

Stehrenberg, Shari; Volk, Johannes

**Article**

## Trusted Smart Surveys: Fragebogen trifft neue digitale Daten

WISTA - Wirtschaft und Statistik

**Provided in Cooperation with:**

Statistisches Bundesamt (Destatis), Wiesbaden

*Suggested Citation:* Stehrenberg, Shari; Volk, Johannes (2023) : Trusted Smart Surveys: Fragebogen trifft neue digitale Daten, WISTA - Wirtschaft und Statistik, ISSN 1619-2907, Statistisches Bundesamt (Destatis), Wiesbaden, Vol. 75, Iss. 1, pp. 59-68

This Version is available at:

<https://hdl.handle.net/10419/268967>

**Standard-Nutzungsbedingungen:**

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

**Terms of use:**

*Documents in EconStor may be saved and copied for your personal and scholarly purposes.*

*You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.*

*If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.*

# TRUSTED SMART SURVEYS: FRAGEBOGEN TRIFFT NEUE DIGITALE DATEN

Shari Stehrenberg, Johannes Volk

↳ **Schlüsselwörter:** Trusted Smart Surveys – Trusted Smart Statistics – neue digitale Daten – Erhebungsmethoden – Sensordaten – ESSnet

## ZUSAMMENFASSUNG

„Smart Surveys“ sind Befragungen, die mithilfe von mobilen Geräten, typischerweise Smartphones oder Tablets, durchgeführt werden. Sie kombinieren interaktive Daten, die explizit von den Auskunftgebenden zur Verfügung gestellt werden, mit passiven, im Hintergrund von Sensoren erfassten Daten. Dadurch bieten Smart Surveys das Potenzial, die Qualität von Befragungen zu verbessern und den Befragungsaufwand bei zusätzlichem Informationsgewinn zu verringern. Das europäische Projekt ESSnet Smart Surveys leistete vorbereitende Arbeiten für die Entwicklung einer europaweiten Plattform zum Austausch von Smart-Survey-Lösungen und -komponenten. Pilotprojekte in den Bereichen Konsum, Zeitverwendung, Gesundheit und Lebensbedingungen ergänzten das ESSnet-Smart-Surveys-Projekt.

↳ **Keywords:** Trusted Smart Surveys – Trusted Smart Statistics – New Digital Data – survey methodology – sensor data – ESSnet

## ABSTRACT

*Smart Surveys are surveys conducted with the help of mobile devices, typically smartphones or tablets. They combine interactive data explicitly provided by respondents with passive data collected in the background by sensors. In this way, Smart Surveys offer the potential to improve the quality of surveys and reduce the response burden while collecting additional information. The European project ESSnet Smart Surveys carried out preparatory work for the development of a Europe-wide platform for the exchange of Smart Survey solutions and components. In addition, pilot projects were conducted in the areas of consumption, time use, health and living conditions.*



**Shari Stehrenberg**

hat einen Master of Science Wirtschaftspsychologie und ist Referentin im Referat „Europäische und internationale Koordination“ des Statistischen Bundesamtes. Zuvor war sie als Referentin im Institut für Forschung und Entwicklung in der Bundesstatistik tätig und in dieser Position für die Gesamtprojektleitung des ESSnet-Smart-Surveys-Projekts verantwortlich.



**Johannes Volk**

studierte Politologie und Statistik und ist Referent im Referat „Design und Weiterentwicklung von Erhebungsinstrumenten“ des Statistischen Bundesamtes. Einer seiner Arbeitsschwerpunkte in den letzten Jahren war das ESSnet-Smart-Surveys-Projekt.

## 1

---

### Einleitung

---

Die rasante und transformative Entwicklung des Internets der Dinge und sensorgesteuerter smarterer Geräte sowie das Vordringen digitaler Technologien in den Alltag haben zu einer regelrechten Datenflut geführt. Das Volumen, die Menge und die Dimensionalität dieser Daten sind um ein Vielfaches höher als bei Datenerhebungen im vordigitalen Zeitalter (Ricciato und andere, 2019).

Dadurch bieten neue digitale Daten die Chance, sich als dritte Datenquelle neben Umfragedaten und Verwaltungsdaten zu etablieren.<sup>1</sup> Zudem eröffnen sie die Möglichkeit, die umfragebasierte Datenerhebung durch den Einsatz von Trusted Smart Surveys maßgeblich zu ergänzen und zu verbessern.

Dieser Beitrag geht einleitend allgemein auf Trusted Smart Surveys ein und stellt danach das ESSnet<sup>2</sup>-Smart-Surveys-Projekt vor, bei dem das Statistische Bundesamt neben der inhaltlichen Beteiligung auch mit der Gesamtkoordination betraut war.

## 2

---

### Chancen und Herausforderungen von Trusted Smart Surveys

---

Der Begriff „Smart Surveys“ bezeichnet Befragungen, die mithilfe von mobilen Geräten, typischerweise mit Smartphones oder Tablets, durchgeführt werden. Sie kombinieren explizit von den Auskunftgebenden zur Verfügung gestellte interaktive Daten (beispielsweise Antworten auf Fragen) mit passiven Daten, die im Hintergrund von Sensoren (zum Beispiel von Beschleunigungsmessern, Temperatursensoren, Mikrofonen) erfasst werden (Ruppert und andere, 2018).

