Making personal data accessible while strictly private - Design challenges in the KRAKEN project

KRAKEN is an ongoing three-year project that TX has embarked on as part of a consortium of 10 European tech companies and research institutions. The aim of this project which is funded by EU's Horizon2020 programme is to build a privacy preserving marketplace for personal data in full compliance with the GDPR regulation, and based on novel technologies such as blockchain, SSI (self-sovereign identity), and SMPC (secure multi-party computing).

The project is currently in its second year and approaching its first prototype launch, which gives an opportunity for an overview of intermediate results and learnings. A recent blog post by Rob Holmes delves into a more specific description of the project and its goals, while the point of view in this post is that of a designer working on KRAKEN's frontend user experience.

What's so difficult in trading (access to) personal data?

Trading access to someone's personal data, obviously with the informed consent of that person, does not need to be more complicated than regular ecommerce. However, current data marketplaces are merely at the first stage in their evolution and thus comparable to webshops in the turn of the millenium. Many users (both data providers and consumers) are first timers and need guidance in why and how to share data, including more abstract topics related to privacy and legality.

A web survey conducted last year in the KRAKEN project indicated that many institutions in the healthcare sector have potentially valuable data to be shared, but struggle to do it because they lack either a strategy or policy to share data, or a clear consent from the data subjects. The typical way to exchange data is still directly between the data providers and consumers, without brokers or platforms that could help to form the market, connect the demand of data with its supply, and support both parties in fulfilling the legal requirements.

For legal reasons, personal data cannot be sold like any ordinary commodity. We need to think of selling in a limited sense by providing access to personal datasets instead of passing their ownership to the buyer. Another consideration is that to be legally compliant, the data marketplace should probably not have the data in its possession and thus become a so-called data controller. When combined, these two limitations lead to an architecture and service concept in which the marketplace facilitates the trade process by creating a connection between data providers and consumers, without the data ever passing through the marketplace infrastructure.

The data providers who monetise data often need to protect the privacy of data subjects through anonymisation and follow the general principle of minimising data to only the absolutely necessary, while data users eg. in biomedical research may have the opposite

wish to gain access to datasets which are rich in background and personal profile data. This is an in-built difference in expectations that should be mediated by the marketplace. In order to service the needs of both sides of the market, marketplaces may feature data product formats that are made to fine-tune the level of privacy, such as the SMPC.

Many existing data marketplaces operate strictly in the B2B market and serve institutional clients only, while data sharing mobile apps are often aimed at individuals who want to share their own data. One of the starting points in KRAKEN is to break this strict division of the data market and approach both user groups. As pointed out later in this post, extra efforts are required for individuals to connect to the marketplace in the form of innovative data sharing technologies (Data Unions), frontends such as mobile or smartwatch apps, and service concepts which are tailored for them.

Owing to today's data-hungry research, innovation and business activities, especially artificial intelligence and machine learning, the use of personal data has the potential to expand rapidly. However, the concerns of privacy and legality, and lacking data sharing strategies and platforms contribute to making the availability of suitable datasets a bottleneck for many data users. KRAKEN promises to tackle these concerns.

The KRAKEN design approach

The above challenges are related to both technology and design and can be approached by analysing the data provider and consumer needs carefully and tailoring the marketplace design accordingly. The project follows roughly the paradigm of user-centered design in which the designers focus on users and their needs at every stage of the project, typically by involving them in an iterative process in which the design is improved in incremental development cycles.

Each year of the KRAKEN project has a different focus from the design point of view as follows:

Year 1 was spent on user research and modeling the marketplace user flow through wireframing. User research and service design was done with a web survey, interviews, personas, and design workshops. The results from this process such as user stories and wireframes were deployed as input to the software development.

In **Year 2** the focus is on finalising the UI and UX design, launching successive marketplace prototypes and gathering feedback from them through user testing.

In **Year 3** the goal will be to improve the marketplace experience iteratively and based on quick feedback from possible platform users, and thus accelerate the design cycle. The design work will be connected to marketing and sales efforts which aim at the launch of the marketplace.

