KRAKEN

BROKERAGE AND MARKET PLATFORM FOR PERSONAL DATA

D6.9 First Advisory Board

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

D6.9 First Advisory Board

Grant agreement	837854
Work Package Leader	InfoCert
Author(s)	Pasquale Chiaro (InfoCert)
Contributors	Francesca Podagrosi (InfoCert)
Reviewer(s)	Sara Diez (Atos) , Simon Guggi (SIC)
Version	Final
Due Date	30/11/2020
Submission Date	1/12/2020
Dissemination Level	Public

Release History

Version	Date	Description	Released by
v0.1	18/11/2020	Initial version	Pasquale Chiaro
v0.2	26/11/2020	Version reviewed by Atos and TUG Graz	Pasquale Chiaro
v1.0	30/11/2020	Submitted version	Atos

Table of Contents

Executive Summary....8
1 Introduction....9
2 The Role of Project Advisory Board....10
2.1 Members of PAB....10
3 First Project Advisory Board Meeting....12
3.1 Meeting organization, agenda and participants....12
3.2 Question for advisors....13
3.3 Presentations and live feedback....15
3.3.1 Business Approach: market analysis, value proposition and initial business model....15
3.3.2 Technical Approach: SSI, Crypto and Blockchain....20
3.3.3 Ethical and legal aspects....22
4 Recommendations and work packages impacted....25
5 Conclusions....27

List of Tables

Table 1: PAB meeting agenda 13
Table 2: Recommendations from advisors 26

List of Figures

Figure 1 - Ethical and legal aspects of research *23*
Figure 2 - Ethical and legal aspects of development *23*

List of Acronyms

Acronym	Description
LSTS	Science, Technology and Society Studies
EC	European Commission
EU	European Union
GDPR	General Data Protection Regulation
HTML	Hyper Text Markup Language
ESSIF	European Self Sovereign Identity Framework
IoT	Internet of Things
APIs	Application Programming Interface
Mx	Month 1, Month 2…
PAB	Project Advisory Board
ZKPS	Zero-Knowledge Proof Systems
EBSI	European Blockchain Services Infrastructure
MPC	Multi-Party Computation
FE	Functional Encryption
FHE	Fully Homomorphic Encryption
Tx.x	Task 6.1, Task 6.2…
WP	Work Package
WPx	Work Package 1, Work Package 2…
EHR	Electronic Health Records
SSI	Self-Sovereign Identity

Executive Summary

As an integrated part of the project, the Consortium has involved several external parties to share with them the approach, methodology and results of the entire project. The “Project advisory board”, is composed of recognized experts in a variety of fields including blockchain, self-sovereign identity (SSI), security and data privacy, law, ethics and business.

Within the project we planned to organize three meetings to show our work and collect feedback to better address the various next steps in the different work packages. On November 13th, 2020 we held the first Project Advisory Board meeting, online, during which the advisors had the chance to get in touch with all work package leaders and contributors. This was also the first time the project Ethics Board met. As agreed at the beginning of the project and explained in D8.4 GEN - Requirement No.6, delivered in February 2020, both project Boards meet together. During the meeting we first presented our business approach, in terms of market analysis and business model; then we continued with a technical presentation showing our approach in SSI, Crypto and Marketplace development, with a focus on the architecture designed. Finally,the discussion focused on the ethical and legal aspects related to the processing of personal data in the project.

The overall feedback of the advisors has been very positive. They agreed on Consortium’ choices, both on business and technical aspects, suggesting the Consortium to address the challenge **of “who really owns the data”** and **“what data belongs to who”** in future project choices. This is an important discussion topic that can generate relevant inputs on data privacy for the implementation of both pilots. Who is the Data Owner for education data, the university or the students? And in the healthcare sector, is the hospital or the patient? In both cases we must put in place an effective management system of end users consents to allow the exchange of personal data.

Regarding business models in the healthcare domain, advisors encouraged the Consortium to pay attention to the value of telemedicine and healthcare home support for patients, confirming that is an important area to investigate. Moreover, in both use cases we should be able to better clarify the type of incentives we want to offer to Data Owners, especially if they are “private” users.

Finally, on the technical side, advisors suggested to align with ESSIF, European Self Sovereign Identity Framework, which is part of EBSI - European blockchain service infrastructure- and aims at facilitating cross-border interactions with SSI. The consortium has been strongly committed to establish collaboration with ESSIF since the beginning of the project.

1 Introduction

The present document has the purpose to report the Project Advisory Board feedback regarding KRAKEN's research and results so far. This is the first of three deliverables (D6.9, D6.10 and D6.11) due respectively in November 2020, November 2021 and November 2022 that will report the outcomes of the three Advisory Board meetings foreseen in the description of action and distributed along the project life.

The document is structured in several sections devoted to introducing the advisors and their expertise profile, the organization and agenda of the meeting and the discussions.

In chapter 4.3 the online meeting and all the live feedbacks collected in each round table after presentations are described.

Finally, the summary of the PAB recommendations is presented in a table with a specific reference to the Work Packages impacted.

2 The Role of Project Advisory Board

As mentioned within grant agreement, the Project Advisory Board is an external body to the project, devoted to transmitting to the Consortium feedbacks and suggestions. To achieve this goal, three official meetings with the Advisory Board have been scheduled as part of T6.5 – Project Advisory Board activities. The reports of these meetings will be submitted as deliverables including the PAB recommendations considered as a relevant input for the achievement of project goals and the technical management of the project. Of course, the collaboration with the Advisory Board will not be limited to these events, at the end of each year of the project, but involves a constant communication with KRAKEN WP leaders.

