# BROKERAGE AND MARKET PLATFORM FOR PERSONAL DATA

## *D5.1 Initial Pilot Marketplaces User Stories*

www.krakenh2020.eu

This project has received funding from the European Union's Horizon 2020 (H2020) research and innovation programme under the Grant Agreement no 871473

# D5.1 Initial Pilot Marketplaces User Stories

| | |
|---|---|
| **Grant agreement** | 871473 |
| **Work Package Leader** | LYNKEUS |
| **Author(s)** | Simon Guggi (SIC), Franco Nieddu (SIC) |
| **Contributors** | Juan Carlos Pérez Baun (ATOS), Anna Rizzo (LYNKEUS), Davide Zaccagnini (LYNKEUS), Markku Nousiainen (TEX), Rob Holmes (TEX), Stefan More (TUG) |
| **Reviewer(s)** | Wim Vandevelde (KU Leuven), Rosa Maimone ( FBK) |
| **Version** | Final |
| **Due Date** | 31/08/2020 |
| **Submission Date** | 25/08/2020 |
| **Dissemination Level** | Public |

**Copyright**

© KRAKEN consortium. This document cannot be copied or reproduced, in whole or in part for any purpose without express attribution to the KRAKEN project.

**Release History**

| Version | Date | Description | Released by |
|---|---|---|---|
| v0.1 | 05/06/2020 | ToC | Franco Nieddu |
| v0.2 | 12/06/2020 | Initial draft | Simon Guggi |
| v0.3 | 30/07/2020 | First revision to review | Franco Nieddu |
| v0.4 | 20/08/2020 | Final version after partner review | Simon Guggi |
| v1.0 | 25/08/2020 | Submitted version | Atos |

# Table of Contents

# List of Figures

# List of Acronyms

| Acronym | Description |
|---|---|
| WP5 | Work Package 5 |
| T5.1 | Task 5.1 |
| D5.1 | Deliverable 5.1 |
| SSI | Self-Sovereign Identity |
| DID | Decentralized Identifiers |
| QR | Quick Response |
| CV | Curriculum Vitae |

# Executive Summary

This document states the initial report describing the User Stories of the two **KRAKEN marketplaces**. It is the first deliverable of Task 5.1 **User Stories refining**, whereas T5.1 is the initial step of Work Package 5 **(WP5)**. The objective of WP5 is to implement pilot marketplaces for the KRAKEN infrastructure, which can be used for the eHealth and education Pilot.

Within this context, T5.1 intends to create User Stories that should reflect the core features of KRAKEN. These User Stories will then help the final refinement of the definition of the use cases of the two pilots. As premised by the task description, we used a user-centered approach during our work. We conducted multiple interviews with potential customers for qualitative information and launched a web-survey to gather quantitative data. We got essential insights for the needs and goals of real-world institutions and managed to get in touch with future users of the KRAKEN platform. Additionally, we were able to create multiple Personas from the held interviews, which will help the immediate next steps of T5.1.

Please note that this is an initial version of the report, and contains the current results. The work on the User Stories is still in progress. The final version of the report will be given in Deliverable 5.2 Final Pilot Marketplaces User stories, due in October 2021.

# 1 Introduction

## 1.1 Purpose of the document

This document aims to provide an initial report of the progress on describing the KRAKEN educational and biomedical marketplace User Stories. To that end, it explains the methodology used and then presents and discusses the initial results.

According to the Description of Work, this task aims to create User Stories that reflect the core features and functionalities of KRAKEN. These User Stories need to be easily understandable. They serve as a basis not only for communication and collaboration but also as a source for the development of software components.

## 1.2 Structure of the document

In the first section of the document, we introduce the deliverable. Section 2 explains some background knowledge, including the methodology we used to derive the User Stories. Next, in section 3, we present our initial results and discuss the insights we obtained. Finally, section 4 concludes the document and frames the next steps in T5.1.

# 2 Background

In this section, we explain the methodology used to derive the initial User Stories for refining the use-cases for the two pilots, namely the eHealth pilot and the educational pilot. We decided to use a user-centered approach where decisions and relevant information is obtained from real-world, potential users of our system, including Universities, Career Centers, and Health Care institutions.

## 2.1 Concept

As already mentioned, we internally agreed to use a user-centered methodology to derive User Stories because it complements the agile methodology used for software development in KRAKEN. User-centered in this context means that we focus on the needs, wishes, and inputs of interested companies, organizations, and individuals that may use the KRAKEN platform in the future.

Our identified main goals were:

- get in touch with real-world institutions for their expertise and for creating contacts
- refine the use-cases identified by the initial KRAKEN proposal
- identify future partners, clients, and users of the KRAKEN platform
- portray the relevant characteristics and usage patterns of the KRAKEN platform's users
- understand the service experience that the users should get on a data exchange marketplace and the touchpoints in which they interact with the KRAKEN platform

## 2.2 Methodology

The first step in our methodology was to **identify User Groups**. We focused on all possible stakeholders that may interact with the KRAKEN marketplace as Data Buyers or Data Providers. As the use case description for the educational and eHealth pilots implies that different actors will be involved, we identified the User Groups for both pilots separately. User Groups act as the starting point of our user-centered approach; therefore, they are a rough accumulation of potential users. A detailed list of the User Groups can be found in Section 3.1.

After the initial step, we **derived actual real-world institutions** according to our previously defined User Groups. With the help of our partners, we identified some organizations as well as some individuals to get in touch with, including already existing contacts of the partners involved.

We conducted several **interviews** with the identified organizations and individuals that were willing to answer our questions. These interviews gave us some initial insight regarding potential users of our system and their needs. The discussions varied in length and detail, depending on the preferences of the interviewee. However, we focused on specific main questions that we identified as fundamental to go on with the next steps.

Parallel to the interviews, we **prepared a web survey** to get in touch with a bigger audience than it would be possible by just contacting institutions. We used the information we gathered during the interviews to define the survey. On our existing channels, we advertised the survey.

With these first results, we created some early **Personas**, also called User Profiles. The Personas help to get an overview of the different user types we will cater to our platform. These Personas are fictional characters that want to interact with our platform. They are used to represent the precise needs of users and should help create a more detailed understanding of the pilots' requirements.

Our next steps include holding more interviews with companies to get an even better insight into our future users and getting more responses to our survey.

After that, we can create **Customer Journey Maps**, which will help us understand the experience of our users better and to identify their pain points while using our platform.

When this is done, we can create Epics based on the Personas and Customer Journey Maps.

## 2.3 Interviews

One of the central parts of our work was to get in touch with institutions that work in the respective sectors of the two pilots. To get some fundamental understanding of the challenges that real-world organizations face to share or monetize their data, we approached institutions and conducted interviews that were individually customized. Although we tailored the discussions for institutions, it was essential to obtain a certain level of comparable information. For that reason, we created a template from which we derived the individual interviews.

Naturally, depending on whether the interviewee acts as Data Buyer or Data Provider, a different template is used.

**Questions for Data Buyers:**

1. Does your company/organization currently acquire data for its operations from outside sources?
    a. If yes, what kind of data and in what format (e.g., batch, real-time)? Also, in what area (personal health/well-being, multiple medical data, disease-specific, medical images, academic, etc.)
    b. If not, do you have a plan about acquiring data from outside, and what could be the kind of data and in which areas of healthcare/academics?
    c. Are there other types of data you're currently seeking/trying to procure?
2. How and with what tools is the data from outside sources currently utilized in your company/organization? OR: How could it be used?
3. What business needs are the data used for? OR What business needs could it be used for?
4. Are there any challenges faced by your organization with regards to data acquisition? If yes, what type of problems?
5. Did you already try any measure to overcome this challenge regarding data acquisition?
6. How do you expect the usage of data change in the near future (next 6-12 months)?

