Arian Rajh[1]

ARCHIVING BEYOND DOCUMENTS: HEALTH AND SPORTS DATA FROM WEARABLE SENSORS IN A DATA-INTENSIVE WORLD

Purpose: *Wearables in the health and sports domains have spread significantly over the past period, and this article analyses several wearable devices and systems used in the treatment of diabetes and in tracking endurance sports.*

Method/approach: *The theoretical framework this research uses is a Foucauldian analysis of datafied reality, alongside both discursive and materialist views of data.*

Results: *The author examines data generated by users and wearables as material-discursive phenomena that rely on data imaginaries and enable specific users' actions over data. After analysing threats to the continued use of data from wearables, the author emphasises the importance of preventing data loss. The author explores whether wearable data should be archived as data or as documents.*

Conclusion/findings: *The author discusses the value of data, recommends using archival strategies to restore user control over their data, and proposes developing a data-driven archival apparatus.*

Keywords: *data imaginary, data archiving, deplatforming, diabetes, health and sports data, personal data governance, sensors, wearables.*

1 Arian Rajh, PhD, Assoc. Prof., University of Zagreb, Faculty of Humanities and Social Sciences, Ivana Lučića 3, Zagreb, Croatia, email: arajh@ffzg.hr.

1 INTRODUCTION

A document is information with a fixed constellation of certain elements of form. Following Duranti (1989, 15; 1991; 2010, 81) and Lemieux (2024, 144), a document's constituents are its form, the action and activities of agents, an archival bond with other documents, and adherence to technological and social contexts. This article explores questions about data archiving, particularly when data rather than documents are generated. Of course, documents still have a place in today's world as fixed-information objects of a particular type, but there is no reason for an archivist to neglect other forms of personally and socially significant information. Although the global quantity of data by the end of 2025 has reached approximately 180 zettabytes (Statista, 2025), making data the primary form of information today, and will continue to grow at even higher rates than before, many archivists still refuse to deal with data on the theoretical and practical level and continue to focus just on documents. Data is not even considered a proper topic of our discussion or archival scientific communication. Maintaining this ignorant attitude may lead to losing oversight of data collected by large tech companies, as they have their own approaches to managing and protecting it. It also renders our profession largely irrelevant in the age of "data-driven capitalism" or "platform capitalism" (Acker, 2025; Srnicek, 2016). Tech companies appear to have recognised the merits of data earlier than other key stakeholders and have assumed an overly large role in the user-engineer-archivist "three-way relationship" (Moss et al., 2018, 146). We live in an era of the "connective turn" and the omnipresence of digital technologies, networks, and storage (Hoskins, 2011; Hoskins, 2017), yet we witness the failure of archiving (Hoskins, 2017, 4). This article points to the pitfalls of this conservative archivist position, which neglects data generated by monitoring conditions or diseases and by tracking fitness, sports achievements, and general bodily activities. Data from patients and athletes that are generated and collected by contemporary wearable technologies.

In this article, wearable tracking technologies, or wearables, refer to Internet of Things (IoT) devices with sensors, connected to software, linked to systems and ecosystems, and used for health or sports purposes. The proliferation of wearable technologies in sports over the last few decades has been noteworthy (Toner, 2023, 19, 26–27, 29). There has also been an explosion of wearable technologies for monitoring various health conditions (Majdinasab in Mitsubayashi, 2024, 42). Wearables and their application environments, or devices that continuously collect personal health and performance data, fall under the Internet of Things (IoT), Big Data, and AI spheres. Talboom and Huentelman (2018) describe them as continuously attached or applied devices that are minimally or not at all invasive for the user and use sensors to gather and transmit data. Domains that use wearables, according to Talboom and Huentelman, include disease-related domains, childhood development domains, and general physical monitoring, but other domains could also be metricated using wearables.

Wearables are worn directly on or in the body (e.g., glucose sensors, 2D fabrics and patches, implanted wearables) or as separate devices (e.g., rings, bands, sports watches and smartwatches). They could also include clothes, contact lenses, insoles, glasses, goggles, shoes, monitoring suits, and so on (Vaghasiya et al., 2023, 2; Vijayan et al., 2021, 1; Xue et al., 2024, 4). There are also combinations of implantable and wearable solutions, such as the Eversense glucose monitoring system, which plans to eliminate the separate wearable transmitter component and enable Bluetooth connectivity between the sensor and a transmitter implanted under the patient's skin. A typical wearable ecosystem consists of one or more devices from a single vendor, each with accompanying software. This can be expanded into a broader system that includes other vendors' products. Such a system includes a wearer with accumulated data, one or more wearables with sensors and transmitters, smartphone software, the first vendor's cloud environment, and other vendors' cloud environments and apps. As stated, this article examines two types of wearable technology, both widely used today: health condition-monitoring and sports-tracking wearables.

1.1 WEARABLES FOR MONITORING SPECIFIC CONDITIONS OR CHRONIC DISEASES – USE CASE OF DIABETES

This group of wearable devices includes various sensors and transmitters for tracking specific health parameters. Processes involving this type of data in the literature are also referred to as "sensor data analytics" (Karthick, 2026). As stated in the literature, they are used both in and out of hospitals, facilitating the transition of treatment to patients' homes, either by patients themselves or under the supervision of healthcare professionals (Vijayan et al., 2021, 3). While more accurate methods exist for determining immediate treatment, such as invasive glucometers for measuring blood glucose levels in patients with diabetes, data from wearables are frequently used to manage treatment continuously and to keep patients in low-risk conditions (e.g., glucose levels of 4-10 mmol/L for patients with diabetes).

1.2 WEARABLES FOR TRACKING FITNESS, SPORTS ACHIEVEMENTS, AND GENERAL CONDITIONS

This group of wearable devices includes sports or smartwatches, shoes and other type of sensors for tracking the number of steps, jogging, running, swimming or cycling speed, foot balance while running, running pace, heart pulse while resting or physical activity, pulse variability, calories burned in activities, minutes of exercise, exercise effects, altitude difference between start and end of activity, sleep time and quality, etc. Their sensors collect static, kinetic, and kinematic data about the body and its activities (Toner, 2023). Professional and recreational athletes use them to monitor conditions during activity and track their athletic development. "Quantified-self" population (Hill, 2011; Lupton, 2013, 25), people who do not have to be athletes, also voluntarily use sports and other wearables to track their overall health or parameters of interest.