2.1 Members of PAB

Andrea Migliavacca (male), degree in Business Administration (1988), 26 - year experience in ICT projects. Since 2009 Senior consultant at Lombardia Informatica (Research, Innovation and Financed Projects Area). Andrea was team leader in LISPA for Palante and Salus Projects and he currently is the CEO of Think4Future.

Carlos Pastor (male), currently working for Bolsas y Mercados Españoles (the Spanish Stock Exchange) in the Open Innovation Area and collaborating with Alastria as Digital Identity Commission Leader. More than 25 years working experience in national and multinational companies like Telefónica, or Sun Microsystems linked to then emerging technologies like Intelligent Buildings, Electronic Banking, e-Commerce, Internet Gaming, Social Networks, Voice over IP, SWIFT Communication, Federated Identity, Public Key Infrastructure, Electronic Signatures (advanced including biometric voice & voice recognition signature), Self-Sovereign Identity and Blockchain.

Melek Önen (female) is an assistant professor in the Digital Security Department at EURECOM. Her current research interests are the design of security and privacy protocols for cloud computing, Big Data and IoT. She was involved in many European and national French research projects. Melek Önen holds a PhD in Computer Science from ENST (2005).

J. Peter Burgess (male) is a philosopher and political scientist. He is Professor and Director of the Chair in Geopolitics of Risk at the Ecole Normale Supérieure, Paris; Professor at the Centre for Advanced Security Theory (CAST) at the University of Copenhagen; and Research Professor at the Centre for Law, Science, Technology and Society Studies (LSTS) of the Vrije Universiteit Brussel. His research and writing have focused mainly on the theory and ethics of security and insecurity, and more recently on questions of fundamental rights in relation to digitization, data protection and privacy. He is at present Chairman of the Ethics Advisory Group of the European Data Protection Supervisor and co-authored its recent report Toward a Digital Ethics.

Mr Harald Zwingelberg (male) is head of the "Privacy Technology Projects" division at Unabhängiges Landeszentrum für Datenschutz (ULD), the office of the Data Protection Authority of Schleswig-Holstein. On behalf of ULD he participated in a series of EU-funded and national research projects with relation to data protection, privacy and identity management. His focus resides with legal aspects of data protection.

Dr. André Kudra (male) has more than 13 years of information security consulting experience. In his career he held various key positions in major information security projects of global enterprise organizations. He studied business administration at the European Business School (EBS) in Oestrich-Winkel, Germany, and computer science at the James Madison University (JMU) in Harrisonburg, Virginia, USA. Since 2013 André is CIO of esatus AG, a consulting company specialized in information security matters, with its headquarter near Frankfurt in the Rhine-Main area and offices in Hamburg and Munich. André is a strong advocate of Self-Sovereign Identity and a Sovrin Technical Governance Board member.

3 First Project Advisory Board Meeting

3.1 Meeting organization, agenda and participants

Due to Covid-19 constraint, the first Project Advisory Board has been organized as an online meeting. The meeting took place on Friday, November 13th, since 9:00 am to 01:40 pm and the platform used was Teams, by Microsoft. The meeting was recorded with the consent of all participant.

The agenda had the aim to show to the advisors the project and Consortium activities as a "Company Presentation" addressing the business point of view first, and then going in depth with all technical, legal and ethical aspects.

The agenda covered all the Work Packages of the project and to make the meeting as interactive as possible all the sessions were followed by a feedback discussion with the advisors.

13 th November 2020			
Time	**Description**	**Responsible**	**Duration**
9:00-9:05	**Conference opening**		5'
9:05–9:15	Welcome, presentation of the agenda and meeting objectives.	**INFOCERT**	10'
09:15-09:30	**Project Overview** • Project Overview (10') • 6 objectives (10') • Partners and organization (5')	**ATOS**	15'
09:30-09.50	**Market Analysis** • State of the Art about blockchain and SSI market trends • Market overview and benchmark on HealthCare and Education	**INFOCERT**	20'
09.50-10.05	Round table for Feedback Session on market Analysis	**INFOCERT**	15'
10.05-10.35	**Value proposition and Business Model**	**TEX - LYNKEUS** **TUG GRAZ**	30'
10.35- 10.50	Round table for Feedback Session on Value proposition	**TEX**	15'
10.50-11.00	**Break**		
11.00 -11.45	**Technical Approach** • Overview of platform ○ SSI Components	**INFOCER**	15'
	○ Crypto aspects	**TUG GRAZ**	15'
	○ Blockchain	**TEX**	15'
11.45 – 12.10	Round table for Feedback Session	**INFOCERT**	25'
12.10-12:50	**Ethical and legal aspect**	**KUL**	40'
12.50-13.10	Round table for Feedback Session	**KUL**	20'

13 th November 2020			
Time	**Description**	**Responsible**	**Duration**
13.10-13.20	Pilot focus (WP5): Healthcare: Initial marketplace use cases	**LYNKEUS**	10’
13.20-13.30	Pilot focus (WP5): Education: Initial marketplace use cases	**TUG GRAZ**	10’
13.30	Final feedback from advisors	**INFOCERT**	10’
13.40	**End of the meeting**		

Table 1: PAB meeting agenda

The six project external advisors, the project management board and representatives from most project partners attended the meeting.

3.2 Question for advisors

To better let the advisors, know about our expectations, the consortium prepared a document with several questions regarding each session. These questions were shared with the advisors before the meeting. The intention was to contribute to the success and the effectiveness of the meeting because all advisors were aware of the specific discussion topics before the meeting. The following list shows all the question collected by the WP Leaders. Some of them, have been used to create a final questionnaire, shared via google forms, with the aim to collect high-level written feedbacks from the advisors.