**Questions for Data Providers:**

1. What kind of data does your organization currently have, or produce?
2. Does your company/organization currently share data?
    a. If yes, what kind of data and in what format (e.g., batch, real-time)?
    b. If not, does your company think about sharing or monetizing data?
3. How is this data produced / how does it originate in your company/organization?
4. Are data consented to be utilized by possible buyers?
5. What is the biggest challenge faced by your organization concerning data sharing or monetization?
6. How do you expect the monetization of data in your company/organization change in the near future (next 6-12 months)?
7. Who are the possible buyers?
8. Is data sharing/selling done by Intermediates (marketplace) or directly with buyers?

For the results of the conducted interviews, see Section 3.2.

# 3 Initial Results

In this section, we analyze the initial results. We describe the identified **User Groups** and present the initial **interviews** with some additional insights. Furthermore, we showcase some early analysis of the conducted **survey**. Then we present the created **Personas**. This section concludes with the initially drafted **User Stories**.

## 3.1 User Groups

To organize the process of exploring KRAKEN User Stories, we first created a set of User Groups. This was done by the members of the consortium with expertise in the respective field. For example, TUG contributed insights into the education sector, while LYN defined User Groups using their experience in the eHealth sector. Based on these User Groups, we conducted interviews with members of each group to explore their perspectives and needs, and in the end, to derive Personas.

Due to the structure of the project and their distinctive needs, we separated the involved User Groups between the two pilots. Furthermore, we assigned them to a sub-group based on their role in the KRAKEN ecosystem (Data Provider, Data Buyer).

Data Providers are at the beginning of every KRAKEN process. The term Data Seller is also sometimes used for a subgroup of this group.

While Data Buyers are at the consuming side of a KRAKEN process, they are not necessarily at the end of it since some Data Buyers aggregate and publish authenticated data and therefore are also Data Providers.

The following sections give an overview of the created User Groups.

### 3.1.1 eHealth pilot User Groups

#### Data Providers

- **Individuals** carrying personal data on mobile apps or personal data storage systems
- **Private** and **public institutions** (e.g., hospitals, app and wearable device companies, data brokers, patient associations, research institutions, healthcare authorities, aggregators of individuals' data [data union]) storing individual data, either consented or where consent can be obtained from individuals using a dynamic consent application

#### Data Buyers

- Market stakeholders (e.g., health-tech companies, research institutions, insurers, public authorities, and well-being service providers) interested in acquiring aggregated data sources

### 3.1.2 education pilot User Groups

#### Data Providers

- **Universities** store a lot of data concerning a student's study process and graduation, and support students with access to the job market by connecting them to potential employers. Also, they process credentials for incoming exchange students. Besides, they need to provide student data to statistical institutions
- **Students** want to retrieve their study data and education credentials and provide them to various Data Buyers

#### Data Buyers

- **Human Resources Departments** are looking for talent in the job market and also need to rank the candidates and verify their credentials.
- **Universities** need to extend means to evaluate applications of incoming students and potential employees. Besides, they process certificates from outgoing exchange students after their return
- **Educational Research Institutions** are interested in access to data sources and combining data from them

## 3.2 Interviews

To learn more about the involved stakeholders, we used the mentioned User Groups to identify several organizations and conducted exploratory interviews with them. The interviews used the template described in Section 2.3 as a starting point, and then, depending on the interviewee's answers, a more in-depth discussion started.

For every interview, we pinpointed the most interesting statements relevant to the refinement of the KRAKEN User Stories. We identified certain key elements that would have to be answered by the interviewee. Every interview and the corresponding result will be given in a separate section. We provide some information about the institution and its potential role in the KRAKEN platform. Furthermore, the paragraphs following the introduction of the institution describe some knowledge we obtained from the interview. Lastly, we summarize the interview and the insights we obtained in bullet points.

### 3.2.1 University Career Center

**Role in KRAKEN ecosystem**: Data Provider

**Sector**: Education

**User Group**: Universities

Career Center X works closely together with the corresponding university and is also located in the same facility. Their goal is to bring students and graduates who are looking for jobs and companies who are looking for employees together. From now on, we refer to students and graduates as "students" for simplicity.

**Providing a dedicated job platform:** The primary business model of the Career Center is a specialized job platform that is self-hosted. This platform enables organizations that want to hire a specific student to browse through anonymized curriculum vitae (CVs). Organizations are then able to contact students and only obtain personal data of the student if the student consents. Furthermore, if preferred, organizations can send out a job description via the platform to all possible applicants who met their defined criteria, e.g., all graduates from a specific major who graduated last year. There is no third party, e.g., a marketplace involved so far. The interested companies purchase the CVs directly from the Career Center. Organizations pay a fair amount to get access to the platform.

**Consent of data:** The Career Center obtains its data from students that voluntarily use their services. All data is consented for further use, although there may be a need for explicit consent to be used in our platform. However, as students already gave their consent, the Career Center anticipates that the usage of the data for the KRAKEN platform should not face any trouble.

As already mentioned, the CVs that are bought by companies consist of the data students provide. Grades from the university are not compulsory and are not embedded in the CV by default.

**Authenticity of data:** Interestingly, the career center does not provide any form of authenticity of the data it obtains from the students. According to the experience of the Career Center, there is also no demand for authentication from the companies that purchase the data.

**Compensation for students:** Companies who are already conducting surveys with students are probably interested in our platform if they can reach a large number of students. To address the issue of compensation for filling the survey, according to the Career Center, not only real-world money and cryptocurrencies are used but first and foremost, vouchers and raffles. Another possible Data Buyer could be research centers if the possibility of analyzing and comparing the data from multiple universities exists. Anyhow, for this, a large portion of the students must have their data on the platform, and the data has to be up to date.

**Insights**

- The authenticity of academic data is not that important when hiring
- Students show a willingness to consent their data for further use
- Students like filling out surveys for voucher/entries in raffles
- Companies have a high interest in student's CV and are ready to pay for them
- Data could be useful to research centers if data from multiple universities exist

### 3.2.2 University Job Guidance

**Role in KRAKEN ecosystem**: Data Provider

**Sector**: Education

**User Group**: Universities

Job Guidance is a career service division of the university, whose goal is to facilitate the match between job/stage offers by companies and students/graduates' identification of relevant opportunities for getting into the job market. Students have free access to the institutional services and support provided by this office of the university. Companies also can get free access to most of the office services by signing an accreditation form, which allows access and downloading of about 100 CVs per year through a cross-universities platform (serving 76 national universities).

**Filtering of participants and registration:** There is a certified procedure for registering to this platform through the university portal, enabling to send personal data securely. A single sign-on is provided from the university portal to the platform, which also provides statistics, infographics, and CVs filtering/browsing features to help identify relevant candidates by companies.

**Intermediary for companies and students:** There is also an Alumni portal of the university, where graduates can post their CVs, and companies can access them. Job Guidance plays an essential role in supporting, monitoring, and certifying stage completion by students (to obtain credits). By contrast, it does not provide any authentication of students' CVs, since companies do not request this.

**Sharing data on an international scale:** The interview with a representative of Job Guidance provided initial insights on the potential interest that this organization might have on KRAKEN's platform and marketplace. As a public institution and service, it is not interested in selling students' data. Still, it

would be interested in sharing data with platforms providing access to job/stage offers at international rather than just the national level.

This could bring additional value to the services they offer, and the possibility of providing incentives/tokens to students uploading their data may raise their motivation in updating their CVs more regularly.

**Insights**

- Companies are interested in accessing students' CVs from large datasets
- The authenticity of CVs is not so important, but CVs should be updated
- Career services are willing to go international if security/privacy of data is ensured
- Incentives/tokens are likely to raise students' motivation to upload/update their CVs
- Quality and innovativeness of digital platforms for data sharing are vital to ensure usage

### 3.2.3 Data consulting company

**Role in KRAKEN ecosystem**: Data Buyer & Data Provider

**Sector**: Health

**User Group**: Several (depending on current client case)

This is a medium-sized innovation company, developing solutions around data and using data to solve problems their clients are facing. The company is focused on Fintech and health tech technologies and also has some experience in data privacy, sharing, and exchange. They are currently working on some Covid-19 related projects. In our interview with a representative of the company, our primary focus was on the challenges in obtaining high-quality data sets for the clients, and in data sharing and monetization.

