Forensic Science International: Digital Investigation 44 (2023) 301504

Contents lists available at ScienceDirect

Forensic Science International: Digital Investigation

journal homepage: www.elsevier.com/locate/fsidi

DFRWS 2023 EU - Selected papers of the Tenth Annual DFRWS Europe Conference

Interpreting the location data extracted from the Apple Health database

Luke Jennings [a, *], Matthew Sorell [a], Hugo G. Espinosa [b]

[a] The University of Adelaide, Adelaide, SA, 5005, Australia
[b] Griffith University, Nathan, QLD, 4111, Australia

ARTICLE INFO

Article history:

Keywords:
Apple health
Tracking
Locations
Coordinate
Time zone
Digital trace
Interpretation

ABSTRACT

Smart fitness tracking devices are becoming more prevalent in criminal investigations. The Apple Health database for Apple Watch and iPhone fitness activity contains a diverse range of digital traces such as exercise or location information. In this paper, we provide an explanation of the database structure and an SQLite query-based approach for as complete database structure oversight as possible. We extract and link time zones to timestamps and limited geo-coordinate data to workout activity, in the context of case studies using a rich 5+ year long real database of one of the authors. Additionally we investigate the impacts of a major iOS change with iOS 16 on these insights to deepen the understanding of the data at hand. We found that the accuracy of geolocation coordinates, time zone entries, and workout information may not necessarily be accurate or consistent and requires contextual interpretation. We demonstrate the forensic value of the database in the context of informing an investigation by establishing context to a victim's or suspect's activities.

1. Introduction

The appearance of smart devices, in particular fitness tracking smart watches, is becoming more prevalent in investigations (Opie, 2019; Carter, 2021; D'Mello, 2019; Clover, 2021) and the need to understand and document their forensic potential and limitations is important for practitioners and researchers. Each year, the number of smart fitness watches getting on the wrists of customers increases, with Apple Watch being the current leading brand with an estimated 36% of the global market in Q1 2022 (Counterpoint, 2022).

The underlying application for Apple Watch and iPhone is Apple Health. Apple Health stores a range of personal health, activity, and biometric data in the **healthdb_secure.sqlite** database (Apple, 2022; Edwards, 2016). This data includes, but is not limited to:

- Step counts, stairs climbed, heart rate and energy burned;
- Workout routes and time zones;
- Personal information such as name, date of birth, weight, and height.

This paper demonstrates how workout geolocation information, time zones and other data can be used to interpret and track a person's location and activity. Up until iOS 15 the only geolocation information recorded in the database were for weather conditions which our review suggests is the start location of a workout. iOS 16 makes significant changes to the geolocation information stored in the database which are discussed in this paper.

Apple Health is an unexpected source of location data for two reasons. Firstly, workout activity may be mapped, including both static locations and journeys. Secondly, health-related logs keep track of time zones, providing an unexpected additional layer of macro-scale geolocation data which can be exploited to establish when the phone (and by implication, the owner) is travelling. Travel data is particularly valuable because network cell location logs may be limited or unavailable during periods of international roaming or SIM card substitution.

The dataset used in this research is a real 5+ year-long rich dataset of one of the authors and can be made available upon request to the author subject to usage policy". Due to the nature of the topic being location tracking and activity interpretation of a person, it is important that the dataset being analysed is real with all the anomalies and difficulties that come with it. In reality, the datasets that would appear in real investigations come with similar limitations, so the focus of this paper is not to provide one clear

* Corresponding author.
E-mail address: luke.jennings@adelaide.edu.au (L. Jennings).

https://doi.org/10.1016/j.fsidi.2023.301504

answer or quantitative accuracies, but to explore the potential of this database.

This dataset was also utilised as the basis for the DFRWS APAC 2022 forensic rodeo CTF (https://forensicrodeo.ctfd.io/, 2022). While tools exist (iLEAPP, 2022) to extract tables from **healthdb_secure.sqlite**, using such a tool itself requires validation to ensure that all relevant records are extracted. Feedback from participants in the rodeo suggested that the size and complexity of this specific database can be overwhelming without guidance and documentation. The CTF itself acts as a scaffolded learning exercise in navigation and exploration of our specific sample database. In this paper, we consider a small number of specific tables relating only to geolocation records and one set of CTF challenge questions, as examples of how to navigate, extract, isolate and analyse this rich source of location information, with reference to known ground truth. These select tables are sufficient enough to answer a lot of relevant questions described in the case studies explored.

1.1. Contributions

This paper details an SQLite query-based approach for extracting geolocation data from the Apple Health database **healthdb_secure.sqlite**, noting also significant changes to the database structure over time, and particularly with the introduction of iOS16. Using real health data accumulated by one of the authors over five years, alongside verifiable ground truth, we provide an example of real-world scenarios to illustrate the forensic value of geolocation data stored in the Apple Health database. Hence, we demonstrate both the potential of this data, as well as some discussion of its limitations.

This is one paper in a series in which we shall be addressing a diverse range of data series in the health data, and how we are validating that data against long term longitudinal ground truth. The overall detailed structure of the database is out of scope for this paper, and we are only concerned with the specific tables that address geolocation, using both workouts and timezones, together with associated selected case studies.

1.2. Paper outline

Section 2 provides the current position of the validation of health data from fitness trackers, previous work involving the analysis of location information derived from other smart devices or applications. Section 3 describes the process for accessing the **healthdb_secure.sqlite** database, the structure of the database and how to find and export the different location data the database can provide. Section 4 will present two examples of real events to explore the health data in an investigative mode. These examples will then be approached using the methods described in Section 3, and their accuracy analysed in comparison to the known established ground truth. Lastly, Section 5 summarises the findings and interpretations presented in Section 4 and establishes the limitations and possibilities for future work.

2. Related work

Before the Apple Health database can be relied on in an investigation, the accuracy of its devices' (smart watches and iPhones) sensors that record a user's health information must be validated. This section addresses the previous research involving the validation of the biometric data and sensors in Apple devices, location data from other sources within the ecosystem and without, and location data of Apple's competing brands.

2.1. Validation of biometrics

The iPhone health app has been previously researched in the context of digital forensics (Peter van Zandwijk and Boztas, 2019), where the authors investigated the use of steps and distances registered during walking or running as digital evidence. The authors found that the step counts recorded were within an acceptable average error range of about 2% but the distances recorded had much higher deviation up to 30–40%. Thus, the authors believed that digital traces from the Apple Health app can be used for evidential purposes.

The sensors and pre-processed step counts and heart rate measurements from Apple Watch devices have also been validated (Espinosa et al., 2020). The authors found that "the inertial and optical sensors from the Apple Watch counted steps and measured heart rate with a minimum error and performed as expected".

2.2. Other sources of geolocation data

In addition to health information, mobile devices have been known to record geolocation metadata in different ways. Examples include, EXIF metadata in photographs (Ramesh Kumar et al., 2016; Albrecht and McIntyre, 2014), map location applications (Chuang et al., 2016; Moore and BaggiliFrank, 2017; Maus et al., 2011) and Wi-Fi access point MAC addresses (Cunche, 2014). Off-device, location data can also be retrieved from Google Maps (Rodriguez et al., 2018) and network cell location logs (Trogh et al., 2019).

2.3. Validation of geolocation data

The accuracy of smartphone GPS data has been researched previously for iPhone (Merry and Bettinger, 2019), and Android (Spreitzenbarth et al., 2012). In (Merry and Bettinger, 2019), the authors found that the "Overall average horizontal position accuracy of the iPhone 6 (7–13 m) is consistent with the general accuracy levels observed of recreation-grade GPS receivers in potential high multi-path environments." The authors from (Spreitzenbarth et al., 2012) found that "location data stored on the phones is often much more precise than the rather coarse-grained data stored by network operators. However, the availability of location data on smartphones varies considerably compared with the data stored by network operators." These findings suggest that the accuracy of the GPS data is relatively good but has some obvious limitations and thus its use as evidence is potentially limited unless supplemented by other sources such as those listed in 2.2 above.

In section 4 the limitations of the GPS data stored within the specific Apple Health application are demonstrated. This data is supplemented by interpretations from other tables within the dataset, OSINT (Open-Source Intelligence) and ground-truth.

