universität wien

MASTERARBEIT | MASTER'S THESIS

Titel | Title

Dynamic Anonymisation for textual location information and digital traces derived therefrom in metropolitan areas: a protective software application for sensitive user data.

verfasst von | submitted by

Aldina Kovacevic BSc

angestrebter akademischer Grad | in partial fulfilment of the requirements for the degree of

Master of Science (MSc)

Wien | Vienna, 2025

Studienkennzahl lt. Studienblatt | Degree programme code as it appears on the student record sheet:	UA 066 856
Studienrichtung lt. Studienblatt | Degree programme as it appears on the student record sheet:	Masterstudium Kartographie und Geoinformation
Betreut von | Supervisor:	Assoz. Prof. Dr. Ourania Kounadi BSc MSc

Contents

List of Figures

List of Tables

Kurzfassung

In der vorliegenden Masterarbeit wird die Anonymisierung von textuellen Standortinformationen in digitalen Anwendungen untersucht. Ziel der Arbeit ist die Entwicklung einer benutzerfreundlichen Lösung, die Datenschutz und praktische Anwendbarkeit vereint.

Zu Beginn des Forschungsprojekts wurden die Datenschutzpraktiken von Google Maps und WienMobil analysiert. Sowohl die eine als auch die andere Plattform bedienen sich der De-Identifizierung und der Differential Privacy, jedoch ist die Transparenz für die Nutzer*innen begrenzt. Google Maps verfolgt einen umfassenden, ökosystemweiten Ansatz, während WienMobil sich auf lokale Mobilitätsdaten und Datenminimierung konzentriert.

Als Alternative wurde die Android-Anwendung DeLocator entwickelt, die einen innovativen Substitutionsansatz verfolgt. DeLocator substituiert private Adressen durch zufällig selektierte Points of Interest (POIs) aus sechs Kategorien innerhalb eines Radius von 500 Metern. Die Anwendung bietet kontextbezogene Anonymisierung für verschiedene Lebensbereiche und gewährleistet durch lokale Datenverarbeitung vollständige Autonomie für die Nutzer. Ein wesentliches Designmerkmal ist die dauerhafte Zuordnung der ursprünglichen Adressen zu anonymisierten Adressen.

Die Evaluation mit sechs Teilnehmern resultierte in überaus positiven Resultaten, wobei die durchschnittliche Usability-Bewertung bei 4,88/5,0 Punkten lag. Die TeilnehmerInnen zeigten ein vollständiges Verständnis des Konzepts der Anonymisierung und bewerteten den Kompromiss zwischen Datenschutz und praktischer Anwendbarkeit positiv.

Die vorliegende Arbeit leistet einen wissenschaftlichen Beitrag, indem sie Nissenbaums Theorie der kontextuellen Integrität operationalisiert und ein Paradigma der "pragmatischen Privatsphäre" etabliert. In der praktischen Anwendung präsentiert DeLocator eine funktionale Open-Source-Alternative und demonstriert Designprinzipien für benutzerfreundliche Datenschutztools.

Die Ergebnisse der Untersuchung legen nahe, dass eine effektive Umsetzung von Standortdatenschutz und eine hohe Benutzerfreundlichkeit durch nutzerzentrierte Designansätze in einer erfolgreichen Wechselbeziehung zueinanderstehen.

Abstract

The present master's thesis examines the anonymisation of textual location information in digital applications. The objective of this thesis is to develop a user-friendly solution that combines data protection and practical applicability.

At the inception of the research project, an analysis was conducted of the data protection practices of Google Maps and WienMobil. It is evident that both platforms employ de-identification and differential privacy techniques, yet the extent to which user transparency is facilitated remains constrained. Google Maps employs a comprehensive, ecosystem-wide approach, while WienMobil focuses on local mobility data and data minimisation.

Consequently, an alternative solution in the form of the Android application DeLocator was developed, adopting an innovative substitution approach. DeLocator substitutes private addresses with randomly selected points of interest (POIs) from six categories within a radius of 500 metres. The application offers context-based anonymisation for different areas of life and ensures complete autonomy for users through local data processing. A fundamental design element pertains to the permanent allocation of the original addresses to anonymised addresses.

The evaluation with six participants yielded extremely positive results, with an average usability rating of 4.88/5.0 points. The participants demonstrated a comprehensive grasp of the concept of anonymisation and expressed a favourable opinion of the compromise between data protection and practical applicability.

This work makes a scientific contribution by operationalising Nissenbaum's theory of contextual integrity and establishing a paradigm of "pragmatic privacy". In practical application, DeLocator presents a functional open-source alternative and demonstrates design principles for user-friendly privacy tools.

The findings of the study indicate that effective implementation of location privacy and high usability are successfully interrelated through user-centered design approaches.

1 Introduction

This chapter establishes the fundamental context for research into the anonymisation of location data in digital communication applications. The widespread use of location-based services and mobile technologies has created new opportunities for personalised digital experiences. However, this has also brought significant privacy risks for users who share location data across various platforms and applications.

The chapter starts by presenting the essential background to the current situation regarding the use of location data and the associated challenges for data protection. It highlights the relevance and urgency of this research in the broader context of digital data protection and data security. Research objectives are clearly defined research questions that guide the investigation. The methodological approach is clear: a combination of theoretical analysis, practical software development and empirical evaluation. This approach tackles the challenges identified and provides meaningful solutions to the field of location data protection.

1.1 Background

Location privacy can be defined as the right to protect an individual's location information from collection, use, or disclosure without the knowledge or consent of the data subject. This may encompass data pertaining to an individual's physical location, in addition to their online location. This includes information such as IP addresses, location data gathered through the utilisation of mobile devices or GPS-enabled devices, and textual location data disseminated across digital platforms that are an integral aspect of our daily lives. The protection of location data is a significant issue, as this information can be exploited to identify individuals and track their movements. This has the potential to have significant implications for the privacy of individuals (Abdelmoty & Alrayes, 2017; Kounadi & Resch, 2018; Kounadi et al., 2018; Seidl et al., 2020).

Geoprivacy initiatives in geographic research predominantly encompass the utilisation of geomasking methodologies, which deliberately render geographical data inaccurate with the objective of safeguarding both confidentiality and spatial distribution. The objective of geomasking applications is to facilitate the sharing of geographic data while ensuring the protection of individual data subjects' privacy. These applications propose a structure in which trusted experts make decisions on the appropriate parameters to safeguard the confidentiality and spatial distribution of the data (Seidl et al., 2020).

Recent years have seen the development of novel methodologies that leverage the unique properties of masked data. To illustrate, Rao et al. (2020) have devised a pioneering privacy-focused approach for safeguarding trajectories, while Polzin and Kounadi (2021) have proposed an adaptive technique for residential masking. Other strategies have been developed, including a modified k-anonymity approach, and a verified neighbour approach (McKenzie et al., 2022).

The degree of privacy protection afforded by an anonymised dataset is evaluated through the utilisation of the k-anonymity metric, as pioneered by Samarati and Sweeney (1998) for tabular data. K-anonymity signifies a methodology wherein a minimum of a specified number of individuals (referred to as 'k') must possess a congruent set of characteristics (e.g., sex, gender, date of birth, etc.). Consequently, the dataset does not allow for the unique identification of a particular individual (McKenzie et al., 2022; Zhang et al., 2019).

1.2 Relevance of the thesis

As previously stated, a number of methods have been proposed with the objective of enabling data subjects to obscure their geographical location. The extant tools take geographical location information into account, but not textual location information. Textual location information refers to data that describes the location of an individual or object, such as the address of a residence or the name of a neighbourhood. In recent years, the progressive proliferation of location tracking technologies and the associated publication of textual location information has led to a growing awareness of the importance of protecting the privacy of geographic locations.

A substantial body of research has been conducted on the privacy and security implications of location tracking, particularly concerning textual location information. Studies have demonstrated that even minimal location data can lead to user identification. For instance, de Montjoye et al. (2013) showed that four spatio-temporal points are sufficient to uniquely identify 95% of individuals in a mobility database .

This research has identified a need for the development of masking methods for textual location information and tools for data subjects to use daily on common digital platforms that collect such information. Data masking involves replacing sensitive information with fictitious but realistic data, thereby protecting privacy while maintaining data utility. Techniques such as pseudonymization and spatial cloaking are commonly employed (Camtepe & Yener, 2005).

Furthermore, empowering individuals with tools to manage their location data is crucial. Guidelines for public administrations on location privacy recommend transparency and media literacy to help citizens control their data (Li et al., 2018).

Additionally, various data masking tools have been developed to assist users in protecting sensitive information on shared digital platforms. These tools help obscure personal location details while allowing continued use of location-based services (Camtepe & Yener, 2005).

Moreover, assessments of the privacy threats posed by shared digital platforms have been conducted, emphasizing the need for robust confidentiality and access control mechanisms to secure user information.

To address these challenges, several strategies are recommended:

- **Implementing Data Masking Techniques:** Employing data masking tools can help protect sensitive information by disguising it, thus maintaining privacy while allowing safe data usage for testing, analytics, or sharing (Camtepe & Yener, 2005).

- **Enhancing User Awareness:** Educating users about the importance of privacy settings and the potential risks associated with sharing location information can empower them to make informed decisions (Li et al., 2018).

- **Developing Privacy-Preserving Technologies:** Investing in technologies that provide location-based services without compromising user privacy is essential. For example, spatial cloaking techniques blur users' exact locations into cloaked regions to minimize the disclosure of private information (Bilogrevic, 2018).

By implementing these strategies, individuals and organizations can better navigate the complexities of location privacy in the digital age.

A substantial body of research has been conducted on the privacy and security implications of location tracking, particularly in the context of textual location information. This research has identified a need for the development of a masking method for textual location information and a tool for data subjects for their daily usage of common digital platforms that collect textual and discrete location information. Furthermore, an assessment of the privacy threats posed by shared digital platforms and the provision of guidance on them is conducted.

1.3 Objectives and Research Questions

The **first scientific objective (SO1)** of this master's thesis is to examine various digital platforms that collect textual and discrete location information, with a view to identifying differences in practice regarding the collection, storage, protection, sharing, and terms of use of such data. Consequently, an evaluation and formulation of recommendations concerning the spatial privacy risks associated with prevalent digital platforms is undertaken.

The **second scientific and primary objective** (**SO2**) of this master's thesis is to develop a user-friendly software application for data subjects that protects textual location information and prevents (re-)identification. Consequently, the developed open-source software applications are made available to the scientific community and content developers.

In order to achieve these objectives, the following research questions are posed:

SO1

1.1. What are the different approaches in terms of collection, storage, protection, sharing, and terms of use of user specific textual location information in commonly used digital platforms?

1.2. Do commonly used digital platforms employ location protection methods or user privacy enhancement tools for data subjects to lower the risk of re-identification? If so, what algorithms are being used and how can they be evaluated in terms of privacy protection offered?

SO2

2.1.What are the usability requirements for a user-friendly location protection software application that balances both the protection and the quality of a requested service?

2.2 To what degree can the developed software applications prevent re-identification?

2.3 How is the protection level dependent on properties such as characteristics of the study area, user activity, or type of digital platform?

1.4 Methodical approach

SO1 is examined with the aid of a literature review. This section of the thesis examines the various digital platforms that collect textual location information and identifies the differences in practice regarding the collection, storage, protection, sharing, and terms of use of this data. Furthermore, an evaluation and formulation of recommendations concerning the spatial privacy risks posed by common digital platforms is provided in Chapter 5.2.

In order to examine the practical aspects of this study, a Python-based mobile application is developed. This application handles the given textual location information of the data subject and protects it from (re-)identification by providing non-sensitive address data as a placeholder.

Accordingly, a dynamic anonymisation method is developed. Furthermore, the open-source software application is evaluated in terms of its user-friendliness and made available to the scientific community and content developers.

2 Literature Review

This chapter provides a comprehensive review of extant research and current practices related to the privacy and security of text-based location data. The review synthesises insights from several interrelated areas to establish a solid theoretical foundation for understanding both the challenges and potential solutions in the field of location data protection.

The chapter systematically examines the specific privacy and security risks of text-based location data, including a detailed analysis of re-identification and inference attack methods and their potential consequences. The text undertakes a comprehensive examination of established privacy protection techniques, ranging from traditional anonymisation approaches to modern methods such as differential privacy. A particular focus is placed on analysing the current practices of large digital platforms, with a specific examination of how Google Maps and WienMobil handle the collection, storage and protection of text-based location data.

The study also covers the legal and ethical framework for the processing of location data and concludes with an analysis of the existing privacy tools available to users. This comprehensive study identifies critical gaps in current solutions and establishes the research context that motivates the development of more effective, user-oriented privacy mechanisms.

2.1 Privacy and Security in textual location information

The increasing use of location-based services (LBS) in various sectors has drawn attention to privacy and security issues related to textual location information. Textual location data, including addresses, place names and informal descriptions of places, has the potential to divulge significant information about an individual's activities and habits. This, in turn, can engender risks of privacy breaches and security vulnerabilities. This chapter undertakes a detailed examination of the privacy and security challenges posed by LBS, with a particular focus on the risks of re-identification, security threats, privacy preservation techniques, and ethical considerations.

2.1.1 Definition and Scope of Textual Location Information

Textual location information is defined as any descriptive or textual representation of geographic data, including, but not limited to, the names of cities, street addresses and informal phrases such as “near the public library”. In contrast to precise geographic coordinates, which provide exact location data, textual descriptions offer a flexible yet equally accurate representation of a person's location (Sweeney, 2002).

