Dominik Leibenger, Frederik Möllers, Anna Petrlic, Ronald Petrlic, and Christoph Sorge

Privacy Challenges in the Quantified Self Movement – An EU Perspective

Abstract: The gathering of data about oneself (such as running speed, pulse, breathing rate, food consumption, etc.) is rapidly becoming more popular, and has lead to the catch phrase "Quantified Self" (QS). While this trend creates opportunities both for individuals and for society, it also creates risks, due to the data's personal and often sensitive nature. Countering these risks, while keeping the benefits of QS services, is a task both for the legal system and for the technical community. However, it should also take users' expectations into account. We therefore analyze the legal situation of QS services based on European law and the privacy policies of some major service providers to clarify the practical consequences for users. We present the result of a study concerning the users' views on privacy, revealing a conflict between the user's expectations and the providers' practices. To help resolve the conflict, we discuss how existing and future privacy-enhancing technologies can avoid the risks associated with QS services.

Keywords: Quantified Self, Privacy

DOI 10.1515/popets-2016-0042
Received 2016-02-29; revised 2016-06-02; accepted 2016-06-02.

1 Introduction

The term *Quantified Self (QS)* arose first in 2007 within the context of personal data collection. It was initially coined by Wolf and Kelly for a meetup of 30 people interested in "new practices" like *life logging* and *location tracking*. [37] Today, such meetups are held in 125 cities. [28] According to a study from June 2014, *health* and *fitness* apps—considered to be QS apps—are used daily by users. Their usage has increased by 62 % within six months. [12]

Our starting point is the observation of two developments: The increasing demand for data privacy (emerging especially after the NSA spying scandal became known), and the ongoing growth of the QS movement in recent years. At first glance, both seem to be incompatible developments. While data privacy is about *imposing constraints* on collection and dissemination of personally identifiable information, the QS movement *wants* to collect and analyze as much personal data as possible. We show that these goals are in fact compatible, but research is required to bring them together in a practical manner. We focus on privacy aspects arising from the use of QS services rather than the sensors themselves or their local (e.g. body area) network connectivity. The interested reader is referred to publications by Barcena, Wuesst and Lau [5] and by Hilts, Parsons and Knockel [18] for a discussion of these aspects.

2 Quantified Self

QS is about the collection of all kinds of different data by users about themselves. Imagine a user, Alice, who stores data about her *workouts*, *weight*, *health condition*, etc., using her smartphone. Smartphones are commonly used for QS purposes, as they provide appropriate sensors and are able to retrieve data from external devices (e.g., digital scales that transmit users' weights). Gathering these data provides multiple opportunities, as shown next.

2.1 Benefits

We can think of numerous different use cases that are supported by analyzing the collected data. The following overview shows just a few possible use cases belonging to four categories that cover a wide range of benefits. They are organized according to the users (one or multiple users) and the data (one- or multi-dimensional) covered by the respective analysis (see Table 1) and described below.

Dominik Leibenger: CISPA, Saarland University, Germany, E-mail: dominik.leibenger@uni-saarland.de
Frederik Möllers: CISPA, Saarland University, Germany, E-mail: frederik.moellers@uni-saarland.de
Anna Petrlic: Independent Researcher, E-mail: anna@petrlic.eu
Ronald Petrlic: Commissioner for Data Protection Baden-Württemberg, Germany, E-mail: petrlic@lfd.bwl.de
Christoph Sorge: CISPA, Saarland University, Germany, E-mail: christoph.sorge@uni-saarland.de

		Users	
		Individual	Multiple
Data	One-dimensional	Reports (I)	Competition (II)
	Multi-dimensional	Correlation (III)	Research (IV)

Table 1. Benefits of quantified self services

I *Reports (1-dimensional data, 1 user):*
Alice could log her workouts just for *fun* (e.g., in order to see how many miles she ran in a year). Such reports, based on health data, can also be used to *control behavior* and to *live healthier*. When, for example, Alice notices that she is gaining weight, she might be motivated to increase her workouts.

II *Competition (1-dimensional data, n users):*
If Alice starts to compare her workout data with Bob's data, she might be able to *improve her sports performance*, e.g., by copying him when she recognizes that Bob is more successful.

III *Correlation (n-dimensional data, 1 user):*
In the big picture, analyzing correlations and dependencies of different data provides even more possibilities. For example, Alice can observe the impact of her sports performance on her weight. Also, by correlating a log of her mood she might be able to see the impact of a healthy life to her mood in the long run—constituting a new finding about herself.

IV *Research (n-dimensional data, n users):*
The huge amount of QS data voluntarily collected and shared by users can also constitute a tremendous source of data for research. A health research institute interested in finding out about the correlation of sports, weight, mood, etc., could "simply" consolidate the collected data from thousands of users and analyze their correlations. Alice may support this because she also benefits from the research (e.g., when learning that many users doing regular uphill runs like herself develop knee problems).

2.2 QS Services: State of the Art

QS services available today already cover the whole range of these benefits. An overview of services associated with different categories via *tags* is given in the QS community's *Guide to Self-Tracking Tools* [30]. As this overview is expansive, we tried to identify a representative subset of popular services covering many areas of a user's personal life. We could not determine the numbers of users for all services, so we selected services from the guide both by means of numbers of users and by the "popularity" as provided by the guide in April 2014.

Runtastic is a popular (85 million users at the time of writing) platform for tracking / competing in sports performance.[1] *Withings*[2], the second-most popular tool tagged as "gadget", provides hardware for fitness data tracking (e.g., weight, pulse). Health condition and success of treatments can be tracked and analyzed with *CureTogether*[3], the most popular tool under "medicine". A tool for tracking user's emotions is *MoodPanda*[4]—the most popular service for each tag "lifelogging", "lifestyle", "location", "mood", "social", and "web app". Also popular (50 million users) is *Foursquare*, a social network for tracking and sharing location-based information.[5] To cover tools affecting other areas of a user's personal life, we further analyze *Goodreads* (a social network for book readers with 50 million users)[6], the expenditures-tracking platform *Mint* (second-most popular tool with tag "money", 20 million users)[7], and the time-tracking service *RescueTime*[8] (second-most popular tool tagged with "productivity").

Regarding the benefits listed in Sect. 2.1, each analyzed service provides a user with *reports* based on her individual data. Runtastic, for example, generates extensive reports that allow the user to analyze both concrete sporting activities in detail as well as the development of her fitness over time. Except Mint and RescueTime, all services encourage *competition* between users by allowing them to compare data. All services employ *correlation* of different data of a specific user or give recommendations based on data collected among different users. The only benefit still rarely provided is *research*, which only CureTogether states as its primary goal. CureTogether's popularity, however, confirms that users are indeed willing to share personal data for research purposes. Some QS apps have provided a glimpse into potential research benefits—a prominent example being Jawbone, which published aggregated data on the effect of an earthquake on their users' sleep [11].

1 https://www.runtastic.com/; see [33] for number of users. All referenced webpages have been accessed on May 17, 2016
2 https://www.withings.com/
3 http://www.curetogether.com/
4 http://www.moodpanda.com/
5 https://www.foursquare.com/; see https://foursquare.com/about for number of users
6 https://www.goodreads.com/; see https://www.goodreads.com/about/us for number of users
7 https://www.mint.com/; number of users claimed on https://blog.mint.com/credit/mint-by-the-numbers-which-user-are-you-040616/
8 https://www.rescuetime.com/

With the exception of Runtastic, which allows using a limited subset of its features offline, all investigated QS services require the transmission of a user's personal data to a cloud backend, which provides the user with access to and analysis of the logged data.

Beyond that, a recent trend shows QS service providers are not only interested in processing and storing personal data at some central location. Instead, they also share data with each other, creating an even more comprehensive data repository. In 2014, Apple introduced *HealthKit* [3] in iOS 8 as a centralized storage of health data obtained through QS services and for sharing the data throughout different QS health services. Shortly afterwards, several QS service providers (*Runtastic*, *Nike*, *Withings*, and *RunKeeper*, among others) stated interest in using it. [23] Samsung announced a similar service earlier that year: *Samsung Architecture Multimodal Interactions (SAMI)*, a "data broker" allowing users to upload their QS data to the cloud. [34] Google presented a similar service dubbed *Google Fit*[9] at its annual developer conference in 2014. [31]

2.3 Risks

Regardless of concerns about specific tools, we highlight some general risks of processing personal data in the cloud, as is done by nearly all available QS services.

The risks a user faces when using QS services certainly depend on the *sensitivity* of the collected data—a term defined more precisely in the context of our user study presented in Sect. 4, and also reflected in legal considerations (Sect. 3.1). For non-sensitive data, the main concern might be loss or manipulation of previously stored data. Sensitive data, however, additionally require the user to strongly trust the QS service provider both to not give away data and to maintain an infrastructure that is secure (e.g., against hacker attacks).

As shown in Sect. 2.1, QS services are usually not limited to processing data of specific users in isolation. Instead, they provide comparisons between and statistics among different users, as well as features to share data (sub)sets with specific users or the public. Those features bear the risk of leaking personal data to others. This might be for obvious reasons like lack of access rights management or mismanagement by the data owner, or due to targeted attacks that try to extract information about an individual from published statistics or other data derived from the QS provider's data set.