3. Methodology

3.1. Background

The Apple health database healthdb_secure.sqlite is an SQLite database that incorporates not just tables of data, but also describes the overall structure through what is known as the Schema. The Schema describes which tables are interconnected through which indices and may also include pre-programmed queries as Views and Triggers. The detailed structure of an SQLite file is outside the scope of this paper but for the interested reader, is discussed in (A visual explanation of sqlite, 2022). The specific understanding of the tables and codes in this database are a consequence of previous work (Edwards, 2016; Peter van Zandwijk and Boztas, 2019), with both systematic and ad-hoc experimental validation of the

database by the authors over time. Our analysis does not consider the entire database structure and is limited only to those tables necessary for the geolocation case studies presented here.

3.2. Extraction

The Apple Health dataset under consideration, was captured between 2017 and 2022. The source author is a frequent international traveler. The dataset is available on request from the authors, subject to usage policy.

Health information can be extracted in multiple ways from the iPhone. The most direct method is a direct export from the application interface (Afonin, 2018). This process supplies the user with export.xml files detailing the user's various health and activity information and a workout route folder consisting of.GPX files containing GPS coordinate location information for the users' workouts. These files can be imported into software, such as Google Earth (Google, 2022), to view the full workout route. However, a key limitation of the direct export is that Apple Health takes a shortcut, exporting all data in the local time zone at the place and date of export. This means that data recorded, for example, in winter, will be exported during summer in daylight savings time; and data recorded in Europe but exported in Singapore will be exported in the Singapore time zone (UTC+8). The export time zone is included in the exported times, but the source time zone is not. Another limitation is that each workout and hence each route of geolocation data is in separate GPX files. This means that if all locations were to be analysed to determine patterns and common locations, the analysis can become bogged down. In the database, these geolocations are listed in a single table instead.

Apple Health information can also be retrieved from iCloud (Afonin, 2018) with the correct tools or using a dedicated forensic tool such as MSAB's XRY (MSAB, 2022). Finally, the raw SQLite database can also be extracted from the encrypted backup of an iPhone (Reincubate Ltd, 2022), which is the method utilised in this study.

The SQLite database, **healthdb_secure.sqlite**, contains numerous tables which contain health, activity, workout and, we demonstrate, geo-location information as well. Data within SQLite tables can be abstracted away from related fields and must be re-joined through SQLite queries, using the join command to link tables that have a shared ID field (A visual explanation of sqlite, 2022). For this study, we work as directly as possible with the database using an SQLite database browser (Piacentini, 2022) and SQL commands, similar to tools such as those provided in the iLEAPP suite (iLEAPP, 2022), recognizing that browsers have their own limitations. By working directly with the database, we maintain as complete oversight of the database structure and schema as possible, rather than relying on a third party's analysis, which we would then need to validate separately.

3.3. Time zone analysis

In the Apple health database, the data_provenances table contains time zone information under the field *tz_name*. This table is connected to the objects table which contains a *creation_date* timestamp, and through objects is connected to the samples table which contains a *start_date* and *end_date* timestamp. These timestamps are stored as Apple cocoa core UTC. This structure is demonstrated in Fig. 1 below.

When there is more than one source device such as an iPhone and an Apple Watch, the provenance is identified separately and the time zone can drift, particularly if the devices are in flight mode. Hence, we only consider a *data_type* which is common to both devices. Since distance travelled is derivative of step count, we limit our analysis to step count only, which allows us to consider time zones according to different devices independently. In the database, step count is identified as a *data_type* '7', identified in (Edwards, 2016) along with other *data_types*. It should be noted that other *data_types* can be used in the same way, such as heart rate, but this would only apply to one type of device.

In establishing the timestamp interval during which a time zone applies, we identify the earliest and latest step count record for either the Watch or the iPhone. The time zone transition occurs between records, but we make the reasonable observation that the interval will be shorter, and therefore provide higher accuracy, if the iPhone and Watch are being carried by a physically active user. The time zone is expressed in the form of Country or Region/City, i.e., Australia/Adelaide or Europe/Berlin.

To express the timestamp in local time, it is still necessary to translate this information into a local UTC offset, considering daylight savings if applicable. Fig. 2 below demonstrates the SQLite query used to extract timestamped time zone information with step counts, as reflected in the table structure highlighted in Fig. 1 above.

3.4. Workout location analysis

Case study 1 in section 4.1 utilises coordinate data associated with workout information stored in the Apple health database. The coordinate data stored in the database is detailed as metadata keys called _HKPrivateWorkoutWeatherLocationCoordinatesLongitude and _HKPrivateWorkoutWeatherLocationCoordinatesLatitude which are found in the metadata_values and metadata_keys tables. Our review suggests that these weather locations are the start location of a workout. Up to and including iOS 15 this was the only source of coordinate data found within the database. However, with iOS 16 came significant database restructures and new tables with other coordinate information as well as the already existing weather coordinate location in metadata_values.

3.4.1. Pre-iOS 16

As mentioned above, up to and including iOS 15 the only source of coordinate information in the database were weather locations from metadata_values. Fig. 3 below illustrates how the metadata_values table is connected to the metadata_keys and samples tables, and hence to workouts and workout_events.

Fig. 4 below demonstrates the SQLite query used to extract timestamped longitude and latitude information with the workout data from the **workouts** and **workout_events** tables described in Fig. 3 above.

3.4.2. Post-iOS 16

The process described in 3.4.1 above still works for iOS 16, but the database structure has now changed and is illustrated in Fig. 5 below and the query is demonstrated in Fig. 6 below.

With iOS 16 there is now a new table called **location_series_data** which contains the full workout route information that was missing in iOS 15 and below. However, there are still some limitations preventing us from using that table for our case studies in the next section. The **location_series_data** table contains a *timestamp* field and a *series_identifier* field, meaning that we can link a sequence of times and places directly from this table. However, the *series_identifier* index does not align with the index's for any of the **workout** or **samples** tables, and requires manual alignment which can't be done with a query. We have a robust start location with the weather coordinates found in **metadata_values**, but even though the full workout routes are now in the database, they have not been fully integrated with the other tables in the database. It is important to acknowledge and document these

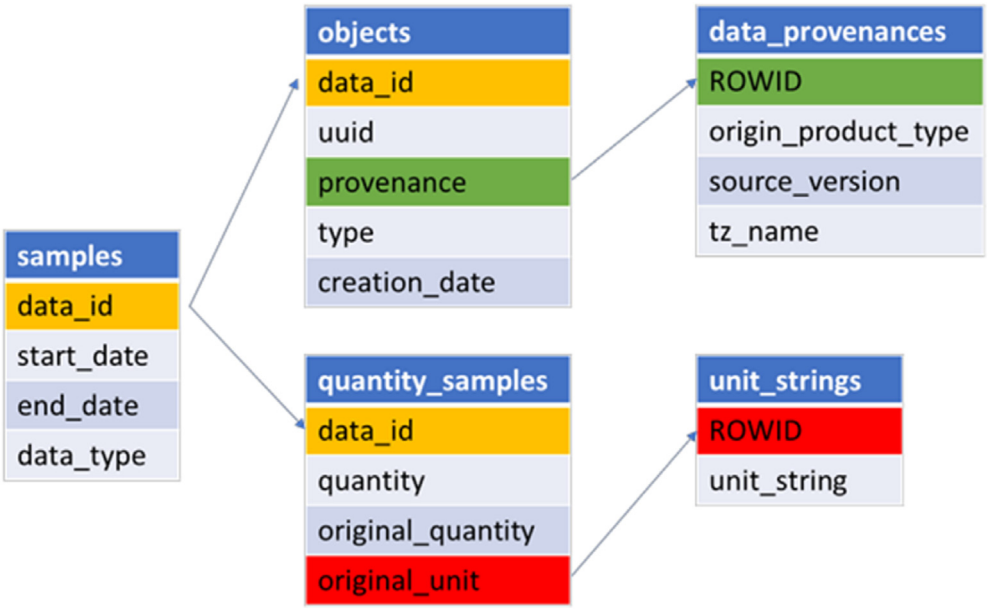

Fig. 1. Time zone structure.

```
1   select samples.data_id ,
2   datetime(samples.start_date+978307200,'unixepoch') as "start date",
3   datetime(samples.end_date+978307200,'unixepoch') as "end date",
4   samples.data_type,
5   quantity_samples.quantity,
6   data_provenances.origin_product_type,
7   data_provenances.source_version,
8   data_provenances.tz_name
9   from samples
10  left outer join quantity_samples on samples.data_id=quantity_samples.data_id
11  left outer join objects on samples.data_id=objects.data_id
12  left outer join data_provenances on objects.provenance=data_provenances.ROWID
13  where data_type = 7
```

Fig. 2. Query – Time zones and samples.

changes, as it is possible for practitioners to come across an older dataset which has not been updated to the newest iOS version. It may also be possible that the indexes may change in a future iOS release allowing for the connection between workout and workout route. As with *data_types* in section 3.4.1, workouts have their own identifier called *activity_type* which is found in the **workouts** table in iOS 15 and the **workout_activities** table in iOS 16. These integers can be identified from (Apple, 2022), and that for this sample dataset only five types of workouts are recorded, including type 35 which is identified as “Rowing”.

