Call identifier: H2020-ICT-2016 - Grant agreement no: 732907

Topic: ICT-18-2016 - Big data PPP: privacy-preserving big data technologies

Deliverable 7.4

Study report

Due date of delivery: December 31st, 2019

Actual submission date: January 1st, 2020

Start of the project: 1st November 2016

Ending Date: 31st December 2019

Partner responsible for this deliverable: Lynkeus

Version: 5.0

Document classification

Title	Study report
Deliverable	D7.4
Reporting period	Second
Authors	Ludovica Durst, Davide Zaccagnini
Work package	WP7
Security	Confidential
Nature	Report
Keyword(s)	Study, report

Document history

Name	Remark	Version	Date
Ludovica Durst	First draft	1.0	18/12/19
Davide Zaccagnini	Second draft	2.0	23/12/19
Ludovica Durst	Third draft, including final statistics	3.0	31/12/19
Anna Rizzo	Statistical analysis review	4.0	31/12/19
Antonella Trezzani	Final review and submission	5.0	01/01/20

List of contributors

Name	Affiliation
Ludovica Durst	Lynkeus
Davide Zaccagnini	Lynkeus
Anna Rizzo	Lynkeus

List of reviewers

Name	Affiliation
Davide Zaccagnini	Lynkeus
Anna Rizzo	Lynkeus
Antonella Trezzani	Lynkeus

Table of contents

1 Introduction

This deliverable describes activities in WP7 and their results primarily around the goal of assessing public perceptions and attitudes towards privacy and data security, in contrast with actual behaviour as observed in the use of the MHMD platform. The WP was built on a logic of consequential steps, starting from the conceptual model of key societal issues addressed by MHMD, stemming from the study design and preliminary analysis conducted in D7.1, in which the issues of *trust, privacy, ethics and control over data* as understood and lived in day to day life were historically, culturally and sociologically framed. The goal in the line of work was *"to assess public perceptions, attitudes and sensibilities of different subjects towards privacy and data security in health, taking particularly into account the ethical and sociological implications related to the use of advanced technological tools and infrastructures"*.

To this end, an extensive literature review was carried out and operational definitions of those four themes were instantiated along with their interrelationships and the methods to practically assess and wherever possible quantify them in European populations. This analysis was also framed in the context of the GDPR and its emphasis on these aspects. On this basis and taking into account the individual user requirements described in D1.1, a questionnaire was designed to quantify the prevalence of certain attitudes and to support an evidence-based study of users' behaviours, stated values and preferences, as detailed in D7.2.

In the meanwhile, the opportunity to conduct an additional study arose in the context of the Wannacry Ransomware attack in the summer of 2017, which severely impacted a number of EU hospitals, including our project partner QMUL. Given the stringent relevance of this event to the themes and objectives of the MHMD project, an additional line of research was initiated to specifically investigate public responses to such a major privacy and data security event using sentiment analysis techniques on both traditional and social media. The sentiment analysis work was included in D7.2 highlighting few relevant trends, among which that the expected decrease in trust among the UK citizenship toward institutions' ability to protect personal data was not as sharp as predictable. After an initial drop, expressed through widely shared disparaging or alarmed tweets in relation to the event, the NHS response mediated by public media managed to restore levels of trust that were just below neutral on the sentiment scale. The trend also reached a steady plateau soon after with an overall, only slight deflection despite the pervasive vulnerabilities of UK healthcare institutions were extensively reported, suggesting a pre-existing disillusionment in the citizenship on the reliability and trustworthiness of medical IT systems and at the same time only marginal interest in the fate of own's personal data.

As detailed in D7.3, a user enrolment strategy to disseminate the questionnaire, embedded in the MHMD app, was implemented. Individual users were invited to download MHMD app through a dedicated marketing campaign involving health apps providers Digi.me and Medicus, to then perform data mining on their selected preferences comparing users' actual behaviours with preferences and opinions from the questionnaire. The results of this analysis are the subject of this deliverable.

2 Individual users questionnaire analysis

The work performed in WP7, *"Platform-driven assessment of attitudes and sensibilities with regard to ethical, privacy, and data security issues"* has led to a clear conceptualisation of four main issues: *privacy, trust, data control and ethics* and the definition of indicators to assess them. From these definitions and their interrelations Lynkeus, in collaboration with P&A and Digi.me, has derived research questions on which the questionnaire on users' preferences and the notifications system on user's behaviours were built.

The hypothesis underlying the social study is that, in a highly dynamic information ecosystem, citizens should be aware of both the sensitivity and value of health data while researchers, public health officials, businesses and policy makers should be able to access those data efficiently as they pursue their legitimate agendas. And yet, people diffusely deplore a lack of control over their data, while paradoxically showing little interest in taking direct responsibility in managing them, mostly because of the complexities and time-consumption the data protection process entails. All this while uncontrolled aggregation and exploitation of personal data by corporations has taken the centre in public debate. Mistrust at times translates into a general, undirected reluctance to share data, and will increasingly do so, impeding research and innovation. These issues seem to lay in poor understanding of how the personal data ecosystem works and, despite the GDPR which clearly emphasizes individual empowerments, a lack of tools to enable citizens to make the most out of their data.