Apart from that, it is questionable to what extent the assumption of a trustworthy QS provider is realistic. First, the provider might be forced to give away personal data (e.g., to law enforcement agencies). Second, it might be tempting to sell such data to third parties (e.g., for financing a free-of-charge QS service). Besides advertisers, conceivable clients might also be insurance companies or employers interested in health-related data about specific persons. This may violate those persons' personal rights, and also have direct financial consequences. An insurance company could decide not to accept a customer based on the collected data, or could charge an increased premium; an advertiser could use the data to try selling tailored products to the user—a practice considered useful by some persons, but bothersome by others.

Clearly, these risks are amplified if data are shared between different providers. Such sharing might happen not only due to explicit usage of tools like HealthKit, due to law enforcement, or for financing purposes. QS service providers could also decide to merge their services or be acquired by possibly larger services. Runtastic, for example, has recently been acquired by Adidas, a large manufacturer of sporting goods that has already been active in the field of wearables and QS before. [1]

While the mentioned risks are, in practice, broadly accepted today, there are more subtle risks one can think of: A QS service provider in control of collected data and interpretations thereof could try to maliciously manipulate such data (e.g., with the goal of analyzing and/or influencing a user's behavior). This is not an unlikely scenario as is shown by a recently published study in which the mood of Facebook users was successfully influenced just by manipulating the choice of entries presented in their news feeds. [25]

3 Law and Privacy Practices

In order to verify our so far theoretical study, we investigated the different QS services mentioned in Sect. 2.2. By examining relevant legal regulations and comparing them to the privacy policies as well as actual practices of the providers, we determined a real-world view of the risks that go hand in hand with the benefits offered.

3.1 Legal Situation: Theory

In member states of the European Union, personal data are legally protected by Directive 95/46/EC and its na-

9 https://developers.google.com/fit/

tional transpositions. Its successor, the Regulation (EU) 2016/679 of 27 April 2016, is known as the General Data Protection Regulation (GDPR) and will be applicable from May 25, 2018. In contrast to a directive, a regulation is directly enforceable law in the member states. Contradicting national legislation is overridden. The GDPR contains some clauses that require or allow member states to regulate specific aspects, e. g. the minimum age of a child for consenting to data processing.

The directive and the GDPR are not only applicable to European service providers. The national transpositions of the directive apply also if data are processed using "equipment, automated or otherwise, situated on the territory of [...]" a member state of the European Union by a non-European company according to article 4, paragraph 1 (c) of the directive. The GDPR goes further by explicitly stating (in its Article 3) that it applies to any data processing related to "the offering of goods or services" within the EU, or to the monitoring of behavior in the EU—even if the data controller or processor is not established in the EU.

A core concept of both the directive and the GDPR is to allow processing of personal data only if certain conditions are met; otherwise, such data processing is forbidden regardless of the specific nature or relevance of the data. The conditions which allow processing of personal data are quite far-reaching, but when interpreting data protection law, it is important to keep the general principle in mind. Neither the directive nor the GDPR apply to data processing "by a natural person in the course of a purely personal or household activity" (Article 3, paragraph 2 of the directive; Article 2, paragraph 2 (c) of the GDPR). Thus, tools used only locally by individuals will not normally be problematic. We therefore focus on data processing by service providers in this section, starting with considerations concerning service provision, and then discussing further usage of the data. The latter is relevant since many QS services are provided in the cloud (see Sect. 3.2), giving the service provider complete access to the data.

3.1.1 Data Processing for Service Provision

Processing of personal data is allowed (according to article 7 of the directive), among others,

- with the data subject's unambiguous consent
- if it is necessary for "the performance of a contract to which the data subject is party", or
- if it is necessary for "purposes of the legitimate interests pursued by the controller", unless these interests "are overridden by the interests for fundamental rights and freedoms of the data subject".

The last item appears overly broad, but is partly concretized in national legislations. It is not generally interpreted as a carte blanche for data processing, but requires an actual weighing up of the respective interests. This means that in most cases, service providers will need the respective users' explicit consent for any processing that goes beyond what is contractually required. Alternatively, they can process anonymized data.

The GDPR does not substantially change this assessment despite a change in wording (see article 6, section 1, of the GDPR): The unambiguity requirement has been moved to the definition of "consent" (article 2, item 11 of the GDPR). In addition to data processing necessary for "the performance of a contract", processing "in order to take steps at the request of the data subject prior to entering into a contract" is now also explicitly allowed. The "legitimate interests" clause now specifically mentions the case of a child as data subject, since children are considered as particularly vulnerable.

The specific nature of QS services may add another complication. As a basic principle, processing of "data concerning health"[10] is prohibited in the EU member states. Of the few exceptions defined in the directive (and, similarly, in the GDPR), only the "explicit consent to the processing of those data" is likely to apply for QS scenarios. This affects all services that process health data such as a person's pulse. The consequence—the need to ask for the users' consent—is not problematic per se. QS services are not intended as secret surveillance, but consciously chosen by their users. Services can therefore ask for consent when necessary.

There are, however, some potentially problematic conditions. For instance, there are services whose collected data are not only related to their respective user, but also to others. From a legal perspective, the relevant issue is whether *personal* data about others are collected. A scale by Withings, for example, measures—in addition to the user's weight—room temperatures and CO_2 concentrations. These data are not specific to the owner or user of the scale, but as they cannot be attributed to any other individuals, this does not pose a problem from a data protection perspective. CureTo-

10 These—and some other—data are considered as "special categories of data".

gether collects only information about registered users, but in case of genetic diseases, this information may also apply to family members. This issue has been mentioned by a few authors (e.g., Kayen [24]); a detailed analysis of the implications of sharing genomic data is provided by Humbert et al. [20]. We believe that users are still allowed to disclose the information—considering both the fact that information about family members is not systematically collected, and that the data also refers to the respective CureTogether member himself. Foursquare uses other users' information to provide a service for a user, but all users are registered on the platform and can thus be asked for their consent[11]. Mint, which provides the user with an overview of her financial transactions, may be more problematic; transaction partners may be identifiable individuals who do not use the service themselves. As Mint does not currently target European users[12], the classification according to the European directive is not a practical issue. However, assuming that a service like Mint is useful, finding ways for the legal and privacy-preserving handling of third-party information remains an interesting research issue both from a legal and a technical perspective.

The same reasoning can also be applied to data that are not collected, but computed by a QS service (e.g., statistics generated for a user). These data might still contain (as they depend on) personal data about persons who have not given their consent. The investigated services provide only sufficiently aggregated or anonymized information of people who have not given their consent, though, so no additional issues arise.

To summarize, there is no general issue preventing the provision of QS services under European law. This holds both for the directive and for the GDPR. However, practical implementations could still violate data protection rules.

3.1.2 Other Usage of QS Data

As pointed out above, the usefulness of data processed by QS tools is not limited to the individual user. Firstly, comparisons to other users—or similar statistics—can be provided. Such a usage is not normally problematic, as it is part of the service for the users, and will usually be covered by the contract about service provision.

Secondly, data can be used for other purposes, that are not part of the service for its users. To protect the privacy of "data subjects", European law requires personal data to be "collected for specified, explicit and legitimate purposes and not further processed in a way incompatible with those purposes"[13]. There are several ways to get around that restriction:

- As the directive only applies to personal data, the use of anonymized data is not problematic. In line with the definition of personal data, recital 26 of the directive (and, similarly worded, recital 26 of the GDPR) states "the principles of protection shall not apply to data rendered anonymous in such a way that the data subject is no longer identifiable". Whether or not a method of anonymization is sufficient to prevent identification of a data subject is not purely a legal issue, but requires a technical assessment as well: Recital 26 of the GDPR explicitly states that "the available technology at the time of the processing and technological developments" must be taken into account when judging whether information is sufficiently anonymized.
- The processing of data for other purposes than fulfillment of the contract (with the QS provider) could be specified as a purpose beforehand, and the users could be asked for their consent concerning that processing.
- National transpositions of the directive can allow exceptions for scientific research. Article 6, paragraph 1b continues: "Further processing of data for historical, statistical or scientific purposes shall not be considered as incompatible, provided that Member States provide appropriate safeguards". The German legislator, for example, provides such an exception. For private bodies, it can be found in section 28, subsection 2, item 3 of the Federal Data Protection Act[14]. Under that regulation, personal

11 Note, however, that users are not prevented from uploading information about third parties, e.g., photographs showing other people than registered users.

12 It was chosen for our analysis anyway since it is a popular service within the QS community's guide to self-tracking tools that represents a rather specific area that is not yet covered by any service also focusing on the European market.

13 Article 6, paragraph 1b of the directive; the principle is also found, in almost the same wording, in Article 5, section 1b of the GDPR.

