Changes, challenges and transitions over time in collecting time use data

Joeri Minnen • Theun Pieter van Tienoven

Time reveals the "many interesting patterns of social life [that] are associated with the temporal distribution of human activities, with the regularities in their timing, duration, frequency, and sequential order" (Szalai, 1972, p. 1). These elements – that is, *timing*, *duration*, *tempo*, and *sequence* – are often referred to as the parameters of time (Zerubavel, 1982) and all data-collection techniques that gather information about at least one of these four parameters are referred to as "time use studies". The time diary methodology (often referred to as time use surveys) is capable of capturing all four parameters of time at once and is therefore believed to be one of the most profound and valuable ways to capture human behaviour. Time use surveys draw a picture of the ways individuals use their time by using a log or a time diary during at least twenty-four consecutive hours (Pronovost, 1989).

From the outset of time diary studies the focus laid on socio-economic issues and what started as small observational diary studies quickly grew into international comparative studies. Along with this, the number of stakeholders also increased: from individual, pioneering scholars to large research groups (e.g., the Centre for Time Use Research – CTUR, Tempus Omnia Revelat – TOR, and the Maryland Time Use Laboratory), the International Association for Time Use Research (IATUR), and from pioneering policy planners such as Stanislav Strumilin to a coordinated network of National Statistical Institutions.

While there is a great deal of consensus about the usefulness and qualities of the time diary *methodology*, there was less consensus about (the design of) the *method* by which and the way in which (i.e., *mode*) data is collected. The complexity of the diary methodology in terms of the principles of the European Statistics Code of Practice (e.g., respondent burden, cost efficiency, accuracy and reliability, and timeliness and punctuality) (Eurostat, 2018) plays a major

role in this. Concessions and choices in methods and modes to meet these principles can often be traced back to discussions about the consequences of certain choices for the reliability and validity of the collected data. With these discussions in mind, this chapter looks at the past, present and future of time use research and aims to identify five eras of time use research characterised by distinctive changes, challenges and transitions in the way time use research is conducted.

To better understand the distinctive characteristics of the different eras of time use research, we need to introduce two concepts that are relevant in all research methodologies but in particular in time use research: reliability and validity.

Reliability and validity

Time use research is a quasi-observational research method in which the respondents' own observations approximate the ways they spent their time (Juster, 1986, pp. 398-399). The discrepancy between their actual time expenditure and their estimated time expenditure is the measurement error. Regarding time use research, it is, in the words of Scheuch,

> very hard to collect answers that correspond to reality with at least some degree of accuracy. Representing the expenditure of time is one of those subject matters where the reliability and validity of data are extremely sensitive to details in the manner of data collection (1972, p. 69).

The total measurement error therefore relates to the reliability or unsystematic error and validity or systematic error of the measurement method.

Reliability

The reliability of a research method presents itself in different forms. A first form of reliability refers to the ability of a measurement instrument to produce comparable results for various samples with the same characteristics. Reliability in this sense thus refers to the stability of the measurement. The random error is then not directly linked to the method per se but relies on erroneous observations by the respondents that would not repeat itself in the same way under the same circumstances.

A second form of reliability is typical of time use research and refers to random errors that are related to the number of observed days or the observation

length of the study. Suppose time use research that consists of a single diary day. The reliability of the sleep time estimates will be high (i.e., result in fewer random errors) because sleep is highly likely to occur every day. However, the reliability of time spent on cultural participation will be low because cultural participation is much less likely to follow a daily rhythm of occurrence. The unsystematic errors are then related to zero-observations (Gershuny, 2012).

The size of the random error then depends not so much on the research method itself, but on the research design, since, in addition to the number of diary days, the reliability of the measurement is also influenced by the length of the fieldwork period (e.g., the inclusion of seasons), the distribution over different days (e.g., weekdays and weekend days), and the sample size (Harvey, 1993).

Validity

Validity subdivides into external and internal validity (te Braak, van Droogenbroeck, Minnen, van Tienoven, & Glorieux, 2022). External validity refers to the generalisability of the research results to the population from which the sample is drawn and the ability to draw conclusions about the real-world implications. Since external validity deals with ensuring that the results are not only applicable to the participating group of respondents, dealing with external validity relates to a large extent to sampling strategy (i.e., size, characteristics).

Internal validity focuses on the research method itself and the effect of the measurement tools on the collection of the research data and subsequently the results. Internal validity, thus, refers to the ability of a measurement instrument to produce accurate or valid data. A systematic error occurs when the measurement instrument does not measure what it is intended to measure.

Over the years, different research methods have been developed and employed to measure how people spend their time. Most common are

(1) the diary method in which respondents keep a time diary in close to real time and record their activities chronologically for at least 24 hours;
(2) the yesterday-recall method in which respondents are interviewed about their time expenditure on the day before over a period of 24 hours; and
(3) the beeper-method in which respondents are beeped several times throughout the course of a day to record what they have been doing, often in the past hour.

In the absence of an objective evaluation method for the internal validity of these research methods, the internal validity of the different methods is often

evaluated in relation to each other. To do this, face validity and content validity are often used. Face validity refers to the subjective acceptance and judgement of the credibility of the research method and is often a useful first indicator to assess internal validity. Content validity refers to the comprehensiveness of the research method by representing all the relevant dimensions of the construct being measured. To assess the internal validity of a diary method, several indicators have been developed (Juster, 1986). These indicators are constructed using the collected diary data, hence *content* validity. One diary method is less valid than the other diary method when respondents (1) record fewer primary activities, (2) record fewer secondary activities, (3) describe time spent using a limited number of different activities, (4) record more unspecified time, and (5) have more activities start and end at round times (Juster, 1986). These indicators are useful but should not be considered conclusive. As Scheuch points out, there is "no absolute measure to judge the results of any of the techniques used during the pre-test as 'true'. […] No one technique known to us will result in 'perfect' data" (1972, p. 74).