This data can originate from various sources, including social media posts, online directories, review platforms and government datasets. To illustrate this point, consider a scenario where an individual mentions their visit to a café in a blog post, thereby providing textual location data that can be matched with public datasets. Such an act, however, has the potential to compromise their privacy (Narayanan & Shmatikov, 2008). Textual location data has a significant role across various domains, e.g. targeted advertising, urban planning, and health monitoring (Zang & Bolot, 2011). Textual data can be categorised into three types: voluntary, observed and inferred data *(as shown in Figure 1)*. Voluntary data includes information that is intentionally shared, such as checking in at a restaurant on a social media platform. Observed data is collected automatically. An example of this would be location metadata embedded in images. Derived data can be obtained using analytics models, such as predicting a user's location based on location information in multiple posts (Cao et al., 2017).The area of textual location data also includes hybrid forms where textual descriptions are combined with other metadata, such as timestamps, which have the potential to raise privacy concerns due to the combination of temporal and spatial data often increasing the risk of re-identification (Dwork, 2008).

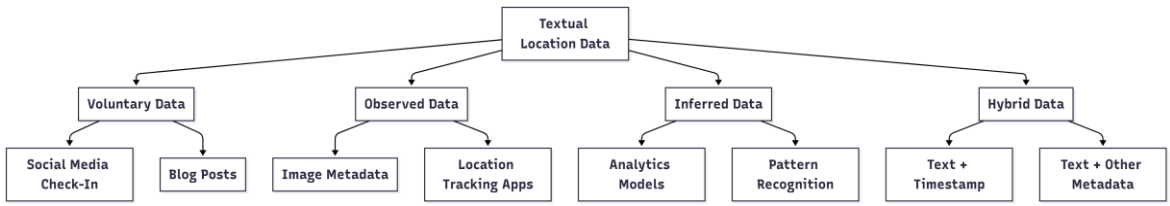

Figure 1: Types and origins of textual location data.

2.1.2 Overview of Privacy and Security Challenges

The utilisation of textual location information gives rise to intricate challenges pertaining to data protection and security that extend beyond the protection of conventional numerical location data. In contradistinction to coordinate-based location systems, textual location descriptions contain rich semantic information that can reveal sensitive personal details

through indirect inferences, thus rendering conventional anonymisation techniques insufficient (Narayanan & Shmatikov, 2008; Ohm, 2009).

Core Privacy Challenge Categories

The privacy risks associated with textual location data can be systematically divided into three primary attack vectors, each posing a different threat to individual privacy and data security *(as shown in Figure 2)*.

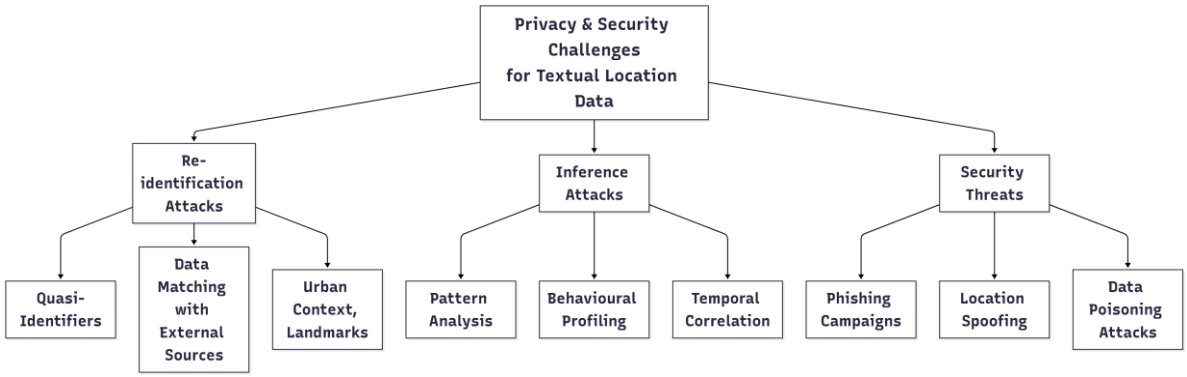

Figure 2: Overview of privacy and security challenges with textual location data.

The most immediate threat to privacy is posed by re-identification attacks, which involve the matching of anonymised textual location data sets with external data sources to uncover the identities of individuals. These attacks leverage the extensive contextual information embedded within location descriptions, encompassing cultural references, informal landmarks, and temporal patterns. When amalgamated with publicly accessible data sets, this information can function as quasi-identifiers (Sweeney, 2002). The efficacy of such attacks is particularly pronounced in urban environments, where location descriptions often refer to unique local features.

Inference attacks pose a more complex threat, as they extract sensitive personal data through pattern analysis without requiring direct identifiers. These attacks leverage the semantic diversity of text-based location data to infer behavioural patterns, residential areas, work locations, and personal preferences through temporal correlation and contextual analysis (Calderoni et al., 2015). The probabilistic nature of such attacks renders their detection and prevention particularly challenging.

The potential for security threats is increased by the fact that compromised text location information can enable a range of malicious activities, including targeted phishing campaigns, location spoofing for fraudulent purposes, and data poisoning attacks that compromise location-based systems (Jagatic et al., 2007; Zhang et al., 2019).

Unique Challenges of Textual Location Data

Textual location information possesses a number of characteristics that differentiate it from numerical coordinate data and impede conventional approaches to privacy protection. The semantic complexity of location descriptions in natural language engenders difficulties in the standardisation of anonymisation, as identical locations can be referenced by multiple linguistic variations, cultural contexts, or temporal qualifiers. The unique challenges of textual location data are summarised in Table 1.

Table 1: Comparison of Numerical and Textual Location Data Characteristics

Characteristic	***Numerical Location Data***	***Textual Location Data***
Data format	Coordinates (e.g., lat/lon)	Natural language desciptions
Semantic complexity	Low	High
Context dependency	Low	High
Anonymisation	Standardised methods possible	Difficult, meaning is context-dependent
Temporal persistence	Single points	Patterns and behavioural profiles possible
Risk of inference	Low	High, due to implicit context

The potential risks associated with contextual inference are highlighted when seemingly innocuous location descriptions implicitly disclose sensitive information pertaining to an individual's activities, health status, or personal relationships.

Furthermore, the temporal persistence of textual location data poses long-term privacy risks, as patterns of location references over time can reveal detailed behavioural profiles, even if individual data points appear innocuous (Shokri et al., 2011).

Regulatory and Ethical Implications

The data protection issues associated with textual location data have significant implications under the current data protection framework. The General Data Protection Regulation (GDPR) categorises location information as personal data, necessitating explicit consent and purpose limitation. The semantic diversity of textual descriptions may trigger additional

requirements under Article 9 for special categories of data when locations reveal sensitive personal information (Voigt & Bussche, 2017).

This overview forms the basis for a detailed examination of specific attack methods and corresponding defense strategies, which are analysed in the following chapters to provide a comprehensive understanding of the data protection landscape in the context of textual location information.

2.2 Re-identification Attacks and Methods

Re-identification attacks are characterised by the practice of cross-referencing anonymised textual location data with external datasets, with the objective of identifying individuals. These attacks leverage the exploitation of overlapping data points across multiple datasets, frequently resulting in the revelation of identities through indirect evidence.

To illustrate this point, consider a dataset comprising anonymised records of frequent visits to a gym, which could be linked with a membership list from the same establishment. This would result in the identification of specific individuals (Narayanan & Shmatikov, 2008).

2.2.1 Techniques for Re-identification Attacks

- **Quasi-Identifiers:** Quasi-identifiers are defined as non-unique data points that can be combined with other datasets to identify individuals. A quasi-identifier is understood as a non-unique data point that is capable of being combined with other datasets so as to identify individuals. Examples of quasi-identifiers include ZIP codes, dates of birth and gender. While each attribute, when considered alone, is not adequate for identification, when combined they can significantly reduce anonymity. To illustrate, Sweeney (2002) demonstrated that, in the United States, the combination of ZIP code, birthdate, and gender was sufficient for the identification of 87% of the population.

- **Linkage Attack**s: These are defined as the combining of anonymised datasets with external data sources to reveal identities. One such instance can be seen in the linking of publicly available voter registration records, which often include names, postal codes and dates of birth, with anonymised health datasets. This has been demonstrated to result in the identification of individuals, as evidenced in the research by Golle & Partridge (2009). The efficacy of linkage attacks is amplified in scenarios where datasets exhibit overlapping quasi-identifiers.

- **Record linkage algorithms:** Record linkage involves the employment of statistical algorithms and machine learning models for the purpose of linking datasets containing common identifiers. Algorithms such as k-means clustering, decision trees, and logistic regression can be utilised to identify patterns across datasets and thereby re-identify individual persons (Sweeney, 2002).

- **Data Fusion:** The process of data fusion entails the combination of multiple data sources to create a more comprehensive profile of individuals. To illustrate this point, consider the scenario of a social media post mentioning a visit to a café. When this is amalgamated with loyalty programme data from the same café, it becomes possible to ascertain the personal routines and identity of the individual in question (Sweeney, 2002).

Various techniques have been developed for re-identification attacks, including quasi-identifiers, linkage attacks, record linkage algorithms, and data fusion. The main characteristics and examples of these methods are summarised in Table 2.

Table 2:Comparison of Techniques Used in Re-identification Attacks

Technique	***Description***	***Example***
Quasi-Identifiers	Non-unique data points that, when combined, can identify individuals	ZIP code + birthdate + gender (Sweeney, 2002)
Linkage Attacks	Combining anonymised data with external sources to identify individuals	Voter registration + health dataset (Golle & Partridge, 2009)
Record Linkage Algorithms	Statistical/ML models to link datasets with common identifiers	k-means, decision trees, logistic regression (Sweeney, 2002)
Data Fusion	Combining multiple sources for comprehensive individual profiles	Social media + loyalty program (Sweeney, 2002)

2.2.2 Consequences of Re-identification

Re-identification attacks have the potential to result in severe consequences, including:

- **Loss of Privacy:** Information of a personal nature, including health records, travel history and home addresses, has the potential to be accessed by unauthorised entities.

For instance, health data integrated with public records has the capacity to divulge the medical history of an individual, thereby infringing on their right to confidentiality (Sweeney, 2002).

- **Targeted Advertising and Manipulation:** The re-identification of individuals can have far-reaching implications, including the potential for targeted advertising and manipulation. Once identified, individuals can be targeted with personalised advertisements based on their location patterns. This phenomenon has led to concerns regarding manipulation through the inference of personal preferences without consent (Zang & Bolot, 2011).

- **Discrimination and Profiling:** Re-identification can lead to profiling, where individuals are grouped based on inferred traits such as health conditions or purchasing behaviour. Discriminatory practices, such as the imposition of higher insurance premiums on individuals identified as having pre-existing conditions, have been documented (Ohm, 2009).

- **Security Threats:** The exposure of personal data has the potential to result in identity theft and social engineering attacks. To illustrate this point, consider the case of linking social media check-ins with personal identifiers, a practice which can facilitate phishing attacks (Narayanan & Shmatikov, 2008).

- **Legal and Regulatory Penalties:** Failures in the prevention of re-identification may incur substantial financial penalties in accordance with regulatory frameworks such as the General Data Protection Regulation (GDPR) and the California Consumer Privacy Act (CCPA). The General Data Protection Regulation (GDPR) stipulates the minimisation and anonymisation of data, and non-compliance can result in penalties of up to 4% of annual global turnover (Voigt & Bussche, 2017).

2.2.3 Mitigation Strategies for Re-identification Attacks

Mitigation strategies for re-identification attacks are devised with the objective of reducing the risk of identifying individuals from anonymised datasets while maintaining the utility of the data for analytical processes. The techniques employed to achieve this often require a careful balancing of privacy protection and the necessity to preserve statistical relevance, making them critical for applications in public health, urban planning and social research. The following strategies are commonly used to mitigate the risks of re-identification:

- **k-Anonymity:** As proposed by Sweeney (2002), the k-anonymity procedure ensures that, within a dataset, each data record is indistinguishable from a minimum of k-1 other records. The functionality of the procedure involves the generalization or suppression of quasi-identifiers so that they repeat across multiple data records. Nevertheless, the k-anonymity procedure is vulnerable to homogeneity attacks, where groups share identical sensitive attributes.

- **l-Diversity:** As proposed by Machanavajjhala et al. (2007), the concept of l-diversity ensures the enhancement of k-anonymity by guaranteeing the presence of a minimum of l distinct values for sensitive attributes across a given group. This prevents the ability to make inferences, even in instances where the group in question shares common quasi-identifiers. However, it is important to note that l-diversity may not be effective in the event of an imbalanced distribution of sensitive attributes.

- **t-Closeness:** The concept of t-closeness was introduced by Li et al. (2007), and it refines the concept of l-diversity by ensuring that the distribution of sensitive attributes in each group is statistically close to the distribution of the entire dataset. This measurement can be calculated using Earth Mover's Distance (EMD).

- **Differential Privacy:** Dwork's seminal work in 2008 introduced differential privacy, a foundational framework in the field that utilizes the introduction of noise drawn from a Laplace or Gaussian distribution to query results, thereby enhancing the challenge of individual data point inference. The ε-differential privacy parameter serves to modulate the trade-off between the safeguarding of personal information and the efficacy of data utilisation.