14 1. Federal Data Protection Act (BDSG) in the version promulgated on 14 January 2003 (Federal Law Gazette I, p. 66), last amended by Article 1 of the Act of 14 August 2009. We use a translation provided by the Federal Ministry of Justice and Consumer Protection in cooperation with juris GmbH at http://www.gesetze-im-internet.de/englisch_bdsg/englisch_bdsg.html

data can be used for a different purpose "if necessary in the interest of a research institution for the purpose of scientific research, where the scientific interest in carrying out the research project significantly outweighs the data subject's interest in ruling out the possibility of collection, and the purpose of the research cannot be achieved in any other way or would require a disproportionate effort." Whether the conditions are met, especially considering the "scientific interest in carrying out the research project", is a question that can be answered only individually, but not in general for all QS tools.

The GDPR contains a similar exception to the directive, and uses a broad definition of the term research, including "technological development and demonstration" (recital 159). Article 89 (1) of the GDPR regulates the required safeguards "for the rights and freedoms of the data subject" in a very broad manner. Technical and organizational measures are required, with pseudonymization mentioned as an example. The article also requires data to be anonymized if the research purpose can be fulfilled this way.

– The assessment whether data processing is *compatible* with the original purpose provides some leeway. While the directive does not specify how to assess compatibility, the GDPR does (Article 6 (4)): If the change of purpose is not based on the data subject's consent, or on a law of the EU or its member states, the data controller must take five criteria into account for the compatibility assessment. These include "any link" between the original and the new purpose, the context of the data collection, the possible consequences for the data subjects, and the use of safeguards like encryption or pseudonymization. It is an open question how wide or narrow the new rule will be interpreted; however, if it went further than the research exception, the latter would not make sense. We therefore assume a narrow interpretation of the compatibility of a changed purpose with the original one.

As stated above, data processed by QS tools will often belong to the "special categories" of data defined in article 1 of the directive. With the subject's explicit consent, processing these data is also allowed (article 8, paragraph 2a). The mentioned exception for scientific research in the German law also exists, in a similar wording, for special categories of personal data (section 28, subsection 6, item 4 of the Federal Data Protection Act)[15]. The sensitive nature of the data will, however, have to be considered when striking a balance between the scientific research interest and the subject's right to informational self-determination. Similar to the directive, the GDPR (Article 9) also allows processing of such data with the data subject's consent, and—with appropriate safeguards—for scientific research. The sensitivity of the data must also be taken into account when assessing the compatibility of a changed data processing purpose with the original one.

3.1.3 Legal Situation in the United States

The United States' approach to data protection is different to the European one: As discussed above, the EU (and a number of other European countries) regulate the processing of personal data in a very general manner, and only allow such processing if there is a legal basis for it. The United States, in contrast, allow data processing in general, though it can be forbidden under specific laws. [35] An example for such a sector-specific law is the privacy rule found in the Health Insurance Portability and Accountability Act (HIPAA), which regulates the processing of health information by "covered entities" such as medical service providers. Though gathered data may reveal information about someone's health, QS service providers are not likely to be affected by this rule. An example for privacy legislation on the state level is the California Online Privacy Protection Act (OPPA), which requires web sites and mobile apps that collect "individually identifiable information about an individual consumer" to post a privacy policy that meets certain conditions. A complete analysis of the federal and state laws which may apply to QS services is out of the scope of this article.

As a side note, Daly [9] provides an overview of legal aspects of QS tools based on Australian law. The results differ significantly from our European perspective, and the author concludes that an adaptation of Australian data privacy legislation might be necessary.

15 Section 28, subsection 6, item 4 does not explicitly mention that the research must be in the interest of a research institution (as subsection 2, cited above, does). However, Gola and Schomerus ([15], comment on section 28 of the Federal Data Protection Act, recital 74) point out that *independent* research is required in both cases.

3.2 Legal Situation: Practice

After having worked out some *theoretical* legal aspects for QS tools in general, we were interested in how user privacy is treated in *practice* by today's popular QS services. For this we examined the privacy policies, the terms of use, and some practices of the 8 service providers mentioned in Section 2.2. The terms of service and policies we examined are available at https://hyperion.cispa.saarland/quantified-self/policies.tar.xz. The results are summarized in Table 2 and explained below.

The structure of our analysis is based on categories of data protection law, most importantly the European Directive 95/46/EC. The first question is whether data protection law is applicable at all. Obviously, this is only the case for data actually processed by the respective service provider; local data processing under the end users' control is not covered. We analyze which providers process data themselves in Sect. 3.2.1.

Data protection law only applies to *personal data*, defined by the directive (article 2 (c)) as "any information relating to an identified or identifiable natural person". Therefore, the next aspect listed in Table 2, and discussed in Section 3.2.2 is the information that is available to the service provider and has the potential to identify a person.

Article 10 of the directive[16] provides the structure for the following investigated aspects. When collecting information from the "data subject", the service provider ("data controller") must provide the following information (as far as "[...] necessary, having regard to the specific circumstances in which the data are collected, to guarantee fair processing in respect of the data subject"):

- "the identity of the controller and of his representative, if any" (this information is provided by all analyzed services and therefore left out in Table 2)
- "the purposes of the processing for which the data are intended" (discussed in Section 3.2.3)
- "the recipients or categories of recipients of the data" (i.e., the question to which parties data are transferred, which is discussed in Section 3.2.4)
- "whether replies to the questions are obligatory or voluntary, as well as the possible consequences of failure to reply" (most of the data is not collected by asking questions, and the few questions (like for the e-mail address) are clearly marked as mandatory or optional by all service providers)
- "the existence of the right of access to and the right to rectify the data concerning him"

Afterwards, we discuss additional aspects noticed during our analysis of the privacy policies, which do not fit into this structure (Sect. 3.2.5). The table also lists the place of business of the respective service provider, as this influences both the applicable law and the chances to enforce that law.

3.2.1 Data Processing by the Service Provider

We noticed that in general, service providers leave users no choice whether to save the data on their servers or on their own devices. Thus, the users have no control about how exactly the data are processed. Users have to trust the provider to ensure the security of the data. The only exception to this is Runtastic as it allows users to choose which data to upload and which to keep saved only locally. Obviously, data protection law is still applicable for any data the provider processes (like e-mail address etc.).

3.2.2 Personal Data

Provision of an e-mail address is mandatory for the registration to each service (and, in many cases, already sufficient to identify a person). Due to its unambiguity among the users, it is also an obvious choice as a user's identifier for the login procedure. All service providers employ a regular login via email address and password, some also offer the possibility to log in via third party authentication providers such as Amazon, Google or Facebook. While this makes it easy for users, it also means that more data about the user are disclosed: The *service provider* may obtain information stored at the *authentication provider*[17] and the *authen-*

16 The list is modified and extended in the GDPR, but as it is not yet applicable, service providers cannot be expected to take the modified information duties into account.

17 The service providers can specify what information to request from the third-party authentication provider and the users can review and decide on this request. In practice this ranges from only public information (e.g., the user's public Facebook profile), to a combination of email address, birthday and list of friends.

QS Tool	Data Exclusively Stored in Cloud	Login			Additional Mandatory Data				Data Uses						Data Transfer								Place of Business	Policy Features			
		Email	Telephone Number	3rd Party Provider	Email Address	Real Address[1]	Real Name[1]	Other Personal Data[2]	Notifications	Improvement of Service[3]	Personalized Advertisement	Direct Advertisement	... Explicitly Stated	Research	Public by Default[4]	To Service Providers	Service Providers Bound by Policy	To Involved 3rd Parties[5]	To Uninvolved 3rd Parties[6]	... Only Anonymized/Aggregated	Google Analytics/Social Media Buttons	Google Analytics IP Anonymization		Email Notification on Changes	Data Usage Rights Revokable	Provider May Modify Information	User May Access/Rectify Information
Runtastic		✓		✓	✓		✓	✓		✓	✓	✓	✓		✓				✓	✓	✓	✓	**Austria**			✓	
Withings	✓	✓			✓		✓	✓	✓	✓		✓		✓		✓		✓			✓	✓	**France**		✓		✓
Foursquare	✓	✓	✓	✓	✓		✓	✓	✓	✓	✓	✓			✓	✓	✓	✓	✓		✓		**USA (NY)**	✓	✓		✓
Goodreads	✓	✓		✓	✓		✓		✓			✓			✓	✓	✓		✓	✓	✓		**USA (CA)**				✓[7]
Mint	✓	✓			✓	✓			✓	✓	✓	✓	✓	✓		✓	✓		✓	✓			**USA (CA)**		✓		✓[7]
RescueTime	✓	✓		✓	✓				✓												✓		**USA (WA)**	✓	✓		
CureTogether	✓	✓			✓				✓	✓		✓		✓					✓	✓	✓		**USA (CA)**	✓	✓		
MoodPanda	✓	✓		✓	✓		✓	✓	✓	✓	✓	✓	✓		✓	✓	✓		✓	✓	✓		**UK**		✓		✓

Table 2. Comparison of QS Tools' Privacy Policies and practical handling of data (as of June, 2015).

1 or parts thereof
2 e.g., birthday
3 includes purely technical improvements and bug fixes
4 whether any information is publicly available with the default settings
5 e.g., a shop where the user checked in
6 3rd parties not further specified in the policy
7 provider explicitly limits information that users can access/rectify

tication provider may learn about the registration with the *service provider*.