1.3 RESEARCH QUESTIONS

This research on data from health and sports wearables covers the following questions:

RQ1: What are the implications of leaving the preservation or archiving functionality to technology companies that create wearables for collecting user data?

RQ2: Is there a professional responsibility of the archival community in these cases?

RQ3: Should the archival strategy be focused on archiving documents or data?

RQ4: Should archival discourse change to archive data generated by wearable technologies in the targeted form?

2 APPROACH AND METHODOLOGY

The methodology of this research comprises a review of the literature on social changes today, focusing on the intensification of data use in general, the use of data in domains of sports and health activities and conditions, and document and data objects in archives, as well as analyses of use cases regarding data formats and their ecosystems. Secondly, both approaches, i.e., management that focuses on data and management that focuses on data derivatives such as documents, will be analysed in cases involving wearable-generated data and discussed. In the case of wearables, the documents are derivatives of the original data. Thirdly, the author will present the answers to the research questions. Finally, after discussing these answers, the author will draw a conclusion and evaluate the potential impacts on data archiving and archival science.

2.1 REVIEW OF LITERATURE AND THEORETICAL FRAMEWORK FOR THE RESEARCH

Today, almost everything is followed by data or "datafied" (Mayer-Schönberger & Cukier, 2014, 78, 91). Many authors describe today's data-rich or data-intensive world, ranging from its dataism ideology to the social impact of the changes it has brought (Beer, 2019; Couldry, 2020; Mejias & Couldry, 2024; Zuboff, 2019). For this article, some newer articles and books on an archival perspective on data have been consulted (Acker, 2025; Mordell, 2019; Moss et al., 2018; Payne, 2018; Rajh, 2025). The author also consulted the literature on wearable technologies (De Arriba-Pérez et al., 2016; Karthick, 2026; Mitsubayashi, 2024; Szeto et al., 2024; Talboom & Huentelman, 2018; Toner, 2023; Vaghasiya et al., 2023; Vijayan et al., 2021; Xue et al., 2024). The theoretical framework for this research draws on neo-Foucauldian theories, including those of David Beer and Amelia Acker, as well as Karen Barad's agential realism within the New Materialism framework.

Beer draws parallels between power and social structures, and between the clinical-scientific practices described in Foucault's "The Birth of the Clinic" (2003) and the datafication movement of our own time. Data is now included in all aspects of our lives, from the workplace to health and leisure, and it enables the formation of knowledge and insights into social life (Beer, 2019, 4–5). Beer explains the infrastructure stat supports data processes and analytical practices, aiming to understand the data industry's role in our societies. His "data gaze" is the concept that relies on the ideology of the "data imaginary," or the promise of what could be accomplished through data, and connects today's data analytics and power structures. IT companies and data centres create new landscapes for data gaze, for its rationality and expanding worldview (Beer, 2019, 130–132).

Acker's "Archiving machines" also examines data and platforms as means and channels of new relations of power (Acker, 2025, 157). With the advancement of IT and telecommunications, the users, as Acker states, gave up control of our data for the convenience of technical solutions and their functionality. In her view, platform-based companies separated users from their data (externalisation), changed users' expectations and behaviour (the grammar of action), and imposed their own meaning of the term "archiving." This relationship with private companies is neither collaborative (Farrugia, 2024, 31) nor healthy.

Finally, Barad's agential realism brings together Foucauldian poststructuralism, realism, feminist theory, and related strands of critical thought with the physics-philosophy of Niels Bohr, who addressed the problem of light's wave-particle duality through the principle of complementarity, arguing that different scientific apparatuses capture different aspects of light. While Barad adopts a realistic ontology, she rejects both representationalism and pure constructivism of the world (by the discourse) in favour of performativity (Barad, 2007, 46–49). She reconciles the materialistic and discursive nature of the world in which humans participate, rather than occupy a central position. Within this framework, phenomena of the world consist of objects and scientific apparatuses that we use to measure and comprehend them. In this case, measuring is not a representation of the object but a material-discursive intra-action. Agency involves actively participating in the creation of phenomena that delineate the boundaries between subjects and objects. Archival materials do not merely record events but actively shape them as material configurations that contribute to the creation of new meaning. Digital data and records are not just discursive; they are also physical phenomena.

Barad's theory and similar theories are beginning to influence archival science (Goudarouli & Prescott, 2025). There is common ground among all these theories, which constitute the framework for this article. For example, the automatic save function in software is, for Acker, a grammar of action and, for Barad, an active agency of software. Beer's "data imaginary" corresponds to Acker's notion of a common belief in data's predictive power, so both outline the ideology behind the data movement. They all see not just raw data but information-related phenomena, products of apparatuses used today. The SWOT analysis presented in the results of this work was broadened in the discussion through a less technical, more philosophical literature on wearables, such as that by De Arriba-Perez et al. (2016), Sharon (2017), and Toner (2023).

2.2 ANALYSIS OF WEARABLES AS DEVICES AND THEIR ECOSYSTEMS

Following the literature review for this article, this section aims to analyse systems, data-related processes, and data as phenomena related to Type 1 diabetes (T1D) wearables, such as Continuous Glucose Monitoring (CGM) devices, from the first group. It also examines performance-tracking sports watches from the second group. Due to their availability and widespread presence in the EU market, three CGM products are being analysed in the first group of devices: Abbott FreeStyle Libre (generations 2 and 2 Plus), Dexcom One+, and Sibionics. All these CGMs can be used with Android and iOS smartphones, providing an easy-to-use experience, along with the additional LibreView application for family members and healthcare professionals. The analysed systems employ a flexible filament that is minimally invasive and reaches the fluid surrounding cells in the tissue under the skin to measure glucose levels.

The FreeStyle CGM ecosystem, manufactured by Abbott, includes a device and software for continuous glucose monitoring and diabetes treatment, featuring a sensor and transmitter integrated into a single device, a smartphone application, and a cloud services solution. It can use third-party vendor information on smart insulin pens, such as the Novo Nordisk NovoPen Echo Plus, to track daily insulin intake. The Libre sensors measure glucose levels in interstitial fluid, which reflects changes in blood glucose levels with time delays of approximately 15 minutes (Libre 2) or 2 minutes (Libre 2 Plus). The manufacturer calibrates the sensor, but the wearer cannot calibrate it independently. Collaborating with third-party partners, Abbott has used the Libre sensor and their insulin pumps in specific configurations. PC/Mac application represents glucose level statistics, average glucose levels, glucose variability, daily glucose profiles, and glucose management indicators (GMI). GMI indicates or estimates the average three-month glucose level and should be confirmed by laboratory analysis.

