self-management of chronic diseases. The system collects data from users' devices, explicit users' inputs, or from the external environment (e.g., facts of the world) and interacts with users by using a goal-based metaphor. Interactive dialogues are used for proposing set of challenges to users that, through a mobile application, are able to provide the required information and to receive contextual motivational messages helping them to achieve the proposed goals.

[1]For a summary of the first instance of SWH, please refer to [15].

### 2.2 Personal Consent in Data Management

Semantic web technologies are inherently suitable to serve the role of providing the common shared vocabularies for data sharing intentions and agreements, together with the algorithmic machinery that is needed to process these agreements. Nowadays, there are several approaches that use knowledge graphs to express aspects of data sharing agreements, building on top of more general schemas used to describe persons, personal data, or even healthcare and medical imaging metadata.

These knowledge graphs are great steps towards a vision where users or parties encode their preferences and intentions of data usage in a machine process-able way and data processing algorithms automatically respect these preferences. In order to achieve this, the developed vocabularies have to be backed by the development of generic and re-applicable algorithms; possibly borrowing from data integration [4, 12] or ontology based query answering [14].

## 3. PAPER PRESENTATIONS

### 3.1 Dialogue Management in Healthcare

The development of methods that implement automated planning to manipulate human-machine dialogue is still in its early stages, but it has gained attention in recent years (e.g., [13]). [19] proposes a novel approach for supporting dialogues. The novelty of the approach has to do with the combination of reasoning and planning for supporting dialogues. These two techniques allow to dynamically update the behavior of conversational agents based on the data provided by users. The reasoner is responsible for inferring the most suitable status of a user (or patient). This activity is performed by exploiting not only the user data and the integrated conceptual model, but also the proper resources of the Linked Open Data cloud. On the other hand, the planner generates the interactions for supporting a multi-turn conversation with users in order to acquire the missing information enabling the classification of the users' status.

### 3.2 Self-Management of Diabetes Patients

The interest in designing smart platforms for supporting the self-management of chronic diseases significantly growths in the last years. One of the chronic diseases that most attracted the attention of the research community is diabetes. [7] presents the TreC-Diabetes system, a smart platform aiming to create a continuous link between clinicians and patients for supporting the self-management of diabetes. The novelty of the approach lies in supporting real-time stream reasoning of information provided by patients (e.g. glycemic index, food intake, sport activities) to detect possible critical situations and to inform clinicians about them.

### 3.3 Education and Emotion Based Semantic Recommendations for Health

FairGRecs [18, 17] is a system focusing on recommending interesting health documents selected by health professionals, to groups of users, incorporating the notion of fairness [16], using a collaborative filtering approach. It is the first time such technologies are combined for health recommendations. The overall approach is based on a notion of semantic distance between documents and user profiles. The goal is to offer a list of recommendations to a caregiver who is responsible for a group of patients. The recommended documents need to be relevant to the patients profiles, i.e., to the patients personal health-care records (PHR). However recommendation algorithms so far ignore the fact that patients profiles are multifaceted. For example, recommending the proper document should not only focus on the patients relevant problems but also on their health literacy (namely, the ability to obtain, read, understand, and use health care information in order to make appropriate health decisions and follow instructions for treatment), educational level and psychoemotional status, as emotions can greatly affect the cognitive processes. [11] explores these dimensions, as well paving the way for a new system incorporating all aforementioned aspects.

### 3.4 From Chronic Diseases to Behavior Change

HeLiS [5] is an ontology aiming to provide in tandem a representation of both the food and physical activity domains. [8] presents two extensions of HeLiS modeling for the first time information about food risk levels and self-management barriers. As such, the first extension provides a conceptual model representing the risk level of the food categories already defined in HeLiS associated with the onset or worsening of the most common five chronic diseases (i.e., diabetes, kidney diseases, cardiovascular diseases, hypertension, and obesity). The second extension provides an abstract layer of a conceptual model representing the barriers that a user may encounter during the self-management of his/her lifestyle or of his/her chronic disease (e.g., knowledge representing why a diabetes patient is not able to check his/her glycemia constantly).

### 3.5 Modeling Context in Knowledge Graphs of Diagnostic Reports

Typically, the NLP-based informatics pipelines that target at converting free text to structured text, lack the ability to recover and convey implicit information, found in diagnostic reports. Such information is readily perceived and taken into account by a human reader. [9] develops a unique method in terms of modeling such contextual information for recovering implicit relationships among structurized diagnostic entities. This method enables structurization of contextual information into a cohesive and holistic representation of free text diagnostic reports. Specifically, for doing so, [9] models the context of a diagnostic report in relational triple resource description framework RDF-like format, which is the building block of the model's knowledge base. Triples that share subject or object induce a graph linked using the n-ary relation schema of the semantic web.

## 4. CONCLUSIONS

A number of key observations and research directions emerged in the discussions that we summarize below.