---

1 Mehrere Forschungsaktivitäten und Pilotprojekte im Europäischen Statistischen System (ESS) haben bereits das Potenzial der Nutzung neuer digitaler Datenquellen (dazu zählen Mobilitätsdaten von Mobilfunknetzbetreibern, Satellitenbilder oder durch Webscraping automatisiert gewonnene, frei verfügbare Daten aus dem Internet) für die amtliche Statistik aufgezeigt. Als einschlägige Beispiele sind die Projekte ESSnet Big Data I und II zu nennen. Siehe [ec.europa.eu](https://ec.europa.eu) und [ec.europa.eu](https://ec.europa.eu)

2 ESSnet = European Statistical System Network.

Darüber hinaus werden bei „Trusted Smart Surveys“ technische Lösungen zum Schutz der Privatsphäre (Privacy Enhancing Technologies) eingesetzt. Sie sollen personenbezogene Daten besser und automatisiert schützen und eine strenge Kontrolle über die Datenverarbeitung und die Extraktion von Informationen aus den Daten etablieren, sodass vollständige Transparenz und Nachprüfbarkeit der Datenverarbeitungsmethoden und -algorithmen gewährleistet sind (Ricciato und andere, 2020). Der Einsatz solcher Technologien ermöglicht es zudem, Daten zu analysieren, ohne dass diese im Auswertungssystem vollständig vorliegen müssen. Diese Maßnahme soll nicht nur die rechtlichen Anforderungen an den Datenschutz erfüllen, sondern auch die Akzeptanz und Teilnahmebereitschaft bei den Auskunftgebenden fördern. Dies ist besonders wichtig, da bei smarten Erhebungen viel mehr Daten erhoben werden als bei herkömmlichen Erhebungen, diese Daten weniger spezifisch für die endgültige Statistik sind und viele zusätzliche Informationen über die Befragten enthalten. Dar- aus resultiert ein größeres Risiko für die Privatsphäre der Auskunftgebenden als bei herkömmlichen Umfragen, was die Notwendigkeit sicherer Systeme unterstreicht.

Smart Surveys bieten das Potenzial, sehr viele, sehr präzise Daten verarbeiten zu können und somit eine bessere Datenqualität zu erreichen als mit herkömmlichen Umfragen. Diese können durch negative Effekte wie schlechtes Erinnerungsvermögen, schlechtes Schätzvermögen oder soziale Erwünschtheit verzerrt sein. Allerdings ist zu bedenken, dass verschiedene Fehlerquellen (wie mangelnde Repräsentativität, technische Probleme, ausgeschaltete Geräte oder geringe Qualität der Sensormessungen) die Qualität von Sensordaten ebenfalls negativ beeinflussen können (Struminskaya und andere, 2020).

Neben den erwähnten Risiken liegen die Chancen von Smart Surveys vor allem in der Möglichkeit, passiv und aktiv erhobene Daten miteinander zu kombinieren. Passiv gewonnene Sensordaten sind zwar umfangreich und können beispielsweise das Reiseverhalten (Standorte, Dauer, Geschwindigkeit, Verkehrsmittel) sehr genau erfassen, sie ermöglichen aber beispielsweise keinen Einblick in die Motivation, die zu einer Reise führte. Die Kombination der Sensordaten mit Umfragedaten, zum Beispiel zum Grund der Reise, bietet den großen Mehrwert, sehr genaue Messdaten mit zusätzlichen, qualitativen Informationen analysieren zu können (Struminskaya und andere, 2020).

Zudem können Smart Surveys die Belastung der Befragten reduzieren: Da entsprechende Informationen passiv über Sensoren gewonnen werden, müssen weniger Fragen gestellt werden (Keusch und andere, 2019). Dies gilt insbesondere, wenn die Datenabfrage sehr komplex und/oder mit großem Zeitaufwand verbunden ist.<sup>13</sup> Einen weiteren positiven Effekt für die Befragten könnte die Möglichkeit einer kontinuierlichen Interaktion mit ihnen darstellen, zum Beispiel indem sie individualisierte Berichte und Auswertungen als Feedback erhalten (Ricciato und andere, 2020).

### 3

## Das ESSnet-Smart-Surveys-Projekt

---

### 3.1 Ziel und Aufbau

---

Mit dem von der Europäischen Kommission geförderten Projekt „ESSnet Smart Surveys“ hat ein Konsortium von zwölf nationalen statistischen Ämtern die Einsatzmöglichkeiten von Trusted Smart Surveys in der amtlichen Statistik untersucht (Stehenberg/Giannakouris, 2022). Das Konzept von Trusted Smart Surveys ist dabei ein Teil der europäischen Trusted-Smart-Statistics-Initiative. Hierbei wird Trusted Smart Statistics als ein von intelligenten Systemen bereitgestellter Dienst verstanden, der prüfbar und transparente Datenlebenszyklen umfasst, die Gültigkeit und Genauigkeit der Ergebnisse gewährleistet und die Privatsphäre der betroffenen Personen respektiert sowie einen hohen Datenschutz gewährleistet (Ricciato und andere, 2020).<sup>14</sup>

Das Projekt „ESSnet Smart Surveys“ leistete in diesem Kontext vorbereitende Arbeiten, um eine europaweite Plattform zu entwickeln. Eine solche Plattform dient zum Austausch und zur Wiederverwendung von Trusted-Smart-Survey-Lösungen und -Komponenten, die sowohl unabhängig von bestimmten Statistikbereichen als auch

flexibel und modular sein sollen. Die Vorentwicklungsphase umfasste folgende Schritte: Anforderungen sammeln, Designprinzipien identifizieren, einen konzeptionellen Rahmen und eine Referenzarchitektur entwickeln, Spezifikationen und Proof of Concepts formulieren (Stehenberg/Giannakouris, 2022).