**The process of obtaining good quality patient data is slow and cumbersome**. The interviewee gave an example of a project where they were struggling to find a good dataset in the UK and ended up with an open dataset that was only mediocre. It can be a slow process to get hold of high-quality patient data in the UK, often having to wait for three months.

**Anonymization**: University researchers are often struggling with how to get high quality aggregated data that isn't anonymized. The advantages of non-anonymized datasets are that you have the ability to reach out to the Data Provider, and to bring it together meaningfully in a full story with further information about the individuals or patients such as any kind of activity that a clinical doctor records on a patient visit and aggregated with other sources. This means going outside of the realm of health data to lifestyle data (fitness trackers, Google health data, records on food consumption, dieting, exercise, and social demographics, e.g., how many children the person has, where they live, what they do). This is all very interesting information, but right now, not in the reach of researchers to do any machine learning modeling on it.

**Origin of the data is important**: If we don't know where the data comes from and how it was collected, it's worrying (e.g., when researching a disease like COVID) - It is important to be able to trace the data and where it comes from. This also provides the ability to integrate with different data sources.

**Inaccuracies in data**: We don't have much information to draw conclusions on inaccuracies. The company can, however, get some data from the National Health Service (NHS) in the UK, which they have more confidence in.

**Real-time data sets**: The interviewee does not see any need at this stage for real-time data with regards to their current work. Data streams or time-series data are useful for them and important. It doesn't, however, need to be in real-time but can be historical batches of time-series data.

This doesn't mean that the use cases for real-time data don't exist. Often technology availability leads to innovation. It could just be that as the technology is not currently available, the use cases and solutions around this data type have not yet been developed or discovered.

**Challenges with regards to health data sharing and monetizing**: The problems brought up in the interview were: How do we make sure that the data is valid and that an individual can protect his/her data? E.g., how to prevent it being copied or stored somewhere else without his/her permission. How can an individual put his/her ownership to a potential dataset and benefit from it? Data is so diverse that it's hard to create a single standard for protection on data sharing unless we can implement federated user access to it.

**Comments on other approaches to obtaining health-related data - i.e., surveys**: A problem with surveys is that people are not always willing to do them. They can be ashamed of sharing certain health information, e.g., if it reveals something about their behaviors or sexually transmitted infections history. For this reason, they wouldn't want to share this information via an app. Some specific subsets of the population would, however, do it.

**General thoughts on cash or tokens for data**: The interviewee has some skepticism about rewards as they can promote wrong motivations for sharing data in individuals. It could also lead to issues with the demographics of the data, attracting only certain types of people, which can skew the datasets. Now is an interesting time, however, as they are currently seeing (e.g., with the example of the COVID pandemic) that people appear to be more willing to give away their data for research purposes than before. This also means that there could be some degree of motivation also to receive a reward for doing so.

**Insights**

- Quality of data sets is a significant concern to researchers
- Non-anonymized data sets with known origin have advantages to researchers in that they can link datasets to, or request further information on, background (such as lifestyle) data of the individuals to form a full story or picture of the persons within the dataset, including defining potential causation factors
- In the interviewee's opinion, real-time data sets do not have yet many use cases since many of the same results can be obtained from historical time-series data sets
- Survey data can be skewed because of people not willing to share a specific type of health information, or because only certain groups of people participate

### 3.2.4 Research Center

**Role in KRAKEN ecosystem**: Data Buyer & Data Provider

**Sector**: Health

**User Group**: Research institutions

This is a large research center comprising different departments, including one dedicated to the design, implementation, and validation of digital solutions for public health interventions, within national and international research projects. The interview was carried out with a representative of the research center working in the area of computational and predictive modeling for eHealth, whose interest is particularly focused on using health data to evaluate the performance of machine learning methods and models for predicting events in intensive care, such as patient deterioration or mortality. The interview was mainly focused on understanding his potential role as a Data Buyer of health marketplaces.

**Challenges in finding open datasets of good quality**: The interviewee mentioned several problems found in his research in obtaining access to public datasets of good quality needed in his work. Often the ones available are in different countries (e.g., Canada), raising the problem of diversity in the patient population, as well as difficulties in exporting these data outside of national borders.

The main challenges in accessing relevant datasets are legal and privacy issues, but technical issues were also mentioned, such as data quality and completeness.

**Batch data and Real-time data**: The interviewee reported a need for both batch data, such as clinical data, data from diagnostic tests (blood, imaging), clinical notes, including disease-specific data (e.g., SEPSIS) as well as for real-time data, like behavior pattern, self-reported data from study participants (e.g., well-being, mental and physical health). Both types of data, if available, are deployed in predictive modeling to achieve high levels of accuracy and reliability of predictions.

The interviewee also expressed interest in obtaining socioeconomic data better to understand the correlation between health conditions and socioeconomic status.

**Considerations regarding potential sharing and monetizing of health data**: The interviewee remarked that his organization is focused on data sharing for research purposes and public health, less on monetizing data. Also, the data collected during his studies are consented only for non-commercial use. However, a possible approach to monetize health data in the future could be by creating spin-off companies based on the research results.

**General comments on the availability of marketplaces for health data access**: The interviewee is very interested in becoming a user of marketplaces for accessing the health data he needs. He thinks such a platform would bring huge support to overcome the challenges mentioned above in research studies. He also thinks that soon, health data will become more open, as stakeholders (e.g., healthcare institutions, policymakers, and governments) realize their potential to transform not only healthcare but society at large.

**Insights**

- Access to open health data presents legal, privacy and technical issues to researchers
- Both batch data and real-time data are needed to improve the accuracy of public health interventions
- Health data sharing for research purposes is key for research organizations
- A marketplace for accessing and sharing health data would be of great value for researchers

### 3.2.5 Data producer company

**Role in KRAKEN ecosystem**: Data Provider

**Sector**: Health

**User Group**: Private institutions

The interviewee is a start-up who has developed an app-based system to support people with chronic disease (i.e., Parkinson's disease, genomic diseases, chronic obstructive pulmonary disease) to manage their disease, through self-reports (to be in turn utilized for telemedicine or generation of clinical reports) and notifications, simplifying their life and the life of people directly involved in the management of their condition (formal or informal caregivers, like family).

**Data formats**: home-based monitoring and self-reporting data streams (never hospital-based monitoring) generated by the patient, or caregiver, or parents, bridging the gap between hospital data. Kind of info includes: self-reported data from the patient, drug intakes (notifications, the patient confirms taking drugs), symptom reporting, physiological parameters (e.g., measuring of temperature, the app notifies and the patient register information), support people in managing their plan (e.g., drug therapy, physiological measurements, drinking), questionnaires (specific questionnaires to evaluate the perception of the situation or assess the rank, to check if the situation is good or they need to contact the hospital/physician).

**Data usage**: in case there are telemedicine protocols, clinicians can monitor the data; else, people just receive automatic feedback and they can generate a report for the physician when needed or send data personally in various forms.

**Consent and anonymization**: all the data registered by the app are available for the company, consent given for research purpose and statistical analysis, even for commercial purposes. All data cannot be identified by the company as: they do not rely with the specific person, as users just have user id and password so personal identification is impossible, only hospitals have personal info and anagraphics that can be used to identify the patient. In case a specific consent is needed, the company can contact the patient anonymously through the app for asking for consent, or they can be contacted by the hospital.

**Scopes and challenges**:

- Connection with other apps (under consideration): e.g., (1) home exercise: at the moment, they are working on a project to monitor some specific home exercise with features (angles, amplitude) of the movement; issue of integration with the movement generated with the smartphone (not very precise); (2) food and diet: for another project (not approved yet) they propose to utilize lifestyle info focused on diet and the food perception. Challenges: identify the data, the device to capture the data; the big issue is identifying the device to be sure that all the people in the cluster have the same device and type of device.
- Contact with additional public/private data customers (e.g., pharma for drug vigilance): it is one of their future goals but it is complex at the moment; they are working in small projects with some hospitals where the focus is evaluating the adherence to therapies and incoming problems and issues related to non-adherence to the therapies. Challenges: they are not clinical people so they have problems to discuss with pharma companies, they are involved in some project where they are providing their infrastructure, where they are getting info from patients, but they are not considered as Data Providers because pharma expects to have relations with the clinicians rather than an IT company, a correct intermediary is lacking at the moment.