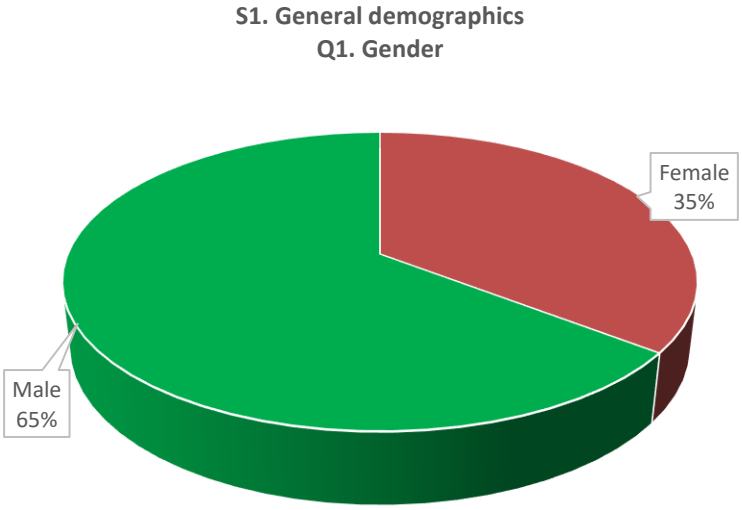
The survey on privacy preferences, privacy control and privacy agency of health data aimed at addressing these issues, guided by the following to research questions:

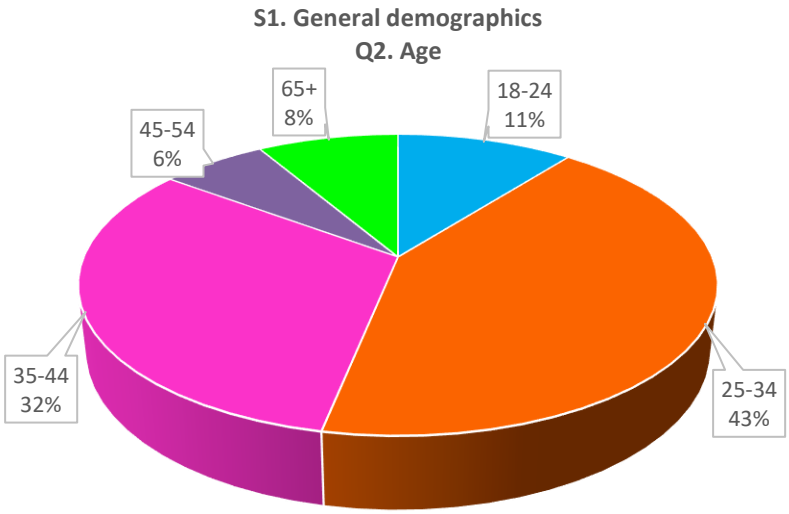
1. *Do patients prefer to handle themselves their personal data, or do they prefer to hand over the control to a proxy?*
2. *What are the attitudes and preferences of the clinicians and patients about data donation and data solidarity?*
3. *Which type of health data have most value (i.e., are most shared), under what conditions, in the MHMD platform?*
4. *Can the privacy paradox be quantified by analysing the inconsistencies between what people say about their privacy and their actual privacy-protecting behaviour?*

A relatively modest sample of users (n=48), still in line with the sample size specified in the DoW, completed the questionnaire. As the MHMD application remains available to users, additional data may become available in the coming weeks, allowing more robust statistical analysis. At the same time, the results gathered so far provide valuable insights into cognitive and cultural dynamics around personal data and value that can be derived from them.