- **Data Minimization and Purpose Limitation:** In order to reduce re-identification risks and comply with General Data Protection Regulation (GDPR) principles, it is recommended that only the minimal data necessary for analysis be collected and that data usage limits be clearly defined (Voigt & Bussche, 2017).

- **Data Masking and Tokenization:** The process of masking involves the alteration or obfuscation of data by employing techniques such as randomisation or shuffling. The process of tokenization replaces sensitive data with non-sensitive tokens, which can only be converted back into their original form with a secure key. These methods are predominantly employed in the domain of financial transactions and the management of confidential information.

The following table (Table 3) summarises the main approaches to mitigating re-identification attacks, together with a brief description of each method's advantages and vulnerabilities.

Table 3: Overview of Mitigation Strategies for Re-identification Attacks

Strategy	***Main Idea***	***Strengths***	***Limitations/ Vulnerabilities***
k-Anonymity	Each record is indistinguishable from at least k-1 others	Simple to implement; widely used	Vulnerable to homogeneity and background knowledge attacks
l-Diversity	At least l distinct values for sensitive attributes in each group	Prevents inference when group shares quasi-identifiers	May not work with imbalanced attribute distributions
t-Closeness	Distribution of sensitive attributes in each group close to overall dataset	Reduces risk of attribute disclosure	Complexity in calculation; may reduce data utility
Differential Privacy	Adds noise to query results to protect individual data points	Strong theoretical privacy guarantees	May affect data utility; requires tuning of ε parameter
Data Minimization / Purpose Limitation	Only necessary data is collected and used for clearly defined purposes	Reduces exposure and risk	May limit analytical possibilities
Data Masking / Tokenization	Obfuscates or replaces sensitive data with non-sensitive tokens	Protects confidential info; useful in finance	May impact usability; token reversal risk if key compromised

Combining these strategies with techniques like regular privacy impact assessments (PIAs), encryption, and synthetic data generation can offer comprehensive protection against re-identification attacks.

2.3 Inference Attack Methods

Inference attacks, defined as the extraction of sensitive information from textual location data through adversarial analysis of patterns, correlations and indirect clues (even in the absence of direct identifiers), are a major concern. Such an attack is based on a probabilistic process of inference, in which hitherto non-expressed particulars are used to deduce the individual's personal data, activities and routines (Calderoni et al., 2015).

2.3.1 Techniques for Inference Attacks

Inference attacks rely on the ability to identify patterns and correlations within textual location data in order to deduce sensitive information about individuals without the use of explicit identifiers, thus allowing the attacker to bypass conventional data protection methods. Exploiting non-obvious relationships between data points, including repeated location mentions and temporal patterns, enables attackers to infer individuals' behaviours and identities. This section delineates the key techniques employed in inference attacks, emphasising their methodologies and the privacy risks associated with them.

- **Location Clustering:** Clustering techniques are employed to group data points on textual location based on proximity, frequency, or contextual similarity. Algorithms including k-means (Lloyd, 1982) and DBSCAN (Ester et al., 1996) are capable of identifying patterns within a dataset where the presence of multiple references to the same location may indicate a residence or a frequently visited place. The application of clustering techniques to textual data, incorporating terms such as "home" or frequent check-ins to the same café, can facilitate the identification of primary places of interest.

- **Temporal Analysis:** Temporal correlation is defined as the process of examining the timestamps associated with textual location data in order to derive behavioural patterns (Shokri et al., 2011). For example, consistent mentions of a location during morning hours may be indicative of a workplace, while consistent mentions in the evening may indicate a residence. Temporal inference models leverage time-series analysis and probabilistic reasoning to identify patterns within timestamped data sets.

- **Contextual Information Fusion:** This methodology incorporates additional metadata (e.g. timestamps, device details, and user behaviour patterns) with a view to enhancing the accuracy of the resulting inferences. Through the amalgamation of diverse datasets, such as social media activity with location tags, malefactors can

generate comprehensive profiles of individuals (Zhao et al., 2020). The utilisation of graph-based models and Bayesian networks is a common practice in the field, with the objective being to enhance the precision of these inferences.

- **Semantic Analysis:** Natural Language Processing (NLP) models are capable of analysing the semantics of locations ascribed by textual descriptions, and inferring such semantics based on the presence of particular descriptive linguistic elements. To illustrate this point, consider the following example: if a user were to describe a visit to "the tallest building in Chicago", it could be inferred that they were referring to the Willis Tower (Camtepe & Yener, 2005).

The principal techniques employed in inference attacks, together with their underlying methodologies and associated privacy risks, are summarised in Table 4.

Table 4: Overview of Techniques Used in Inference Attacks

Technique	***Main Principle***	***Example/ Algorithm***	***Privacy Risk***
Location Clustering	Grouping location mentions to identify places of interest	k-means, DBSCAN	Identification of home/workplace
Temporal Analysis	Analyzing timestamps to infer behavioural patterns	Time-series analysis	Revealing routines, workplace, residence
Contextual Information Fusion	Combining metadata (timestamps, device info, etc.) for better inference	Bayesian networks, graph models	Comprehensive behavioural profiles
Semantic Analysis	Using NLP to interpret location semantics	NLP, linguistic elements	Inferring identity through description

2.3.2 Consequences of Inference Attacks

Inference attacks have the potential to result in significant violations of privacy, given that they can expose sensitive information that was presumed to be secure. These attacks frequently go unnoticed due to their reliance on probabilistic patterns extracted from textual location data, as opposed to direct identifiers. Consequently, individuals may encounter privacy risks even when datasets have undergone anonymisation processes. The subsequent discussion will therefore outline several consequences of this kind, so as to demonstrate the significant repercussions for both personal privacy and security. The potential for serious

privacy invasion is highlighted by these attacks since the exposure can occur without the need for direct identifiers, instead leveraging probabilistic patterns derived from textual location data. Consequently, the risk of privacy violation persists even in cases where data sets have been subjected to anonymisation procedures. The consequences of such attacks are manifold and include:

- **Behaviour Profiling:** Through meticulous examination of recurring location references and temporal fluctuations, malefactors are able to construct comprehensive behavioural profiles of individuals. This accumulated information can be exploited for the purpose of targeted marketing or behavioural manipulation. For illustration, an examination of a user's recurrent visits to wellness centres could be leveraged for the purpose of health-related advertising (Zang & Bolot, 2011).

- **Location-based Discrimination:** The assumption of a given socioeconomic status or community affiliation on the basis of locational data has the potential to result in discriminatory practices. To illustrate this point, consider the potential use of a dataset indicating frequent visits to low-income areas by financial institutions. Such a dataset could be employed to make adjustments to credit scores or to deny services in an unfair manner (Hartzog & Selinger, 2013).

- **Security Risks:** Disclosing the locations at which an individual is habitually present, such as their place of residence or place of educational study, has been shown by Golle and Partridge (2009) to result in the exposure of said individual to a heightened degree of risk with respect to physical security, including but not limited to the possibility of stalking, harassment, and burglary.

- **Privacy Violations:** It has been demonstrated that even anonymised data may reveal sensitive information about lifestyle choices through the use of inference attacks. This, in turn, undermines the perceived privacy of datasets which are shared publicly (Ohm, 2009).

2.3.3 Mitigation Strategies for Inference Attacks

It is imperative to mitigate the risk of inference attacks in order to ensure the confidentiality of data sets that contain textual location information. Since such attacks depend on the identification of patterns and correlations rather than on the use of explicit identifiers, effective strategies must concentrate on the disruption of these patterns whilst ensuring the preservation of the utility of the data. The following methods have been developed to minimise the risk of privacy breaches while maintaining data utility for research and

analysis. A number of techniques that preserve privacy aim to reduce the risk of such attacks without compromising the utility of datasets containing textual location information:

- **Geographical Generalization:** This process entails a reduction in the precision of textual location data, whereby exact references are substituted with more expansive geographic areas. To illustrate this principle, consider the example of replacing a precise mention of a specific location, such as "Main Street Café", with a more general designation, such as "City Center Area". Techniques such as spatial k-anonymity have been developed to assist in the reduction of location precision (Chatzikokolakis et al., 2013).

- **Noise Injection:** This approach involves the introduction of controlled randomness in datasets through the perturbation of location data, thereby obscuring any precise patterns. In the context of differential privacy frameworks, the Laplace and Gaussian mechanisms are widely employed for the purpose of noise injection. Nevertheless, it is important to note that excessive noise has the potential to compromise the utility of the data for legitimate research purposes (Dwork, 2008).

- **Context-Aware Privacy Filters:** In the field of information technology, machine learning techniques have been employed to detect and filter sensitive patterns in data prior to its dissemination. This process involves the implementation of mechanisms such as differential privacy and pattern detection algorithms, which facilitate the automated identification and isolation of sensitive information. The objective of these methods is to detect high-risk clusters and apply additional anonymisation techniques selectively (Abowd, 2018).

- **Differential Privacy:** Differential privacy mechanisms are a formal method of introducing privacy guarantees, accomplished through the incorporation of noise which is proportional to the degree of sensitivity exhibited by the dataset in question. This mathematical process is intended to limit, as far as possible, the probability of any individual being revealed in a dataset, while achieving an optimal balance between privacy and data utility (Dwork, 2008).

- **Data Minimization:** It is vital to limit the collection of textual location data so that only the minimum required for analysis is retained; this will significantly reduce the risk of inference attacks. This approach is consistent with the concept of data minimization as defined under the General Data Protection Regulation (GDPR) (Voigt & Bussche, 2017).

Table 5 provides a comparative overview of commonly used mitigation strategies against inference attacks, including a summary of their main principles, strengths, and limitations.

Table 5: Comparison of Mitigation Strategies for Inference Attacks

Mitigation Strategy	***Principle***	***Strengths***	***Limitations***
Geographical Generalization	Reducing location precision	Decreases risk of identification	May reduce utility for analysis
Noise Injection	Adding randomness to data	Obscures patterns	Excessive noise reduces utility
Context-Aware Privacy Filters	ML-based pattern detection and selective anonymisation	Automated detection of sensitive clusters	May require complex models, risk of false positives/negatives
Differential Privacy	Formal privacy guarantees via noise injection	Strong privacy protection	Can reduce data accuracy, parameter tuning required
Data Minimization	Only collect minimum needed data	Reduces exposure	Limits analytical potential

The combination of these techniques, applied with proper privacy impact assessments (PIAs) and encryption frameworks, can help mitigate inference attacks while preserving the analytical value of textual location data.

2.4 Security Threats and Countermeasures

Beyond privacy violations, textual location information can introduce several security threats, including phishing attacks, data poisoning, and location spoofing.

2.4.1 Phishing and Social Engineering

The practice of sending deceptive communications to individuals by means of the manipulation of textual location data is termed 'spear phishing' in the literature. To illustrate this point, consider a scenario in which an assailant is aware of an individual's regular

patronage of a specific café. The assailant could then craft a deceptive email promotion, purporting to be from the aforementioned café. This tactic would serve to enhance the probability that the victim would engage with the message (Hong & Landay, 2004; Jagatic et al., 2007). Furthermore, the use of location-based information can enhance the perceived authenticity of such deceptive communications, such as by incorporating details from a recipient's recent online shopping history into a fraudulent package delivery notification. The exploitation of users' trust in personalised information by means of social engineering techniques has been demonstrated to render them more susceptible to deception.

2.4.2 Data Poisoning and Fake Data Injection

Data poisoning, as defined in the relevant literature, is the act of injecting false or misleading textual location data into datasets, with a view to disrupting the accuracy of data-driven systems. This threat can be executed by malefactors seeking to corrupt location-based decision-making processes. To elaborate, the injection of falsified location data into traffic monitoring systems is a potential modus operandi for creating artificial congestion patterns (Xiao et al., 2017), with the potential to mislead urban planners and navigation services. A related illustration is the potential for the injection of fictitious health data to create the impression of infection hotspots during a public health crisis, with the consequent misdirection of the authorities and disruption to responses (Barreno et al., 2006).

2.4.3 Location Spoofing and Fraudulent Activities

Location spoofing is defined as the deliberate falsification of an individual's location information with the intent to gain unauthorised benefits or commit fraud. Examples include the manipulation of ride-sharing or food delivery applications to exploit pricing algorithms (Zhang et al., 2019). In addition to the above, the practice of spoofing has been observed in the realm of location-based gaming. Specifically, individuals employ deceptive tactics to falsify their GPS positions, thereby circumventing the typical challenges associated with augmented reality applications. Furthermore, the security implications of spoofing are significant in location-based authentication systems, where the falsification of data can be utilised to gain physical or digital access (Wernke et al., 2014).

Spoofing attacks can be perpetrated through the use of specialized software tools that alter GPS signals or by manipulating device location settings. Consequently, the implementation of robust countermeasures against such attacks is paramount. Such countermeasures frequently entail the validation of location data through multi-source verification,

encompassing the cross-referencing of IP addresses, Wi-Fi access points and cellular data (Li & Ren, 2009).

2.5 Privacy Preservation Techniques

Privacy preservation techniques have been developed for the purpose of mitigating the risks associated with the collection, storage, and sharing of textual location information. At the same time, these techniques are designed to ensure that the information remains useful for the purposes of analysis and decision-making. These methods are developed to achieve a balanced compromise between the conflicting imperatives of data privacy and usability. This is accomplished by employing sophisticated algorithms and theoretical models in the field of data science. A number of techniques have emerged and been refined through research in both academic and industrial settings, as discussed further below (Dwork, 2008; Fung et al., 2010; Li & Ren, 2009; Machanavajjhala et al., 2007; Sweeney, 2002).