4. Case studies

4.1. Case study 1 - Workout coordinate data

Disclaimer: The raw coordinate data as well as the dataset's owner's name cannot be shared here on account of privacy. The owners home address is omitted.

This section implements the methodology and queries from section 3 into a benign but real scenario to demonstrate the process and interpretations that can arise out of investigating the Apple Health database.

4.1.1. The narrative

On the 15th of October 2020 in Adelaide Australia, the author claimed to be at St Peter's Cathedral in the afternoon for approximately an hour.

Is there any data in the Apple Health database that can verify this claim? If so, what were they doing there and why?

4.1.2. The approach

The database was imaged prior to iOS 16's release, so we will apply the query from Fig. 4. Fig. 7 highlights the output table consisting of workouts from as early as 2018-01-07 up to 2022-01-17 with 1552 database entries.

The resulting table is still quite large to go through each entry, so we can extend the WHERE statement in the query to include the specific start date of 2020-10-15, giving the result in Fig. 8.

The result is one workout with a pair of location coordinates. Note the date is in UTC and not local Adelaide time.

We can take those coordinates and enter them into a mapping software. For this paper we use Google Maps (Google, 2022).

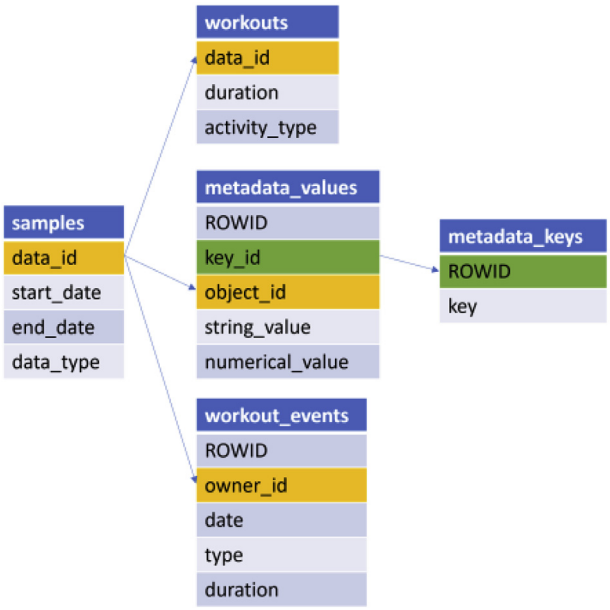

Fig. 3. Workout location structure Pre-iOS 16.

We can see in Fig. 9 that the geolocation coordinates logged in Apple Health for a workout coincide with the claim that they visited St Peter's Cathedral.

Besides confirming the claim, the database tells other information as well. Firstly, the workout was a type 35 which is a "Rowing" type workout, and secondly the total energy burned was significantly higher than others with a value of 397.3 kJ (units are probably kJ but the units are not specified in the database). Now if you're unfamiliar with exercise, churches and cathedrals don't typically have canoes or rowing machines in them. This creates a question then of “If the person wasn't actually Rowing as the health data suggests, what were they actually doing?".

Going back to the query, we can now filter the workouts by activity type 35 ″Rowing" to see all the person's rowing workouts. The newly exported rowing table contains 87 unique entries, of which 78 are at the person's home address. Excluding the home address, we now have a working set of 9 rowing workouts, listed in Table 1 below, which is much easier to work with. It can be inferred that due to the vast amount of rowing workouts at their home, that they own a personal rowing exercise machine.

Uploading this csv file into Google Maps, we get the following maps (Figs. 10–12):

We have two rowing workouts in southern England, one in Norway and six in Adelaide Australia.

First, we will consider the Adelaide Rowing workouts, shown in Table 2 below.

Data_id 2294753, 1/3/2020, shown in Fig. 13 below, the coordinate places the person in the middle of the street. Considering the accuracies expected from previous work outlined in section 2, we can see that the St Cuthberts Anglican Church is close.

Data_id's 3067817, 17/4/2021 and 3186896, 26/6/2021, shown in Fig. 14 above place the persons coordinates at St Andrew's Anglican Church, with id 3186896 being slightly inaccurate and across the road.

Now we are starting to see a pattern arise out of the persons workout data - they do "rowing" at multiple churches in Adelaide. Using OSINT, we identify that the person rings church bells as a hobby. This explains exactly what the person was doing at these churches that caused their Apple Watch to record a "Rowing" workout, as Bell Ringing has a similar motion to rowing, but vertically instead of horizontally.

However, the remaining locations are not so accurate or clear as to where the person actually was. Consider the final remaining Adelaide entry of id 1207724 shown in Fig. 15 above. Unlike the other five Adelaide workouts there doesn't seem to be a church or cathedral close enough to suggest the person was bell ringing.

The final three locations we consider, the two UK entries and the Norway entry show similar vagueness and inaccuracy as the previous Adelaide entry. For id 1050973 shown in Fig. 16 above, the nearest church is St Mary's Church, Twyford and for id 1058300 shown in Fig. 17 below the nearest church is St Johns, New Alresford. Utilising OSINT as we did previously, we can confirm that they also rang bells at these churches on these dates, but unlike the Adelaide churches the GPS coordinates are significantly further away. If those earlier Adelaide churches had not been identified and the person related to bell ringing, these GPS locations may have had not much meaning to an investigator.

Finally, the Norway entry with id 2065589 shown in Fig. 18 below places them at the Oslo Airport. As it stands, there is nothing at the airport that suggests bell ringing and thus the interpretation of this particular entry, limited.

```
1   select workouts.data_id,
2   datetime(samples.start_date+978307200,'unixepoch') as "start date",
3   datetime(samples.end_date+978307200,'unixepoch') as "end date",
4   workouts.duration,
5   workouts.activity_type,
6   workouts.total_basal_energy_burned,
7   workouts.total_energy_burned,
8   workouts.total_distance,
9   metadata_keys.key,
10  metadata_values.numerical_value
11  from workouts
12  left outer join workout_events on workouts.data_id=workout_events.owner_id
13  left outer join metadata_values on workouts.data_id=metadata_values.object_id
14  left outer join metadata_keys on metadata_values.key_id=metadata_keys.ROWID
15  left outer join samples on workouts.data_id=samples.data_id
16  where metadata_keys.key like '%coordinate%'
```

Fig. 4. Query – workouts Pre-iOS 16.