Section 1. General demographics

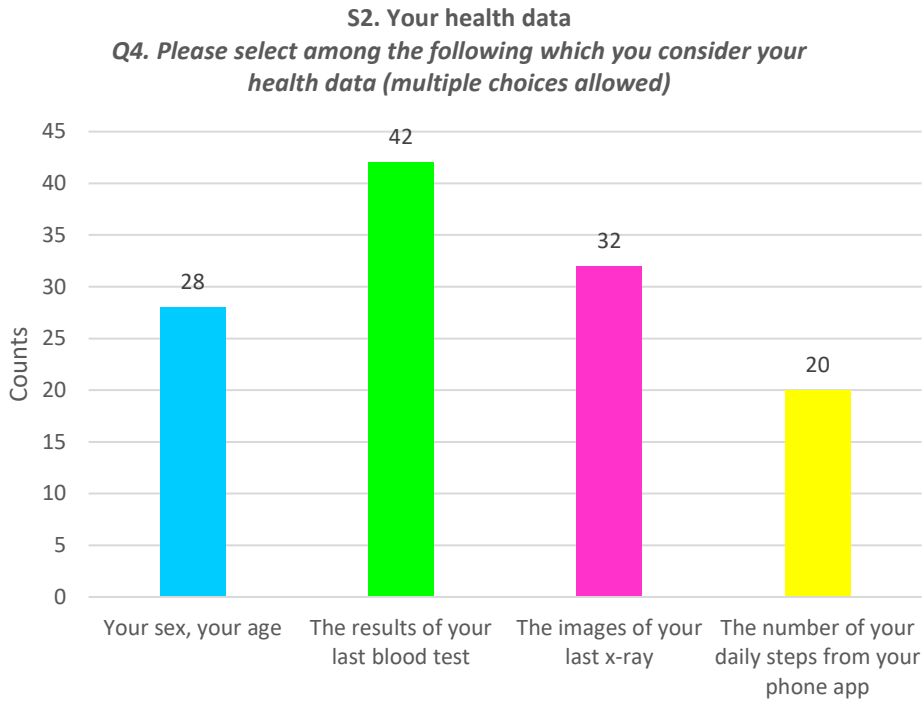
Also, the general demographic of the sample is the same described in D7.3: the total number of respondents was 48 (50 was the assessed value for the study), 35% female and 65% male. The biggest portion (75%) of the sample is made up of people aged between 24 and 44.





Section 2. Your health data

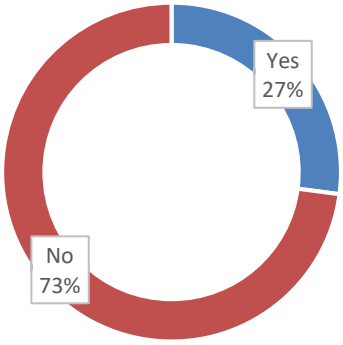
The second group of survey questions aimed at understanding the level of awareness about what can be considered 'health data' from the point of view of an individual user. While respondents have almost no doubts that such definition applies to blood test or X-ray, it appears less clear that other kind of data, such as sex, age and data collected via phone app, like number of daily step would be considered health data. In the context of the "quantified self" movement powered by now ubiquitous wearable devices this has important implications. While in fact corporations are actively using this type of data for medical and health-related applications, only half of the sample, for instance, seems to be deeming their physical activity data medically relevant, pointing to a gap in public awareness of how their data can be and are actually used.



The vast majority of enrolled MHMD app users (73%) was trying a medical data management app for the first time, even if 44% were already using some mobile app related to health or medical care. Among them, only 24% declared to use it/them every day, 24% once to few times a week, 38% once to few times a month, 10% very rarely and 5% never (data not shown).

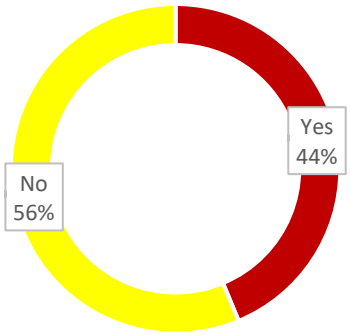
S2. You health data

Q6. Do you already have an app on your mobile device that allows you to access your medical data from your hospital or family doctor?



S2. Your health data

Q7. Do you have on your mobile device any type of application related to your medical care (e.g., about medications, sugar levels, blood pressure, diet)?