**Insights**

- App-based data stream providers have the capability to provide data which are already consented for research and commercial purposes and anonymized by default, without the

need to further processing. If needed, consent for additional data uses in a direct, rapid and easy way to the data subject, making this data as ideal for the KRAKEN platform;

- App-based data stream providers would welcome the opportunity to be part of such a marketplace, be engaged in research collaborations, and integrate their data with other app-based Data Providers, but often lack contacts with suitable companies, **organizations** or institutions. KRAKEN should help them with that as an intermediator.

## 3.3 Survey

To complement the interview results, we created a survey that targets a broader spectrum of institutions and aims to get a more general knowledge base. In contrast to the interviews, where the firm goal is to produce qualitative information, the survey creates quantitative results.

To better understand our potential users' needs in the future, we created an online survey targeted on the eHealth pilot. We separated between questions that are targeted for Data Providers and Data Buyers. At the time of writing this report, the survey is not finished yet and is still running. Therefore, we can only give preliminary trends and highlights of the results so far.

**Does your organization currently acquire data for its operations from outside (from data sellers or some other source such as open data)?**

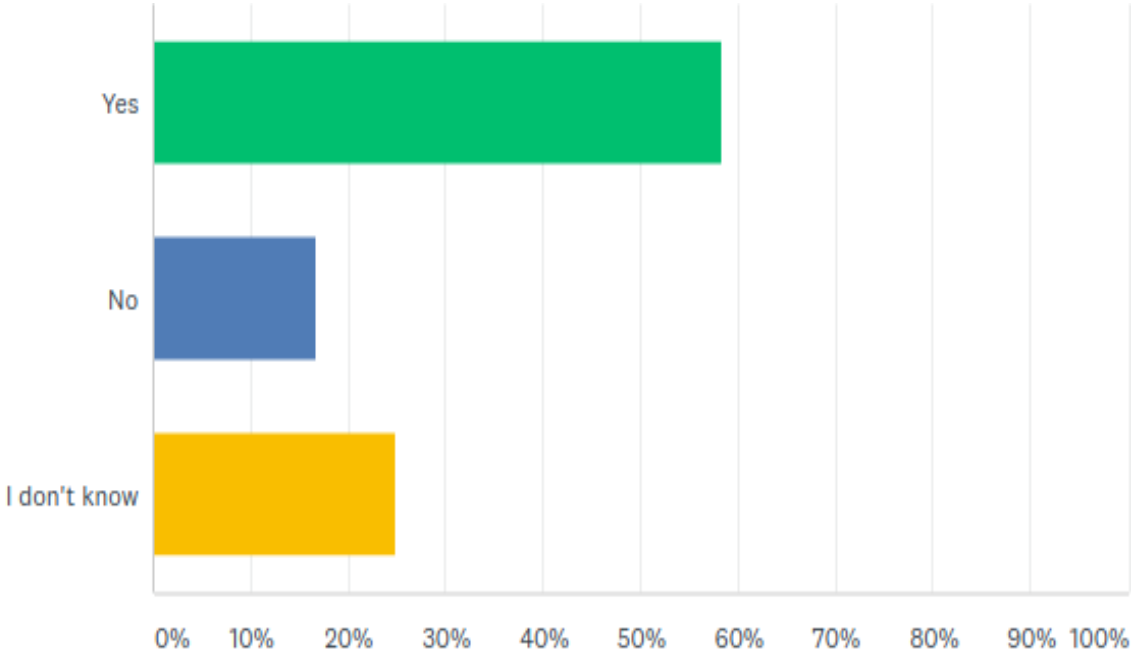


**Figure 1: Survey result for whether organizations obtain outside data**

Of course, it is essential to examine whether our platform's potential users already acquire data from outside sources. Depending on that, we can deduce where to focus our strategies to get in touch with institutions. Surprisingly, more than 50% of respondents already acquire data that does not originate internally.

**How do you acquire the data needed for your work?**

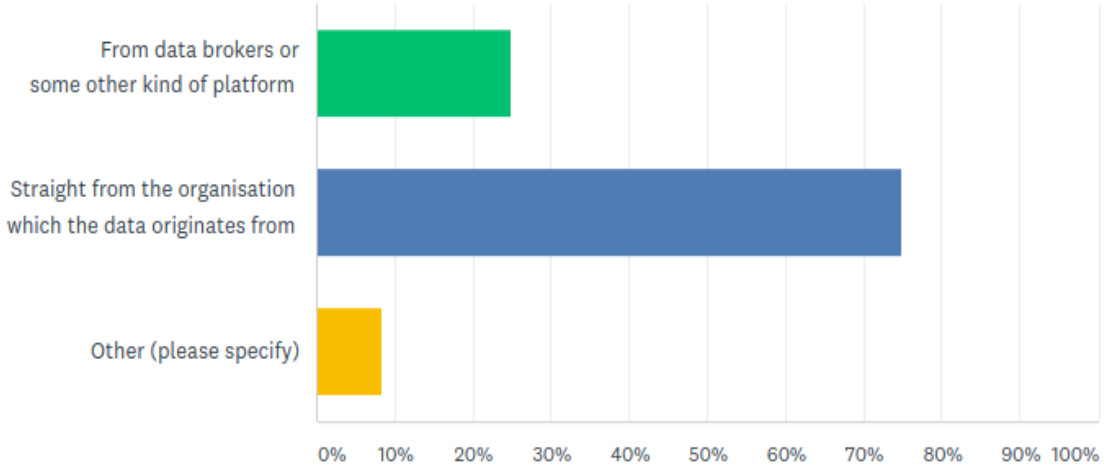


**Figure 2: Survey result for source of data acquiring**

Complementing the previous question, we also asked the respondents whether they are dependent on third-parties to acquire data, like a marketplace or data brokers, or obtain the data straight from the data source. The most significant proportion of the used data is obtained from the organization the data originates from. Nevertheless, there is already the willingness to use data brokers or marketplaces for data.

**What kind of data does your organization currently have, or produce?**

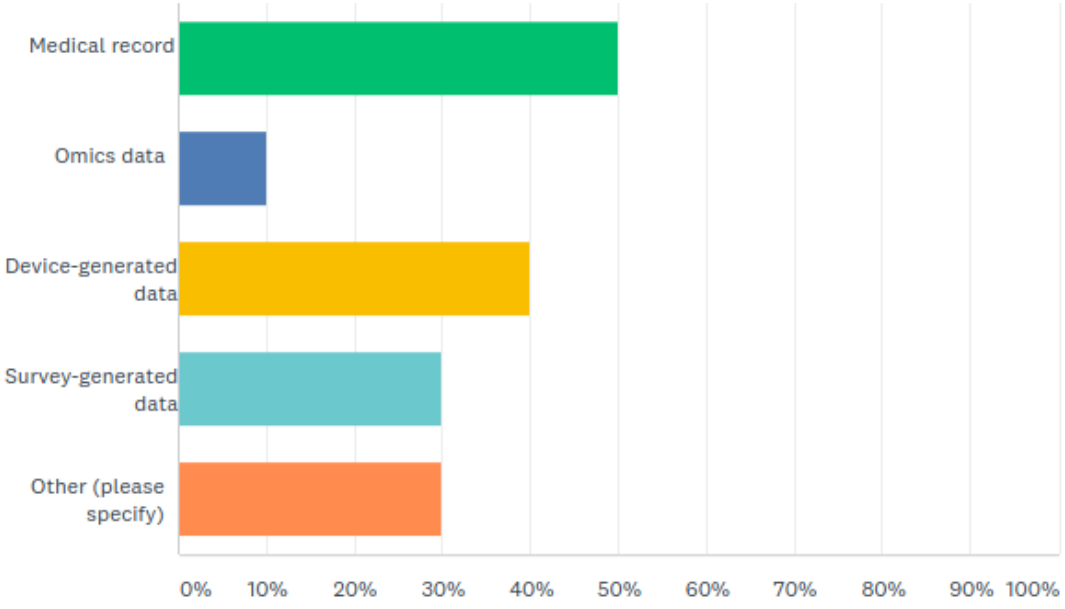


**Figure 3: Survey result for what kind of data is owned/produced**

Of great interest in our market analysis and, in turn, for the design of the marketplace is the question of what kind of data the Data Providers produce. The produced and processed type of data is evenly distributed. We specifically asked for medical records (e.g., clinical histories, lab results, etc.), omics data (e.g., genomics, transcriptomics, proteomics, etc.), data that was generated on devices (e.g., physiological measures, physical performances, etc.), and survey generated data.