Das Projekt gliederte sich in drei Arbeitspakete:

- › Kommunikation und Koordinierung (Arbeitspaket 1; Leitung: Statistisches Bundesamt)
- › Pilotprojekte (Arbeitspaket 2; Leitung: CBS [Centraal Bureau voor de Statistiek – Statistikamt der Niederlande])
- › Konzeptioneller Rahmen für eine Plattform für Trusted-Smart-Survey-Komponenten (Arbeitspaket 3; Leitung: Istat [Istituto Nazionale di Statistica – Statistikamt Italiens])

Neben diesen Arbeitspaketen gründeten sich im Verlauf des Projekts Arbeitsgruppen, um die Arbeiten zwischen den Arbeitspaketen 2 und 3 besser zu verknüpfen. Der folgende Abschnitt 3.2 stellt die Arbeitsgruppe zu rechtlichen Fragestellungen vor.

Dieser Beitrag kann nicht auf alle Teilaspekte des Projekts eingehen. Die detaillierten Projektergebnisse der beiden Arbeitspakete 2 und 3 stehen auf der [CROS-Seite des Projekts](#) zur Verfügung; darauf basieren die Abschnitte 3.3 bis 3.5.

### 3.2 Arbeitsgruppe zu rechtlichen Fragen

---

Zu Beginn des Projektes zeichnete sich ab, dass rechtliche Aspekte eine ernst zu nehmende Herausforderung für das Projekt, aber auch generell für die europäische Zusammenarbeit im Bereich von Trusted Smart Surveys darstellen. Bei der Vertragsgestaltung zur Nutzung einer App zur Zeitverwendungserhebung zeigte sich, dass die einzelnen Mitgliedstaaten die für alle Beteiligten geltende europäische Datenschutz-Grundverordnung (DSGVO) unterschiedlich interpretieren. Zudem stellt nationales Recht, zum Beispiel das deutsche IT-Grundschutz-Kompodium, weitere Anforderungen.

Um dieses Thema umfassender zu berücksichtigen, wurde während der Projektlaufzeit beschlossen, eine zusätzliche Arbeitsgruppe zu rechtlichen und ethi-

---

3 So kann beispielsweise bei der Einkommens- und Verbrauchsstichprobe das Abfotografieren eines Kassenzettels zur Erfassung der Konsumausgaben einfacher und zeitsparender sein, als jede Position des Belegs einzeln einzutippen. Außerdem können Zeitpunkt und Inhalt der gestellten Fragen dynamisch auf der Grundlage von aus den Sensordaten abgeleiteten Kontextinformationen festgelegt werden.

4 Siehe zudem auch die Informationen auf der Seite von Eurostat, dem Statistischen Amt der Europäischen Union: [ec.europa.eu](https://ec.europa.eu)

schen Fragestellungen einzusetzen. Sie setzte sich aus Rechts- und Erhebungsdesignexperten und -expertinnen zusammen.

Die Arbeitsgruppe hat einen ausführlichen Bericht erstellt über neue Aspekte, die bei Trusted Smart Surveys im Vergleich zu herkömmlichen Umfragen zu beachten sind, und welche technischen und organisatorischen Maßnahmen dazu zu ergreifen sind. Der Bericht enthält zudem einen Entwurf für eine Datenschutz-Folgenabschätzung für Smart Surveys.

In einem Diskussionspapier setzte sich die Gruppe mit zwei wichtigen Designentscheidungen bei Trusted Smart Surveys auseinander:

- (1) mit der Einbeziehung von Dritten, beispielsweise Vertragspartnern, zur Bereitstellung von Apps, und
- (2) der Datenverarbeitung auf den zur Datenerhebung eingesetzten smarten Geräten gegenüber der Verarbeitung auf Servern, zum Beispiel in den Statistikämtern.

Zwar waren diese Arbeiten im Voraus nicht geplant, sie stellen jedoch wichtige grundlegende Überlegungen für die weitere Zusammenarbeit im Bereich Trusted Smart Surveys dar, auf denen andere Projekte aufbauen können. Dies ist nicht nur von erheblicher Bedeutung, um die rechtlichen Anforderungen an den Datenschutz zu erfüllen, sondern auch, um die Akzeptanz und Teilnahmereitschaft in der Gesellschaft zu fördern.

### 3.3 Arbeitspaket 1: Kommunikation und Koordinierung

---

Das Projekt startete Anfang 2020, also etwa zu der Zeit, als das Coronavirus Europa erreichte. Die Pandemie sorgte dafür, dass persönliche Treffen während fast der gesamten Laufzeit nicht möglich waren. Die Koordinations- und Halbzettreffen mussten daher alle virtuell stattfinden. Zudem kam es aufgrund der Pandemie zu Verzögerungen, da Personal in den nationalen statistischen Ämtern anderweitig gebunden war. Persönliche Befragungen mithilfe von Interviewerinnen und Interviewern waren in Arbeitspaket 2 geplant, konnten aber nicht im ursprünglich vorgesehenen Zeitrahmen durchgeführt werden, sodass die Projektlaufzeit um ein halbes Jahr verlängert wurde.

Schließlich war es im Mai 2022 möglich, das Projekt mit einem [Abschlussworkshop](#) in Wien zu beenden. Das Projekt wurde darüber hinaus auf einigen (virtuellen) internationalen Konferenzen vorgestellt, etwa dem Weltkongress 2021 des Internationalen Statistischen Instituts (ISI)<sup>5</sup>, der [NTTS 2021](#)<sup>6</sup> sowie der [IAOS-Konferenz 2022](#)<sup>7</sup> und der [Q2022-Konferenz](#)<sup>8</sup>, dort mit einer eigenen Special Session.