The expanded ecosystem for Dexcom One+ CGM (shown in Figure 1 below) comprises a smartphone app, the Apple Watch Dexcom app, and Dexcom Clarity cloud software. The sensor measures interstitial glucose every 5 minutes. The system includes the Dexcom Clarity application for home users (patients or parents) and healthcare professionals. Dexcom Clarity provides reports that include average glucose values; time-averaged glucose values considered appropriate for persons with diabetes (4.0 to 10.0 mmol/L); information on insulin intake; the best day for maintaining glucose levels; data visualisation; trends; comparisons of values for selected days; and ambulatory glucose profile information. Wearers can calibrate the sensor on demand after measuring their blood glucose levels. The EU-marketed Dexcom device does not integrate with smart insulin pens; the patient must manually enter insulin doses. When integrated with tubeless insulin pumps, Omnipod 5 features an automated insulin delivery system that uses advanced Dexcom CGMs available in the US (Dexcom G7) to automatically adjust insulin doses. The Dexcom One+ sensor-transmitter is fully waterproof (up to approximately 2.5 metres), which can be critical for wearers with diabetes who lead active lifestyles, such as swimmers or triathletes, and is beneficial for managing their condition.

Figure 1: Dexcom One+ is combined with a device accompanying the sensor, a smartphone application, and a smartwatch

Sibionics CGM consists of the sensor, additional patch, and smartphone app. The sensor sends data to the app at 5-minute intervals, and patients can view values in 3-, 6-, 12-, or 24-hour segments. The Sibionics app provides diary functionality with time-in-range, time-above/below-range percentages, daily trends, and events such as meals, exercises, additional blood glucose measurements, and insulin intakes. The user doesn't estimate carbohydrate intake; instead, they choose meals from a database that provides calorie and nutrient values, like calorie-counting applications.

Figure 2: Garmin Fenix6 sports watch and HR monitor (parts of the Garmin ecosystem).

The Apple Watch and Garmin ecosystems belong to the second group of wearables being analysed in this article. Both sports tracking systems allow third parties, such as the social network Strava, to ingest data they collect. The Apple sports-tracking ecosystem consists of a smartwatch, the Apple Fitness app and Health app on the phone, and iCloud synchronisation. The literature describes

Apple's Health system (De Arriba-Pérez et al., 2016, 6). Fitness and sports tracking functions include pedometer data, calorie data for activities and rest, daily walking distance and activity data, running and walking paces, cadence, geolocation data for activities, various heart rate (HR) data, etc.

Garmin's ecosystem includes various sports watches and additional sensors that track data more precisely (e.g., running dynamics and HR via a separate heart rate monitor connected to the Fenix watch, as shown in Figure 2). Garmin Fenix 6 and 8 watches and triathlon HRM were analysed. The system applications are the Garmin Connect smartphone application and the Garmin Connect cloud solution. Although sensor accuracy is not perfect, many sports watches offer sufficient accuracy for personal use (Jamieson et al., 2024, 11)3-min step test (3MST. Garmin is reputed for its reliable algorithms, and accuracy increases with the addition of sensors to its ecosystem. Its race preparation plans and race PacePro strategies utilise predictive analytics. Transferring data from a Garmin watch to Strava can be considered an indirect warehouse data transfer (De Arriba-Pérez et al., 2016, 8–10), in which the data flows through intermediate systems without direct communication between the Garmin watch and Strava. The watch records activities, and the data are synced to Garmin Connect. If the wearer links their Connect and Strava accounts, Garmin's backend systems push activity data to Strava's servers. There was a dispute between Strava and Garmin in 2025, but Garmin's sports data can be displayed on both systems. Garmin Connect acts as a warehouse, aggregating and processing data before enabling its use in the Strava environment. Strava also enriches the data by applying its algorithms and adding comments from other community members, thereby satisfying users' need to share data and compare with others (Grüning & Richlan, 2026), as well as their general "compulsion of connectivity" (Hoskins, 2017, 2).

Several types of sensors used by wearables are categorised and listed in the literature (Vaghasiya et al., 2023, 3–4; Xue et al., 2024). The showcased CGMs use electrochemical sensors for glucose monitoring (Majdinasab in Mitsubayashi, 2024, 41–43), and the sports watches described in this article rely on a combination of motion sensors (accelerometer, gyroscope), environmental and navigation sensors (ambient temperature, barometer, GPS), and optoelectronic sensors (photoplethysmography for HR). These sensors gather data that must be processed by accompanying software. Data are transmitted and stored within the system. The article's upcoming section 2.3. will focus on long-term preservation tasks that may extend beyond the original system's borders, beginning with the format issue as a foundation for further preservation efforts.

2.3 ANALYSIS OF FORMATS FOR PRESENTING HEALTH AND SPORT DATA

Data on specific health conditions and recording file formats were analysed using the primary example of the Dexcom Clarity application (version 3.51.0) with the export function. Two other CGMs were also considered, both from literature (Roze et al., 2021; Visser et al., 2024) and practical cases. Data from Dexcom Clarity can be exported as Excel files and thus converted to CSV files. The Excel file contains glucose levels, times, warning events, insulin and carbohydrate intake, and technical metadata for wearable devices.

Dexcom Clarity records sent by mail provide weekly summaries of patients' time within appropriate glucose levels, with incremental or decremental changes compared to the previous week's summary, and visualised trends. The following reports from the Dexcom Clarity application can be exported as PDF files: overview report, pattern report, overlap report, daily reports, comparison report, daily statistics, hourly statistics, and Ambulatory Glucose Profile (AGP). The overview report includes average glucose over 2 weeks, time in range, sensor usage, percentages of fast- and long-acting insulin, average total daily insulin intake, best glucose day, and device metadata. The patterns report provides information about the patient's optimal day, while the overlay displays daily glucose graphs stacked on top of each other. The Daily report presents glucose graphs and event information for each day. A comparative report assesses biweekly periods, and a statistics report displays daily and hourly measurements in tabular form, accompanied by visualisations. The AGP provides insights about the time the patient spends in a good glucose range, average glucose levels, the Glucose Management Indicator (GMI) that estimates laboratory values expressed as a percentage, the coefficient of variation, the timing of CGM activity, a visually represented ambulatory glucose profile, and daily glucose profiles for all days within the reporting period. Libre CGM (version 2.11.2.8275) provides reports with statistics, time-in-range data, AGP data, daily glucose profile, trends, trends after meals, monthly summaries, weekly summaries, and a diary with all events in the reporting period. Sibionics' reports (app version 01.12.00.00) include time-in-range data as percentages and hours, average glucose value, GMI, glucose variability (compared to the goal), and daily glucose profiles.

