Data collection using new information technology

Recommendations on data quality, data management, research ethics, and data protection

German Data Forum (RatSWD)

Data collection using new information technology

Recommendations on data quality, data management, research ethics, and data protection

Contents

Executive Summary

■ Data collection with smartphones, wearables, and other sensors – or simply data collection using new information technology – offers enormous potential for researchers. Sensors can gather data in people's everyday lives and in real-time, including location, movement, noises, light, media use, videos, and voice recordings. They also facilitate the repetition of surveys. In this report, the German Data Forum (RatSWD) outlines a quality assurance framework for using new information technology for research purposes.

The report covers three main aspects:

1) The quality of sensor-generated data

Researchers should note that most sensors are created as consumer products and their data quality does not meet the scientific standards of reliability and validity per se. Researchers should therefore critically assess and document the data collection process, measurement accuracy, and the validity of their inferences. Producers of sensors should ensure transparency on measurement accuracy and archive the algorithms used by their sensors for processing data.

2) Issues of research ethics and data protection

Due to the high temporal and spatial resolution of the mostly personal data collected with new information technology, researchers should take appropriate measures to ensure the anonymity of study participants, e.g., pseudonymisation strategies, coarsening data, distorting voices, and adhering to standards of data minimisation. They should also safeguard privacy (e.g., by giving participants the opportunity to edit and delete data). When researchers collect data from third parties or document illegal behaviour of study participants, they must make sure to handle the data in an appropriate way and in accordance with the law. Participants (and third parties) should be informed about the purpose of data collection and the related risks, e.g., in the informed consent form.

3) Data management

These new data collection methods often result in large volumes of (raw) data containing sensitive information. This creates new challenges in storing, preparing, and making data available for secondary use, as well as in managing metadata. Here, too, it is important to adhere to and to implement data protection regulations and standards of research ethics.

The report is addressed to:

- researchers and students who are already using these new methods or plan to use them in the future,
- ethics commissions and research funding agencies assessing relevant projects, and
- the general public, policymakers, and companies using new technology (possibly in cooperation with science and research).

1 Introduction

The development of new information technology in the past years has been characterized by digital mobility, miniaturisation, and the increasing use of such technology by the general population. These developments also affect how the social, behavioural, and economic sciences collect research data. An example of this is **experience sampling**, a method involving self-reporting of thoughts, emotions, and/or environments over time, using smartphone-based questionnaires. Another example of these new methods is **mobile sensing**, which can include capturing location, movement, noises, light, and smartphone usage by means of smartphone sensors. Gathering data using wearables has also increased. Wearables use sensors to capture physiological parameters and activity without using a smartphone. Last but not least, **ambient monitoring** uses sensors integrated into the environment to capture movement, sound, images, and other parameters.

For decades, gathering data with such information technology was a niche topic in the social, behavioural, and economic sciences (Fahrenberg and Myrtek 1996). However, the advent of the smartphone has resulted in a stark increase of these data collection methods and their use in various fields of empirical research (Harari et al. 2016; Yetisen et al. 2018). Due to their numerous benefits compared to previous data collection methods, these methods have the potential to fundamentally change the way we do research, as well as the way we apply these devices in other fields (e.g., diagnosing how people experience and behave in everyday life; using them for interventions in the health sector).

Firstly, these methods facilitate the measurement of certain attributes in everyday, real-life situations. Compared to classic laboratory research, capturing phenomena of interest in individuals' everyday lives increases ecological validity. It helps to generalise research findings to real-life settings, which, after all, is the context closest to people. This has an extraordinary significance for empirical research, because many research findings are poorly transferrable. This is exemplified by the so-called doctor's office or white coat hypertension (Salles/Cardoso/Muxfeldt 2008; Stergiou et al. 2018), the phenomenon that people exhibit a higher blood pressure when measuring it in a doctor's office setting, which can result in unnecessary prescription of antihypertensive drugs. This situational context is relevant for many research issues, because context affects the behaviour of individuals and groups in a multitude of ways. (Mischel and Shoda 1995). Mobile sensor technology makes it possible to better understand the effect of context on behaviour (e.g., by gathering data on location, social and physical environments) (Böhmer et al. 2011).

Secondly, new information technology allows to capture data in real-time (Trull and Ebner-Priemer 2013). Relevant episodes, which can be fleeting events embedded in a larger context, can be identified based on objective parameters such as movement and localisation (depending on the research question, e.g., entering a mall or polling station, the use of public transportation, or social interactions). In such situations, smartphones can automatically send out specific inquiries in real-time. Capturing data in real-time (e.g., using electronic diaries) can minimise recall biases. Recall biases refer to the phenomenon that memory is an active process and prone to systematic distortion (Schwarz 2012).

A **third** advantage of collecting data using new information technology is that it is objective and multimodal. The development of wearables thus contributes to the objective measurement of interesting attributes in the social, behavioural, and economic sciences. An example of this is the measurement of physical activity using accelerometers. While miniaturisation and mass dissemination of such sensors (e.g., in cars, cameras, etc.) has given way to objective and cost-efficient data collection, they also facilitate a precision in measuring these attributes that was not previously possible using subjective survey methods (e.g., questionnaires) (Prince et al. 2008). Moreover, wearables facilitate the long-term monitoring of attributes of interest.

Fourth, the use of new information technology also makes it possible to repeatedly collect individual data. This improves specifically the way we capture intra-individual variability, changes, and processes. Research on intra-individual variability has greatly increased in recent years, particularly in psychology (Diehl/Hooker/Sliwinski 2015). Data collection using new information technology offers very promising prospects for this field.

Smartphones – a ubiquitous instrument for technology-driven research data collection

Smartphones are equipped with an array of sensors that can be applied in research in the social, behavioural, and economic sciences. They include:

Localisation technology such as the **Global Positioning System** (GPS; to record a person's whereabouts, for example, to investigate the influence of context on individual behaviour)

Cameras (for taking pictures of a person's context, or to perform physiological measurements, for example, measurement of the heart rate on a finger using fluoroscopy)

Log files for documenting smartphone activity (phone calls and the use of instant messaging to assess social activity, or assessing sleep and waking hours based on smartphone off time)

Microphones (for capturing traffic noise, social interactions, etc.)

Accelerometers (for assessing physical activity)

Using data collected with new information technology for research purposes also creates a range of challenges. In the following chapter of this report, we discuss these challenges. Then, we derive recommendations on how to assess the quality of sensor-based data capturing methods and how to verify and secure the validity of conclusions based on these data. Furthermore, we will show how to meet the demands of research ethics and data protection and how to approach data management.

By doing so, this publication serves as an initial guide for researchers and students who are already using or planning to use new information technology in research projects. It also aims at helping ethics commissions and research funding agencies to evaluate such projects. Lastly, the report addresses target groups outside of (academic) research, including the general public, policymakers, and start-ups using new technology (possibly in cooperation with science and research).

This publication focuses on the collection of data from individuals (individual assessment) for scientific studies using mobile information technology, such as wearables, smartphones, and other mobile sensors. Other interesting issues that are not covered by this publication, include: 1) the possibilities for analysing such data (incl. big data technology, see also RatSWD 2020), 2) virtual reality, 3) data that can be collected in the context of the internet of things, and 4) other data collection technology that is similar to laboratory situations. Independent of the new information technology discussed in this report, recording video, image, and voice data has traditionally been a tool used in qualitative-interpretative social research. Here, the technical preparation and analysis of (now mostly digital) text, video, and image data takes a back seat to hermeneutic approaches and other forms of (text-) interpretative analysis – often integrated into more complex research designs. On a structural level, the forms of scientific use and processing methods for these types of material have hardly been changed by the increased use of new recording technology (smartphone replacing the cassette recorder; the webcam replacing the VHS tape). In this very heterogeneous field, researchers apply qualitative social research's own quality criteria, which are based on this interpretative paradigm (Strübing et al. 2018). These are not part of this report.

2 Challenges of data collection using new information technology

■ Data collection using modern information technology opens new paths for empirical research. However, it also creates specific challenges. These include data gathering, the reliability of measurements as well as the validity of conclusions based on these data. Moreover, it creates specific demands to research ethics, data protection, and data management.

Studies based on new information technology: exemplary results and challenges
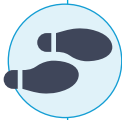
Example 1:
Althoff and colleagues (2017) used a mobile sensing method to access the software of a smartphone's step counter app, collecting acceleration data of over 700,000 individuals in over 100 countries (number of tracked days: 68 million). This enabled the authors to highlight differences in activity between countries and predict values for obesity. Methodically it was impressive that the study tracked such a large number of individuals and days as well as the fact that the researchers could draw on an existing dataset from the consumer sector. Some challenging questions raised regarding the quality criteria applied to these data: did the subjects carry the smartphone all of the time or were there uncaptured movements, because, for example, the smartphone was lying on the desk all day? How is the data quality? Did the app provide the data in a raw format or as aggregated steps per day, processed by an unknown algorithm? The study also raises research ethical and privacy-related questions on using such data: can these data be used for research purposes given that they were originally collected for a different purpose?

Example 2:
A study authored by Tost et al. (2019) used the experience sampling method to demonstrate context dependency and real-time analysis. While pursuing their regular everyday activities for one week, the study participants had their location tracked in real-time via GPS, which prompted location-based diary requests. The study found that its subjects displayed higher subjective well-being in the vicinity of green areas (e.g., forests, parks, and other larger green spaces). The methodological and legal challenges of this study include data protection issues related to location tracking, and how to make the raw data available for replication studies.

Example 3:
Brown et al. (2014) conducted a study on cognitive processes with over 20,000 respondents using a smartphone app. The app contained small games replicating four classic experiments. The underlying question was to what extent the smartphone-based experiments led to similar conclusions as comparable laboratory experiments. It was shown that the smartphone is a viable survey instrument and that the method can help obtain large samples. Measurement reliability is a crucial challenge of such studies, since respondents use the app without supervision and with different hard- and software. Moreover, out of the 40,000 people who had downloaded the app, only 20,000 took the cognitive tests. Here, a logical step would be to investigate the reasons and conditions for this lack of participation to better assess a possible selection bias in the results.

Example 4:
Harari et al. (2019) used mobile sensor technology to compare differences in sociability among 926 young adults. To do so, they used smartphone data on conversations, including calls, messages, and social media communication, and contextualised them with self-reported personality traits. The results of the study supported previous findings which showed, to a relevant degree, that sociability depends on personality. An advantage of collecting data using smartphones is the real-time availability of data. However, since the data were collected from students using a certain operating system, possible selection bias should be critically examined. Additionally, inaccuracies when capturing conversations using voice recordings are also possible, because the app in question could not distinguish between conversations on TV and actual conversations of the participants.

2.1 Data collection, reliability, and validity

Smartphones, wearables, and other sensors were originally developed and optimised as consumer products and not for research purposes in the narrower sense. To date, investigations of the measurement quality and validity of such technology have rarely been carried out (Peake/Kerr/Sullivan 2018; Shcherbina et al. 2017). Smartphones as well as their built-in sensors are continuously developed and perform regular software updates that could put into question whether results from previous validation studies are transferable to other contexts. Device manufacturing companies rarely disclose the specifics of software changes as well as the algorithms used to obtain specific values (Evenson/Goto/Furberg 2015).