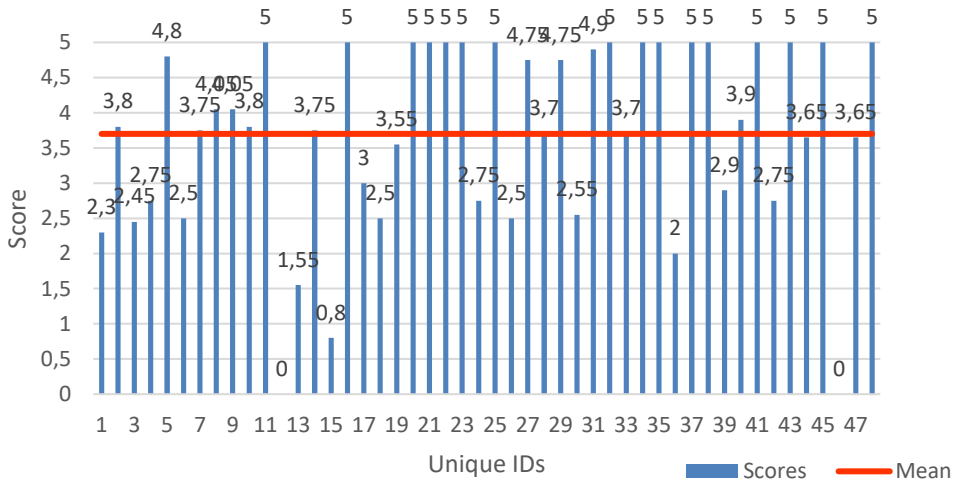


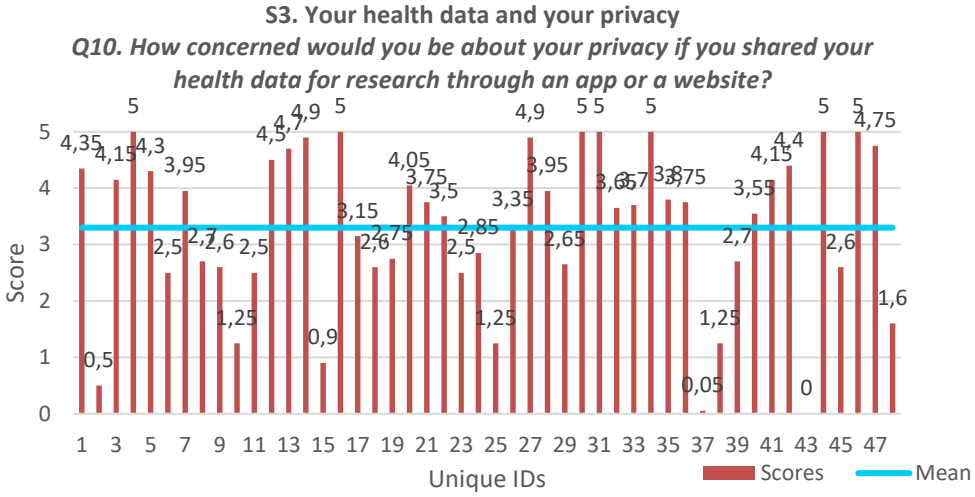
Section 3. Your health data and your privacy

In order to assess the relation between personal data sharing and privacy concerns, users were asked how happy they would be, in a range from 1 to 5, to share their health data for medical research, and how concerned they would be about their privacy if they had to share their health data for research through an app or a website (with 5 being the most concerned). Results show that users have a good propensity to share their data for medical research (mean = 3.7) even though most of them were rather concerned about their), despite privacy (mean = 3.3). Not surprisingly, 98% of them (data not shown)concerns, while almost 100% would feel more confident to share their health data if they knew they were properly de-identified *(i.e., data is processed in a way that it is not possible to track back the identity of a data subject anymore)*. This finding, while not surprising, points to a key trend in the general discussion around how to protect data. Full anonymization seems to be the preferred method for individuals to make their data available, in line with the focus spent during the project on synthetic data as a tool to decouple individual identities from the information content of the data.

S3. Your health data and your privacy

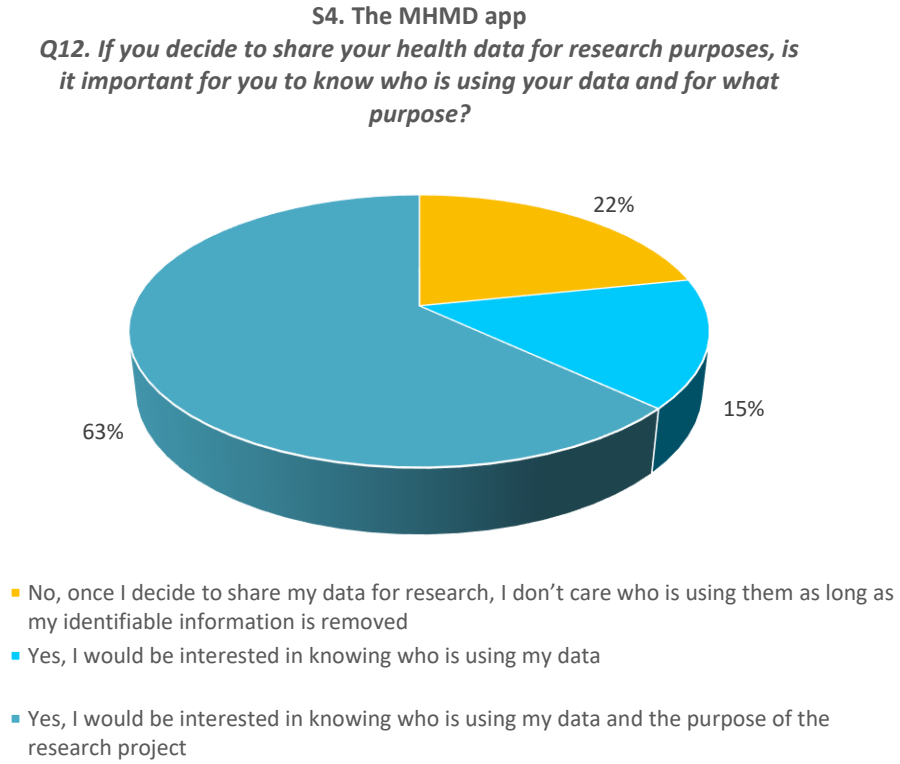
Q9. Researchers use large amounts of data to derive information about a medical condition. Would you be happy to share your health data for medical research (...)?