Sports performance data and record types have been analysed using an example of a running session recorded by a Garmin sports watch. Exporting data from Garmin Connect as a binary FIT file (Flexible and Interoperable Data Transfer (2019)) is possible. TCX (Training Centre XML) and CSV (Comma-Separated Values) files are also available. There are also XML-based files containing geospatial and sports data, as well as GPX (GPS Exchange Format) and KML (Keyhole Markup Language) digital files. An author's half-marathon recorded activity is an example to analyse the file formats below (the 'Starek' race, held in November 2024, Zagreb, Croatia). The FIT file for the race includes a file identifier, software version information, and timestamps for the event start, markers, and stop moments. It contains details about the wearable device and the wearer's personal data, including sex, height, weight, and HR (resting, maximum, threshold). It also provides information about the type of sports session, the wearer's activity (speed, vertical oscillation), external conditions (temperature, altitude), and details about laps and segments of activity and the overall session, as in the Fit File Explorer application (v. 3.5, 2023, developed by Brice Rosenzweig). A CSV file contains less information, organised by running segments or laps: lap time, distance, average pace, average and maximal HR, ascent and descent data, average power, average cadence, stride length, and other running-related data. A TCX file is an XML file with the root TrainingCenterDatabase element that lists Garmin's namespaces and includes elements for the activity and the software agent responsible for recording it. The activity element consists of sub-elements for describing laps: lap start time, total time in seconds, distance, maximum speed, calories spent, HR values, intensity, trigger method, and tracks. Tracks are smaller segments of geospatial data, including position, altitude, and distance from the start of a run to the end of a segment, along with the wearer's data, such as speed and running cadence. The GPX files (shown in Figure 3 below) contain information about the running path (the trk element), including a track segment (trkseg) sub-element and lower-level sub-elements, such as track points (trkpt). Track points contain geospatial data, elevation data, a specific time point, and extended HR and cadence data.

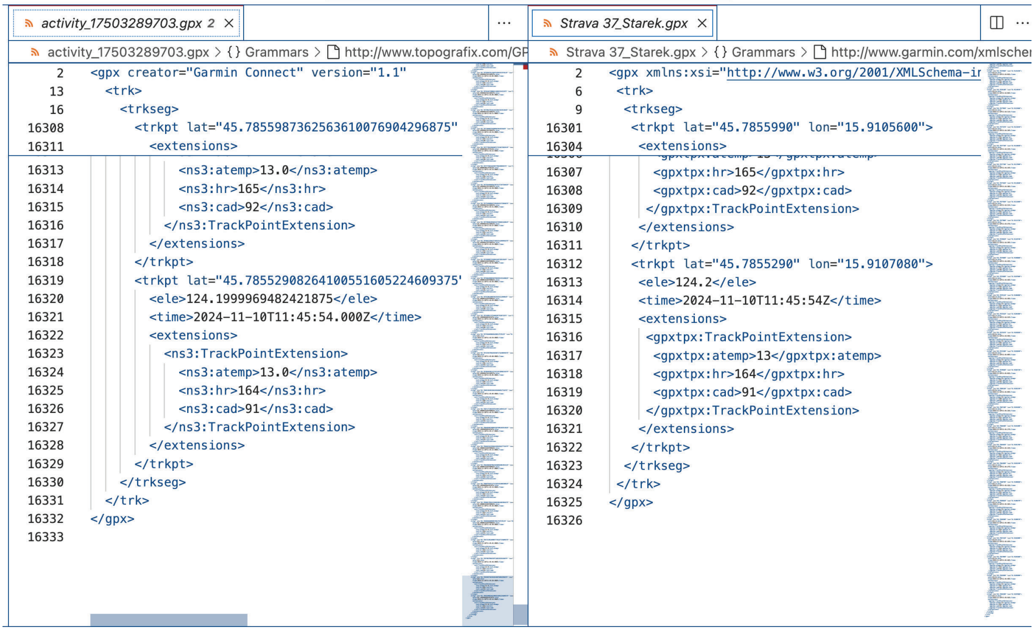

Figure 3: GPX files from Garmin (left side) and Strava applications describing the same activity

The Garmin half-marathon activity was also shared on Strava, from which FIT and GPX files were exported. GPX files generated by both applications were similar, as expected, since Garmin data was processed by Strava and inserted into an XML template following the same standard. FIT files were based on the same data, but there were minor differences in structure and visualisation, as shown in Figure 4. It is also possible to create PDFs from Strava by using a browser's PDF export feature.

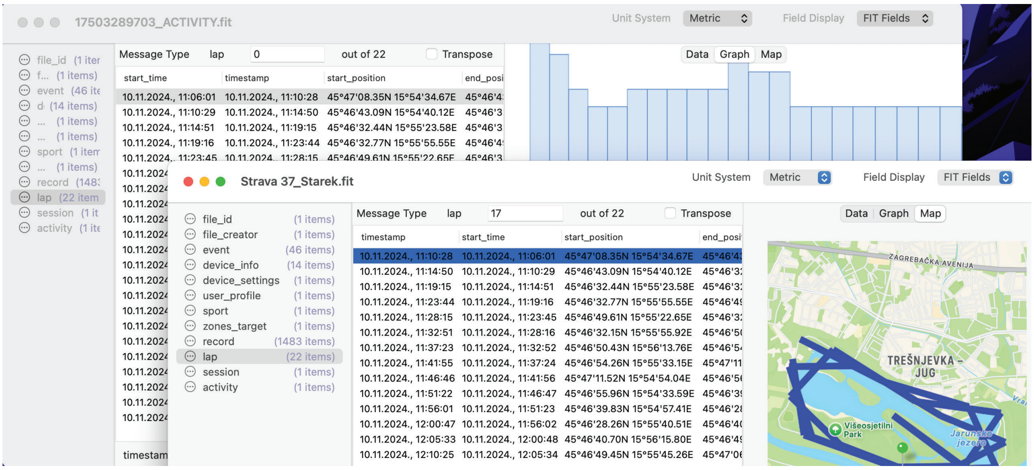

Figure 4: Garmin and Strava FIT files compared (the file created by Strava in the front), presented in the Fit File Explorer application