Collecting data using mobile information technology aims at making claims on how people experience and behave in their everyday life. The technical requirements can lead to specific distortions during data collection, limiting the validity of the research findings. To cite one example, data collection cannot be carried out at any time and place, because signals cannot be received (e.g., limited GPS in buildings or limited microphone recordings due to a pocketed phone; see Harari et al. 2019). Also, people cannot always supply information (e.g., self-reporting data while driving a car, in professional meetings, or in situations that require privacy). Physiological measurements can be ambiguous and often require additional information on a measurement's situational context for correct interpretations (Schmidt et al. 2018). Regarding the group of individuals that a study aims at investigating, selection bias can also be a result of differences in information technology competence among the respondents.

2.2 Research ethics and data protection

Sensors can capture a range of different attributes in high temporal and spatial resolution, which, in some combinations, make the identification of a respondent very likely. This results in a range of issues relating to research ethics and data protection: how can the anonymity of respondents be ensured? How must data be stored to prevent the danger of deanonymisation? How can people be enabled to make an informed decision about which data to make available? Which requirements result from the EU General Data Protection Regulation (GDPR), particularly concerning Article 35 on data protection impact assessments? How (extensively) should an informed consent form be designed, also in light of the willingness to participate in the studies? What are the consequences of documenting criminally liable behaviour?

Voice and image recordings often collect data on persons and objects other than those who consent to being studied. This results in additional challenges in terms of research ethics and data protection: can the data of bystanders be collected without specifically informing them about a study's goals and obtaining their informed consent? How can privacy and anonymity of bystanders be protected?

2.3 Data management

Sensors provide a multitude of data points that can create specific challenges for data storage and, with that, to data management (e.g., data storage, data preparation, data provision for secondary use, metadata). Moreover, we observe a paradigm shift towards open science. This has created new research approaches aimed at making data that individuals collect in their everyday life through, e.g., wearables or smartphones, available to researchers on online platforms. In addition, data are increasingly made available for re-analysis. These trends connect to challenges to safeguarding data privacy when storing data and making them available through repositories of research facilities and research data centres.

3 Quality of sensor-based data collection methods

■ The following chapter examines issues and recommendations relating to the quality of sensor-based data collection methods with a focus on three aspects: 1) data collection, 2) reliability, and 3) validity. Since these aspects are interrelated to some extent, we cross-reference between sections.

3.1 Data collection

The use of wearables and smartphones in social, behavioural, and economic research is challenged by the fact that most soft- and hardware products are produced as consumer products. Scientific measurement accuracy often plays a minor role compared to primary requirements such as design, price, and user-friendliness. Consequently, the consumer sector is replete with products that are of questionable use for research purposes. These include, for example, the mobile recording of heart rate variability.

Measuring heart rate variability

Heart rate variability in a certain frequency range is a good indicator for the stimulation of the parasympathetic nervous system. It is determined by measuring the milliseconds between two consecutive R-peaks in an ECG (electrocardiogram) and analysing the difference. R-peaks are the ECG's most dominant peaks, indicating a heart muscle's maximum excitation. Consumer products, such as specific watches and wristbands, use the wrist-pulse wave to estimate heart rate variability. Unlike the R-peak of an ECG, the pulse wave is a flat signal, which makes determining its peaks highly inaccurate. Capturing the pulsating signal is subject to a variety of influences, e.g., changes in sensor position by moving the arms and hands, contraction of hand musculature, changes in vasoconstriction (e.g., by exposure to cold), or venous return. These interferences are even stronger when looking at mental stress or athletic activity (Georgiou et al. 2018). Accordingly, it is impossible to precisely measure heart rate variability, in a scientific way, using the pulse wave. Still, there are countless products that offer heart rate variability measurement. Measurement accuracy in a scientific sense is currently not ensured by these products.

A second problem is the widespread (pre-)preparation of raw data performed by such products. Raw data, here, refer to primary data that are directly collected by the device and are not processed in any way. Some consumer products, including so-called activity trackers, do not store any raw data to save storage space. In order to determine the derived attributes (e.g., steps taken), the high-resolution acceleration values are filtered through algorithms, integrated, and aggregated at a minute level. Typically, sensor manufacturers do not publish their data processing methods, e.g., the applied algorithm for derived data (that are used, e.g., for localisation or estimated calorie consumption), because they are considered trade secrets (Evenson/Goto/Furberg 2015). For this reason, it is hardly possible to assess these algorithms, or to ascertain the conditions under which the devices, or algorithms, function correctly. Heart rate variability is also a good example in this context. Chest straps equipped with ECGs, commonly used for sports, flatten out the heart rate graph to eliminate large, unlikely spikes. Incidentally, this also reduces the measured heart rate variability. However, when the central interest of a scientific study is exactly measuring high heart rate variability, such algorithms distort the central variable of interest. Moreover, it is often unclear which of a wearable's sensors is used to measure which attribute. Attributes are often calculated using measurements of several sensors. It can also be unclear whether periods of strong movement are automatically

excluded from the calculation of heart rate variability, i.e., whether the values generated by the accelerometer are also fed into the ECG signal processing.

A third problem of using consumer products for research is the short product cycles of their hard- and software. Consumer products are rarely on the market for longer than 12 months. The most recent products are usually not scientifically validated. Additionally, regular software updates can invalidate the comparability of hardware used by a study over time, particularly when these updates change the software's measurement method. Moreover, companies aim at extending the battery life of their products by using more energy-efficient data collection methods and algorithms, i.e., reducing the frequency of data collection (sampling frequency) following a software update. This also affects measurement accuracy. Some devices reduce the sampling frequency based on remaining battery life, i.e., reducing the sampling frequency when the battery is low and, in this way, the scientific measurement accuracy.

Moreover, some devices adjust the sampling frequency to the stimulus material. Google Clips (a small portable camera), for example, only records pre-defined 'interesting' situations (e.g., animals or children). When using this or a similar technology for research, it is important to document the criteria by which the (automated) data collection is triggered, if this information is available at all.

It is therefore essential to document the conditions during a study (e.g., which device, filter, sampling rate, algorithm, etc. were used) in a data sheet (see also the recommendations of the German Council for Scientific Information Infrastructures (RfII) 2019, 32, 84). This is particularly important when using data from sensors on different aggregation levels, depending on whether these come in the form of raw data (the unfiltered acceleration signal) or derived data (e.g., localisation, calorie consumption). Moreover, contextual information (e.g., annotations on processes and activities when recording movement to train and evaluate algorithms) is of vital importance for interpreting and understanding the collected data. Detailed documentation of the data collection process is also necessary to facilitate the re-analysis of data that are made available in repositories.

Recommendations on data collection

1) Data collection requires comprehensive **documentation of the data collection process**. Documentation of data collection using new information technology should contain information on the following aspects:

   - All **sensors and software used** (manufacturer, type, production year, software version)
   - **Raw data** that were collected and stored
   - All the **data processing steps** leading to the derived data
   - **Contextual information** (e.g., the situation in which the data were collected)

2) Researchers should **weigh the benefits of consumer products against more expensive, scientific products**. Measuring attributes of interest without access to raw data or knowledge of the product's signal processing is hard to justify in a scientific context.

3) Sensor data should be recorded as **raw data in a standard format** and in their original (native) resolution. The use of lossless types of compression is advisable.

4) If the device already processes sensor data into higher-quality information, it is important to know the way it does so. If **data analytics procedures** such as machine learning are employed for such processes, training data and the respective model should be documented, and programme codes and programme versions should be provided.

5) Manufacturers should **archive the documentation of all their sensors and data processing algorithms for all versions of the sensors**. However, this is unlikely with common consumer products. It is therefore advisable to archive the data sheets when buying a product. Researchers should also try to document hard- and software updates.

3.2 Reliability

Reliability refers to the precision of a measurement and shows to what extent it is not prone to the influence of measurement errors. Reliability is therefore a central quality criterion for evaluating data collection methods (Eid and Schmidt 2014). Unlike design, costs, and user-friendliness, reliability is not a priority when developing smartphones and wearables. In contrast to medical products, it cannot be assumed, without further verification, that a wearable reliably measures the attributes of interest. Studies on a device's precision and on the extent of the impact of measurement errors are rare (Evenson/Goto/Furberg 2015; Shcherbina et al. 2017). However, it is very important to assess the reliability of measurements when evaluating research results as well as when diagnostic decisions with respect to individuals are made.

In principle, it is possible to apply tried-and-tested procedures for determining reliability of data collected using new information technology. Evenson/Goto/Furberg (2015) distinguish between 'interdevice reliability' and 'intradevice reliability' in this context. Interdevice reliability is tested by wearing multiple devices of the same production type (the same manufacturer) at the same time and determining the consistency of the measurements. For example, convergence of measurements can be tested by simultaneously wearing multiple fitness wristbands or using multiple location trackers at the same time. This method is equivalent to the parallel-forms method used to determine reliability (parallel-forms reliability; Eid and Schmidt 2014). The method used to determine intradevice reliability is based on determining measurement quality through repeated testing using the same device (test-retest method; ibid.). However, this method has its limitations regarding wearables, because the recording of dynamic processes in everyday life produces fluctuations over time that not only represent the (unsystematic) error variance but also the attribute of interest (systematic fluctuations). For this reason, reliability cannot be determined based on repeated measurements using simple statistical methods, e.g., correlation of measurements, but requires more complex statistical methods. According to Evenson/Goto/Furberg (2015), for example, this is one reason why there are no studies on intradevice reliability of activity trackers.

It is important to note that audio, video, and text data are frequently coded for further processing. This requires ensuring and evaluating the quality of the coding system, e.g., by assessing the level of agreement between different coders.

Recommendations on ensuring reliability

1) Information on the **measurement precision of sensors** should be made transparent by manufacturers and documented in scientific publications that rely on sensor data.

2) Empirical research projects should make transparent the **methods used to determine measurement precision**. This can be done using a range of methods (e.g., parallel-forms method, testing-retesting method).

3) When reducing complex data (e.g., audio, video, text data) using coding procedures, the **quality of the reduction techniques** should be **documented** (e.g., through assessing inter-coder reliability, i.e., the level of agreement between different coders).

3.3   Validity

Validity refers to the correctness of conclusions that can be drawn from the data (Messick 1989). There are different facets of validity in empirical research (AERA/APA/NCME 2014; Messick 1995; Rammstedt et al. 2014) (see box below). Each of these facets play a role and must be observed when collecting data using new information technology. The following chapter focuses on the specific validity requirements when collecting data with new information technology. It pays particular attention to two facets of validity, namely, construct validity and external validity, while also briefly addressing internal validity.

> **The facets of validity referenced in this report**
>
> **Construct validity** refers to the extent to which inferences can be made based on the collected values in terms of the construct they aim to measure (Messick, 1995).
>
> **External validity** is achieved when the results of a study can be generalised to other locations, people, situations, and times (e.g. Eid/Gollwitzer/Schmitt 2017). One important facet of external validity is ecological validity. This refers to the question of whether the data gathered in an empirical study correspond to real-world settings.
>
> A study is said to be **internally valid** when a causal influence of the independent variable on the dependent variable can be ensured, i.e., if all relevant confounding variables can be controlled for (e.g., Eid/Gollwitzer/Schmitt 2017).