2.5.1 Data Anonymisation Algorithms

The process of data anonymisation involves the alteration or removal of personal identifiers from textual location data, with the objective of preventing re-identification. A range of methods are employed for this purpose.

- **Pseudonymisation:** Pseudonymisation, a widely employed technique for anonymous data processing, involves the replacement of direct identifiers with either pseudonyms or hash values derived using cryptographic hash functions, thereby obfuscating the original data. SHA-256 is the most common cryptographic hash function employed in the pseudonymisation process. This function transforms input data into a fixed-length, irreversible hash value. The process entails the application of the aforementioned function directly to sensitive fields, including but not limited to names and phone numbers *(as shown in Figure 3)*. Nevertheless, it is important to note that the security of pseudonymization can be compromised by linkage attacks if identical identifiers are reused across multiple datasets or if auxiliary datasets containing partial identifiers are available (Narayanan & Shmatikov, 2008).

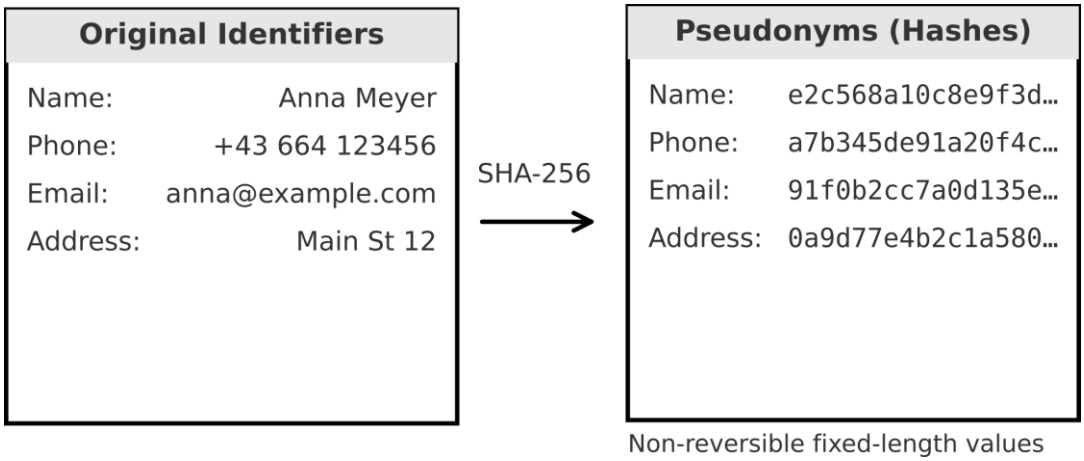

Figure 3: Example for Pseudonymisation by SHA-256

- **Generalization:** The act of generalisation can be defined as the process of converting sensitive data into categories that are less precise. For illustrative purposes, exact ages can be converted into age ranges (e.g., 20-30, 30-40), or ZIP codes can be truncated (e.g., 12345 – 123*) as illustrated in Figure 4. It should be noted that the process of generalisation is implemented through a hierarchical structure where data undergoes successive aggregation to ensure that each modified value accurately reflects a more populous group. This technique frequently enables k-anonymity (i.e. the process of ensuring that each record in a dataset shares common attributes with at least k other records; thereby reducing the risk of individual re-identification) (Machanavajjhala et al., 2007).

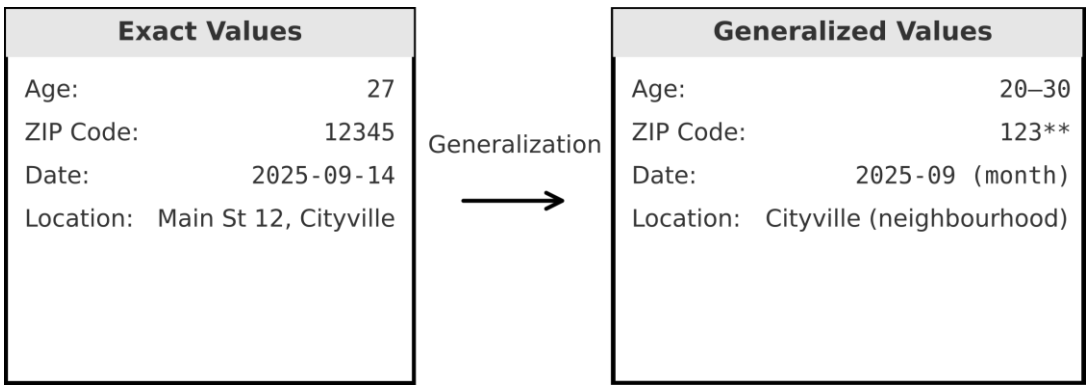

Figure 4: Example for Generalization

- **Suppression:** Suppression can be defined as the process of removing highly sensitive attributes from a dataset with the objective of preventing re-identification. The utilisation of suppression is most commonly employed when the efficacy of data anonymisation techniques, such as generalisation or pseudonymisation, is deemed insufficient to adequately safeguard the data at hand. To illustrate this point, consider the example of removing full date of birth fields while retaining only the year of birth or no information about the date of birth in general as shown in Figure 5. This can

help to reduce privacy risks. Nevertheless, suppression can result in a substantial diminution of data utility, rendering it less appropriate for datasets necessitating detailed analysis (Li et al., 2007).

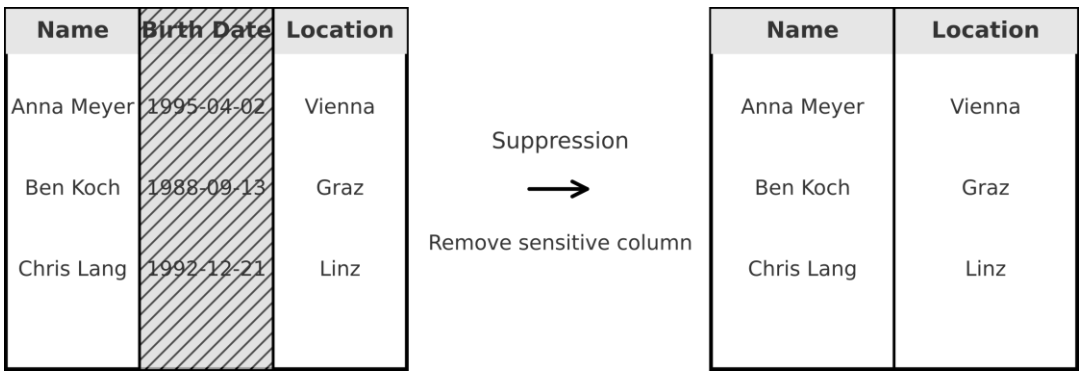

Figure 5: Example for Suppression

- **Clustering:** Clustering is a technique used for data anonymisation that involves the grouping together of data points that are similar, with a view to reducing the likelihood of individual re-identification within a given dataset. The process entails the aggregation of data points based on their similarity, which is frequently gauged through the utilisation of distance metrics or density measures. This renders the identification of specific records a challenging endeavour (Aggarwal & Yu, 2008; Fung et al., 2010).

  Clustering techniques can be broadly categorised into two distinct approaches: partition-based clustering, which includes k-means as a prominent example, and density-based clustering, which exemplified by DBSCAN. Each method employs distinct strategies for the aggregation of data points while ensuring the confidentiality of personal information.

  - *Partition-Based Clustering:* This approach involves the use of algorithms such as k-means, which aim to partition the data into k clusters, with each data point belonging to the cluster that has the nearest centroid *(as shown in Figure 6 with k=3)*. This approach is effective when the number of clusters is predefined and when the data is sufficiently separated (Jain et al., 1999).

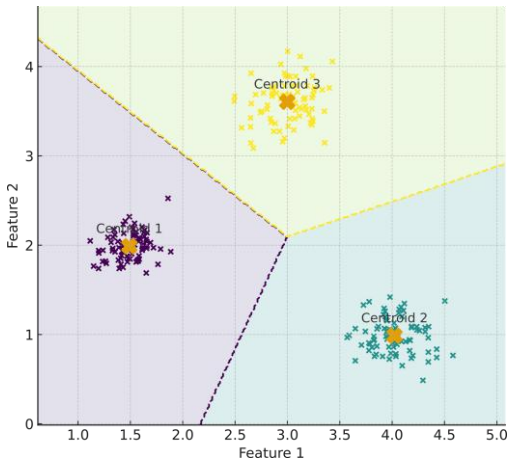

Figure 6: Example for Partition-Based Clustering

- *Density-Based Clustering:* The DBSCAN (Density-Based Spatial Clustering of Applications with Noise) method is an algorithm that clusters data points that are closely packed together and identifies points in low-density regions as outliers. Based on these principles, it was only possible to identify two DBSCANs in the example shown in Figure 7. This methodology is advantageous when managing spatial information and clusters of varying shapes (Ester et al., 1996).

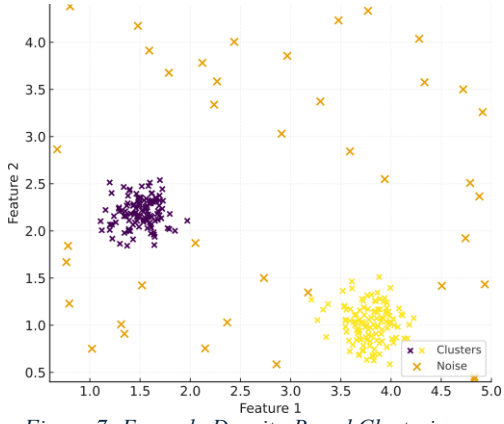

Figure 7: Example Density-Based Clustering

Table 6 summarises the principal data anonymisation algorithms, outlining their underlying principles, advantages, and limitations.

Table 6: Comparative Overview of Data Anonymisation Algorithms

Method	***Principle***	***Advantages***	***Limitations***
Pseudonymisation	Replace identifiers with pseudonyms/hash	Simple, reversible, scalable	Vulnerable to linkage attacks
Generalisation	Aggregate data into broader categories	Enables k-anonymity	Reduces data utility
Suppression	Remove sensitive attributes	Strong privacy for key fields	Data utility loss
Clustering	Group similar data points	Harder individual identification	May affect analytical precision

2.5.2 Geographical Masking Algorithms

Geographical masking algorithms have been developed for the purpose of protecting the privacy of location data. These algorithms obfuscate or generalise spatial information whilst preserving its utility for analysis. These techniques are employed in a variety of contexts, including location-based services, mobility studies, and public data releases, with the objective of mitigating the risk of re-identification from geospatial data (Chatzikokolakis et al., 2013; Fung et al., 2010; Gruteser & Grunwald, 2003). Geographical masking involves altering the precision of location data. This ensures that individual locations can no longer

be uniquely identified. However, statistical analysis and service functionality are still enabled.

- **Spatial Cloaking:** In the context of geographical information systems (GIS), spatial cloaking is a technique for obscuring precise location data by replacing a single point with a region that contains a multitude of potential points. This method frequently employs grid-based partitioning, whereby the geographic region is segmented into cells of predefined dimensions, and all points within a cell are consolidated into a composite representation. A common approach in implementing spatial cloaking involves the utilisation of the QuadTree algorithm, which recursively subdivides space into quadrants, continuing this division until the density of points within each cell attains a predefined threshold witch is illustrated in Figure 8 (Gruteser & Grunwald, 2003).

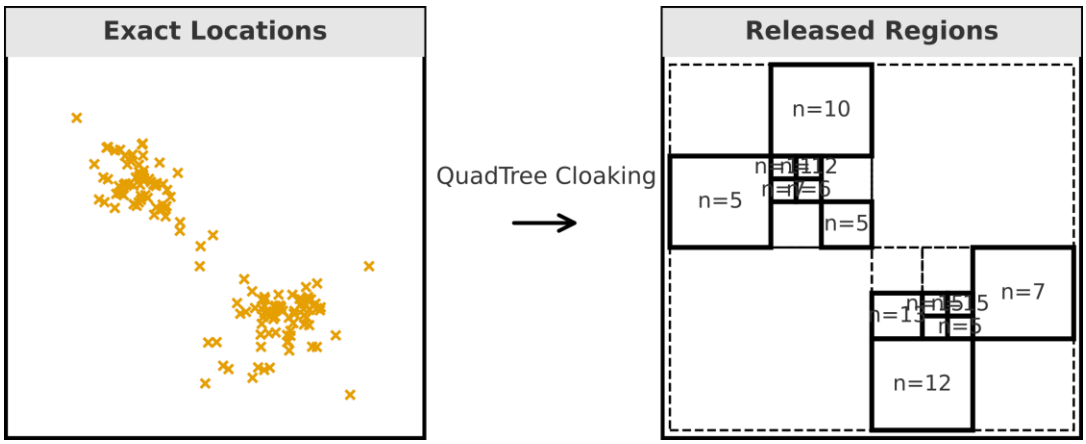

Figure 8: Example Spatial Cloaking by QuadTree algorithm

- **k-Anonymity for Geospatial Data:** The concept of k-anonymity for geospatial data involves the process of generalisation or modification of a location dataset so that each data point is indistinguishable from at least k-1 other points within the same geographic area. This technique is designed to prevent the possibility of uniquely identifying an individual on the basis of their location data, thereby enhancing privacy protection. From a mathematical perspective, the term 'k-anonymity' is employed to denote the condition whereby a dataset satisfies the condition of k-anonymity if, for each data point in the dataset, there exist a minimum of k other data points in the dataset such that the location of the data point in question falls within the same generalized region as the other data points (Chatzikokolakis et al., 2013; Fung et al., 2010; Sweeney, 2002).