Despite the fact that none of the analyzed services need users' real names and addresses to perform the advertised tasks, 5 providers request a real first and/or last name, and Mint even requests information about the place of residence upon registration. This is critical as no provider visibly states how this information is used or why it is needed. The MoodPanda website states that the name is displayed next to the user's comments, but it does not explain why a pseudonym does not suffice for that. The address required for Mint is later used to suggest businesses in the area for which bills and periodic money transfers can be set up, but this is not stated anywhere in the policy or during the registration process. Practical implementations differ from the privacy policies, though—in our experience, no provider formally verifies the names given during registration. The question whether European law allows the provider to enforce a real-name policy is yet to be decided by court.

3.2.3 Purpose of Data Collection

Regarding the purpose of the data collection (which corresponds to the intended usage of the collected data), we investigated how the users' data are processed in addition to the regular operation of the service. In the case of Runtastic, for example, it is obvious that the tracks and timings of a user are used to measure her fitness and to compare her performance with other people's. In their privacy policies, though, all providers require their users to grant them additional rights.

Notifications are a common purpose for the usage of contact information, and it makes sense to notify users directly about changes in the service's operation. How-

ever, most services do not specify the nature of the notifications and while 3 providers explicitly state that they will send advertisements to the users, all policies but RescueTime's would allow this ("notifications about new products and services").

A very interesting aspect about the notifications is that only 3 of the 8 analyzed providers state that they will notify users about changes in their privacy policy by email. The others merely state that they will notify about them on the website without specifying the nature of this notification. This might be problematic for users if the notification is not prominent enough. Some providers even go as far as giving no notification at all and relying on the users to periodically check the privacy policy on the website, as changes are effective immediately after publication. This practice is concerning, as it theoretically allows a provider to grant itself the rights to sell user data to any other company and then do so immediately before any user has had the time to terminate her account. In the member states of the EU, such a unilateral change cannot justify any data processing, including data transfers, that would require the data subject's consent: The directive (article 2 (h)) defines that consent as "freely given specific and informed indication of his wishes by which the data subject signifies his agreement to personal data relating to him being processed".

As pointed out above, data processing by QS service providers will usually require the users' consent, so unilateral changes of the privacy policies will usually be illegal in the EU; the practice of not notifying users, or at least not clearly stating how users will be notified, is concerning.

3.2.4 Transfer of User Data

We also looked at the rights to transfer user data to other parties. 4 providers make parts of users' data publicly available by default. Any user of the service (or, in some cases, any website visitor) can see some information about registered users. This means that it is possible for adversaries to automatically collect data from the website without the user's knowledge. While contact details like the email or real address are not publicly visible by default, pseudonyms or real names allow any viewer to map details to individuals. From a legal perspective, this means that user information is passed to uninvolved third parties, though almost all services allowing this transfer state that such information is only shared in an anonymous, aggregated manner. However, all services allow users to customize their privacy settings and have access only to data explicitly entered by users, implicitly specifying the amount of information that can be shared.

For 5 of the 8 services, the user has to grant the provider the right to transfer arbitrary data to partners if these partners perform specific tasks ("Data Transfer To Service Providers"). To protect the data from misuse, 4 of them restrict these partners' rights ("Service Providers Bound by Policy"), so the data may only be used for the provided services. While the usage of Google's *Analytics* software and the inclusion of social media buttons (such as the Facebook *Like*) can be seen as a sharing of individual information with uninvolved third parties, this can also fall under the transfer of data to service providers. Interestingly, only 3 providers state this in their policy while 7 include Google Analytics on their websites. Only 2 providers use a feature of Google Analytics which anonymizes visitors' IP addresses to a certain degree (i.e., the last byte of an IPv4 address is stripped). In this regard, only Runtastic complies with their own policy to pass only anonymized or aggregated data to third parties (here: Google) not bound by any policy.

3.2.5 Other Noteworthy Aspects

Finally, we looked for distinct and noteworthy aspects in the terms of service and privacy policies that have not been covered by the categories above. Of the 8 providers we investigated, Runtastic's policy included an interesting paragraph. In addition to the usual rights (to use the data in order to provide the service), the user grants Runtastic the right not only to collect and store, but also to modify the user's data. No statement is made whether this applies to collected service-related data (e.g., traveled distances and times) or identity-related data (e.g., email address or name). Furthermore, according to the policy, the rights to collect, use, store and modify information are irrevocable. This means that in theory, the provider can arbitrarily use and change information about current and previous users in any way and is not bound to give accurate or correct information. We contacted Runtastic about this and inquired how this paragraph is applied in practice, but did not receive any comment within half a year.

Another precarious practice concerns the user's legally guaranteed right to access and amend the information stored about him. Only 5 out of 8 providers

mention this in their policies and 2 of these 5 limit the information that users can access and/or rectify.

In conclusion, we found that while many aspects of privacy policies and the actual usage of data are similar to what can be expected of a free, ad-sponsored service, each of the analyzed services' policies showed concerning or even dangerous practices. These practices (i.e. allowing modifications to the policy without notifying the users, or passing personal data to uninvolved third parties) have the potential to cause harm to oblivious users. It remains to be seen how widely these are used throughout the variety of QS services and how badly they threaten the privacy of the users.

The results are in line with previous research (not specifically related to QS services). Jensen and Potts [21] published a survey of privacy policies in 2004. They investigated the privacy policies of 64 popular websites and found similar problems to the ones identified in the paper at hand. They especially criticized readability and the notification practices—providers expect users to re-read the policies periodically. We still observe the latter a decade later. Borghi et al. [6] performed a similar study in the UK in 2013, once again leading to very similar results. Specifically for QS services, the study by Hilts, Parsons and Knockel [18] contains an analysis of privacy policies, complementing the one at hand regarding the investigated services and the aspects under investigation. In addition, the authors tested the responses of nine QS service providers when a user of the respective service requested information about his stored personal data. While Canadian law requires them to provide the requested information within thirty days, only five companies answered in a meaningful manner.

4 User Study

Given the fact that Runtastic alone has 85 million users as of May 2016 [33] and the numbers for the other services range in similar orders of magnitude[18], the aforementioned problems seem to be no concern for most of the target audience. To gain a better understanding of the users' preferences and practices, we performed a survey[19] among persons residing in the United Kingdom, with different educational backgrounds (see Tab. 3), of different ages (see Tab. 4) and different genders (215 female, 172 male, 6 unspecified). Interestingly, the majority of QS users is under the age of 30, whereas the majority of the non-users is older than 30.

Education	Users	Non-Users
Agriculture	0.9 %	0.4 %
Art	3.4 %	6.2 %
Business	5.2 %	9.5 %
Computer Science	12.1 %	14.4 %
Education and Training	10.3 %	9.1 %
Engineering	7.8 %	4.9 %
Health	8.6 %	7.0 %
Humanities	10.3 %	15.2 %
Law	4.3 %	3.7 %
Services	0.0 %	3.7 %
Social Science	16.4 %	12.3 %
Other	20.7 %	13.6 %

Table 3. Education of survey participants

Age	Users	Non-Users	Age	Users	Non-Users
15-19	9.9 %	11.0 %	35-39	13.2 %	8.1 %
20-24	23.1 %	15.8 %	40-44	6.6 %	7.0 %
25-29	28.9 %	19.9 %	45-49	2.5 %	11.4 %
30-34	14.9 %	14.3 %	$\geq$ 50	0.8 %	12.5 %

Table 4. Age distribution of survey participants

We conducted our study with the online survey service Prolific Academic[20] in February 2016. The service allows anyone to register as a survey participant who provides some basic demographic data; researchers can reject obvious bogus answers, and are required to pay a reward (in our case, 0.85 GBP) to the remaining participants. Our survey was open to the first 400 persons who gave the UK as their country of residence, and whose answers in previous studies had an approval rate of more than 85%. After removing incomplete and duplicate answers, 393 participants remained.

4.1 Reasons for QS Usage

While serving as a potential explanation of the users' behavior, hardly any data are available about *why* users use QS services. Therefore, we were interested in their

18 https://foursquare.com/about; https://www.goodreads.com/about/us; https://blog.mint.com/credit/mint-by-the-numbers-which-user-are-you-040616/

19 The complete questionnaire can be found in Appendix A.2.

20 http://prolific.ac/

reasons for QS usage. We asked QS users and people who do not (yet) use QS services.[21] The results are given in Fig. 1.

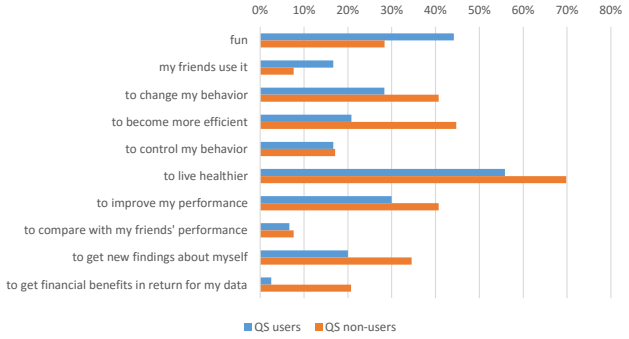

Fig. 1. Reasons for QS services usage

People could answer with multiple reasons to those two questions. We received reasons from 119 QS users and 274 QS nonusers. The results shown in the figure were normalized in the number of respondents to allow for comparison.