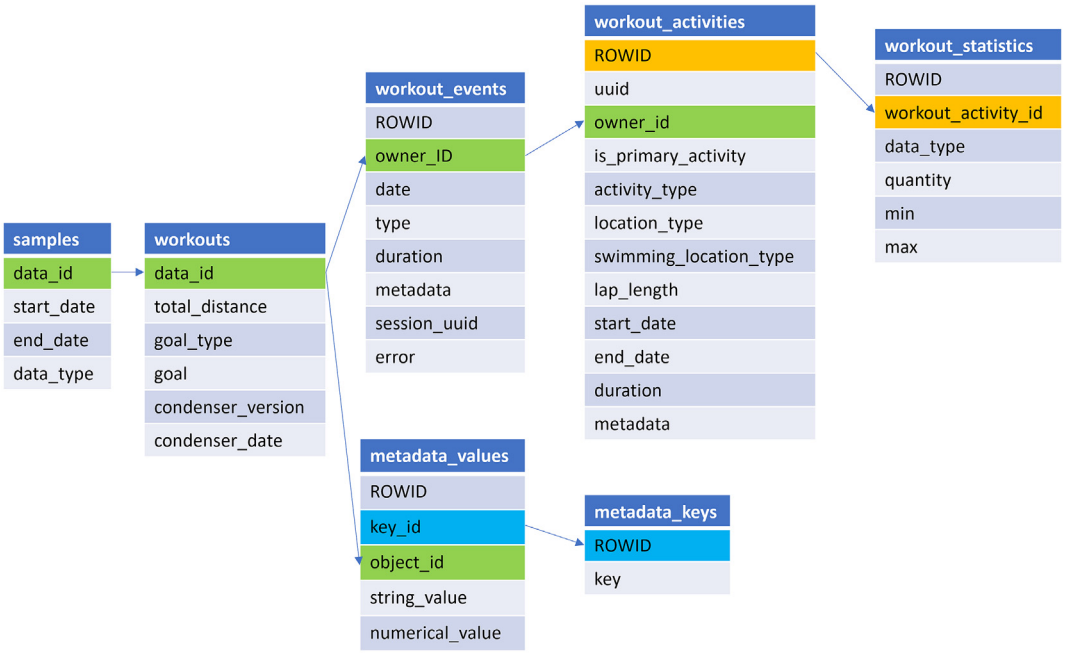

Fig. 5. Workout location structure Post-iOS 16.

```
SELECT workouts.data_id,
datetime(workout_activities.start_date+978307200,'unixepoch') as "start date",
datetime(workout_activities.end_date+978307200,'unixepoch') as "end date",
workout_activities.duration,
workout_activities.activity_type,
workout_statistics.data_type,
workout_statistics.quantity,
metadata_keys.key,
metadata_values.numerical_value
from workouts
left outer join workout_events on workouts.data_id=workout_events.owner_id
left outer join metadata_values on workouts.data_id=metadata_values.object_id
left outer join metadata_keys on metadata_values.key_id=metadata_keys.ROWID
left outer join samples on workouts.data_id=samples.data_id
left outer join workout_activities on workouts.data_id=workout_activities.owner_id
left outer join workout_statistics on workout_activities.ROWID=workout_statistics.workout_activity_id
where metadata_keys.key like '%coordinate%'
```

Fig. 6. Query – workouts Post-iOS 16.

data_id	start date	end date	duration	activity_type	total_basal_energy_burned	total_energy_burned	key
97254	7/1/2018 3:39	7/1/2018 6:18	9552.913055	52	284.222	423.869	_HKPrivateWorkoutWeatherLocationCoordinatesLongitude
97254	7/1/2018 3:39	7/1/2018 6:18	9552.913055	52	284.222	423.869	_HKPrivateWorkoutWeatherLocationCoordinatesLatitude
515802	22/9/2018 5:19	22/9/2018 6:43	4848.719144	52	144.4481135	299.8723186	_HKPrivateWorkoutWeatherLocationCoordinatesLatitude
515802	22/9/2018 5:19	22/9/2018 6:43	4848.719144	52	144.4481135	299.8723186	_HKPrivateWorkoutWeatherLocationCoordinatesLongitude
527363	29/9/2018 9:49	29/9/2018 9:54	301.979615	35	8.89446088	30.74622462	_HKPrivateWorkoutWeatherLocationCoordinatesLongitude
527363	29/9/2018 9:49	29/9/2018 9:54	301.979615	35	8.89446088	30.74622462	_HKPrivateWorkoutWeatherLocationCoordinatesLatitude

Fig. 7. Query table output.

data_id	start date	end date	duration	activity_type	total_basal_energy_burned	total_energy_burned	key	numerical_va
2763540	25/10/2020 6:21	25/10/2020 7:20	3557.37599	35	105.6338724	397.322308	_HKPrivateWorkoutWeatherLocationCoordinatesLongitude	138.5982
2763540	25/10/2020 6:21	25/10/2020 7:20	3557.37599	35	105.6338724	397.322308	_HKPrivateWorkoutWeatherLocationCoordinatesLatitude	-34.91296

Fig. 8. Query table output for 2020-10-15.

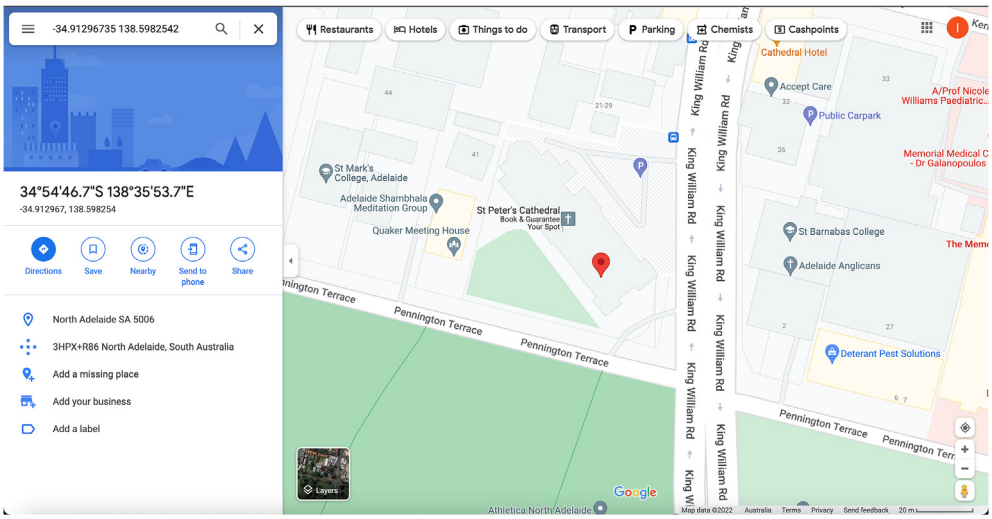

Fig. 9. Workout Location claim – St Peter's Cathedral.

Table 1
Workout rowing locations.

data_id	Latitude	Longitude	start date	end date	duration	total_basal_ energy_burned	total_energy_ burned
1050973	51.01702731	−1.316199791	10/3/2019 15:04	10/3/2019 15:13	561.750406	16.64791015	53.14128571
1058300	51.08966494	−1.161684508	10/3/2019 17:01	10/3/2019 18:00	3535.026181	105.5758413	376.405094
1207724	−34.91632864	138.6210555	6/4/2019 23:43	7/4/2019 1:29	2165.546736	64.25992857	228.6037254
2065589	60.19669991	11.10275407	13/11/2019 18:37	13/11/2019 18:43	361.1245749	10.60905317	19.85767049
2294753	−34.89228102	138.5944175	29/2/2020 22:36	1/3/2020 0:27	6625.569902	196.6999237	722.0203738
2763540	−34.91296735	138.5982542	25/10/2020 6:21	25/10/2020 7:20	3557.375986	105.6338724	397.322308
3067817	−34.89189430	138.6161569	17/4/2021 3:30	17/4/2021 4:22	3091.881259	91.68703598	312.3386591
3186896	−34.89163634	138.6164316	26/6/2021 2:30	26/6/2021 3:20	3011.692324	89.39493935	322.8654958
3206492	−34.91315428	138.5983448	4/7/2021 7:26	4/7/2021 9:04	5871.812864	174.2975827	702.395

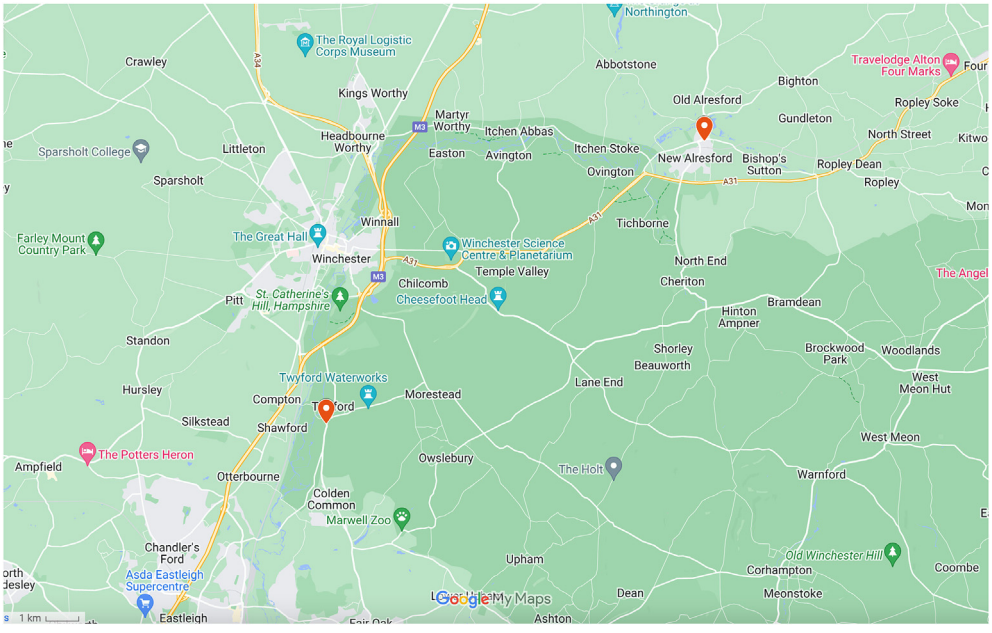

Fig. 10. UK rowing locations.

