



**BROKERAGE AND MARKET PLATFORM  
FOR PERSONAL DATA**

*D2.6 Marketplace Technical  
Specification*

[www.krakenh2020.eu](http://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 837854



## D2.6 Marketplace Technical Specification

<b>Grant agreement</b>	837854
<b>Work Package Leader</b>	InfoCert
<b>Author(s)</b>	Rob Holmes (TEX), Donato Pelegrino (TEX)
<b>Contributors</b>	Michael Malka (TEX), Davide Zaccagnini (LYN), Minos Garofalakis (LYN), Edwin Morley-Fletcher (LYN)
<b>Reviewer(s)</b>	Franco Nieddu (SIC), Juan Pérez Baún (Atos)
<b>Version</b>	Final
<b>Due Date</b>	30/11/2020
<b>Submission Date</b>	25/11/2020
<b>Dissemination Level</b>	Confidential

### 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	01/09/2020	Initial version	Rob Holmes
v0.2	30/10/2020	Version for internal review (including Lynkeus contributions)	Rob Holmes
v0.3	19/11/2020	Version after peer review with amendments	Rob Holmes
v1.0	25/11/2020	Submitted version	Atos

## Table of Contents

1	Introduction .....	10
1.1	Purpose of the document.....	10
1.2	Structure of the document.....	10
1.3	Glossary used in this document.....	11
1.4	Background Technologies.....	12
1.4.1	The Streamr Technology Stack .....	12
1.4.2	The My Health My Data Platform .....	15
2	Marketplace Scope.....	16
2.1	User Categories.....	17
2.2	Data Types.....	17
2.2.1	Data Streams.....	17
2.2.2	Batch Datasets .....	17
2.3	Data Sharing Modalities .....	17
2.3.1	Sharing of Anonymous Datasets .....	18
2.3.2	Secure Sharing of Data .....	21
2.3.3	Privacy Preserving Analytics Results (By Secure Multi-Party Computation) .....	23
2.4	Data Products .....	24
2.4.1	Real-Time Data Product.....	26
2.4.2	Data Unions Product .....	27
2.4.3	Batch Data Product .....	27
2.4.4	Data Analytics Product .....	28
2.5	Data Sources.....	28
2.5.1	Health Data.....	28
2.5.2	Education Data.....	29
2.6	Consent .....	30
2.7	Payment .....	30
2.8	Target Countries .....	31
3	Technical Requirements .....	32
3.1	Functional Requirements .....	32
3.2	Performance Requirements .....	33
3.3	Testing .....	34
3.4	Environments.....	35
3.5	Logging .....	36
3.6	Monitoring.....	36
3.7	CI Tools / Process .....	37

4	Marketplace Specifications.....	38
4.1	System Architecture.....	38
4.1.1	Marketplace.....	39
4.1.2	Streamr Data Transfer & Payment Subsystem .....	40
4.1.3	User System and Cloud.....	41
4.2	Interfaces with other KRAKEN Subsystems .....	42
5	User Interface .....	43
5.1	Marketplace Web Frontend .....	43
5.1.1	Registration and Login.....	43
5.1.2	Explore Data Products.....	43
5.1.3	Data Product Description Page.....	44
5.1.4	Create Data Products .....	44
5.1.5	Manage Data Products.....	44
5.1.6	Subscribe / Purchase a Data Product .....	45
5.1.7	Transaction History .....	45
5.2	Mobile App Frontend .....	45
6	Conclusion .....	46
7	References .....	48

## List of Tables

<i>Table 1: Glossary.....</i>	<i>12</i>
<i>Table 2: List of existing marketplace functional requirements .....</i>	<i>33</i>
<i>Table 3: List of initial marketplace performance requirements .....</i>	<i>34</i>
<i>Table 4: List of initial marketplace testing requirements .....</i>	<i>35</i>
<i>Table 5: List of initial marketplace environments requirements .....</i>	<i>35</i>
<i>Table 6: List of initial marketplace requirements for logging .....</i>	<i>36</i>
<i>Table 7: List of initial marketplace monitoring requirements .....</i>	<i>36</i>
<i>Table 8: List of initial marketplace requirements for the CI tools / process .....</i>	<i>37</i>

## List of Figures

<i>Figure 1: Components of the Streamr Technology Stack (in blue)</i> .....	13
<i>Figure 2: An example of raw event data from a data source travelling through the Streamr Network to a subscribing Decentralized app (Dapp).</i> .....	13
<i>Figure 3: The existing Streamr marketplace</i> .....	14
<i>Figure 4: Three main data sharing modalities and their levels of complexity and scalability</i> .....	18
<i>Figure 5: Anonymized data sharing modality for batch datasets</i> .....	19
<i>Figure 6: Anonymized data sharing modality for real time data streams</i> .....	20
<i>Figure 7: Secure sharing of data modality for batch datasets</i> .....	22
<i>Figure 8: Secure sharing of data modality for real-time data streams</i> .....	23
<i>Figure 9: Privacy-preserving analytics results modality for real-time data streams</i> .....	24
<i>Figure 10: Example of a Data Product in the existing Streamr marketplace</i> .....	25
<i>Figure 11: Example of a Data Unions Product in the existing Streamr marketplace</i> .....	26
<i>Figure 12: Example of a subscription to a real-time data product in the existing Streamr marketplace</i> .....	27
<i>Figure 13: Existing marketplace Subsystem Architecture</i> .....	38
<i>Figure 14: Sub-components of the backend component</i> .....	40

## List of Acronyms

Acronym	Description
AWS	Amazon Web Services
CI	Continuous Integration
ERC-20	Ethereum Request for Comment-20
GDPR	General Data Protection Regulation
IoT	Internet of Things
MHMD	MyHealthMyData
P2P	Peer-to-Peer
PoC	Point of Care
QA	Quality Assurance
SSI	Self-Sovereign Identity
SMPC	Secure Multi Party Computation
UI	User Interface



## Executive Summary

The KRAKEN marketplace builds upon and enhances existing technologies developed in two projects; the Streamr and MyHealthMyData (MHMD) projects. The Streamr Project is a well-known web3 project that is building a fully decentralised and open infrastructure for real-time data sharing and data monetization. The MHMD project is the result of a Horizon 2020 research and innovation action which aimed at establishing an open biomedical information system centred on the connection between organisations and individuals.

The aim of the KRAKEN marketplace is to realise a General Data Protection Regulation (GDPR) compliant infrastructure that allows individuals and organisations within the project's two pilot sectors to protect, share and trade trusted biomedical, wellbeing and educational data with interested third parties. It securely connects providers and consumers of high-quality batch datasets and real-time data streams and leverages a dual-blockchain solution to enforce the business and legal logic behind all data transactions and facilitate payment between data providers and data consumers.

This document presents the initial technical specification for the core component of the KRAKEN platform; the marketplace subsystem. The KRAKEN project has adopted an Agile approach based on the SCRUM methodology with periodical sprints and sessions to develop the user stories for the marketplace. Therefore, at the end of each sprint, the marketplace design and specifications are re-evaluated and refined. This document therefore forms the initial basis for the final design and specification of the marketplace which will be presented in D2.7 - Design for Marketplace Reference Implementations, which is programmed for January 2022. Included within this document is an initial definition of what is within the scope of the KRAKEN marketplace, its functional requirements and performance requirements, its technical specifications and architectural design, and the additional features and functionality to be added to the marketplace UI.

The main goal of this deliverable is to guide the implementation of the marketplace by the design and development teams within TEX and Lynkeus. But it shall also serve as a reference point for the Self Sovereign Identity (SSI) and crypto teams to aid their development and integration with the marketplace.

# 1 Introduction

## 1.1 Purpose of the document

This document is a WP2 deliverable. It details and describes the initial progress made to date by the marketplace team on specifying the technical aspects of the marketplace component within KRAKEN. It starts off by describing and providing information on the background technologies to be integrated and further developed within the project. It then identifies the initial scope of the KRAKEN marketplace subsystem, its technical requirements and specifications, and finally provides an initial indication of the functionalities to be made available to users in the application's frontend via the marketplace UI.

The marketplace subsystem is developed within WP5 and will be realised through the integration of two established blockchain-based data platforms, MHMD[1] and Streamr[2]. The subsystem will be integrated with two additional subsystems within the wider KRAKEN architecture which are being developed as part of WP3 and WP4: the SSI and Crypto subsystems.

This document is to be used as a reference point to assist the design and development teams within TEX and Lynkeus who are responsible for the implementation of the KRAKEN Marketplace. As this is a month 12 submission, it is stressed that this deliverable is based on the existing knowledge and progress made in the first year of the project by the marketplace team. Further refinements are to be made to the integrated marketplace architecture which will be communicated in D5.3 - Initial KRAKEN marketplace Integrated Architecture and D5.4 - Final KRAKEN marketplace Integrated Architecture in June 2021 and December 2021. The final design of the marketplace will also be presented in D2.7 - Design for Marketplace Reference Implementations, which is programmed for January 2022.

## 1.2 Structure of the document

This document commences in [Section 1](#) with a short introduction to the background technologies to be used for the development of the KRAKEN marketplace. It then gives an overview of the key features, functionalities and data types that fall within the scope of the KRAKEN marketplace within [Section 2](#). This Section also provides a description of the three main data sharing modalities that will be implemented within the data exchange infrastructure during the project; 1) Sharing of anonymised datasets; 2) Secure sharing of data, and; 3) Privacy-preserving analytics results (by secure multi party computation).

Following on from the opening Sections on background technologies and scope, [Section 3](#) provides an overview of the initial functional requirements identified to date for the KRAKEN marketplace. The initial functional requirements are fed by the legal requirements and implementation guidelines identified within D7.2[3], the market analysis within D6.2 and the user requirements identified within D5.1[4] and will be refined as these activities progress during the course of the project. In addition, [Section 3](#) also details the performance requirements and the approach towards testing of the software, hosting environments, error monitoring and logging and the Continuous Integration (CI) process.

[Section 4](#) describes the marketplace specifications, detailing the various sub-components of the two main components of the marketplace architecture; 1) the marketplace software, and; 2) the data transfer and payment system. It also describes the interfaces with the other two key components of the KRAKEN platform, the SSI and crypto.

The KRAKEN marketplace User Interface (UI) will use the Streamr marketplace UI as a base for adding the additional features and functionalities identified during the lifetime of the project. [Section 5](#) completes the marketplace technical specification by detailing the specific additional features and

functionalities to be added to the marketplace UI that have been identified based on the current understanding of the requirements.

### 1.3 Glossary used in this document

Table 1 below provides a glossary of terms used within this document and their definitions.

Term	Definition
Agreements	Agreements occur when a data purchase or access request in the marketplace is confirmed by the Agreements smart contract as compatible with the policies set by the data provider. The records of all agreements are stored in the Agreements smart contract.
Batch Data	Batch data is a static record or collection of files.
Data Catalogue	The Data Catalogue is a component of the KRAKEN marketplace subsystem that will collect metadata about the Data Products published on the marketplace.
Data Products	A Data Product is a single stream or bundle of real-time data streams, or single file or bundle of files, or analytics results on encrypted datasets which are either sold or made available for free on the marketplace.
Data Streams	A data stream is simply a sequence of data points in time.
DATACoin	The cryptocurrency which powers the Streamr ecosystem in various ways. It is an ERC-20 token on the Ethereum blockchain <a href="#">[5]</a> .
Data Union	A Data Union is a structure that allows separate entities who generate data to pool their data in order to achieve a common goal, typically to increase its value. They share the following common components: a group of members contributing their data, an entity providing the application layer to manage the data and a discovery mechanism or marketplace for the data.
ERC-20 Token	ERC-20 tokens are blockchain-based assets that have value and can be sent and received, they are issued on the Ethereum blockchain.
Marketplace	A marketplace is an application that enables users to buy and sell access to data content. It can be filled with both paid and free products, offering data providers an opportunity to either monetise their data or make it freely available.
Metadata	Metadata is descriptive data and is separate to the content data within a Data product for sale on the marketplace. In the case of KRAKEN, it provides information about the Data Products for sale such as title, description, tags and price.
Secure Multi-Party Computation (SMPC)	A cryptographic protocol that distributes the computation of a function on input data across multiple parties (or nodes) without revealing their own individual private inputs or outputs to another party.
Peer-to-Peer (P2P)	Peer-to-Peer is a distributed application architecture that partitions tasks or workloads between various peers within a network.
Policies	Policies refer to the data access rights, or preferences set by data sellers within the KRAKEN marketplace.
Publish-Subscribe (Pub-Sub)	The pub/sub messaging pattern is a highly useful primitive for data-driven applications. It decouples services or devices that publish data from services or devices that subscribe to data. Publishers send messages to a topic, and subscribers create a subscription to a topic in order to receive messages from it.

Smart Contracts	A smart contract is a computer program that can be executed by a blockchain.
Streamr	Streamr is a distributed open-source software project building a new data economy through the establishment of a global, open, peer-to-peer network for data transportation and data marketplace for data monetization.
Streamr marketplace	The Streamr marketplace is an application that enables users to buy and sell access to real-time data content on the Streamr Network. The marketplace is filled with both paid and free products, offering data producers an opportunity to either monetise their data or make it freely available to everyone. Products have time-based subscription periods, and the currency used on the marketplace is DATACoin.
Streamr Network	The Streamr Network is a scalable real-time messaging system, which enables applications and devices such as Internet of Things (IoT) sensors to share and trade their data.
Verifiable Credential	A tamper-evident credential that has authorship that can be cryptographically verified.

