

## **Course Project Task - Data Space design**

In our ecosystem, data comes from a few different sources, which are then utilized for a few different purposes and then visualized in a user-friendly format. It is important to edit and merge the data because the data we need is not directly available, so that data needs to be partially created.

### **Data sources**

#### **The queue of emergency times in the City of Oulu health stations and well-being centers**

The first data to be used is the queuing situation of health stations in the city of Oulu. This data is generated by the City of Oulu and is the interface from which constantly updated information is obtained. The data is licensed under a Creative Commons Attribution 4.0 license, so it is free to use and modify, but the source of the data must be mentioned. I would say the source of the data is secure because these queuing situations are not going to go anywhere, they can at the most change to another interface or file type, etc. The interface is open to everyone and is built on the PostgREST platform. The web service provides JSON format messages for API calls. In addition, there are good documents and instructions for using the data. In my opinion, there is nothing essential or value missing from the data.

#### **Statistics Finland's free-of-charge statistical databases**

In our App and Ecosystem, the program aims to provide targeted tips to the user if he or she wants them, or he or she can browse different statistics using the program, for example. Data sources that can be helpful for our analysis are obtained from Statistics Finland's free-of-charge statistical database. Numerous different statistics are obtained from these databases, which provide data of

interest to us. the data that interests us is health-related and even better if it is limited to the Oulu region or its surrounding municipalities. It is difficult to say anything about this data in mere tabular form, but it can be visualized for the user to make it more understandable. The data is available through Statistics Finland's system, but the system also offers a PxWeb API, through which the data is easier to obtain. Statistics Finland's open data materials and the public content of the online service are subject to the Creative Commons Attribution 4.0Transfer to another service. license, which is a license following the JHS 189 recommendation for public administration open data. It gives you the right to copy, modify and redistribute the materials in either their original or modified form. The materials may also be combined with other materials and used for both commercial and non-commercial purposes. I am sure there will be more fresh data at least every year, but monthly reports are also available related to health.

## **The data the platform will create for others**

Because our platform doesn't need a lot of data to work, it also doesn't produce large amounts of new data. The data we produce is very comparable to normal on-site health care visits. Our platform data consists of information related to health care visits, such as duration of individual visits, date, generic reason, general patient information. However, much of the health data generated by our platform is unstructured data, such as notes about a patient and audio recordings.

Our platform does not use a standard data model, as we think it would be difficult to use them in our context where a lot of weight is on Unstructured data. Such data is difficult to format and transfer between different systems. Extensive work and challenges will occur if we would have to translate the source data language into a Generic model and then translate it again into a vocabulary

used in the final presentation layer for users to make sense of it. The final data format will be known at a later stage when we will know in more detail what the data to be stored will be. However, it is likely to be comparable to other healthcare platforms.

Healthcare organizations are the Greatest single beneficiaries from the data generated by our platform. This data makes it easy to see the health history of individual patients in health care, such as data from previous visits. This makes it easier to maintain patient data when the platform automatically generates recordings of remote visits that are easy to review afterward. Patients benefit in the same way from the data generated by our platform, as they also have easy access to their health history. They can also look at the recordings of the visits in retrospect if they don't remember everything they went through. External stakeholders can also benefit from the data in our system. For example, it can be used to extract general information about, the prevalence of certain diseases or generic information about patient groups, such as mean height and weight.

Patients and healthcare organizations can easily access the data through the different applications they use. Patients naturally only have access to their data, and even individual healthcare professionals only have access to the data of the patients they care for. External stakeholders do not have direct access to the data but need separate rights to access them via API or use data published separately from our platform.

Our platform data largely adheres to FAIR data principles. The data is easily found for both humans and computers, and most of it is machine-readable. Once the user has found the required data, they will know how it can be accessed, including what authentications and authorizations they will need. The data will be largely integrable with other data, although some of it is unstructured. Perhaps the most important FAIR principle, or reusability, is also well managed

because the data is stored so that it is also readily available later for different stakeholders and various purposes.