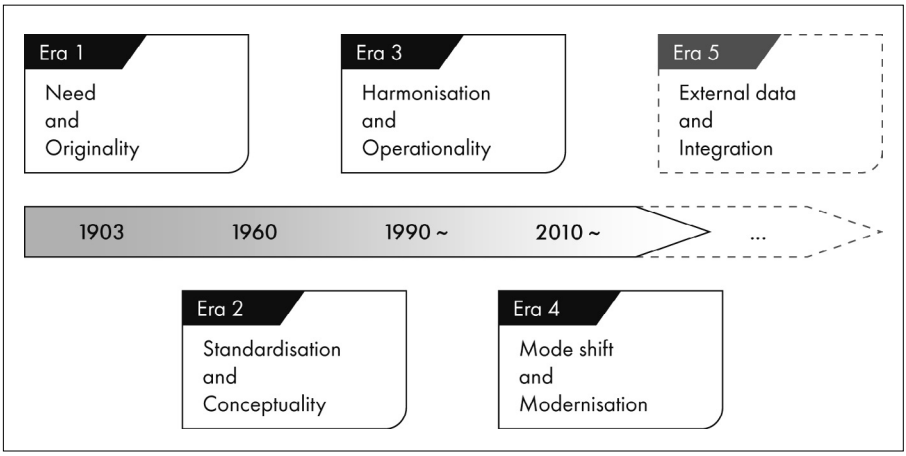
Reliability and validity studies between different research methods are scarce, often only apply to relatively small samples, and are seldom conclusive. Diary methods seem to have the advantage over survey methods (Bonke, 2005; Gersbuny & Sullivan, 1998; Robinson & Godbey, 1997), although occasionally survey estimates concur with time spent on activities in a diary (Frazis & Stewart, 2010). Similarly, a combination of a "yesterday survey" method for day one and a diary method for days two to seven to produce weekly time estimates led Huysmans, Lammers, and Wester (1997) to conclude that there was no difference between the two methods for their study topic (i.e., media use). However, any evaluation of reliability and validity should always be interpreted with a certain caution because "ultimately, we cannot provide conclusive proof of the validity of our results, but their technical reliability can be tested, at least in part" (Fisher, Gershuny, Gauthier, & Victorino, 2000, p. 20).

Eras of time use research

As a research discipline, time use research appears to be dynamic, which allows for both different research questions and changes in the way of doing time use research over time. Historically, it is possible to speak of five eras of time use research (Figure 1). The first era arises from the need for time use research and is characterised by the originality of the time use research methodology. The second era is characterised by standardisation of the methodology and the conceptuality of what constitutes the diary method. In the third era, the

harmonisation of time use research is central in function of an upscaling of the operationalisation of time use research. The fourth era is the era which the research community is now largely in and is characterised by a change in the way (i.e., mode) in which time use research is conducted in response to the need to modernise. The fifth era is seen as the next evolving stage where external data are captured and these different data streams are integrated into the overall data-collection strategy. In this section we discuss the first four eras. The discussion of the fifth age – the future of time use research – concludes this chapter.

Figure 1. Five eras of time use research



First era: need and originality

At the start of the twentieth century, the need arose to collect reliable data on the relationship between paid work, on the one hand, and unpaid work and leisure, on the other. Before that, Frédérique Le Play (1877) published *Les ouvriers européen*, which reported on 57 time reports from workers and their families from various industries across Europe and focused on paid work and how the family as a unit provided an income based on the number of hours worked. However, the origin of the time use research methodology based on diaries can be attributed to the books *How working men spend their time* by George Bevans (1913) and *Round about a pound a week* by Maud Pember-Reeves (1913). Both studies question social (in)justice. Bevans was interested in the way leisure time was spent by men from four different industrial cities and how their working conditions affected their leisure time. Feminist Pember-Reeves documented the life of working-class families in London from the perspective of the family and

the woman in that family. Her work can be said to have initiated the use of time use research to study the gender division of labour at the household level.

However, it is the USSR that validated the diary method by using it to collect data on a representative scale for planning purposes and to make comparisons between groups of industrial workers. In the communist USSR, under the leadership of Strumilin (1921-1923), time diaries were used to design and assess economic and social planning (Zuzanek, 1980). America followed later (1925-1931), with the Department of Agriculture (USDA) asking women to keep time diaries to chart work in the field and beyond (Stinson, 1999, pp. 12-14). Later, the academic world also jumped on the bandwagon of the diary method.Sorokin, who was a former student of Strumilin, published, for example, "Social time: A methodological and functional analysis" (Sorokin & Merton, 1937) and *Time-budgets of human behavior* (Sorokin & Berger, 1939), which made the link between theory of time and empirical data.

The reliability and validity of the diary method continued to be valued after the Second World War. Once again, the Russians were the first to move forward with time use research. More than a hundred studies were carried out between 1958 and 1968 (Pronovost, 1989, p. 74). At about the same time, public media companies in the United Kingdom, the United States and Japan started using the diary method to collect data on media use and thus chart emerging activities such as listening to the radio and watching TV (Robinson & Converse, 1972).

The methodology of time use research and the diary method to collect data in a reliable and valid way have secured their place in policy and scientific research.

Era 2: Standardisation and conceptuality

The real international breakthrough for time use research, though, came from the *Multinational Comparative Time-Budget Research Project*, funded by UNESCO and coordinated by Alexander Szalai. Between 1964 and 1966, respondents in twelve medium-sized cities in different European countries were surveyed using the same diary format. The underlying goal was to understand and reduce the divide between Western European and Eastern European countries. The details and results of this study are described in the so-called "bible of time use research", *The use of time* (Szalai, 1972). The then-current decisions about the format of the diary and the organisation of the fieldwork still have great resonance in contemporary time use research (Minnen & Glorieux, 2011).

For this project to succeed, the methodology of time use research had to be standardised to yield comparable results. This standardisation entailed the acceptance of the diary methodology as the most reliable and most valid methodology to capture the ways in which people spend their time. The diary methodology consists of a chronological record of daily activities and their context such that daily life can be described in terms of the timing, duration, tempo, and sequences of actions. The American Sociological Association had the same providence and founded the Task Force on Time Budget Research in 1965 with the aim of new and, above all, comparable data collections. One of the results was the American Heritage Time Use Study (AHTUS) led by John P. Robinson, which provided a detailed understanding of the way Americans use their time. It showed that Americans generally spend significantly more time on work-related activities than measured by stylised questionnaires, but also that those who work less than average underreport and those who work more than average overreport in the time diaries.

Although the diary methodology became standardised, the methods varied. Szalai used the time diary method, while AHTUS used the yesterday recall method or, similarly, the daily reconstruction method. Yet others used the beeper method. With this arsenal of methods available, the question of the most reliable and (internally) valid method must be assessed in the context of the research question. A choice between methods is then based on the choice between intra- and inter-person variation, more participants versus more diary days per participant, and the need for detail (activities and spatial and social context). It would appear that when the goal is to collect data with a focus on inter-subject variation and comparisons between groups of respondents, continuous recording seems more appropriate, while with the focus on within-person variation and a broader activity definition, daily recall records are likely to be a better strategy.

Era 3: Harmonisation and operationality

The success of the diary methodology in producing reliable and valid estimates of how people spend their time led the United Nations to popularise time use research in the 1980s. A clear added value of this research at that time was that it could provide a picture of invisible and largely undervalued unpaid work (Juster & Stafford, 1991, p. 472; Robinson & Godbey, 1997, p. 97). The diary methodology allows unpaid work to be included in the System of National Accounts. This made visible the contribution of women to the economic development of societies (Gershuny, 2003; Juster & Stafford, 1991).