Table 1: Glossary

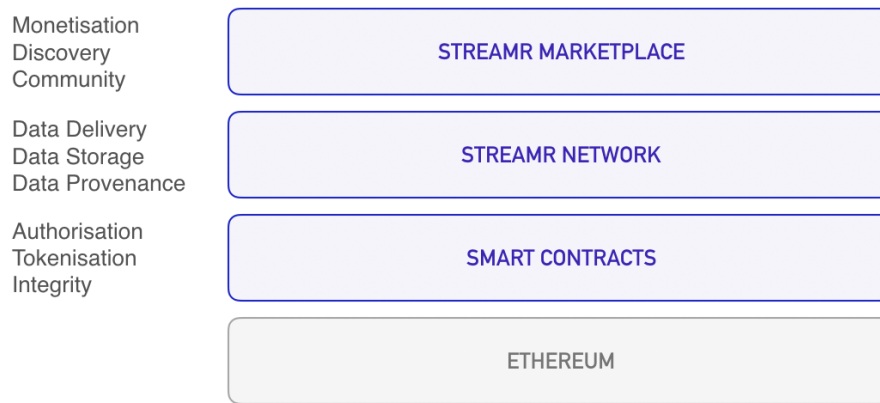
## 1.4 Background Technologies

This Section provides a short introduction to the background technologies that are being used for the development of the KRAKEN marketplace. It includes a description of the Streamr Technology Stack and MHMD platform.

### 1.4.1 The Streamr Technology Stack

The Streamr Project's mission is to build a fully decentralised and open infrastructure for real-time data sharing and monetisation. It has already made significant progress towards this mission, having established a Peer-to-Peer (P2P) publish-subscribe (pub-sub) network for low latency real-time data transfer between cross-sector IoT sensors and applications (the Streamr Network), and a data marketplace application that is built on top of this network and enables **anyone** to share and monetise their data streams (the Streamr marketplace).

The layers of the Streamr Technology Stack are built on top of the Ethereum Blockchain, which supports the operation of the Streamr Network and marketplace via smart contracts. Figure 1 below shows each of these different layers of the Streamr Technology Stack and lists at a high level their functionalities.



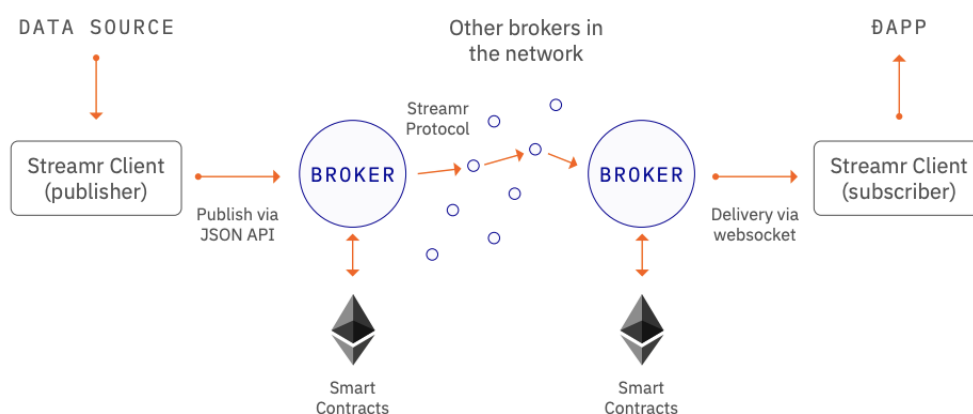
**Figure 1: Components of the Streamr Technology Stack (in blue)**

#### 1.4.1.1 The Streamr Network

The public Streamr Network consists of primitives (events and streams) and Streamr broker nodes which establish a P2P network. It hosts a publish/subscribe mechanism and supports decentralized storage and decentralized messaging.

The network is a result of multiple broker nodes, which collectively form a distributed peer-to-peer network and pub/sub messaging service. The Broker is a piece of software, intended to be installed and run on always-on machines with stable bandwidth. It connects to other nodes in the network and starts to route data traffic through them.

Figure 2 below shows an example of raw event data from a data source travelling through the Streamr Network. The network uses the underlying Ethereum Blockchain for its operations. It utilises Ethereum smart contracts to co-ordinate the broker nodes, store stream metadata and for permissioning / access control. However, when transporting data streams from publishers to subscribers all of the raw event data remains off-chain to overcome the problems associated with scalability and cost of data being on-chain. Throughput of the network scales logarithmically with the number of participating nodes.



**Figure 2: An example of raw event data from a data source travelling through the Streamr Network to a subscribing Decentralized app (Dapp).**

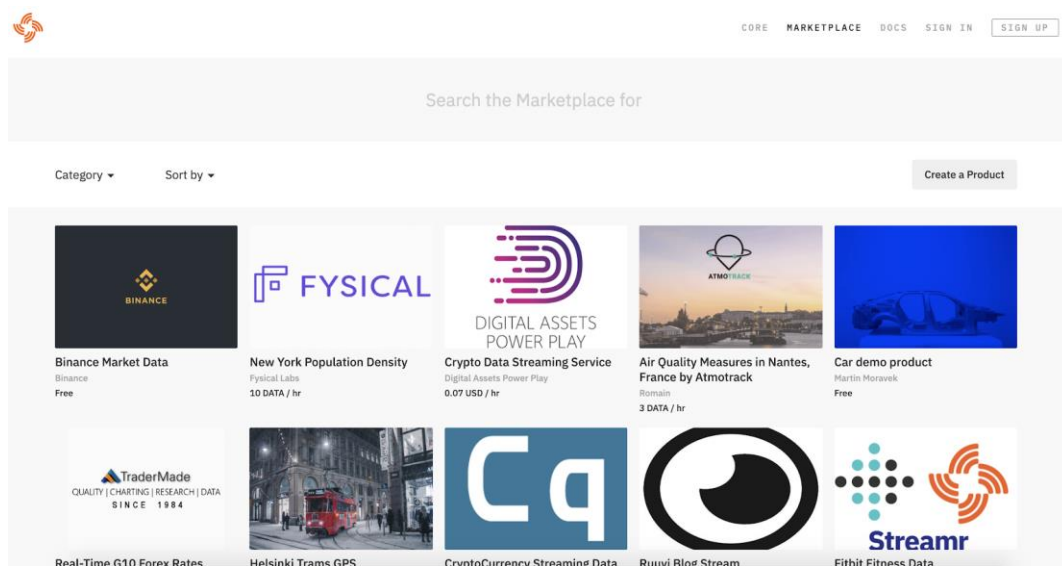
The network is being released in four stages; Monk, Corea, Brubeck and Tatum. In the first stage (Monk), the network was a centralized pub-sub infrastructure. In Q4 of 2019 the network reached the second stage (Corea), which was the first implementation of the P2P network. At this stage, broker

nodes are all hosted by Streamr. Over the course of the next two years (2021 and 2022) and next two milestones (Brubeck and Tatum), the focus is on moving the P2P network towards a fully decentralized model. The Brubeck stage is programmed to be launched in Q1 of 2021 and will include end-to-end encryption and the first step towards full decentralization: trusted parties running nodes in the network. The final stage (Tatum) is programmed for launch in 2022, this will include built-in network token economics where node runners receive Streamr's ERC-20 token (DATAcoin) as an incentive and payment for the data transportation service, allowing anyone to act as a node runner in the network, and as a result, full decentralization.

#### 1.4.1.2 Streamr marketplace

The KRAKEN marketplace will be built by extending a forked code base of the Streamr marketplace. This KRAKEN marketplace will serve as the basis for which all of the necessary integrations, extensions and additional functionalities required by the health and education pilots will be built upon.

The existing Streamr marketplace (see Figure 3 below) is a web-based application that makes a wide selection of timestamped data discoverable through its user interface, and available for subscription. The marketplace allows anyone with an Ethereum address to publish their data streams as Data Products and monetise them.



**Figure 3: The existing Streamr marketplace**

Functionality is provided to integrate a user's data streams, provide the necessary descriptive metadata, and package their data stream as a Data Product. A data seller can set the name, description, category and price for the data, and publish it to make it discoverable to interested buyers within the marketplace's UI.

Payments for data subscriptions are made using an ERC-20 token called DATAcoin, and subscriptions and access rights are implemented and controlled by smart contracts on the Ethereum Blockchain. This offers a decentralized way to ensure that payments are made as agreed before the data consumer is given access to the data streams contained within the advertised Data Product.

In its present form, it is an open marketplace where anyone can publish or subscribe to data streams, but there is no dynamic consent mechanism for controlling who can access data products advertised within the data marketplace.

### 1.4.2 The My Health My Data Platform

The My-Health-My-Data project was established to interconnect individuals, hospitals, research centres, pharmaceutical companies, universities, among others, within a biomedical information network that allows them to control access to healthcare data for research purposes in line with the GDPR.

It relies on a distributed secured blockchain network based on Hyperledger Fabric, instead of a central authority, in order to appropriately manage the data-sharing pipeline and regulate data access on the basis of user-defined permission/consent settings. It uses smart contracts to enforce different complex business rules according to the data category, security level, purpose of use, network members and GDPR requirements. It includes the My-Health-My-Data mobile application which acts as a dynamic consent interface and enables data owners to allow, refuse and withdraw access to their data according to different types of potential usage.



## 2 Marketplace Scope

Through the integration of three major subsystems (Self Sovereign Identity (SSI), data marketplace and Crypto tools) the KRAKEN platform aims to realise a GDPR compliant infrastructure that will allow individuals and organisations to protect, share and trade trusted biomedical, wellbeing and educational data with interested third parties via the KRAKEN data marketplace.

The KRAKEN marketplace subsystem will act as an open and decentralised exchange for data within the wider KRAKEN platform; securely connecting providers and consumers of high-quality datasets and data streams, and leveraging a blockchain network to enforce the business and legal logic behind data transactions. There will be one single marketplace for both the Health and Education pilots.

The marketplace subsystem will leverage the Streamr Network for the delivery of data streams in real-time between individual citizen and institutional/organisational data providers and interested data consumers. Through the use of this P2P pub-sub network, the KRAKEN marketplace will have access to an open, neutral and scalable data streams pipeline. When organisational or institutional participants interested in the legal and ethical sharing and monetisation of data streams join the marketplace, although not strictly necessary, they will in future also have the option to host their own network nodes once the Network reaches its final Tatum milestone, resulting in a truly common and shared infrastructure for the sharing and trading of data streams.

Users of the KRAKEN marketplace will also be able to monetize and package Batch datasets, or static records or files, as Data Products. In this case, the marketplace will not store or be able to access the datasets, and they will be stored and accessed from users' own cloud-based storage.

At a high level, the key features in the scope of the KRAKEN data marketplace are outlined below:

- Users (data providers and data consumers) will be able to **register** and **authenticate** to the KRAKEN data marketplace by using their SSI wallet to prove their identity.
- A **Data Catalogue** will be provided and act as a “shop window” that lists the available datasets for interested data buyers to browse.
- Data providers will be provided with the functionality to **package and advertise** their batch datasets and data streams as Data Products within the KRAKEN data marketplace.
- The KRAKEN data marketplace will **restrict access** to Data Products created by data providers based on their preferences, which will define the parts of a user's data that can be accessed, by whom and for what purpose, whilst also providing the option to withdraw consent at any moment.
- Data providers will be able to select one of three different methods for **securing their data** and **preserving their desired level of privacy** prior to sharing data. These will be anonymous data exchange, secure sharing of data and privacy-preserving analytics results (by Secure Multi Party Computation).
- Data providers will be able to use the KRAKEN data marketplace to **set a price** for their tradable data and **receive payment** from data consumers.
- Upon completion of payment, agreement for data access, and checks of legal compliance of the transaction, data streams will be **delivered** from data providers to data consumers using **Streamr's P2P pub-sub network**, whilst batch datasets will be accessed via an **encrypted link** to a users' own cloud data storage provided by the Data Provider.

A more comprehensive list of requirements that will drive the design and development of the KRAKEN marketplace subsystem are identified in [Section 3](#) Marketplace Requirements.



The following subsections expand in more detail on the scope of the KRAKEN marketplace Subsystem, covering the categories of user that the marketplace is targeting, types of data to be shared, potential data sources to be supported, data sharing modalities, registration and authentication, consent, payment methods and the target countries from which users will be able to access the marketplace.

## 2.1 User Categories

There are three types of users that fall within either the category of Data Providers or Data Consumers which must be catered for within the marketplace. These are described within the KRAKEN Grant Agreement and are repeated here:

1. **Individuals** producing device data and carrying personal data on mobile apps or personal data storage systems who want to monetize such data in a secure and privacy protected way.
2. Both **private and public institutions or organisations** who store individual data on which consent has been given by original data owners or can be obtained by having individuals using a dynamic consent application.
3. **Market stakeholders**, in particular health-tech companies, insurers, public authorities and wellbeing service providers interested in acquiring aggregated data sources.

## 2.2 Data Types

The KRAKEN marketplace will offer for sale or to be shared for free both batch datasets and real-time data streams. The Sections below set out the major differences in the definitions of these two data types.

### 2.2.1 Data Streams

A data stream is simply a sequence of data points in time. The data may originate, for example from devices such as Point of Care (PoC) devices in a hospital or personal health or wellbeing devices such as wearables that are carried by an individual citizen and are streamed in **real-time**. Data points contained within a data stream each consist of a timestamp and further data fields providing measurements such as heart rate and blood pressure. Data streams also have the following properties:

- Any kind of real-time data can be stored in it.
- The data will always be retrieved in the correct order.
- The data is identifiable by a unique ID.

### 2.2.2 Batch Datasets

In the context of the KRAKEN marketplace, we refer to batch data as a **static record or collection of records**. An example of this for the health pilot could be personal health records like lab results or medical histories. For the educational use case, this may be, for example, an aggregation of grades bundled together, resulting in a digital curriculum vitae for a student. It could also be a historical batch of streaming data from a wearable device or wellness application related to a specific period of time with a fixed start and end point that is stored within a user's cloud storage.

