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 anytime, anywhere. On top of that, they can also share it with external stakeholders for research and policy-making – typically referred to as the "common good" – and contribute actively, in a data privacy compliant way, to the emerging European Health Data Space.

AIDAVA starts with AI – what is the role of AI in this project?

Remzi Celebi: In this project, AI is very important for sure. Firstly, we are planning to facilitate curation with AI-based tools through a Virtual Assistant. AI-based tools will support patients and data stewards during the curation process.

In order to create an interoperable health data record, we should be able to integrate personal health data (which comes in various formats and structures and varying quality) into

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

a shareable format with other systems and individuals. We need to solve many problems, including deduplication of data that refers to the same entity, extracting structured data within narratives in different languages, as well as mapping them to a common and international ontology.

We are planning to develop or use AI-based tools for each of these problems. With the recent growth in popularity of Large Language Models (LLMs), we also anticipate that they can replace some curation tools.

What are the key technological and research innovations envisioned in AIDAVA?

Remzi Celebi: As I mentioned, AIDAVA actually aims to automate the curation process as much as possible. Just as a data scientist subjects data to preprocessing and transformation processes to make it ready for analysis, transforming health data for sharing typically requires these transformations to be carried out by a human. Usually, individuals or data stewards need to clean and add semantics to data, perform the necessary mapping and transformations.

However, this process can actually be automated up to a certain level, and we plan to reduce the curation tasks that a data steward or patient needs to perform by integrating a tool they can use. This will not only make the data steward's job easier but will also enable patients to improve the quality of their own data.

In order to achieve this, we are planning to develop questions and interfaces tailored to the user's profile. Through the AI-Human interaction module that we will develop in the project, we aim to make it easier for the user to understand the questions asked and define interactions suitable for it, thus obtaining curation with less time and effort without compromising the quality of health records.

Why use knowledge graphs to store personal health

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

records?

Remzi Celebi: Despite the existence of many data formalisms and exchange formats, knowledge graphs are still the easiest and most convenient data model to share data if it is correctly done. The best way to do that is to follow the FAIR principles, which are a set of guidelines on how to publish and share data with other people and systems. Personal health records are scattered in various places and should be linked and shared in an interoperable and reusable format (i.e., FAIR). We know very well that the FAIR principles are influenced by the Linked Data Principles, which play a significant role at the core of knowledge graphs.

Some features of knowledge graphs are the key to adopting them. Supporting semantic web technologies that facilitate data integration, having a flexible structure like a graph, and providing tools to control data inconsistency through reasoning are crucial. In particular, in situations where storing personal data in one place would be problematic, knowledge graphs enable easy linking and querying of data, taking a step in this direction.

Furthermore, when these data are meaningfully combined, it will be possible to make new discoveries about individuals. Just as we discover new information from interconnected websites on the internet, we will be able to learn new information about ourselves by linking medical records accumulated about us. For example, we will be able to know how the food we eat or an existing genetic mutation affects our health.

What are the biggest obstacles to overcome on the path to an ideal solution to personalized healthcare?

Remzi Celebi: For an ideal solution in personalized healthcare, AI systems require access to a diverse range of health data. However, it is crucial that this data is consolidated as consistently as possible as well as standardized. This is because the accuracy of the analyses performed by these AI systems depends on the quality of the data. Achieving interoperability, where different healthcare systems can seamlessly exchange and use data, is also crucial for creating a comprehensive view of a patient's health. Without interoperability, valuable insights may be missed, and care coordination becomes challenging.

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

Healthcare data is well-known for its complexity and variability, including structured data from electronic health records (EHRs), unstructured data from clinical notes, and information from various devices and sources. It can often contain errors, duplications, or inconsistencies. Cleaning and integrating this data to create a unified and accurate patient profile is essential.

Ensuring the quality and accuracy of this data presents a monumental challenge. Inaccurate or inconsistent data can lead to incorrect treatment decisions, potentially negatively impacting patient care. Furthermore, it is of utmost importance that this data is as complete as possible.

How do you plan to scale the AIDAVA solution across EU countries and use data in various formats and languages?

Remzi Celebi: The AIDAVA project's strategy for scaling its health data curation solution across EU countries is a comprehensive and collaborative effort that addresses both the technical and ethical challenges. By focusing on data accessibility, standardization, automation, and linguistic adaptability, AIDAVA aims to create a robust ecosystem that benefits both patients and healthcare providers while respecting privacy and data security regulations.

For example, to make health data accessible and easily integrable for individuals across EU countries, we will create user-friendly interfaces and tools that allow patients to access and manage their own health records effortlessly. By simplifying the process of inputting personal health data into the AIDAVA system, we enable individuals to actively participate in the curation process and enhance the quality of their data.

We also put a strong emphasis on data standardization and automation. Standardized healthcare data formats and ontologies, such as SNOMED, HL7 FHIR, play a crucial role in enabling interoperability. Additionally, the use of AI and machine learning algorithms automates various data curation processes, including data mapping and translation. This not only reduces the workload on human curators but also enhances the accuracy and efficiency of the system.

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

One of the most significant challenges in health data curation is dealing with unstructured data. Unstructured data includes free-text clinical notes, medical reports, and other narrative documents that do not conform to a structured format. The AIDAVA project recognizes the importance of extracting valuable information from these unstructured sources. Through advanced multilingual Language Models and Natural Language Processing (NLP) techniques, the system can identify and extract relevant clinical insights from unstructured data in different languages.

The linguistic diversity across EU countries necessitates language adaptability within the AIDAVA system. The project will employ state-of-the-art machine translation tools for extraction and mapping of medical records and documents in multiple languages.

Above all, we recognize the importance of a patient-centric approach. Patients' consent, data ownership, and control over their health information are fundamental principles. Transparent consent processes empower individuals to make informed decisions about how their data is used and shared for research and healthcare improvement.

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

AI-powered Solutions to Personalized Healthcare Using Knowledge Graphs: An Interview with Remzi Celebi

Author: Ontotext Interviews

This project 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 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