Section 4. The MHMD app

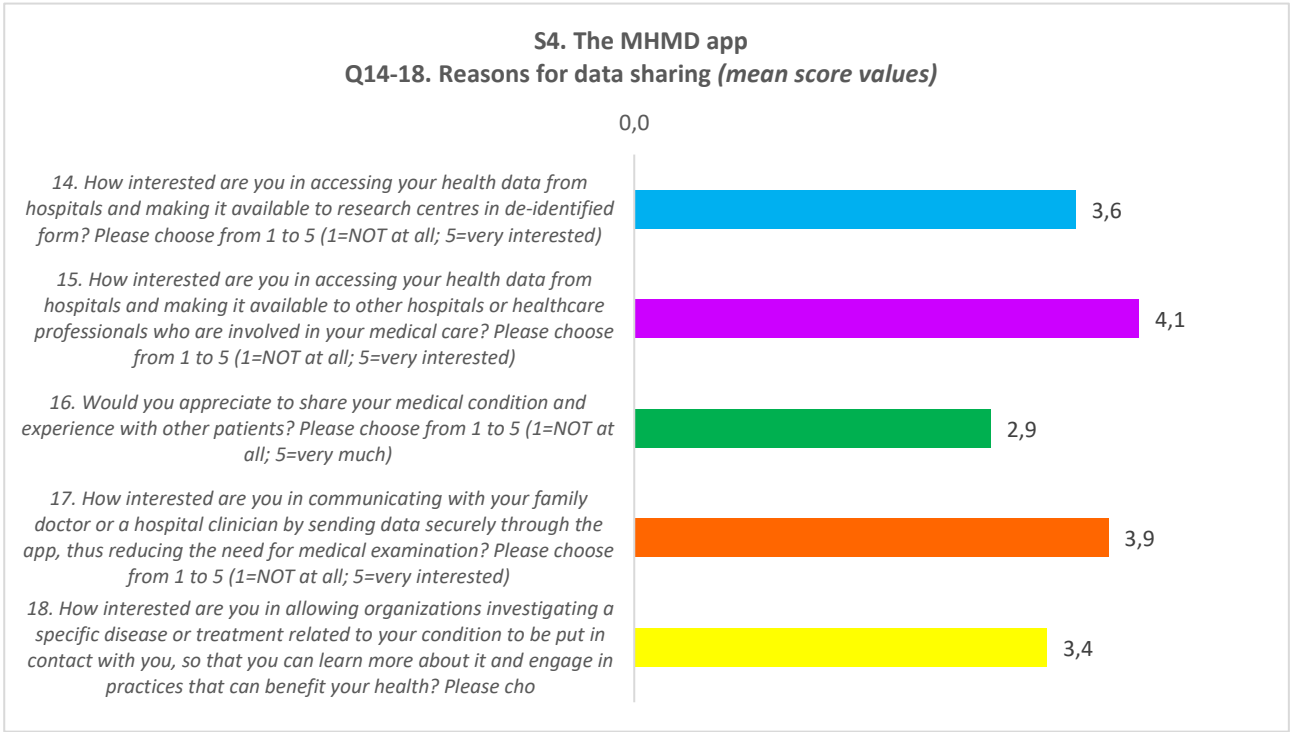
Some questions directly addressed the possibility to set and personalize different consent options, in order to understand how useful these could be for an individual user to actively engaging them in managing their data and their medical care by using the app. For 63% of users it is important to know both who is using the data and for what purpose, a smaller portion (15%) would be interested in knowing who is using the data, while 22% of users believe that once they decided to share their data , they don't care who is using them as long as any identifiable information is removed.



On one hand, these results show the patients' will to be effectively engaged and in control of their data, being enabled to decide who to share their data also on an ethical level. Privacy remains essential but, even