4.1.3. Results summary

From this case study, we have been able to identify where a person has been, when they went there and what they were doing to some extent using primarily their health data and not taking its results at face value. The health data suggested that this person was doing rowing workouts, when in reality they were doing bell ringing at cathedrals and churches that the Apple Watch had registered as a rowing workout. Using a supplementary source such

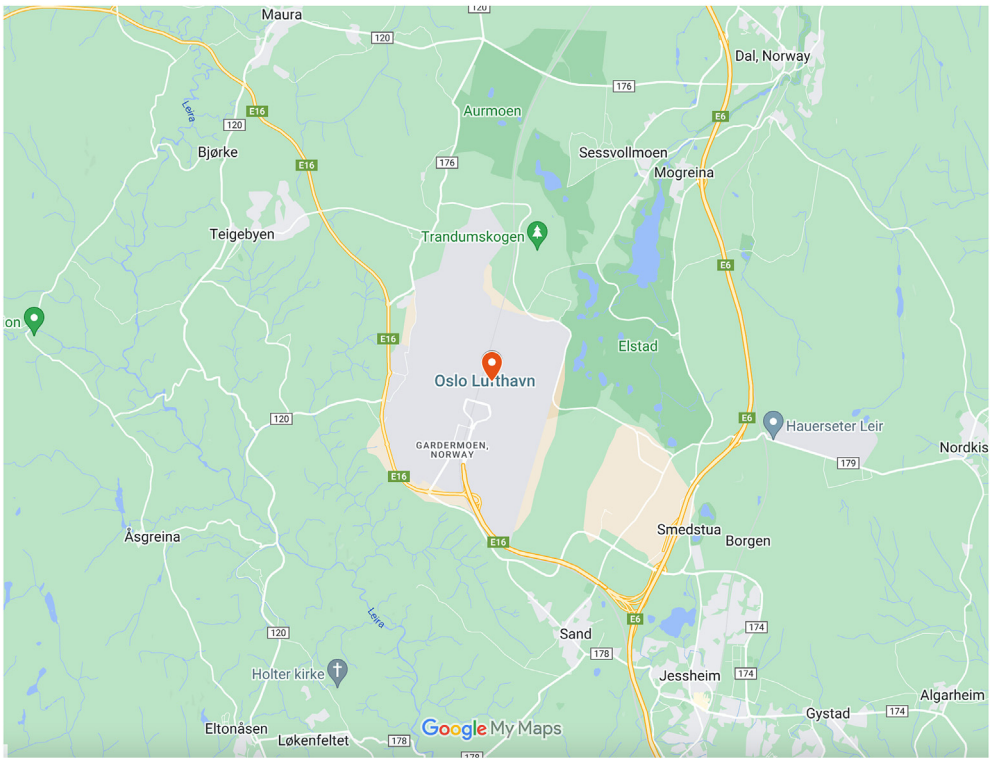

Fig. 11. Norway rowing location.

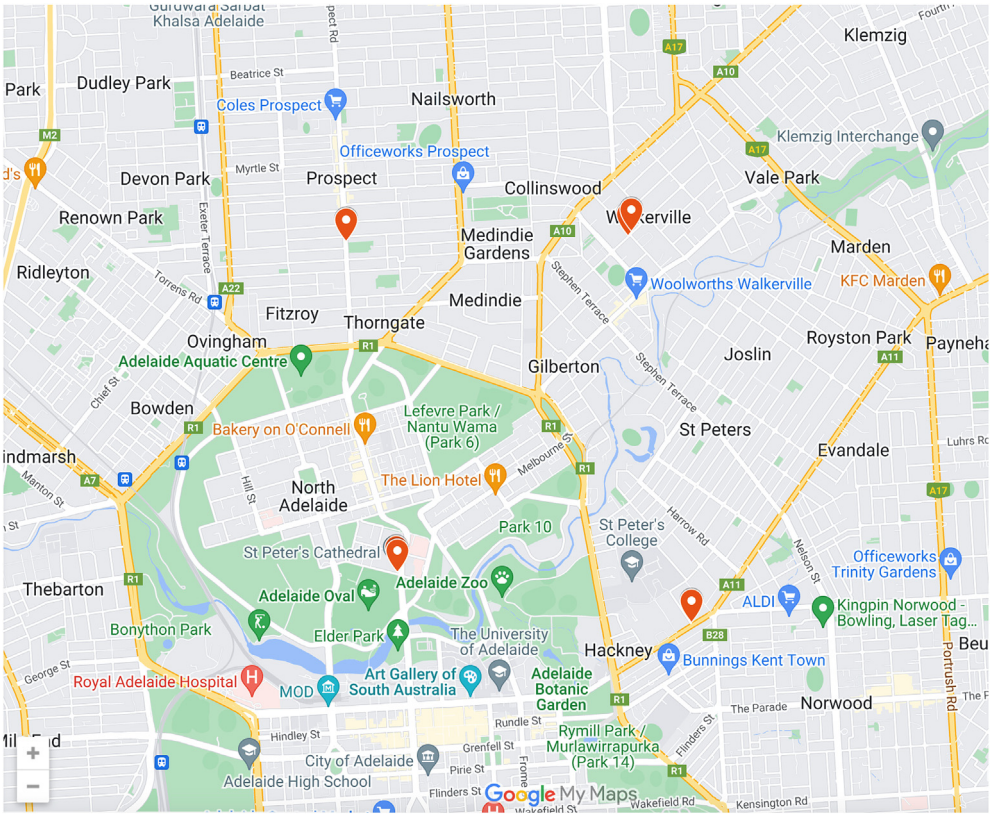

Fig. 12. Adelaide rowing locations.

Table 2
Adelaide rowing locations.

data_id	Latitude	Longitude	start date	end date	duration	total_basal_ energy_burned	total_energy_ burned	Location
1207724	−34.91632864	138.6210555	6/4/2019 23:43	7/4/2019 1:29	2165.546736	64.25992857	228.6037254	?
2294753	−34.89228102	138.5944175	29/2/2020 22:36	1/3/2020 0:27	6625.569902	196.6999237	722.0203738	St Cuthberts
2763540	−34.91296735	138.5982542	25/10/2020 6:21	25/10/2020 7:20	3557.375986	105.6338724	397.322308	St Peters
3067817	−34.89189430	138.6161569	17/4/2021 3:30	17/4/2021 4:22	3091.881259	91.68703598	312.3386591	St Andrews
3186896	−34.89163634	138.6164316	26/6/2021 2:30	26/6/2021 3:20	3011.692324	89.39493935	322.8654958	St Andrews
3206492	−34.91315428	138.5983448	4/7/2021 7:26	4/7/2021 9:04	5871.812864	174.2975827	702.395	St Peters

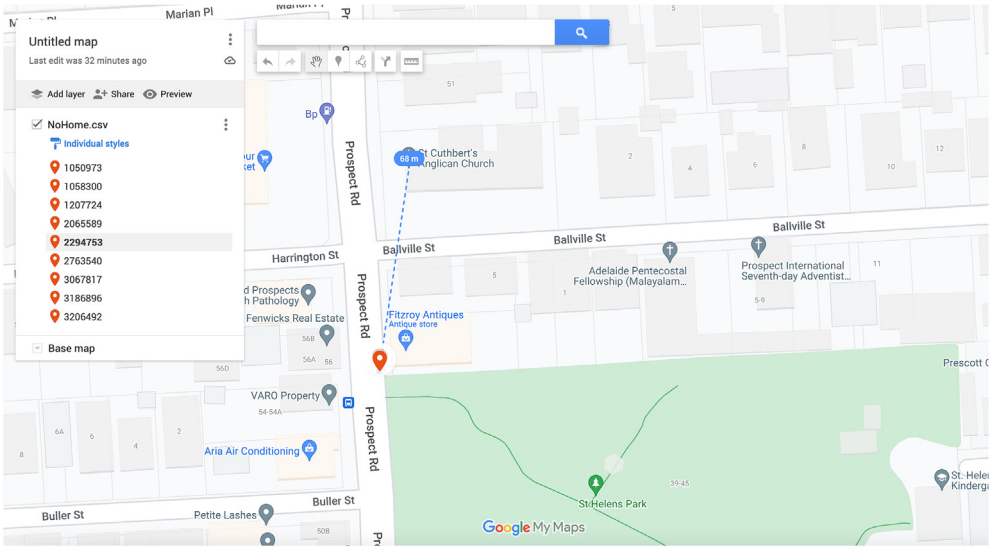

Fig. 13. Adelaide rowing location – St Cuthberts.