For this to happen, time use research needed to be operationalised at a much larger scale and harmonised across different countries. Europe took a leading position in the *pre*-harmonisation of time use research. Under the leadership of Eurostat and after a decade of debates and decision-making, the guidelines on Harmonized European Time Use Surveys (HETUS) were formulated (Eurostat, 2020). These guidelines include instructions for the diary method (e.g., activity coding list, interval of time recording, number and assignment of diary days, length of observation period) but also on the construction of the sample selection, the training of interviewers, and data entering and cleaning. Eurostat promoted time use research in its member states and associated countries, which resulted in comparable datasets of 20 countries, several of which are available in different waves.

The success of time use research also triggered the interest of academics. More and more academics started taking up the diary method to analyse a wide variety of social and economic issues. This led to a plea for more internationally comparable time use data – including those beyond Europe (Harvey, 1993) – which also fuelled a *post*-harmonisation project. This project has largely been realised by the Centre for Time Use Research (CTUR) and resulted in an open-access database of Multinational Time Use Study (MTUS), containing harmonised time use data across 30 countries in over 70 different waves all over the world (Fisher, Gershuny, & Gauthier, 2012).

Through guidelines and international collaboration, time use research became a reliable and standardised method for making valid comparisons between countries and regions and studying trends over time.

Era 4: Mode shift and modernisation

It is safe to say that this is the current era of time use research. It signifies a change in the mode of collecting time use data under the wings of the process of modernisation. This process is not only fuelled by technological developments and the rise in use of ICT in an increasingly connected world but is also seen as an answer to the waning willingness to participate and the high costs of time use research (Minnen, Rymenants, Glorieux, & van Tienoven, 2023). Essentially, this era begins by progressing through the previous three eras, but at a much faster pace – not only at the level of time use methodology, but also at the level of a mode shift in the way time use data are collected. First, there is a need for modern, connected tools and platforms that are again subject to the question of whether they lead to reliable and valid data. Second, the standardisation and harmonisation of the diary method is questioned. Are these tools and platforms a literal translation of the paper-and-pen mode of data collection or not? And

if not, how do data collections remain comparable? The answers lie in concepts such as flexibility, modularity, reusability, and shareability. Third – and this is relatively unique to this era – how are privacy and security handled?

Need for new tools and platforms

At this point, the standardised and harmonised diary method consists of an interviewer conducting household and individual surveys and leaving behind two paper time diaries per eligible household member with the dates on which both time diaries must be completed. One diary concerns a weekday and another diary concerns a weekend day. All eligible household members need to complete the same days. The interviewer also leaves behind a drop-off questionnaire, which is to be completed by all eligible household members after the time diaries. At a prearranged date, the interviewer returns to check and collect the time diaries and the drop-off questionnaire. The paper-and-pencil mode of data collection has at least three downsides: (1) it is very expensive due to interview visits, printing costs, and data entering and coding costs; (2) it is very burdensome for respondents to participate in, and (3) processing the data is time-consuming and no longer answers to the quest of understanding the rapidly changing context of modern societies (Cai & Zhu, 2015).

As time use research yields highly valid and reliable data, rich in information and contexts, and allows regions, countries and cultures to be compared, the momentum of the "Big Data" challenged researchers to modernise and digitise time use research to collect data in increasing volumes with greater speed and more variation. Eurostat is again an accelerator in promoting this modernisation, showing their ambition in the DGINS Wiesbaden Memorandum 2011. This Memorandum is grounded in Eurostat's responsibility to provide reliable, valid, and comparable statistical information to the institutions of the European Union (E.U.). An important tool to achieve this is to promote – as far as possible – harmonised statistical methods in all Member States. The Memorandum emphasises the need for better data, in terms of coverage and comparability. The process of modernising the production of official statistics should contribute to this by a mode shift from paper-and-pencil to online data collection with the aim of (1) improving the responsiveness of respondents, (2) better integrating new ways of data collection and new sources of information, and (3) collecting data more efficiently.

Bonke and Fallesen (2010) were among the first to develop a prototype for data collection via a web interface, with funding from the Rockwool Foundation in Denmark. In addition to being online, the diary featured a search tag selectable pre-coded list and reported for a weekday and a weekend day divided

into ten-minute time slots. The first (native) mobile application on a smartphone was created in 2011 by the Netherlands Institute for Social Research with the HETUS guidelines in mind. The app was tested in a feasibility study with a representative sample of the adult Dutch population via the LISS panel in 2012 (Sonck & Fernee, 2013). The backbone of the application was reporting on pre-stated HETUS-based activities in ten-minute intervals over two fixed days (a weekday and a weekend day) along with context on location and mode of transport and social context of the activity. The app also got into reality mining by using prompts to ask questions about mood at different times of the day.

In 2012 and after receiving a Hercules grant, the Research Group TOR of the Vrije Universiteit Brussel also translated the pencil-and-paper method into a web-based environment. In 2013, the MOTUS web application was tested in a yearlong large-scale data collection parallel to the then ongoing Belgian Time Use Survey (BTUS). This concurrency made it possible to differentiate between two designs and modes: a pencil-and-paper two-day diary, ten-minute intervals, post-coding, and household mapping (BTUS), on the one hand, and an online seven-day diary, continuous time recording, pre-coding and individual sampling, on the other hand. Compared to the Danish and Dutch development, MOTUS also takes into account the full ecosystem of a time use survey where respondents are invited to complete their questionnaires and diaries in the same application without the help or support of an interviewer because all communication is managed by the platform (Minnen et al., 2014). These strengths were further embodied in updates and new releases of MOTUS. MOTUS currently combines a web application with a mobile application. The big difference to the Danish and Dutch application is that MOTUS is a platform instead of a native application. The platform currently consists of a back office where studies are designed and a front office where studies are conducted.

More recently, the Centre for Time Use Research of the London School of Economics (at the time located at Oxford University) and the Time Use Laboratory at the University of Maryland developed and released their online implementation. CTUR's web-based CaDDI tool takes its name from its Click-and-Drag principle of reporting an activity by "dragging a pointer across a horizontal timeline bar to create a record of the duration" (Sullivan, Gershuny, Sevilla, Walthery, & Vega-Rapun, 2020, p. 8). The activity list to choose from is a light version with broader categories than known from the HETUS activity list, but the context questions are similar. These context details populate only after all the primary activities throughout a day have been registered instead of per primary activity, as in the previous applications. In line with the HETUS

guidelines, recording is done in ten-minute intervals on a weekday and a weekend day.

The Time Use Laboratory starts from the daily reconstruction method where, as the name suggests, respondents are asked to reconstruct their previous day. They use mytimeuse.com for this, which is a responsive web application developed with a grant from the National Science Foundation (Rinderknecht, Doan, & Sayer, 2022). Activities are recorded continuously (i.e., not in intervals but with exact start and end times) and include a primary activity and the associated context of a secondary activity (as a percentage of the primary activity) as well as where the time was spent and with whom. The context is further expanded by also scaling emotions when doing the activity. The selectable activities are presented to the respondent as they type the activity into a search field. Respondents may keep their own wording as a custom activity, which will be saved to the list for later use; but they must relate it to an already existing activity from the list.