- Although recent technological advancements allow data collection from personal devices, off-the-shelf wearable sensors, and external sources, exploiting these data requires combining and reasoning on a considerable amount of knowledge from different domains (e.g. user attitudes, preferences and environmental conditions, etc.). Semantic technology is a key to this purpose. Besides structured data, semantic data integration should be generalized to unstructured information as well (e.g. discharge letters, pathology reports etc.) as still, such information is widely used in the healthcare domain capturing essential information.
- This semantic integration besides static should also include dynamic data, as multiple streams of data such as glycemic index, food intake, and performed sport activities, constantly arrive and their processing could highly benefit CDS systems.
- In order to generate effective personalized health recommendations, traditional recommendation approaches are not enough. Contextual, psychological and other information should be considered as well, motivating people to adopt healthy lifestyle and better management of their chronic conditions. To this direction persuasion techniques could also be exploited, whereas explainable AI and more

specifically explainable recommendations will pave the way for systems that end-users will actually trust and use daily.

- Another interesting direction in the health domain is dialogue management. Task-oriented dialogues can give advice to patients, offering guidance on the patient's treatment. However, in dialogue systems for the healthcare domain, making the right question at the right moment is a relevant challenge, whereas efficient and effective semantic reasoning are required for providing intelligent discussions.
- Finally, as the usage of personal data is key to in achieving the vision of precision medicine, methods are required to describe smart contracts of data usage in a formal, machine-processable language. Semantic Web technologies can have a central role in this approach by providing the formal tools and languages required.

This second instance of the Semantic Web Meets Health Data Management Workshop made clear that a lot of research work still needs to be done in the area of semantic health data management. Given the growing interest in industry and academia, the third version of the workshop will be held in Athens along with ISWC 2020 in Athens, Greece[2], with a renewed list of topics such as explainable AI in health through semantics, blockchain solutions etc.

# 5. REFERENCES


[1] What information does an electronic health record (EHR) contain?, 2019. https://www.healthit.gov/faq/what-information-does-electronic-health-record-ehr-contain.

[2] What is precision medicine?, 2019. `https://ghr.nlm.nih.gov/primer/precisionmedicine/definition`.

[3] A. Adadi and M. Berrada. Peeking inside the black-box: A survey on explainable artificial intelligence (XAI). *IEEE Access*, 6:52138–52160, 2018.

[4] V. Christophides, V. Efthymiou, and K. Stefanidis. *Entity Resolution in the Web of Data*. Synthesis Lectures on the Semantic Web: Theory and Technology. Morgan & Claypool Publishers, 2015.

[5] M. Dragoni, T. Bailoni, R. Maimone, and C. Eccher. Helis: An ontology for supporting healthy lifestyles. In *ISWC*, 2018.

[6] M. Dragoni, T. Bailoni, R. Maimone, M. Marchesoni, and C. Eccher. HORUS.AI - A knowledge-based solution supporting health persuasive self-monitoring. In *ISWC*, 2018.

[7] M. Dragoni, C. Eccher, S. Forti, S. Puccini, B. Purin, and A. Valentini. Trec diabetes: A semantic platform for supporting the self-management of patients. In *SWH*, 2019.

[8] M. Dragoni and V. Tamma. Extending helis: From chronic diseases to behavior change. In *SWH*, 2019.

[9] P. S. Giannaris, C. Tang, O. Kholod, S. Hanson, C. Shyu, R. Hammer, D. Xu, and D. Shin. Modeling of contextual information in knowledge graphs of diagnostic reports. In *SWH*, 2019.

[10] HL7. Fast healthcare interoperability resources, 2019. `http://hl7.org/fhir`.

[11] H. Kondylakis and K. Stefanidis. Towards education and emotion based semantic group recommendations for health. In *SWH*, 2019.

[12] G. Konstantinidis and J. L. Ambite. Scalable query rewriting: a graph-based approach. In *SIGMOD*, 2011.

[13] K. Lee, Y. S. Lee, and Y. Nam. A model of fsm-based planner and dialogue supporting system for emergency call services. *The Journal of Supercomputing*, 74(9):4603–4612, 2018.

[14] H. Pérez-Urbina, E. Rodríguez-Díaz, M. Grove, G. Konstantinidis, and E. Sirin. Evaluation of query rewriting approaches for OWL 2. In *SSWS+HPCSW*, 2012.

[15] K. Stefanidis, H. Kondylakis, and P. Rao, editors. *Proceedings of the First International Workshop on Semantic Web Technologies for Health Data Management, SWH 2018*, volume 2164 of *CEUR Workshop Proceedings*. CEUR-WS.org, 2018.

[16] M. Stratigi, H. Kondylakis, and K. Stefanidis. Fairness in group recommendations in the health domain. In *ICDE*, pages 1481–1488, 2017.

[17] M. Stratigi, H. Kondylakis, and K. Stefanidis. The fairgrecs dataset: A dataset for producing health-related recommendations. In *SWH*, 2018.

[18] M. Stratigi, H. Kondylakis, and K. Stefanidis. Fairgrecs: Fair group recommendations by exploiting personal health information. In *DEXA*, 2018.

[19] M. S. Teixeira, M. Dragoni, and C. Eccher. A planning strategy for dialogue management in healthcare. In *SWH*, 2019.


[2] `https://sites.google.com/view/swh2020`