Some reasons of QS users differ from those of nonusers. For example, users answered more often that they use QS for fun and that their friends use it than nonusers. On the other hand, nonusers answered more often that they could imagine to become more efficient, to live healthier and to get new findings about themselves. Interestingly, there is especially one huge difference between nonusers and users: the nonusers answered far more often that the reason for using QS would be financial benefits that they get in return for their data. The majority of QS users use sports services. A lot of the users use QS to keep track of their hobbies (especially reading), their location and their financial assets. The reasons also reflect the benefits of QS discussed in Sect. 2.1 and the state-of-the-art of QS discussed in Sect. 2.2.

We were also interested in users' opinions concerning risks of QS. For users to perceive risks as being privacy-related, they must perceive the used personal data as sensitive in the first place. Therefore, we asked them to classify data categories according to their sensitivity.

4.2 Users' View on Privacy

Concerning the users' perception of privacy in QS services, we found out that data concerning *health* are sensitive, which is in line with processing of health data being prohibited by law as described in Section 3.1. Data concerning *location* and *finance*, however, are even more sensitive, which is not considered by legal regulations. The results are given in detail in Fig. 2[22].

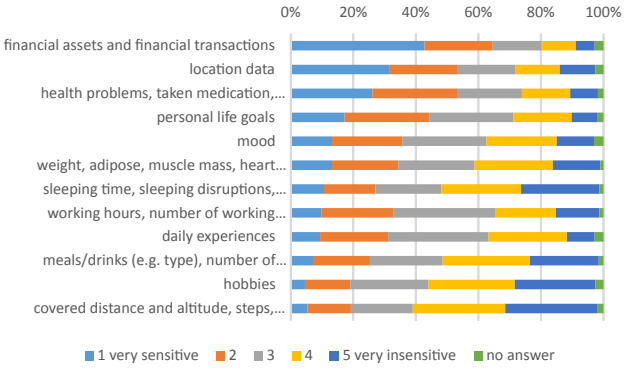

Fig. 2. How sensitive are the following data for you?

To find out the differences concerning the perceived sensitivity of QS data between users and non-users, we performed the Mann-Whitney-U-Test[23]. There is no significant difference between users and non-users concerning location ($p = 0.217$) and financial ($p = 0.491$) data. This highlights the high awareness for the sensitivity of these QS data—among non-users as well as among users. There is a significant difference between users and non-users concerning health data ($p = 0.010$), though. Although both groups perceive health data as sensitive, QS users perceive health data as more sensitive than non-users[24]. In general, the differences in the perceived sensitivity between users and non-users are small.

Apart from the general sensitivity classification, the majority of the respondents is worried[25] about the transfer of health data (75 %), location data (76 %) and financial data (88 %) *under their real identity*. Interestingly, respondents are even worried about the transfer of

21 The questions were derived from the theory of planned behavior [2, 8]; in case of non-users, we asked for reasons they could imagine to start using a QS service.

22 see Appendix A.1 for the non-shortened list of categories

23 We have chosen the Mann-Whitney-U-Test because the data is not parametric.

24 The mean rank of users is 172.87; for non-users, it is 203.95. The median is 2 for both.

25 Cumulative frequency of those who answered "very much worried" and "much worried", at a scale from "very much worried" to "not worried at all" (five-step scale).

their data under a random pseudonym (for health data 51 %, for location data 55 %, for financial data 64 %). Concerning the transfer of these data under the real identity and under a random pseudonym, there is no significant difference between QS users and non-users.[26] Furthermore, there is no significant difference concerning the transfer of health, finance and location data between QS users who use a pseudonym and QS users who do not use a pseudonym.

While this shows that trust in pseudonyms is limited, their use is obviously seen as advantageous by at least some of the users. The obvious question arises whether users themselves use pseudonyms to protect their identities. Our survey results show that 39 % of QS users choose a pseudonym for services with a real name policy, instead of using their real names (12 % did not answer the question; among the non-users, 47 % would use a pseudonym, with 18 % not answering)[27]. Almost all QS service users who use a pseudonym answered the question about the reason for doing so—besides 2 who apparently misunderstood the question and provided their respective pseudonyms. 2 users mentioned usability arguments (like "shorter name is better to use"), 3 provided general reasons like "habit", and the remaining 40 users stated privacy-related reasons—some of them being specific (e. g. "to prevent anyone from linking my different accounts online"), but the majority referring generally to anonymity, privacy, and identity protection. A similar result was obtained for the non-users. Interestingly, 12 out of 57 QS users who stated a reason for *not* using a pseudonym had never thought of the idea; 6 users considered using a pseudonym as "dishonest", did not want to violate the policy, or feared to lose access to the service.

45 % of the respondents are also worried about the *storage of the data by the QS service providers*. Even more respondents are worried about the *consolidation of the data* (56 %) or the *storage of the meta data of the service usage* in addition to QS data (65 %).

The results show that respondents are worried about the use of their data; however, only a minority among the QS service users have read the privacy policy, as shown in Fig. 3. While 47% admitted not to have read the privacy policy, we suspect this is true for a much larger percentage given the number of participants who chose not to answer this question. The discrepancy of users stating to be concerned about privacy but not actually reading policies has been confirmed in previous surveys [22]. Interestingly, 59% of the QS non-users state that they would read the privacy policies before starting to use a QS service. The different results between users and non-users concerning the privacy policy reading behavior can be generalized.[28]

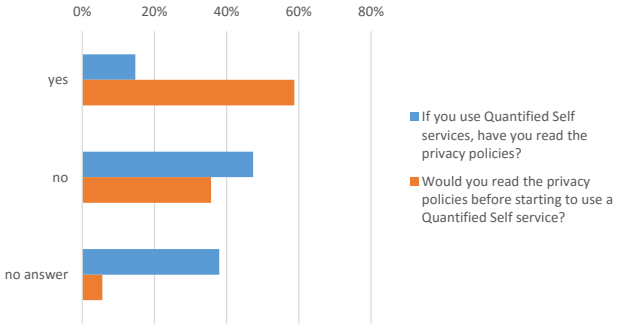

Fig. 3. Portion of QS users reading the services' privacy policies

The survey also gives hints that QS users are concerned about the security of their data. 74 % of the respondents are afraid that their data can be stolen by hackers.

4.3 Opposing Users' View Against Reality

It is interesting to see that users' fears in terms of privacy protection do not necessarily correlate with how they act.

An obvious example concerns the use of data. For example, four of the QS service providers listed in Tab. 2 explicitly state that they use personal data for personalized advertisement. We asked the respective users whether they expect these providers to use their data for that purpose. For each provider, at least half of the users did not expect such behavior. Most of them even stated they would not continue using the respective provider if it used their data for personalized advertisement (see Tab. 5).[29] Similar results were obtained for services which make user data public by default; once

26 Transfer of data type: health 1.) under pseudonym: $p = 0.837$ 2.) under real identity: $p = 0.156$; location 1.) $p = 0.590$, 2.) $p = 0.983$; financial 1.) $p = 0.672$, 2.) $p = 0.091$.

27 The Chi-Square test confirms that the different results concerning the pseudonym usage between QS users and non-users can be generalized ($p = 0.022$)

28 Chi-Square test, $p < 0.001$.

29 In case of Mint, there was a user who both stated he expected Mint to use his data for personalized advertisement, and that he would not continue using Mint in that case.

Service	Users	Expect personalized advertisement		Would continue using service	
		Yes	No	Yes	No
Runtastic	32	10	22	12	20
Foursquare	11	5	6	6	4
Mint	11	5	5	4	6
MoodPanda	4	1	3	1	3

Table 5. Expectations concerning personalized advertisement

Service	Users	Expect data to be public		Would continue using service	
		Yes	No	Yes	No
Runtastic	32	6	26	7	24
Foursquare	11	2	8	4	7
Goodreads	30	8	22	10	17
MoodPanda	4	1	3	1	3

Table 6. Expectations concerning default visibility settings

again, a majority of users did not expect this behavior (see Tab. 6).

The apparent contradiction can be explained by a lack of awareness, which is not surprising: As previously mentioned and illustrated in Fig. 3, at least 47 % of QS users did not read the privacy policies before starting to use the service.

As another example, as shown in Tab. 2, 7 out of 8 QS tools exclusively store the gathered data in the cloud. 14.5% of the survey participants use one of these 7 QS tools; within that group, 37% (compared to 45% when considering all survey participants) are worried about their data being stored by the service provider. Both lack of awareness or a conscious decision to accept the drawback are possible explanations why these persons continue to use the respective services.

4.4 Willingness to Pay for Privacy

Only a minority of respondents are willing to pay a one-off fee for the privacy of their data (i.e. to prevent the non-anonymized transfer of their data to third parties): 30% for the privacy of their health data, 30% and 39% for the privacy of location and financial data respectively.[30] 30% stated that they would pay a fee for a technical solution that anonymizes their data before uploading them to the QS service provider. Surveys and experiments in literature also indicate a low willingness to pay for privacy [16]. A laboratory study by Tsai et al. [36] indicates that prominently displayed privacy information may have a more significant impact on purchasing behavior. With that information, users tend to purchase from online retailers that better protect their privacy. The study even indicates users would be willing to pay a premium to purchase from privacy protective websites. The difference to our own study may be explained by the lack of reliable and accessible privacy information provided by service providers in real-world scenarios (as opposed to the laboratory study).