## 2.3 Data Sharing Modalities

In deciding the data sharing modalities offered by the KRAKEN marketplace consideration has been given to legal, ethical and privacy constraints, and business requirements. The specific ethical and legal

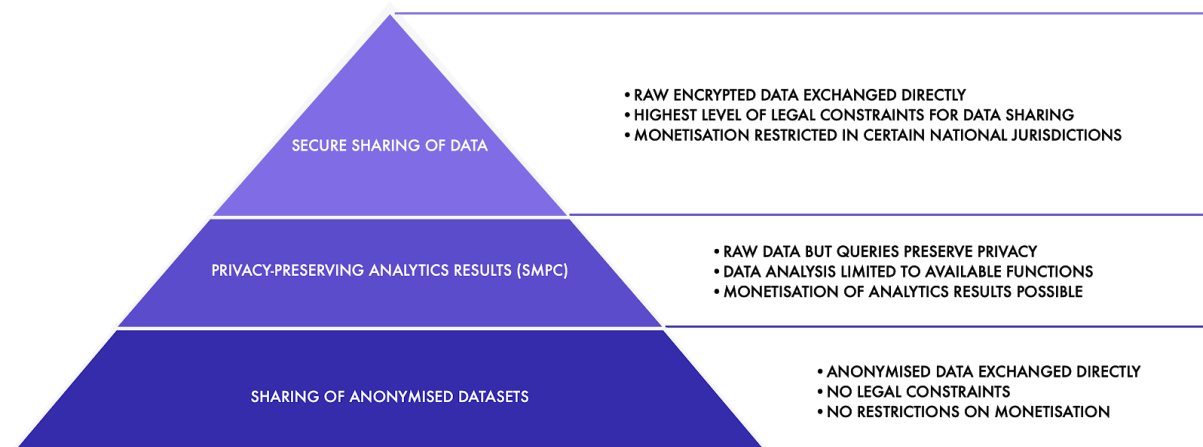
requirements and their implementation guidelines which need to be taken into account in the course of the marketplace design and development are being developed as part of the ongoing efforts on T7.2 - Ethical and Legal Analysis and Evaluation.

There are a number of legal and ethical questions that the marketplace team must ask itself during the definition of the marketplace design and the data transfer / sharing modalities to be supported by the marketplace. These include:

- What data can be legally shared and what data can be sold under both the scope of the EU GDPR and individual restrictions at a national jurisdiction level;
- What data must be protected; and
- How data owners must be allowed to exercise control over their personal data.

The KRAKEN data marketplace will implement three main data sharing modalities. Each modality has a different level of complexity for implementation and difficulty level associated with its scalability from a commercial standpoint. The three data sharing / transfer modalities currently identified are listed below and described in Figure 4:

1. Sharing of anonymised datasets;
2. Secure sharing of data;
3. Privacy-preserving analytics results (By secure multi-party computation).



**Figure 4: Three main data sharing modalities and their levels of complexity and scalability**

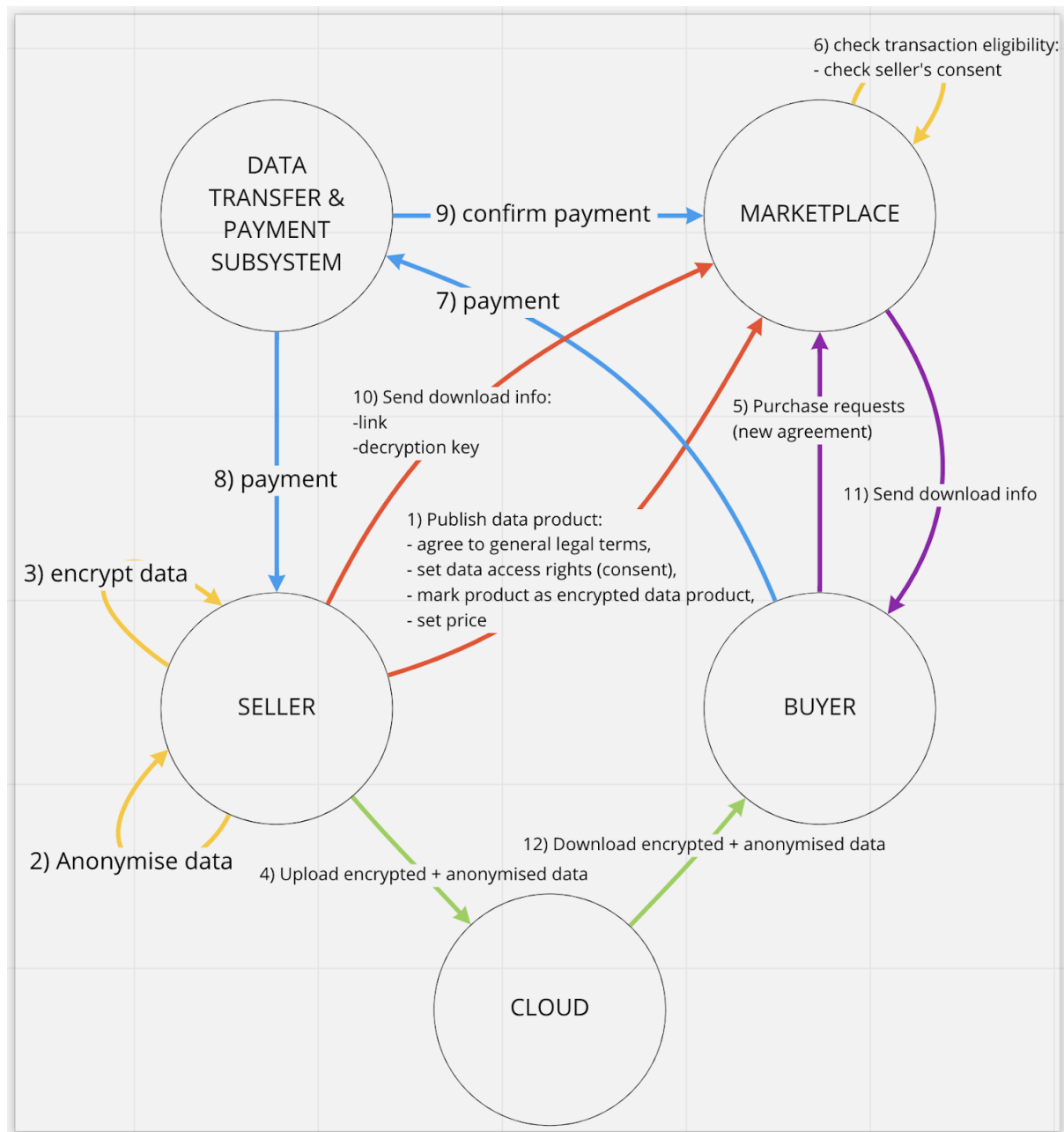
The KRAKEN marketplace will be geared towards the sharing and monetization of both batch datasets and real-time data streams. The first two data sharing modalities of anonymized data and encrypted data are planned to be applicable to both batch datasets and real-time data streams. However, for the time being, real-time data streams will not be included in the privacy-preserving analytics modality. A decision will be made at a later stage in the project as to whether this type of functionality is possible for real-time data streams.

### 2.3.1 Sharing of Anonymous Datasets

Within this data sharing modality, data producers or data owners, will be able to share batch datasets that are anonymized at source and that have been legally validated as anonymous. This is the most scalable mode of data sharing and direct data monetization from a commercial perspective, as it does not constitute personal data, and so it can be shared freely outside of the scope of the GDPR and without legal constraints on commercializing the data.

For aggregated batch datasets, it is currently planned that the platform's front end will be designed to allow a user to download external software or tools for anonymization services. The consortium is also currently exploring the possibility to implement an anonymizer within SMPC to provide a secure and privacy-preserving cloud solution to the user.

Figure 5 below describes the process of sharing of anonymous batch datasets.



**Figure 5: Anonymized data sharing modality for batch datasets**

A user (data provider) that wants to register an aggregated batch dataset on the marketplace, will begin by publishing a Data Product on the marketplace, agreeing to the general legal terms, entering the necessary descriptive information and metadata, setting their access rights and price, and marking the Data Product as an anonymized dataset. Within this workflow the user will be presented with the possibility to download the external anonymization tool in order to perform the anonymization before finalising the registration of the Data Product in the marketplace.

A buyer will then be able to make a purchase request (new agreement), that will be sent to the marketplace to check the transaction eligibility based on the seller's consent preferences. Once transaction eligibility is confirmed, the buyers' payment will be sent to the seller via the data transfer and payment subsystem (Section 4.1.2) which will trigger the anonymized and encrypted data transfer. In this case, the system will expose a link to the individual user's personal data store to the buyer.

Anonymized data streams will be directly transferred from the data provider to the data consumer over the marketplace's data transfer and payment system, which uses the Streamr Network (Section 4.1.2.2). Anonymization of these data streams will be the responsibility of the data provider. He/she will be required to ensure that the data stream is anonymised before publishing the Data Product on the marketplace and pushing the data stream over the Streamr Network. This can be performed by the data provider at various levels, for example at the device level or at the level of the third-party server that collects information from the devices (e.g. in the case of a wearable like Fitbit - on the manufacturer's servers).

Figure 6 below describes the process of sharing of anonymous real-time data streams.

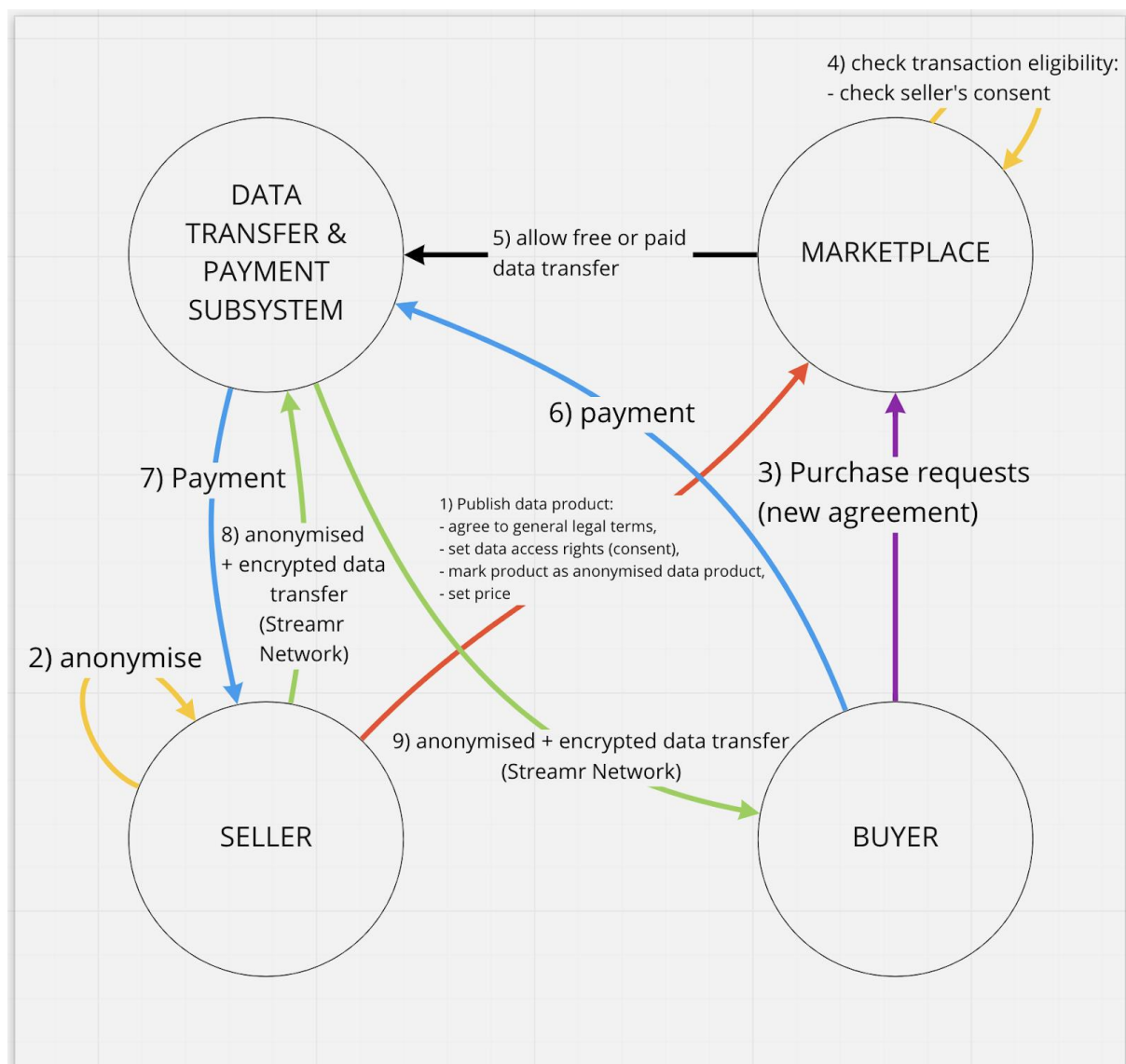


Figure 6: Anonymized data sharing modality for real time data streams

A user (data provider) of the marketplace, first, will begin by publishing a Data Product on the marketplace, agreeing to the general legal terms, entering the necessary descriptive information and metadata, setting their access rights and price, marking the Data Product as an anonymized dataset and connecting the already anonymized data stream to the Streamr Network. The end-to-end encryption of the stream is a functionality provided automatically by the Streamr Network.