### 3.4 Arbeitspaket 2: Pilotprojekte

---

Arbeitspaket 2 umfasste die empirische und angewandte Komponente des Projekts und war in vier Pilotprojekte zu den Themen Konsum, Zeitverwendung, Gesundheit und Lebensbedingungen gegliedert. Befragungen zum Konsum (Beispiel: Einkommens- und Verbrauchsstichprobe – EVS) und zur Zeitverwendung (Beispiel: Zeitverwendungserhebung – ZVE) sind sehr aufwendig für Befragte und gleichzeitig anfällig für fehlerhafte Angaben, zum Beispiel aufgrund von Erinnerungslücken. Dies zeigt sich beispielsweise in den Erhebungen, die einen tagebuchgestützten Erhebungsansatz verfolgen. Gerade in diesen Fällen erscheint der Einsatz von Apps angebracht. Fragen rund um die Verwendung solcher Apps waren daher Thema in den ersten beiden Pilotprojekten: Wie unterschiedlich sind die Anforderungen der verschiedenen nationalen Statistikämter? Wie können smarte Instrumente in den Apps die Belastung der Befragten reduzieren? Wie müssen Apps gestaltet und aufgebaut sein, damit die Befragten sie leicht und gut nutzen können? Kann eine App auch für andere, ähnliche thematische Fragestellungen genutzt werden? Wie ist die Datenqualität der per App erhobenen Daten? Wie können Befragte am besten für die Nutzung einer App gewonnen werden?

Die Pilotprojekte drei und vier waren kleiner und experimenteller. In beiden wurden Sensordaten mittels anderer Geräte als dem Smartphone erfasst, die Qualität der

---

5 International Statistical Institute World Statistical Congress.

6 Die New Techniques and Technologies for Statistics (NTTS) ist eine internationale zweijährliche wissenschaftliche Konferenzreihe, die vom Statistischen Amt der Europäischen Union (Eurostat) organisiert wird.

7 IAOS: International Association for Official Statistics.

8 Q2022: 10th European Conference on Quality in Official Statistics.

Daten bewertet, aber auch die Bereitschaft von Befragten für solche Erhebungen thematisiert.<sup>9</sup>

### Konsum

Im größten Pilotprojekt aus dem zweiten Arbeitspaket stand das Erfassen von Konsumausgaben, vor allem von Ausgaben des täglichen Gebrauchs, wie Einkäufen im Supermarkt, im Mittelpunkt. Es war daher thematisch eng mit dem europäischen Household Budget Survey (HBS) (in Deutschland: EVS), verbunden. Das Pilotprojekt ist das einzige im ESSnet-Smart-Surveys-Projekt, in dem auch groß angelegte, länderübergreifende Feldtests durchgeführt wurden: Die nationalen Statistikämter der Niederlande, Spaniens und Luxemburgs testeten die vom CBS (Niederlande) entwickelte @HBS App<sup>10</sup> mit insgesamt über 4 000 Personen in der Brutto-Stichprobe.

Zentraler Bestandteil der App sind die beiden Möglichkeiten, Ausgaben zu erfassen: manuell durch eigenes Eingeben der Ausgaben über eine Suchfunktion und automatisiert durch das Abfotografieren von Kassenzetteln. Hinter beiden Möglichkeiten verbergen sich smarte Funktionen: zum einen der Suchalgorithmus bei der Freitext-Eingabe, zum anderen das OCR-Scanverfahren zum Auslesen des Textes auf den Kassenzetteln und die anschließende Kodierung der Produkte anhand der [COICOP-Klassifikation](#). Etwa ein Viertel der in den Feldtests gemeldeten Einkäufe wurde durch die automatisierte Eingabe erfasst. Dabei zeigte sich, dass Befragte zusätzlich die Möglichkeit haben sollten, die eingescannten Daten zu sichten und bei Bedarf zu bearbeiten.<sup>11</sup>

Das Pilotprojekt hat zudem Folgendes verdeutlicht: Die Entwicklung von auf maschinellem Lernen basierten Modellen für die korrekte Kodierung und COICOP-Klassifizierung der Produkte sowie das Erstellen der umfangreichen Produktlisten in den jeweiligen Landessprachen

stellen entscheidende Herausforderungen dar für eine erfolgreiche Verwendung einer Haushaltstagebuch-App.

Die App bietet personalisierte Feedback-Möglichkeiten in Form von Auswertungsübersichten, die die eigenen Ausgaben der Befragten grafisch aufbereiten. Bei den Feldtests ließen sich jedoch keine eindeutigen Anzeichen für höhere Antwortquoten, eine geringere Abbruchquote und/oder eine bessere Datenqualität erkennen, wenn ein personalisiertes Feedback angeboten wird. Nichtsdestotrotz wurde es in den Usability-Tests sehr geschätzt und sollte daher beibehalten werden.

### Zeitverwendung

Das zweite Pilotprojekt beschäftigte sich mit der Erfassung aller Aktivitäten im Tagesablauf und war daher inhaltlich mit dem Harmonised European Time Use Survey (HETUS) (in Deutschland: ZVE) verknüpft. Ähnlich wie beim ersten Pilotprojekt kam auch hier eine bereits zuvor entwickelte App zum Einsatz, die getestet und weiterentwickelt wurde: die [Motus-App](#). Kernelement dieser App ist ein Tagebuch, in dem Befragte ihre Aktivitäten einschließlich einiger Zusatzinformationen eintragen. Zentral ist auch hier eine Suchfunktion, um die entsprechenden Tätigkeiten zu finden und der korrekten Kategorie zuzuordnen.

In mehreren Usability-Testrunden wurde die App getestet. Der Fokus variierte dabei: Neben der klassischen Zeitverwendungserhebung wurden auch Versionen getestet, die eher die Themen Arbeitszeit und Mobilität aufgriffen. In Richtung einer breiteren Einsatzmöglichkeit zielte auch der Test einer neu entwickelten Funktion, die mittels „Geo Fencing“ Geodaten erfasst und eine Aktion auslöst, sobald sich das Smartphone in der Nähe eines bestimmten Ortes befindet.<sup>12</sup> Zwar befindet sich diese Funktion noch in einem frühen Entwicklungsstadium und die Tests zeigten deutlichen Verbesserungsbedarf<sup>13</sup>, jedoch ist das grundlegende Motus-System insgesamt hinsichtlich Funktionalität und Nutzerfreundlichkeit bereits weit ausgereift. So konnte bei der deutschen Zeitverwendungserhebung 2022 dieses System mit einer angepassten App erstmals eingesetzt werden.