An advantage of collecting data with new information technology is that it facilitates the studying of individuals in their everyday lives. On the one hand, unlike research in a laboratory, data collected with new information technology have a much larger potential for ensuring ecological validity. On the other hand, they also result in specific problems related to other facets of validity. An example: the activity of two individuals is recorded over the course of two weeks using an activity tracker. Person A has a higher mean activity level than Person B. The validity of the conclusion that Person A is more active than Person B is threatened by several issues. Construct validity is threatened if it is not scientifically evaluated whether an activity tracker precisely measures physical activity. External validity is threatened if, instead of wearing the activity tracker for the full time, say, two weeks, the subjects take the tracker off in certain situations (e.g., while playing football). The following chapter looks at threats to construct validity and external validity that are specific to data collection using new information technology.

3.3.1 Construct validity

3.3.1.1 Sensor quality

Firstly, the quality of the used sensors plays an important role in evaluating a measurement's construct validity. Are the devices generally able to perform physiological measurements such as heart rate variability? There is a lack of such validation studies for many commercially available wearables (Peake/Kerr/Sullivan 2018). According to Shcherbina et al. (2017), for example, it is widely unknown how commercially produced wearables ensure valid recording of cardiovascular parameters (e.g., the heart rate and its variability). Quality of sensors can be examined in validation studies. They do so by comparing data collected by the sensors with data collected using gold-standard methods (e.g., physiological measurements in a standard laboratory environment). To obtain a value for measurement accuracy, the so-called error value – the difference between the value measured by the sensor and the gold-standard value, divided by the gold-standard value – can be calculated (Shcherbina et al. 2017, S. 5):

$$\blacktriangleright \text{ Error} = \frac{(\text{sensor measurement value} – \text{gold standard value})}{\text{gold standard value}}$$

This error value can also be affected by measurement errors and thus depends on the reliability of the measurements (see Section 4.2). However, if such errors are averaged across many measurements, it is possible to reduce the impact of unsystematic measurement errors. Moreover, it is assumed that the gold-standard values provide valid results. Previous validation studies show that sensors (wearables) indeed provide valid measurements to some extent. However, this does not apply to all available products and to all constructs of interest in the same way (Evenson/Goto/Furberg 2015; Peake/Kerr/Sullivan 2018; Shcherbina et al. 2017). The validity of measurements of energy expenditure by commercially available wristbands, for example, is questionable, whereas the measurement of heart rates appears to be reliable (Shcherbina et al. 2017). Thus, quality studies to ensure construct validity are indispensable. Because devices are continually developed, quality studies too must be conducted continually, and their results should be published. Moreover, researchers should present proofs of accuracy of such devices when using them for scientific purposes.

Data obtained from sensors can be ambiguous. For example, when interpreting heart rate data, it is important to adequately take movement data into account. Additionally, factors like posture, circadian effects, temperature, and behaviour (e.g., eating, drinking, smoking) can affect measurements and threaten validity. It is particularly important to note that the impact of such confounding factors are often several times higher than the effect of interest, e.g. intervention effects. For example, an increase of the heart rate by 20 beats through a change of position (getting up) and other movements is normal. Psychological manipulation does not usually result in effects of this magnitude (20 beats). When these confounding variables are unsystematic and not related to the intervention, this merely affects how precisely the effect is measured. However, when these confounding variables are systematic and related to the intervention, it threatens an additional facet of validity, namely, internal validity. Then, a change in the heart rate may not be attributable to the intervention but to confounding factors. Controlling for movement and body position is indispensable to ensure validity when recording the heart rate as a measurement in the behavioural sciences (Houtveen and de Geus 2009). A multi-method approach is recommended when reviewing construct validity. For example, the validity of conclusions based on the activity of the nervous system can be assessed by converging data from other information sources (e.g., catecholamine in urine, proportional pulse pressure, and electrodermal activity).

Taking the conditions of a measurement into account is not only important for research involving cardiovascular attributes. Movement data can also only be adequately interpreted when the situational context is documented during data collection (Schmidt et al. 2018). This can be done, for example, by using questionnaires or through observation.

3.3.1.2 Quality assurance for processing raw data through algorithms

Raw data from different sensors are frequently processed by algorithms to capture a construct of interest. Movement measurements, for example, are generated by algorithms processing data from accelerometers and by measuring gyration (using gyroscopes) (Klucken et al. 2019). So-called technical validation studies are required to verify whether these algorithms work correctly (ibid.).

Algorithms are also used when collecting audio and image data. The interpretation of such data is often based on these machine-generated data. To examine and ensure the validity of such data, these must be compared to real-world observations ('ground truth'). When recording audio, for example, which involves automated signal processing on the device (so-called behavioural signal processing; Narayanan and Georgiou 2013), the quality of the machine-generated behavioural data from a part of the participants can be compared to a manual code based on raw audio signal recordings. If the raw audio signal was not or could not be recorded during a study, the precision of the algorithm used for measuring behaviour can be compared against manual coding in a pilot study with raw audio recordings (in a limited context and with the consent of all participants, but under realistic conditions). When coding raw data, validation studies can also be used to examine the consistency of the coding.

3.3.2 External validity

External validity of studies based on new information technology is primarily threatened by selection biases related to the situations and persons under investigation as well as by reactive biases.

3.3.2.1 Selection biases: situations

Collecting data with new information technology offers the advantage of capturing data in individuals' real-world environments, ensuring ecological validity (Harari et al. 2019). However, ecological validity can be threatened, because certain real-life situations are prohibitive to data collection (e.g., smartphone-based questionnaires cannot be taken while driving a car because this would constitute a criminal offence). Additionally, participants might not want to answer questions on a smartphone or wear a wearable in specific situations (Brown/Blake/Sherman 2017). Moreover, some environments are unfavourable to technology-based data collection (e.g., GPS data on an underground train or inside of a building; microphone recordings with a pocketed device). When the recording of the data is selectively suspended in certain situations (e.g., during a conflict, possibly by activating a device's privacy button), the representativeness of a sample is no longer given and thus the results cannot be generalised to other situations.

3.3.2.2 Selection biases: persons

The use of new information technology requires specific skills (e.g., knowledge of how to use a smartphone) on behalf of study participants. This can result in sampling bias, which must be documented (e.g., response rate, documenting reasons for non-participation). Even then, it remains questionable to what extent the results can be transferred to other populations. However, concerns that certain studies mainly attract young, male individuals with an affinity to technology have been clearly refuted by studies in ageing research (Brose and Ebner-Priemer, 2015). The validity of results is also threatened when people share devices with others (e.g., smartphones, fitness trackers, etc.) or – in the extreme case – devices are used exclusively by someone other than the person selected for the study.

3.3.2.3 Reactivity effects

Inferences to everyday behaviour can also be threatened by reactivity resulting from individuals' awareness of the data collection process (Brown/Blake/Sherman 2017). The data collection process might prompt people to exhibit (socially) desirable behaviour, for example, by exercising more than usual. The same applies to consciously refraining from (socially) undesirable behaviour while participating in the study, which also limits the validity of the results. Previous empirical studies on reactivity have shown that these are largely unproblematic (Stone et al. 2003). While the scope for eliminating reactive biases is limited, it does tend to decrease after a few hours, or days, depending on the study's design. It is therefore recommended to examine the initial period of measurement of an analysis separately (Shrout et al. 2018). To date, there are no concrete recommendations on the time range of those initial measurements. The range can differ depending on the used sensor, the research interest, or study design (Clemes and Deans 2012). Moreover, it is recommended to draw on other data collection methods as a standard for comparison. The effect of eliciting socially desirable responses is more significant with questionnaire-based studies than with studies that collect data in everyday life situations of participants.

Recommendations on evaluating validity

Construct validity

1) For sensors used in scientific studies, validation studies should demonstrate the **convergence with gold-standard methods**.

2) The validity of results based on sensors should be examined in **theory-driven validation studies** (including the use of other data sources, e.g., behavioural observations and questionnaires).

3) In the case of ambiguous signals, the **validity of inferences** should be verified – to the extent that this is consistent with standards of data protection and research ethics (see Chapter 4) – by including additional methods as well as by documenting the data collection context. This is particularly important when collecting data on physiological processes in everyday life situations.

4) Whenever data are generated or evaluated using algorithms, technical validation studies should be used to assess the **correctness of algorithms**. Moreover, these data should be matched with real-world data ('ground truth'). To the extent possible, this comparison should be performed and documented in every data collection process (using subsamples). Pilot studies should at least be used where this is not possible in real-life applications.

External Validity

5) It must be documented when **selective samples of situations lead to missing data** and – if possible and reasonable – it must be reconstructed and corrected using appropriate methods. These can include the following:

   a) **Subsequent interviewing of participants** (e.g., using appropriately implemented questions on a smartphone); this can give some indication of reasons for missing data and possibly facilitate the reconstruction of missing data. However, such subsequent questioning can be viewed negatively by the respondents.

   b) Assessing selectivity in the recorded situations can be made possible by **comparing** the frequency of occurrence of such situations in datasets collected **through different methods** (mobile sensing, experience sampling, field observations, questionnaires). Studies which do this systematically are extremely rare, however, there are some examples to be found in research on illegal drugs (Linas et al. 2016).

c) If viable (depending on the burden on the participants and existing resources), the **data collection period or data collection frequency can be increased**. Data can be collected every 60, 30, or 15 minutes, for a brief duration in certain situations. It is also possible to survey levels of acceptance, or annoyance, to estimate the influence on survey compliance. Generally, such procedures may lead to higher compliance (Trull and Ebner-Priemer 2013). Yet, systematic studies on this issue are also still very rare (Stone et al. 2003).

d) To estimate the amount of missing data, it can help to document how often participants resorted to **editing and censoring** their data.

e) Providing **incentives** for participation (e.g., financial compensation) can increase compliance (Göritz 2014).

6) **Missing data** through **selective sampling of individuals** must be documented and – if possible and reasonable – **reconstructed and corrected** using appropriate methods. Researchers can rely, for this purpose, on established methods from survey research (Kreuter et al. 2018; Schupp and Wolf 2015):

a) In order to be able to analyse selectivity of participants and to adjust for it (using statistical techniques such as weighting) it will be useful to **randomly select participants from data sources** that contain relevant information on participants and non-participants. These can be administrative data, such as population registers, or existing representative surveys. The advantage of sampling from surveys (cf. Kreuter et al. 2018) is that researchers can bring in questions that measure important target variables of the following technology based survey, e.g., their usage intensity or survey equivalents of the variables later to be measured using new technology. Models can then show whether a study's participants differ from non-participants regarding the study's key measures.

b) Data from participants (and, if applicable, the control group) can be **compared with external data sources and statistics**.

c) Data from participants and non-participants can be **linked with commercial microgeographic data** (e.g., infas 360) (for the survey context, see Sinibaldi/ Trappmann/Kreuter 2014).

d) Non-participants can be interviewed about **reasons for non-participation**.

e) If the researchers themselves oversee recruiting, they can have recruiters **observe or estimate information on participants and non-participants** (cf. West 2013).

f) It must be documented whether and to what extent devices (e.g., smartphones, fitness trackers, etc.) are shared with other individuals. This can be done by:

- depending on the aim of the analysis, explicitly **excluding the sharing of wearables** used for the study in the **informed consent** form for the scientific study,

- interviewing participants, if sensible, whether and to what extent devices were shared or used by others. **Devices used by several individuals can be excluded** or – depending on the aim of the analysis – be given special consideration,

- using machine-learning-based algorithms to identify **variations in user behaviour** and to assign the collected data to various users (Ochoa/Bort/Porcar 2018). This must take data protection regulations into account (see chapter 4).