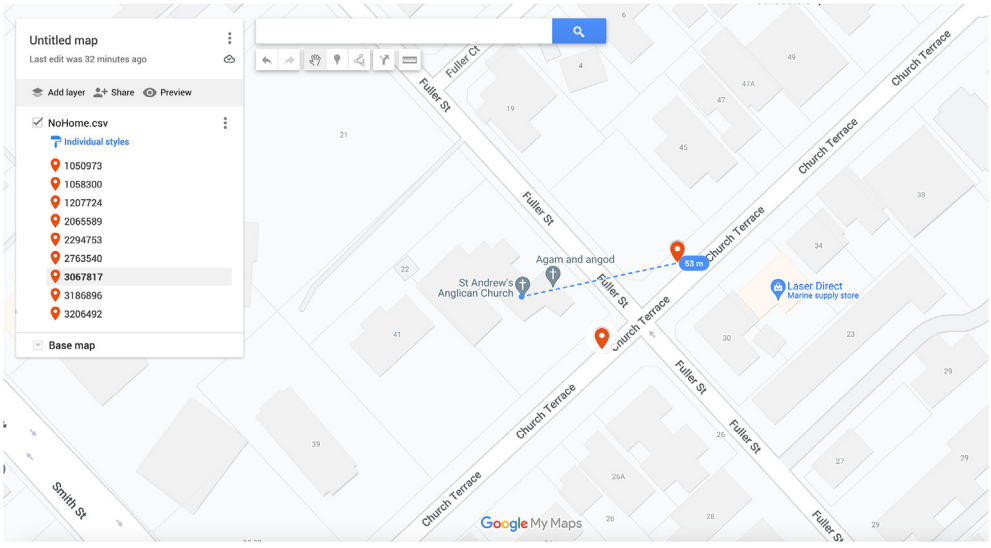

Fig. 14. Adelaide rowing location – St Andrews.

as OSINT, we were also able to confirm seven of the nine rowing workout activities as bell ringing.

4.1.4. Limitations

There are a few limitations with using this health data. The first limitation is the accuracy of the GPS coordinates logged in the database. The GPS coordinates are found to vary quite significantly, with the St Peters coordinate being on top of the church, the St Cuthberts coordinate being some meters away, and the St Mary's church in Twyford being approximately 650+ meters away (according to Google Maps), demonstrated in Fig. 16. The second limitation is that it can be quite easy to make assumptions. Apple Health recorded all these forensic artefacts as rowing workouts, which at face value most people would accept and not investigate any further. Just because the data says that they were rowing, walking, running etc. does not mean that the person was actually

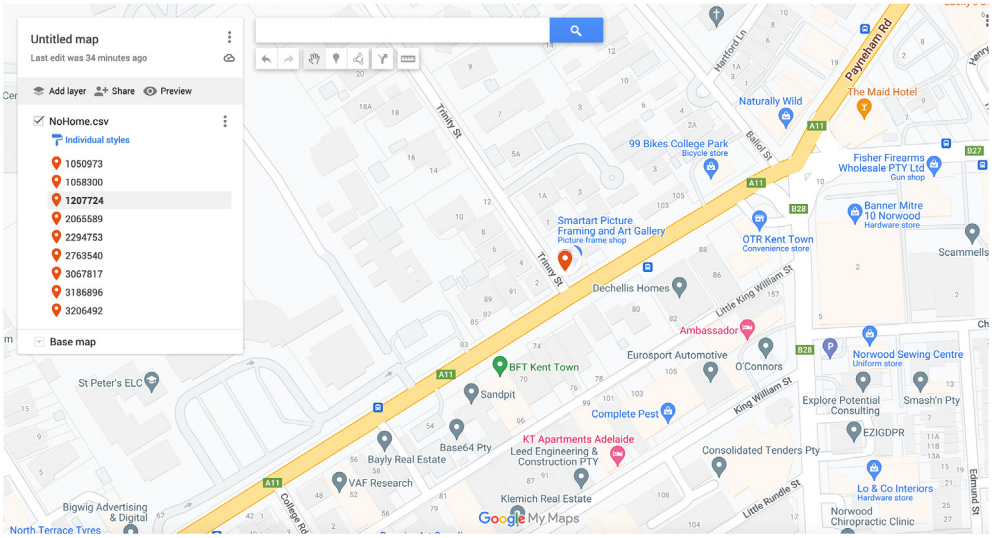

Fig. 15. Adelaide Rowing Location — Unknown location.

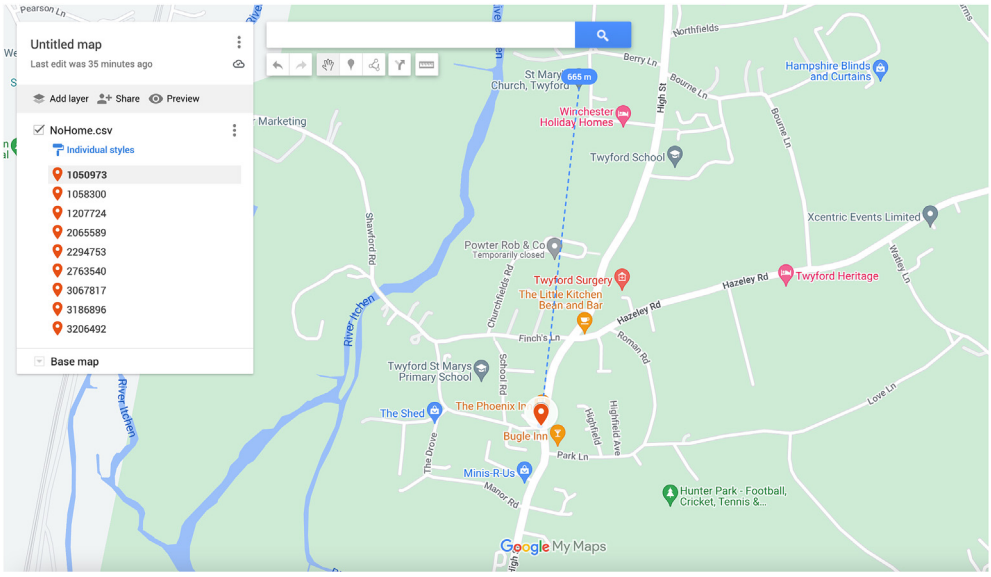

Fig. 16. UK Rowing Location — St Marys offset.

rowing, walking or running. The heath data requires some other supplementary source of information such as a witness statement, OSINT, CCTV, etc. to be able to inform the investigator as to what they might actually be seeing in the health data. The final limitation is that these coordinates are only logged for workouts in the database. From the OSINT source of the Adelaide Bell Ringers, the person of interest also rings bells at several other churches in Adelaide that did not appear as rowing workouts in the database. The following case study will explore what happens when there are no workout coordinates available when trying to track a person's location.

4.2. Case study 2 - time zones

4.2.1. The narrative

On the 19th of May 2019, the author claimed to have visited Frankfurt, Germany. Is there any data in the Apple Health database that can verify this claim?

4.2.2. The approach

Initially we apply the query from Fig. 4 to the dataset with an updated WHERE clause to view any geolocation coordinates and workouts during the period of May 2019. The resulting table only produces three results, which are walking workouts done in Adelaide Australia near their home address.

Hence, we apply the query from Fig. 2 to the dataset to view the time zones associated with step counts (data type = 7). This resulting table is now 4460 entries, and two distinct times can be identified that the person left Australia during May 2019. The first being on the 7/5/2019 from Adelaide and the second being 20/5/2019 from Sydney. Due to the size of the dataset, we display an abridged version as Table 3 below, considering only the initial travel on 7/5/2019 and showing only when the time zone changes for this paper.