**In which format is the data?**

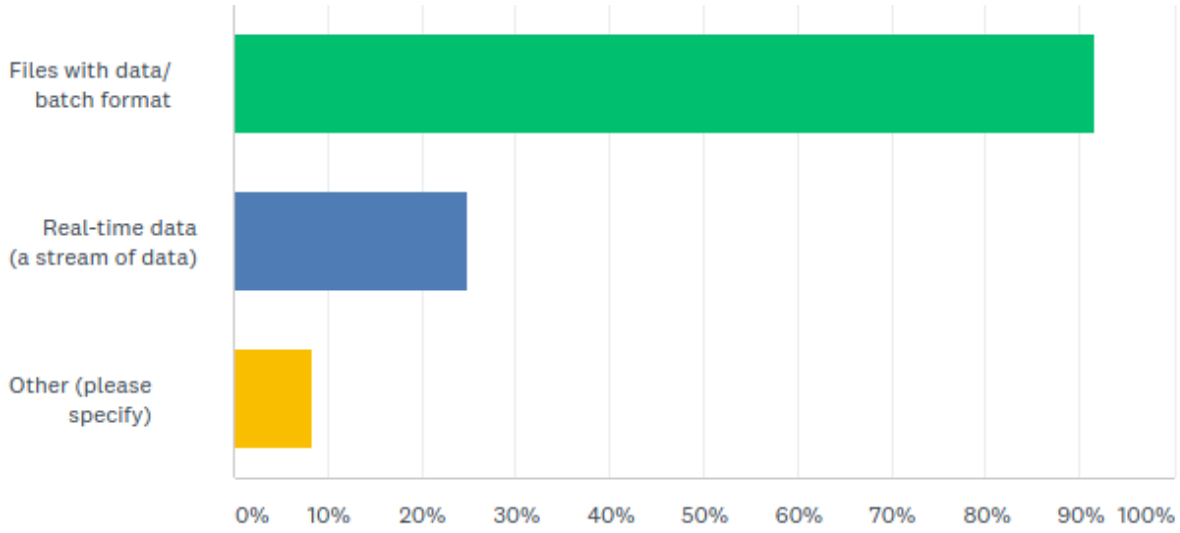


**Figure 4: Survey result for the format of used data**

Although we are not yet at the end of the survey's lifetime, we can estimate that most of the data used are in batch format (>90% of respondents). A small portion also uses streamed data; therefore, we can deduce that if our marketplace offers the possibility to purchase a stream of data, there will be interest.

**Would you be willing to receive cryptocurrency in exchange for data?**

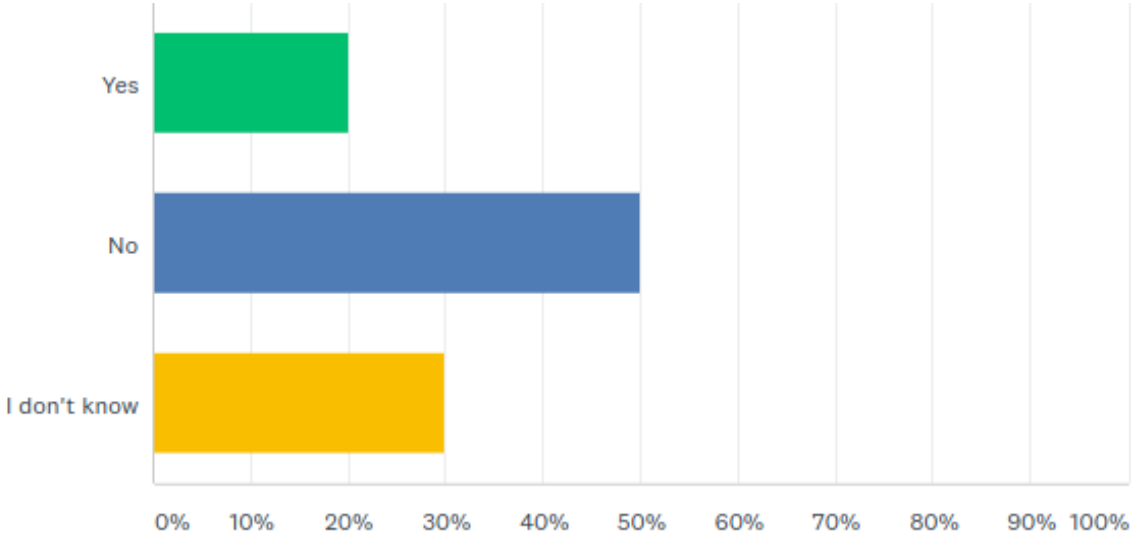


**Figure 5: Survey result for whether organizations are willing to use cryptocurrencies**

This question tries to estimate the willingness of stakeholders to use cryptocurrencies for monetization or buying data via the KRAKEN marketplace. As for now, it seems that the respondents tend to alternate ways for payment. We can give a definite answer when the survey is finished.

## 3.4 Personas

With the insight, we obtained from the initial interviews and the first survey results, we created multiple Personas. In user-centered methodologies, Personas are fictional characters that interact with the system we want to describe. In our case, these characters have different goals they want to achieve. Furthermore, every Persona has a list of challenges that hinders the characters from doing so. These challenges and goals should help in the next steps to fully formulate the User Stories and the refinement for the pilot's marketplaces.

In the upcoming paragraphs, we showcase some of the Personas we created.

### 3.4.1 Victor

**Role in KRAKEN ecosystem**: Data Buyer & Data Provider

**Sector**: Health

**User Group**: Research institutions

This Persona ('Victor') was created based on the research center interview (Section 3.2.4). He is a representative of researchers working in the area of computational and predictive modeling for public health interventions, a large User Group that can potentially use the KRAKEN platform as a Data Buyer and/or for data sharing for research purposes. Victor is well aware that soon, health data will have to become more open and that stakeholders like healthcare institutions, policymakers and governments are going to realize their potential to transform not only healthcare but society at large.