4.5 Limitations of the User Study

Like all empirical research, the survey must be carefully interpreted, taking into account any potential biases. One limitation of the study concerns the sample. We have conducted an online study. Although online studies are less error-prone in terms of the result analysis, the sample consists of Internet users only and, thus, might exclude some groups of the population and endanger the representativity of the results. However, this risk is considered to be negligible today. Most studies in the technical field are performed online. Furthermore, as shown in Tab. 3 and Tab. 4, diverse groups concerning age and educational background are represented in our study. Another issue might be generalizability. This limitation is due to the number of registered, potential study participants in the survey system that we used. If we had opened the survey for participants from different EU countries, we would not have reached the adequate number of participants for each country that is necessary to achieve representative results—especially in terms of comparability between different countries. We opted for reliable results from one country instead of trying to cover the EU as a whole.

4.6 Other QS User Studies

User studies previously performed by other authors can provide insights in addition to the results discussed in

30 Even fewer users would pay a monthly fee for the privacy of their data: 21% for the privacy of health data, 21% and 30% for the privacy of location and financial data, respectively.

the paper at hand. Zhang et al. [40] investigate influence factors of the decision to disclose data collected by Fitbit, thus contributing to a better understanding of privacy choices and complementing our user study.

The fact that Quantified Self does not only cover personal data about oneself but also has implications for people in the surroundings, is illustrated by Hoyle et al. [19] for the case of wearable cameras that constitute another approach towards lifelogging. They conducted a user study with 36 participants in order to find out which kind of images are considered as sensitive by users: A combination of other people shown on the picture, time and location is considered as sensitive by most participants. As the authors also point out, users would prefer to manage privacy during the process of lifelogging (i.e., while capturing images), and not after the collection has been finished.

5 Towards Privacy-Aware Quantified Self Tools

Our user study has shown that privacy concerns play a role for many users. Still, not all users act accordingly—the majority does not even read the privacy policies. Most users are not willing to pay for better privacy protection.

As a consequence, providers may not gain much from charging their customers for improved privacy. However, the privacy-aware minorities are significant, and there are also other good reasons to invest in privacy-enhancing technologies. Legislation in some regions, such as the European Union, may otherwise restrict the potential usages of gathered data—in particular health-related data. In this context, the concept of "privacy by design and by default" plays a crucial role in the GDPR. Moreover, customers with a low willingness to pay for privacy may still prefer a service provider that protects their data. We therefore suggest requirements for privacy-aware QS services, and discuss some potential approaches to fulfill them.

Suitable privacy requirements are highly dependent on the benefits that should be provided by a specific QS tool (cf. Sect. 2.1), ranging from basic requirements and straightforward solutions for simple services to complex requirements for sophisticated services.

5.1 Reports Based on Individual Data

In the simplest case, if a QS tool aims only at reports based on data collected about an individual user, privacy could be respected by employing an offline solution. Obiously, this assumes that measures against local attacks—like tracking devices with a static Bluetooth address [5, 18], or attacking a body area network [26]—are in place.

If all calculations are performed locally (e.g., by an app running on the user's smartphone), it is completely up to the user to decide whether data are shared with others or not. Although this is rather simple to achieve, none of the QS services we analyzed but Runtastic actually supports this today.

A privacy-respecting solution could, however, also be easily achieved if storage and/or processing of personal data were performed within a cloud backend, e.g., by employing pseudonymization. *Pseudonymity* is achieved if data is mapped to a pseudonym, but not the full (real-world) identity of the user. Achieving pseudonymity does not usually require technical changes to the service, as users can simply enter pseudonyms instead of their real names. Even for fee-based services, persistent links of real names to accounts are not necessary. Prevention of multiple registrations by the same person is harder when using pseudonyms, but it is disputable whether the collection of real names provides an adequate countermeasure. Pseudonymity could, thus, already be achieved with the analyzed QS services in principle, but it is actually prevented in most of the service's terms of use for inexplicable reasons as pointed out in Section 3.2.2.

Regardless, pseudonymity is a weak requirement. In multiple cases, pseudonyms could be successfully mapped to persons using external information [29]. Given the kind of data processed by QS tools, finding such a mapping could be especially simple in our context. A location history of a pseudonymous user, for example, is likely to reveal her place of residence, which significantly lowers the necessary efforts to ascertain her identity.

A much stronger requirement would be *unlinkability*. Unlinkability of values means that they cannot be successfully linked with a sufficient probability—most commonly, we consider the values to be linked if they refer to the same person. As an example, different service providers should not be able to collaborate in order to link different data sets belonging to the same user.

Possibilities to create such a link are communication channels (e.g., a user connecting to different service

providers from the same source IP address) and authentication schemes. Concerning communication, unlinkability of different interactions between users and service providers is offered by anonymity networks like Tor [10]. Concerning authentication and authorization, anonymous credential schemes [7] can be used. Unlinkability for the data themselves can, in principle, be achieved by reducing their precision. While the problem has been investigated for location privacy [17], there is no general, application-independent solution.

At the extreme, unlinkability means that nobody shall be able to tell whether any two data points are linked or not. To this extent, however, the requirement is not suitable for the QS scenario: Most services rely on processing of time series data, and removal of timing information—as is required to provide unlinkability—would render the data useless. In some cases (such as the gathering of location data or pulse), data from consecutive time periods will also be highly correlated, allowing to observe links even if explicit timing information is removed. Assume a system measuring the users' pulse in one-minute intervals, and providing the following data about three users: $t_0 : \{61; 75; 140\}$; $t_0 + 1\,\text{minute} : \{145; 85; 57\}$. Even if any explicit links are removed and no further information is available, a mapping between values from the two time periods could be found with high probability.

To account for that, suitable definitions of unlinkability for the QS scenario still have to be found. The characteristics and concrete requirements of each individual service have to be analyzed with respect to its acceptable and/or required information leakage, and tailored technical solutions have to be developed that enforce these requirements.

As long as the data affect only a single user and no complex computations are required, however, a cloud-backed QS service with better privacy guarantees could be realized much more easily. Any computations could be performed locally by a device in control of the user, and end-to-end encryption could be used to store results and other data within the QS service's cloud backend. Unfortunately, this is implemented by none of the services we analyzed.

5.2 Competition

If competition is to be supported, restricting access to personal data to their owner is not an option anymore. Instead, the goal is to provide *some* information to others to allow comparisons, while maintaining some degree of privacy for the user. More sophisticated approaches are thus required which account for the specific kinds of comparisons that are desired for competition between users. Two dimensions are of special importance here: The group of users within which comparisons should be performed, and the granularity of data involved in comparisons.

If a user wants to compare and share her data with the general public, the data are obviously not considered sensitive, so no privacy-preserving measures are required. Limiting the scope of the comparison to a couple of users, on the other hand, can be realized by employing end-to-end encryption and sharing the corresponding key(s) among the group of legitimate users. For larger group sizes and dynamic group memberships, which we expect for QS services, key management is likely to become a challenge that could be tackled by adaptation of schemes from the related field of secure group communication [38].

An end-to-end encrypted exchange of personal data between legitimate users (e.g., their weights at different points of time) is sufficient as long as a fine-grained comparison of each individual data point is desired. Competition is also possible, though, if users consider these data too sensitive for being shared within a respective group.

If users are, for example, only interested in relations between their data (e.g., who has the lowest weight, whose pulse is below a certain threshold, whose pulse is lower than Alice's, who slimmed down most within a certain time range, ...), there is no need for sharing exact numbers. Instead, users should be able to reveal only limited information suitable for a certain comparison, and nobody else should be able to learn more than the information voluntarily disclosed. Protocols solving Yao's millionaires' problem [39] could be candidates for that purpose, allowing comparisons between values without revealing them.

Similarly, users could be interested in only finding out commonalities (e.g., which users are running the same distance, which users have similar weights, ...). Private set intersection protocols [13] could be used to show results only in the case the respective users have certain data in common, i.e. without revealing further information.

Alternatively, or in addition, privacy needs might also be satisfied by fuzziness or aggregation. For many comparisons, exact numbers might not be required, but it might be sufficient to round values or work with clusters of similar values (e.g., compare only intervals of weights), or to aggregate data collected within cer-

tain time spans (e.g., compare users' monthly average weights). Such mechanisms are already applied in practice, but only with respect to data that are shared with third parties as discussed in Section 3.2.4.

5.3 Correlation

In contrast to the previous benefits, the benefit of correlation involves *multi-dimensional data* and, thus, requires much more complex processing by the QS service. From a privacy perspective, however, this benefit is easy to deal with as only a single user is involved. Essentially, the same requirements and approaches can be established as described under *reports based on individual data*. If QS services offer a method to export individual data in a well-known format, even data collected by different QS services could be brought together for local processing in a privacy-preserving manner. If local processing is not an option for any reason (e.g., due to limited capacity of the devices used by end users), processing of aggregated and/or fuzzy data as described above could be sufficient to acquire new findings.