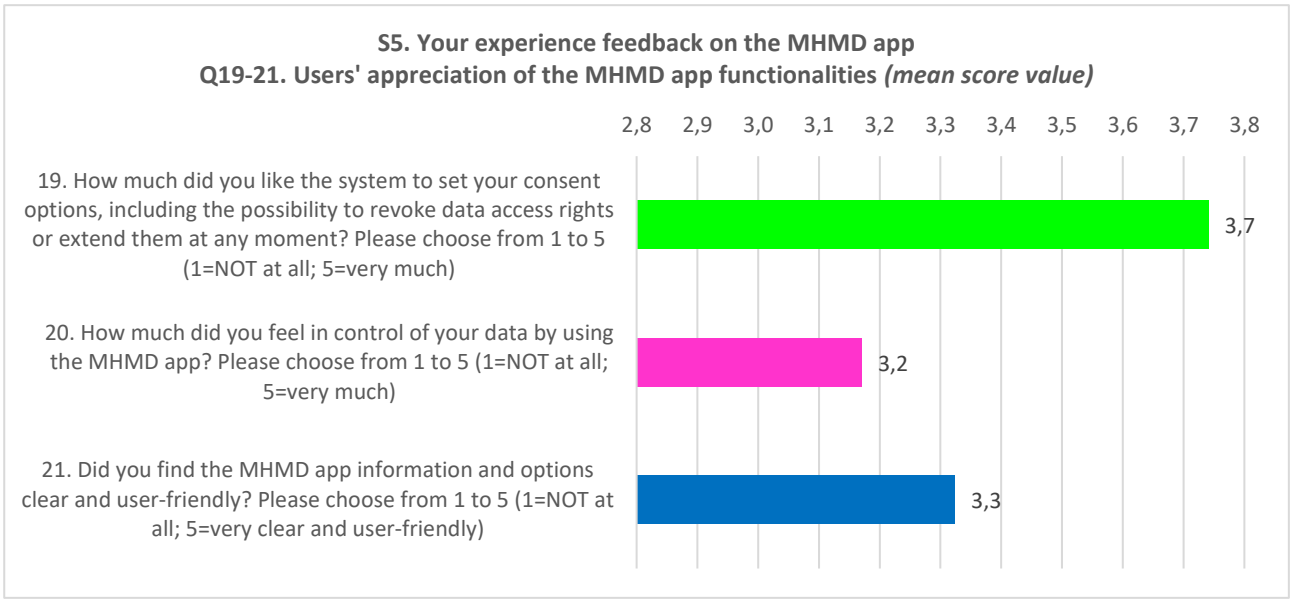
if guaranteed, it does not remove the desire to see who and how it uses one's data, pointing to a level of recognition of the value inherent to the data. Especially in regard to this, 98% of users would prefer to share their data knowing that they could obtain services or other rewards for that (data not shown).

With regard to the purposes in which they are most interested, by asking to assign a value from 1 to 5 (with 1 = NOT at all, and 5=very interested), MHMD app users put, as first, the possibility of *accessing their health data from hospitals and making it available to other hospitals or healthcare professionals who are involved in their medical care* (mean score=4.1), followed by the possibility of *communicating with family doctor or a hospital clinician by sending data securely through the app thus reducing the need for medical examination* (mean=3.9) and by the possibility in general of *accessing their health data from hospitals and making it available to research centres in de-identified form* (mean=3.6). Also, allowing *organizations investigating a specific disease or treatment related to the patients' condition to be put in contact with the users, so that they can learn more about it and engage in practices that can benefit their health*, raises good interest (mean=3.4), even if probably users don't like so much the idea of being contacted, or rather disturbed. Finally, below average and not much appreciated is the possibility to *share medical condition and experience with other patients* (mean=2.9). In general, there seems to be more confidence in patient data sharing than with institutions, which means, in other words, that there is more trust in organizations, when they are credible, than in any citizen.