Market Analysis

- What is your overall impression about the market analysis we conducted?
- Do you think there is some important segment/use case we have to consider?
- Healthcare: How do you consider our analysis about use cases and business model?
- Education: How do you consider our analysis about use cases and business model?

Business Model

- What do you think about the business model created?
- Is the value proposition solid enough in your opinion?
- Are the characteristics of our offering coherent with SSI and Blockchain values?
- Do you think HR departments are willing to pay for access to (anonymized) student data, or computations/statistics on student data?
- How do you feel about the use of crypto tokens as a means for payment?
- Would having the ability to be able to pay for data access with credit card be interesting? And would this be a viable way to pay for access to data within these two markets?
- Do you feel our offering on Data Unions (where value is shared between organisations and individuals) would be interesting to companies or organisations holding data on individuals?
- Would stakeholders be willing to pay a fee to gain access to a marketplace where they can both sell and discover datasets available for purchase?

- Which of the two KRAKEN revenue models sound most appealing? (Subscription vs. share of data sales / transaction fees)

Technical Approach Session

- Does buying (or requesting access to) data need to be real time or is it acceptable to have a delay (e.g., of a few hours/days.) to wait for active consent from the Data Owner?
- Where would you see room for improvements in terms of privacy aspects in SSI, like identity assertions, etc.?
- We are enhancing SSI with:
    - Multi-wallet synchronization
    - integration with European Regulation for digital identities (eIDAS)
    - authentication and registration processes for the marketplace

  Do you see other relevant possible enhancement?
- How can we claim that we are compliance with EBSI/ESSIF?
- Which resources can we use from the EBSI/ESSIF community?
- What will EBSI/ESSIF expect from us?
- Does EBSI/EBBSI foresee to incorporate new ledger/s to the infrastructure (such as H Indy)?
- We chose Multi-Party Computation (MPC) and Functional Encryption (FE), accompanied with Group Signatures and Zero-Knowledge Proof Systems (ZKPS), to perform the data analytics in a privacy- and authenticity-preserving way. We did not choose Fully Homomorphic Encryption (FHE) because of its lack of practical performance for general computations. Besides the mentioned techniques, do you know other approaches which would achieve that goal?
- We chose group signatures for having authenticity of the data while still ensuring the privacy of the users; except an authority which can inspect in doubt OR we thought of throwing the master secret key away. We didn't choose Ring Signatures as they do not offer updates. Do you also know other techniques to ensure authenticity of the data while still having the guarantee of privacy?
- To ensure the integrity during MPC, we chose ZKPS. Do you have any recommendations about this approach?
- In terms of MPC we rely on the fact that, e.g., at least 1 node is honest. If, e.g., a hospital is one of the nodes, this is easy to argue, as they normally have a strong interest in keeping their data private. However, for other users, which cannot act as a node (yet), nodes are run on some servers, for instance 3 cross-country nodes of KRAKEN partners. Now, how would you evaluate the trust of such an MPC node?
- For FE, the case of a single user providing a ciphertext that can evaluate a function is straightforward. However, in KRAKEN we also want to evaluate functions on data of multiple users. For this multi-user scenario, we currently rely on a trusted Key Generator, which can get the function output of single users and only then computes the aggregated analysis result. How would you improve this aggregated multi-user data analysis with FE?
    - What is the current state of performance FHE; did it significantly change this year? Is it comparable with e.g. Application Programming Interface or FE in some scenarios?
- With respect to advanced sharing of data we evaluate now the following techniques: Puncturable Encryption (PE), Attribute-base encryption, (Hierarchical) Identity-based encryption, Proxy Re-Encryption, and Functional Encryption. Currently we consider PE, as the tagging of ciphertexts fits our use case well. Besides the mentioned techniques, do you know other approaches which would achieve that goal?

- Considering the needs of your customers, which (parts of the) [meta]data must be protected, which may remain unprotected towards backend/storage/…?
- Payments on Ethereum blockchain are public, but transactions recorded on the blockchain are pseudonymous. What should be the required / minimum level of confidentiality for payments?
- What should be the required / minimum level of privacy with regards to the identities of sellers and buyers in the marketplace? For example, should a data seller be able to know the identity of a buyer and vice versa?

Pilot focus (WP5): Healthcare

- Which of the value propositions we presented resonated with you the most and why?
- As a seller what amount of revenues would make the sale of your data on the marketplace cost-beneficial, attractive or very attractive, and what domain/sector (ex. pharma, wellbeing industry etc.) would you see as the readiest to purchase your data?
- As a buyer what price would your pay for data relevant to your business and what (very approximate) ROI would you see from that investments?
- Do you, or did you in the past, have any active discussions with market players around the commercialization of your data, or, as a buyer, the purchase of data. With whom? what was the outcome of that discussion and why did it/did not come to fruition?
- As a potential seller, are you collecting data from your business with specific purposes of is the data collection a by-product of your operations?

Pilot focus (WP5): Education:

- Do you think it is important that qualification/education credentials can be authenticated, to make sure they are not forged and not issued by a fake issuer (e.g. diploma mill)?
- Ideas for other use cases?

3.3 Presentations and live feedback

3.3.1 Business Approach: market analysis, value proposition and initial business model

3.3.1.1 Market analysis [1]

<u>Why digital Identity is the new money</u>

A digital identity reduces the level of bureaucracy and increases the speed of processes within organizations by allowing for a greater interoperability between departments and other institutions. But if this digital identity is stored on a centralized server, it becomes a honeypot for hackers. Digital Identity solved the need of a portable and verified identity within a standard format.