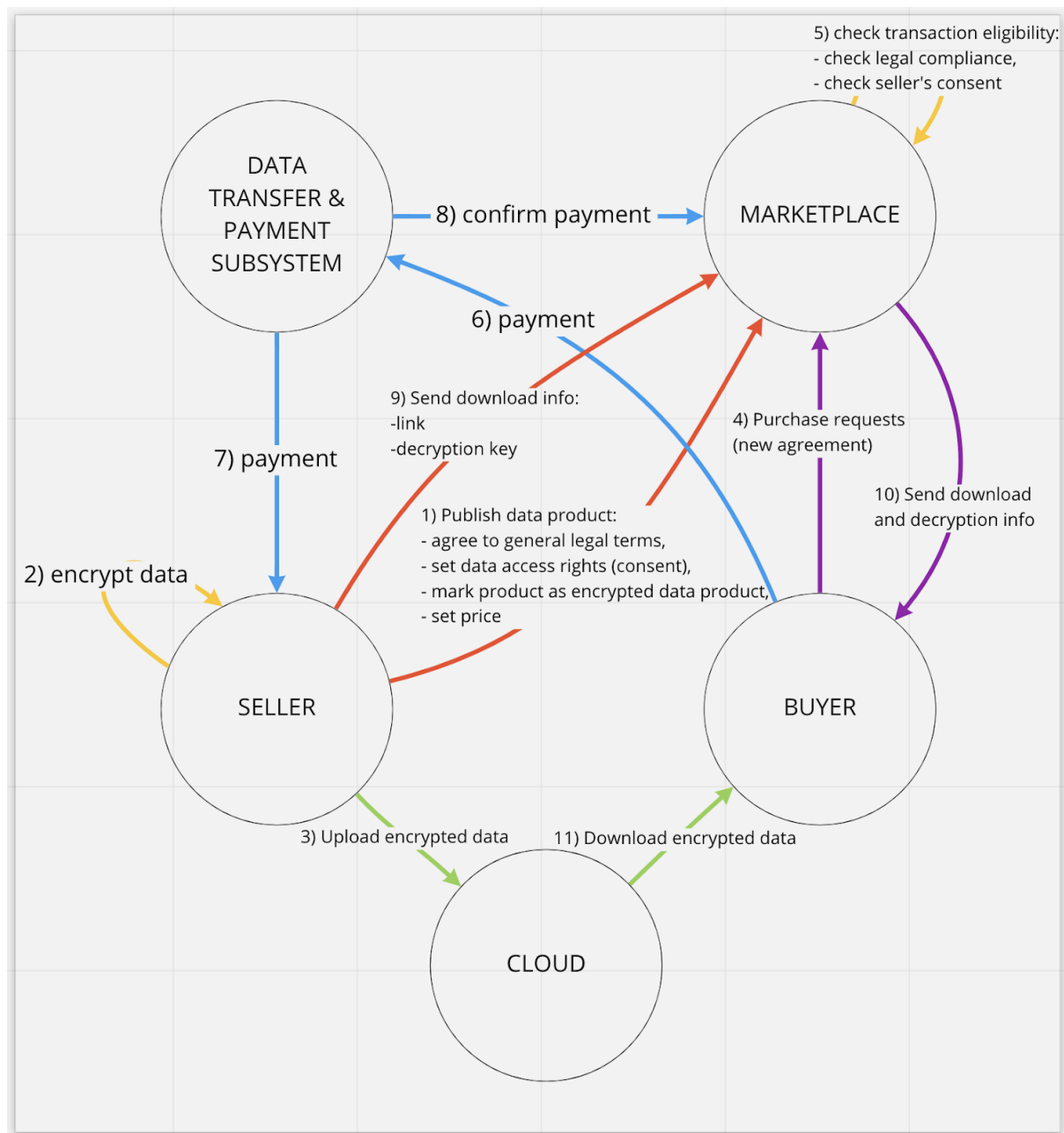
A buyer will then be able to make a purchase request (new agreement), that will be sent to the marketplace to check the transaction eligibility based on the seller's consent preferences. Once transaction eligibility is confirmed, the buyers' payment will be sent to the data transfer and payment subsystem which will trigger the encrypted data transfer via the Streamr Network to the data buyer.

### 2.3.2 Secure Sharing of Data

Data sharing via the "secure sharing of data" mode will enable data producers or owners to encrypt their data. In this mode, it is still to be determined by the legal and ethical analysis whether or not it will be possible to commercialize personal data through direct data monetization. It is understood by the marketplace team that there are various legal restrictions placed on the sharing of personal data at an EU level and also different restrictions at a national level. There is also ambiguity and uncertainty about the possibility of monetizing personal data, which must first be considered by the legal team before a decision is made on how to proceed.

Of the two planned KRAKEN pilots, data concerning a person's health falls into the special category of personal data, and therefore requires extra consideration as to the specific circumstances upon which they may be shared and/or monetised. Within the Education pilot, the data to be used is considered non-sensitive personal data, which means that apart from the general restrictions placed upon the sharing of personal data within the GDPR, no further restrictions are expected to be placed upon the data involved. It should be noted that in the case of both pilots the ethical requirements, such as the need to provide informed consent, will be considered.

Figure 7 below describes the process of sharing of encrypted batch datasets.

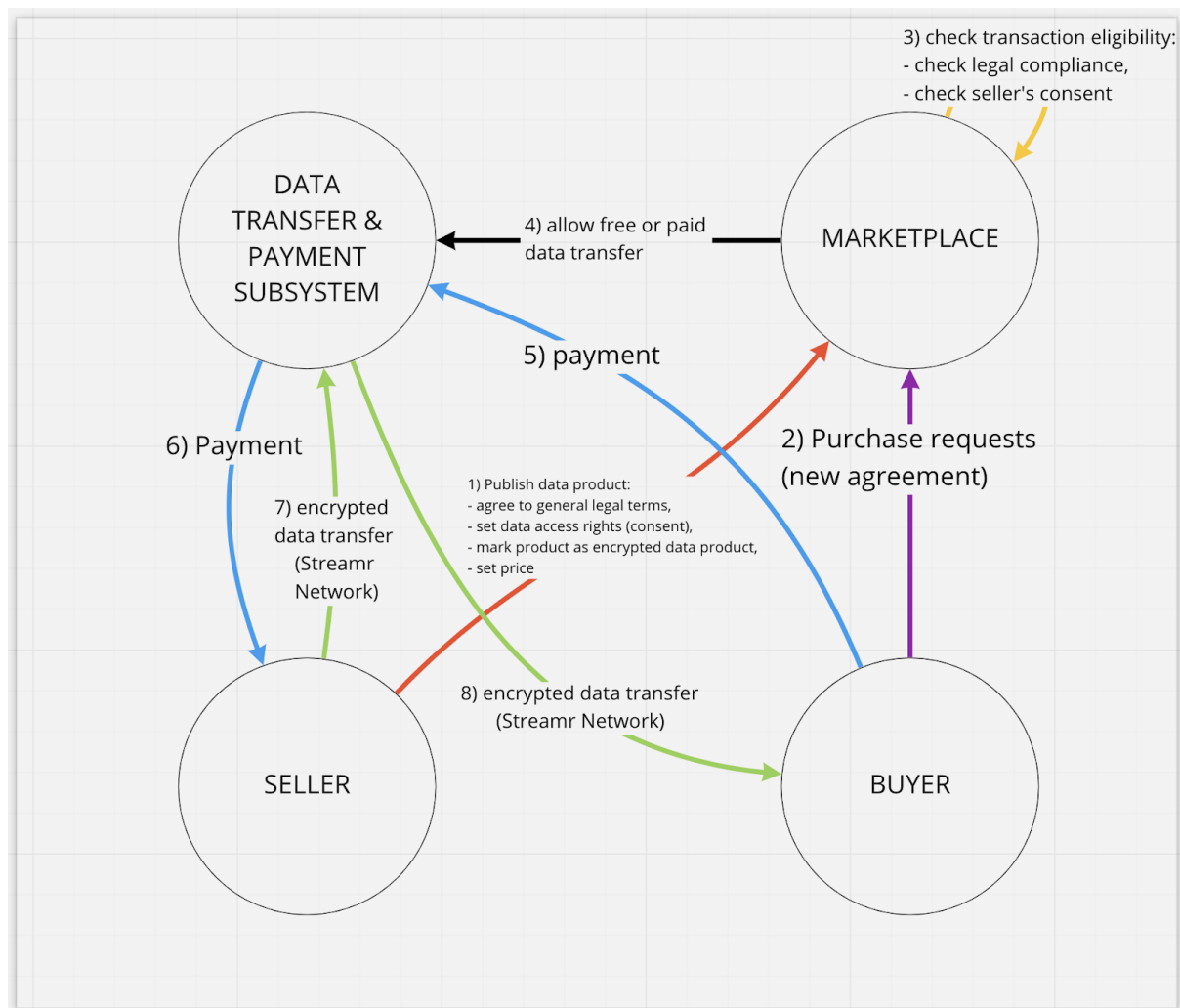


**Figure 7: Secure sharing of data modality for batch datasets**

A user (data provider) of the marketplace will first of all begin by publishing a dataset on the marketplace, agreeing to the general legal terms, entering the necessary descriptive information and metadata, setting their access rights (consents) and (if possible) price, and marking the data product as an encrypted dataset.

A buyer will be able to make an access request (new agreement), that will be sent to the marketplace to check the transaction eligibility based on the data owner's consent preferences. Once transaction eligibility is confirmed, this will trigger the encrypted data transfer. The information about the download and decryption of the dataset will be encrypted for the consumer using his/her public key. In this way no one, except the data consumer, will be able to access the data, not even the marketplace itself. The buyer will then legally become Data Controller of the received dataset.

Figure 8 below describes the process of sharing of encrypted real-time data streams.



**Figure 8: Secure sharing of data modality for real-time data streams**

For encrypted data streams, these will be directly transferred from the data provider to the data consumer over the marketplace's data transfer and payment system, which uses the public Streamr Network that provides encryption features. For encrypted datasets such as static records, the system will expose a link to the locally stored dataset or an individual user's personal data store.

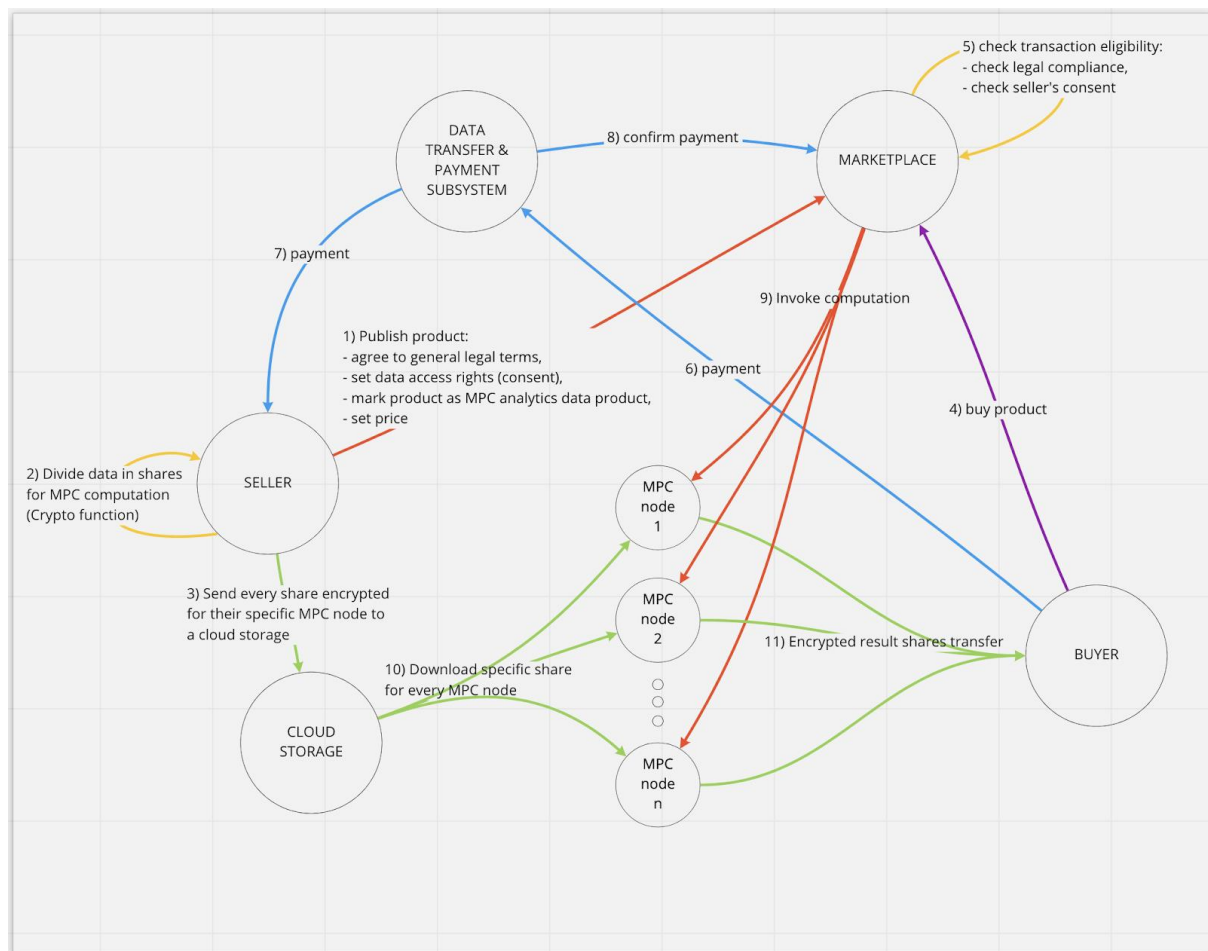
### 2.3.3 Privacy Preserving Analytics Results (By Secure Multi-Party Computation)

Within the KRAKEN marketplace, data providers or data sellers will be able to consent for their datasets to be made available for privacy-preserving analytics using secure Secure Multi Party Computation (SMPC) and data consumers or data buyers will be able to specify their desired analysis and pay for this service.