At the University of New England in Australia, Michael Bittman continued to work with the beeper method. The "Intensive Hour" app or Random Time Sampling (RTS) method "beeps" respondents at random times, asking them to reconstruct the past hour into ten six-minute entries (Wong et al., 2022). The advantage of this method is that it is less invasive and allows, in particular, to survey time spent on paid work in more detail. The HETUS guidelines examine paid work only in broad, generic categories to prevent participation in time use research from leading to conflicts in the workplace. In addition, recording randomised, beeped hours of work time prevents sensitive company information from being obtained or individuals from being monitored.

Another application worth mentioning is the i-log application from the University of Trento, available for smartphones with Android and recently also with iOS as an operating system (Zeni, Bison, Reis, Gauckler, & Giunchiglia, 2020). This application aims to collect sensor streams in a passive way (i.e., mainly to determine the position of the device but also to capture para-data) in addition to asking respondents to fill in a small questionnaire and then a time diary for a weekday and a weekend day. Activities are selected from a pre-coded activity list and context recording is similar to the HETUS guidelines. No additional activities can be registered. As with the Danish, Dutch, Belgian and American applications, each episode repeats the same logic of questions over and over. The application started as part of an EU Horizon project and was used, among other contexts, as part of the Big Data Hackathon 2019 in Brussels.

There are many more applications that entered a development path, but many did not reach the pilot or test phase. An overview can be found via the

Eurostat inventory.[1] More importantly, all these different applications mean that the internal validity and comparability of the diary method have once again been called into question. While the method itself has remained relatively standardised and harmonised (i.e., activity lists, recording intervals, context query, diary days), the mode is anything but. To harmonise this, the focus must be on the modularity and shareability of the modes.

Modularity and shareability

Harmonising the outcome of the modernisation of time use research (and other surveys) is part of the European Statistical System (ESS). While the Member States are responsible for collecting the data and compiling the statistics for national and EU knowledge building, Eurostat's role is to support the ESS to create networks that strive towards harmonised procedures. One way to guide this process in a standardised manner is to use a generic production architecture such as GSBPM. The GSBPM or Generic Statistical Business Process Model was developed by UNECE, Eurostat and OECD to provide a standardised overview of the way official statistics are produced (Kuonen & Loison, 2019). The GSBPM considers a total of eight phases, of which the design phase (phase 3), the construction phase (phase 4) and the process phase (phase 5) relate to the set-up and implementation of the data collection itself, including the handling of the data files. This also applies to time use research (Minnen et al., 2023).

As mentioned above, the need to modernise not only arises from the need to produce data more cheaply and faster, but also to visualise the ever faster-changing society. Consider, for example, mapping the effects of the COVID-19 pandemic on daily life (Gershuny et al., 2021; van Tienoven et al., 2023). The statistical process should therefore be sufficiently modular in design to meet country-specific requirements but at the same time be sufficiently harmonised to ensure comparability (Glorieux & Minnen, 2009; Salgado, Esteban, Novás, Saldaña, & Sanguiao, 2018; Stodden, 2014).

This underlines the importance of using platforms. Platforms are more supportive to tailor the research design to the needs of the research question, while native or one-off applications are quite rigid about making adjustments in favour of the setting. It is the task of the GSBPM to then communicate the opportunities for designing, building, and implementing the data collection and the means of processing the collected data in a standardised way.

1 Accessible through https://webgate.ec.europa.eu/fpfis/wikis/display/ISTLCS/TUS+TOOLS+MENU.

However, this is only half the story, because in order to support standardisation it is also important that these platforms can be shared. Shareability here refers to the ability of a platform to be integrated into a data-collection architecture and/or process of another institution (e.g., NSI, academic institution, …). The simpler the implementation, the more easily the environment can be shared. A platform that is modular and has a high degree of (re)use, sharing and collaboration also has minimal development and maintenance costs.

Shared platforms mapped to a business architecture such as the GSBPM combine the power of modularity (internal validity) and the power of standardisation (reliability). The result is an upgrade of the time use research toolbox to configure the right approach to collect the best fit data for a given research problem within the same IT environment. Evolved modernisation and digitisation provide the opportunity to embed differences into a stronger and interoperable validation of time use practices.

Privacy and security

Privacy and data security are not new concerns, but the modernisation of data collection based on technological progress and digitisation has brought them explicitly to the fore. Particularly in the light of the decreasing willingness of the population to participate, data collectors can no longer act overnight. When collecting data, personal data are essential, visible and stored in databases. In addition, the output of the participation is a detailed collection of activities timestamped and contextualised with additional personal information.

This is even more true when modernisation also means that external databases can be linked or when passive data registration, for example via sensors, becomes part of a collection process. As well-intentioned as it is to reduce the registration burden on the respondent and to increase the ease of use of the applications, concerns about privacy and security must first be dispelled. Although hard work can be done in the background on all kinds of documentation about privacy statements, data-protection impact assessments, data management plans, and data protection policies, the main challenge remains to gain and maintain the trust of the respondents (Keusch, Struminskaya, Antoun, Couper, & Kreuter, 2019; Revilla, Couper, & Ochoa, 2019; Ricciato, Wirthmann, & Hahn, 2020).

One way to do this is to shift from "privacy by legal" to "privacy by design". For data-collection platforms, this means that privacy and data security are taken into account from the very beginning of the platform's design. This involves matters such as pseudonymisation, encryption, two-step authentication,

and limiting default positions (e.g., omitting unnecessary profile information), on the one hand, and, using ISO-certified servers, penetration tests, and load and performance tests on the other hand.

Privacy and security are an essential part of the development of the platforms, not only in the back office but also in the front office applications (web, mobile) to bolster participants' confidence to start and continue their respondent journey in collecting highly detailed data. Tools with a higher resolution of privacy and security will benefit from better response rates and data accuracy compared to tools that only declare privacy and security from legal documentation.

MOTUS

Our added value to the field of time use research lies most profoundly in development of the MOTUS data-collection platform. MOTUS stands for Modular Online Time Use Survey and the development builds on a long-standing history in time use research at the Research Group TOR of the Vrije Universiteit Brussel. After having conducted multiple pencil-and-paper time use surveys (in 1984, 1988, 1999 and 2004) and having played an advisory role in the pencil-and-paper time use surveys of 1999, 2005 and 2013 of Statistics Belgium, TOR won a HERCULES grant to develop "An Infrastructure for a Continuous Modular Online Time Use Survey". It marked the start of the MOTUS project in 2012.