3 RESULTS

Data elements created by two groups of wearable technologies, described in the previous section and summarised here, capture different phenomena, as shown in Table 1 below. The table shows an implicit imaginary (Beer) of information collected by those wearables, which is involved in the creation of the patient or athlete as datafied subjects (Barad), material-discursive phenomena created by these wearables (Barad) that serve to reshape patients' and athletes' bodies, and grammars of user actions and excluded actions (Acker) with the data and phenomena. Data imaginary is what is promised, phenomena are what are being created as information objects, and grammars of actions refer to what is permitted and banned.

Wearable	Promised	Created	Permitted	Excluded
Abbot Libre	Continuous diabetes monitoring; routine disease management; quality of life optimisation	Interstitial glucose; hypo/hyperglycaemic events; GMI; glucose variability; time-in-range; glucose profiles	CSV export (partial); report downloads (document-based)	User calibration; full raw data export
Dexcom			Report downloads (document-based)	Automated insulin data integration; full raw data export
Sibionics				User calibration; automated insulin data integration; raw data export
Garmin Fenix	Performance optimisation; longevity; self-improvement; preventing threats; return of investment in an athlete	Training readiness; HRV status; race prediction; training load/effect	Performance analysis; historical trend review	Full raw data export
Apple Watch	Holistic life management; everyday self-tracking	Activity metrics, trends, awards, and aggregated health indicators	Data visualisation; limited export via ecosystem	

Table 1: Data imaginary, phenomena, and grammars of actions for the analysed wearables

Toner critically examines the proliferation of wearables in the sports domain and links it to the perception of analytical apparatus that provides insights and predictions about the wearer's performance. Toner sees this as the approach to managing athletes in the current biopower landscape, alongside its monitoring, risk-mitigating, and discipline systems (Toner, 2023, 30–31). According to Toner, the hidden goal of using wearables is to prevent athletes from failing to reach their potential, as identified through predictive analytics, thereby avoiding the risk of not generating profit in today's competitive world. The literature on medicinal wearables highlights the potential benefits for patients (Inoue, Yokota, and Takewa & Mitsubayashi in Mitsubayashi, 2024, 19). The final column in Table 1 highlights the actions excluded from the grammars of action governing the analysed wearable ecosystems. Although these systems generate data ostensibly for patients and athletes, access to that data remains restricted, rendering it "asymmetrical" and increasingly „distant" from users (Acker, 2025, 137). When limitations are embedded at the point of data creation, concerns regarding long-term preservation and archival responsibility inevitably arise. This raises a fundamental question: whether the archiving function can be left to the technology companies that design and control wearable platforms, or whether archivists bear responsibility for intervening in the governance and preservation of such data (RQ1, RQ2).

Table 2 below presents strengths, weaknesses, opportunities, and threats related to the preservation of data or documents that contain processed data. Data from the examined systems can be exported and preserved in several ways. Diabetes-related data can be exported from various CGM systems as PDF files or CSV container files. Sport-related data can be exported as XML files (TCX and GPX) or in binary FIT format. Diabetes-related data could not be ingested into another CGM system, but sports-related data can be ingested into third-party systems like Strava (TCX, GPX, and FIT files). Examined wearables that generate sports-related data demonstrated greater flexibility in using export and preservation formats outside their original ecosystems.

<table>
<tr><th>Approach</th><th>Strengths</th><th>Weaknesses</th><th>Opportunities</th><th>Threats</th></tr>
<tr><td>Preservation of health data as documents</td><td>Preservation of data in a standardised PDF file format.</td><td>Data is encapsulated in static files that are not easily processable.</td><td>PDFs can be transformed straightforwardly into archival PDFs; using the archival document format adds reliability.</td><td>Static representations may become impossible to reuse as analytical needs evolve, information is discontinued, interoperability is limited, or an archival strategy is lacking.</td></tr>
<tr><td>Preservation of health data as data</td><td>Exported data in standard CSV format.</td><td>Limited health data formats hinder continuity of monitoring.</td><td rowspan="2">Data could be reused and processed in more advanced ways at some point.</td><td rowspan="2">Data is represented as neutral, although it is already pre-processed; limited interoperability and functionality of the chosen data containers; a lack of context; and a lack of an archival strategy.</td></tr>
<tr><td>Preservation of sports data as data</td><td>Data export (binary and XML containers) supports continuity and interoperability as files can be manually uploaded to some third-party systems.</td><td>Data derived from company-dependent metrics and algorithms.</td></tr>
</table>

Table 2: SWOT analysis of different approaches to archiving sports and health data

The current situation, characterised by varied preservation approaches, indicates a lack of a systematic archiving strategy, posing a threat to hundreds of millions of patients with both types of diabetes (World Health Organisation, 2024) and to athletes, who might lose their valuable health and performance data. Another group of sensor data producers includes ageing populations, which further exacerbates the problem, but this is outside the scope of the article. Archivists are committed to preserving and assuring access to information valuable to individuals, groups, or societies. Information in the past was typically shaped as a document. In contrast, data generated by wearable technologies today is a different type of informational object – dynamic, deeply system-dependent, and continuously produced stream of data. This divergence calls into question whether such data, created as data, should be preserved as typical, static archival documents, usually in PDF files, or as data (RQ3), and whether existing archival discourse and practice are adequate for preserving these forms in their native, data-centric configurations (RQ4). The approach of archiving documents by the wearables' software has few advantages but more severe weaknesses and threats (as shown in Table 2), so preserving personal sports activities by archiving fitness-related datasets and formats would make more sense. FIT and TCX files focus primarily on fitness-tracking data, while GPX and KML files integrate geospatial data and fitness and sports-session information. FIT and TCX file formats are industry standards that are not globally accepted. Globally accepted file types are GPX and CSV. The XML files for describing the half-marathon activity obtained from the Garmin ecosystem were relatively large. Still, the GPX file was approximately 40% smaller (by line count and disk size) than the KML and TCX files derived in one of the analysed use cases and compared. Preservation of sports data as documents was not analysed in Table 2 because the creation of a PDF with data processed by Strava was possible by using a browser's PDF saving feature. Rather than a purely technical problem, the issues analysed point to a broader need to reorient archival theory and practice towards the preservation of data as information in adequate forms, as this affects individuals and patient communities.