There are 3 models of digital Identity:

- Siloed Identity: a digital credential issued by an organization to an individual
- Federated Identity: the digital identity is issued by a Digital Identity Provider to an individual and can be used to access to several services
- Self-Sovereign Identity: Self-sovereign identity is a two-party relationship model, between the organization and individual now considered as “peer”

[1] Presentation about market analysis showed the results of D6.2 “Initial market analysis”

The SSI addresses the problem of individual control, security, and full portability of Digital Identity. The individual can manage his/her identity by giving express consent to the Data Buyer. Privacy and security are guaranteed by design, SSI is based on Blockchain technology with a decentralized approach.

Digital Identity on blockchain bring benefits in several industries, key strategic values are:

- Reducing transactions and their costs (intermediaries or administrative effort) with private and permissioned blockchain architecture.
- Generating new revenue streams and new business model

Healthcare market analysis

Healthcare data ecosystem is facing three important issues:

- Data are widespread and stored in silos
- Lack of safety and security of data
- Lack of Data Ownership and control

To address these problems, it is necessary to create an ecosystem where different actors can exchange data in a frictionless way, breaking down data silos, enhance the security and privacy management and give a clear ownership of data.

We analyzed 11 companies from all over the world with different level of maturity and different type of approach. Most have a B2B and B2C approach.

Analyzing the value propositions of several companies, main topic to leverage on are:

- Empower Data Owners as data controllers enhancing control over their health data
- Provide tools for sharing the data in an easily and compliant way
- Monetize data enabling sharing of revenues from data sales with original producers to encourage wider sharing of personal data for the benefit of research, innovation and society as a whole

There are 2 main monetization approaches:

- "Pay per use": Service providers reward Data Owner once he gives consent to share data and Marketplace generally takes a transaction fee
- "Subscription fee": Service providers pays a fee to marketplace for their services

Generally, B2B monetization approach is based on a platform fee, cut of sales and subscription for added value services (such as analytics)

Educational Market Analysis

Educational data ecosystem is facing three important issues:

- Lack of data resources for ensuring customized learning experiences and lack of effective track
- Lack of knowledge exchange
- Paper-based certifications

To address these problems, it is necessary to create a safe infrastructure to exchange data in a decentralized way, make student performance track easier and provide validation of quality of courses.

We analyzed 7 companies from all over the world with different level of maturity and different type of approach. Most have a B2B and B2C approach.

Analyzing the value propositions of several companies, main topic to leverage on are:

- Empower Data Owners by provision and enforcement of learners or organizations enhanced control over their educational data

- Provide a platform to share aggregated and certified data with hiring companies/organizations
- Reduce costs giving free access to services to promote personal portfolio and instant access to global pool of talent at low cost

There are 2 main monetization approaches:

- "Sponsorships": HR companies, organizations, service providers buy tokens and provide sponsorships to get data from students and marketplace. The Marketplace spends tokens to provide incentives to students and content providers
- "Transaction fee": Service providers pays a fee to marketplace for specific services

Feedbacks from advisors:

André Kudra: Data Marketplaces have a clear potential, especially for Healthcare sector that is suffering of data quality and management. This topic it is also fundamental for pharmaceutical companies since they are data driven but they are heavily regulated. Access in a simply and compliant way to different types of data is very valuable for both industries. Educational is a good market to address too guaranteeing transparency and privacy.

A key topic for SSI is also to consider soft skill such as ethical consideration of individual. SSI is full of "Technocrats", the risk is to focus only on technologic aspects considering the individual as a source of data. The marketplace should be developed incorporating this thinking. SSI enables the exchange of data in a secure way.

The business model is interesting and fits to a data-driven economy. However, it requires the actors in these business models to be aware of it and be willing to participate in it. The understanding of customers in such a model has to mature first.

Harald Zwingelberg: SSI is a strong asset to address the need of self-control of identity. The terms Data Owner and Data Buyer can have different interpretations, a clarification of the meaning is needed.

I think bringing the data subjects in the position to decide about the (secondary) use of their data is a step in the right direction. One thing to consider is, where the data is stored and if this system will create a single point of failure with far worse damage for data subjects as individual data silos with the controllers. Even if the data is stored with the data subject or encrypted to a private key of the data subject, it may still be a high risk.

Business model: I am not sure, if data subjects are willing to share data for money in return but I know that some or for the feel-good-factor e.g. in large-scale medical studies (look for the German "Nationale Kohorte").

J. Peter Burgess: There're still a lot of problems not solved regarding blockchain. The challenge is to mix the identities, who owns data, which type of data belongs to who. A marketplace should specify which kind of data individuals are going to share with who, and who is going to manage them. Individual should be enabled to share their data preserving privacy and having a transparent view of who and how the data are managed. The digital identity should be inseparable by the moral part of identity. Moral issues should be considered in developing the marketplace

3.3.1.2 Business Model Canvas [2]

Why we are delivering this project to address three key issues:

- Data Owners face a trade-off between the benefits of sharing personal data and the privacy risks of exposing it.
- Missing an open ecosystem where organisations can securely share, trade and gain access to personal data whilst complying with the GDPR and national legislations.

[2] Presentation regarding business model showed the first results of D6.4 "Initial Exploitation Plan"

- Individual citizens are ignored in today's data economy - No direct way for them to control access to their data and no incentives to share it.

Educational

In the educational context it is important to exchange these types of data: academic data of students, such as graduation certificates, certificates for courses, and the enrolment status for individual terms.

The User Groups involved in the data exchange are:

- Student: use student data for job applications and share selected data without compromising other data
- University: exchange student data while ensuring the students' privacy.
- Recruiter: make analysis of student data sets, e.g., gain insight into the academic performance of an individual student by comparing it with analytic of the data set
- Statistical Agency: Combine data sets from multiple sources and compute statistics on that data while preserving student's privacy

Perform analytics is fundamental for the educational user groups.