9 Die im Folgenden dargestellten Ergebnisse basieren auf den Projektergebnissen des Arbeitspaketes 2. Diese sind auf der CROS-Seite des Projektes zu finden: [ec.europa.eu](https://ec.europa.eu)

10 Siehe Endbericht zu Arbeitspaket 2, Seite 6: [ec.europa.eu](https://ec.europa.eu)

11 In den Feldtests war dies eine der Experimentierbedingungen: Manche Testpersonen hatten die Möglichkeit, ihre eingescannten Daten zu bearbeiten, andere nicht. Die Auswertung zeigt, dass die Datenqualität bei manueller Nachbearbeitung besser ist. Ausführlicher thematisiert wird dies in [Deliverable 2.3](#).

12 Ein Beispiel für eine solche Aktion ist das Senden einer Push-Nachricht mit dem Hinweis, einen kurzen Fragebogen zu beantworten.

13 [ec.europa.eu](https://ec.europa.eu)

## Gesundheit

Beim Thema Gesundheit lag der Fokus auf der Messung der körperlichen Aktivität. Kleinere Machbarkeitsstudien wurden in Belgien, den Niederlanden und Polen mittels eines einfachen Fitnesstrackers, der nur über einen Beschleunigungsmesser verfügt, durchgeführt. Daneben gab es drei größere Online-Befragungen in den Niederlanden<sup>14</sup>. Sie zielten darauf ab, die Akzeptanz solcher Erhebungen in der Öffentlichkeit zu ermitteln, Strategien zu untersuchen, mit denen Menschen zu einer Teilnahme animiert werden können, sowie herauszufinden, ob möglicherweise private Fitnesstracker genutzt werden können. Hierbei zeigte sich, dass die Befragten grundsätzlich bereit sind, an solchen Erhebungen teilzunehmen, dass aber zusätzliche Maßnahmen nötig sind, um die Akzeptanz und damit die Teilnahmebereitschaft zu erhöhen.

Für die Auswertung der Daten wurde versucht, eigene Modelle zu trainieren und Algorithmen abzuleiten mit dem Ziel, körperliche Aktivität erfassen und zuordnen zu können. Die Ergebnisse waren noch nicht zufriedenstellend. Hier ist weitere Forschung unumgänglich, da bereits aus den oben angeführten Gründen zu Vertrauen, Verantwortung und Transparenz eine alleinige Nutzung der von kommerziellen Anbietern zur Verfügung gestellten Analyse-Algorithmen nicht sinnvoll ist.

## Lebensbedingungen

Im vierten und letzten Pilotprojekt von Arbeitspaket 2 beschränkten sich die untersuchten Lebensbedingungen auf die Luftqualität in geschlossenen Räumen. Da Befragte in der Regel nicht über Messgeräte zum Erfassen der Luftqualität verfügen, wurden für diesen Pilotversuch entsprechende Geräte ([uHoo](#)) verteilt, die die Teilnehmenden für eine gewisse Zeit in ihrer privaten Wohnung aufstellen sollten.

In mehreren kleineren Machbarkeitsstudien – durchgeführt in Deutschland, den Niederlanden und Belgien – standen die Akzeptanz und die Benutzerfreundlichkeit der Messgeräte im Fokus. Dabei gaben manche Teilnehmende an, dass die Werte, die ihnen auf einer App zu ihrer Luftqualität angezeigt wurden, bei ihnen zu (dauer-

haften) Verhaltensänderungen führten, vor allem bezüglich des Lüftens und Heizens. Das ist für eine objektive Datenerfassung, welche die Situation ohne den durch die Messgeräte verursachten Einfluss erheben will, problematisch.

In der Wahrnehmung der Befragten steigt die Sensibilität der Daten abhängig von der Art und Menge der Verknüpfung mit Kontextdaten.<sup>15</sup> Zudem zeigte sich ein Unterschied zwischen den Ländern hinsichtlich der wahrgenommenen Sensibilität der Daten, wobei Befragte aus Deutschland und den Niederlanden die Daten als sensibler einschätzten als jene aus Belgien.

## 3.5 Arbeitspaket 3: Konzeptioneller Rahmen für eine Trusted-Smart-Surveys-Plattform

---

Das Hauptziel von Arbeitspaket 3 war, einen konzeptionellen Rahmen und technische Spezifikationen für eine europäische Plattform zum Austausch und zur Wiederverwendung von Smart-Survey-Lösungen und -Komponenten zu entwickeln. Dieser allgemeine Rahmen betrachtet die beteiligten Erhebungsprozesse aus verschiedenen Blickwinkeln (methodisch, technologisch, architektonisch) und bietet ein neues und nützliches Referenzschema für die nationalen statistischen Ämter.

Der Entwurf einer Referenzarchitektur ging von bestehenden Rahmenwerken aus. Im nächsten Schritt wurden Proofs of Concepts in Form von modularen Prototyp-Elementen für wesentliche Aspekte der Architektur durchgeführt, deren Ergebnisse zur Weiterentwicklung des konzeptionellen Rahmens dienten. Im letzten Schritt erfolgten auf Basis dieser Erkenntnisse Anpassung und Erweiterung des zuvor entworfenen konzeptionellen Rahmens für die Plattform.