Unique to the MOTUS project was not only the digitisation of the pencil-and-paper method. The main aim was to translate all the accumulated knowledge about the design, implementation, and execution of time use studies and the known consequences of design and implementation choices for the reliability and validity of time use data into a platform that makes it possible to make different choices regarding different parameters. MOTUS did this (and still does) by using the concept of "modularity". It means that the front office application (i.e., the application that is used by respondents) is (largely) defined by the content and the different settings in the back office application (i.e., the application that is used by researchers). In other words, the front office application is not a rigid time diary tool but remains an "empty box" until linked to an active (time use) study that has been designed in the back office application. A major additional advantage is that the front office can host multiple studies, even in one and the same respondent.

To achieve such a platform, we identified (at least) four development challenges:

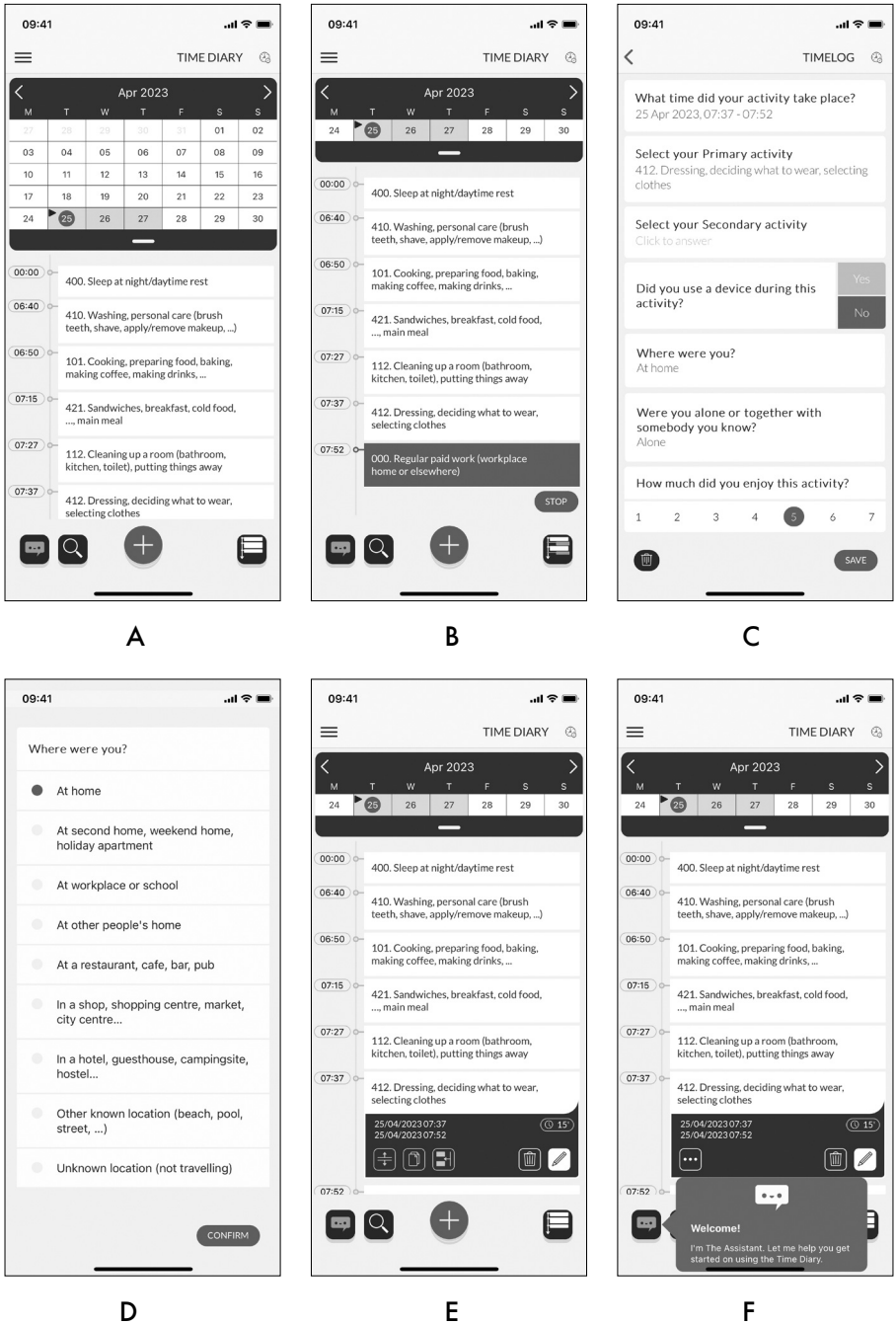
(1) the user interface (UI) and user experience (UX) design of a tool to collect data (i.e., the front office);
(2) the back office software or platform design to manage and organise data collections;
(3) the creation of a shareable architecture to run the tool and the platform with respect to privacy and security requirements; and
(4) the ability to connect the architecture to other environments.

MOTUS front office

The front office is the application for the respondent to participate in surveys, register data, consult data, and provide additional data. Through a well-balanced UI and UX of MOTUS, visual elements and functionalities, it supports a less burdensome task on the part of the respondent, for both the mobile and web application. The mobile application is available in the Appstore (iOS) and the Play Store (Android). The web application can be accessed via https://www.motusresearch.io using any conventional internet browser. Information is interchangeable between the two applications, while multiple devices can also be used to participate in the surveys.

Figure 2 shows some of the trademarks of MOTUS for the mobile application. Diaries start from a calendar that highlights the days that need to be recorded (Figure 2A). On a registration day, the application shows the timeline overview with, if selected, the activity that is currently tracked at the bottom (Figure 2B). The activity recording is sequential (Figure 2C) and allows for more detailed context questions (Figure 2D). Tapping on an activity unfolds additional options to edit the timeline, such as deleting, copying, splitting the activity, or inserting another activity (Figure 2E). Finally, the sequence of registering an activity is supported by the on-screen Assistant that can be toggled on or off in the bottom left corner (Figure 2F). The web application has the same look and feel as the mobile application to make it easier and more recognisable for respondents to switch between applications and devices. More visuals of the mobile and web application are available on https://www.motusresearch.io.

Figure 2. Visuals of some of the trademarks of the MOTUS application



MOTUS back office

The modular character of MOTUS is related to the flexibility with which investigations can be designed in the back office. For this MOTUS uses the concept of "builders". Each builder allows to shape certain elements of a study. The "survey builder" provides for the preparation of questionnaires, the "diary builder" for the preparation of surveys based on the diary method, such as time use research, but also household budget research. Surveys conducted at the household level and involving multiple members of the household (simultaneously) can use the "grid builder" that synchronises the research tasks of groups of respondents. The "communication builder" provides for setting up communication with the respondent (e.g., via email, via static information pages, or via notifications). The so-called "survey flow builder" brings everything together – for example, when a study consists of several sequential tasks (e.g., questionnaire and time diary).

Other builders support related processes. For example, there is a "translation builder" that allows to offer a survey in multiple languages and an "invitation builder" that manages respondents (i.e., import, invite). Data processes, such as real-time quality checks, are supported by the "R builder" using the `motusR` package. Finally, the "event builder" is under construction: it enables certain (passive) data streams to initiate tasks (cf., the beeper method).