The value propositions to address the needs of data seller are: providing a marketplace to easily find Data Buyers, exposing student profiles' and certified information in secure and privacy way. The user is enabled to express consent to manage data and the interoperability should be guaranteed

While the value propositions to address the needs of Data Buyer are save time and improve efficiency finding standardized and certified information for an easy comparison, analysing personal data sets in a privacy-friendly way and verifying consent easily. Privacy and integrity by design is a common need.

Feedbacks from advisors:

Andrè Kudra: digital agents scout for opportunities for individuals acting on behalf of users, are you considering them?

J. Peter Burgess: there are two levels of data, one the data exchange on marketplace, the other, the management of data after the marketplace, who is going to manage them? How can the Data Owner control them? The marketplace should cover also this level, data continues beyond the transaction

Andrè Kudra: the marketplace should enable users to give consent on specific topic, not a large and generic consent.

Andrea Migliavacca: it is important to inform the Data Owner of the consent in a clear a detailed way in order to enable them to understand the content of the consent and know which the consequences of the consent are. This is a big issue when we decide to exchange data with individual, in which way are you working on this?

Davide Zaccagnini: developing a dynamic consent where we specify who will access data, for which aim, and these parameters can be changed. This solution is complaint with GDPR. The platform is native designed to enforce complaint with regulation and privacy.

Healthcare

The healthcare data ecosystem is full of barriers in exchanging data between hospitals, insurance and pharma companies, government institutions, data app and individuals. In this context, individuals are considered a source of data, not a peer counterpart.

All these parties would like to destroy these barriers in a secure a GDPR compliant way.

The aim of this section is to interactively approach to receive a quick feedback on value proposition proposal answering if the value is perceived as "not interested" "maybe" or "very interested":

1. *Find and access new data sources safely, easily, in compliance with the GDPR and national legislations*
    a. *Minimize legal liabilities through strict and automatic enforcement of legal and ethical constraints for every transaction*
    b. *Identify data assets on the KRAKEN catalogue*

Andrea Migliavacca: maybe or very interested

Andrè Kudra: Very interested

Harald Zwingelberg: Not interested in buying data but interested to see which part of the agreement is transferred, interesting tool of audit

J. Peter Burgess: maybe or very interesting, cooperation is interesting

2. *Connect and collaborate with Data Owners*
    a. *Post on KRAKEN your "looking for" data announcements*
    b. *Connect privately and securely with Data Owners to collaborate on enhancing, creating data products*
    c. *Incentivize end-users to engage with intermediaries and enrich data products*

Andrea Migliavacca: very interesting. The intermediary should understand well the role. The intermediary (eg: Data Union) provides an added value of data to connect individual to others, a monetization and rewarding approach should be clear.

J. Peter Burgess: a data aggregator definition could be helpful

3. *Monetize data safely, easily, in compliance with the GDPR and national legislations*
    a. *Minimize legal liabilities through strict and automatic enforcement of legal and ethical constraints for every transaction*
    b. *Post your data product in minutes*
    c. *Data remain behind your firewall and are transacted under the strongest security and privacy-preserving measures*

Andrea Migliavacca: nowadays the attention is focus on telemedicine and support healthcare to digitalize process. Healthcare data exchange market could represent a 1/3 of our business, with focus on specific pathological area.

Monetization

KRAKEN project will be released in two phases:

- First Release: Users will be able to sell and buy access to anonymized datasets and make payments using Streamr's cryptographic currency DATACoin (also possible using ETH, DAI). Still uncertainties around enabling payments for encrypted / pseudonymous data - regulations are still evolving.
- Second Marketplace Release: Possibility to implement credit card payments for fiat-based transactions. Users will be able to sell and buy access to anonymized datasets and make payments using Streamr's cryptographic currency DATACoin (also possible using ETH, DAI). Still uncertainties around enabling payments for encrypted / pseudonymous data - regulations are still evolving.

Within the Marketplace it is possible to directly share the value of data sales with your customers. Data Unions allow received crypto payments to be automatically divided and distributed between a company and its customers. The company or organisation acts as the Data Union manager or administrator and receives a percentage of the overall data sales. Each of the individual customers receive a share of the remaining crypto payment.

Initially, KRAKEN aims to attract more buyers and data sellers, without applying a revenues model. Once KRAKEN is attractive on the market, a revenues model could be activated. WP6 analysed two different types approach:

- Subscription fee: Users of the marketplace pay subscription fees to join the marketplace. It is possible to introduce "tiered" subscription packages (free with limited transactions, paid with unlimited transactions)
- Transaction fee: The marketplace takes a percentage fee of every transaction between buyers and sellers.

Feedbacks from advisors:

Andrè Kudra: The Data Owner should be incentivized, for example, the clinical trials are generally rewarded. The marketplace should also consider that the way data are exposed can influence the habits of users. The consent management should be defined, mainly there could be two different approach: one with a large and generic consent (such as Facebook) and the other with a "paranoia" policy with specific consent approach based also on algorithms to control every transaction. To apply a specific value of a data, a curve of analysis and dynamic prices.

J. Peter Burgess: Data Buyers would prefer a wide consent management like Facebook. If we add a cost for data aggregators, we should expect they would app charging the user upstream.