  Algorithmic Steps for Achieving k-Anonymity:

  1. *Partitioning the Space:* The geographic space under consideration can be subdivided into regions, with this division being performed by way of a hierarchical partitioning technique. Such techniques may take the form of a QuadTree, or alternatively a grid-based partitioning system (Gruteser & Grunwald, 2003).

  2. *Cluster Formation:* The subsequent stage of the process is the grouping of all points within each individual partition. In instances where a partition contains an insufficient number of points, it should either be merged with a neighbouring partition or the points within it should be suppressed (Chatzikokolakis et al., 2013).

  3. *Generalization:* The specific location of each point should be replaced with a region identifier that corresponds to the partition containing at least k points (Machanavajjhala et al., 2007).

  4. *Suppression:* In instances where the process of merging partitions is unsuccessful in meeting the k-anonymity requirement, the points belonging to the partition are subject to removal with a view to safeguarding privacy (Li et al., 2007).

- **Differential Privacy for Spatial Data:** The differential privacy framework offers a rigorous mathematical foundation for ensuring the confidentiality of individual data while enabling statistical analysis on data sets. The objective is to guarantee that the inclusion or exclusion of a single data point does not substantially influence the outcome of a query. This, in turn, ensures the protection of personal privacy. Differential privacy builds on this by adding mathematically calibrated noise to datasets, ensuring individual data points cannot be reverse-engineered (Abowd, 2018; Dwork, 2008).

  Two common methods for adding noise are:

  - **Laplace Noise:** Laplace noise is a type of noise generation that employs a specific distribution, known as the Laplace distribution. This distribution is characterised by a pronounced peak at its mean value and more substantial tails than the typical normal distribution. Figure 9 provides a visual representation of the alterations to a curve that has been caused by the introduction of Laplace noise. The Laplace distribution is advantageous for use with smaller data sets, where its ability to introduce significant random variations around the true data points effectively obscures the data. This property makes Laplace noise a common choice in scenarios where enhanced privacy protection is a priority, such as in small survey datasets that include information on hospital visits. In such contexts, Laplace noise can ensure the confidentiality of individual visits while still facilitating the calculation of accurate summary statistics (Abowd, 2018; Dwork, 2008). The impact of adding Laplace noise to the data is demonstrated in Figure 9, which compares the original counts to the values obtained after the application of Laplace noise.

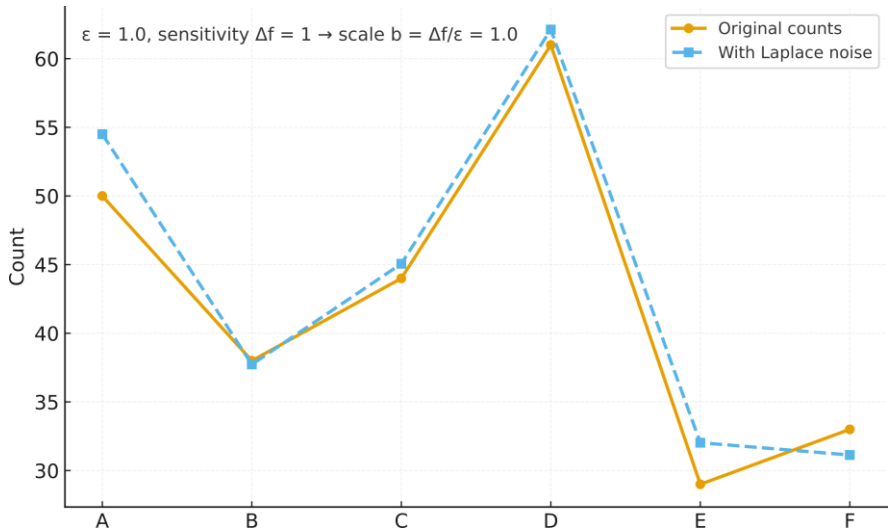

Figure 9: Example adding of Laplace Noise

- **Gaussian Noise:** Gaussian noise is characterised by its ability to follow a normal distribution, which is also known as the Gaussian distribution. This distribution is characterised by a bell-shaped curve that is commonly referred to as the normal distribution. It is considered more appropriate when dealing with larger datasets, as well as in situations where a lower level of sensitivity to outliers is required. This is due to the fact that the distribution of Gaussian noise is more balanced around the mean than other distributions. To illustrate this, consider a public transportation dataset that tracks passenger counts. In such a scenario, the utilisation of Gaussian noise facilitates a more extensive trend analysis while preserving the privacy of individual passengers (Abowd, 2018; Dwork, 2008). The effect of adding Gaussian noise to the data is illustrated in Figure 10, where the original counts are compared to the values obtained after the application of Gaussian noise.

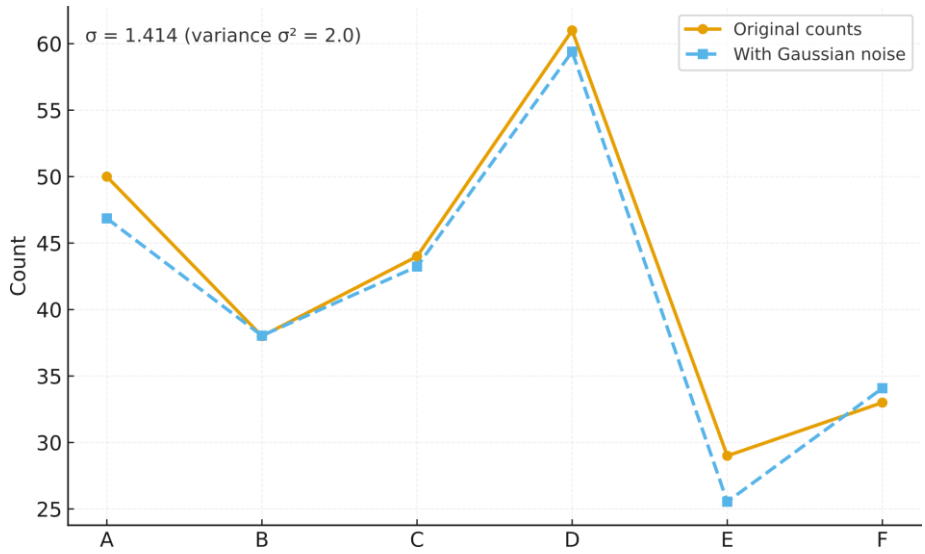

Figure 10: Example adding of Gaussian Noise

An overview of geographical masking algorithms, including their main approaches and associated trade-offs, is provided in Table 7.

Table 7: Geographical Masking Algorithms and Their Characteristics

Algorithm	Principle	Advantages	Limitations
Spatial Cloaking	Replace location with region/cell	Preserves privacy, simple	Reduces spatial accuracy
k-Anonymity	Ensure each location is indistinguishable from k others	Strong privacy guarantee	May require suppression/merging
Differential Privacy (Laplace)	Add Laplace-distributed noise	Good for small datasets	May distort small samples
Differential Privacy (Gaussian)	Add Gaussian-distributed noise	Better for large datasets	Less effective for outliers

2.5.3 Access Control Algorithms

Access control mechanisms have been identified as being of paramount importance in the context of safeguarding textual location data. These mechanisms are designed in such a way as to limit access to the relevant data to authorised individuals, thereby ensuring secure data sharing practices. Common techniques include:

- **Role-Based Access Control (RBAC):** The principle RBAC is predicated on the allocation of authorisation levels according to organisational roles. A role is defined as an attribute that determines the specific rights and privileges associated with an individual within an organizational structure. These privileges are typically determined by the role's responsibilities, with users being assigned roles that align with their respective job functions or duties. To illustrate this, consider a healthcare institution that may grant data access to medical staff based on their roles, thereby ensuring that only authorised personnel can access sensitive location data (Sandhu et al., 1996).

- **Attribute-Based Access Control (ABAC):** ABAC can be regarded as an extension of the more limited scope of the more widely applied Role-Based Access Control (RBAC), insofar as it incorporates additional attributes, including but not limited to location, time of access and device type, in the decisions made regarding access

authorisation. These attributes are then subject to evaluation by the policy engine prior to the granting or denial of access. The enhanced level of detail afforded by this process renders ABAC particularly well-suited to dynamic environments where data sensitivity is subject to variation based on contextual factors (Yuan & Tong, 2005).

- **Encryption-Based Access Control:** The encryption-based approach to access control employs cryptographic methods, such as homomorphic encryption, to facilitate the execution of computations on encrypted data without the necessity of data decryption. This facilitates secure data sharing, especially in collaborative environments where the disclosure of unprocessed data is deemed unacceptable (Gentry, 2009).

2.6 Ethical and Legal Framework

The ethical implications of textual location data call for a nuanced approach, wherein both data utility and privacy protection must be judiciously balanced. The key ethical principles that emerge from an examination of this issue are as follows:

- **Informed Consent:** It is imperative that data subjects possess a comprehensive understanding of the manner in which their location data is collected, processed and utilised. Obtaining explicit consent prior to the collection of textual location information is of the essence (Nissenbaum, 2010).

- **Transparency:** Organisations that process location data are obliged to make their data handling practices transparent. This includes the duration for which data is retained and the policies surrounding its sharing (Solove, 2008).

- **Data Minimization:** The collection of data must be limited to that which is absolutely necessary in order to fulfil the intended purpose, thereby reducing exposure to privacy risks (Cavoukian, 2009).

- **Accountability:** It is imperative that data handlers be held accountable for any misuse or unauthorised access to location data, with adherence to data protection policies being essential. The utilisation of mechanisms such as privacy impact assessments (PIAs) has been posited as a means of enhancing these accountability measures (Wright & Hert, 2012).

2.6.1 Compliance Frameworks

A range of legal frameworks and industry standards regulate the ethical use of textual location data.

- **General Data Protection Regulation (GDPR):** The GDPR exerts a substantial influence on data management practices, by way of the enforcement of stringent consent mechanisms, the stipulation of data minimisation, and the requirement for organisations to furnish data usage policies that are easily comprehensible (Voigt & Bussche, 2017). Additionally, it grants data subjects a range of rights, including the right to access their personal data, request its rectification or erasure, and object to its processing. Failure to comply with these regulations can result in substantial financial penalties, underscoring the importance for organisations to adhere to the stipulated GDPR mandates.

Article 6 of the GDPR delineates the lawful bases for processing data, while Article 25 stipulates the implementation of privacy by design and default (Voigt & Bussche, 2017).

- **California Consumer Privacy Act (CCPA):** The CCPA provides consumers with the ability to exercise control over their personal data, and it mandates transparency in data collection practices. The CCPA underscores the right to be informed about the collection of personal information and the right to have it erased upon request (Richards & Hartzog, 2020).

- **Health Insurance Portability and Accountability Act (HIPAA):** The protection of health-related location data in healthcare settings, ensuring the secure management of individually identifiable health information, is of paramount importance (Rothstein, 2010).

2.6.2 Best Practices for Ethical Compliance

In order to guarantee that they adhere to the established frameworks and maintain ethical standards, organizations are advised to consider the following strategies:

- **Data Protection Impact Assessments (DPIAs):** DPIAs are a critical component of contemporary data management practices. It has been demonstrated that regular DPIAs can assist organisations in evaluating potential privacy risks and mitigating them proactively (Wright & Hert, 2012).

- **Privacy by Design:** The integration of privacy features during the development phase of data systems is imperative to ensure that privacy measures are proactive rather than reactive (Cavoukian, 2009).

- **Data Retention Policies:** The establishment of clear guidelines concerning the retention and deletion of textual location data is imperative (Voigt & Bussche, 2017).

- **Third-Party Audits:** The implementation of independent audits constitutes a methodology for the verification of data management practices and the assurance of compliance with established standards (Richards & Hartzog, 2020).

Adherence to ethical considerations and compliance frameworks is pivotal for organisations seeking to balance the pursuit of location-based insights with the safeguarding of individuals' privacy rights. Such adherence enables these organisations to continue benefiting from location-based insights in a responsible manner.

2.7 Digital platforms that collect textual location information

The selection of Google Maps and WienMobil as the research platforms for the study on the collection and utilisation of location-specific data is based on a number of considerations that impact the geographical significance, as well as the diverse approaches to data usage in Vienna.

Google Maps is a globally recognised and widely used platform that offers a multitude of features, including map views, route planning and traffic conditions information. Google Maps is of particular utility to users in Vienna due to its versatility, as it not only furnishes local information but also establishes a global standard for map services. This enables a comparison of local circumstances in Vienna with Google's global practices regarding the collection, storage and use of location data.

In contrast, WienMobil is designed to address the specific mobility needs of the Viennese population and visitors to the city. WienMobil provides bespoke information on public transport, cycle paths and pedestrian routes via a local mobility application. The focus on urban space and the integration of diverse modes of transportation make WienMobil a significant contribution to the advancement of sustainable mobility in Vienna. Furthermore, WienMobil, in its capacity as a platform, may adhere to more rigorous data protection guidelines, which is a pertinent consideration in the context of my study on the responsible utilisation of location data.

Furthermore, both platforms demonstrate disparate perspectives on the utilisation of location data. Google Maps is a global service that employs in-depth data analysis, whereas WienMobil is a local solution that addresses the particular mobility requirements of a city. The evaluation of these two platforms enables an understanding of the diversity of approaches to the utilisation of location-specific data in Vienna, while also facilitating an exploration of the challenges and opportunities associated with the collection and utilisation of this data.