A comprehensive overview of both the builders and the way they support the design, construction, and collection phase of the GSBPM is given in Minnen et al. (2023). Within the theme of reliability and (internal) validity of this chapter, we only go a little deeper into the "diary builder". After all, within this builder two important elements of time diary research are designed: the activity list and the parameters of the diary.

The activity list can be designed up to three levels deep. A first relevant element of the activity list is the selection of the activities. MOTUS offers respondents four options to select an activity: they can

(1) search for their activity in a tree structure of categories that expand to the lowest (selectable) detail;
(2) search for their activities using search terms they type in the search field. For this, search tags must be assigned to each activity in the back office;
(3) "star" activities so that they are available in a personal list of favourites and can be selected from there; or
(4) use a list of suggestions made by MOTUS based on previous entries and depending on time and place.

Which of the options is available to the respondent is indicated in the back office.

A second relevant element of the activity list is registering the context. MOTUS allows (in theory) a separate context questionnaire to be added to each activity. These questionnaires are created in the "survey builder". This shows an important advantage of digital time use research over pencil-and-paper. After all, in printed diaries, all context questions must be visible and it is not possible to vary them per activity (group). MOTUS allows, for example, the questions about transport modes to be displayed only when a displacement is registered, or no questions to be displayed when sleeping time is registered (cf., HETUS guidelines), or additional questions to be asked about the content when media-related activities are recorded.

Modularity also plays out at the level of the diary's parameters. We previously described that the choices about the number of days surveyed, the fieldwork period, the size of the registration intervals, and so on, (can) influence the reliability and (internal) validity of the collected data. MOTUS allows respondents to set several parameters in advance according to their own wishes and insights (see Table 1). These parameters are divided into the diary settings and closing criteria. Diary settings are the granularity of the recording, the length and the way it is calculated when the recording starts and how the 24-hour cycle is offered. There is also the option to allow a learning period. The diary starts, for example, at midnight, but it is available from 19:00 to enable the respondent to explore and practice. Closing criteria enable respondents to close the diary themselves or not. If this is conditional, then a number of quantitative criteria can be indicated, such as the extent to which indefinite time is allowed and whether there is a minimum number of registered activities. Quality criteria can also be indicated, such as a minimum of different activities, the mandatory registration of sleeping time, and the mandatory registration of some eating and/or drinking activities.

All in all, MOTUS therefore offers the opportunity to design time use research in a well-considered manner as a function of the concessions or requirements for the reliability and validity of the data. Moreover, by means of the back office and the underlying builders, MOTUS facilitates automated data collection without the intervention of an interviewer.

Table 1. Overview of adjustable time diary parameters in MOTUS

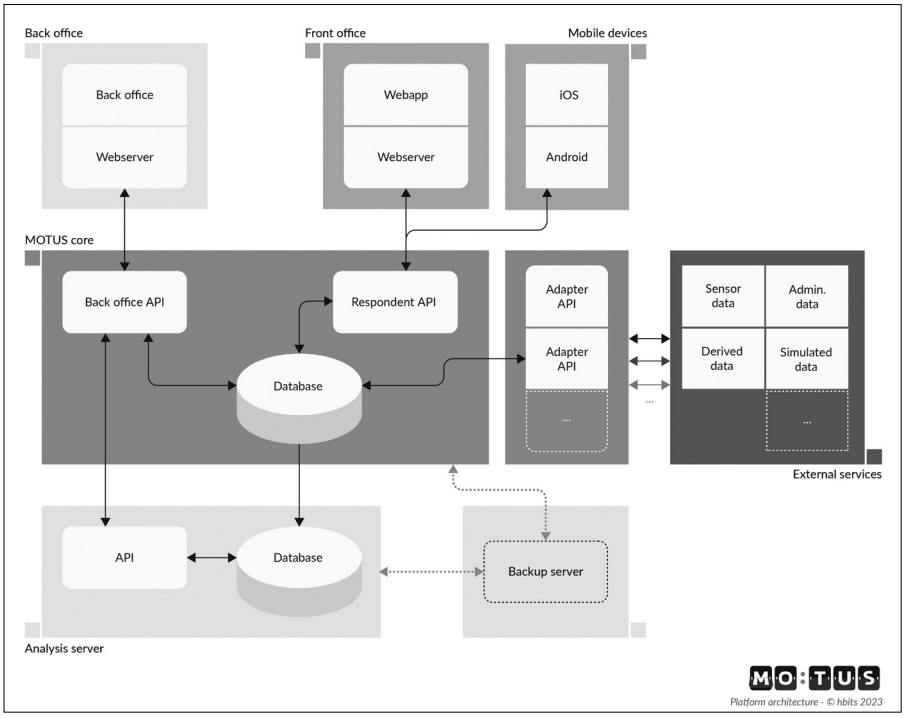
Diary settings		
Item	***Options***	***Details***
Granularity	Continuous	
	Fixed	5/10/15/20/30 minutes
		1/2/3/4 hours
Length	Week	
	Day	
	Custom	Specified as number of hours
Length calculation	Sum of logged time	
	Start of diary to end of last activity	
	Start of first activity to end of last activity	
Diary start	Immediately	
	Fixed	
	Retrospective	Define number of retrospective days
Diary cycle	Midnight to midnight	
	16:00 to 16:00	
	Custom	Specify start time
Learning cycle	None	
	Custom	Specified as number of hours
Closing criteria		
Item	***Options***	***Details***
Manual closing	Allowed	
	Not allowed	
	Allowed under conditions	
Quantity criteria	Undefined hours per day	Unlimited or specified as number of hours
	Number of activities per day	None or specified as number of activities
	Undefined hours per week	Unlimited or specified as number of hours
	Number of activities per week	None or specified as number of activities
Quality criteria	Number of different activities	None or specified as number of activities
	Duration of sleep	None or specified as number of hours
	Occasional eating and/or drinking	Required/not required

MOTUS architecture

With privacy by design in mind, the MOTUS architecture is separated into three levels (see Figure 3). The first level presents the web and mobile interfaces to the respondents (i.e., front office) and the web interface to the researchers and administrators (i.e., back office). Both the front office and the back office are connected to the MOTUS core via Application Programming Interfaces (APIs). The core holds the database with all information required to build a study and collect data. A separate analysis server holds a replica of the database from the

core and facilitates the processing of information in the back office. The back-up server is a replica of the core and analysis server.

Figure 3. Overview of the MOTUS platform architecture



Adapter APIs serve to adapt external information so that it can be processed in the core, enabling the inclusion of, for example, passive data collected from integrated sensors or connected devices, administrative or secondary data available from external data sources, or other processed data. For optimisation, data security, and privacy reasons, these data are handled and organised into standalone microservices.

The MOTUS architecture is set up using Docker containers. This enables installation within a data-collection environment of another institution in a simple, rapidly scalable, manner with certified and tested privacy and security. These four pillars characterise a good architecture and increase the divisibility of MOTUS so that it can grow into a modern and stable platform.