Nach einem Blick auf die Referenzarchitektur thematisiert dieser Abschnitt noch den Einsatz von Modellen mit maschinellem Lernen und Technologien zum Schutz der Privatsphäre.<sup>16</sup>

---

14 Die drei Projekte in den Niederlanden waren Online-Befragungen, an deren Ende auch die Bereitschaft zur Teilnahme an Studien mit Fitnesstrackern abgefragt wurde. Genauere Informationen finden sich im Endbericht des Arbeitspaketes 2 ([ec.europa.eu](#), hier: Seite 10).

15 Beispielsweise mit sozioökonomischen Daten der Befragten oder Adressdaten.

16 Die im Folgenden dargestellten Ergebnisse basieren auf den Projektergebnissen des Arbeitspaketes 3. Diese sind auf der CROS-Seite des Projektes zu finden: [ec.europa.eu](#)

### Architektur

Die Gestaltung der Referenzarchitektur für Trusted Smart Surveys zielte darauf ab, die Beziehungen zwischen der Architektur, methodischen Komponenten und technologischen Überlegungen zu verdeutlichen. Dazu wurden ein „Business Layer“ und ein „Application Layer“ modelliert und die „Building Blocks“ beschrieben.

Das Hauptziel der Modellierung des Business Layers ist, die wichtigsten Aktivitäten und Verhaltensweisen zu identifizieren und zu standardisieren, die von den nationalen statistischen Ämtern im Zuge der Datenerfassung und -verarbeitung bei Trusted Smart Surveys durchzuführen sind. Dabei standen die folgenden Fragen im Vordergrund: Welche Auswirkungen haben Trusted Smart Surveys auf die GSBPM-Phasen<sup>17</sup> und Prozessschritte? Welche Schritte des Prozesses können standardisiert und in einer gemeinsamen Infrastruktur ausgeführt werden? Welche Tätigkeiten können in der Plattform oder auf den Geräten der Auskunftgebenden ausgeführt und welche Aufgaben müssen von den nationalen statistischen Ämtern intern durchgeführt werden? Was ist der beste Ansatz, um die spezifischen nationalen Bedürfnisse und die Besonderheiten der an einem Prozess beteiligten Akteure zu harmonisieren?

Der Entwurf des Application Layers ist der erste Schritt zur Definition einer vorläufigen Reihe von Bausteinen (Building Blocks), die für die Ausführung der Prozessschritte eines allgemeinen Trusted Smart Surveys erforderlich sind. Diese Bausteine werden in Form von Funktionalitäten beschrieben, die von der Plattform angeboten werden könnten. Sie dienen als Basis, um Annahmen über die Bereitstellung und gemeinsame Nutzung von Diensten abzuleiten.

Durch die Arbeiten im Arbeitspaket 3 zeigte sich, dass eine allumfassende, von den Fachgebieten unabhängige Plattform für Trusted Smart Surveys, wie sie anfangs angedacht war, äußerst komplex und unrealistisch ist. Zudem wäre sie nicht für alle Bereiche notwendig. Komplett unabhängige, auf einzelne Erhebungen zugeschnittene Anwendungen zu entwickeln, erscheint jedoch ebenfalls nicht zielführend. Der Vorschlag ist daher, die Anwendungen für Trusted Smart Surveys flexibel zu ent-

wickeln und klar zwischen allgemeinen Komponenten und dem, was bereichsabhängig ist, zu unterscheiden. Die konkrete Ausgestaltung der Erfassung und Verarbeitung von Smart-Survey-Daten je nach verwendeten Sensoren und Datenquellen sollte angepasst werden, ebenso Aspekte wie das Monitoring der Erhebung.

Die Modularität ist dabei der zentrale Punkt, um ein Gleichgewicht zwischen Verallgemeinerung und Spezifität herzustellen und so die Implementierung wiederverwendbarer Komponenten zu fördern. In der konkreten Umsetzung könnte mit gemeinsam nutzbaren Bausteinen begonnen werden.

Die technische Infrastruktur muss dabei dieser Modularität Rechnung tragen. Das bedeutet, dass sie für die verschiedenen Arten von Anwendungsschnittstellen (APIs) für die Bereitstellung von Daten (verschiedene Sensoren, Umfragedaten) ausgelegt sein muss. Zudem muss sie es ermöglichen, technische Komponenten für Trusted Smart Surveys einheitlich zu spezifizieren. Und zwar unabhängig davon, ob die Daten im Gerät oder auf der Plattform verarbeitet werden oder ob die bereitgestellten Daten unstrukturiert oder strukturiert sind.

### Machine Learning in Smart Surveys

Modelle, die auf maschinellem Lernen (Machine Learning – ML) basieren, können zum einen eingesetzt werden, um von smarten Geräten gelieferte Eingabedaten (zum Beispiel Bilder, Signale, Sprache) während der Datenerfassung zu verarbeiten. Zum anderen können sie hilfreich sein, um Sensordaten zusammenzuführen und Fehler (Anomalien in den Daten, fehlende Daten und Ähnliches) in der Datenverarbeitungsphase zu beheben. Die ML-Anwendungen können dabei entweder auf den Geräten der Befragten oder zentral auf Systemen der nationalen statistischen Ämter (beispielsweise in einer Cloud oder auf deren Servern) ausgeführt werden. Welche der beiden Lösungen zu wählen ist, hängt von den Vor- und Nachteilen des jeweiligen Ansatzes und den Anforderungen der Erhebung ab.

Für ML-Anwendungen direkt in der App spricht zum einen der Schutz persönlicher Daten, da Rohdaten das Gerät der Nutzenden gar nicht erst verlassen, sondern nur die relevanten Daten in aufbereiteter Form. Zum anderen sind ML-Anwendungen direkt in der App nicht auf eine Internetverbindung angewiesen. Die direkte Verarbeitung vor Ort bietet außerdem schnellere Reak-

17 Generic Statistical Business Process Model, siehe hierzu [statswiki.unece.org](https://www.statswiki.unece.org). Das deutsche GMAS-Konzept (GMAS = Geschäftsprozessmodell Amtliche Statistik) ist hiervon abgeleitet.

tionsmöglichkeiten, zum Beispiel zur Qualitätskontrolle oder für Rückmeldungen an Befragte.