4 DISCUSSION

Although dataism ideology and data imaginary were criticised in the literature (Acker, 2025:129–130; Beer, 2019:18), both health and sport data collection and processing rely on data for predictions at the individual level, e.g., early detection and individualised treatment for a diabetes patient (Karthick, 2026, 1) and race-distance predictions for an athlete. The reality, as always, is in the middle: despite all the processing techniques, sensor data can still produce errors, but it also makes life much easier for a person with diabetes. Data from wearables can be beneficial for planning rest and training of an athlete or recreational athlete. Yes, wearables are a part of today's platformed world. Predictive possibilities have limitations; "data imaginary" is a set of promises, but data can be used to predict some consequences of the state from which the data is collected. This research does not place itself within the dataist literature; instead, it highlights the benefits of sensor technologies and encourages their use with self-awareness. But more importantly, given the recognised threats, it calls for the addition of an archival strategy to the export, preservation, interoperability, and use of data collected by wearables.

The literature also identifies additional positive and negative aspects of wearables, complementing the SWOT analysis presented in Table 2. Sharon (2017, 106–107) describes wearable technologies, identifying empowerment to manage one's health, individual responsibility for personal health, and the facilitation of self-knowledge as key positive attitudes. The values Sharon associates with wearables are *solidarity*, *autonomy*, and *authenticity*. Preventing unnecessary costs for overburdened health systems contributes to overall social solidarity, autonomy refers to gaining control over one's personal health, and understanding oneself leads to a more authentic self. On the other hand, loss of autonomy and the imposition of control can be observed when wearables are not used voluntarily. Further loss of solidarity, as Sharon states, occurs when healthcare is left to individuals, and the quantification of various states can further distance individuals from genuine experiences. Importantly, Sharon's framework focuses on wearable systems rather than on the data they produce.

This article extends that discussion by shifting the focus from devices to the nexus of wearable systems and data. Wearables, the domain discourse (diabetes, running, cycling, swimming, triathlon, etc.), and the body of a patient or athlete comprise an apparatus in Barad's sense. The data produced are not raw but material-discursive, processed phenomena (as shown in Table 1) that must be maintained and preserved over time. We need to include an archival strategy within those apparatuses to ensure this. The empowerment that wearables offer ends when users lose control of their data. The literature also indicates that autonomy diminishes as users increasingly lose control of their data (Acker, 2025, 82, 110, 169). In addition to Sharon's values, the author of this article would identify continuity, interoperability, and reliability as guiding principles for data selection and management that archivists should adopt. These principles directly concern how wearable data persist and remain reusable over time.

The first guiding idea concerns maintaining *continuity*. Continuous monitoring or tracking of medical and sports phenomena aids in diagnosis, treatment, and performance planning. However, interruptions and data discontinuity caused by platform obsolescence, device and technology switching, or static data in documents undermine these advantages. Wearable data can be exported as static PDF documents, which prevents recombination with newly generated datasets and results in information that is not readily reusable.

Valuing *interoperability* is also a key guiding idea. Efforts toward standardisation facilitate users' ownership of data and third-party integration, which is critical to switching systems and, again, user independence. Weak interoperability increases threats and undermines continuity. Sensor data fusion, i.e., combining multiple data streams (Karthick, 2026, 14–15), is essential for advanced analytics. This also favours the argument that wearable data should not be preserved merely as static documents. Interoperability was also discussed by Arriba-Pérez, Caeiro-Rodriguez, and Santos-Gago (De Arriba-Pérez et al., 2016, 3).

Finally, *reliability* is a very valuable data characteristic, and we need to maintain it. The guiding idea here concerns the seamless integration of wearables into daily life and the sensor accuracy that supports well-informed health- or performance-related decisions. Unreliability can result from sensor inaccuracies or battery loss. A broader context loss when data is discontinued and segmented also affects reliability. Wearables do not directly replace human medical or sports coaching expertise; instead, they depend on insightful contextual interpretation of trustworthy data. Unreliable data and misinterpretation can make data unusable and destroy users' trust.

There is additional literature that addresses additional threads related to wearables and their data and describes mitigation measures (Apple, 2023; Talboom & Huentelman, 2018; Vijayan et al., 2021). Despite some standardisation efforts, the domain is still too heterogeneous. Proprietary formats, such as Garmin's FIT and TCX, coexist with open standards, such as Topografix's GPX format. Health-related data rely on condition-specific standards (e.g., HL7, DICOM, IEEE 11073-10417-223 for glucose monitoring systems (2023), and ISO medical device standards). However, there is no comprehensive framework that addresses data containers and directs the preservation of data generated by wearables. Leaving standardisation solely to technology companies risks further entrenching the domains' deep dependence on platforms, as Acker aptly noted (Acker, 2025, 156). From an archival perspective, wearable data should be actively "de-platformed" (Acker, 2025, 147). Platforms are not the enemy here; they drive forward progressive data use, but they are not the sole stakeholders and should certainly not own private data. To address these challenges, the development of semantically rich, machine-readable, and interoperable data formats should be prioritised and incorporated into archival strategies. Linked data approaches offer advantages over overly simple, yet voluminous, XML-based formats for managing complex sensor data. Relevant ontologies usable for sports data include WGS84 Geo, GeoSPARQL, the Time Ontology in OWL (2022), the Semantic Sensor Network Ontology (2017), the Quantity-Units-Dimensions-and-Types ontologies, Schema.org, and PROV-O (2013). Although there are standards and terminologies in the health super domain (SNOMED, DICOM, etc.), ontology frameworks suitable for many disease- and condition-specific monitoring processes still need to be developed.