Era 5: external data and integration

We conclude with a look into the near future, because the fifth era that is coming is a result of the previous era where we arrive at the integration of different data streams. The fifth era will be centred on the apparent contradiction that, on the one hand, national and scientific institutions are finding it increasingly difficult to obtain permission and cooperation from respondents to collect data while, on the other hand, more and more behavioural data are being tracked via internet-connected devices and are even being combined by algorithms and artificial intelligence to gain a better understanding of (individual) human behaviour. The questions that are therefore central to this fifth era are: Who owns such external data? How can this data be integrated for the production of official statistics and/or scientific research? And, how can the introduction of external data reduce the registration burden of respondents be reduced?

Much of these types of data will be owned by market-oriented parties. They collect data through connected devices as a continuous stream of embedded personal data (Groves, 2011). These data are collected passively without the continuous and active participation of respondents. Today, however, most time use data are still collected through active participation, even when this is done through web and mobile applications developed for the domain of time use research. Looking back at what sets time use data apart from other data-gathering information about people's daily lives, it is the validity of the data. If we want to understand why people engage in an activity, we must focus on the "beneficial, immediate approach" (Hamermesh, 1999), which understands what people do in their temporal, spatial, social, and motivational contexts. A large measure of such observations can be supported by passive data streams such as perceptual (body), environmental, and even administrative data.

Sensor data

It is therefore expected that sensor data will play an important role in this fifth era. Through sensors, the state of respondents or their environment can be observed and measured, continuously or intermittently, or even at a defined level (e.g., entering geographic location, exceeding noise level, from a certain point in time). The variety of sensors is extensive and almost every physical element can be captured by temperature sensors, pressure sensors, proximity sensors, accelerometer and gyroscope sensors, humidity sensors, CO_2 sensors, and many more. If these sensors are connected to the internet, the output is available in real time and can be used as input for another system or used to

control a process of actions. It is this "If This Then That" (ITTT) application that is the true added value of the "Internet of Things" (IoT).

However, the current downside of sensor data is they are too fragmented, of too high velocity and too scarce on subtlety (Marr, 2015). At this point, a link between sensor data and behavioural data (e.g., time use) seems ideal. After all, sensory data are timestamped and sequential. They can therefore easily be linked to the activities that are registered in the time diary. As a result, these data can provide additional context not only for research itself (e.g., data enrichment through sensor data), but also for the way in which respondents can participate in research (e.g., sensor data can make activity suggestions). The question remains how data enrichment and respondent support can be realised.

Microservices and tentative and committed data

Sensors are often embedded in other devices, the smartphone being the most prominent. The data that are captured are processed into meaningful output using developed algorithms supported by Machine Learning and/or models based on Artificial Intelligence. Smartphones are often recognised as a proxy for the individual using it (place, sound, temperature ...). The most common way to send data to another environment is via a microservice. Such a microservice is often specialised in one specific function, such as geotracking or energy consumption. A characteristic of a microservice is that it is independent from a platform and can communicate with other platforms by means of an API. A data-collection platform on which time use research runs can therefore retrieve and link the data via this API.

The use of APIs facilitates the exchange of data between environments, including communication with front office applications used by respondents to participate in surveys. This also makes it possible to go beyond the rather rigid duality between active and passive data (collection), by making use of so-called tentative versus committed data in practice. Tentative data are passively collected from an individual but have not yet been validated by the individual. The data are considered a proxy of (the behaviour of) the individual and need interaction with the individual before they are turned into committed data. The big difference with the active registration of data is therefore that tentative data are first presented to the respondent for confirmation, addition, or a specific question, whereas otherwise the respondent must provide the input entirely himself. This business process, where an API exchanges tentative data between the microservices and the platform which in turn allows the individual to confirm the data through the front office application(s), is designed to keep the respondent central to the data-collection strategy. As such, it supports the

trustworthiness of the collection strategy since the respondent retains control over the data collection because the data becomes part of the research database only when the respondent records (i.e., commits) the data. At the same time, the front office application provides the option to the respondent to delete the tentative data stored in the microservice.

Sensor data provide extra information which is often difficult for an individual to grasp in such detail and with such precision. When this data stream is committed by human beings in the data-collection loop, both systematic error and privacy concerns are reduced.

This is the trajectory MOTUS currently follows and will continue to follow in the future. Microservices are seen as external environments which can be developed by external partners (e.g., using wearables like a Fitbit) but which can also be developed internally in connection with the MOTUS platform (e.g., the inclusion of sensors in the MOTUS applications). The development of microservices gives priority to the flexibility, on the one hand, and to privacy, on the other hand, as sensor data can contain sensitive information. The MOTUS core can communicate with different satellite microservices via adapter APIs to collect the tentative data that are defined in the study design in the back office of MOTUS, and which are presented to the respondent in the front office application for inclusion, adaptation, or deletion.

References

Bevans, G. E. (1913). *How working men spend their spare time*. Columbia University Press.

Bonke, J. (2005). Paid work and unpaid work: Diary information versus questionnaire information. *Social Indicators Research*, *70*, 349-368.

Bonke, J. & Fallesen, P. (2010). *The impact of incentives and interview methods on response quantity and quality in diary-and booklet-based surveys.* Paper presented at the Survey Research Methods.

Cai, L. & Zhu, Y. (2015). The challenges of data quality and data quality assessment in the big data era. *Data Science Journal*, *14*.

Eurostat. (2018). *European statistics code of practice: For the national statistical authorities and Eurostat (EU statistical authority).* Luxembourg: Publications Office of the European Union.

Eurostat. (2020). *Harmonised European Time Use Surveys (HETUS) – 2018 guidelines – Re-edition*. Luxembourg: Publications Office of the European Union.

Fisher, K., Gershuny, J., & Gauthier, A. (2012). *Multinational time use study: User's guide and documentation.* Oxford: Centre for Time Use Research, University of Oxford.

Fisher, K., Gershuny, J., Gauthier, A. H., & Victorino, C. (2000). *Exploring new ground for using the Multinational Time Use Study*. Study No. 2000-28. ISER Working Paper Series.

Frazis, H. & Stewart, J. (2010). Why do BLS hours series tell different stories about trends in hours worked? In K.G. Abraham, J.R. Spletzer, & M. Harper (Eds.), *Labor in the New Economy* (pp. 343-372). University of Chicago Press.

Gersbuny, J. & Sullivan, O. (1998). The sociological uses of time-use diary analysis. *European sociological review*, *14*(1), 69-85.

Gershuny, J. (2003). *Changing times: Work and leisure in postindustrial society*. Oxford University Press.

Gershuny, J. (2012). Too many zeros: A method for estimating long-term time-use from short diaries. *Annals of Economics and Statistics/Annales d'Economie et de Statistique*, 247-270.