Bei ML-Anwendungen an einem zentralen Ort, beispielsweise in einer Cloud oder auf den Servern der nationalen statistischen Ämter, können die Modelle komplexer sein. Sie können mehr Speicherplatz und Rechenleistung in Anspruch nehmen, da beides weniger beschränkt ist als auf mobilen Endgeräten. Dies könnte sich positiv auf die Qualität der Ergebnisse auswirken. Wartung und Aktualisierung der Algorithmen sind intern möglich, ohne dass die Anwendung der Nutzenden geändert oder aktualisiert werden muss.

Im Proof of Concept zum maschinellen Lernen wurde eine verallgemeinerte ML-Komponente entwickelt, die einem unabhängigen Ansatz folgt, um die Daten von verschiedenen Sensortypen zu verarbeiten. Dabei gelang es, aus den verschiedenen Sensordaten mithilfe des ML-Verfahrens mehrere Variablen abzuleiten, zum Beispiel die Art der körperlichen Aktivität oder das Verkehrsmittel der befragten Person. Hierfür wurde eine Wavelet + Convolutional-Neural-Network(CNN)-Pipeline entwickelt. Der Proof of Concept hat gezeigt, dass dieses sehr breit angelegte Instrument genutzt werden kann, um Sensordaten von Endgeräten so zu verarbeiten, dass nützliche, interpretierbare Variablen entstehen.

### Technologien zum Schutz der Privatsphäre

Wie bereits erläutert, enthalten Trusted-Smart-Surveys-Daten ein Übermaß an Informationen über die Befragten, was ein größeres Risiko für die Privatsphäre mit sich bringt als herkömmliche Umfragen. Traditionell werden Methoden wie die Pseudonymisierung verwendet, um zu verhindern, dass die privaten Daten der Befragten den Mitarbeiterinnen und Mitarbeitern in den nationalen statistischen Ämtern zugänglich sind. Außerdem werden digital erhobene Daten während der Übertragung verschlüsselt. Diese Maßnahmen sollen natürlich auch bei Smart Surveys angewendet werden. Möglicherweise verursacht die erhöhte Sensibilität der von den intelligenten Sensoren erfassten Daten jedoch strengere Datenschutzerfordernungen. Um diesen gerecht zu werden, könnten zusätzliche Maßnahmen erforderlich sein, zum Beispiel die Datenminimierung: Hierbei soll die Menge der in den gesammelten Daten enthaltenen Informationen auf das für einen bestimmten Zweck erforderliche Maß reduziert werden. Die Datenminimie-

rung schützt die Privatsphäre, indem sie die Möglichkeit ausschließt, aus den Daten zusätzliche sensible Informationen zu gewinnen. Dieses Konzept ist nur von begrenzter Bedeutung, wenn es sich um herkömmliche, fragebogengestützte Erhebungen handelt. Es ist jedoch äußerst relevant, wenn es um sensorgestützte Smart Surveys geht, da die Sensordaten (wahrscheinlich) weit mehr als die benötigten Informationen bieten.

Als zweite Methode sei die Verarbeitung und Analyse von Daten genannt, ohne dass die darin enthaltenen Informationen zugänglich werden. Dies ist die bei weitem stärkste Maßnahme zum Schutz der Privatsphäre, aber schwierig zu realisieren. Es gibt verschiedene Techniken, die eine verdeckte Analyse für unterschiedliche Anwendungen und Szenarien ermöglichen. So kann beispielsweise das Training von ML-Modellen durchgeführt werden, ohne dass es notwendig ist, die Daten zentral zu erfassen und zu speichern.<sup>18</sup>

Da solche Anwendungen zum Schutz der Privatsphäre für verschiedene GSBPM-Phasen (beispielsweise die Phasen 4 [Daten gewinnen], 5 [Daten aufbereiten] und 6 [Ergebnisse analysieren]) relevant sein können, ist die Entwicklung dieser Maßnahmen bereits bei der Konzeption der Erhebung mitzudenken. Es ist jedoch zu beachten, dass die Implementierung solcher Technologien mit einer erhöhten Komplexität der Plattformarchitektur und des Umfragedesigns einhergeht. Daher ist es wichtig, die Anforderungen an den Schutz der Privatsphäre auf Grundlage der Sensibilität der Daten genau zu bestimmen, um die Anwendung übermäßig komplexer Technologien zu vermeiden.

Auch im Bereich der Privatsphäre wurde ein Proof of Concept durchgeführt. Dessen Hauptziel bestand darin, die Anwendbarkeit von Technologien zum Schutz der Privatsphäre zu testen, um die Erstellung von Statistiken ohne Zugang zu Mikrodaten der Befragten zu ermöglichen. Dabei wurden einige Überlegungen zur Nutzung von Aggregatdaten in einer Multiparty-Computation(MPC)-Umgebung aufgezeigt und demonstriert, dass die Nutzung einer solchen Technologie grundsätzlich möglich ist.

Bei der Umsetzung zeigte sich, dass die Verwendung von Open-Source-MPC-Software viele zusätzliche Implemen-

---

18 Nähere Informationen dazu bietet der Abschnitt 3.2.3 in Deliverable 3.2.

tierungen erfordert. Einige dieser Implementierungen gehen über das übliche Software-Engineering hinaus und setzen Expertenwissen in Kryptografie oder Statistik voraus, um sie ordnungsgemäß durchführen zu können. Die entwickelte Anwendung erfüllt die für den Proof of Concepts gestellte Anforderung weitgehend, Auswertungen zu ermöglichen, Einzeldaten dabei aber zu schützen und nicht zugänglich zu machen. Allerdings umfasste die statistische Analyse nur sehr grundlegende Aggregationen. Um zu prüfen, ob ein solches Design alle Anforderungen der modernen statistischen Analyse erfüllt, sind weitere, tiefergehende Tests erforderlich.