According to the health database we can see what resembles the persons travel itinerary. They left Adelaide Australia on 7-5-2019 and arrived in Berlin Germany on 7-5-2019 approximately 10 h

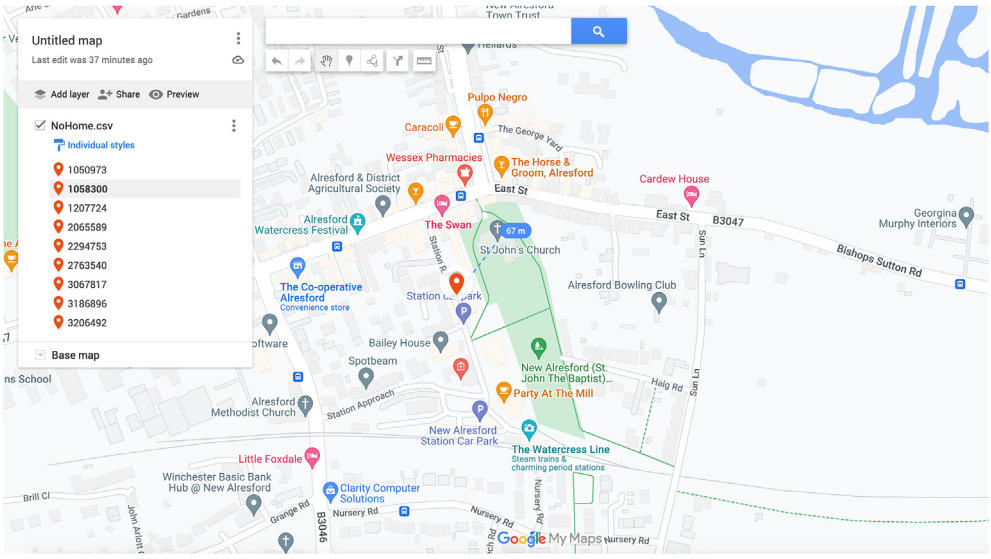

Fig. 17. UK rowing location – St Johns, new Alresford.

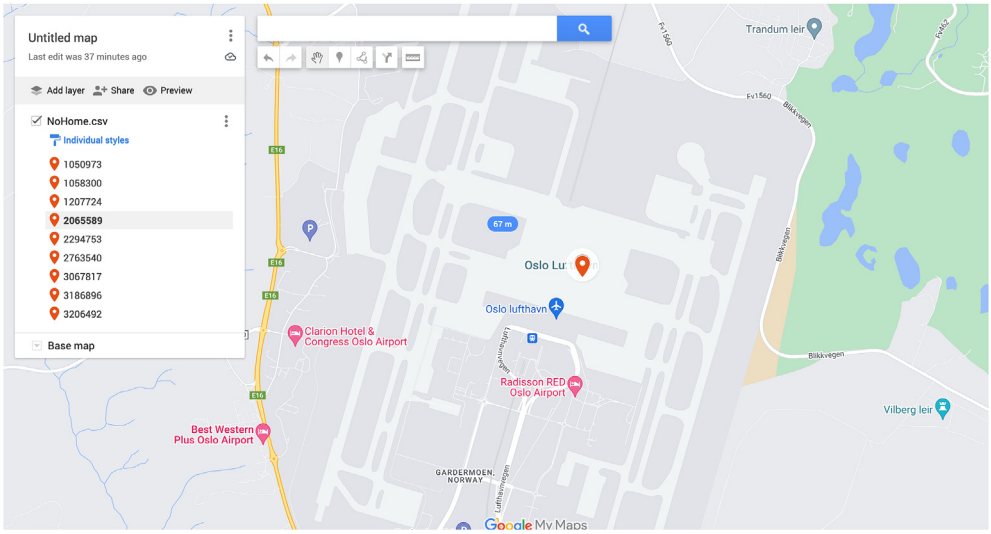

Fig. 18. Norway rowing location – Oslo airport.

Table 3
May 2019 travel itinerary.

data_id	start date	end date	data_type	quantity	origin_product_type	source_version	tz_name
1277615	7/5/2019 1:07	7/5/2019 1:09	7	55	Watch3,2	5.1.3	Australia/Adelaide
1277639	7/5/2019 1:32	7/5/2019 1:32	7	16	Watch3,2	5.1.3	Asia/Dubai
1278314	7/5/2019 10:09	7/5/2019 10:09	7	16	Watch3,2	5.1.3	Asia/Dubai
1278320	7/5/2019 10:44	7/5/2019 10:53	7	19	iPhone9,4	12.2	Europe/Berlin
1285147	11/5/2019 19:58	11/5/2019 19:58	7	4	Watch3,2	5.1.3	Europe/Berlin
1285155	11/5/2019 20:00	11/5/2019 20:10	7	549	iPhone9,4	12.2	Asia/Dubai
1286168	12/5/2019 8:53	12/5/2019 9:00	7	97	Watch3,2	5.1.3	Asia/Dubai
1286233	12/5/2019 10:07	12/5/2019 10:07	7	8	Watch3,2	5.1.3	Australia/Adelaide

later. We can see that their connecting flight took them to Dubai for their layover before reaching Europe and their return route was the same.

The issue here is that the person claims they were in Frankfurt but according to the database they were in Berlin, so how do we determine which is the truth? Around this time they posted this image to social media showing their hire car, shown in Fig. 19.

Analysing the properties or metadata of that image reveals the following information in Fig. 20:

As indicated by the metadata there is a pair of latitude and longitude coordinates for when the photo was taken. Putting these coordinates into Google Maps and we get the location shown below, Frankfurt Airport, shown in Fig. 21.

Fig. 19. OSINT Travel car hire photo.

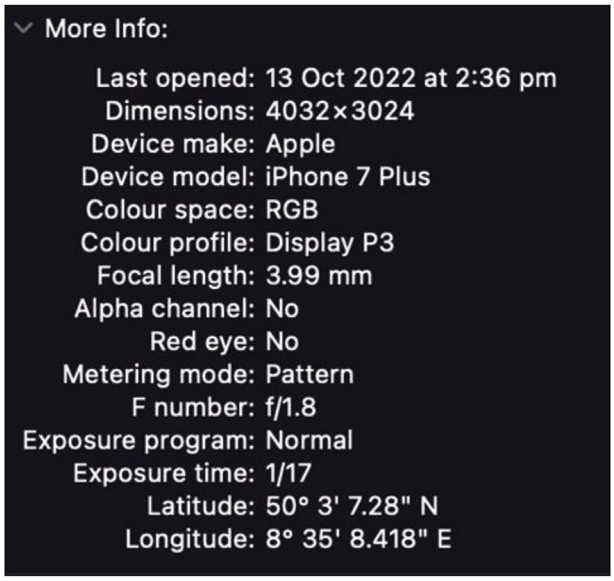

Fig. 20. OSINT Travel car hire photo metadata.

4.2.3. Results summary

From this case study we have been able to identify a persons international travel itinerary, as in their destination and times, from their health data. This becomes a powerful tool especially when that person has not done any workouts to trigger logging of geo-location coordinate data in the database, allowing us to still retrieve some macro level location data. However similarly to section 4.1 this is not without some limitations.

4.2.4. Limitations

As in section 4.1, the time zones logged in the health data are not entirely accurate. Domestically such as in Australia there isn't much of an issue as the health data separates the time zones into the different states for the country e.g. Australia/Adelaide. However as the owner of the health data travels internationally, the health data stops considering the individual states and summarises with the continent and capital of the country, e.g. Europe/Berlin instead of Germany/Frankfurt. This limitation means that for its use in an investigation, practitioners must rely on supplementary sources of information such as witnesses, image data, CCTV, etc. to validate the actual city a person travels to if they travel internationally, or if the state they travel to has more than one airport for domestic flights. Another aspect to consider is the issue of tampering. With a short thought experiment, the authors have seen that manually changing the time zone in the iPhones settings to Pakistan, is reflected in the database for new exercise records. However, with the limitations on accuracy already discussed such that time zones alone can not be solely relied upon and another source of information is required for validation, the effect of time zone tampering is minimal.

5. Conclusions

We provide a detailed explanation of the Apple Health database structure, demonstrating how to link time zone provenance with time stamps to show macro-level location information for a persons international travel. We also demonstrate the impacts of a major iOS version change with iOS 16, detailing the changes to the database structure that comes with it and the potential for more location information to be found within the database.