Harald Zwingelberg: ethical and legal aspect of data management, freedom of choice is fundamental

3.3.2 Technical Approach: SSI, Crypto and Blockchain

SSI components

Within this presentation we described an initial view of the KRAKEN contest in which SSI components are inserted, introducing the main concepts of Verifiable Credentials (like Issuer and Verifier) and Data (like Data Producer and Data Consumer); Moreover, we described the exchanging of Verifiable Credential and Data. After these initial concepts, we depicted a more detailed description of flows and connection between Issuer components, SSI wallet and Verifier components. Then followed a description of the backup feature of encrypted SSI wallet. The presentation, hence, proceeded with a detailed description of how KRAKEN will implement the identity derivation from national identity systems acting as a Service Provider connected to the eIDAS nodes. Another concept descripted is how KRAKEN will implement the one-shot qualified digital signature based on a Verifiable Credential - Q-cert provisioning will leverage on a q-signed identity credential (eIDAS Bridge / DSS verification). And the last topic has been about how SSI will implement the registration and authentication layer of the marketplace.

Blockchain

The main aspects of this presentation have been focused on how blockchain will enable personal data sharing and trading of the marketplace. The marketplace ecosystem leverage on two main blockchain components: MHMD blockchain (Lynkeus), which is based on Hyperledger Fabric and an Ethereum Blockchain (TEX). The different roles of these blockchains have been described as MHMD Blockchain will act as an access control gateway and Ethereum Blockchain will provide smart contract management for the marketplace.

Feedbacks from advisors:

Carlos Pastor: introduced ESSIF and said that they are in the phase of finishing the official documentation like functional specification about ESSIF and it's important to use the APIs that are going to be published (new version is due by March 2021 and then APIs will be available). ESSIF will not provide a wallet and they are trying to foster the market to provide wallets compatible with ESSIF. If a wallet is compatible with ESSIF and also with KRAKEN, it could be a good idea. Focus of ESSIF and KRAKEN are different, but he thinks they are compatible. There are meetings for stakeholders of ESSIF that KRAKEN team could attend and it's important to read the official documentation in order to be compliant. If KRAKEN team needs, it's possible to arrange a meeting with ESSIF team. There are specifications regarding policies use and most of all the APIs (again not worth to use present version, better to wait for the new one by March 2021) ESSIF will provide some tools to check the compatibility of other software with the ESSIF reference implementation.

CRYPTO components

Within the crypto-related part of WP4 we presented our approaches on data sharing and data analytics. For data sharing, we presented first, the standard or trivial approach; and second, possible advanced approaches. In the trivial scenario, Data Owners need to give active consent whenever a Data Consumer purchases their data; thus, for real-time purchasing Data Owners essentially need to be online and available "24/7". In the advanced scenario, Data Owners use cryptographic means to protect their data in a way that only certain consumers can get their data. Essentially, the Data Consumer would buy access to the data via the KRAKEN Marketplace. Moreover, these advanced cryptographic means allow the Data Consumer to purchase the data essentially in real time, so the Data Owners can also be offline, while keeping the privacy of the Data Owners intact.

For data analytics we presented first, the approach and an example workflow using Multi-Party Computation (MPC); and second, the approach using Functional Encryption (FE). In the MPC scenario, Data Owners split their data in shares and encrypt these shares for the individual MPC nodes. Then upon request of a Data Consumer via the KRAKEN Marketplace, the MPC nodes fetch the corresponding data and compute together the analytics function. The result is encrypted and sent to the Data Consumer. Our security assumption in this MPC scenarios is that, e.g., at least one MPC node is honest. If this is the case, the Data Consumer gets the analysis result and the KRAKEN Marketplace some metadata, and besides that no one really learns something about the Owners' data. Furthermore, Data Owners can define permitted functions. With these permitted functions, the honest MPC node(s) would abort the computation on an invalid request. In the FE scenario, users provide an FE ciphertext and a corresponding FE key. Upon request, the Data Consumer gets the FE Key and FE Ciphertext to apply the analytics during the decryption of the ciphertext. Also, in this FE scenario, Data Consumers only learn the analysis result. Though, please note that the concrete approach for the aggregated multi-user FE scenario is ongoing work. After our presentation, we asked the advisers for feedback. On the one hand in a generic way, and on the other hand with our prepared questions.

Feedbacks from advisors:

Melek Önen, as a crypto expert, gave us valuable feedback in an abstract way on our data-analytics approaches. She mentioned that there exists a company called ZAMA[3] which focuses on efficient privacy-preserving computation on encrypted data - as we do it, e.g., for the data analytics - leveraging Homomorphic Encryption (HE). And with respect to the used technology of ZAMA, it might be worth to give at least an overview of the different techniques which help to perform privacy-preserving computation on encrypted data: (mainly) MPC, FE, and HE. For instance, where each of the techniques

[3] https://zama.ai/technology

has its advantages and disadvantages, and why we chose MPC and FE. Apart from that, she mentioned that there might be use cases where hybrid approaches are useful.

Furthermore, Melek liked the idea of giving the Data Owner the power to define permitted functions on the data analytics; like user-defined policies on their data. And with this regard, she thinks that a user-defined threshold of the data analytics in terms of the amount of overall Data Owners, being part of the respective analytics computation, seems to make the most sense.

With respect to checking the validity of the Data Owners' data within the MPC scenario, she pointed to a possible solution for the case when the Data Owner participates in the computation as a node. Moreover, Melek mentioned that the presented approaches on data sharing and data analytics look quite interesting and she is looking forward to seeing them further developed and integrated in the KRAKEN System.

Andre Kudra: Browser of available data and statistics about the data pool is required. A waiting period may be acceptable, depending on the business case. MPC hosting also depends on the business and use case. It is quite common that "own hosting" is chosen because of misunderstandings regarding rights and technical possibilities in the "foreign hosting" case (e.g. organizations want to become "Stewards" because they believe it is the only way to get write access).