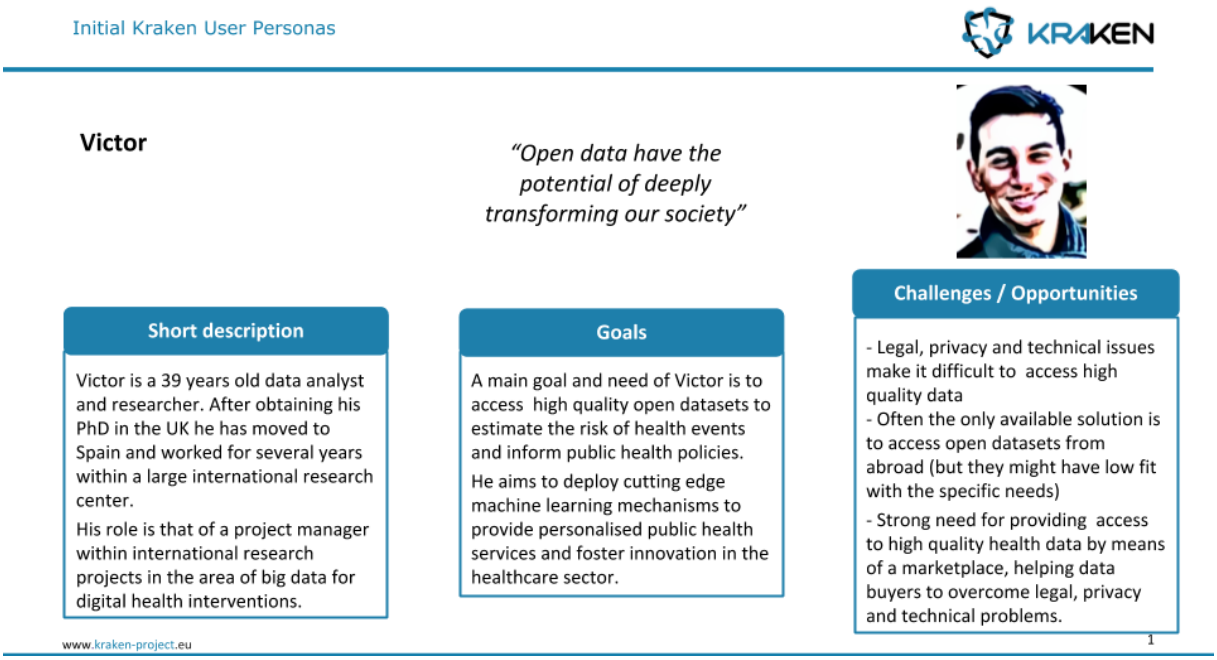


**Figure 6: The Persona Victor**

### 3.4.2 Kateryna

**Role in KRAKEN ecosystem**: Data Buyer & Data Provider

**Sector**: Health

**User Group**: Several (depending on current client case)

This Persona ('Kateryna') was written based on the data consulting company interview (Section 3.2.3). She represents data consultants providing R&D and innovative data-fueled solutions in the health tech sector, which is a User Group that can potentially use the KRAKEN platform by adopting different roles and use cases based on their current client and project. Kateryna is a power user who understands the concerns of her different client groups, such as research scientists, related to the availability and quality of suitable datasets, and the difficulties associated with gaining access to complimentary background and lifestyle data.

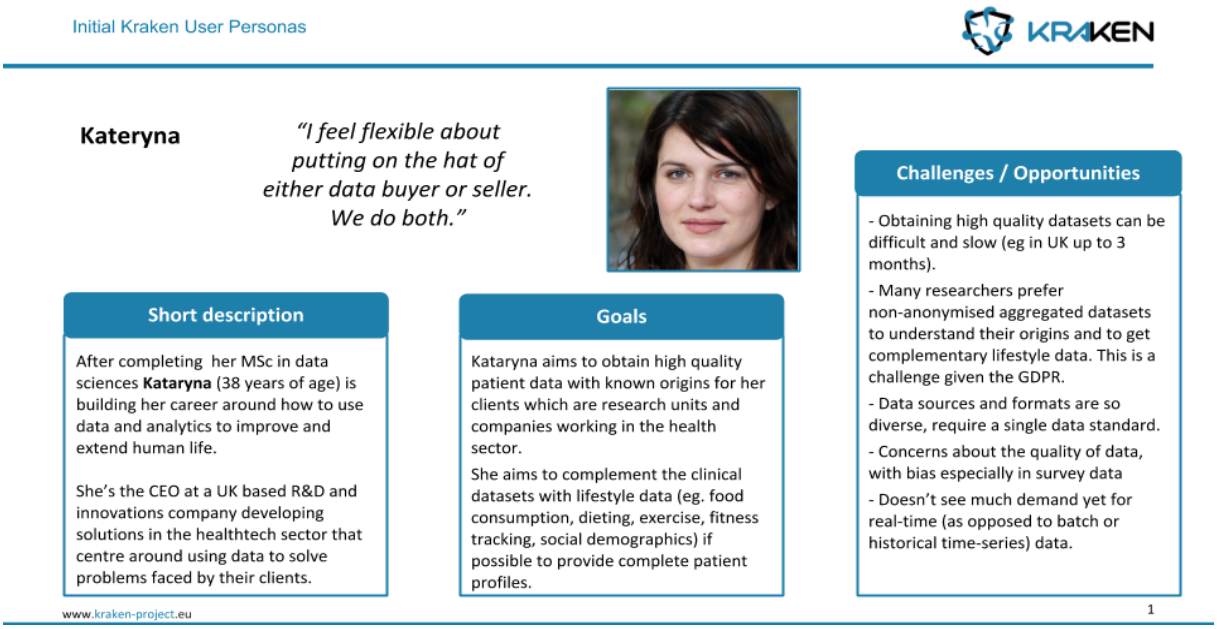


**Figure 7: The Persona Kateryna**

### 3.4.3 Joe

**Role in KRAKEN ecosystem**: Data Provider

**Sector**: Education

**User Group**: University

This Persona ('Joe') was based on the interview with the University Career Center (Section 3.2.1). He represents career service employees providing a platform to connect students and graduates with companies. All of which are potential future users of the KRAKEN platform to offer and buy academic data. Joe is experienced in handling the student data and doesn't want to share any data without the consent of the students. His focus is also only to show relevant job offers to the students.

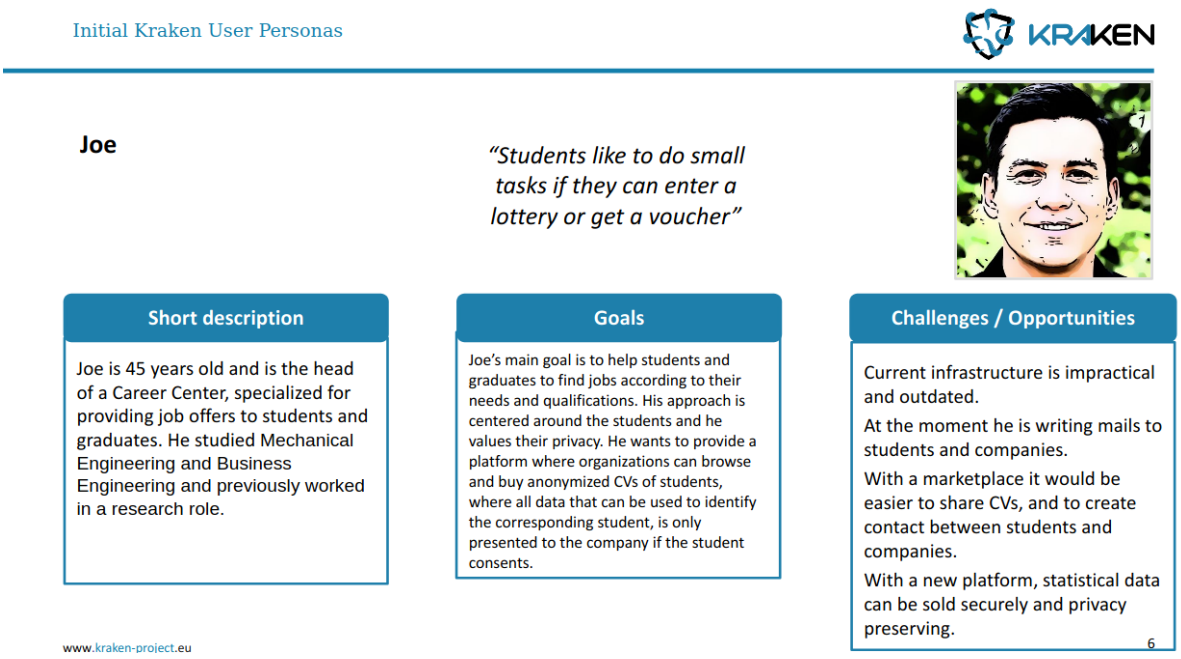


**Figure 8: The Persona Joe**

### 3.4.4 Sara

**Role in KRAKEN ecosystem**: Data Provider

**Sector**: Education

**User Group**: University

This Persona ('Sara') was created based on the Job Guidance Office interview (Section 3.2.2). She is a representative of career service managers at public universities, supporting students and graduates to access the job market, a User Group that can potentially use the KRAKEN platform as a Data Provider to facilitate the identification of stage/job positions at international level. Sara is aware that global digital platforms for data sharing have great potential to improve the quality and innovation of job guidance services, primarily if they can ensure security and privacy of personal data and reduce cost and load of administrative procedures.