The combination of Google Maps and WienMobil offers valuable insights into the digital platforms that are significant in Vienna, including data usage and user requirements. This combination allows for a comprehensive analysis of the issues surrounding data collection, storage and security in an urban environment.

The emerge of location-based services (LBS) has fundamentally changed the way people interact with their environment, find their way around, optimise their daily journeys and discover new places. Google Maps is one of the world’s leading services in this field, known for its broad functionality and global reach. WienMobil, on the other hand, is geared towards

the city of Vienna's local public transport network. Handling text-based location information, including user-generated searches, addresses and places of interest, is critical for the functioning of both platforms.

This text provides a detailed comparison of how Google Maps and WienMobil handle the collection, storage, protection, sharing and use of text-based location information. It focuses on data protection concerns, the impact of the General Data Protection Regulation (GDPR) and the different business models of both applications. Scientific literature also provides insights into the underlying technologies, data handling methods and data protection aspects.

Table 3 summarises the key characteristics of Google Maps and WienMobil regarding the collection, storage, protection, and sharing of textual location data.

2.7.1 Data Collection: Scope and Mechanisms

In the aforementioned section, we undertake an analysis of the various approaches to data collection, together with the scope and mechanisms involved. The collection of data is a pivotal aspect of the provision of digital services, as it enables enhancements in the user experience, personalisation of services and effective functionality. The scope of data collection encompasses both user-generated data, such as personal account information, and system-collected data, including location and usage patterns. Each platform employs its own data collection methods and tools, which are tailored to its particular service objectives and user requirements.

2.7.1.1 Google Maps

Google Maps is part of a comprehensive Google ecosystem, which integrates data from various services to provide a personalised user experience. In general, there is a distinction made between data created by the user and data collected by Google. Any data shared by the user is considered personal data, including information such as the stored name, password, email address, and, in some cases, a telephone number and payment information (Google Privacy & Terms, 2025).

When utilising Google services, such as Google Maps, further data specific to the service in question may be recorded and collected. In the case of Google, specific location data collected includes GPS and other sensor data from the device used, the IP address, search queries made from specific addresses or already labelled addresses, and information about objects located in the vicinity of the device used, such as Wi-Fi access points, radio masts or Bluetooth-enabled devices. This data is collected and stored from various sources,

including cookies, pixel tags, browser web storage or application data caches, databases and server logs (Google Privacy & Terms, 2025).

2.7.1.2 WienMobil

In contrast, WienMobil is an application designed for the purpose of facilitating access to public transportation in Vienna. Its data collection is specifically oriented towards the provision of information that can be utilized in the development of transportation planning and ticketing services within the city. It is important to note that WienMobil is an application operated via Google Firebase. Google Firebase is a component of the Google Cloud Platform, which provides a range of services for app developers, including support for app development, quality control, and improvement (WienMobil, 2025).

According to Wiener Linien, the objective of utilising the Google Firebase service is to undertake continuous analysis of the WienMobil application, with a view to ensuring and enhancing its user-friendliness. This process entails the storage and subsequent analysis of anonymised user interactions. Wiener Linien has asserted that no discernible correlation exists between the available and processed data and the respective users (WienMobil, 2025).

The utilisation of WienMobil functionalities entails the transmission of user-specific data (e.g. location and movement data), the IP address and the origin and destination points entered to Google upon the creation of a route. In order to restrict the utilisation of this data to the sole purpose of map creation, an additional agreement has been concluded, which has been designated as a data protection policy for Google controllers. This policy is defined by Google and named “Google Controller-Controller Data Protection Terms” (Google Privacy & Terms, 2025). The Google Controller-Controller Data Protection Terms are predicated on the General Data Protection Regulation (GDPR) and regulate data processing between Google and its partners when both parties act as independent controllers in accordance with Art. 4 No. 7 GDPR (Google Privacy & Terms, 2025; Voigt & Bussche, 2017). With regard to data transfers to third countries, Google employs standard contractual clauses (SCCs) in accordance with Art. 46 GDPR to ensure an adequate level of data protection, especially in the aftermath of the Schrems II judgement (ECJ, C-311/18) (Google Privacy & Terms, 2025; Voigt & Bussche, 2017).

The agreement obliges Google to fulfil transparency requirements in accordance with Art. 12-14 GDPR by providing partners with information to support their information obligations. The liability provisions are based on Art. 82 GDPR, whereby each party is liable for its own processing activities. Amendments to the contractual terms are made in accordance with Art. 24 and 32 GDPR, which require adaptation to legal and technical developments (Bygrave, 2014; Google Privacy & Terms, 2025).

In order to justify the processing of data, Wiener Linien invokes Article 6(1)(b) and (f) of the GDPR (WienMobil, 2025).

> *Article 6(1)(b) of the General Data Protection Regulation (GDPR) constitutes a lawful basis for the processing of personal data, provided that it is indispensable for the performance of a contract to which the data subject is party. This condition is applicable when the processing of personal data is imperative for the delivery of a service or product requested by the data subject (Cavoukian, 2009; Voigt & Bussche, 2017).*

> *Article 6(1)(f) sanctions the processing of personal data on the basis of the legitimate interests of the controller or a third party, provided that these interests are not overridden by the fundamental rights and freedoms of the data subject. This legal basis is typically employed in instances where data processing serves a discernible business objective, such as combatting fraud or enhancing services, without compromising the user's data protection rights (Solove, 2008; Wright & Hert, 2012).*

2.7.2 Data Storage and Protection

This chapter presents an analysis of the data storage and data protection mechanisms implemented by Google Maps and WienMobil. In particular, the study examines the manner in which both platforms reconcile the safeguarding of user privacy with the demands of operational necessity and the requirements of the law. Google Maps employs a differentiated approach to data storage, categorising data according to specific deletion periods. This approach ensures the protection of user preferences, compliance with legal regulations and the stability of the services. In contrast, WienMobil adheres to the principle of data minimisation, storing data for only as long as is necessary for operational requirements or legal obligations. Specific measures are taken to anonymise user interactions within the app. This section provides an overview of the individual platforms' unique approaches to data storage, deletion protocols and the data protection precautions they use to effectively protect user data.

2.7.2.1 Google Maps

Google categorises data storage into four distinct types, each of which is subject to a different deletion timeframe.

1. The shortest deletion period is that of data deleted immediately upon user deletion; this does not require account deletion. Such data encompasses stored personal information, activities, photos stored under documents, and even products stored in

the Google account. When any of these elements are deleted, Google performs a concurrent delete (Google Privacy & Terms, 2025).

2. Data that is stored for a specified period of time. Google clarifies that this period of time is determined by Google and is deemed necessary. While the exact nature of this data is not specified, examples are provided, including the interaction between browser window height and width and their effects, as well as the storage of advertising data for a designated period before anonymisation (Google Privacy & Terms, 2025).

3. The data in question is retained until the Google account is deleted. This encompasses data pertaining to interactions with specific Google services. To illustrate, the option to delete an address from the route planner (encompassed by point 1) is available, yet the data associated with the route planner interaction persists until the Google account is deleted (Google Privacy & Terms, 2025).

4. Data that is stored for a specific, limited purpose over an extended period of time. The aforementioned points encompass data that is stored for longer periods of time and is applicable to this definition (Google Privacy & Terms, 2025):

    - Protection and prevention of fraud and abuse
        - In order to safeguard third parties, Google and users themselves from fraud, misuse and unauthorised access.
    - Financial documentation obligations
        - In the event of Google being involved in a financial transaction, including instances where Google processes a payment for users or where users make a payment to Google, it is frequently necessary to retain such data for a longer period in order to comply with accounting regulations, facilitate dispute resolution and ensure adherence to tax, fraud, anti-money laundering and other financial regulations.
    - Ensure compliance with all legal and statutory provisions
        - In order to ensure compliance with the relevant legislation, regulations and legal processes, or to comply with any enforceable

governmental requests, or to enforce the applicable terms of use, including the investigation of potential violations.

- Ensure uninterrupted access to Google services
    - In order to guarantee the uninterrupted provision of services to users.
- Communication is conducted directly with Google
    - In the event that users have engaged in direct communication with Google via a customer support enquiry, feedback form or bug report, Google may retain the requisite records of such communication.

The generalisation and addition of noise are typically employed as data protection mechanisms by Google (see Chapter 2.8 for further details). However, Google states that in certain instances, the application of protection mechanisms or anonymisation occurs only after a period of several months. To illustrate, the anonymisation of advertising data, as outlined in point 2, is only partially carried out after a time interval of 9 months. Subsequently, part of the IP address is removed from the server log and the associated cookie information is deleted after a total of 18 months (Google Privacy & Terms, 2025).

No information regarding further time intervals can be found in Google's privacy policy and terms of use.

2.7.2.2 WienMobil

Wiener Linien adheres to the principle of data minimisation, ensuring that personal data is stored for only as long as is necessary to fulfil the respective purpose. Furthermore, data is stored in accordance with legal retention obligations or when necessary to protect legal claims. Purchase and order data that is subject to tax retention obligations is deleted after a period of seven years, calculated from the calendar year in which the purchase was made. This also applies to orders that have been cancelled. Photographs submitted for the creation of an annual pass will be deleted three months after the cancellation of the contract. Browser data is deleted after the order process has been completed and IP addresses are deleted after a period of twelve months. Data from purchase processes that are used to analyse errors is deleted after a period of six months at the latest (WienMobil, 2025).

As previously stated, Wiener Linien employs the Google service Firebase for the maintenance and servicing of the WienMobil app. During this process, a range of components within this service result in the collection of personal data. The deletion intervals for this data are dependent on the components in question (WienMobil, 2025).

1. Google Firebase components, including Firebase Cloud Messaging, Firebase Remote Configuration, Firebase Crashlytics, Firebase Performance Monitoring, and Dynamic Firebase Links, operate with an ID that was assigned at the time of WienMobil app installation. It should be noted that personal data is not stored in conjunction with this process and is not linked to the assigned ID. Upon deleting the WienMobil app, the stored ID is retained for a defined period and reassigned upon subsequent reinstatement of the app (WienMobil, 2025).

2. Google Analytics for Firebase is employed for the purpose of analysing usage data. In order to optimise the application, data such as usage frequency, dwell time, functions utilised and flow analyses are utilised. The data is deleted after a period of 14 months (WienMobil, 2025).

In order to guarantee the anonymisation of data, Wiener Linien guarantees that the installation ID generated by Firebase will not be associated with any personal information (WienMobil, 2025).

2.7.3 Data sharing: scope and justifications

The following section presents an analysis of the scope and justifications for the data exchange practices of Google Maps and WienMobil. Furthermore, the respective approaches to international data transfers and the protection of user privacy are examined. As a global entity, Google Maps adheres to the established international frameworks for the protection of personal data in cross-border transfers, particularly to the United States. Concurrently, it guarantees that its data protection measures are applied consistently across all locations. In contrast, WienMobil employs data exchange procedures that prioritise essential functions for users. This is achieved through the use of cookies and selected third-party services that enhance user-friendliness. Both Google Maps and WienMobil strive to achieve a balance between the functional utilisation of data and compliance with data protection standards. This is intended to fulfil the requirements of the General Data Protection Regulation (GDPR) and other legal obligations to protect user data.

2.7.3.1 Google Maps

Google Maps engages in data sharing practices with the aim of facilitating its core services, including navigation, traffic updates, and personalised recommendations. The platform collects various types of personal data, including location data, search history, and device identifiers. These data are processed with the objective of improving the functionality of the service and enhancing the user's experience (Google Cloud, 2025).

Google disseminates user data to external service providers in designated circumstances. This data may encompass third-party platforms, such as Uber or Lyft, for integrated transport options, or content providers, including YouTube, for embedded media within the service. The aforementioned data sharing practices are governed by several international data protection frameworks, including the EU-USA Data Privacy Framework and the Switzerland-USA Data Privacy Framework. Additionally the UK extension of the EU-USA framework is also applicable. These frameworks ensure compliance with established data protection standards (European Commission, 2025).

Google relies on standard contractual clauses (SCCs) and Binding Corporate Rules (BCRs) to ensure lawful cross-border data transfers, particularly in cases involving the transfer of data outside the European Economic Area (EEA). These mechanisms are designed to minimise data, impose limitations on its intended use, and implement secure processing measures to mitigate privacy risks (Google Privacy & Terms, 2025). The European Commission similarly affirms the importance of these safeguards, emphasising data minimisation, purpose limitation, and secure processing measures to mitigate privacy risks.

It is evident that consent is a pivotal component in Google's data sharing methodologies. Prior to the sharing of personal data with third parties for non-essential purposes, explicit consent is mandatory, in accordance with the General Data Protection Regulation (GDPR). Google furthermore provides sophisticated privacy management options, enabling users to customise their cookie preferences, disable location history, and opt out of personalised data sharing (Google Privacy & Terms, 2025).

Google employs a multifaceted approach to data protection encompassing technical and organisational security measures. These measures include data encryption, anonymisation, and regular compliance audits. In addition to the implementation of technological and organisational safeguards, Google engages in periodic reviews of its data protection practices in order to align with the evolving regulatory frameworks and technological advancements (onlinesicherheit.at, 2025).

