AIDAVA (short for AI-powered Data Curation & Publishing Virtual Assistant) is a Horizon Europe project, which brings together 14 partners from 9 EU countries. Their shared goal is to maximize automation in the curation and publishing of heterogeneous and scattered personal health data with the support of the patients, while minimizing their input. The end goal is to give control of health data to patients, so that they can build a comprehensive, interoperable and reusable health medical record, which they can share with their treating physician any time, anywhere. On top of that, they can also share it with external stakeholders for research and policy-making – typically referred as the "common good" – and contribute actively, in a data privacy compliant way, to the emerging European Health Data Space.

What is the Personal Health Knowledge Graph (PHKG) and why is it important?

Isabelle de Zegher: A Personal Health Knowledge Graph is a physical representation of the patient's medical record that could hold potentially all health data from a single individual and therefore constitute a comprehensive longitudinal health record.

There are 2 aspects here that are important. First of all, A PHKG – like any knowledge graph – is based on an ontology. If all PHKGs are compliant with the same ontology, then all these PHKGs share the same exact precise semantics and are interoperable and reusable. This makes the PHKG not just a potential longitudinal health record, but an interoperable and reusable one. This opens tremendous opportunities.

Also, a personal health record evolves over time with unexpected events. It is not always possible to predict beforehand how to represent that information.

- A PHKG can easily be extended by adding triples with any type of semantics; it may require adding concepts to the reference ontology mentioned above but this does not require structural change into the underlying data repository.
- In the relational database management system (RDBMS) / SQL world, which is still today the main repository of electronic medical records, adding information might require modification of the data store and of the supporting application; this can be quite heavy.

Copyright: Ontotext New York | Sofia | Basel
Reach out to us: info@ontotext.com

What is the added value to the patient that semantic technologies (and knowledge graphs in particular) can bring to improve personal healthcare?

Isabelle de Zegher: Today, for a patient, the main interest of a PHKG is that by combining all their information they would be able to provide to their treating physician a holistic dossier to make a clinical decision. Currently this is not possible – as health information is scattered across systems in different formats, from physical paper, to narrative text to structured table – and busy clinicians need to make effort to compile a complete patient picture from this disparate information. This is time consuming; in addition, clinicians may not have access to all the necessary information (e.g., from ambulatory monitoring devices). Such a difficult and time-consuming process is particularly painful for patients with serious conditions and a long medical history.

Europe's Digital Decade declaration targets for 2030 outline the digital rights and principles complementing data protection, privacy legislation and other rights. This includes Principle 4, "citizens able to engage and have control over their own data" (including their health data). As of today, health data is a kaleidoscope of multiple sources in different formats with inconsistencies, redundancies and so on. While it is nice for EU citizens to have control and access to their personal (health) data, there is limited value in its current form.

A PHKG – supporting a stepwise building of a comprehensive patient's medical record – will facilitate increasing the quality of the eHealth record, and ensure it is not only accessible but it also brings benefits to the patient, to their treating physician and to the research community (with patient's content)

What is the added value to Healthcare providers (HCPs) that semantic technologies (and knowledge graphs in particular) can bring to improve personal

Copyright: Ontotext New York | Sofia | Basel
Reach out to us: info@ontotext.com

healthcare?

Isabelle de Zegher: It is not just for HCPs – but also for clinical researchers.

For treating physicians – the first value is having a comprehensive dossier of the patient, allowing him or her to get a very good idea of the patient's problem at a glance as the basis of clinical decision making. Without a complete dossier, he or she may miss an important point and make a sub-optimal decision

For clinical researchers, whether in hospital or Pharma, it is critical to have access to high quality data across multiple patients to understand diseases, assess safety and efficacy of new medications. Having interoperable PHKGs allows to extract "just in time", and with patient's consent, the needed data to build so-called "real world data" (RWD) that can be used in statistical models or Artificial Intelligence (AI) algorithms. Today delivering this RWD is costly: it requires a lot of manual work often managed by specialized companies.

What is your dream application of knowledge graphs in the Healthcare domain and how is the AIDAVA project a step in this direction?

Isabelle de Zegher: My dream: To have all health data of a patient stored in a potentially gigantic personal health knowledge graph managed by the patient or by their deputy. Each patient would have a "data intermediary " (per Data Governance Act) managing their PHKG, similar to how a bank manages their money, and would be able to use this data for their own benefit such as personalized medicine – and share this data for the common good. The PHKG would be linked to Images and to genome sequences to support additional analysis (and truly enable preventive and personalized medicine).

There are many challenges related to this:

- heterogeneity of data – more than 40% of the data are in narrative format and require Natural Language Processing to extract the PHKG,
- more than 50% of the data is redundant (e.g., a copy from a previous encounter to maintain a consistent discharge summary)

Copyright: Ontotext New York | Sofia | Basel
Reach out to us: info@ontotext.com

- many tools are available to support curation and data quality improvement but there is no automation for using these tools

AIDAVA is a research project that aims to demonstrate the feasibility of automating the curation of personal health data, by integrating these data, data source by data source, into a PHKG which is interoperable as it is constrained by the same reference ontology. It is definitely a first step towards that dream

AIDAVA has received funding from the European Union's Horizon Europe research and innovation programme under grant agreement No 101057062. Views and opinions expressed are however those of the author only and do not necessarily

Copyright: Ontotext New York | Sofia | Basel
Reach out to us: info@ontotext.com

reflect those of the European Union. Neither the European Union nor the granting authority can be held responsible for them.

Copyright: Ontotext New York | Sofia | Basel
Reach out to us: info@ontotext.com