Under GDPR, all data relating to a person's physical or mental health is considered personal and protected data. This includes any details about the type of care they've received, information about a person's Genetic makeup or lab findings, or even facial photos and Fingerprints. We will start by complying with this by asking users for consent to their data being collected and explaining how their data will be used. We are also prepared to delete user data when they request it.

Creative Commons Attribution 4.0 is used for the data we create. The data may be redistributed and modified, but the sources must also be mentioned, as the data we produce has also been imported from elsewhere. However, the intention is that as many people as possible would benefit from it so that the data can be freely used and exploited in the future.

It is never difficult to say 100% certain whether data is accurate or true if it has not been collected by itself or at least involved in its creation. However, I believe that Statistics Finland's free-of-charge statistical Databases, for example, is a reliable database because it is a government agency. Because of this, the data is reliable and can be trusted. The data generated by the city of Oulu is again real-time data that indicates the amount of the queue, and it is generated by the city, so it is reliable.

## **Design principles for the data spaces**

The principal “public-private governance” is not relevant to our app, so we do not consider it.

Regarding data sovereignty, it is about data owners being able to monitor and control how their data is used and the purpose for which their data is used. It is especially applicable when the data

space contains personal or private data and the data owner is concerned with purposes the data is used for, so do not give free access to data, so imposes a restriction on the data.

As it is mentioned in previous sections, one of the data spaces our app uses its data is the queuing situation of health stations in the city of Oulu. As the data is licensed under a Creative Commons Attribution 4.0 license, so it is free to be used and modified, but the source of the data must be mentioned. This means data sovereignty is not relevant here.

The other dataspace our app uses is Statistics Finland's free-of-charge statistical database. This dataspace is subject to the Creative Commons Attribution 4.0 too. Data sovereignty is not relevant to this dataspace either.

However, data sovereignty is relevant to the data our platform produces. Our app contains data related to the health history of individual patients in health care, such as data from previous visits and patients' physical or mental health. Such data is considered personal data, the patients are told that their data is being collected and how it is being used, they can monitor and control the usage of their data, and there is the possibility to delete users' data if they request it. patients and healthcare organizations have easy access to the data, but external stakeholders do not have direct access to the data.

Regarding the level playing field, it is about facilitating the use of the platform for other service businesses or other entities, so that they can create additional services. It means that the platform should be open to everyone to enter and use the services that the platform provides.

The level playing field is relevant to our platform, as our platform is open to as many patients and healthcare organizations that would like to join it and use its services and add additional services,

without any barriers. So, if there are any new patients or any new healthcare centers, they can easily become a part of our platform and get services or serve through the platform.

Decentralized soft infrastructure, is more about architecture; it is about where the actual data resides, and how different parties of the ecosystems bring in their data. It should also be clarified which party keeps which data and what interfaces are.

Decentralized soft infrastructure is relevant to our platform. The data our platform uses comes from separate sources; even the data that the platform produces and saves for further use is kept on different servers belonging to different parties of the ecosystem. There is an API through which users can have access to the data. There is not a single central source for the data that our platform uses, instead, each party brings in its data.

## **Platform expansion and evolution**

The platform will allow expansion and other entities to join, for instance pharmacies can join to consult prescriptions and provide medication for patients, or nutritionists can create optimal diet plans for people who are suffering from diabetes or heart disease, so there will be a chance for a lot of parties to join and cooperate to help improve the health of the users in general.

New Data will always be created with this platform, since the health status of patients will always change, and they will get cured of a disease, or get diagnosed with a new disease, or get new prescriptions, there will be also new data about appointments and remote consultations, and this data can go through different processing, like cleaning, visualizing, classification, analysis

Participants can be replaced without affecting the data chain since all the data will be directly uploaded to the cloud, so whenever a doctor, pharmacist, dietitian get replaced, the patient records

will still be available in the cloud and will be accessible by the new participant, so it will not affect the data value chain

Each participant can evolve independently, since all of them are independent , and they are just connected with each other, so they are not deeply dependent on the app, since all the participants will be practicing professional with their own offices and workspace, which will give them the ability to evolve independently and still connect and interact with other participants.