Gershuny, J., Sullivan, O., Sevilla, A., Vega-Rapun, M., Foliano, F., Lamote de Grignon, J., . . . Walthery, P. (2021). A new perspective from time use research on the effects of social restrictions on COVID-19 behavioral infection risk. *Plos One*, *16*(2), e0245551.

Glorieux, I., & Minnen, J. (2009). How many days? A comparison of the quality of time-use data from 2-day and 7-day diaries. *Electronic International Journal of Time Use Research*, *6*(2), 314-327.

Groves, R. M. (2011). Three eras of survey research. *Public opinion quarterly*, *75*(5), 861-871.

Hamermesh, D. S. (1999). The timing of work over time. *The Economic Journal*, *109*(452), 37-66.

Harvey, A. S. (1993). Guidelines for time use data collection. *Social Indicators Research*, *30*, 197-228.

Huysmans, F., Lammers, J., & Wester, F. (1997). Radioluistertijd: Dagboekmeting, yesterday-meting en directe schattingen vergeleken. *Tijdschrift voor Communicatiewetenschap*, *25*(3), 253-276.

Juster, F. T. (1986). Response errors in the measurement of time use. *Journal of the American Statistical Association*, *81*(394), 390-402.

Juster, F. T. & Stafford, F. P. (1991). The allocation of time: Empirical findings, behavioral models, and problems of measurement. *Journal of Economic Literature*, *29*(2), 471-522.

Keusch, F., Struminskaya, B., Antoun, C., Couper, M. P., & Kreuter, F. (2019). Willingness to participate in passive mobile data collection. *Public opinion quarterly*, *83*(S1), 210-235.

Kuonen, D. & Loison, B. (2019). Production processes of official statistics and analytics processes augmented by trusted smart statistics: Friends or foes? *Statistical Journal of the IAOS*, *35*(4), 615-622.

Le Play, F. (1877). *Les ouvriers européen*. Alfred Mame et fils.

Marr, B. (2015). *Big Data: Using SMART big data, analytics and metrics to make better decisions and improve performance*. John Wiley & Sons.

Minnen, J. & Glorieux, I. (2011). Two days a week? A comparison of the quality of time-use data from 2-day, 7-day diaries and a weekly work grid. In J. Carrasco,

S. Jara-Díaz, & M. Munizaga (Eds.), *Time use observatory* (pp. 105-118). Grafica LOM.

Minnen, J., Glorieux, I., van Tienoven, T. P., Daniels, S., Weenas, D., Deyaert, J., . . . Rymenants, S. (2014). Modular Online Time Use Survey (MOTUS) – translating an existing method in the 21st century. *Electronic International Journal of Time Use Research*, *11*, 73-93.

Minnen, J., Rymenants, S., Glorieux, I., & van Tienoven, T. P. (2023). Answering current challenges of and changes in producing official time-use statistics using the data collection platform MOTUS. *Journal of Official Statistics*, Online First.

Pronovost, G. (1989). The sociology of time. *Sociologie Contemporaine (La)*, *37*(3), 1-124.

Reeves, M. S. R. & Reeves, M. P. (1913). *Round about a pound a week*. G. Bell and Sons, Limited.

Revilla, M., Couper, M. P., & Ochoa, C. (2019). Willingness of online panelists to perform additional tasks. *Methods, Data, Analyses: A Journal for Quantitative Methods and Survey Methodology (MDA)*, *13*(2), 223-252.

Ricciato, F., Wirthmann, A., & Hahn, M. (2020). Trusted smart statistics: How new data will change official statistics. *Data & Policy*, *2*.

Rinderknecht, R. G., Doan, L., & Sayer, L. C. (2022). MyTimeUse: An online implementation of the day-reconstruction method. *Journal of Time Use Research*, *1*.

Robinson, J. P. & Converse, P. (1972). Social change as reflected in the use of time. In A. Campbell & P. Converse (Eds.), *The human meaning of social change* (pp. 17-86). Russell Sage Foundation.

Robinson, J. P. & Godbey, G. (1997). *Time for life: The surprising ways Americans use their time*. Penn State Press.

Salgado, D., Esteban, M. E., Novás, M., Saldaña, S., & Sanguiao, L. (2018). Data organisation and process design based on functional modularity for a standard production process. *Journal of Official Statistics*, *34*(4), 811-833.

Scheuch, E. (1972). The time budget interview. In A. Szalai (Ed.), *The use of time* (pp. 69-87). Mouton.

Sonck, N. & Fernee, H. (2013). *Using smartphones in survey research: A multifunctional tool*. The Hague: The Netherlands Institute for Social Research.

Sorokin, P. A. & Berger, C. Q. (1939). *Time-budgets of human behavior* (*Vol. 2*). Harvard University Press.

Sorokin, P. A. & Merton, R. K. (1937). Social time: A methodological and functional analysis. *American Journal of Sociology*, *42*(5), 615-629.

Stinson, L. L. (1999). Measuring how people spend their time: A time-use survey design. *Monthly Lab. Rev.*, *122*, 12.

Stodden, V. (2014). The reproducible research movement in statistics. *Statistical Journal of the IAOS*, *30*(2), 91-93.

Sullivan, O., Gershuny, J., Sevilla, A., Walthery, P., & Vega-Rapun, M. (2020). Time use diary design for our times-an overview, presenting a Click-and-Drag Diary Instrument (CaDDI) for online application. *Journal of Time Use Research*, *15*(1), 1-17.

Szalai, A. (1972). *The use of time: Daily activities of urban and suburban populations in twelve countries*. Mouton.

te Braak, P., van Droogenbroeck, F., Minnen, J., van Tienoven, T. P., & Glorieux, I. (2022). Teachers' working time from time-use data: Consequences of the invalidity of survey questions for teachers, researchers, and policy. *Teaching and Teacher Education*, *109*, 103536.

van Tienoven, T. P., Minnen, J., Glorieux, A., Laurijssen, I., te Braak, P., & Glorieux, I. (2023). Locking down gender roles? A time-use perspective on gender division of household labour during the COVID-19 pandemic lockdown in Belgium. *Journal of Family Issues*, *44*(3), 654-680.

Wong, S., Harrison, L. J., Gibson, M. L., Press, F., Bittman, M., Crisp, K., & Ryan, S. (2022). Measuring early childhood educators' time at work using an electronic random time-sampling approach. *International Journal of Research & Method in Education*, 1-16.

Zeni, M., Bison, I., Reis, F., Gauckler, B., & Giunchiglia, F. (2020). Improving time use measurement with personal big data collection – the experience of the European Big Data Hackathon 2019. *Journal of Official Statistics*, *37*(2), 341-365.

Zerubavel, E. (1982). *Hidden rhythms: Schedules and calendars in social life*. University of Chicago Press.

Zuzanek, J. (1980). The work-leisure relationship in Soviet sociological discussion. *Canadian Slavonic Papers*, *22*(1), 122-128.