## 4

---

### Ausblick

---

Der Einsatz von Apps und das Erheben von Sensordaten öffnet die Tür in eine Welt voller neuer Erhebungsmöglichkeiten. Trotz aller großen Potenziale kann dieser Weg aber nur dann erfolgreich zu beschreiten sein, wenn sichere Systeme und Vorgehensweisen bereitstehen, denen die Befragten vertrauen. Schließlich sind sie es, die in ausreichender Zahl und Repräsentativität an den Erhebungen teilnehmen müssen und ihre Daten zur Verfügung stellen sollen – nicht selten durch die Installation einer App auf ihren Smartphones. Nur mit ihrer Mitwirkung ist eine verlässliche Statistikproduktion möglich.

Das Essnet-Smart-Surveys-Projekt war ein erster Schritt. Durch grundlegende Überlegungen und Arbeiten wurden in Arbeitspaket 3 Möglichkeiten und Grenzen einer kombinierten Plattform und einzelner Datenverarbeitungsinstrumente betrachtet. Die vier Pilotprojekte aus Arbeitspaket 2 testeten mögliche Einsatzbereiche und Anforderungen aus der Sicht von Befragten. Insgesamt wurde deutlich, dass es noch einige Herausforderungen auf dem Weg zu einem sicheren, verlässlichen und transparenten Smart-Survey-Datengewinnungs- und -verarbeitungssystem gibt. Gleiches gilt für die Interaktion mit den Befragten an sich, um sie für diese neue Art der Datenerhebung zu gewinnen. Einen weiteren, nächsten Schritt in diese Richtung geht Eurostat mit dem Projekt „Smart Survey Implementation“, das an das ESSnet-Smart-Surveys-Projekt anknüpft. Zusammen mit zehn weiteren Projektpartnern wird sich das Statistische Bundesamt in den nächsten zwei Jahren an diesem Projekt beteiligen. Mit ersten Ergebnissen ist Mitte 2024 zu rechnen. 

## LITERATURVERZEICHNIS

---

Cukier, Kenneth/Mayer-Schoenberger, Viktor. *The Rise of Big Data: How It's Changing the Way We Think About the World*. In: Foreign Affairs. Ausgabe Mai/Juni 2013, Seite 28 ff.

Keusch, Florian/Struminskaya, Bella/Antoun, Christopher/Couper, Mick P./Kreuter, Frauke. *Willingness to Participate in Passive Mobile Data Collection*. In: Public Opinion Quarterly. Ausgabe 83/2019, Seite 210 ff.

Ricciato, Fabio/Wirthmann, Albrecht/Giannakouris, Konstantinos/Reis, Fernando/Skaliotis, Michail. *Trusted smart statistics: Motivations and principles*. In: Statistical Journal of the IAOS. Ausgabe 35/2019, Seite 589 ff.

Ricciato, Fabio/Wirthmann, Albrecht/Hahn, Martina. *Trusted Smart Statistics: How new data will change official statistics*. In: Data & Policy. Ausgabe e7/2020.

Ricciato, Fabio/Giannakouris, Kostas/Wirthmann, Albrecht/Hahn, Martina. *Trusted Smart Surveys: a possible application of Privacy Enhancing Technologies in Official Statistics*. Conference paper: 50th Scientific Meeting of the Italian Statistical Society (SIS 2020). Rom 2020.

Ruppert, Evelyn/Grommé, Francisca/Ustek-Spilda, Funda/Cakici, Baki. *Citizen Data and Trust in Official Statistics*. In: Economics and Statistics. Ausgabe 505–506/2018, Seite 179 ff.

Stehrenberg, Shari/Giannakouris, Konstantinos. *Trusted Smart Surveys: Solutions for the European Statistical system: An overview of the objectives and the main challenges*. Conference paper: European Conference on Quality in Official Statistics (Q2022). Vilnius 2022.

Struminskaya, Bella/Lugtig, Peter/Keusch, Florian/Höhne, Jan Karem. *Augmenting Surveys With Data From Sensors and Apps: Opportunities and Challenges*. In: Social Science Computer Review. Ausgabe Dezember 2020, Seite 1 ff.

## RECHTSGRUNDLAGEN

---

Verordnung (EU) 2016/679 des Europäischen Parlaments und des Rates vom 27. April 2016 zum Schutz natürlicher Personen bei der Verarbeitung personenbezogener Daten, zum freien Datenverkehr und zur Aufhebung der Richtlinie 95/46/EG (Datenschutz-Grundverordnung) (Amtsblatt der EU Nr. L 119, Seite 1).

**Herausgeber**  
Statistisches Bundesamt (Destatis), Wiesbaden

---

**Schriftleitung**  
Dr. Daniel Vorgrimler  
Redaktion: Ellen Römer

---

**Ihr Kontakt zu uns**  
[www.destatis.de/kontakt](http://www.destatis.de/kontakt)

---

**Erscheinungsfolge**  
zweimonatlich, erschienen im Februar 2023  
Ältere Ausgaben finden Sie unter [www.destatis.de](http://www.destatis.de) sowie in der [Statistischen Bibliothek](#).

---

Artikelnummer: 1010200-23001-4, ISSN 1619-2907

---

© Statistisches Bundesamt (Destatis), 2023  
Vervielfältigung und Verbreitung, auch auszugsweise, mit Quellenangabe gestattet.