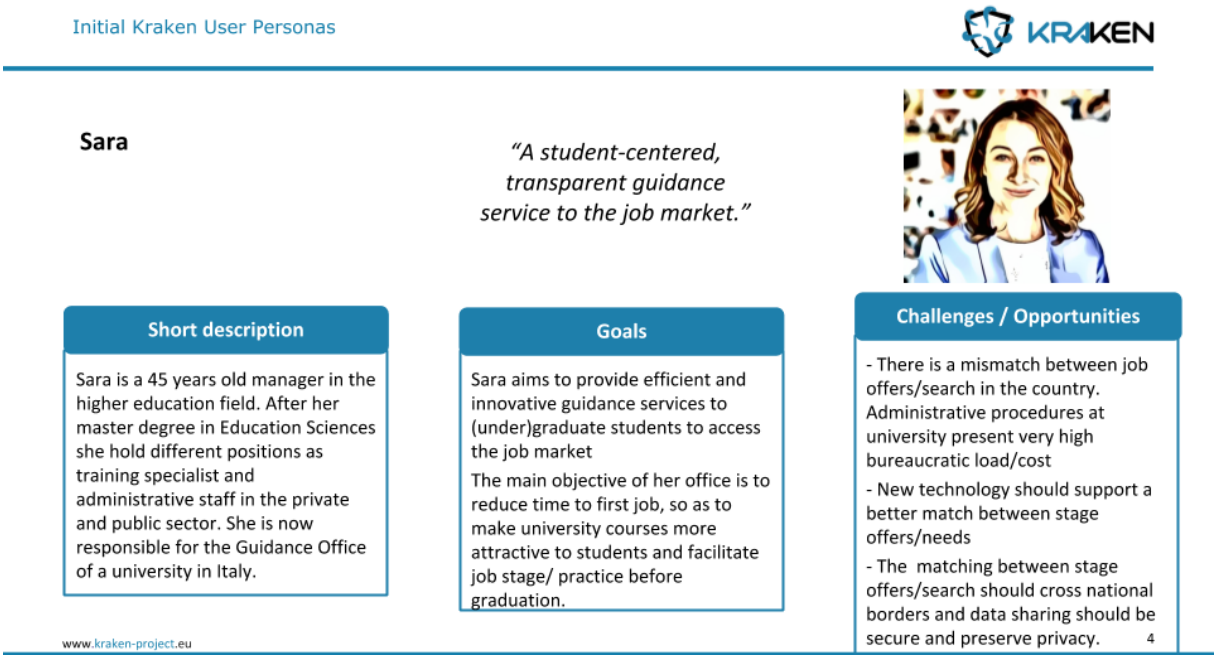


**Figure 9:The Persona Sara**

### 3.4.5 Timothy

**Role in KRAKEN ecosystem:** Data Provider

**Sector:** Education

**User group:** University

The "Timothy" Persona is based on the experience of students involved in the KRAKEN project. He is a student looking for a job in the international job market. While doing so, he needs to prove his academic accomplishments. Authentic academic evidence (e.g., graduation certificates) is an integral part of the KRAKEN platform. In addition, his records are part of the data sets provided by his university to agencies performing education statistics and other measurements.

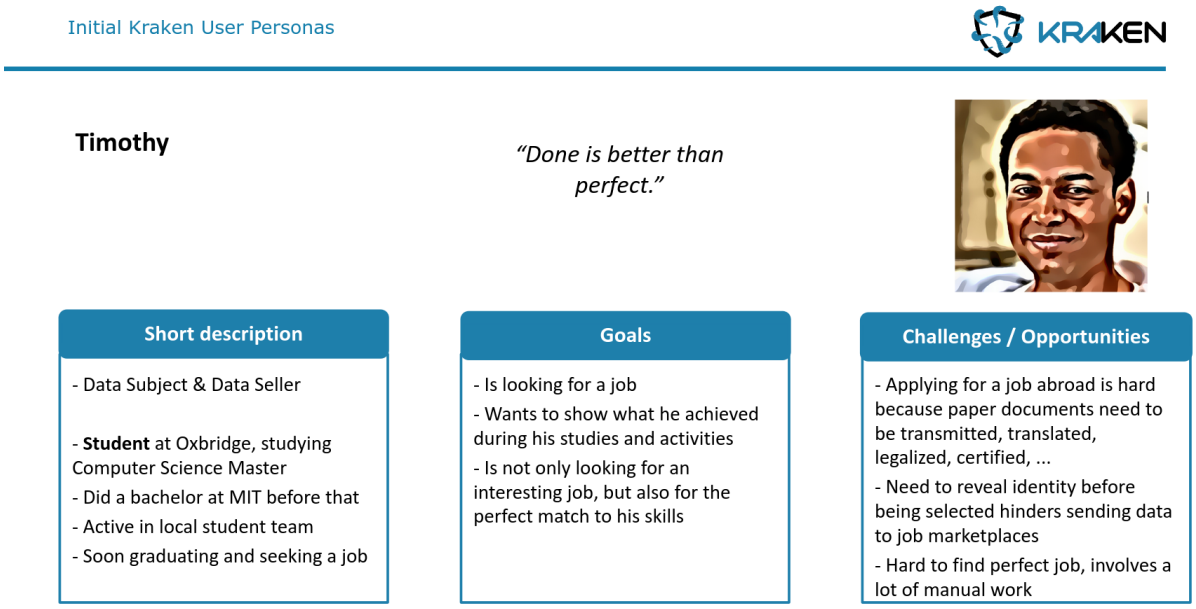


**Figure 10: The Persona Timothy**

### 3.4.6 Sabine

**Role in KRAKEN ecosystem:** Data User

**Sector:** Logistics Industry

**User group:** Human Resources

The "Sabine" Persona was derived from the interview with the University Career Center (Section 3.2.1) since she represents the actor on the other end of the same transaction. She is the head of an HR department. While looking for talent on the international job market, she needs a system to verify the credentials provided by possible hires and applicants.

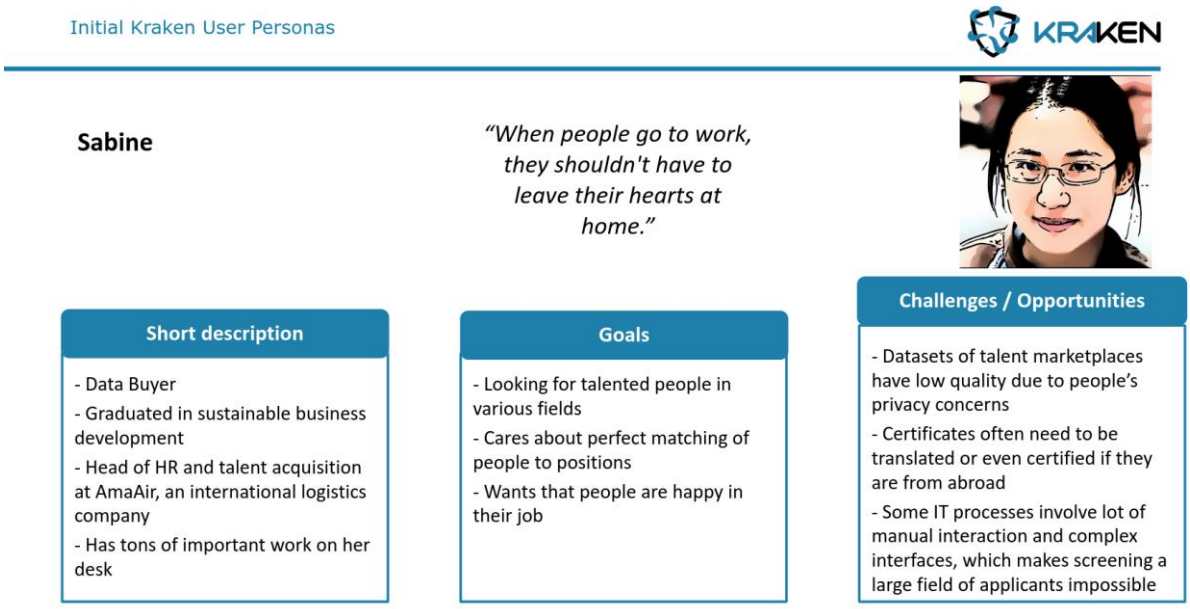


**Figure 11: The Persona Sabine**

### 3.4.7 Thomas

**Role in KRAKEN ecosystem:** Data User

**Sector:** Government

**User group:** Educational Research

The "Thomas" Persona is based on conversations with government agencies in the field of education statistics. Thomas works at a government agency that is responsible for evaluating the education system. In addition, he wants to promote the permeability of education as a whole. For both goals, he needs access to various sources of data. Since data privacy regulations often don't allow the combination of data from different sources, only privacy-enhancing technology, as used in KRAKEN, would allow him access to this data.