2.7.3.2 WienMobil

In order to guarantee the smooth navigation of the WienMobil website and the associated shop, cookies and local storage are employed to facilitate the utilisation of user functions and navigation on the site. In this process, personal data is stored in cookies with the objective of recognising user devices on subsequent visits. These cookies are text files stored on the device and thus saved locally. The functional cookies that are required for the basic functions of the website enable the user to log in, manage their shopping basket and complete the checkout process, among other things. The utilisation of the website may be constrained in the absence of these cookies. The storage of these functional cookies is safeguarded by legislation in accordance with Section 165 (3) TKG 2021 and Article 6 (1)(f) GDPR, as they are indispensable for the fundamental operations of the website (WienMobil, 2025).

Furthermore, third-party services are utilised for the purpose of processing data for the purpose of displaying content. Such services include Google Maps, which facilitates interaction with geographical representations and the collection of personal data for the creation of a unique user identifier. Video sharing platforms such as YouTube and Vimeo offer the option of playing videos and use advertising cookies to identify users. This data is then processed within the advertising networks of the respective providers. The transfer of data to the United States is subject to Art. 49 para. 1 lit. a i.V.m. Art. 6 para. 1 lit. a GDPR. Art. 6 para. The aforementioned Article establishes consent as the fundamental legal basis for the processing of personal data, which requires the informed, voluntary and explicit consent of the data subject for one or more specific purposes. In the context of data transfers to third countries, Art. 49 para. 1 lit. a GDPR provides for an exception in the event that no adequacy decision or other guarantees exist. In such circumstances, data transfers may only take place if the data subject has consented to them explicitly after being made aware of the risks associated with the absence of an adequate level of data protection in the destination country (Article 49 (1)(a) of the GDPR). It is incumbent upon users to be made aware of the fact that data protection in the USA does not reach the standards of the EU, and that data held by US authorities may be accessed by those same authorities without judicial oversight. This exception is categorised as a high-risk exception under Art. 49 para. 1 lit. a GDPR requires a clear affirmative action by the data subject, and ticked boxes or tacit consent are not sufficient, which emphasises the need for transparent and specific information about the processing purposes and the associated risks (WienMobil, 2025).

Wiener Linien employs the Consent Management Tool from LEGALWEB.IO for its website. This tool enables users to manage their consent to cookies and store them for up to 1095 days without reference to personal IDs. Furthermore, Wiener Linien asserts that it employs technical and organisational measures to safeguard the data, preventing

unauthorised access or disclosure. This is often achieved by storing the data within Europe (WienMobil, 2025).

Additionally, consent can be granted for the storage of supplementary data, including the last searched locations, for up to 365 days. This data is also subject to optimisation of the search function (WienMobil, 2025).

2.8 Privacy protection methods/strategies/concepts of common digital platforms

The following chapter elucidates the functionality of the privacy protection algorithms utilised by Google Maps and WienMobil, showcasing their operational capabilities through the utilisation of pertinent literature. The aforementioned analyses have identified two principal privacy protection methods, namely de-identification and differential privacy.

2.8.1 De-identification

The fundamental principles and the significance of anonymisation were elucidated in Section 2.2. In the context of digital platforms, anonymisation processes require specific adjustments to handle dynamic data flows and diverse user interactions. This section will examine how Google Maps and WienMobil implement these principles to balance data protection and service functionality.

By implementing systematic removal or masking of personally identifiable information, the risk of re-identification and data misuse is mitigated. Nevertheless, it should be noted that the efficacy of this process is contingent upon a comprehensive understanding of data identifiers and the implementation of appropriate techniques (El Emam & Arbuckle, 2013). This section will therefore proceed to examine the fundamental components of de-identification, beginning with a detailed examination of data identifiers, their classifications, and their role in the broader context of data protection. Subsequent sections will explore de-identification techniques (2.8.1.2) and de-identification types (2.8.1.3), thus providing a comprehensive understanding of the effective anonymisation of personal data and their protection.

2.8.1.1 Data identifiers

Identifiers are of critical importance in data protection, as they function to establish both the identity of an individual and the links between that person and their data. The accurate categorisation, as well as the judicious administration of such identifiers, is foundational to ensuring privacy and adherence to regulatory standards (Cavoukian, 2009).

As outlined in Section 2.2.1, identifiers are classified as either direct or quasi-identifiers. In the specific context of digital platforms, these identifiers require special attention due to the scale and speed of data processing. In platform environments, direct identifiers such as user accounts and device IDs, as well as quasi-identifiers such as activity patterns and preference data, pose particular challenges that extend beyond traditional data management scenarios.

- **Direct identifiers:**

  As outlined in Section 2.2.1, direct identifiers enable the immediate identification of an individual without additional information. In the context of digital platforms, these identifiers pose particular challenges due to their persistence across user sessions and their potential for cross-platform tracking. In the context of platform environments, direct identifiers frequently encompass user accounts, device identifiers, and authentication tokens. These elements necessitate safeguards that extend beyond the scope of conventional anonymisation methods. The regulatory implications for such identifiers are of particular significance, as frameworks such as the General Data Protection Regulation (GDPR) and the Health Insurance Portability and Accountability Act (HIPAA) mandate specific protections for these high-risk data elements (Voigt & Bussche, 2017).

- **Quasi-identifiers:**

  The utilisation of quasi-identifiers in isolation is inadequate for the exclusive identification of an individual; nevertheless, when integrated with supplementary data elements, they can exert an identifying effect. Examples of quasi-identifiers include gender, postal code, and date of birth. Consequently, for instance, a data record containing a person's postal code and gender may not be sufficient as a standalone identifier, but integrating this data record with another that includes additional attributes and unique data may facilitate identification. In instances where a substantial quantity of quasi-identifiers are amalgamated within a solitary data record, they have the capacity to coalesce into a distinctive synthesis, thereby enabling the identification of an individual even in the absence of a direct identifier (Sweeney, 2002).

The significant re-identification risks associated with combined quasi-identifiers, as documented in Section 2.2.1, are particularly relevant for platform operators. Digital platforms process vast datasets that amplify these risks, necessitating sophisticated protection methods beyond standard anonymisation approaches.

In the domain of privacy research and data protection practice, it is imperative to distinguish between direct and quasi-identifiers. The effective anonymisation of data necessitates a

comprehensive understanding of the manner in which these identifiers interact within datasets. This understanding must be accompanied by the application of privacy measures such as K-anonymity, L-diversity and differential privacy. These measures are instrumental in minimising the risk of re-identification. Common techniques employed to obscure quasi-identifiers include generalisation and suppression, whilst direct identifiers can be either removed or substituted with pseudonyms (El Emam & Arbuckle, 2013).

Furthermore, data controllers must take into account the context of data utilisation, as specific quasi-identifiers may transform into direct identifiers when connected to external data sources (El Emam & Arbuckle, 2013). This aspect of identifiers' dynamic nature underscores the imperative for ongoing assessment and the establishment of contextual data processing policies in the realm of data protection (Nissenbaum, 2010).

2.8.1.2 De-identification Techniques

In the context of anonymising information or removing as many direct and quasi-identifiers as possible in order to prevent the identification of individuals, two fundamental techniques can be identified: The two basic techniques are generalisation and randomisation.

- **Minimization:**
  Data minimisation principles, as established in Sections 2.3.3 and 2.5, are foundational to platform privacy strategies. In large-scale digital environments, these principles manifest through sophisticated data lifecycle management systems that balance service functionality, analytical requirements, and privacy protection.

  The minimisation process is a data management technique that is employed when collecting or storing data. The purpose thereof is to ensure that only the essential attributes necessary for a specific task or operation are retained. To illustrate this principle, consider the example of a complete date of birth; rather than storing the entire record, only the year of birth can be recorded. A similar approach is adopted for health data, wherein only the necessary diagnostic codes are stored, with the rest of the personal information excluded (Cavoukian, 2009).

  A variety of data minimization techniques exist, the effectiveness of which varies depending on the data context. Examples include:

  - *Attribute reduction:*
    The removal of superfluous attributes that are not relevant to the specified analysis, such as the exclusion of middle names from a customer database (Samarati & Sweeney, 1998).

- *Data aggregation:*
  This involves the conversion of individual data points into summarized values, e.g. storing average income per age group instead of individual income (Aggarwal, 2005).

- *Value truncation:*
  This process involves the reduction of data field detail, for instance, by storing only the first three digits of a postal code rather than the complete code in its entirety (Sweeney, 2002).

The purpose of these techniques is to ensure that the data that is retained is limited solely to that which is necessary (*as shown in Table 8*), whilst concomitantly reducing the risk of re-identification and the potential misuse of personal data.

Table 8: Example for Data Minimization

Attribute	***Full Collection***	***Minimized Collection***
Name	Full name (e.g. "Anna Müller")	Initials only ("A.M.") or omitted
Address	Full address ("Main Street 12, Vienna")	City only ("Vienna")
Date of Birth	Full date ("1998-09-28")	Year only ("1998")
Location History	Complete movement profile	Only most recent location
Email	Full email ("anna.mueller@mail.com")	Not collected unless required
Purchase Details	All past purchases	Only open transactions

A foundational step in data minimisation is the elimination or conversion of direct identifiers before further utilisation of the data. Pseudonymization (replacement of identifiable information with a unique code) and encryption (conversion of data into an unreadable format that requires a decryption key) are frequently employed to bolster data minimization efforts (Cavoukian, 2009).

Once implemented, data minimization can facilitate compliance with privacy models such as purpose limitation and data proportionality by ensuring that the data collected or retained is strictly aligned with its intended use. The efficacy of data minimisation is contingent upon meticulous evaluation of the configuration and

intent of the data collection, necessitating an equilibrium between data privacy and its utility (Voigt & Bussche, 2017).

- **Generalisation:**
  While the theoretical foundation of data generalisation is discussed in Section 2.5.1, its practical application in digital platforms involves complex algorithmic decisions about abstraction levels. Platform-specific generalisation approaches must accommodate diverse data types and varying privacy requirements across different jurisdictional frameworks.

  The concept of data generalisation may be defined as a type of dynamic data masking. A specific data value is substituted with a less precise one. This may appear disadvantageous, but it is, in fact, a prevalent and extensively utilized approach for data mining, analysis, and secure storage. An Example for data generalisation is shown in Table 9 (Google Privacy & Terms, 2025; Kavitha S et al., 2015; Mascetti & Bettini, 2007; Samarati & Sweeney, 1998).

Table 9: Example for Data Generalisation

Attribute	***Original Value***	***Generalised Value***
Age	27	20–30
ZIP Code	12345	123**
Date of Birth	1998-09-28	September 1998
Location	Main Street Café	City Center Area
Timestamp	2025-09-28 18:50:19	September 2025

In the event that an analysis of the data collected is required, while ensuring the protection of the individuals whose data is included, one of the most crucial applications is data generalisation. It is an effective method of abstracting personal data while maintaining the utility of the data points (Google Privacy & Terms, 2025; Kavitha S et al., 2015; Mascetti & Bettini, 2007; Samarati & Sweeney, 1998).

There exists a wide variety of techniques for generalising data, the effectiveness and ability to maintain data integrity of which vary significantly. Examples of generalisation methods include data aggregation, binning and truncation, which aim to reduce the specificity of the data while preserving its analytical value. In instances where multiple identifying data points are available, but only a subset is pertinent to the desired outcome, it is frequently feasible to implement more robust

generalisation techniques on the irrelevant data points while preserving the integrity of the significant data points (Google Privacy & Terms, 2025; Kavitha S et al., 2015; Mascetti & Bettini, 2007; Pingley et al., 2011; Samarati & Sweeney, 1998).

A fundamental stage in the procedure of anonymisation pertains to the process of masking direct identifiers. The process of data masking entails the rendering of data unrecognisable or altered in such a manner that it cannot be employed to directly identify an individual. There are several techniques that are employed for masking data. These include the use of pseudonymisation, whereby identifiable information is replaced by a unique code, and the process of encryption, whereby data is converted into a format that can only be decrypted with a specific key (Cavoukian, 2009).

Once the direct identifiers in the data have been masked, the data can be further generalised by means of the k-anonymity method. Building on the k-anonymity concept introduced in Sections 2.2.3 and 2.5.2, digital platforms implement specialized variations of this privacy measure to accommodate real-time data processing requirements. These implementations must balance computational efficiency with effective privacy protection in dynamic user environments. The key to effective k-anonymity is an optimised process that takes into account all relevant privacy and legal considerations, such as ensuring that pseudonyms do not indirectly reveal sensitive information (Mascetti & Bettini, 2007).

When generalising data, it is essential to follow established privacy guidelines. These guidelines determine the extent to which identifying data about individuals can remain unchanged, balancing the protection of privacy with the utility of the data (Pingley et al., 2011; Samarati & Sweeney, 1998).

- **Randomisation:**
  Randomising data refers to deliberately altering data within a dataset by introducing controlled randomness to protect sensitive information. This technique alters data points non-deterministically to reduce the risk of identifying individuals while maintaining the overall statistical relevance of the dataset (Dwork, 2008). Because it changes the data but preserves the analytical value, data randomisation is often considered a dynamic data masking technique. It is often used for the protection of privacy in statistical analysis, for the secure storage of data, and for the sharing of anonymised data (Aggarwal, 2005).

  Randomisation becomes an important method in contexts where data analysis is required while preserving the identity of the individual. The controlled alteration of

data points allows for the abstraction of personal data whilst retaining the utility of the data for analysis and decision making (Samarati & Sweeney, 1998).