If we assume that at least one node in the SMPC network is trustworthy, then it resolves trade-offs between the benefits of making personal or private data available for certain analyses and the risks and privacy concerns associated with exposing it.



Figure 9 below describes the process of sharing of privacy-preserving analytics results.



**Figure 9: Privacy-preserving analytics results modality for real-time data streams**

This data sharing modality works as follows: a data provider signs their data to give their consent to the computation and to ensure the data's provenance - that the data originates from the data provider. They perform a secret-sharing of the data for all SMPC nodes and encrypt the shares specifically for respective nodes. The SMPC nodes decrypt their shares, verify the signature to ensure that they are allowed to perform the computation, and then perform the actual computation. Once the SMPC nodes obtain their shares of the result, they encrypt the result and expose it to the customer. After receiving the encrypted shares of the results, the customer can decrypt them and reconstruct the actual result. Such result, even if computed in a privacy preserving way, could lead to the disclosure of information; a classic example is computing the average of two individuals' salaries which can be done privately by SMPC (without either individual exposing their salary) but, clearly, by looking at the result, each participant knows the other's salary. We will explore solutions to address this problem, for example automatic detection of the risk before computation and anonymization of the dataset or SMPC result.

## 2.4 Data Products

Within the KRAKEN marketplace a user (data seller) will be able to package data streams and batch datasets into what are termed Data Products. Similar to how any item for sale on Amazon, such as an item of clothing, must be productised and advertised for discovery, a data seller must be able to do the same for their data streams and datasets. Data Products are therefore how data sellers using the KRAKEN marketplace will productise and commercialise their data assets to be advertised and made discoverable within the marketplace's data catalogue.

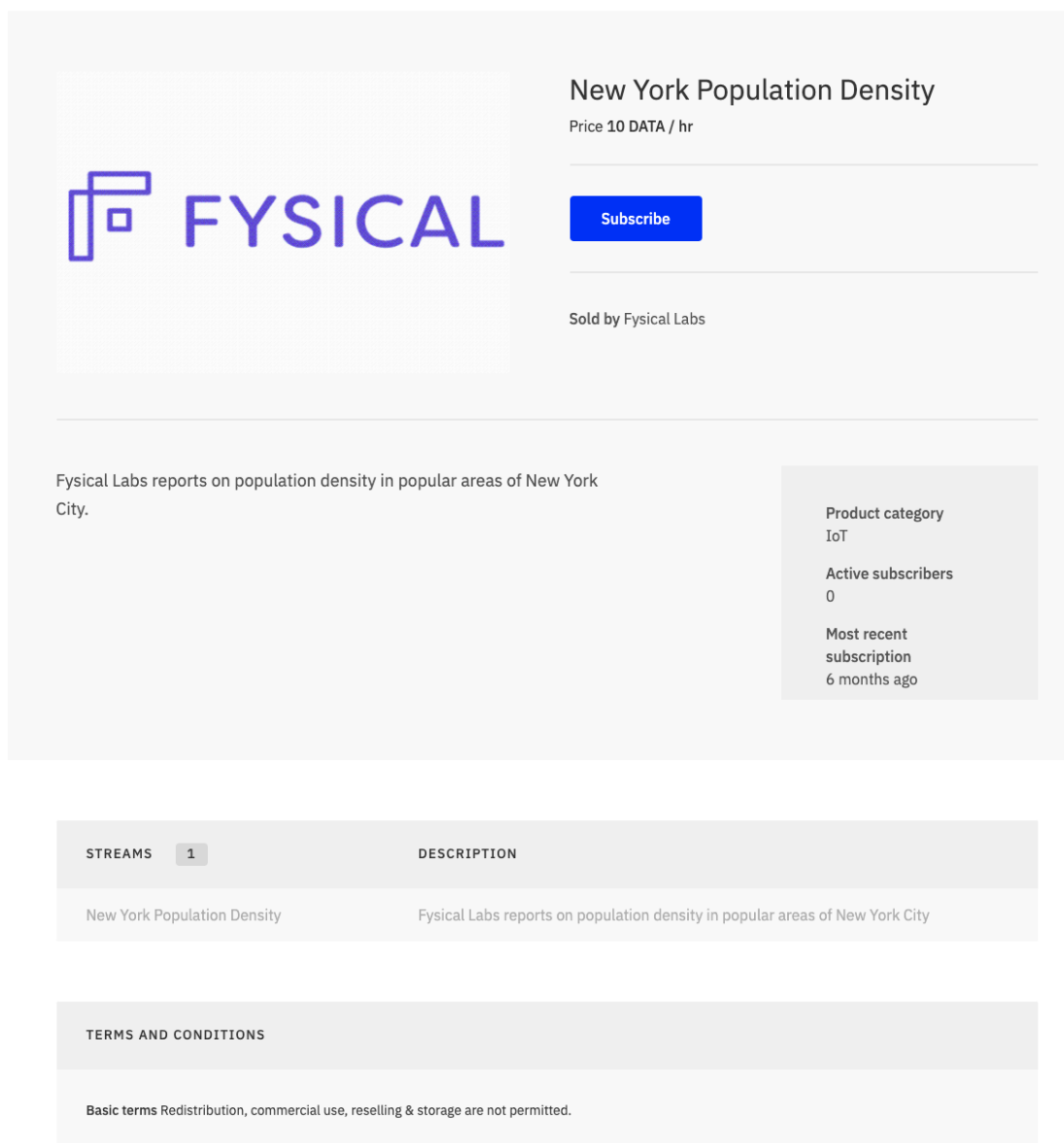


As the KRAKEN marketplace expands on the existing sharing and monetisation functionality for real-time data streams within the Streamr marketplace to enable the sharing and monetisation of batch datasets and the results of privacy-preserving analytics, it will require the KRAKEN user to be able to choose additional types of Data Product.

The types of Data Product currently identified for inclusion in the KRAKEN marketplace are as follows:

- Real-time Data Product (Existing Streamr Data Product)
- Data Unions Product (Existing Streamr Data Product)
- Batch Data Product (New Data Product to be added)
- Data Analytics (by secure Multi-Party Computation) Product (New Data Product to be added)

Examples of how a real-time Data Product and a Data Unions Product appear in the existing Streamr marketplace are included in Figures 10 and 11 below.



**New York Population Density**  
Price 10 DATA / hr

**FYSICAL** [Subscribe](#)

Sold by Fysical Labs

Fysical Labs reports on population density in popular areas of New York City.

Product category  
IoT

Active subscribers  
0

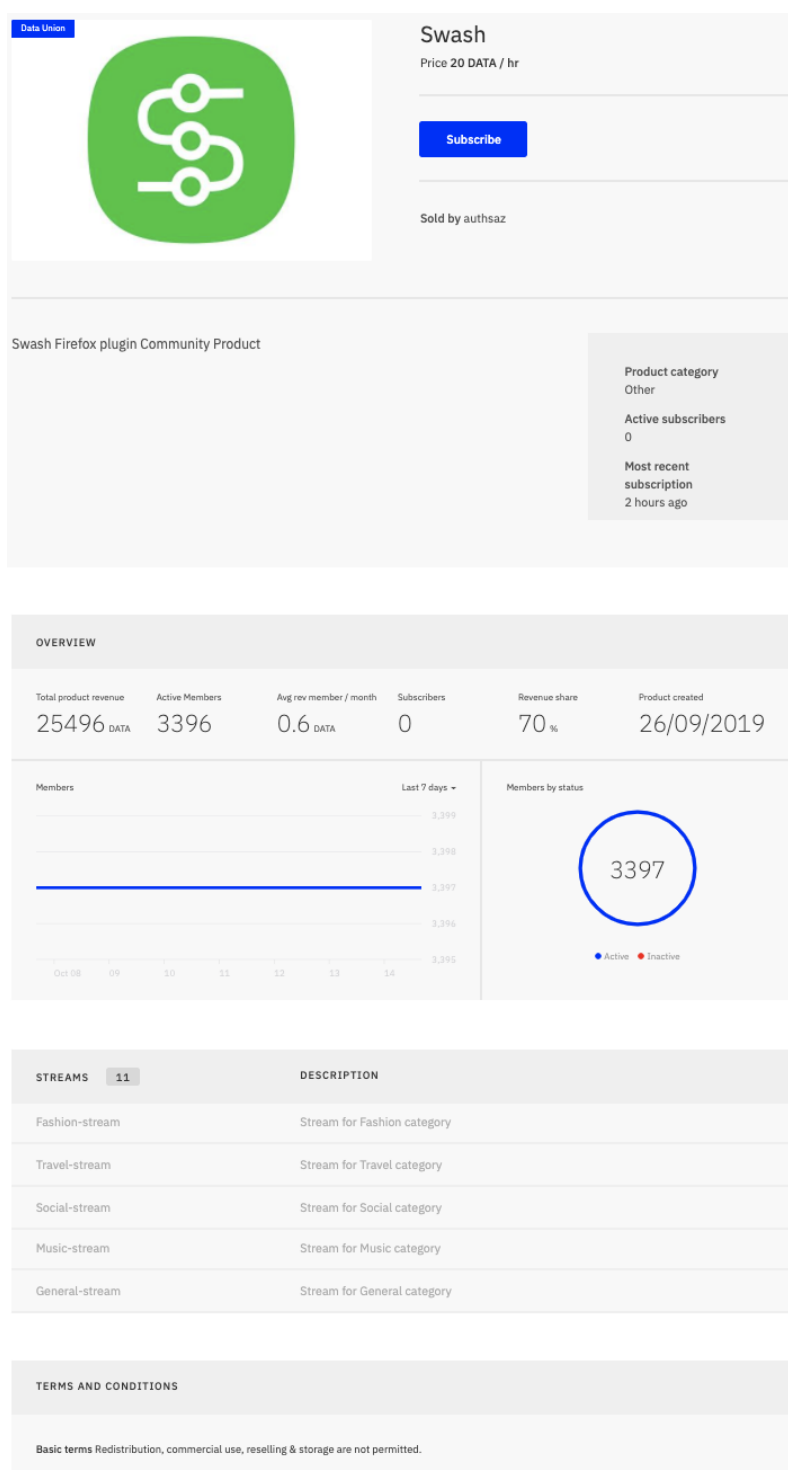
Most recent subscription  
6 months ago

STREAMS	DESCRIPTION
New York Population Density	Fysical Labs reports on population density in popular areas of New York City

**TERMS AND CONDITIONS**

Basic terms Redistribution, commercial use, reselling & storage are not permitted.

**Figure 10: Example of a Data Product in the existing Streamr marketplace**



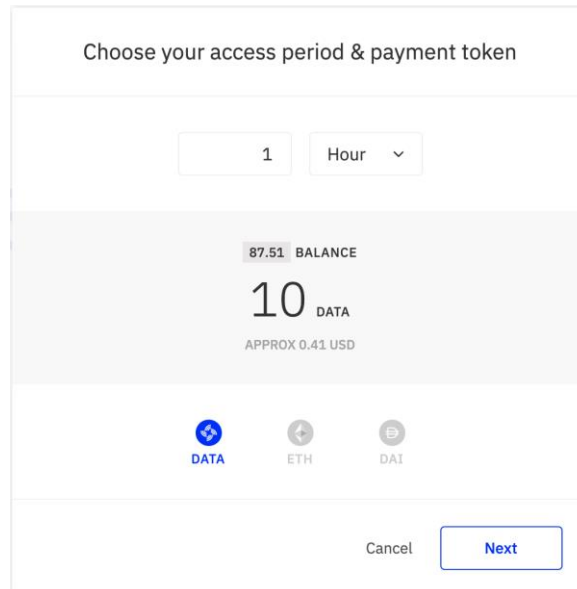
**Figure 11: Example of a Data Unions Product in the existing Streamr marketplace**

The following subsections describe these different types of Data Products in more detail.

### 2.4.1 Real-Time Data Product

A real-time data product originates from a data source that provides access to a real-time data feed. When a data point is added to the real-time data source an event is triggered in the Streamr Network that distributes the data through its publisher-subscriber (pub-sub) model. When a real-time data product is bought within the marketplace the user is granted a time-limited access to the related data stream for a specified period of time.

The seller will be able to set the price of access to the data product per hour, day, week or month and similarly the buyer will be able to specify their subscription period in hours, days, weeks or months when purchasing the Data Product. Figure 12 below shows an example of how a data buyer subscribes to a real-time Data Product in the existing Streamr marketplace.



**Figure 12: Example of a subscription to a real-time data product in the existing Streamr marketplace**

### 2.4.2 Data Unions Product

A Data Unions Product (see [Figure 11](#)) is a concept created and realised by the Streamr Project. It aims to provide a fairer and more ethical framework for the monetization of data streams produced by users of applications and devices. It enables the pooling of mainly real-time data from multiple sources and also provides the means to share the revenue from the data sales to all of the individual data providers from which it was originally sourced.

Data Unions are managed by Data Union Admins, who set up the Data Union by building an integration between a device or application and the Streamr Stack. They then manage and market the Data Union to potential buyers, and in return can set a percentage fee or commission for their efforts.

A data seller, in this case a Data Unions Admin will be able to use the marketplace UI to create a Data Unions Product, set the price of access to the Data Union and set the admin fee they will receive for managing and maintaining the Data Union.

### 2.4.3 Batch Data Product

A Batch Data Product consists of data that is accessed in batches and constitutes a set of data that is bought as such. It does not generate and provide data in real-time or live, and by its nature is a snapshot within a fixed period of time.