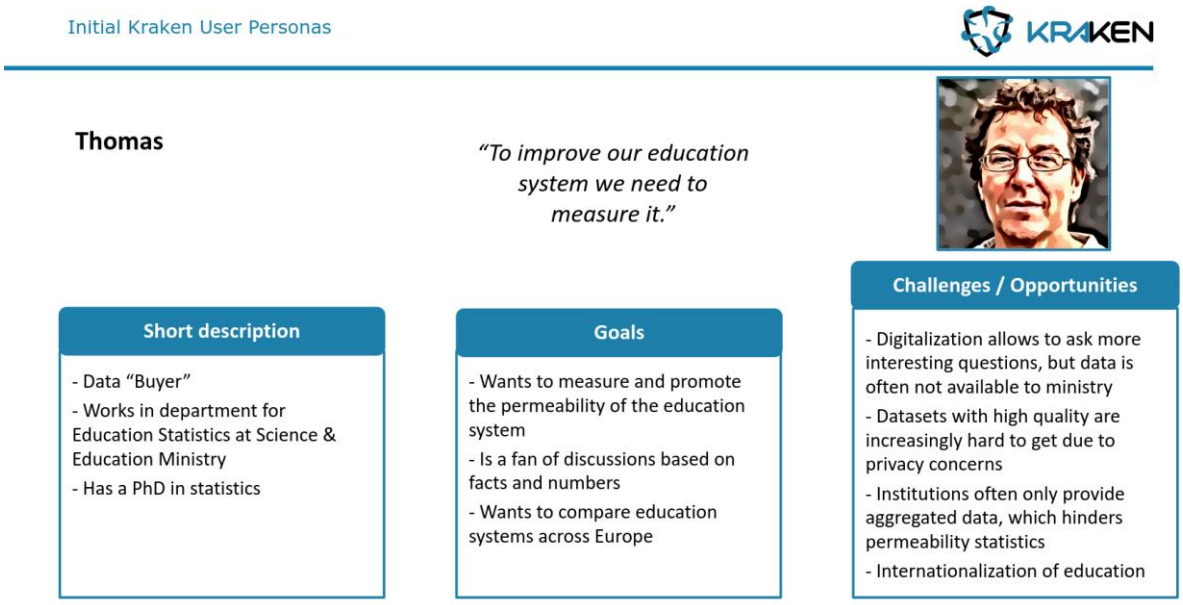


**Figure 12: The Persona Thomas**

## 3.5 Initial User Stories

This section describes the process we followed for creating the initial User Stories and gives some examples of the identified User Stories. These User Stories were derived during several workshops devoted to the creation of the User Stories, where technical and legal partners following our agile methodology participated. The drafted User Stories were stored on Taiga[1] and will be refined, detailed, and split into tasks in the scrum meetings over the next months to create the final User Stories as a source for defining development specifications.

The steps followed for creating this initial version of the User Stories are the following:

- The User Stories were created based on this standard template [2,3]:
  - **As a** <type of user>
  - **I want to** <do something>
  - **So that** <I can achieve some business value>

[1] https://taiga.io/

- Besides the basic structure of the User Story, we also defined some Acceptance Criteria. These criteria state some prerequisites for the integration of the User Story and are structured as follows:
    - **Given** <context>
    - **When** <action>
    - **Then** <expected result>
- Based on these templates, the initial User Stories have been created during several dedicated workshops, conducted by a scrum expert, where the scrum teams of the main KRAKEN pillars (Self-Sovereign Identity, crypto analytics and marketplace) participated.
- During these workshops, we defined a specific work-flow to create the User Stories exhaustively:
    - Identify the users and their needs
    - For each user:
        - Identify the interactions with the product (journey)
        - Split the interactions into small activities (steps)
    - For each step:
        - Split it into tasks

Over the next paragraphs, we give some of the initial User Stories that were derived during the workshops.

**Marketplace User Registration**

- As a user of the KRAKEN platform, I have to complete a registration process to use the KRAKEN's services.

**Generate Key material**

- As a selling user, I want to create key material, to sell data in a privacy-preserving way.

**Authentication – Decentralized Identifier (DID) connection with a backend SSI agent**

- As a user, I want to perform a first anonymous enrollment to a service (web site) using an invitation to be able to authenticate to the service many times subsequently.
- Acceptance criteria:
    - The authentication page (which I see on a browser) must, therefore, provide a way for me to connect to the service (e.g., showing a Quick Response (QR) code)
    - The authentication must grant that the service will recognize me at any subsequent access the web site will not require any information about me - it will only know me by a blind identifier

**Authentication - Full authentication with mobile**

- As a user, I need to be able to run a full authentication process starting from a web site using my mobile.
- Acceptance criteria:
    - The establishment of the new login and authentication must use the DID protocol.
    - The app must be able to read the QR code presented by the marketplace.

# 4 Conclusion

This document provides an initial report of the progress for creating the User Stories containing the main characteristics of KRAKEN from the user perspective. To derive the User Stories, we used the methodology described in Section 2, and the scrum methodology [1] applied to the KRAKEN project.

Based on these pillars, a group of relevant users has been identified. We contacted some organizations and individuals, and in turn, we conducted multiple interviews. Additionally, we launched a web survey. As a result, we were able to create numerous Personas that will be used in the final refinement of the two marketplaces and the respective User Stories.

These preliminary results gave us an initial picture of the specific needs and aspects of the actors involved in the data sharing process. According to this first approach, the initial Persona profile was defined.

For the next document of this series, which is to be delivered by October 2021, we plan to conduct more face-to-face interviews. We do this to obtain more precise information from the different bodies and individuals based on the data retrieved from the initial interviews performed for the current report. Moreover, a large number of responses to the survey are expected by that time, which will give a comprehensive user experience perspective.

The aim is to have a clear view of the needs of the different actors participating during the data exchange activities. This is done to provide enough information to the technical partners for developing a trusted and secure KRAKEN environment to put on the market, which covers the needs and expectations of the Data Providers and data consumers beyond the end of the KRAKEN project.

# 5 References

[1] Schwaber K., & Sutherland J. (2017). The Scrum Guide, https://www.scrumguides.org/docs/scrumguide/v2017/2017-Scrum-Guide-US.pdf, retrieved July 25, 2020.

[2] Klingenberg T., & Draba T (2020). What is a User Story, https://t2informatik.de/en/smartpedia/user-story/, retrieved Aug 21,2020.

[3] Lucassen G., Dalpiaz F., Werf JMEM, Brinkkemper S. (2016) The Use and Effectiveness of User Stories in Practice. In: Daneva M., Pastor O. (eds) Requirements Engineering: Foundation for Software Quality. REFSQ 2016. Lecture Notes in Computer Science, vol 9619. Springer, Cham. https://doi.org/10.1007/978-3-319-30282-9_14, pp. 205–222












@KrakenH2020

Kraken H2020

**www.krakenh2020.eu**

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 871473