4 Research ethics and data protection

■ Data collection using new information technology is faced with specific issues of research ethics and data protection. Due consideration of these issues, ultimately, ensures data quality and the potential for (secondary) data (re-)use. They lead to specific requirements to the design of informed consent forms. The recommendations are based on German law and might have to be modified in other countries.

4.1 Issues of research ethics and data protection

Data collection using new information technology is often conducted in individuals' everyday lives. It can be directed at situations concerning individuals' privacy, understood here as 'a sphere where one can live a self-directed life without being influenced or observed by others' (Deutscher Ethikrat 2018; 190, translation by the authors). Even more importantly, data collection can take place in intimate situations, understood by the German Ethics Council (Deutscher Ethikrat 2018, 29) as those areas of life 'reserved only for those immediately involved, and of which any details are made available to selected third parties only with expressed consent, if at all.' Ensuring privacy and freedom of choice regarding which data are collected and stored from which areas of life is a pivotal aspect of research ethics (Langheinrich and Schaub 2019). Additionally, study participants might not be fully aware of the ways in which their privacy is affected by data collection using new information technology. This is a challenge in the way participants are informed, how to ensure their sovereignty over their data, and how to structure informed consent forms (Deutscher Ethikrat 2018; Kelly et al. 2013; Langheinrich and Schaub 2019).

The sheer volume of data and the possibilities for data linkage created by using new information technology makes safeguarding the anonymity of study participants increasingly challenging (Langheinrich and Schaub 2019; Miller 2012). This is especially true when collecting high-precision location data using GPS (Apte et al. 2019) as well as audio and video data (Kelly et al. 2013). When data are prepared and processed within the project, members of the project team might be able to identify study participants, e.g., based on their voices (Robbins 2017). As soon as data facilitate reconstructing personal information of individuals, they can no longer be deemed anonymous. The risk for re-identification is considerable when using new technology and must be minimised by taking appropriate measures. Moreover, it is not possible to predict to what extent future data analytics methods will facilitate re-identification. When anonymity cannot be ensured, data collection and processing fall under the General Data Protection Regulation (GDPR). Considering a very high re-identification risk, it is recommended, when in doubt, to apply the General Data Protection Regulation, even when data appear anonymous by today's standards. When the rights and freedoms of study participants are at a high risk, it is recommended to apply Article 35 GDPR and perform a data protection impact assessment. According to the GDPR, this risk is particularly high when processing high volumes of sensitive data, e.g., health data, and when 'new technologies' are used extensively (GDPR Article 35 (3), Recital 91, see box). When personal data are collected, according to the GDPR and as specified in Chapter 4.2., obtaining informed consent is essential.

§

Excerpt from the General Data Protection Regulation (GDPR)

Article 35 Data protection impact assessment

1) Where a type of processing, in particular using new technologies, and taking into account the nature, scope, context, and purposes of the processing, is likely to result in a high risk to the rights and freedoms of natural persons, the controller shall, prior to the processing, carry out an assessment of the impact of the envisaged processing operations on the protection of personal data. A single assessment may address a set of similar processing operations that present similar high risks.

2) The controller shall seek the advice of the data protection officer, where designated, when carrying out a data protection impact assessment.

3) A data protection impact assessment referred to in paragraph 1 shall in particular be required in the case of:

- a systematic and extensive evaluation of personal aspects relating to natural persons which is based on automated processing, including profiling, and on which decisions are based that produce legal effects concerning the natural person or similarly significantly affect the natural person;
- processing on a large scale of special categories of data referred to in Article 9 (1)[1] or of personal data relating to criminal convictions and offences referred to in Article 10; or
- a systematic monitoring of a publicly accessible area on a large scale.

(…)

7) The assessment shall contain at least:

- a systematic description of the envisaged processing operations and the purposes of the processing, including, where applicable, the legitimate interest pursued by the controller;
- an assessment of the necessity and proportionality of the processing operations in relation to the purposes;
- an assessment of the risks to the rights and freedoms of data subjects referred to in paragraph 1; and the measures envisaged to address the risks, including safeguards, security measures and mechanisms to ensure the protection of personal data and to demonstrate compliance with this Regulation taking into account the rights and legitimate interests of data subjects and other persons concerned.

Data collection in a person's social environment can also create challenges. Data collection may be clearly aimed at gathering data on a person's social environment, but not necessarily. However, in any case, data from bystanders may be collected. Video and audio recordings, for example, can capture data from people who never explicitly consented to taking part in the study (Brown/Blake/Sherman 2017; Kelly et al. 2013; Robbins 2017). It is therefore necessary to take precautions by making people aware that their data are being collected, ensuring that their participation is voluntary, and giving them the opportunity to refuse the collection of their data or to withdraw their consent (Article 7 (3) GDPR, this is especially relevant for specific categories of personal data pursuant to Article 9 GDPR). The situation is further complicated, because, in addition to obtaining consent of the 'recorded' participants, it is also necessary to obtain consent from individuals they interact with (so-called 'third parties'). Acquiring this consent is often tedious or, in some research contexts, simply impossible. The General Data Protection Regulation (GDPR) does not apply where it is possible to gather data on third parties anonymously. However, researching third parties without their knowledge can still be problematic

---

1 Processing of special categories of personal data according to Article 9 (1) GDPR refers to processing of data 'revealing racial or ethnic origin, political opinions, religious or philosophical beliefs, or trade union membership, and the processing of genetic data, biometric data for the purpose of uniquely identifying a natural person, data concerning health or data concerning a natural person's sex life or sexual orientation'.

from a research ethical perspective. As mentioned above, maintaining complete anonymity is often not possible due to the large volumes of information-dense data that is facilitated by new information technology.

Researchers may also encounter situations where they record illegal or criminally liable behaviour. Such situations put into question whether (1) the responsible authorities must be notified, (2) researchers must release the data on request, and (3) the participants must be notified in advance of this possibility. Generally, researchers are not obliged to notify the authorities when they acquire knowledge of illegal or criminally liable behaviour during a study, except in the case of particularly serious offences according to the German Criminal Code (*Strafgesetzbuch*, StGB) (§ 138 StGB). These include murder (§ 211 StGB), manslaughter (*Totschlag*) (§ 212 StGB), offences against personal liberty (§§ 232 ff. StGB), high treason (§§ 81 ff. StGB) and treason (§§ 94 ff. StGB), counterfeiting of money or bonds (§§ 146 ff. StGB), robbery and extortion (§§ 249 ff. StGB), and offences constituting public danger (§§ 306 ff. StGB). In the case of such particularly serious criminal offences, researchers must notify the responsible law enforcement agencies (police, state prosecutors) even if they only gain knowledge about such offences, because failure to report a planned offence itself constitutes a criminal offence (§ 138 StGB). If requested during a criminal investigation, a study's documents must be released to the law enforcement agencies in any case – i.e., also in the case of non-serious crimes. Law enforcement agencies may also confiscate them, where necessary (§ 94 (2) Code of Criminal Procedure (*Strafprozessordnung*, StPO)), because researchers are not bound by the obligation of professional secrecy laid out in § 203 StGB. From an ethics perspective, study participants should always be informed about this possibility.

When saving data on smartphones or transmitting data via the internet, there is a danger of third parties gaining unlawful access to these data, e.g., through hacking (Carter et al. 2015). This danger should be considered, and care should be taken to prevent this as far as possible, and to document the steps taken (GDPR Article 24. (1)). In the following section, we will summarise and supplement the recommendations formulated in the relevant literature on issues of research ethics and data protection.

Recommendations relating to anonymity

1) In general, **strategies to render data anonymous or pseudonymous** should be applied. According to the law, data must generally be made anonymous after the research purpose has been achieved and, until then, be stored in a way that prevents de-identification, namely, by using pseudonymisation. (§ 27 (3) Federal Data Protection Act (BDSG)).

2) Where anonymisation is not ensured, the legal requirements of the GDPR must be implemented. Especially the **need for a data protection impact assessment** must be reviewed and, where necessary, carried out.

3) Data should only be **stored, processed, and transmitted through secure pathways**, e.g., by using encryption (Brown/Blake/Sherman 2017; Carter et al. 2015).

4) Generally, one-way hash functions can be used to ensure **(formal) anonymity of study participants** to prevent inferences about specific individuals (see Naor and Yung 1989).

5) Researchers should refrain from storing data that facilitate re-identification of individuals (e.g., location data) if they are not necessary to answer the concrete research question. Geodata, for example, should only be collected, directly (e.g. via GPS, wi-fi, or mobile phone data) or indirectly (e.g., through context data), if they are necessary for answering the research question. When using other data sources, '**location tracing**' (geo-inferencing using Twitter data) **should be avoided if it is unnecessary for answering the research question** (Johnson et al. 2016).

6) Researchers should make **use of data type-specific methods for anonymisation** (which may lead to information loss), including voice transformation (distortion) in audio recordings and transformation of coordinate systems (while maintaining the topology) in location data (Mainali/Shepherd/Petitcolas 2019).

7) To reduce the risk of re-identification, data (e.g., location data, audio data) can be **saved in a less detailed or processed form** (e.g., through behavioural signal processing). Raw data can be deleted right off the device (e.g., Feng et al. 2018; Wyatt et al. 2011). This creates some **tension with the quality-assuring recommendations** on construct validity (see recommendations in Chapter 3.3.2). It is sensible to balance these interests and to document the process.

8) Data containing **identifiable information** (e.g. full names, addresses, bank account details), e.g., in audio and video recordings, should be **deleted or edited** (e.g., beeped out) during the project/coding process. (Robbins 2017).

9) If **people involved in the processing** of the data (e.g., coders) **identify study participants** (e.g., based on their voice), they should cease processing the data and transfer responsibility to another person (Robbins 2017).

Recommendations relating to the social context/third parties

1) Researchers should **obtain informed consent from individuals** that are in regular contact with a person taking part in a study (e.g., family, friends, co-workers) and from whom data might also be collected (Kelly et al. 2013).

2) The **risk of being recorded should be made transparent** to minimise privacy expectations of third parties. To do so, participants of a study might wear a button indicating the risk of being recorded ('This conversation might be recorded'), enabling researchers to assume passive consent by the test persons and providing ethical justification (Mehl 2017; Robbins 2017). Alternatively, study participants could be instructed to highlight the recording risk to people they interact with (Robbins 2017).

3) **Devices should be turned off** if people from the participant's social context request so (Brown/Blake/Sherman 2017; Robbins 2017).

4) Without the **explicit consent** for personal data use, researchers should use filters to **distort the voices of third parties in audio recordings**.

5) Without the **explicit consent** for personal data use, researchers should **blur the faces of third parties in video recordings**.

6) **Publishing data** that facilitate identifying third parties **requires their consent** (Kelly et al. 2013).

7) Individuals from a study participant's social context, who can be identified through recordings, must be given the opportunity to **change or withdraw their consent for the use of their data for any given purpose at any time** (Kelly et al. 2013).