A user (data seller) will interact with the marketplace UI to create a Batch Data Product and will be able to set a one-time price for a buyer to gain access to the encrypted link stored within their own personal data store or cloud storage and the key to decrypt it.

### 2.4.4 Data Analytics Product

The KRAKEN platform will enable distributed data analytics by SMPC as an additional service. Users (data sellers) of the marketplace will be able to use the marketplace UI to package their batch datasets in a Data Product that allows interested data buyers to perform analytical queries on the dataset and receive the result of the query without granting direct access to the data and exposing any personal or private data to the data buyer. The integration between the marketplace back-end and the SMPC system is currently being defined.

## 2.5 Data Sources

This Section includes an initial estimation of the potential data sources and categories of data that will be shared and monetised within the marketplace during the pilots. These are presented in two sub-sections based on the two pilot market sectors: health and education.

### 2.5.1 Health Data

Health data can consist of data related to the medical and general health situation of an individual or can consist of an aggregated dataset containing information pertaining to populations. This therefore provides the marketplace with two potential data sources acting as data providers:

1. Individual citizens (or patients); and
2. Healthcare and wellness organisations (e.g. hospitals and clinical centres, pharmaceutical companies, insurance companies, device manufacturers and health and wellness application developers).

Individual citizens may hold their own personal health data on mobile apps or personal cloud data storage, whilst organisations are generating data on populations and store this either on premises within their own servers or using cloud data storage. The KRAKEN data marketplace will therefore need to support the sharing and monetisation of data from both of these potential data sources.

Examples of just some of the potential types of health data, or data relevant to the healthcare sector, that may be shared and monetised within the KRAKEN data marketplace include:

- Personal health / Medical Records
- Medical history
- Laboratory and test results
- Medications / prescriptions
- Procedures (maybe need to explain what this actually means)
- Immunizations
- Allergies
- Radiology images
- Health and wellbeing real world data recorded by mobile apps and wearable devices (consumer and medical grade)
- Heart rate
- Dietary
- Physical activity
- Clinical trials data
- Medication adherence data

Another potential type of data seen as useful for example by population health managers is other demographic, socioeconomic and lifestyle data. These will not be included in the initial stage but their integration into the marketplace will be investigated at a later stage. Examples of this type of data include:

- Income
- Home ownership
- Transportation choices
- Education level

Examples of some possible uses for all of the above identified data categories by data consumers may include:

- Biomedical and academic research
- Clinical trials including patient stratification, and identification of patient cohorts and clinical monitoring
- Clinical development
- AI-based clinical decision making such as early diagnosis, patient management and stratification, and therapy assignment
- Value-based care / value-based contracts
- Medical insurance personalized premiums
- MHealth tech e.g. Machine learning / Artificial Intelligence, medical device development
- Population level health interventions and public health policy by authorities
- Wellbeing services and apps

### 2.5.2 Education Data

Education (or academic) data about students is produced and provided by educational institutions (e.g. Universities and colleges). It consists of data related to an individual's academic career, such as their capabilities, qualifications and performance.

Examples of some of the potential types of education data include:

- Certificates of Graduation
- Certificates of each course
- Enrollment status

Examples of some possible uses for this data may include:

- Human resources management systems
- Recruitment companies to identify ideal candidates
- Professional social networks
- Training, education and life-long learning institutions
- Market consultants and advisory companies for research
- Government, education ministries, European networks (Eurydice, European Association for Quality Assurance in Higher Education)
- Public or private statistical organisations

## 2.6 Consent

In the KRAKEN marketplace, consent will form the key legal basis for the sharing of personal data. Users (both individual citizens and organisations) will therefore be provided with an interface for setting consents to access their data within the marketplace. Consent setting will be performed at the level of Data Products and will be set during the workflow process to create a new Data Product in the marketplace. This is because a user may have different preferences around who can access their data and for what purpose for each single Data Product that they publish in the marketplace.

More dynamic consent will also be built into the marketplace, allowing a user to change his or her consent preferences for a Data Product after its publication.

This consent setting interface will be backed by the MHMD blockchain network, which will orchestrate the secure data sharing process, including the management and authorization of data exchange and access on the basis of the user-defined permission/consent settings.

## 2.7 Payment

Special categories of personal data, particularly certain types of health data have different national implementations of the GDPR and restrictions on their processing. In addition, there are still uncertainties surrounding the EUs position on the legality and ethical implications of monetising personal data, i.e. exchanging personal data in return for payment. These restrictions, in addition to the user requirements, must be closely studied before implementing the final approach to payment.

The following initial approach has therefore been decided for payments:

In the first iteration of the marketplace, for fully anonymised datasets, users will be able to use Streamr's token-based payments (DATACoin) functionality. In order to do so, they will require an Ethereum account and wallet. Whilst there are still uncertainties surrounding the legalities of monetising personal data within the EU, anonymous data falls outside of the category of personal data and outside of the scope of the GDPR and so it will be possible to monetise this data in the KRAKEN marketplace.

In the case of non-anonymised personal data, it could also be possible in this first marketplace iteration for users' payments to be handled using a white listing feature. This means that payment can be performed outside of the marketplace system directly within the control of the buyer and seller. When a buyer wants to purchase a Data Product using this feature, the seller will be notified of the purchase request which contains the relevant information, including the billing details. If the buyer fulfills the seller's requirements for accessing their data then the seller sends a bill to the buyer. The seller then waits for the bill to be paid or waits until he/she sees fit to activate the buyer's access to the Data Product within the marketplace.

With regards to the second release of the platform and later iterations of the KRAKEN marketplace, the project will investigate extending token-based payments to non-anonymised personal data and will also investigate the implementation of credit card payments within the marketplace. But this is largely dependent upon the findings of the legal analysis, clarifications around the EU's position on the legality of monetising personal data, and also user requirements.

## 2.8 Target Countries

The KRAKEN marketplace team have identified an initial list of countries within the EU that fall within the scope of the KRAKEN marketplace. These are the countries from which users will be able to access the marketplace. Major economies such as Germany, UK, France, Italy and Spain have been included in this initial list, but also countries which are known for seeing strong performance and growth in the uptake of eHealth / digital health technologies such as Denmark, Estonia, Netherlands, Sweden and Finland. Non-EU countries to be given access to the KRAKEN marketplace will be considered in the second period of the project, under specific integration/translation of their local legal framework with the GDPR.

1. Germany
2. UK
3. France
4. Italy
5. Spain
6. Netherlands
7. Belgium
8. Portugal
9. Sweden
10. Denmark
11. Estonia
12. Finland
13. Austria

It is important to identify this list not only to prioritise countries for scaling the marketplace, but also because there are various national variances in the implementation of the GDPR which must be taken into account for defining which data are processable in each country, as well as national variations in law which determine whether or not it is possible to monetise (and process) the special category of personal health data. The KRAKEN marketplace will need to ensure that transactions of data on the platform fall within the scope of these national variations, which will impact on its design and technical specifications.

### 3 Technical Requirements

This section details the initial technical requirements for the KRAKEN marketplace. It is divided into seven sub-sections and starts with a description of the marketplace's functional requirements, before detailing the requirements for performance, testing, environments, error logging, monitoring and CI tools.

#### 3.1 Functional Requirements

This Section details an initial set of functional requirements for the KRAKEN marketplace. The identification of functional requirements will be performed iteratively and is fed by the market analysis within T6.1 - Market Analysis, the user requirements identified within T5.1 - User Stories Refining, and also complemented by the legal requirements and implementation guidelines documented within D7.2[3].

The existing list of functional requirements are identified in Table 2 below.

No.	Requirement
MKT-FUN-01	The marketplace subsystem should allow a user (data buyer or data seller) to register on the marketplace by using their SSI wallet to transmit the necessary credentials required during the registration process.
MKT-FUN-02	The login process should allow the user (buyer and seller) to authenticate with the marketplace subsystem using their SSI wallet.
MKT-FUN-03	The marketplace should be usable by both individual citizens and organisations and should distinguish between these two high level categories of user.
MKT-FUN-04	The marketplace should allow users to browse the information about available datasets in a Data Catalogue viewer/browser.
MKT-FUN-05	The Data Catalogue viewer within the marketplace should provide users with the facility to search and filter the data.
MKT-FUN-06	The marketplace should provide the functionality to categorise data based on the guidance and input from the health and education pilot teams.
MKT-FUN-07	The marketplace should allow data owners / providers to create a Data Product to be advertised in the Data Catalogue viewer and provide information such as descriptive content that helps the data buyer understand the information contained within the Data Product.
MKT-FUN-08	Working within the legal restrictions identified by the KRAKEN legal team, the marketplace should allow the sharing and transportation of real-time data streams from data owner / provider to data buyer / consumer.
MKT-FUN-09	The marketplace should allow the sharing of batch data (i.e. health records) which will be provided to the data buyer / consumer by exposing an encrypted link to the location of the stored dataset, which may be stored locally or within a user's cloud storage.
MKT-FUN-10	The marketplace should allow the selling and sharing of datasets that have been anonymised at source.



No.	Requirement
MKT-FUN-11	The marketplace should allow sharing of end-to-end encrypted datasets.
MKT-FUN-12	The marketplace should allow sharing of privacy-preserving analytics results (by Secure Multi Party Computation).
MKT-FUN-13	The marketplace should not be able to access, store, manipulate or otherwise be exposed to identifiable information under any of the scenarios in 10, 11 or 12.
MKT-FUN-14	Data access preferences recorded on the blockchain should be anonymous or protected to ensure they remain private.
MKT-FUN-15	All datasets to be shared or monetized in the marketplace (batches and streams) should remain in local or cloud-based storage outside of the KRAKEN platform until consented and authorized by the data owner / provider.
MKT-FUN-16	The provider should have the facility for setting his access preferences / consents (who can access their data and for what purpose).
MKT-FUN-17	The provider should be able to update their preferences at any time and will have a method for retracting their consent to a particular user of the marketplace, however such retractions will only affect future transactions on the Data Product.
MKT-FUN-18	Individual data owners should be provided with the facility for aggregating their data with other individuals so that it can be packaged into a combined Data Product offering (Data Unions).
MKT-FUN-19	Data owners / providers should be able to track certain statistics in real-time, like who is currently accessing their data, transaction history and profit made from selling their data.
MKT-FUN-20	Data buyers / data consumers should be able to track certain statistics in real-time, like who's data they are currently accessing, transaction history and amount spent.

**Table 2: List of existing marketplace functional requirements**

### 3.2 Performance Requirements

The marketplace performance can be divided into three components:

- Marketplace UI response times
- Payment processing times
- Data transfer speeds

Marketplace UI response times should mirror the response times of the Streamr marketplace. The Kraken marketplace frontend and backend will have similar scalability properties to the Streamr marketplace which has been tested to scale up to hundreds of concurrent users. There are numerous scaling options available starting from upgrading the virtual machines running the frontend and backend components. A potential area of interest here is the calls made to the API component that connects to the blockchain to be provided by Lynkeus. Here the strategy is to cache data as needed to provide reasonable response times.

Payment processing is determined by the final agreed approach that will be taken to payments. For token based transactions, Ethereum will be used as the settlement chain which limits the KRAKEN marketplace to the transaction speeds provided by Ethereum. Transactions are only needed when a Data Product is bought, so this is not something that the user needs to go through often in typical marketplace usage situations.

Data transfer speeds are determined in batch data use cases by the relevant transfer speeds of the seller cloud store and the buyers' internet access. For real-time data streams, the data is transferred through the Streamr Network where latencies are typically in the range of milliseconds. For a better understanding of Streamr Network scalability please refer to the Streamr Network Scalability Whitepaper [\[6\]](#).

The initial performance requirements for the marketplace are detailed in Table 3 below. The performance requirements are still open as they are dependent on the business side developments, but from a technical perspective the marketplace should scale for the foreseeable needs during the pilots without any issue.

No.	Requirement
MKT-PER-01	Marketplace UI response times should mirror the response times of the Streamr marketplace.
MKT-PER-02	Transaction speeds for token payments in the marketplace should mirror those of the settlement chain being used - Ethereum.
MKT-PER-03	Data transfer speeds for real-time data streams should mirror those of the Streamr network, where latencies are typically in the range of milliseconds.

**Table 3: List of initial marketplace performance requirements**

### 3.3 Testing

Work Package 5 will handle the user testing within T5.5 including the Quality Assurance (QA). From this perspective we are planning to use the following:

- Unit testing and integration testing on the marketplace backend side will be added as deemed necessary.
- The CI process will be configured so that unit tests and integration tests are run before each release.
- The code base will be in a number of different languages as follows:
  - Backend with Java with Grails Framework; and
  - Frontend with Javascript and NodeJS Framework.
- Language specific testing frameworks will be used for unit tests and integration tests.

The initial marketplace testing requirements are detailed in Table 4 below.

No.	Requirement
MKT-TES-01	Unit testing and integration testing on the marketplace backend side should be added and performed as deemed necessary.
MKT-TES-02	The CI process should be configured so that unit tests and integration tests are run before each release.
MKT-TES-03	The code base should be in the following languages: 1) Backend with Java with Grails Framework, 2) Frontend with Javascript and NodeJS Framework.
MKT-TES-04	Language specific testing frameworks should be used for unit tests and integration tests.

Table 4: List of initial marketplace testing requirements

### 3.4 Environments

The marketplace will be hosted in an Amazon Web Service (AWS) cloud service based in the Milan region.

There will be two environments configured to the AWS cloud service:

1. Production
2. Test

The production environment will contain the production version of the software and it will respond to the <https://marketplace.krakenh2020.eu> url.

The test environment will be a copy of the production environment, and will answer to the url, <https://test.marketplace.krakenh2020.eu>.

The respective environments' production and test environments will be synchronized with other teams' (Crypto and SSI) respective environments as needed.