5 CONCLUSION AND FURTHER DIRECTIONS

Large amounts of personal and personally identifiable data are being generated and must be continually managed if they are to be used over time or throughout an individual's life. There is *more than enough archival science* to address questions about producing fixed digital documents of archival quality, as well as about managing other non-documentary information of significant social and personal value. The current state of wearable data practices implies that responsibility for long-term information preservation should not be delegated exclusively to wearable technology companies. This finding addresses RQ1. Enormous amounts of data are being collected, and a vast population is involved, so archivists can no longer ignore identified needs. Archivists have an institutional or professional responsibility in such cases because information must remain available to users, especially vulnerable groups such as patients, over the long term. This finding addresses RQ2.

Should we archive the original, loosely fixed data in non-documentary form with full functionality, or the secondary, derived, less-functional records that resemble traditional archival records and are stored in static, stable file formats, such as PDFs? As indicated by the SWOT analysis and the discussion of values, it is better to archive data stored in more dynamic data containers, as this enables active reuse of information. Instead of pushing traditional static records, medicinal products or sport-related IT platforms should start adopting data formats that users need. From the traditional perspective on originality, storing data as static documents is also questionable, as the original material-discursive objects were data, not documents. The recommendation, based on the SWOT analysis and discussion presented above, is to archive data in fast-processing formats supported by ontologies. This directs the answer to RQ3.

Not long ago, archival practice shifted its focus from records inseparably fixed to physical media to logical archival objects. Nowadays, archived data remains connected to (predominantly XML) records. However, if the structure of the global infosphere shifts, archival science should take this into account. If the information to be preserved evolves, archival theory and practice should also adapt. It is shown that connecting our profession to wearables' practices and standardising the wearable-archival subdomain are worthwhile and much-needed endeavours. Does this mean the archival profession's attention should shift towards all forms of relevant information? In today's data-intensive world, it's justified because the function is important ("form […] perpetuates the function it serves," Duranti, 1991, 6), and it gives users back control of their data. Data should be fully reusable across various manufacturers' ecosystems, and users shouldn't lose control of their data at any point. Data collecting, processing, and access can be tools of Foucauldian power, as Toner and Acker insightfully wrote, or means of our empowerment. Archivy has always been a double-edged sword for its wielder. Regarding RQ4, yes, the author of the article believes that archival discourse and its material-discursive practices should change to accommodate data as their object and to develop solutions for data preservation at the level of data. Further directions should include adjusting archival terminology to recognise data as its subject, establishing a sound legislative framework and practical professional standards, adhering to the data selection and management guidance outlined in the discussion (continuity, interoperability, data reliability), and promoting good archival practices in the field.

REFERENCES

Acker, A. (2025). *Archiving Machines: From Punch Cards to Platforms*. The MIT Press.

Apple. (2023, May). *Health Privacy Overview*. Retrieved at https://www.apple.com/health/pdf/Health_Privacy_Overview_May_2023.pdf

Barad, K. M. (2007). *Meeting the Universe Halfway: Quantim Physics and the Entanglement of Matter and Meaning*. Duke University Press.

Beer, D. (2019). *The Data Gaze: Capitalism, Power and Perception*. SAGE Publications Ltd. Retrieved at https://doi.org/10.4135/9781526463210

Couldry, N. (2020). Recovering Critique in an Age of Datafication. *New Media & Society*, *22* (7), 1135–1151. Retrieved at https://doi.org/10.1177/1461444820912536

De Arriba-Pérez, F., Caeiro-Rodríguez, M., & Santos-Gago, J. (2016). Collection and Processing of Data from Wrist Wearable Devices in Heterogeneous and Multiple-User Scenarios. *Sensors*, *16*(9), 1538. Retrieved at https://doi.org/10.3390/s16091538

Duranti, L. (2010). Concepts and Principles for the Management of Electronic Records, or Records Management Theory is Archival Diplomatics. Records Management Journal, 20 (1): 78–95. Retrieved at https://doi.org/10.1108/09565691011039852

Duranti, L. (1989). Diplomatics: New Uses for an Old Science, Part I. *Archivaria 28* (January):7-27. Retrieved at https://archivaria.ca/index.php/archivaria/article/view/11567

Duranti, L. (1991). Diplomatics: New Uses for an Old Science, Part V. *Archivaria* 32 (January):6-24. Retrieved at https://archivaria.ca/index.php/archivaria/article/view/11605

Flexible and Interoperable Data Transfer (FIT) Protocol Introduction v 2.4. (2019). Garmin. Retrieved at https://developer.garmin.com/fit/protocol/

Farrugia, C. J. (2024). IT and Archives: Challenges, Threats, and Opportunities. *Atlanti*+ 34 (1): 23-34, Retrieved at https://journal.almamater.si/index.php/atlantiplus/issue/view/45/38

Foucault, M. (2003). *The Birth of the Clinic: An Archaeology of Medical Perception*. Routledge.

Goudarouli, E., & Prescott, A. (Eds.). (2025). *Archival Materialities in a Digital Age*. Oxford University Press.

Grüning, L. S., & Richlan, F. (2026). *"Strava made me do it": Psychological effects of social comparison and (self-)surveillance on a social network for athletes*. Preprint. PsyArXiv. https://doi.org/10.31234/osf.io/4gajv_v1

Hill, K. (2011). Adventures in Self-Surveillance: Fitbit, Tracking My Movement and Sleep. *Forbes*. Retrieved at https://www.forbes.com/sites/kashmirhill/2011/02/25/adventures-in-self-surveillance-fitbit-tracking-my-movement-and-sleep/

Hoskins, A. (2011). Media, Memory, Metaphor: Remembering and the Connective Turn. *Parallax,* 17 (4), 19–31. https://doi.org/10.1080/13534645.2011.605573

Hoskins, A. (2017). The Restless Past: An Introduction to Digital Memory and Media. In A. Hoskins (Ed.), *Digital Memory Studies: Media Past in Transition.* 1st ed. Routledge. https://doi.org/10.4324/9781315637235-1

IEEE Standard for Health Informatics-Device Interoperability Part 10417: Personal Health Device Communication-Device Specialization-Glucose Meter. (No. IEEE Std 11073-10417-2023). (2023).

Jamieson, A., Jones, S., Chaturvedi, N., Hughes, A. D., & Orini, M. (2024). Accuracy of Smartwatches for the Remote Assessment of Exercise Capacity. *Scientific Reports*, *14* (1), 22994. Retrieved at https://doi.org/10.1038/s41598-024-74140-x

Karthick, G. S. (Ed.). (2026). *Sensor Data Analytics for Intelligent Healthcare Delivery* (First edition). Chapman and Hall.