3.3.3 Ethical and legal aspects

One of the ethical requirements imposed by the European Commission to the Consortium at the beginning of the project was the appointment of an Ethics Board composed of relevant independent expertise to monitor the ethics issues in the project. The Board was appointed in February 2020. The Consortium decided to select the board members from among project partners playing a relevant role in the marketplace development and the pilots and from PAB members. Peter Burgess, a recognized expert in the ethics of security, data protection and privacy accepted the position. It was also decided that both project boards: Advisory Board and Ethics Board will meet together three times along the project life.

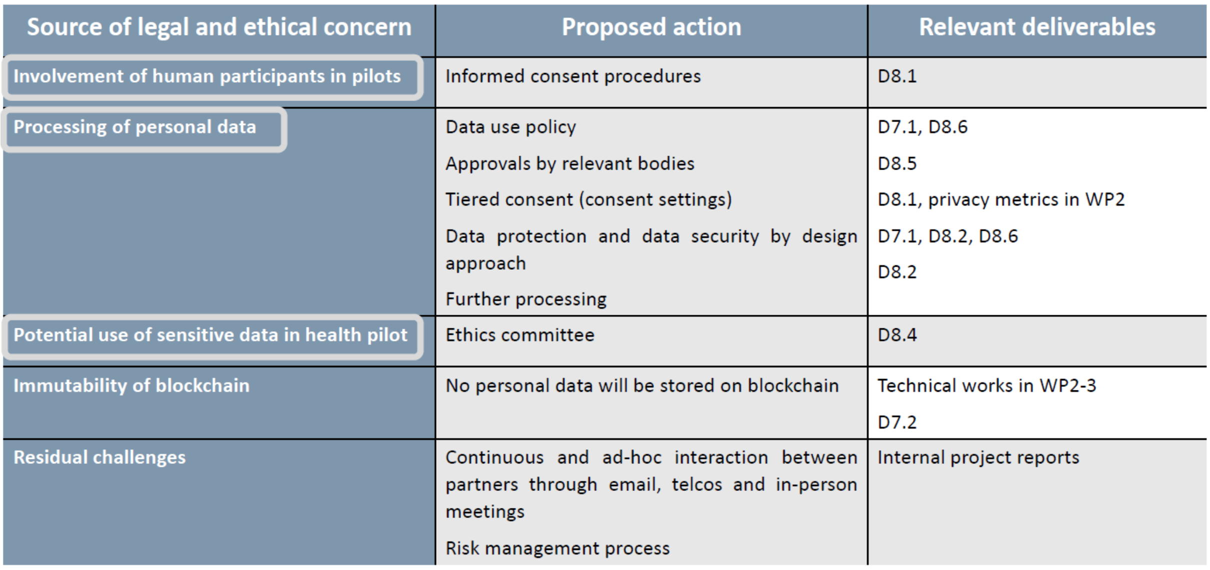
This meeting can therefore be also considered as the first Ethical Board meeting. As it is evident from the previous sections of this document, Ethics is closely intertwined with the other two topics discussed in the agenda: value proposition and business models and technical approach.

Projects like KRAKEN continue to be challenged by the EU General Data Protection Regulation (GDPR) and vice versa. The GDPR holds many challenges which could be addressed by projects like KRAKEN:

- Transparency (information to data subjects) and consent management tool
- Privacy-by-design approach
- Encryption and the GDPR: pseudonymizing and mitigating measures

The Consortium aims to fully respect applicable EU legislation and ethical principles and has undertaken a thorough study and research of the ethical and legal framework applicable to both: (1) research activities and, (2) design and development of the KRAKEN system and final project results.

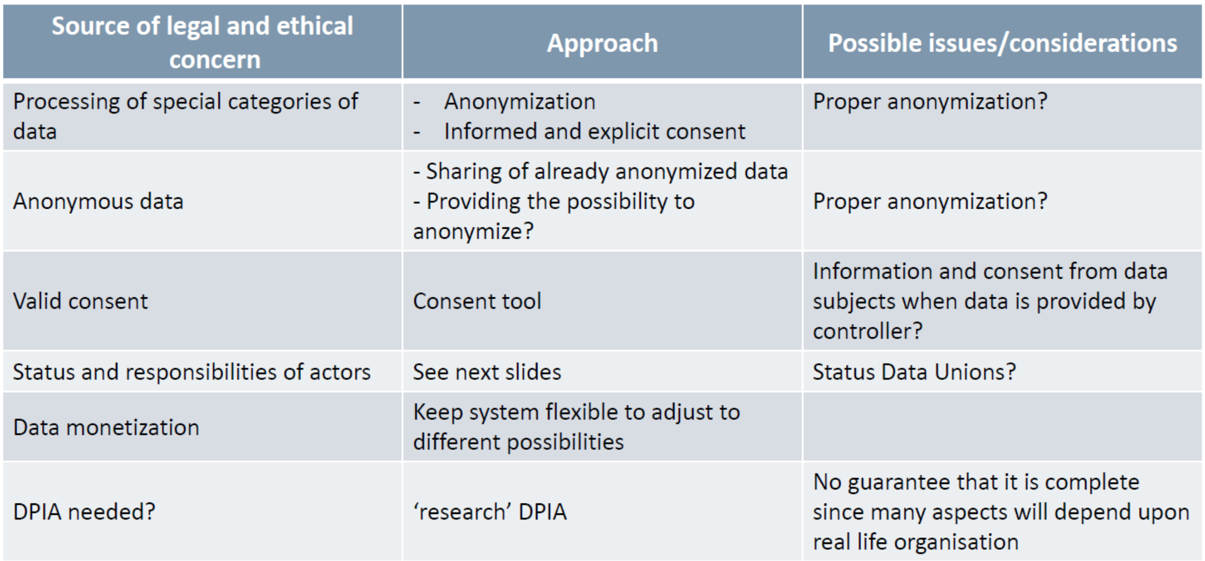
Next figure shows the ethical and legal concerns addressed to comply with the applicable legislation in the project research activities:



Source of legal and ethical concern	Proposed action	Relevant deliverables
Involvement of human participants in pilots	Informed consent procedures	D8.1
Processing of personal data	Data use policy Approvals by relevant bodies Tiered consent (consent settings) Data protection and data security by design approach Further processing	D7.1, D8.6 D8.5 D8.1, privacy metrics in WP2 D7.1, D8.2, D8.6 D8.2
Potential use of sensitive data in health pilot	Ethics committee	D8.4
Immutability of blockchain	No personal data will be stored on blockchain	Technical works in WP2-3 D7.2
Residual challenges	Continuous and ad-hoc interaction between partners through email, telcos and in-person meetings Risk management process	Internal project reports

Figure 1 - Ethical and legal aspects of research

The figure below summarizes the ethical and legal aspects that need to be addressed in relation to KRAKEN development activities and the technologies used to build the project results.



Source of legal and ethical concern	Approach	Possible issues/considerations
Processing of special categories of data	- Anonymization - Informed and explicit consent	Proper anonymization?
Anonymous data	- Sharing of already anonymized data - Providing the possibility to anonymize?	Proper anonymization?
Valid consent	Consent tool	Information and consent from data subjects when data is provided by controller?
Status and responsibilities of actors	See next slides	Status Data Unions?
Data monetization	Keep system flexible to adjust to different possibilities	
DPIA needed?	'research' DPIA	No guarantee that it is complete since many aspects will depend upon real life organisation

Figure 2 - Ethical and legal aspects of development

<u>Feedbacks from advisors:</u>

Harald Zwingelberg: Clarify the "Data Owner" problem, maybe avoid this terminology besides clarifying what it means within KRAKEN and what it does not mean. I wondered why a data subject must be 18 years. Art. 8 GDPR sets 16 years. Certainly, you are on the safe side with 18 years.

André Kudra: Trading data is a sensitive topic in general. The fact that "surveillance capitalism" has become the current norm doesn't make it right. A key challenge is that data stakeholders are made aware of data sharing / selling / exploitation consequences and that they are enabled to make educated decisions when providing data use consent. This is a whole different approach to the data-driven economy. People have to understand their power and use it wisely and cautiously.

Peter Burgess: about pilot of healthcare, thinking about the status of the data that we borrow from hospitals, Hospitals are holding data and are liable of those Data, so the agreement with these subjects could be enough for the Consortium. He suggested also to evaluate to buy “dummy data” to realize the pilot. On development side, about the data monetization, he said that this is very important thing to address for the future of blockchain: it is essential that fundamental rights won’t be touched by the monetization’s model without clear consent of data subject.

4 Recommendations and work packages impacted

The next table shows a summary of PAB recommendations.

Name of Advisors	Work Package	Recommendation Description
André Kudra	WP6	Pharmaceutical companies can be too valuable
André Kudra	WP6	The marketplace should incorporate also soft skill considering individuals not a source of data
Harald Zwingelberg	WP6	Clarify the definition of Data Owner and Data Buyer
J. Peter Burgess	WP6	Improve transparency in marketplace, enabling individuals to control which type of data are shared with who and how they are managed
André Kudra	WP6	Digital agents could be considered in the business model
J. Peter Burgess	WP6	The marketplace should consider that the data continues beyond the transaction on marketplace
André Kudra	WP6	The marketplace should enable users to give consent on specific topic, not a large and generic consent
Andrea Migliavacca	WP6	The consent management of the marketplace should be the most transparent for Data Owner
Andrè Kudra	WP6	The marketplace should consider an incentive for Data Owner
Andrè Kudra	WP6	The consent management should be well defined, if a large or specific consent
Andrè Kudra	WP6	The marketplace should also consider that the way data are exposed can influence the habits of users
Carlos Pastor	WP2- WP3	Compliance with ESSIF Project. Consider using the new ESSIF APIs that will be published in March 2021.

Name of Advisors	Work Package	Recommendation Description
		ESSIF will provide some tools in order to check the compatibility of other software with the ESSIF reference implementation.
Harald Zwingelberg	WP7	Maybe Consortium should evaluate to shift the age of data subject from 18 years to 16 years. Art 8 GDPR sets 16 years
Peter Burgess	WP7 – WP5	Evaluate use of “dummy” data for pilot implementation
Peter Burgess	WP7	Data monetization should take into consideration the protection of data subject rights
Melek Önen	WP 4	Encouraged Consortium to give an overview of the different techniques which help to perform privacy-preserving computation on encrypted data: (mainly) MPC, FE, and HE, bringing as an example a company named ZAMA.

Table 2: Recommendations from advisors

5 Conclusions

The meeting with advisory board members has been very constructive. On one hand the Consortium received a positive feedback regarding the approach and the strategic choices pursued; on the other hand, very interesting suggestions on both business aspects and technical features were examined. The PAB recommendations will impact future project activities and deliverables such as:

- D6.4 - Initial Exploitation Plan and Report M15
- D5.3 - Initial KRAKEN marketplace integrated architecture document
- D3.3 - Data model and ledger for biomedical marketplace first release
- D4.1 - Progress report on cryptographic protocols for privacy-preserving data markets and SSI systems

Finally, as results of this meeting the consortium has found out a way to stay aligned with ESSIF technical specifications and has learnt about ZAMA, a company which uses Homomorphic Encryption to build privacy-preserving solutions.

KRAKEN

@KrakenH2020

Kraken H2020

www.krakenh2020.eu

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