The initial requirements for the marketplace environments are detailed in Table 5 below.

No.	Requirement
MKT-ENV-01	The marketplace should be hosted in an AWS cloud service based in the Milan region.
MKT-ENV-02	Two environments should be configured to the AWS cloud service: 1) production, and 2) test.
MKT-ENV-03	The production environment should respond to the URL - <a href="https://marketplace.krakenh2020.eu">https://marketplace.krakenh2020.eu</a>
MKT-ENV-04	The test environment should respond to the URL <a href="https://test.marketplace.krakenh2020.eu">https://test.marketplace.krakenh2020.eu</a>

Table 5: List of initial marketplace environments requirements

### 3.5 Logging

Logging will be configured in a way that enables Amazon Cloud Watch to monitor the logs and trigger alerts when fatal application errors happen.

Logging will have four levels: FATAL, ERROR, INFO, DEBUG

Fatal errors will trigger alerts for maintenance by email. Debug level will be visible only in development environments.

The initial marketplace logging requirements are detailed in Table 6 below.

No.	Requirement
MKT-LOG-01	Logging should be configured so that Amazon Cloud Watch can monitor the logs and trigger alerts when fatal application errors happen.
MKT-LOG-02	Logging should include four levels: FATAL, ERROR, INFO, DEBUG.
MKT-LOG-03	Fatal errors should trigger alerts for maintenance by email.
MKT-LOG-04	Debug level should be only visible in development environments.

**Table 6: List of initial marketplace requirements for logging**

### 3.6 Monitoring

Amazon Cloud Watch will be used to monitor each container in the production environment. Trigger alerts from errors in logs will be added and performance of the instances shall be monitored.

Server health alerts will be configured to help guarantee that the servers are running in an optimal way, and to react to fatal errors.

The initial marketplace monitoring requirements are detailed in Table 7 below.

No.	Requirement
MKT-MON-01	Amazon Cloud Watch should be used to monitor each container in the production environment.
MKT-MON-02	Server health alerts should be configured to guarantee that the servers are running optimally and in order to react to fatal errors.

**Table 7: List of initial marketplace monitoring requirements**

### 3.7 CI Tools / Process

Github will be used as the code repository and Jenkins will be used as the CI tool to configure CI to enable the automated deployment process.

Deployment will be triggered from the Jenkins console to test and production environments. Jenkins will be configured so that the code from the master branch of the git repository will be deployed. Developers will develop in separate development branches that will be merged to the master branch for releases.

The initial marketplace requirements for the CI tools / process are detailed in Table 8 below.

No.	Requirement
MKT-CI-01	Github should be used as the code repository for the marketplace.
MKT-CI-02	Jenkins should be used as the tool to configure the CI to enable the automated deployment process.
MKT-CI-03	Deployment should be triggered from the Jenkins console to both marketplace test and production environments.
MKT-CI-04	Jenkins should be configured so that the code of the master branch of the git depository is deployed to marketplace production and test environments.
MKT-CI-05	Development should be conducted in separate development branches which are merged to the master branch for marketplace releases.

**Table 8: List of initial marketplace requirements for the CI tools / process**

## 4 Marketplace Specifications

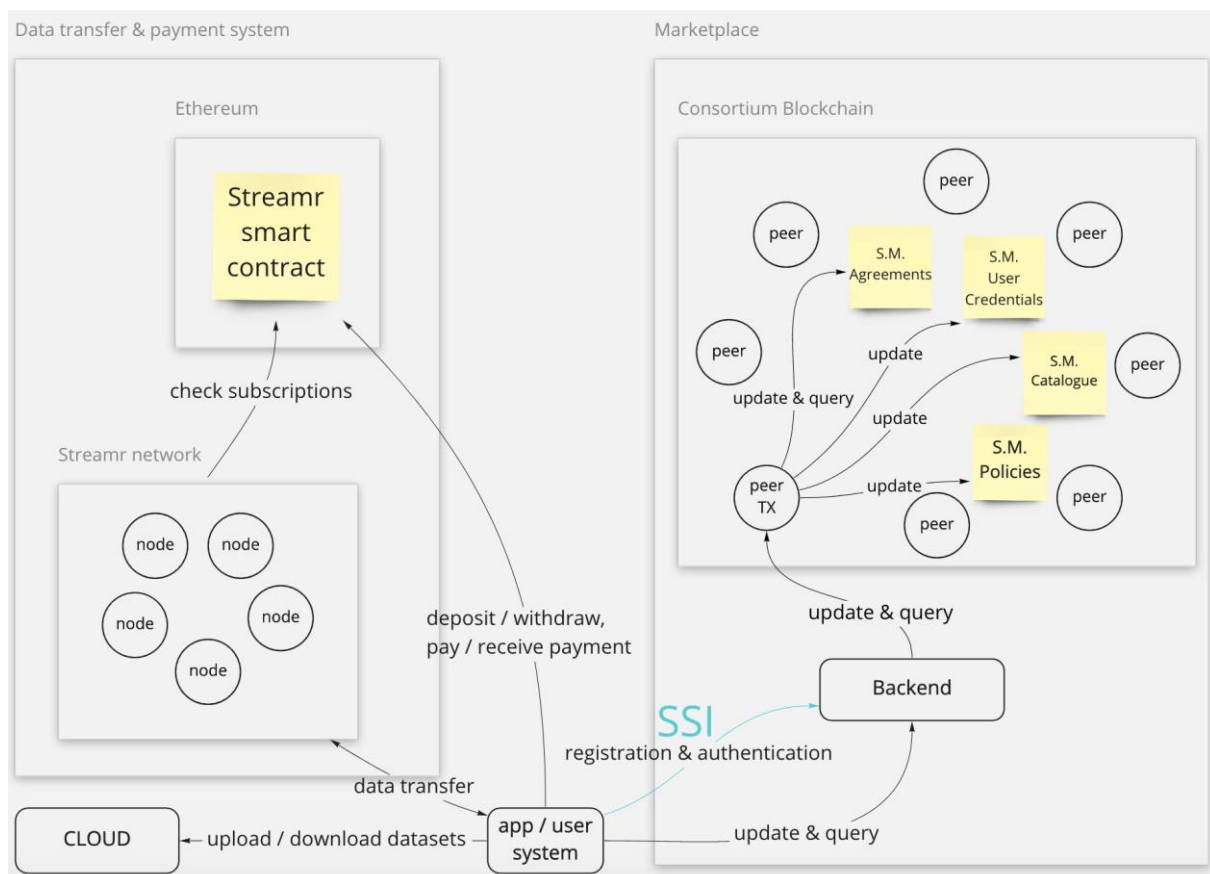
This section provides an initial definition of the system architecture of the KRAKEN marketplace. It outlines the two key subsystems of the marketplace: the marketplace and the data transfer and payment subsystem, and then provides information about their respective components. It concludes by describing the marketplace's interfaces with the two other KRAKEN subsystems: SSI and crypto.

### 4.1 System Architecture

The KRAKEN marketplace subsystem allows the secure control, sharing, transfer and monetisation of data streams and batch datasets in a decentralised and distributed way. It consists of two key subsystems; 1) A decentralized marketplace powered by a blockchain network which controls the business logic for transactions using smart contracts, and 2) a data transfer and payment system which includes a distributed peer-to-peer event messaging network for transporting permissioned data streams and enables a mechanism (StreamrFS) for the exchange of encrypted information about the location of stored batch data from data providers to data consumers.

In the case of data streams, in order to respect a user's privacy and ensure they retain full control, data is never stored on the marketplace system and is only transferred from the data provider to the data consumer once an approved transaction has been executed in the marketplace. In the case of batch data, it is also never stored on the system, and the system will be compatible, in a secure way, with the usage of a users' cloud storage as intermediaries between data holders and data consumers.

The architecture currently proposed within [Figure 13](#) below is divided into two main subsystems: the marketplace and the data transfer & payment system. The following Sections describe the two key subsystems and their respective components in more detail.



**Figure 13: Existing marketplace Subsystem Architecture**

### 4.1.1 Marketplace

The marketplace is the digital market, or shop window, for Data Products. This subsystem is used for the purpose of connecting sellers and buyers. Data Products, along with their textual descriptions, including data sources, data exchange modalities, pricing and permission requirements will be featured on the KRAKEN data marketplace and recorded in the marketplace specific data model. The data model will record a high-level description of data content, source, format, data exchange modality, pricing and access criteria.

#### 4.1.1.1 Consortium Blockchain

The consortium blockchain is the blockchain network that is sustained by the consortium peers and provides the business logic as illustrated in [Figure 13](#). The business and legal logic is implemented using smart contracts. In this specific case the blockchain provides a “catalogue” smart contract, a “policies” smart contract, “user credentials smart contract” and an “agreements” smart contract as described in Sections [4.1.1.3](#), [4.1.1.4](#), [4.1.1.5](#) and [4.1.1.6](#) below.

#### 4.1.1.2 Data Catalogue

The Data Catalogue is a component of the marketplace subsystem that will collect metadata about the Data Products published on the marketplace. We refer to metadata as descriptive information about Data Products such as title, image, description, tags, etc. A Data Product is the unit of sale in the marketplace as described in [Section 2.4](#).

This component will be split between a smart contract on the consortium blockchain and the Backend storage; specifically the Backend storage ([Section 4.1.1.7](#)) will act as storage for the metadata, while the catalogue smart contract ([Section 4.1.1.3](#)) will be responsible for updating the index of all of the Data Products added to the marketplace.

#### 4.1.1.3 Catalogue Smart Contract

The catalogue smart contract is the smart contract with the role of keeping track of all Data Products currently present and being added to the marketplace subsystem, as a kind of index. Whenever a user decides to publish a Data Product, the index is updated and the other components of the marketplace subsystem (such as the policies smart contract ([Section 4.1.1.4](#)) or agreements smart contract ([Section 4.1.1.6](#))) can exploit it to accomplish their functions.

#### 4.1.1.4 Policies Smart Contract

The KRAKEN platform will be at a European level. Non-EU jurisdictions will be considered in light of adoption KPIs. In Europe the current regulation for data protection is the GDPR. Every country must comply with this regulation, however there is a level of autonomy that allows different countries to define further constraints with their own specific laws within their own jurisdictions. Single users, whether that be an individual citizen or an organisation such as a hospital, have a need to set certain data access rights when sharing their data. This information is collectively termed “policies” in the scope of the marketplace subsystem.

The policies smart contract is used to store the policies related to individual Data Products. Every time a new Data Product is published in the catalogue, a new record in the policies smart contract will be linked to the Data Product with the purpose of storing its policies. The policies may be updated by the user whenever needed.

#### 4.1.1.5 User Credentials Smart Contract

In the KRAKEN platform, verifiable credentials are used to check whether a Data Consumer is eligible for buying Data Products. The “User credentials smart contract” is used to store this information that is collected at registration time. The credentials will then be retrieved by the Agreements smart contract to perform the eligibility check.

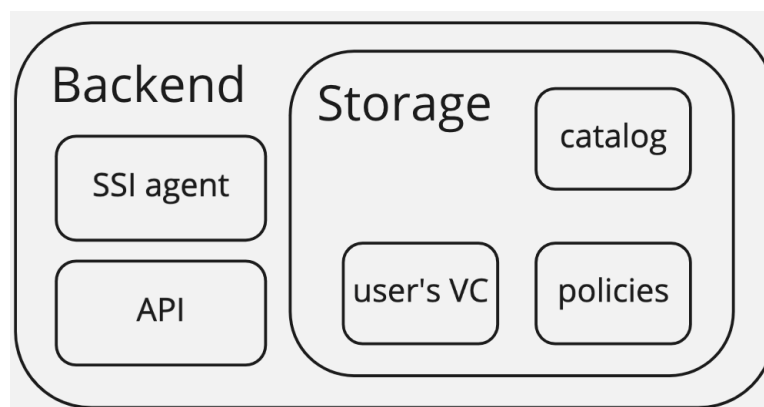
#### 4.1.1.6 Agreements Smart Contract

The “agreements” smart contract is applicable to when a Data Consumer decides to purchase or access a Data Product. The role of this component is to control that the purchase or access request is compliant with the policies set by the seller. To do so, this component will retrieve the policies connected to the Data Product from the Policies smart contract and the credentials belonging to the Data consumer from the User credentials smart contract. If the policies and the credentials are compatible, the purchase is stored in the Agreements smart contract and is defined as an “agreement”.

#### 4.1.1.7 Backend

The backend component will be responsible for providing backend services and is shown in [Figure 14](#) below. Its subcomponents include:

- An SSI agent;
- An API;
- A storage system.



**Figure 14: Sub-components of the backend component**

The SSI Agent will provide registration and login functionalities and the storage component will store the Data Catalogue, the policies and the users’ Verifiable Credentials.

The API will receive the updates from the user system. Updates regarding the users’ policies and Verifiable Credentials will be stored in the backend storage component and forwarded to the consortium blockchain. Updates regarding metadata will be stored in the Data Catalogue in the storage subcomponent. The metadata stored in the Data Catalogue will be used to allow users to browse the published Data Products and also to implement filtering features.

Key exchange for the secure transfer of datasets regarding Batch Data Products will also be performed by the API.

#### 4.1.2 Streamr Data Transfer & Payment Subsystem

This subsystem will be responsible for the transfer of data for the Data Products and will also be responsible for the processing of payments for access to Data Products. The subcomponents of this



system are a smart contract on the Ethereum network, the Streamr Network and the payment manager.

#### **4.1.2.1 Ethereum Smart Contract**

This is the smart contract of the Streamr Network that is responsible for storing subscriptions. A subscription is a record that describes a data transfer that is allowed to happen. This smart contract is updated by the users every time a data transfer needs to occur. Depending on the inputs from the legal team, this smart contract could be substituted by a different one or implemented in a private Ethereum Network. There may be the need to establish an additional smart contract if it is determined that there is a requirement for the consortium blockchain to keep a record of the confirmation of payments on the Data transfer and payment subsystem. This requirement is still to be decided.