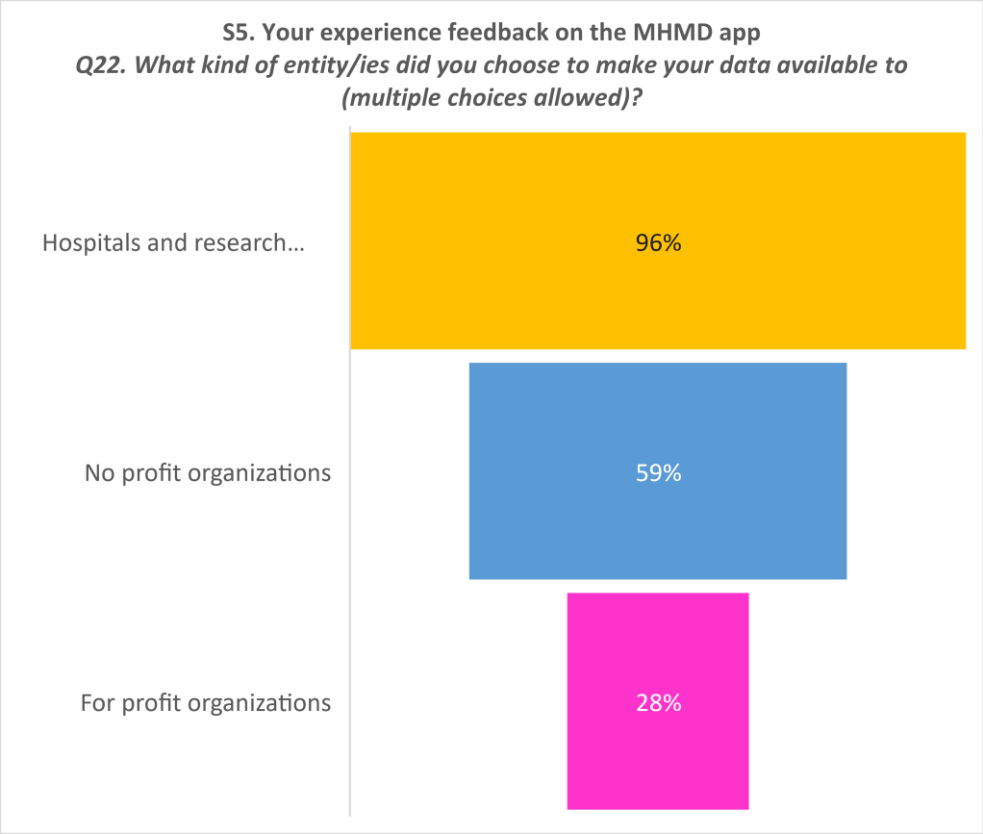


Section 5. Your experience feedback on the mhmd app

Users seem to have very much appreciated the features of the app enabling them to set their consent options. Assigning a score from 1 to 5 (1=NOT at all; 5=very much), they mostly valued the *possibility to revoke data access rights or extend them at any moment* (mean score= 3.7), while the app has achieved less success with reference to the *clarity of information and options and user-friendliness* (mean=3.3) and the *feeling of being in control of the data* (mean=3.2). This highlights the known challenge of simplifying in app-based or even desktop-based workflows the legal and ethical complexities of patient empowerment over their data as, by definition, granular control requires deeper engagement and therefore more time and cognitive burden spent on the controlling mechanism, i.e. the app. While human-machine interaction techniques could mitigate this friction, our opinion is that new approaches to information management should be explored to streamline users interactions with the consent and data management process. Nevertheless, considering that the app released was a test prototype, these results are encouraging.



When asked to express their will through multiple-choice questions, almost all users (96%) decided to make their data available to hospitals and research centres, a majority of users (59%) also allowed no profit organizations, while few only (28%) consented to for profit organizations. Yet the vast majority responded positively to direct data access requests from private companies, in contrast with what specified in the consent and in the questionnaire. It appears thus that it is not only a matter of trust in a specific institutions, but also of the dynamic of the interactions. As organizations present themselves in a transparent way (with name and surname, and for what purposes) users seem to be inclined, at least in the moment, to respond positively.



When asked to express their interest for various possibilities offered as reward to the sharing of their data, with possibility to select different opportunities with multiple-choice questions, MHMD app users greatly appreciated the possibility of *being informed about progress in scientific and medical research performed with their data* (78%) as well as of *being informed about discoveries related to users' health conditions outcoming from their data* (80%). Less favoured was the possibility of *being informed and put in contact with other physicians treating patients with conditions like theirs* (50%). Once again, users seem to prefer and trust more institutions than other patients (data not shown).

Finally, users seem to have most appreciated the possibility of *being informed about research results and data traceability and the transparency through accessible and easy opt-out options, including information on if and when data is used* (63%) and the *possibility of being informed about research results and data traceability* (72%) and , less to be informed about the "altruistic benefit" of sharing data (48%). Most of users (83%) stated that they would use the app shall it be released (data not shown).

3 Results from the comparison between individual users questionnaire and users' behaviours

If we try to compare the results emerging from the actual behaviours of the MHMD app users and their preferences and attitudes as stated in their answers to the survey, some interesting findings might be stressed.

The propensity stated in the questionnaire to consent to use any kind of data for research purposes appears in line with actual users' behaviour, as only few of them chose to select and specify what type of data to be used, for what/not for disease(s).

Users have a good propensity to share their data for medical research (the average value is 3.7), despite privacy concerns, but this propensity decreases slightly (3.3) if data sharing via web or app is required, even if almost 100% would feel more encouraged to share their health data if they knew they were properly de-identified.

First of all, it must be highlighted that privacy concerns are not overwhelming the desire to share data to support scientific research. Also, even if users may have some major concerns and mistrust with regard to industrial usage, profiling, and being re-contacted, as less than 50% of users allowed these options, when reacting to notifications referring to profiling, statistical analysis, industrial usage, they didn't seem to perceive them as directly affecting their privacy, nor particularly worried or concerned by the possible usages that different type of organizations may put in place (be they for industrial usage and statistical analysis, or related to profiling or further contacting the patients/users).

This behaviour is also in line with the results of the questionnaire showing that more than 60% of users believe it is important to know both who is using the data and for what purpose of the research project.

Moreover, almost all users decided to make their data available to hospitals and research centres, the great majority of them also allowed no profit organizations, while few only consented to for profit organizations. This fact is in line with actual consent options selected, allowing both public and private research options, and with the fact that in the case of study requests coming from research centres – regardless they are public or private – and from Pharma companies, users appear more confident in accepting (100% accepted), on the contrary in the case of study requests from private companies denials reached 10%.

Pearson correlation of stated consent for the sharing among different research entities, and actual preferences expressed through the app. It is evident that the correlation among the two is very low, confirming the gap between expressed will and concerns and actual users' behaviour.