As demonstrated in the case studies in section 4, interesting

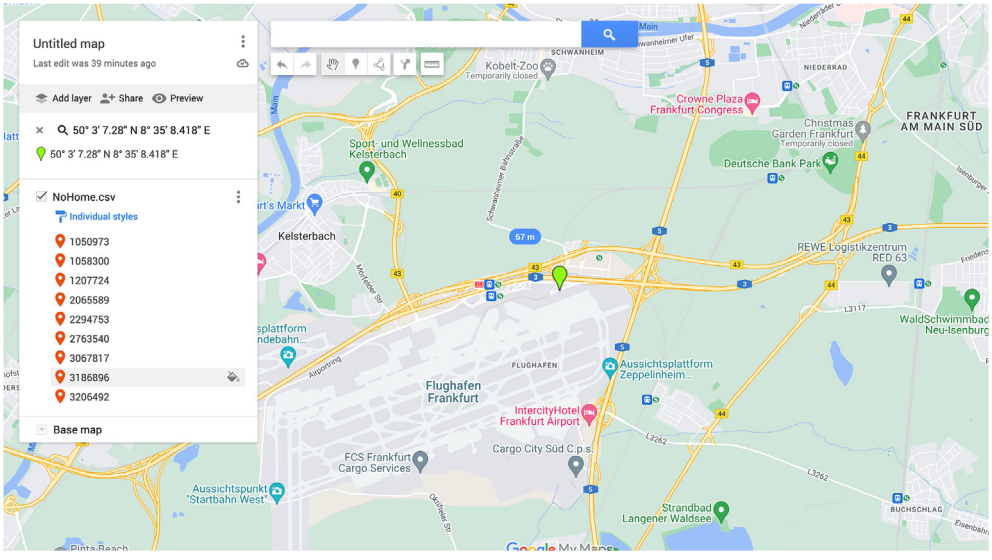

Fig. 21. Metadata location – Frankfurt Airport.

location information can be extracted from the Apple Health database to help inform an investigation, provide leads, and support other evidence such as witness statements. Case study 1 looked at the workout geolocation coordinates, showing that the data recorded by Apple Health, such as a workout being listed as *activity_type* 35 or "rowing" can not be taken at face value and requires some interpretation. This involved plotting the locations out on a map and looking deeper at the intensity of the workout (energy burned) to discover what the person of interest was actually doing, providing more context to their actions and behaviour. Case study 2 looked at establishing a persons travel itinerary when flying both domestically and internationally purely from their health data. Both case studies also demonstrated the limitations of the health data. For geolocation coordinate data, the accuracy of the GPS can vary significantly and for the Time Zones, Apple Health takes a shortcut for international destinations and uses the capital of that country instead of the actual city being visited. While these limitations are certainly significant, the interpretations and context to a persons activities provided by the Apple Health database can be a powerful tool for investigators and researchers.

5.1. Future work

There are many aspects that can be expanded upon to further understand the depth of the limitations of the database. For the first case study regarding coordinate information recorded for workouts, establishing the cause for the variation in GPS accuracy will be extremely useful. For the second case study, establishing if an Apple Device bought, registered and used in Germany has its time zones more accurately represented in the database similarly to how our dataset, bought registered and used in Australia displays country and state. As of iOS 16 there are 108 tables in this dataset. This database is a rich source of potentially valuable forensic information, but we cannot cover every single table and possible case study in a single paper. The authors are actively researching and have papers in the pipeline on other interesting case studies and the tables utilized in that process, as well as other issues involving what stimuli and conditions are necessary to trigger step counts or other activities, such as workouts, to be initiated and recorded into the database.

References

A visual explanation of sqlite joins. https://www.sqlitetutorial.net/sqlite-join/. (Accessed 11 May 2022).

Afonin, Oleg, 2018. Extracting apple health data from icloud. https://blog.elcomsoft.com/2018/11/extracting-apple-health-data-from-icloud/. (Accessed 11 May 2022).

Albrecht, Katherine, McIntyre, Liz, 2014. Psst... your location is showing!: metadata in digital photos and posts could be revealing more than you realize. IEEE Consumer Electronics Magazine 4 (1), 94–96.

Apple. Hkworkoutactivitytype. https://developer.apple.com/documentation/healthkit/hkworkoutactivitytype. (Accessed 12 October 2022).

Apple. https://www.apple.com/au/ios/health/. (Accessed 20 May 2022).

Carter, Mahalia, 2021. Caroline nilsson found not guilty of murder-ing mother-in-law after smart watch case retrial. https://www.abc.net.au/news/2021-10-26/caroline-nilsson-found-not-guilty-of-murder-in-smart-watch-case/100564980. (Accessed 20 May 2022).

Chuang, Hsiu-Min, Chang, Chia-Hui, Kao, Ting-Yao, Chung-Ting Cheng, Ya-Yun Huang, Cheong, Kuo-Pin, 2016. Enabling maps/location searches on mobile devices: constructing a poi database via focused crawling and information extraction. Int. J. Geogr. Inf. Sci. 30 (7), 1405–1425.

Clover, Juli, 2021. Apple health app data helps send a man to prison for his wife's death. https://www.macrumors.com/2021/02/09/apple-health-app-data-crime-investigation/. (Accessed 20 May 2022).

Counterpoint, 2022. Infographic: smartwatch market — q1 2022. https://www.counterpointresearch.com/infographic-smartwatch-market-q1-2022/#:~:text=Smartwatch%20Market%20Share%20by%20Brand,14%25%20YoY%20in%20Q1%202022. (Accessed 10 October 2022).

Cunche, Mathieu, 2014. I know your mac address: targeted tracking of individual using wi-fi. J. Comput. Virol. Hack. Tech. 10 (4), 219–227.

D'Mello, Gwyn, 2019. Police nabbed a killer by looking at gps data from his smartwatch, as evidence of his crime. https://www.indiatimes.com/technology/news/a-uk-mob-hitman-was-jailed-because-his-smartwatch-revealed-his-location-to-the-police-360874.html. (Accessed 20 May 2022).

Edwards, Sarah, 2016. Ios of sauron – how ios tracks everything you do. https://www.digitalforensics.com/blog/ios-of-sauron-how-ios-tracks-everything-you-do/#content-anchor. (Accessed 20 May 2022).

Espinosa, Hugo G., Thiel, David V., Sorell, Matthew, Rowlands, David, 2020. Can we trust inertial and heart rate sensor data from an apple watch device? Multidiscipl. Digit. Publ. Instit. 49, 128.

Google. Importing global positioning systems (gps) data in google earth desktop. https://www.google.com/earth/outreach/learn/importing-global-positioning-systems-gps-data-in-google-earth/. (Accessed 11 May 2022).

https://forensicrodeo.ctfd.io/. (Accessed October 2022).

iLEAPP. https://github.com/abrignoni/iLEAPP/blob/master/scripts/artifacts/Health.py. (Accessed 8 December 2022).

Maus, Stefan, Höfken, Hans, Schuba, Marko, 2011. Forensic analysis of geodata in android smartphones. In: International Conference on Cybercrime, Security and Digital Forensics. https://www.semanticscholar.org/paper/Forensic-Analysis-of-Geodata-in-Android-Smartphones-Maus-Höfken/5cad6c29acc200f2ebbf13efa5c9df4533cc836a.

Merry, Krista, Bettinger, Pete, 2019. Smartphone gps accuracy study in an urban environment. PLoS One 14 (7), 1–19, 07.

Moore, Jason, Baggili, Ibrahim, Frank, Breitinger, 2017. Find me if you can: mobile gps mapping applications forensic analysis & snavp the open source, modular, extensible parser. J. Digit. Foren. Secur. Law 12 (1), 7.

MSAB. Xry – mobile forensics and data recovery software. https://www.msab.com/product/xry-extract/. (Accessed 20 May 2022).

Opie, Rebecca, 2019. Smartwatch reliability to be scrutinised by tech experts in alleged murder case. https://www.abc.net.au/news/2019-04-11/technology-experts-to-give-evidence-about-smartwatch-data/10995210. (Accessed 20 May 2022).

Peter van Zandwijk, Jan, Boztas, Abdul, 2019. The iphone health app from a forensic perspective: can steps and distances registered during walking and running be used as digital evidence? Digit. Invest. 28, S126–S133.

Piacentini, Mauricio. Db browser for sqlite. https://sqlitebrowser.org/. (Accessed 11 October 2022).

Ramesh Kumar, P., Srikanth, Ch, Sailaja, K.L., 2016. Location identification of the individual based on image metadata. Procedia Comput. Sci. 85, 451–454. International Conference on Computational Modelling and Security (CMS 2016).

Reincubate Ltd. Iphone backup extractor. https://www.iphonebackupextractor.com/. (Accessed 11 October 2022).

Rodriguez, Andrea Macarulla, Tiberius, Christian, Bree, Roel van, Geradts, Zeno, 2018. Google timeline accuracy assessment and error prediction. Foren. Sci. Res. 3 (3), 240–255.

Spreitzenbarth, Michael, Schmitt, Sven, Freiling, Felix, 2012. Comparing sources of location data from android smartphones. In: IFIP International Conference on Digital Forensics. Springer, pp. 143–157.

Trogh, Jens, Plets, David, Surewaard, Erik, Spiessens, Mathias, Versichele, Mathias, Martens, Luc, Joseph, Wout, 2019. Outdoor location tracking of mobile devices in cellular networks. EURASIP J. Wirel. Commun. Netw. (1), 1–18, 2019.