How to apply design to empower the users and smoothen the user-experience

Contrary to the personal data protection (of both data subjects and platform users) which is ensured by the underlying technological layers in KRAKEN, I find that the main role of design in this project is to make data sharing actually happen by empowering and educating users and offering them an easy access to the platform.

Making KRAKEN a compelling and easy-to-use platform needs to begin with educating and guiding first time users and enabling their onboarding, while ensuring that more experienced data providers and consumers can reach their goals efficiently. Through design (and some marketing), the underlying technologies can be made understandable, their relevant applications compelling, and their use easy and accessible – although the integration of technical components such as a crypto wallet, SSI credentials, and encryption into a seamless frontend experience can be a challenge.

An integral part of the user experience is the design of search tools for users to find data products of their interest, and the metadata that powers the search. Data products differ from typical webshop items in this regard, as they are often suited to very specific purposes only. The search logic may also differ from one sector or use case to another and thus call for a tailored metadata structure, exemplified by the two KRAKEN pilots in biomedical and education data. A 'one size fits all' search box will probably not work as the users require more specific filtering options. For example, biomedical data products can be searched with a medical taxonomy, while university course grades would be found with such metadata as the name of the university, education programme, and course. It is possible to improve the discoverability of data products with this kind of sector specific tailoring, but a price is paid each time the marketplace is scaled to a new data market sector. An ideal metadata system would thus combine an essentially uniform structure with some modularity and customisation.

Privacy and legal (mainly GDPR) compliance are main selling points of the KRAKEN platform, as the project aims not only to fulfill the requirements, but to excel in these areas. To implement the compliance truly well beyond the platform architecture, there is also a need to make users actually understand their rights and obligations, provide their informed consent, and remind them whenever needed. This all requires careful UX design. It is a challenge to guide users to wise privacy choices with carefully laid out options, explanations, limitations, and help texts. It remains to be seen to which extent the data market will be driven with the same principles of privacy and legal compliance as in KRAKEN, as there always exist consumers that prefer a less privacy preserving approach that poses less limitations.

B2B data marketplaces tend to focus on institutional, professional users and the monetisation of data from their own field of expertise, while the motives and skills of individual persons, or data subjects, to share their data may be very different. It appears that personal data has often more value when applied to the individual's own utility than if it's simply monetised. In the lines of this, many existing data sharing apps aim to create value by empowering their users, giving them control to their own data, and providing options to share the data for their own benefit. In the context of KRAKEN, individuals with a certain shared cause such as a medical condition or shared sports or hobby interests would be good candidates to data sharing. A generally accepted cause to share data (such as to support medical research), or the support of a peer group makes the sharing all the more likely.

The way that KRAKEN will interface with individuals is mainly through Data Unions, which are platforms for crowd-selling data and sharing the revenue between the individuals. To function correctly, Data Unions need their own service providers or administrators who act as intermediaries between individuals and the marketplace. They are also likely to have their own interfaces, eg. mobile apps for the collection of the data, thus circumventing the KRAKEN frontend which is more attuned to institutional users. The integration of Data Unions is however an upcoming topic in KRAKEN as the current first pilot prototype caters mainly for institutional users and the B2B data market.

Tackling the project risk with detailed market knowledge

The three-year timeframe of KRAKEN is long enough for the market to change during the project, which means that some of the initial assumptions may need to be changed. Now that the project is reaching its half-way milestone and beginning the pilot phase, it is time to do these redefinitions if necessary. To give an example, one of the questions under consideration is which currencies, either crypto or 'normal' fiat ones, will be used for trading on the platform. According to our web survey, the adoption of cryptocurrencies is not yet widespread in many of our target groups.

All in all, as a forward-looking initiative that aims to change the modality and terms of data sharing, KRAKEN faces also a risk of failure in its entry to market.

The basic concept of data marketplaces is simple, and the complexity is all in the details. This explains why a project of the size of KRAKEN is needed to make data trading platforms comply with the current and upcoming legal and privacy standards. Another reason to the big challenge is that personal data is a highly specialised product with niche markets, user groups and use cases that should be understood - all of which contributes to making the project heavy in research and design work.