MHMD APP CONSENT SETTINGS Research	SURVEY S5. Your experience feedback on the MHMD app *Q22. What kind of entity/ies did you choose to make your data available to (multiple choices allowed)?*		
	Hospitals and research centres	**No profit organizations**	**For profit organizations**
Public sector	0,16	-0,01	-0,07
Private sector	0,10	0,06	0,06

In all other cases, were high or low monetary reward was proposed, or data required were social media data (and no health data), some requests were declined.

Comparing actual users behaviours and their preferences and attitudes as stated in their answers to the Survey leads to significant findings.

The propensity stated in the questionnaire to generously consent to allow use of data for broad research purposes appears in line with actual users' behaviour, as only few cases chose to select and specify what type of data to be used, for what/not for disease(s).

Privacy concerns are not overruling the desire to share data to support scientific research. Users have shown major concerns and mistrust with regard to industrial usage, profiling, and especially being re-contacted, as less than 50% of users allowed these options, possibly indicating preoccupation with possibly annoying patterns of interactions. At the same time, when directly reacting to notifications referring to profiling, statistical analysis, industrial usage, they didn't seem to perceive them as directly affecting their privacy, nor particularly worried by the possible usages that different type of organizations may put in place (be they for industrial usage and statistical analysis, or related to profiling or further contacting the patients/users).

Rather, actual behaviours demonstrate that most users tend to accept study requests without much concerns, as long as they are informed of who is the subject requesting the data and for what research purposes. Here again, the specific timing and presentation of the request, along with the its source, seems to carry an overwhelming weight against users' stated preferences and opinion, strongly tipping their behaviour toward sharing rather than controlling. This can possibly be explained by looking at general patterns of use of mobile apps in which fine-tuned and hyper-designed user experiences drive user behaviour much more than conscious control.

Nevertheless, some privacy concerns may be related to the lower percentage of users allowing secondary use consent or to be contacted or to have their data used for virtual cohort compositions (all below 40%), even if in using MHMD App, users seem to have most appreciated the possibility of being informed about research results and data traceability and the transparency through accessible and easy opt-out options (including information on if and when data is used), less to be informed about the "altruistic benefit" of sharing data.

While almost 100% of users would prefer to share their data knowing that they could obtain services or other rewards for that, their actual behaviours show once again that users seem to depend more on who is requesting the study than the reward promised. In one case only, in fact, 100% of users accepted the study request: it was the case in which no monetary reward was proposed, but only feedback on personal health conditions.

4 Conclusions

As per tasks 7.4 and 7.5, in the last months the individuals' app was released and the questionnaire was submitted to individual users, recruited among the general public. The objective was to recruit 50 individuals/patients to obtain a direct evaluation of users' stated values and preferences through the surveys: even if the sample is small to perform and assess statistics, it is valid for cognitive interpretations, to be combined with data gathered from users' statistics (e.g., number of users, data logs, type of consent).

These results provide partial but insightful indications of how privacy perceptions and intentions compare to actual behaviours, indicating areas for further research. The key lesson is that no single features would realize the goal of conscious and engaged citizens actively sharing their data under clearly understood ethical and legal references, but that instead a delicate and complex balance between control features, trust creation through open and specific communication, overall simplicity and above all transparency is needed to meet users expectations.

Indeed, the Report by the Academy of Medical Sciences on 'Our data-driven future in healthcare People and partnerships at the heart of health related technologies' (November 2018) includes among the main principles on which data-driven technologies should be designed for health to respect and protect the privacy, rights and choices of patients and the public, with special regard to trust and transparency. As stated in the Report, *"Trust in healthcare professionals and the NHS is high but this needs to be actively maintained and reinforced, particularly with the introduction of innovative data-driven technologies. Transparency is central to achieving this through demonstrating trustworthiness. Thereforerobust but simple communication frameworks are needed to foster clarity about why and how data-driven technologies and associated patient data are being used, by whom, for what, and how decisions about these uses are arrived at"*.

As shown by the work done in this WP, supporting research represents a valid goal for most users, who overall are not really concerned by privacy risks as long as subjects requesting data and research purposes are clearly stated, and users are enabled through consent options to feel in control of their data. In other words, people are inclined to trade personal information for control as long as this can be exercised efficiently. At the same time de-identification is considered indispensable. This is particularly true if we consider that recent statistics (HRA/ University of Sheffield workshops 2019) exploring anonymised patient data use with mixed public and private benefit, showed a shift from 18% support for commercial use of patient data pre-workshop to 45% post-workshop. While most people in this study remain in other words sceptical of commercial data use, transparency provided in the form of workshop based communication more than doubles that predisposition.

As more data may be collected in the interim period before the final review these results may be updated. As of now , in their limited statistical value and scope, they point nevertheless to key principles that should be considered in design and implementation strategies to generate engagement around privacy preserving solutions, namely the paramount need for transparency, the guarantee of full anonymization and of carefully designed user experiences.