#### **4.1.2.2 Streamr Network**

The Streamr Network is the system that performs the data transfer for real-time data streams. It is a decentralized network that can be accessed globally by anyone. This network transfers data on the basis of subscriptions checked against the Ethereum smart contract. A user will be able to publish data (send) or subscribe to a data stream (receive) until the subscription expires. For the batch data transfer, the platform will be interfaced with StreamrFS, a feature of Streamr that, combined with a cloud storage, exploits the Streamr Network to exchange information about the location of batch datasets instead of stream messages. Depending on the inputs from the legal team about routing of personal data inside and outside of the EU, this Network could be substituted with a private instance hosted on nodes within the EU.

#### **4.1.2.3 Payment Manager**

Payments in the initial marketplace release will be executed using DATACoin on the Streamr marketplace smart contract that is currently deployed on the Ethereum network.

Once a new purchase request is commenced, and after the MHMD blockchain validates ethical and legal constraints allowing the transaction, this is first marked as “unpaid”, which prevents data transfer between a data provider and data consumer, unless there is no price. When the buyer pays the set price for subscription, the seller grants the buyer access to the Data Product.

We will investigate ways to facilitate the bridging of fiat payments and tokens after the first implementation of the marketplace release. There are still further user and market needs to be studied within T5.1 and WP6 before the requirements for the full range of payment options is determined.

### **4.1.3 User System and Cloud**

The user system is the software running on the user's machine. In KRAKEN the user can decide to use an app or a web page. On the user system will sit the SSI wallet as well to allow registration and login.

An essential component that will communicate only with the user system is the user's cloud storage, for example the user's Google Drive or Dropbox. This component has the role of storing encrypted datasets and communicates only with users, never with the marketplace or the Data Transfer system.

## 4.2 Interfaces with other KRAKEN Subsystems

The KRAKEN marketplace subsystem will be integrated with the two other major KRAKEN subsystems: the SSI and crypto subsystems. The following sub-sections describe the interfaces between the marketplace and the aforementioned KRAKEN subsystems.

### 4.2.1.1 SSI Subsystem

The registration and login process for the existing Streamr marketplace, which currently uses Ethereum identities or requires a user to input their email address and password to participate within the marketplace ecosystem, will be replaced by the feature provided by the SSI subsystem.

Authentication and registration within the KRAKEN marketplace will be implemented in collaboration with an SSI Agent instance provided by the SSI subsystem, running on the marketplace Backend. A user of the marketplace will be able to authenticate and register by scanning a QR code containing an invitation, presented on the webpage of the marketplace and provided by the SSI agent instance, using a mobile app containing the user's SSI wallet. This will establish a DID connection, that is a secure and private communication channel, between the marketplace SSI agent instance and the SSI mobile app.

Verifiable credentials will be used to provide information about the users to the marketplace.

### 4.2.1.2 Crypto Subsystem

The Crypto Subsystem will be the main component of the KRAKEN platform that enables the privacy preserving analytics results for Data Products that have been made discoverable within the KRAKEN marketplace. In KRAKEN, this subsystem will function within the user system and in the SMPC (Secure Multi Party Computation) network. On the Data Owner system, the web page will perform the SMPC encryption of the dataset. The SMPC network will perform the privacy preserving analytics. Once the computation is complete, the encrypted analytics result will be sent to the Data Consumer where the Crypto subsystem will be responsible for decryption.

The crypto subsystem will be interfaced with the marketplace in order for the marketplace to exploit the privacy-preserving analytics mechanism previously described in [Section 2.3.3](#). Using this mechanism, the Data Holder will be able to send the encrypted dataset to every node of the SMPC network involved in the computation and, in the same way, the nodes will be able to send the encrypted result to the Data Consumer. The API ([Section 4.1.1.7](#)) will be the interface point between the crypto subsystem and marketplace subsystem. It will trigger a notification to the SMPC network nodes of a new computation request received by the marketplace.

## 5 User Interface

The user interface will follow the existing Streamr Marketplace User Interface (UI).

The specific features and functionalities of the KRAKEN marketplace UI will be largely driven by the legal and ethical requirements, business / market requirements and user requirements to be identified during the course of this project. The following Sections therefore provide an initial indication of the functionalities envisioned at this moment in time to be made available to users in the web frontend based on the understanding of the marketplace specific requirements to date. This will be updated throughout the project as further information and feedback is received by the marketplace team.

### 5.1 Marketplace Web Frontend

The KRAKEN marketplace is accessible through the marketplace UI, which visualises the data catalogue and user specific data administration views. The marketplace UI is a web application implemented with a Java backend and React frontend.

The Web User interface will be customized using the Streamr marketplace frontend as the basis for all necessary customizations. The frontend code will be forked and customized to add the required features for the purposes of the use cases defined within the health and education pilots.

The KRAKEN marketplace will be modified to use the Backend component as a data backend for the catalogue information instead of the database used in the existing Streamr marketplace. The Information retrieved will be displayed in the marketplace UI.

#### 5.1.1 Registration and Login

Users of the KRAKEN marketplace (data providers and data consumers) will need a page to register and login to the marketplace in order to join and use the marketplace ecosystem (sell or buy Data Products). For first-time access, a registration page will be required to act as an interface that enables users to perform the registration using their SSI wallet. Once registered, users will be able to use the login page to access the KRAKEN marketplace, using their SSI wallet.

#### 5.1.2 Explore Data Products

Upon registering and logging in to the KRAKEN marketplace, users will be directed to the initial data catalogue webpage, which will allow them to browse and explore the datasets and data streams that have been made available by data providers for free or at a price.

Users will be able to simply scroll through the available Data Products listed within the data catalogue webpage or will be able to use certain functionalities that aim to assist the user in more easily identifying interesting Data Products. These functionalities will include a filtering tool for different categories of data, a sorting functionality (e.g. by price - low to high and high to low), a keyword search functionality, and a policies-based filter to exclude products that a specific Data Consumer does not have the required consent to buy.

Filtering functionalities, such as filtering by category will be customised based on the categorization identified by the health and education pilot teams. Additional functionalities that improve the user experience whilst searching and attempting to identify a Data Product will be considered as the project progresses.

### 5.1.3 Data Product Description Page

When browsing the available Data Products on the initial data catalogue webpage, users will be able to click on a Data Product and navigate to its specific Data Product page. This will provide descriptive content about the Data Product that enables the seller to advertise and market it.

At present the existing Streamr marketplace allows a user to provide the product name, description and price. The full list of descriptive content and information to be included on the Data Product Description page for the KRAKEN marketplace will be based on the identified user requirements by the two pilot teams. But some additional content already being discussed includes the type of data (batch or stream), category (e.g. medical dataset, wellbeing etc.), possible usages for the data (conditions) and an indication as to whether a data buyer meets the criteria to access the Data Product.

### 5.1.4 Create Data Products

A data seller will be able to use the marketplace UI to connect an existing data stream or dataset and publish a Data Product for sale or to share for free. In the current Streamr marketplace, the Create Data Product page allows a user to select between the creation of a regular Data Product for a data stream from a single source or a Data Union Product for data streams from multiple data providers (both are described in [Section 2.4](#)).

In the KRAKEN marketplace the types of Data Products available to the user will need to be extended to include Batch Data Products and Data Analytics Products. A user will also at the very least need to be able to use the marketplace UI to:

- Select the type of Data Product they want to publish
- Add a product description - including name, textual summary of the Data Product and cover image
- Add / connect a data stream or batch dataset
- Add tags to categorise the data stream / batch dataset
- Select the data exchange modality
- Set the allowed jurisdictions
- Set policies (Consents)
- Set price (if applicable)
- Add metadata to the catalogue

### 5.1.5 Manage Data Products

The UI will provide to the user also the possibility to update Data Products. Specifically the user will be able to update the catalog record, policies and price.

Everytime a Data Product is published on the marketplace, it is listed on the Data Catalogue ([Section 4.1.1.2](#)). The record of the Data Catalogue corresponding to the Data Product is not permanent but can be modified whenever the user needs to update the metadata regarding his Data Products.

The same principle applies to policies and price, the UI will allow users to modify already existing policies and price corresponding to their Data Products on the page dedicated to the management of the Data Product.

### 5.1.6 Subscribe / Purchase a Data Product

Data consumers will be able to purchase Data Products from the Data Product Description page. A purchase will be initiated when the user clicks the “buy” or “purchase” button. Once the purchase is triggered, the user will receive a confirmation of the success (or otherwise) of the access control. If the access control is successful, the user will be able to choose the payment option and finally buy the Data Product.

### 5.1.7 Transaction History

The user will be provided with a set of functionalities to improve user experience and transparency around the Data Products they are selling or have bought within the KRAKEN marketplace. A main feature exposed in the user profile will be the transaction history that will show historic purchases and sales performed on the marketplace.

## 5.2 Mobile App Frontend

Design of the KRAKEN mobile application is still in its very early stages as the user registration process is currently being defined. The general assumption for this component is to focus it on the individual data seller (for which correspondent user stories personas have been defined in T5.1 - User Stories Refining, and in the preliminary business analysis in WP6 - Business Plan, Exploitation and Sustainability). This type of user will use the app to register him/herself through the SSI workflow, but also to register his/her data products, to set correspondent policies and consent parameters, and additionally to connect data sources to the platform data management services, especially for streaming data, all using app commands.

Based on preliminary market studies, individual users are not expected to act as buyers, at least not on sizable scales. The buying process will be in fact much more frequently used by institutions and that perspective will be consequently fully implemented in the KRAKEN web-based data catalogue currently under development.

The MHMD App user workflow will be reutilized in the KRAKEN app to allow users (i.e. individual data sellers) to set data access parameters (personal preferences and consent) and integrated with the SSI user authentication workflow for an end to end user experience.

## 6 Conclusion

This deliverable provides a description of the Marketplace Technical Specification based on the progress made during the first year of the project. The KRAKEN marketplace Subsystem acts as an open and decentralised exchange for data. It securely connects providers and consumers of high-quality datasets and data streams in a trusted manner, collecting explicit consent from data providers and leveraging a blockchain network to enforce the logic behind all data transactions. In order to ensure the trusted collection of explicit consent, it leverages the SSI subsystem to provide legal identification of persons and organisations transacting within the marketplace.

The marketplace allows users to package datasets and data streams into Data Products. When a user purchases a Data Product, data streams are delivered by the Streamr Network and batch datasets accessed by buyers directly from the data providers own cloud-based storage. The marketplace allows three different modalities of data sharing, 1) Sharing of anonymised datasets; 2) Secure sharing of data; and 3) Privacy-preserving analytics results (By Secure Multi-Party Computation).

Architecturally, the KRAKEN marketplace is divided into two main subsystems; the marketplace and the data transfer & payment subsystem. At the backend of the marketplace subsystem is the consortium blockchain, which provides the business and legal logic behind all data transactions and is implemented using smart contracts.

There are four main smart contracts, the catalogue smart contract, the policies smart contract, the user credentials smart contract, and the agreements smart contract. The catalogue smart contract keeps track of all existing Data Products and those being added to the marketplace. The policies smart contract stores the policies related to individual Data Products. The user credentials smart contract stores the users' virtual credentials information gathered during registration. The agreements smart contract checks if an access request is compliant with the policies set by the seller and stores all agreements.

The second of the two main subsystems, the data transfer & payment subsystem, is responsible for the transfer of data and the processing of payments for access to Data Products. The subcomponents of this system are the Ethereum Network, the Streamr Network and the payment manager. The Ethereum Network stores information about subscriptions to Data Products and is used by the Streamr Network to determine whether a data transfer can happen. The payment manager allows users to execute payments for data access using DATACoin, exploiting the Streamr marketplace smart contract that is currently deployed on the Ethereum network.

The KRAKEN marketplace UI uses the Streamr marketplace UI as a basis for all necessary customizations in the KRAKEN project. It allows users of the marketplace to register and login using their SSI credentials, explore available Data Products, create new Data Products and set their consents, manage existing Data Products, subscribe to or purchase a Data Product, and view a users' transaction history.

Going forward, refinements will be made to these specifications as further inputs are received from the market analysis activities and the legal and user requirements become better defined. The final design and specification of the marketplace will be presented in D2.7 - Design for Marketplace Reference Implementations, which is programmed for January 2022. Further updates are also expected to be provided for the integrated marketplace architecture which will be communicated in

D5.3 - Initial KRAKEN marketplace Integrated Architecture and D5.4 - Final KRAKEN marketplace Integrated Architecture in June 2021 and December 2021 respectively.

## 7 References

---

- [1] <http://www.myhealthmydata.eu/>
- [2] <https://streamr.network/>
- [3] <https://www.krakenh2020.eu/node/78>
- [4] <https://www.krakenh2020.eu/node/70>
- [5] <https://ethereum.org/en/>
- [6] <https://streamr-public.s3.amazonaws.com/streamr-network-scalability-whitepaper-2020-08-20.pdf>





Atos

Fbk  
FONDAZIONE  
BRUNO KESSLER

AIT  
AUSTRIAN INSTITUTE  
OF TECHNOLOGY



LYNKEUS.  
STRATEGY CONSULTING | BLOCKCHAIN & SMART CONTRACTS | DATA ANALYTICS



TX

KU LEUVEN  
CENTRE FOR IT & IP LAW

CITIP

IAIK  
TU  
Graz

InfoCert  
TINEXTA GROUP

@KrakenH2020



Kraken H2020



[www.krakenh2020.eu](http://www.krakenh2020.eu)



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