Recommendations relating to privacy

1) Study participants should be informed in great detail about the way their data are collected, saved, and processed, and about any risks to their privacy. They should be informed in full detail about **how to request corrections and deletion of their data, and how to make changes to their informed consent** (Kelly et al. 2013; Langheinrich and Schaub 2019) (see Chapter 4.2 – informed consent).

2) Study participants should be given **full control of their data** (informational self-determination). They should be given the opportunity to receive their data in a commonly used format (Article 20 GDPR), which later enables them to request the deletion of specific sequences (**retroactive censoring**) (Kelly et al. 2013; Langheinrich and Schaub 2019; Mehl 2017).

3) Where possible, study participants should be able **to delete personal data** (e.g., audio and image recordings) before the research data are shared with the researchers (Brown/ Blake/Sherman 2017; Robbins 2017).

4) Recording devices should **facilitate proactive censorship**, e.g., by providing a privacy button to suspend data recording (Kelly et al. 2013; Langheinrich and Schaub 2019; Robbins 2017).

5) Study participants should be able to **take off recording** devices if it makes them feel more comfortable in specific situations (Kelly et al. 2013).

6) Study participants should have the **opportunity for pro- and retroactive censorship** to protect themselves against the recording of illegal or reputation-damaging behaviour. Study participants should be informed that (1) particularly serious criminal offences must be reported to the responsible authorities, and that (2) documents relating to the study must be made available for investigation by law enforcement agencies for any type of offence.

7) **Continuous recording** (e.g., audio recordings) **should be avoided** if it is not necessary for answering a research question. Short segments (30 seconds or less) with short, appropriate intermissions are preferable to continuous recording (if the research questions allow this) (Mehl 2017).

8) **Audio and image data** that allow the identification of persons should **not be passed on without the explicit consent** of the person in question (e.g., through an explicit declaration of consent for future data use) (Kelly et al. 2013).

9) **Raw sensor data** with a risk of facilitating breaches of privacy at a later point (e.g., video data) should only be **passed on or published in a processed form** (e.g., in low resolution or with a reduced number of frames per second) (e.g., Muaaz and Mayrhofer 2017; Ranjan and Whitehouse 2015).

10) The **principle of data minimisation should always be applied to audio recordings**. Researchers should only record the audio needed for behavioural analysis (e.g., short segments (30 seconds or less), not high-frequency data collection (Mehl 2017)).

11) When **collecting data using smartphones**, researchers must ensure that access to private data of the smartphone's owner is not possible and that data are only transmitted with explicit consent of the data owner.

4.2   Informed consent

For reasons of research ethics as well as data protection, it is necessary to obtain informed consent of study participants. Informed consent – and, with it, the voluntary decision to take part in a study – requires that people are informed in plain language about the benefits and risks of the data collection and processing (see Section 4.1) (Carter et al. 2015). The complexity of data collecting and processing using new information technology makes it necessary to strike a balance here. On the one hand, it must be ensured that people understand the implications that a study has for them. On the other hand, they must not be overburdened with too-detailed information (Carter et al. 2015; Langheinrich and Schaub 2019). Overly detailed or long informed consent forms run the risk of not being read, or not being read fully (Langheinrich and Schaub 2019).

4.2.1   Concepts of consent

The German Ethics Council (2018) addressed the issue of formulating declarations of informed consent in the context of big data.[2] It did so based on the fact that, when gathering data using big data methods, new questions can arise during the data collection and data analysis process that make data linkage necessary, resulting in additional risks. Classic concepts of information and consent, as used in narrowly defined clinical studies or studies in the social and behavioural sciences, have their limits in a big-data context. Due to their similar data volume and potential for linkage with other sources, the opportunities created by collecting data using new information technology are closely related to issues of big data. Concepts of consent discussed in a big-data context can thus be made fruitful for data collection using new information technology. The German Ethics Council (2018, 183-185) distinguishes three models of consent: blanket consent, dynamic consent, and cascading or meta consent. Additionally, the GDPR facilitates broad consent.

---

2   The German Ethics Council (2018, 54) defines big data as follows: 'Big data refers to the handling of large volumes of data aimed at identifying patterns and drawing new insight from them, which employs innovative and continuously developed information technology in light of the abundance and variety of data as well as the speed at which they are collected, analysed, and linked.' (Translation by the authors)

Concepts of consent (German Ethics Council, 2018, p. 183-185; GDPR)

Blanket consent

Study participants consent to an undetermined future use and passing-on of their data. All other decisions relating to handling the data are not made by the study participants but by those managing the database.

Dynamic consent

Study participants are repeatedly contacted to obtain their consent to specific research questions. They are in regular contact with a database (e.g., via an online platform or telephone).

Cascading consent or meta consent

This is an extension of dynamic consent. While it is not necessary to obtain consent for each individual research question, study participants may choose between a range of options. In addition to the classic dynamic consent model, they can, for example, give consent for different categories of research questions instead of being contacted for each individual question. It is also possible to pass on the responsibility for making the decision (delegating it to a trustee) to other people or bodies (e.g., expert committees), to make one's data available for any use without restriction, or to refuse participation. Individual options can be combined or changed over time.

Broad consent based on the GDPR

Recital 33 GDPR created the concept of 'broad' consent especially for scientific research. By doing so, the European legislator acted on the assumption that the purpose of processing personal data in scientific research cannot always be fully specified at the time when personal data are collected. It therefore facilitates that study participants provide 'broad' consent to a) certain areas of research or parts of research projects, b) to the extent permitted by the purpose, and c) in accordance with ethical standards in scientific research. These 'certain areas,' however, must be related to the original research goal.

By letting study participants adjust options to meet their needs, the cascading model facilitates tailored safeguarding of informational self-determination.
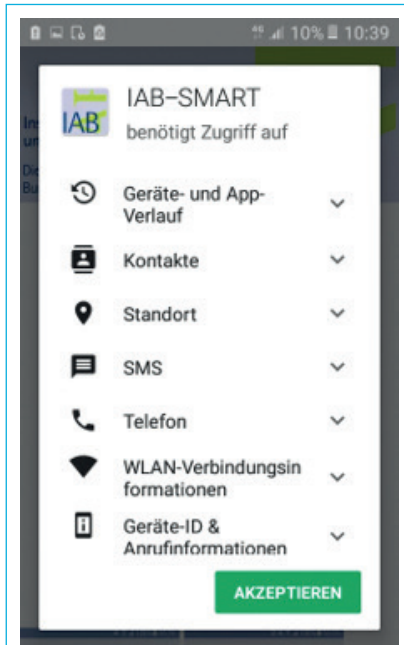
Screenshot 1: Play Store downloading process: granting permissions
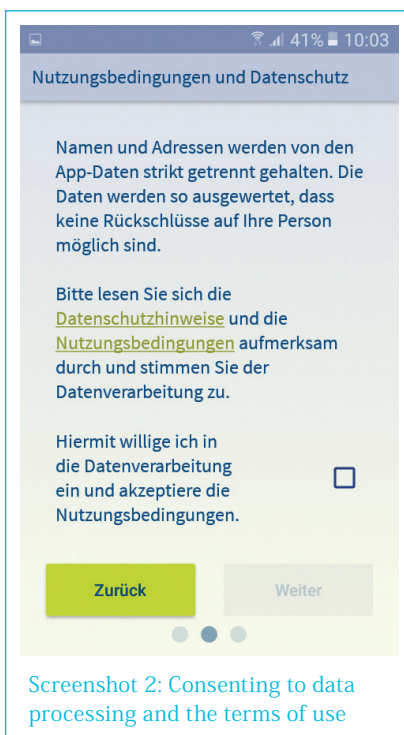
Screenshot 2: Consenting to data processing and the terms of use

Example: Recruiting and consent process of the IAB-SMART study

The following section will outline the recruiting and consent process of the IAB-SMART study (Kreuter et al. 2018) to provide an example of how to implement the requirements of the EU GDPR. During this panel study of the Institute for Employment Research (IAB), participants were asked to install a research app, the IAB-SMART app, on their smartphones. The app collected data in two ways: firstly, through short questionnaires (about every 14 days) and, secondly, by making passive data available to the researchers. The latter required permissions across five categories: 1) mobile phone network quality and location information, 2) interaction history, 3) characteristics of the social network, 4) activity data, 5) smartphone usage. The following information are largely based on Kreuter et al. (2018). The complete documentation of the study is available at https://www.iab.de/smart.

Recruitment was done by postal mail. Letters to potential participants contained a short invitation and a link for downloading the app, an individual registration code, information on data protection, a short description of the data collection functions, information on incentives, and a brochure explaining the installation process stepwise. The description of the data collection functions was published in a separate document to make it more visible. In as plain language as possible, and for each of the later requested data collection functions, it explained which data are collected and why they are needed and for which research purpose.

The admission process was designed based on the requirements of Articles 5 and 7 and the recitals 32 and 33 GDPR. Key issues were documentation, who consented when to what (Article 7 (1)), making the consent request distinguishable from other matters (Article 7 (2)), the right to withdraw consent (Article 7 (3)), transparency through the use of clear and plain language (Article 7 (2)), active consent by way of an opt-in (Recital 32), and purpose limitation (Recital 33 and Article 5 (1b)).

It must be noted that, during implementation, the process had to consider the peculiarities of app installation on Android-based smartphones at the time. The installation process could only be completed if the user, as a first step, granted all permissions the app could potentially require (see Screenshot 1)[3]. However, granting permissions at this point was a mere technical requirement for installing the app and did not initiate data collection. Fortunately, more recent Android versions have improved permission management, asking users for each permission only before using apps for the first time. This avoids unnecessary permission requests and gives the test persons more individual control.

---

3 The app was only available in German. Therefore, the screenshots show the German text.

Upon opening the app, a request for entering the registration code made sure that only invited individuals could take part in the study. Then – again, this is an Android-specific request – the participants were asked to agree to the general terms of use and the data protection disclaimer (see Screenshot 2). Both were linked on-screen and had been previously sent to the invitees with the invitation letter and all other documents by post. Participation and the subsequent data collection were only possible if consent was given at this point.

It is important to note that even at this point the app was not able to collect any passive data from a participant's smartphone after giving consent. Data collection authorisation was given by ticking the boxes on the following screen (see Screenshot 3).

Consent had to be given individually for each research purpose. This required active consent (opt-in): none of the boxes were pre-ticked in the default setting. Data collection had to be activated by the participants by actively ticking a box. All permissions were displayed on a single screen. By clicking on each of the functions, users could display and read the explanations, which were shortened versions of the more detailed explanations in the invitation letter.