Lemieux, V. L. (2024). Chapter 10. Blockchain: The Theory and Practice of Trustworthy Archival Documents. In C. Rogers and A. Wieland. *Archival Science in Interdisciplinary Theory and Practice.* 1st ed. Rowman & Littlefield.

Lupton, D. (2013). Understanding the Human Machine [Commentary]. *IEEE Technology and Society Magazine*, *32*(4), 25–30. Retrieved at https://doi.org/10.1109/MTS.2013.2286431

Mayer-Schönberger, V., & Cukier, K. (2014). *Big data: A revolution that will transform how we live, work, and think* (First Mariner Books edition). Mariner Books, Houghton Mifflin Harcourt (Kindle ed.).

Mejias, U. A., & Couldry, N. (2024). *Data Grab: The New Colonialism of Big Tech and How to Fight Back*. The University of Chicago Press.

Mitsubayashi, K. (Ed.). (2024). *Wearable Biosensing in Medicine and Healthcare*. Springer Nature Singapore. Retrieved at https://doi.org/10.1007/978-981-99-8122-9

Mordell, D. (2019). Critical Questions for Archives as (Big) Data. *Archivaria*, *87*, 140–161.

Moss, M., Thomas, D., & Gollins, T. (2018). The Reconfiguration of the Archive as Data to be Mined. *Archivaria*, *86*, 118–151. Retrieved at https://archivaria.ca/index.php/archivaria/article/view/1364

Payne, N. (2018). Stirring The Cauldron: Redefining Computational Archival Science (CAS) For The Big Data Domain. *2018 IEEE International Conference on Big Data (Big Data)*, 2743–2752. Retrieved at https://doi.org/10.1109/BigData.2018.8622594

Rajh, A. (2025). Arhiviranje podataka kao nova faza postkustodijalne prakse i potreba redefiniranja terminologije. *Arhivski Vjesnik*, *68*, 15–37. Retrieved at https://doi.org/10.36506/av.68.1

Roze, S., Isitt, J. J., Smith-Palmer, J., Lynch, P., Klinkenbijl, B., Zammit, G., & Benhamou, P.-Y. (2021). Long-Term Cost-Effectiveness the Dexcom G6 Real-Time Continuous Glucose Monitoring System Compared with Self-Monitoring of Blood Glucose in People with Type 1 Diabetes in France. *Diabetes Therapy*, *12*(1), 235–246. Retrieved at https://doi.org/10.1007/s13300-020-00959-y

Semantic Sensor Network Ontology. (2017). https://www.w3.org/TR/vocab-ssn/

Sharon, T. (2017). Self-Tracking for Health and the Quantified Self: Re-Articulating Autonomy, Solidarity, and Authenticity in an Age of Personalized Healthcare. *Philosophy & Technology*, *30* (1), 93–121. Retrieved at https://doi.org/10.1007/s13347-016-0215-5

Srnicek, N. (2016). *Platform Capitalism*. Polity. https://mudancatecnologicaedinamicacapitalista.wordpress.com/wp-content/uploads/2019/02/platform-capitalism.pdf

Statista. (2025, November 14). *Volume of Data or Information Created, Captured, Copied, and Consumed Worldwide from 2010 to 2029*. Statista. Retrieved at https://www.statista.com/statistics/871513/worldwide-data-created/?srsltid=AfmBOoqto2fE6tdyUxrwUpcVLg477lf6VxjCo9aNf7dfo6BWuqcDwCDG

Szeto, K., Arnold, J., & Maher, C. (2024). The Wearable Activity Tracker Checklist for Healthcare (WATCH): A 12-point guide for the implementation of wearable activity trackers in healthcare. *International Journal of Behavioral Nutrition and Physical Activity*, *21* (1), 30. Retrieved at https://doi.org/10.1186/s12966-024-01567-w

Talboom, J. S., & Huentelman, M. J. (2018). Big Data Collision: The Internet of Things, Wearable Devices and Genomics in the Study of Neurological Traits and Disease. *Human Molecular Genetics*, *27*(R1), 35–39. Retrieved at https://doi.org/10.1093/hmg/ddy092

The PROV Ontology. (2013). Retrieved at http://www.w3.org/TR/prov-o/

Time Ontology in OWL. (2022). Retrieved at https://www.w3.org/TR/owl-time/#time:Interval

Toner, J. (2023). *Wearable Technology in Elite Sport: A Critical Examination*. Routledge. Retrieved at https://doi.org/10.4324/9781003184409

Vaghasiya, J. V., Mayorga-Martinez, C. C., & Pumera, M. (2023). Wearable sensors for telehealth based on emerging materials and nanoarchitectonics. *Npj Flexible Electronics*, *7*(1), 26. Retrieved at https://doi.org/10.1038/s41528-023-00261-4

Vijayan, V., Connolly, J. P., Condell, J., McKelvey, N., & Gardiner, P. (2021). Review of Wearable Devices and Data Collection Considerations for Connected Health. *Sensors*, *21* (16), 5589. Retrieved at https://doi.org/10.3390/s21165589

Visser, M. M., Van Muylder, A., Charleer, S., Isitt, J. J., Roze, S., De Block, C., Maes, T., Vanhaverbeke, G., Nobels, F., Keymeulen, B., Mathieu, C., Luyten, J., Gillard, P., & Verhaeghe, N. (2024). Cost–utility analysis of Dexcom G6 real-time continuous glucose monitoring versus FreeStyle Libre 1 intermittently scanned continuous glucose monitoring in adults with type 1 diabetes in Belgium. *Diabetologia*, *67*(4), 650–662. Retrieved at https://doi.org/10.1007/s00125-023-06084-2

World Health Organization. (2024, November 14). *Diabetes*. Retrieved at https://www.who.int/News-Room/Fact-Sheets/Detail/Diabetes

Xue, Z., Gai, Y., Wu, Y., Liu, Z., & Li, Z. (2024). Wearable mechanical and electrochemical sensors for real-time health monitoring. *Communications Materials*, *5* (1), 211. Retrieved at https://doi.org/10.1038/s43246-024-00658-2

Zuboff, S. (2019). *The age of surveillance capitalism: The fight for a human future at the new frontier of power* (First trade paperback ed.). PublicAffairs.