Some insight may be gained from research in context-aware services [4] in this scenario. As pointed out by Riboni et al. [32], obfuscation techniques for location privacy cannot be applied in a straightforward manner to other context information. Moreover, the authors point out that such additional context information also makes "identity anonymity" difficult to achieve—a similar issue to the processing of multi-dimensional data in a QS scenario. There is, however, a major difference: Research on context-aware services aims at the (in some cases: privacy-preserving) use of current context data to provide a tailored service; QS tools, on the other hand, are more geared towards provision and analysis of data over a longer timespan.

5.4 Research

Allowing research is a bigger challenge: Here, both multi-dimensional data *and* different users are involved. If sensitive data of several users possibly not knowing each other (and thus, possibly not willing to share their data with each other) should be combined to gain insights, purely local processing is no longer feasible.

The already mentioned approaches, however, can also be adapted to this case: Depending on the specific research goal, fuzzy or aggregated data of single users might be sufficient, and pseudonymization might be performed anyway. As before, unlinkability only makes sense to a certain degree, as different data—particularly multi-dimensional data (e.g., pulse and breathing rate)—need to be linked to be able to gain research insights.

In addition, aggregation is not only possible with respect to different data of a single user, but also with respect to different users (e.g., what is the average weight within a certain group of users). This means that even if exact numbers are required, computations are performed in such a way that the QS service performing the computation (and possibly other users) may only learn aggregated information, but nothing about any individual user's value. To be useful, such an aggregation must cover a sufficiently large group, and an attacker must not have prior knowledge about the group's members. Privacy-preserving data aggregation has been tackled for smart meters [27] and wireless sensor networks [14]. Often, the approaches employ homomorphic encryption to perform calculations on encrypted data. In QS services, we expect dynamic group memberships, more frequent measurements, and multi-dimensional data. Moreover, users may want to reveal different kinds of aggregates (e.g., their improvement over time instead of a current fitness level).

6 Conclusion and Outlook

Although quantified self tools are increasingly collecting more personal data of users, storing them within the cloud and starting to consolidate data from different services, the number of users is ever increasing. The privacy policies of the service providers that we analyzed allow the usage of data for diverse purposes. In the survey we conducted, we found out that users say they care about their private QS data—at least when it comes to data relating to health and finances—and that they are afraid of it being shared with third parties. Only a minority of the respondents would pay for QS services that offer better protection of their personal data from unwanted transfer, though.

Concerning future work, we are investigating applications of privacy-enhancing technologies, and are also analyzing the economic aspects of privacy protection for QS services.

To conclude, it is noteworthy that only little research has yet been conducted in terms of privacy protection in this field—even though it deals with sensitive, personal data. It will be interesting to see whether privacy-preserving QS tools—which do not yet exist—might be better accepted by users in the future.

References

[1] adidas Group. Adidas Group acquires Runtastic. http://www.adidas-group.com/en/media/news-archive/press-releases/2015/adidas-group-acquires-runtastic/, Aug. 2015. Press Release.

[2] I. Ajzen. The theory of planned behavior. *Organizational Behavior and Human Decision Processes*, 50(2):179–211, 1991. Theories of Cognitive Self-Regulation.

[3] Apple Inc. Developer: HealthKit. Webpage. https://developer.apple.com/healthkit/.

[4] M. Baldauf, S. Dustdar, and F. Rosenberg. A survey on context-aware systems. *International Journal of Ad Hoc and Ubiquitous Computing*, 2(4):263–277, 2007.

[5] M. B. Barcena, C. Wueest, and H. Lau. How safe is your quantified self? Technical report, Symantec, Aug. 2014. Version 1.1.

[6] M. Borghi, F. Ferretti, and S. Karapapa. Online data processing consent under EU law: a theoretical framework and empirical evidence from the UK. *International Journal of Law and Information Technology*, 21(2):109–153, 2013.

[7] J. Camenisch and A. Lysyanskaya. Signature Schemes and Anonymous Credentials from Bilinear Maps. In M. K. Franklin, editor, *Advances in Cryptology - CRYPTO 2004, 24th Annual International Cryptology Conference, Santa Barbara, California, USA, August 15-19, 2004, Proceedings*, volume 3152 of *Lecture Notes in Computer Science*, pages 56–72. Springer, 2004.

[8] M. Conner and C. J. Armitage. Extending the Theory of Planned Behavior: A Review and Avenues for Further Research. *Journal of Applied Social Psychology*, 28(15):1429–1464, 1998.

[9] A. Daly. The law and ethics of 'self-quantified' health information: an australian perspective. *International Data Privacy Law*, 5(2):144–155, 2015.

[10] R. Dingledine, N. Mathewson, and P. F. Syverson. Tor: The Second-Generation Onion Router. In M. Blaze, editor, *Proceedings of the 13th USENIX Security Symposium, August 9-13, 2004, San Diego, CA, USA*, pages 303–320. USENIX, 2004.

[11] Eugene Mandel. How the Napa Earthquake affected Bay Area sleepers. Webpage, August 2014. https://jawbone.com/blog/napa-earthquake-effect-on-sleep/.

[12] Flurry. Health and Fitness Apps Finally Take Off, Fueled by Fitness Fanatics. Webpage, June 2014. http://flurrymobile.tumblr.com/post/115192181465/health-and-fitness-apps-finally-take-off-fueled.

[13] M. Freedman, K. Nissim, and B. Pinkas. Efficient private matching and set intersection. In C. Cachin and J. Camenisch, editors, *Advances in Cryptology - EUROCRYPT 2004*, volume 3027 of *Lecture Notes in Computer Science*, pages 1–19. Springer Berlin Heidelberg, 2004.

[14] J. Girao, D. Westhoff, and M. Schneider. CDA: concealed data aggregation for reverse multicast traffic in wireless sensor networks. In *2005 IEEE International Conference on Communications, 2005*, volume 5, pages 3044–3049 Vol. 5, May 2005.

[15] P. Gola, C. Klug, B. Körffer, and R. Schomerus. *BDSG: Bundesdatenschutzgesetz: Kommentar*. C.H.Beck, 11th edition, 2012.

[16] J. Grossklags and A. Acquisti. When 25 Cents is Too Much: An Experiment on Willingness-To-Sell and Willingness-To-Protect Personal Information. In *6th Annual Workshop on the Economics of Information Security, WEIS 2007*, 2007.

[17] M. Gruteser and B. Hoh. On the Anonymity of Periodic Location Samples. In D. Hutter and M. Ullmann, editors, *Security in Pervasive Computing*, volume 3450 of *Lecture Notes in Computer Science*, pages 179–192. Springer Berlin Heidelberg, 2005.

[18] A. Hilts, C. Parsons, and J. Knockel. Every step you fake – a comparative analysis of fitness tracker privacy and security. Technical report, Open Effect, 2016. Version 0.3.

[19] R. Hoyle, R. Templeman, S. Armes, D. Anthony, D. Crandall, and A. Kapadia. Privacy behaviors of lifeloggers using wearable cameras. In *Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing*, UbiComp '14, pages 571–582, New York, NY, USA, 2014. ACM.

[20] M. Humbert, E. Ayday, J.-P. Hubaux, and A. Telenti. Addressing the concerns of the lacks family: Quantification of kin genomic privacy. In *Proceedings of the 2013 ACM SIGSAC Conference on Computer & Communications Security*, CCS '13, pages 1141–1152, New York, NY, USA, 2013. ACM.

[21] C. Jensen and C. Potts. Privacy Policies As Decision-making Tools: An Evaluation of Online Privacy Notices. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '04, pages 471–478, New York, NY, USA, 2004. ACM.

[22] C. Jensen, C. Potts, and C. Jensen. Privacy practices of internet users: Self-reports versus observed behavior. *International Journal of Human-Computer Studies*, 63(1–2):203–227, 2005. {HCI} research in privacy and security.

[23] J. Kahn. RunKeeper, Withings, Strava, & iHealth plan HealthKit integration, excited for medical industry tie-in. Webpage, June 2014. http://9to5mac.com/2014/06/04/runkeeper-withings-strava-ihealth-plan-healthkit-integration-excited-for-medical-industry-tie-in/.

[24] J. Kaye. Abandoning informed consent. In O. Corrigan and R. Tutton, editors, *Genetic Databases: Socio-Ethical Issues in the Collection and Use of DNA*. Routledge, Abingdon, 2004.

[25] A. D. I. Kramer, J. E. Guillory, and J. T. Hancock. Experimental evidence of massive-scale emotional contagion through social networks. *Proceedings of the National Academy of Sciences*, 111(24):8788–8790, 2014.

[26] B. Latré, B. Braem, I. Moerman, C. Blondia, and P. Demeester. A survey on wireless body area networks. *Wirel. Netw.*, 17(1):1–18, Jan. 2011.

[27] F. G. Martinez Perez, C. Sorge, R. Petrlic, O. Ugus, D. Westhoff, and Gregorio. Privacy Enhanced Architecture for Smart Metering. *International Journal of Information Security*, 12(2):67–82, 2013.