At this point, activating a function granted permission to the requested data. To comply with the provision that withdrawal must be as easy as consent (see Article 7 (3) GDPR)), users could return to this screen through the settings menu accessed via the main menu and un-tick the boxes to withdraw consent (see Screenshot 4).
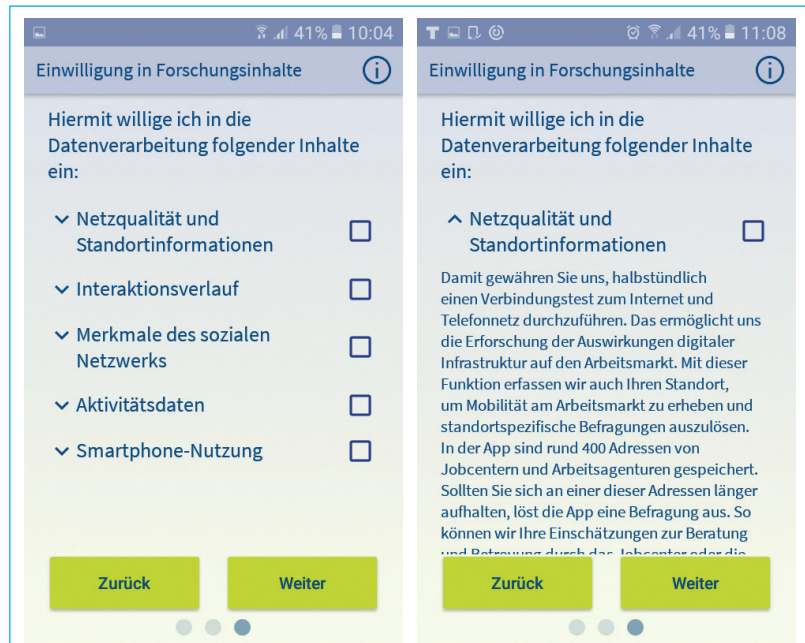
Screenshot 3: Consenting to the collection of passive data

Screenshot 4: Main menu and changing permissions and research purposes

4.2.2 Consent from third parties

In addition to data from participants, it is possible that data from third parties (individuals from a study participant's social context) are collected. Here, it is sometimes not possible or sensible to obtain written consent. However, it must be ensured in any case that individuals are informed about the risk of being recorded in a realistic and practicable manner (e.g., participants wear a button indicating the risk of being recorded). It is then assumed that the third parties have consciously decided to 'act' in front of the microphone or camera and consented to being recorded. Consent, in this context, is understood as a clear affirmative act, e.g., continuing a conversation. Although written consent is not essential, the General Data Protection Regulation requires proof of consent of all contact persons. This proof can also be recorded electronically. Seeing as the General Data Protection Regulation does not contain form requirements for consent[4], this proof can also be provided through an oral declaration. An affirmative act – for example, where the risk of being recorded was made transparent and a third party enters a conversation in awareness of this risk – can also constitute consent. This giving of consent must be documented by researchers.

Recommendations relating to informed consent

1) **Participants must be informed** about the use, risks, and the type of data collection as well as the purpose of data use, data storage, and future data use **in plain language** enabling them to consent to the processing of their data in an informed and voluntary fashion.

2) It should always be made possible to **change or withdraw consent** for data use for predefined purposes.

3) To ensure that **informed consent forms** are read in full, it is possible to provide **short versions and attaching additional information**. Online studies can use videos explaining a study's purpose, procedure, and data protection issues. The box for granting consent then appears after it has been made sure that the video has been viewed in full and with the sound on.

4) **Cascading consent concepts** can be a solution to ensure the data can be used on a long-term basis while duly considering the right to self-determination on behalf of the respondents.

5) P**articipation must remain voluntary**. Study participants must always be able to cancel their participation.

6) Study participants should have the opportunity to **access the collected data and correct them**.

7) When recording video or audio, it is necessary to obtain the **consent of third parties** (e.g., contact persons). This can also be implemented by implied consent, which must be **documented**. The form of documentation is flexible and can include written declarations, electronic declarations (email, ticking a box in an electronic form), or log entries that document that third parties took part in the study in a voluntary and informed manner.

8) **Apps should only ask for the access permissions they require for functioning** (access to contacts, camera, microphone, location data, etc.). They should not ask for blanket permission.

---

4   According to Article 5 (1) (a) GDPR data processing must be lawful. If there is no other legal basis, processing is lawful if consent was given pursuant to Article 6 (1) (a) GDPR. Compliance with lawfulness must be proven based on Article 5 (2) GDPR ('accountability'). Written proof is not required.

5   Data management

■ Studies using new information technology pose challenges to data management, on the one hand, due to the high amounts of data they can produce and, on the other hand, due to the challenges this report has already addressed in Chapters 3 and 4, including data quality, research ethics, and data protection.

5.1   Interoperability

Studies using new information technology can result in extensive datasets. This has been well-illustrated by Miller (2012, 230), who refers to a smartphone study by Kiukkonen et al. (2010), which examined 168 individuals over an average of four months. It gathered the following data: 15 million Bluetooth scans, 13 million wi-fi scans, 15 million GPS logs, 4 million app use recordings, 500,000 accelerometer measurements, 220,000 voice recordings, 130,000 voice calls, 90,000 text messages, 28,000 photos, and 2,000 videos. Such volumes of data pose a challenge to data storage and data management, requiring interdisciplinary cooperation between the social, behavioural, and economic sciences, on the one hand, and between computer and data science, on the other. Modern information technology provides researchers with the tools to develop data storage systems for large datasets and to manage problems related to large datasets (e.g., Doukas and Maglogiannis 2011; Ranjan et al. 2019).

When collecting data using wearables and smartphones, using different devices can be problematic depending on a study's design. How does one deal with data stored in various formats? Researchers have developed solutions for dealing with issues of interoperability relating to wearables when storing data (e.g., Mezghani et al. 2015).

5.2   Data archiving and data storage

The replication crisis in many empirically-based disciplines has given rise to Open Science, a new paradigm that – to the extent possible – calls for making research data and data processing methods (e.g., algorithms) openly available (e.g., Klenk et al. 2019). The German Research Foundation (DFG) has included these principles in its guidelines for good scientific practice (DFG 2019, Guideline 13). Gathering data using new information technology and using new information technology for data storage and data management creates new opportunities for involving the general public in the research process. The development of 'citizen science' – understood as the involvement of citizens on all levels of the research process, including data collection, data analysis, and even interpretation of research results – is a good example for this (Greshake Tzovaras et al. 2019). Through the intense use of smartphones and wearables, the general public produces datasets every day, which could be fed into a common data storage system. An example of this approach is the Open Humans Project, which, in addition to data from wearables, also gleans data from other sources (e.g., the genome) (ibid. 2019). This requires flexible data storage and management systems that make sure that participants can, on the one hand, make their data available and, on the other hand, keep control over their data and independently manage them. It is also possible to develop data storage systems enabling people to maintain ownership over their data and store their data locally, which can then be accessed to perform calculations (e.g., Lai et al. 2018[5]). Data storage and data management systems for new information technology can also make use of other recent technological developments such as cloud

---

5   The data stored in the database are encrypted in a way that requires the user to enter a password every time. This ensures security even if the database is stolen. The procedure implies that even the database's operator cannot access the data without the study participants' consent.

systems (e.g., Doukas and Maglogiannis 2011). These systems go beyond the scope of classic data repositories that archive data in research data centres.

Due to the rapid development of information technology and the complexity of IT systems, this report cannot address these aspects in greater detail. However, the challenges related to data protection, research ethics, and data management, which we discussed in Chapters 4 and 5, result in more general, cross-system challenges. In addition to online storage systems, they also relate to data storage and data management in repositories located at research facilities and research data centres. This leads to some additional recommendations.

Recommendations on data management

1) It is recommended to continue to develop general **standards for data storage** that also ensure interoperability.

2) To assess the quality of the data collection and analysis, and to facilitate re-analysis, raw data as well as **all information on employed devices, and other data processing strategies** (e.g., algorithms), should be archived and **made available** together with the data (as recommended in Section 4.1).

3) If data are made openly available, e.g., online or in open repositories, it must be ensured that **deanonymisation is not possible**. Due to the availability of numerous data sources that facilitate deanonymisation, and a lack of knowledge about the extent to which future data analysis methods might make deanonymisation possible (see, e.g., Rocher/ Hendrickx/de Montjoye 2019), such data should not be made publicly available unless the participants – after having been informed about the risks – explicitly consent to it in a declaration of informed consent.[6] It is recommended to store data in research data centres with limited and regulated data access, either on-site or online using suitable password protection. The passwords can be **managed at research data centres** and given out to interested researchers after consulting with the authors of a publication and with contractual safeguards in place (data use agreements to prevent misuse).

4) **Storage and further use of data** should be included in **declarations of informed consent**. To do so, researchers can use one of the procedures presented in Chapter 4, which can include **delegating responsibilities to a trustee**.

5) It is recommended to further develop data analysis procedures that enable **individuals to maintain ownership of their data**. The data used for calculations are then stored in different locations (locally with the participants).

---

6   Participants can be made aware of the uniqueness of their personal data using resources like the following website: https://cpg.doc.ic.ac.uk/individual-risk

6 References

AERA [American Educational Research Association], APA [American Psychological Association] and NCME [National Council of Measurement in Education] (2014): Standards for educational and psychological testing. Washington, DC, AERA.

Althoff, Tim; Rok Sosič, Jennifer L. Hicks, Abby C. King, Scott L. Delp and Jure Leskovec (2017): Large-scale physical activity data reveal worldwide activity inequality. Nature 547(7663), 336-339.

Apte, Aditi; Vijendra Ingole, Pallavi Lele, Andrew Marsh, Tathagata Bhattacharjee, Siddhivinayak Hirve, Harry Campbell, Harish Nair, Sarah Chan and Sanjay Juvekar (2019): Ethical considerations in the use of GPS-based movement tracking in health research - lessons from a care-seeking study in rural west India. Journal of Global Health 9(1), 010323. https://doi.org/10.7189/jogh.09.010323 (Last accessed: 27.01.2020).

Böhmer, Matthias; Brent Hecht, Johannes Schöning, Antonio Krüger and Gernot Bauer (2011): Falling asleep with Angry Birds, Facebook and Kindle: a large scale study on mobile application usage. In: Markus Bylund, Oskar Juhlin and Ylva Fernaeus (eds.): Proceedings of the 13th Conference on Human-Computer Interaction with Mobile Devices and Services, Mobile HCI 2011, Stockholm, Sweden, August 30 - September 2, 2011. New York, ACM, 47-56. https://doi.org/10.1145/2037373.2037383 (Last accessed: 27.01.2020).

Brose, Annette and Ulrich Ebner-Priemer (2015): Ambulatory assessment in research on aging: Contemporary and future applications. Gerontology 61(4), 372-380. https://doi.org/10.1159/000371707 (Last accessed: 27.01.2020).

Brown, Nicolas A.; Andrew B. Blake and Ryne A. Sherman (2017): A snapshot of life as lived: Wearable cameras in social and personality psychological science. Social Psychological and Personality Science 8(5), 592-600.

Brown, Harriet R.; Peter Zeidman, Peter Smittenaar, Rick A. Adams, Fiona McNab, Robb B. Rutledge and Raymond J. Dolan (2014): Crowdsourcing for cognitive science–the utility of smartphones. PLOS ONE 9(7), e100662. https://doi.org/10.1371/journal.pone.0100662 (Last accessed: 27.01.2020).

Carter, Adrian; Jacki Liddle, Wayne Hall and Helen Chenery (2015): Mobile phones in research and treatment: Ethical guidelines and future directions. Journal of Medical Internet Research mHealth and uHealth 3(4), e95. https://doi.org/10.2196/mhealth.4538 (Last accessed: 27.01.2020).

Clemes, Stacy A. und Nuala K. Deans (2012): Presence and duration of reactivity to pedometers in adults. Medicine & Science in Sports & Exercise 44(6), 1097-1101. http://dx.doi.org//10.1249/MSS.0b013e318242a377 (Last accessed: 27.01.2020).

Deutscher Ethikrat (2018): Big Data und Gesundheit – Datensouveränität als informationelle Freiheitsgestaltung. Stellungnahme. Berlin, Deutscher Ethikrat. https://www.ethikrat.org/fileadmin/Publikationen/Stellungnahmen/deutsch/stellungnahme-big-data-und-gesundheit.pdf (Last accessed: 27.01.2020).

DFG [Deutsche Forschungsgemeinschaft] (2019): Leitlinien zur Sicherung guter wissenschaftlicher Praxis. Kodex. Bonn, Deutsche Forschungsgemeinschaft (DFG). https://www.dfg.de/download/pdf/foerderung/rechtliche_rahmenbedingungen/gute_wissenschaftliche_praxis/kodex_gwp.pdf (Last accessed: 27.01.2020).

Diehl, Manfred; Karen Hooker and Martin J. Sliwinski (eds.) (2015): Handbook of intraindividual variability across the life span. New York, Routledge.

Doukas, Charalampos and Ilias Maglogiannis (2011): Managing wearable sensor data through cloud computing. 2011 IEEE Third International Conference on Cloud Computing Technology and Science. Athens, Greece, November 29 - December 1, 2011, 440-445. https://doi.org/10.1109/CloudCom.2011.65 (Last accessed: 27.01.2020).

Eid, Michael; Mario Gollwitzer and Manfred Schmitt (2017): Statistik und Forschungsmethoden. 5th edition, Weinheim, Beltz.

Eid, Michael and Katharina Schmidt (2014): Testtheorie und Testkonstruktion. Göttingen, Hogrefe.

Evenson, Kelly R.; Michelle M. Goto and Robert D. Furberg (2015): Systematic review of the validity and reliability of consumer-wearable activity trackers. International Journal of Behavioral Nutrition and Physical Activity 12(159). https://doi.org/10.1186/s12966-015-0314-1 (Last accessed: 27.01.2020).

Fahrenberg, Jochen and Michael Myrtek (eds.) (1996): Ambulatory assessment. Computer-assisted psychological and psychophysiological methods in monitoring and field studies. Göttingen, Hogrefe.

Feng, Tiantian; Amrutha Nadarajan, Colin Vaz, Brandon Booth and Shrikanth Narayanan (2018): TILES audio recorder: an unobtrusive wearable solution to track audio activity. WearSys '18 Proceedings of the 4th ACM Workshop on Wearable Systems and Applications. 2018 Munich, Germany, June 10, 2018, 33-38. https://doi.org/10.1145/3211960.3211975 (Last accessed: 27.01.2020).

Georgiou, Konstantinos; Andreas V. Larentzakis, Nehal N. Khamis, Ghadah I. Alsuhaibani, Yasser A. Alaska and Elias J. Giallafos (2018): Can wearable devices accurately measure heart rate variability? A systematic review. Folia Med (Plovdiv) 60(1), 7-20. https://doi.org/10.2478/folmed-2018-0012 (Last accessed: 27.01.2020).

Göritz, Anja S. (2014): Incentive effects. In: Uwe Engel, Ben Jann, Peter Lynn, Annette Scherpenzeel und Patrick Sturgis (eds.): Improving survey methods: Lessons from recent research. London, Routledge, 339-350.

Greshake Tzovaras, Bastian; Misha Angrist, Kevin Arvai, Mairi Dulaney, Vero Estrada-Galiñanes, Beau Gunderson, Tim Head, Dana Lewis, Oded Nov, Orit Shaer, Athina Tzovara, Jason Bobe and Mad Price Ball (2019): Open Humans: A platform for participant-centered research and personal data exploration. GigaScience 8(6), giz076. https://doi.org/10.1093/gigascience/giz076 (Last accessed: 27.01.2020).

Harari, Gabriella M.; Nicholas D. Lane, Rui Wang, Benjamin S. Crosier, Andrew T. Campbell and Samuel D. Gosling (2016): Using smartphones to collect behavioral data in psychological science: Opportunities, practical considerations, and challenges. Perspectives on Psychological Science 11(6), 838-854. https://doi.org/10.1177/1745691616650285 (Last accessed: 27.01.2020).

Harari, Gabriella M.; Sandrine R. Müller, Clemens Stachl, Rui Wang, Weichen Wang, Markus Bühner, Peter J. Rentfrow, Andrew T. Campbell and Samuel D. Gosling (2019): Sensing sociability: Individual differences in young adults' conversation, calling, texting, and app use behaviors in daily life. Journal of Personality and Social Psychology. Advance online publication, May 20, 2019. http://dx.doi.org/10.1037/pspp0000245 (Last accessed: 27.01.2020).

Houtveen, Jan H. and Eco J. C. de Geus (2009): Noninvasive psychophysiological ambulatory recordings: Study design and data-analysis strategies. European Psychologist 14(12), 132–141. http://dx.doi.org/10.1027/1016-9040.14.2.132 (Last accessed: 27.01.2020).

Johnson, Isaac L.; Subhasree Sengupta, Johannes Schöning and Brent Hecht (2016): The geography and importance of localness in geotagged social media. Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems (CHI '16). New York, NY, USA, ACM, 515-526. https://doi.org/10.1145/2858036.2858122 (Last accessed: 27.01.2020).

Kelly, Paul; Simon J. Marshall, Hannah Badland, Jacqueline Kerr, Melody Oliver, Aiden R. Doherty and Charlie Foster (2013): An ethical framework for automated, wearable cameras in health behavior research. American Journal of Preventive Medicine 4(3), 314-319. https://doi.org/10.1016/j.amepre.2012.11.006 (Last accessed: 27.01.2020).

Kiukkonen, Niko; Jan Blom, Oliviel Dousse, Daniel Gatica-Perez and Juha Laurila (2010): Towards rich mobile phone datasets: Lausanne data collection campaign. Proceedings of the International Conference on Pervasive Services. Berlin, ICPS, 68. http://publications.idiap.ch/downloads/papers/2010/Kiukkonen_ICPS,BERLIN,2010_2010.pdf (Last accessed: 27.01.2020).

Klenk, Juergen; Philip R. O. Payne, Rasu Shrestha and Margo Edmunds (2019): Open science and the future of data. In: Margo Edmunds, Christopher Hass und Erin Holve (eds.): Consumer informatics and digital health. Solutions for health and health care. Basel, Springer International Publishing, 337-357. https://doi.org/10.1007/978-3-319-96906-0 (Last accessed: 27.01.2020).

Klucken, Jochen; Till Gladow, Johannes G. Hilgert, Marc Stamminger, Christian Weigand and Björn Eskofier (2019): „Wearables" in der Behandlung neurologischer Erkrankungen – wo stehen wir heute? Der Nervenarzt 90(8), 787–795. https://doi:10.1007/s00115-019-0753-z (Last accessed: 27.01.2020).

Kreuter, Frauke; Georg-Christoph Haas, Florian Keusch, Sebastian Bähr and Mark Trappmann (2018): Collecting survey and smartphone sensor data with an app: Opportunities and challenges around privacy and informed consent. Social Science Computer Review, OnlineFirst. https://doi.org/10.1177/0894439318816389 (Last accessed: 27.01.2020).

Lai, Russell W. F; Christoph Egger, Manuel Reinert, Sherman S. M. Chow, Matteo Maffei and Dominique Schröder (2018): Simple password-hardened encryption services. 27th USENIX Security Symposium (USENIX Security 18), Baltimore, MD, USA, August 15-17, 2018, USENIX Association, 1405-1421.

Langheinrich, Marc and Florian Schaub (2019): Privacy in mobile and pervasive computing. San Rafael, California (USA), Morgan & Claypool Publishers. https://doi.org/10.2200/S00882ED1V01Y201810MPC013 (Last accessed: 27.01.2020).

Linas, Beth S.; Andrew Genz, Ryan P. Westergaard, Larry W. Chang, Robert C. Bollinger, Carl Latkin and Gregory D. Kirk (2016): Ecological momentary assessment of illicit drug use compared to biological and self-reported methods. Journal of Medical Internet Research mHealth and uHealth 4(1), e27. https://doi.org/10.2196/mhealth.4470 (Last accessed: 27.01.2020).

Mainali, Pradip; Carlton Shepherd and Fabien A. P. Petitcolas (2019): Privacy-enhancing context authentication from location-sensitive data. In Proceedings of the 14th International Conference on Availability, Reliability and Securit, 87-97. Osaka, ACM Press. https://doi.org/10.1145/3339252.3340334 (Last accessed: 27.01.2020).

Mehl, Matthias R. (2017): The Electronically Activated Recorder (EAR): A method for the naturalistic observation of daily social behavior. Current directions in psychological science 26(2), 184-190. https://doi.org/10.1177/0963721416680611 (Last accessed: 27.01.2020).

Messick, Samuel (1989): Validity. In: Robert L. Linn (ed.): Educational measurement. 3rd edition, New York, Collier Macmillan Publishers, 13-103.

Messick, Samuel (1995): Validity of psychological assessment: Validation of inferences from persons' responses and performances as scientific inquiry into score meaning. American Psychologist 50(9), 741-749. http://dx.doi.org/10.1037/0003-066X.50.9.741 (Last accessed: 27.01.2020).

Miller, Geoffrey (2012): The smartphone psychology manifesto. Perspectives on Psychological Science 7(3), 221-237. https://doi.org/10.1177/1745691612441215 (Last accessed: 27.01.2020).

Mischel, Walter and Yuichi Shoda (1995): A cognitive-affective theory of personality: Reconceptualising situations, dispositions, dynamics, and invariance in personality structure. Psychological Review 102(2), 246-268. http://dx.doi.org/10.1037/0033-295X.102.2.246 (Last accessed: 27.01.2020).

Mezghani, Emna; Ernesto Exposito, Khalil Drira, Marcos Da Silveira and Cédric Pruski (2015): A semantic big data platform for integrating heterogeneous wearable data in healthcare. Journal of Medical Systems 39, 185. https://doi.org/10.1007/s10916-015-0344-x (Last accessed: 27.01.2020).

Muaaz, Muhammad and René Mayrhofer (2017): Smartphone-based gait recognition: From authentication to imitation. IEEE Transactions on Mobile Computing 16 (11), 3209-3221. https://doi.org/10.1109/TMC.2017.2686855 (Last accessed: 27.01.2020).

Naor, Moni and Moti Yung (1989): Universal one-way hash functions and their cryptographic applications. STOC '89 Proceedings of the twenty-first annual ACM symposium on Theory of computing. Seattle, Washington (USA), ACM, 33-43. https://doi.org/10.1145/73007.73011 (Last accessed: 27.01.2020).

Narayanan, Shrikanth and Panayiotis G. Georgiou (2013): Behavioral signal processing: Deriving human behavioral informatics from speech and language. Proceedings of the IEEE 101(5), 1203–1233. https://doi.org/10.1109/JPROC.2012.2236291 (Last accessed: 27.01.2020).