[28] meetup.com. Quantified Self Meetups. http://www.meetup.com/en-US/topics/quantified-self/all/.

[29] A. Narayanan and V. Shmatikov. Robust De-anonymization of Large Sparse Datasets. In *Proc. IEEE Symposium on Security and Privacy (SP 2008)*, pages 111–125, May 2008.

[30] quantifiedself.com. Guide to Self-Tracking Tools. http://quantifiedself.com/guide/.
[31] Reuters. Google unveils "Fit" health, fitness tracking platform. http://www.reuters.com/article/2014/06/25/google-healthcare-idUSL2N0P61N820140625.
[32] D. Riboni, L. Pareschi, and C. Bettini. Privacy in georeferenced context-aware services: A survey. In C. Bettini, S. Jajodia, P. Samarati, and X. Wang, editors, *Privacy in Location-Based Applications*, volume 5599 of *Lecture Notes in Computer Science*, pages 151–172. Springer Berlin Heidelberg, 2009.
[33] runtastic GmbH. Facts About Runtastic. Available at https://www.runtastic.com/mediacenter/corporate-assets/english/company-overview/facts-about-runtastic_en_may2016.pdf, May 2016.
[34] Samsung. Intelligence for smarter health. Webpage. http://www.samsung.com/us/ssic/innovation_areas/#digital-health.
[35] P. M. Schwartz. The eu-u.s. privacy collision: A turn to institutions and procedures. *Harvard Law Review*, 126:1966–2009, 2013.
[36] J. Y. Tsai, S. Egelman, L. Cranor, and A. Acquisti. The effect of online privacy information on purchasing behavior: An experimental study. *Info. Sys. Research*, 22(2):254–268, June 2011.
[37] G. Wolf. What is The Quantified Self. http://quantifiedself.com/2011/03/what-is-the-quantified-self/, Mar. 2011.
[38] C. K. Wong, M. Gouda, and S. S. Lam. Secure group communications using key graphs. In *Proceedings of the ACM SIGCOMM '98 Conference on Applications, Technologies, Architectures, and Protocols for Computer Communication*, SIGCOMM '98, pages 68–79, New York, NY, USA, 1998. ACM.
[39] A. C. Yao. Protocols for Secure Computations. In *Proceedings of the 23rd Annual Symposium on Foundations of Computer Science*, pages 160–164, Washington, DC, USA, 1982. IEEE Computer Society.
[40] J. Zhang, V. Dibia, A. Sodnomov, and P. B. Lowry. Understanding the disclosure of private healthcare information within online quantified self 2.0 platforms. In *19th Pacific Asia Conference on Information Systems, PACIS 2015, Singapore, July 5-9, 2015*, page 140, 2015.

A QS Survey

A.1 Types of Sensitive Data

The complete list of data categories as used, e.g., in Tab. 2 is as follows:

- weight, adipose, muscle mass, heart frequency, body mass index
- covered distance and altitude, steps, burnt calories, etc.
- sleeping time, sleeping disruptions, movements during the sleep, sleeping quality, snoring, etc.
- meals/drinks (e.g. type), number of chews, calories
- financial assets and financial transactions
- working hours, number of working disruptions, completed tasks, etc.
- mood
- daily experiences
- personal life goals
- health problems, taken medication, allergies, laboratory results, etc.
- location data
- hobbies

A.2 Questionnaire

Our survey was performed using the LimeSurvey[31] survey system. Below is a compact version of the full questionnaire we used. Question groups of the questionnaire are shown as paragraphs and questions are displayed as items. Answer options are given after "→": if separated by comma, we allowed multiple answers; if separated by ⊕, the user had to decide on a single option. The answer option "other" was always combined with a freetext answer field.

Foreword

What is Quantified Self?
Quantified Self is a self-monitoring technology that allows to track person's daily life in terms of inputs (e.g. food consumed, sleeping hours, covered distance), states (mood, blood oxygen levels) and performance (physical and mental).

You are kindly asked to fill in the questionnaire.

Main Part

Before you start...
1. Please enter your Prolific ID
   → *freetext*

31 https://www.limesurvey.org/

QS usage in general

2. Do you use Quantified Self services?
   → Yes ⊕ No

Questions for QS users

only asked if answer for question 2 was "yes"

3. Which QS services do you use?
   → *services from Tab. 2*, other
4. What are your reasons for using QS?
   → *reasons from Fig. 1*, other

Questions for QS nonusers

only asked if answer for question 2 was "no"

5. What are your reasons for not using QS services?
   → I don't know any QS services, I am just not interested, I am worried that they use my data for advertising, I am worried that they forward my data to uninvolved third parties, other
6. Which reasons could you imagine to start using Quantified Self Services?
   → *reasons from Fig. 1*, other

Categories of collected data

7. How sensitive are the following data for you?
   → ∀ *options from Fig. 2:* 1 = very sensitive ⊕ 2 ⊕ 3 ⊕ 4 ⊕ 5 = very insensitive
8. How much are you worried about the transfer of your data to third parties (e.g. insurance companies, employers or business companies) if a random pseudonym (like "user 1234") is transferred instead of your real name?
   → ∀ *options from Fig. 2:* 1 = very much ⊕ 2 ⊕ 3 ⊕ 4 ⊕ 5 = not at all
9. How much are you worried about the transfer of your data to third parties (insurance companies, employers or business companies) if they are transferred under your real identity?
   → ∀ *options from Fig. 2:* 1 = very much ⊕ 2 ⊕ 3 ⊕ 4 ⊕ 5 = not at all

Privacy practices

10. How much are you worried that...
    – ...your data is stored by the Quantified Self Service Providers?
    – ...your data is consolidated? (e.g., health data + financial data)
    – ...meta data of your service usage (e.g., location information, IP addresses, time of access) is stored additionally to your Quantified Self data?
    – ...your data could be stolen by hackers?
    → ∀ *subquestions above:* 1 = very much ⊕ 2 ⊕ 3 ⊕ 4 ⊕ 5 = not at all
11. ∀ *service* ∈ *Tab. 2* ∩ *user's choices in question 3:*
    Privacy practices of *service*
    – Would you expect your data stored by *service* to be public by default?
    – Would you continue to use *service* if your data were public by default?
    – Would you expect *service* to forward your (non-anonymized) personal data to uninvolved third parties?
    – Would you continue to use *service* if it forwarded your (non-anonymized) personal data to uninvolved third parties?
    – Would you expect your data stored by *service* to be used for personalized advertisement?
    – Would you continue to use *service* if it used your data for personalized advertisement?
    → ∀ *subquestions above:* Yes ⊕ No

Pseudonyms

12. Do you use pseudonyms even when the Quantified Self service has a real-name policy?
    only asked if answer for question 2 was "yes"
    → Yes ⊕ No
13. Why do you use pseudonyms?
    only asked if answer for question 12 was "yes"
    → *freetext*
14. Why don't you use pseudonyms?
    only asked if answer for question 12 was "no"
    → *freetext*
15. If you used QS services, would you use pseudonyms even when the Quantified Self service has a real-name policy?
    only asked if answer for question 2 was "no"
    → Yes ⊕ No
16. Why would you use pseudonyms?
    only asked if answer for question 15 was "yes"
    → *freetext*
17. Why wouldn't you use pseudonyms?
    only asked if answer for question 15 was "no"
    → *freetext*

Willingness to pay

18. Would you pay a one-off fee in order to prevent the (non-anonymized) transfer of the following data to third parties (insurances, employer and business companies)?
    → ∀ *options from Fig. 2:* Yes ⊕ No
19. Would you pay a monthly fee in order to prevent the (non-anonymized) transfer of the following data to third parties (insurance companies, employers and business companies)?
    → ∀ *options from Fig. 2:* Yes ⊕ No

Opinion

20. Please give your opinion.
    - Would you pay a one-off fee in order to prevent the transfer of your data to third parties (e.g. insurance companies, employers and business companies)?
    - Would you pay a monthly fee in order to prevent the transfer of your data to third parties (e.g. insurance companies, employers and business companies)?
    - Would you agree to the transfer of your data to third parties (e.g. insurance companies, employers and business companies) if you got the Quantified Self service for free?
    - If you want to share your Quantified Self data with others – would you like to choose those people on your own?
    - Do you use the function for indicating the person you want to share your data with?
    - Would you agree to transfer your financial data to third parties for a reduction of insurance rate?
    - Would you agree to transfer your health data to third parties in order to contribute to prevent epidemics?
    - Would you use a Quantified Self service that finances itself by transferring your data to third parties (e.g. business companies)?
    - If you use Quantified Self services, have you read the privacy policies?
    - Would you read the privacy policies before starting to use a Quantified Self service?
    - Would you pay a fee for a technical solution that anonymizes your data before transferring it to the Quantified Self service provider?

    → ∀ *subquestions above:* Yes ⊕ No

Closing questions

21. Please indicate your age.
    → *options from Tab. 4*
22. Please specify your nationality.
    → *freetext*
23. Please specify your education.
    → *options from Tab. 3*
24. Please specify your gender.
    → Female ⊕ Male