Ochoa, Carlos; Carlos Bort and Josep Miquel Porcar (2018): Who is who with behavioural data. An algorithm to attribute the device's navigation to users sharing the same device. Frontiers of Marketing Data Science Journal 1, 43-62.

Peake, Jonathan M.; Graham Kerr and John P. Sullivan (2018): A critical review of consumer wearables, mobile applications, and equipment for providing biofeedback, monitoring stress, and sleep in physically active populations. Frontiers in Physiology 9, 743. https://doi.org/10.3389/fphys.2018.00743 (Last accessed: 27.01.2020).

Prince, Stéphanie A; Kristi B. Adamo, Meghan E. Hamel, Jill Hardt, Sarah Connor Gorber and Mark Tremblay (2008): A comparison of direct versus self-report measures for assessing physical activity in adults: A systematic review. International Journal of Behavioral Nutrition and Physical Activity 5(1), 56. https://dx.doi.org/10.1186/1479-5868-5-56 (Last accessed: 27.01.2020).

Rammstedt, Beatrice; Constanze Beierlein, Elmar Brähler, Michael Eid, Johannes Hartig, Martin Kersting, Stefan Liebig, Josef Lukas, Anne-Kathrin Mayer, Natalja Menold, Jürgen Schupp and Erich Weichselgartner (2014): Qualitätsstandards zur Entwicklung, Anwendung und Bewertung von Messinstrumenten in der sozialwissenschaftlichen Umfrageforschung. Schmollers Jahrbuch – Zeitschrift für Wirtschaft- und Sozialwissenschaften 134(4), 517-546. Also published as RatSWD Working Paper (RatSWD Working Paper 230/2014). English version available: https://www.ratswd.de/dl/RatSWD_WP_245.pdf (Last accessed: 21.04.2020).

Ranjan, Juhi and Kamin Whitehouse (2015): Object hallmarks: Identifying object users using wearable wrist sensors. Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing. Osaka, ACM Press, 51-61. https://doi.org/10.1145/2750858.2804263 (Last accessed: 27.01.2020).

Ranjan, Yatharth; Zulqarnain Rashid, Callum Stewart, Pauline Conde, Mark Begale, Denny Verbeeck, Sebastian Boettcher, The Hyve, Richard Dobson, Amos Folarin and The RADAR-CNS Consortium (2019): RADAR-base: An open source mHealth platform for collecting, monitoring and analyzing data using sensors, wearables, and mobile devices. Journal of Medical Internet Research mHealth and uHealth 7(8), e11734. https://doi.org/10.2196/11734 (Last accessed: 27.01.2020).

RatSWD [German Data Forum] (2020): Big data in social, behavioural, and economic sciences. Data access and research data management. RatSWD Output 4 (6). Berlin, German Data Forum (RatSWD). https://doi.org/10.17620/02671.52 (forthcoming June 2020).

RfII [Rat für Informationsinfrastrukturen] (2019): Herausforderung Datenqualität – Empfehlungen zur Zukunftsfähigkeit von Forschung im digitalen Wandel. Göttingen, Rat für Informationsinfrastrukturen. http://www.rfii.de/?p=4043 (Last accessed: 27.01.2020). English version available: http://www.rfii.de/?p=4203 (Last accessed: 21.04.2020).

Robbins, Megan L. (2017): Practical suggestions for legal and ethical concerns with social environment sampling methods. Social Psychological and Personality Science 8(5), 573-580. https://doi.org/10.1177/1948550617699253 (Last accessed: 27.01.2020).

Rocher, Luc; Julien M. Hendrickx and Yves-Alexandre de Montjoye (2019): Estimating the success of re-identifications in incomplete datasets using generative models. Nature Communications 10(1), 3069. https://doi.org/10.1038/s41467-019-10933-3 (Last accessed: 27.01.2020).

Salles, Gil F.; Claudia R. L. Cardoso and Elizabeth S. Muxfeldt (2008): Prognostic influence of office and ambulatory blood pressures in resistant hypertension. Archives of Internal Medicine, 168(21), 2340-2346. https://doi.org/10.1001/archinte.168.21.2340 (Last accessed: 27.01.2020).

Schmidt, Philip; Attila Reiss, Robert Duerichen and Kristof Van Laerhoven (2018): Wearable affect and stress recognition: A review. https://arxiv.org/abs/1811.08854 (Last accessed: 27.01.2020).

Schwarz, Nobert (2012): Why researchers should think "real-time": A cognitive rationale. In: Matthias R. Mehl and Tamlin S. Connor (Hrsg.): Handbook of Research Methods for Studying Daily Life. New York, Guilford Press, 22-42.

Schupp, Jürgen and Christof Wolf (eds.) (2015): Nonresponse Bias: Qualitätssicherung sozialwissenschaftlicher Umfragen. Berlin, Springer-Verlag.

Shcherbina, Anna; C. Mikael Mattsson, Daryl Waggott, Heidi Salisbury, Jeffrey W. Christle, Trevor Hastie, Matthew T. Wheeler and Euan A. Ashley (2017): Accuracy in wrist-worn, sensor-based measurements of heart rate and energy expenditure in diverse cohort. Journal of Personalized Medicine 7(2), 3. https://doi.org/10.3390/jpm7020003 (Last accessed: 27.01.2020).

Shrout, Patrick E.; Gertraud Stadler, Sean P. Lane, M. Joy McClure, Grace L. Jackson, Frederick D. Clavél, Masumi Iida, Marci E. J. Gleason, Joy H. Xu and Niall Bolger (2018): Initial elevation bias in subjective reports. Proceedings of the National Academy of Sciences 115(1), E15-E23. https://doi.org/10.1073/pnas.1712277115 (Last accessed: 27.01.2020).

Sinibaldi, Jennifer; Mark Trappmann and Frauke Kreuter (2014): Which is the better investment for nonresponse adjustment: Purchasing commercial auxiliary data or collecting interviewer observations? Public Opinion Quarterly 78(2), 440–473. https://doi.org/10.1093/poq/nfu003 (Last accessed: 27.01.2020).

Stergiou, George; Anastasios Kollias, Gianfranco Parati and Eoin O'Brien (2018): Office blood pressure measaurement. The weak cornerstone of hypertension diagnosis. Hypertension 71(5), 813-815. https://doi.org/10.1161/HYPERTENSIONAHA.118.10850 (Last accessed: 27.01.2020).

Stone, Arthur; Joan Broderick, Joseph Schwartz, Saul Shiffman, Leighann Litcher-Kelly and Pamela Calvanese (2003): Intensive momentary reporting of pain with an electronic diary: reactivity, compliance, and patient satisfaction. Pain 104(1-2), 343-351. https://doi.org/10.1016/S0304-3959(03)00040-X (Last accessed: 27.01.2020).

Strübing, Jörg; Stefan Hirschauer, Ruth Ayaß, Uwe Krähnke and Thomas Scheffer (2018): Gütekriterien qualitativer Sozialforschung. Ein Diskussionsanstoß. Zeitschrift für Soziologie 47(2), 83-100. https://doi.org/10.1515/zfsoz-2018-1006 (Last accessed: 27.01.2020).

Tost, Heike; Markus Reichert, Urs Braun, Iris Reinhard, Robin Peters, Sven Lautenbach, Andreas Hoell, Emanuel Schwarz, Ulrich W. Ebner-Priemer, Alexander Zipf and Andreas Meyer-Lindenberg (2019): Brain circuitry underlying the real-life effect of urban green space on emotional well-being. Nature Neuroscience 22(9), 1389-1393. https://doi.org/10.1038/s41593-019-0451-y (Last accessed: 27.01.2020).

Trull, Timothy J. and Ulrich W. Ebner-Priemer (2013): Ambulatory assessment. Annual review of clinical psychology 9, 151-176. https://doi.org/10.1146/annurev-clinpsy-050212-185510 (Last accessed: 27.01.2020).

West, Brady T. (2013): An examination of the quality and utility of interviewer observations in the National Survey of Family Growth. Journal of the Royal Statistical Society: Series A. Statistics in Society 176(1), 211-225. https://doi.org/10.1111/j.1467-985X.2012.01038.x (Last accessed: 27.01.2020).

Wyatt, Danny; Tanzeem Choudhury, Jeff Bilmes and James A. Kitts (2011): Inferring colocation and conversation networks from privacy-sensitive audio with implications for computational social science. ACM Transactions on Intelligent Systems and Technology (TIST) 2(1), 7. https://doi.org/10.1145/1889681.1889688 (Last accessed: 27.01.2020).

Yetisen, Ali K.; Juan Leonardo Martinez - Hurtado, Barış Ünal, Ali Khademhosseini and Haider Butt (2018): Wearables in medicine. Advanced Materials 30(33), 1706910. https://doi.org/10.1002/adma.201706910 (Last accessed: 27.01.2020).

Contributors to the report

<u>Members of the working group 'Data Collection with New Information Technology'</u>

Prof. Dr. Thomas K. Bauer
Ruhr University Bochum, RWI – Leibniz Institute for Economic Research

Prof. Dr. Ulrich Ebner-Priemer
Karlsruhe Institute of Technology (KIT)

Prof. Dr. Michael Eid, *(Chair)*
Free University Berlin (FU)

Prof. Dr. Anja S. Göritz
University of Freiburg

Dr. Cornelia Lange
Robert Koch Institute (RKI)

Prof. Dr. Kai Maaz
Goethe University Frankfurt, DIPF | Leibniz Institute for Research and Information in Education

Elke Nagel
Federal Statistical Office (Destatis)

Bertram Raum
Expert on data protection

Dr. David Richter
Socio-Economic Panel (SOEP) at the German Institute for Economic Research (DIW Berlin)

Prof. Dr. Mark Trappmann
Institute for Employment Research (IAB) of the German Federal Employment Agency (BA),
University of Bamberg

<u>Consultation</u>

Prof. Dr. Bernad Batinic
Johannes Kepler University Linz

Prof. Dr. Peter Dabrock
German Ethics Council

Prof. Dr. Eco de Geus
Vrije Universiteit Amsterdam

Prof. Dr. Björn Eskofier
Friedrich–Alexander-University Erlangen–Nürnberg

Prof. Matthias Mehl, Ph.D.
University of Arizona

Prof. Dr. Johannes Schöning
University of Bremen

Prof. Kristof Van Laerhoven, Ph.D.
University of Siegen

Prof. Dr. Cornelia Wrzus
University of Heidelberg

<u>German Data Forum (RatSWD) business office</u>

Dr. Mathias Bug
Dr. Tim Deeken
Dr. Nora Dörrenbächer

Imprint

■ Established in 2004, the **German Data Forum** (Rat für Sozial- und Wirtschaftsdaten, RatSWD) is an independent council. It advises the German federal government and the federal states (Länder) in matters concerning the research data infrastructure for the empirical social, behavioural, and economic sciences. The German Data Forum (RatSWD) has 16 members. Membership consists of eight elected representatives of the social, behavioural, and economic sciences and eight appointed representatives of Germany's most important data producers.

The German Data Forum (RatSWD) offers a forum for dialogue between researchers and data producers, who jointly issue recommendations and position papers. The council furthers the development of a research infrastructure that provides researchers with flexible and secure access to a broad range of data. The German Data Forum (RatSWD) has accredited 38 research data centres (as of May 2020) and fosters their interaction and collaboration.

www.ratswd.de