A variety of randomisation techniques exist, each with different effectiveness and impact on data integrity. Common methods include:

- *Additive noise:*
  This involves the addition of small random values to numerical data, such as income or age (Du & Atallah, 2001).

- *Data swapping:*
  A process whereby values are exchanged between data sets, e.g. the date of birth of two individuals can be exchanged while maintaining the distribution's overall integrity (Sweeney, 2002).

- *Differential privacy:*
  This involves the introduction of mathematically calibrated noise to ensure that the presence of a single individual cannot significantly affect the results of a dataset (Dwork, 2008).

The objective of these techniques is twofold: firstly, to ensure that the data retains its analytical value, and secondly, to reduce the risk of re-identification through linkage attacks. Those principles can be applied as shown in Table 10.

Table 10: Example of Randomisation

Attribute	***Original Value***	***Randomised Value***	***Randomisation Method***
Location	Main Street Café	Main Street Café (+300m offset)	Random spatial offset
Age	27	24 or 29	Random value within plausible range
Timestamp	2025-09-28 18:50:19	2025-09-28 18:53:07	Random time shift
Purchase Amount	€49.99	€52.30	Random noise added
Device ID	12345ABC	12X45ZBC	Random character substitution

A foundational step in this process is the masking of directly identifiable information prior to the implementation of randomisation techniques. Combinations of such

methods, for instance, with techniques like pseudonymisation – whereby personally identifiable details are substituted with a token randomly selected on the basis of unpredictable criteria – or encryption, whereby the data is rendered illegible through its conversion into a format requiring a decryption key to be rendered intelligible once again, are utilised in conjunction with randomisation strategies to enhance protection of personal privacy (Cavoukian, 2009).

Following the masking of direct identifiers, further data randomisation can be implemented in order to comply with specific privacy models, for example differential privacy. The model ensures that the contribution of any individual data point to the overall dataset is negligible, even when multiple datasets are linked. The effectiveness of randomisation depends on the selection of techniques appropriate to the dataset's structure and sensitivity, achieving a balance between data privacy and utility (Dwork, 2008).

2.8.1.3 De-identification Types

Two principal forms of data generalisation have been identified. The two principal forms of data generalisation are declarative generalisation and automated generalisation. The following section will analyse the meanings of each of these concepts and examine their practical manifestations.

- **Automated:**

  The process of automated generalisation entails the *utilisation of algorithms for the purpose of ascertaining the optimal level of data abstraction* that is necessary to strike a balance between the principles of privacy protection and the optimisation of data utility. The functionality of these aforementioned algorithms involves an assessment of the structure of the dataset, with the outcome of this assessment being an automatic adjustment of the level of generalisation to achieve a pre-defined privacy threshold. One of the more common ways in which automated generalisation is measured is by the k-value, which is defined as the minimum number of data-points required for the same quasi-identifiers to be shared in order to prevent individual re-identification. K-anonymity can be regarded as a measure of privacy protection rather than a method for achieving automated generalisation. The objective of k-anonymity is to ensure that, for a given dataset, each record cannot be distinguished from a minimum of k-1 other records, based on a set of quasi-identifiers. The achievement of k-anonymity involves techniques such as generalisation, suppression, and data perturbation (Kavitha S et al., 2015; Samarati & Sweeney, 1998).

  To illustrate this, one can posit that a dataset may be designated as 2-anonymous with a k-value of 2. This is to signify that each combination of quasi-identifiers (for example, age and location) is present in a minimum of two records. This level of generalisation serves to reduce the risk of reidentification, while retaining the analytical properties of the data. In order to ensure the maintenance of privacy, the data is to be modified such that each distinct age-location pair occurs a minimum of twice within the dataset, under the assumption that the dataset contains the aforesaid attributes for multiple individuals (Google Privacy & Terms, 2025; Samarati & Sweeney, 1998).

- **Declerative:**
  In the context of declarative generalisation, the *user determines the desired level of data abstraction* prior to the dissemination or analysis of the data. This methodology

empowers data controllers to manually define the parameters of the generalised data groups, contingent on the stipulated utilisation scenario and privacy stipulations. Nevertheless, it is important to acknowledge the inherent limitations of this approach, which may encompass potential distortion or loss of accuracy. One such potential pitfall arises when outlying cases are completely excluded from consideration during data generalisation, which often results in an incomplete dataset. Notwithstanding the above limitations, declarative generalisation can be considered a valuable preliminary step in data anonymisation, especially where the objective is to limit excessive data exposure while retaining the core insights needed for analysis (Mascetti & Bettini, 2007).

2.8.2 Differential Privacy

Section 2.5.2 introduces differential privacy as a mathematical framework for privacy protection. This section extends that foundation by examining how digital platforms operationalise differential privacy through specific parameterisation choices, algorithmic implementations, and integration with broader data governance frameworks. The mathematical formulations and practical trade-offs are of particular importance in scalable platform environments.

2.8.2.1 Key principles of Differential Privacy

Differential Privacy represents an innovative approach to the protection of individuals' privacy in the context of data analysis. In order to guarantee the confidentiality of personal information, this concept is founded upon a number of fundamental principles that work in conjunction to ensure that the results of data analyses cannot be traced back to individuals. The following points elucidate the fundamental tenets of Differential Privacy, encompassing the significance of randomisation, the vulnerability of data, the privacy loss parameter (ε), and the composition of privacy losses. These principles constitute the foundation for the implementation of differential privacy and are of paramount importance in maintaining a balance between the protection of privacy and the utility of data analyses (Dwork & Roth, 2013; Nissim et al., 2016).

- **Privacy by Randomisation:**

  A core tenet of differential privacy is the employment of randomisation as a mechanism to ensure the confidentiality of data. Specifically, the implementation of differential privacy entails the incorporation of calibrated noise drawn from a Laplacian or Gaussian noise distribution into the output of a query. This ensures that

the probability of observing a specific output remains relatively unchanged when an individual is either included or excluded from the dataset. The calibration of noise is informed by the sensitivity of the dataset, thereby ensuring a balance between privacy protection and data accuracy. It is crucial to emphasise that the employment of randomisation ensures that the results of an analysis cannot be directly attributed to individual people's data. To illustrate this point, consider the calculation of an average score for a group of individuals. In this scenario, random noise can be introduced during the calculation process, thereby ensuring the anonymity of the specific contributions of each individual. This technique serves to reduce the probability that an attacker can draw conclusions about a person's identity or specific information from the published data (Dwork, 2008; Dwork & Roth, 2013; Nissim et al., 2016).

- **Data Sensitivity:**

  In the context of differential data protection, another pivotal concept pertains to the categorisation of data according to its sensitivity. The degree of data vulnerability is determined by the extent to which the outcomes of an analysis can be altered by the inclusion or exclusion of an individual's data. The formal quantification of this sensitivity, termed "global sensitivity," is characterised as the maximum possible alteration in a function's output upon the alteration of a single data point. This concept can be represented formally as follows:

  $\Delta f = \max(||f(D) - f(D')||)$

  in this context, D and D' represent data sets that differ by a single data point pertaining to an individual. It is asserted that global sensitivity is a fundamental measure of differential privacy, insofar as it is the parameter that enables calculation of the necessary amount of noise that must be added so as to ensure privacy is adequately safeguarded (Dwork, 2008; Wasserman & Zhou, 2010). Examples for the Sensitivity-Levels could include:

  - *Low Sensitivity:*

    The enumeration of individuals within a metropolitan area serves as a paradigm for low sensitivity, given that the exclusion or inclusion of a solitary entity exerts a negligible influence on the ultimate tally. This characteristic renders it more straightforward to safeguard privacy, given that the impact of a single data point is negligible (Dwork, 2008).

- *Moderate Sensitivity:*

  The calculation of the mean income in a dataset of moderate sample size demonstrates moderate sensitivity. The exclusion or inclusion of a single individual with a notably elevated or depressed income can exert a substantial influence on the mean, to a considerable extent. This observation underscores the necessity for enhanced data protection measures, such as the implementation of additional noise, in comparison to low-sensitivity data, to ensure the preservation of privacy while maintaining data utility (Nissim et al., 2016).

- *High Sensitivity:*
  The medical test results for rare diseases are characterised by a high level of sensitivity. The inclusion or exclusion of a single patient's data in a small dataset has been demonstrated to have a considerable impact on the overall statistical results, such as the disease prevalence rate. This underscores the necessity for substantial randomisation to ensure patient confidentiality while enabling statistically significant analysis (Wasserman & Zhou, 2010).

- **The privacy loss parameter (ε):**

  The parameter ε, the quantification of which is pivotal to the concept of differential privacy, serves to measure the aforementioned loss of privacy. It quantifies the impact of including or excluded an individual on the result of a specified query. Formally, it can be expressed as:

  $$Pr[M(D) = S] \leq e^{\varepsilon} \times Pr[M(D') = S]$$

  where M is the mechanism applied to datasets D and D'. A lower value for ε (e.g. 0.1) indicates stronger privacy protection, while a higher value (e.g. 1) implies weaker protection and greater data accuracy. The selection of an appropriate ε is of paramount importance, as it signifies a trade-off between data privacy and utility. It should be noted that overly low values may distort the outcomes of the study to a significant degree, whereas values that are too high may compromise the privacy of the data (Dwork & Roth, 2013; Nissim et al., 2016).

- **Composition of differential privacy:**

  Another significant principle of differential privacy is composition, which can be defined as follows. This principle delineates the phenomenon by which the privacy of data is diminished when multiple queries are made on the same dataset. Sequential

composition asserts that when multiple analyses are conducted on the same dataset, the overall privacy loss accrues. The advanced composition principle refines this principle, thereby reducing the cumulative privacy loss when multiple low-ε analyses are performed. To illustrate this, consider a scenario where the privacy loss is set at ε = 0.5 for a single analysis. If this analysis is repeated twice, the total privacy loss would accrue to ε = 1.0. However, in parallel composition, when non-overlapping subsets of data are analysed, the privacy loss remains ε for each subset. The management of privacy loss through composition is of particular importance in repeated analyses, necessitating a delicate balance between the number of queries and the desired privacy guarantees (Dwork, 2008; McSherry, 2009).

2.8.2.2 Methodology of Differential Privacy

The methodology of differential privacy is a complex and multifaceted topic that encompasses various mechanisms and considerations to ensure privacy protection in data analysis. Although the fundamental concepts were introduced in Section 2.5.2, their application in the context of digital platforms necessitates specific adjustments and methodological refinements. The Laplace mechanism and the Gaussian mechanism are widely regarded as the two most significant practical implementation approaches (Dwork & Roth, 2013; Nissim et al., 2016).

- **Laplace-Mechanism:**
  The Laplace mechanism is a pivotal method in the context of differential privacy. As outlined in Section 2.5.2, the fundamental properties of the Laplace distribution have been elucidated. However, it is imperative to emphasise that the specific implementation and parameterisation of this distribution on digital platforms is of paramount importance. The mechanism incorporates random noise from the Laplace distribution into a query result, with the magnitude of the added noise (referred to as the "noise level") being determined by the global sensitivity of the dataset and the privacy loss parameter, designated as ε (Dwork, 2008).

  Mathematically, the Laplace mechanism is described as follows:

  $$M(D) = f(D) + Lap(0, \Delta f/\varepsilon)$$

  where Lap(0, Δf/ε) denotes noise drawn from a Laplace distribution centred at zero with scale parameter Δf/ε (Dwork, 2008; Wasserman & Zhou, 2010).

One potential application of the Laplace mechanism is the estimation of the average income of a cohort of individuals. Adding noise from a Laplace distribution is a methodology that has been demonstrated to ensure that the resulting estimate cannot be traced back to individual incomes, whilst concomitantly providing useful statistical information (Dwork & Roth, 2013; Nissim et al., 2016).

- **Gaussian-Mechanism:**
  In contrast to the Laplace mechanism, the Gaussian mechanism utilises noise derived from a normal distribution. This method is frequently employed in scenarios where differential privacy is defined in terms of (ε,δ) differential privacy, which provides a weaker but more flexible privacy guarantee than the pure ε differential privacy of the Laplace mechanism (Dwork, 2008).

  Mathematically, the Gaussian mechanism described as follows:

  $$M(D) = f(D) + \mathcal{N}(0, \sigma^2)$$

  where $\mathcal{N}(0, \sigma^2)$ represents noise drawn from a normal distribution with zero mean and variance σ^2, which is determined based on the dataset's sensitivity and the desired privacy level (Dwork & Roth, 2013; Nissim et al., 2016).

  The Gaussian mechanism may be advantageous in certain scenarios, particularly when the data exhibits a specific structure or when greater accuracy is desired. For instance, it can be beneficial in scenarios where data distributions are continuous, and the objective is to achieve a balance between privacy and statistical significance (McSherry, 2009).

The Laplace and Gaussian mechanisms are employed to introduce noise in data analysis, with the purpose of ensuring the confidentiality of the data. Despite their similarities, there are important distinctions to note between the two with regard to application and statistical properties (*as shown in Figure 11*). Firstly, the Laplace mechanism is generally favoured for use with small datasets and discrete outputs due to the fact that it provides optimal privacy guarantees under the condition of pure ε-differential privacy. By contrast, the Gaussian mechanism is more appropriate for the context of continuous data and larger datasets, with the requirement of (ε, δ)-differential privacy. This is especially the case when the priority is to ensure greater accuracy. The selection of a particular mechanism should be informed by the sensitivity of the data, the privacy guarantees required, and the analytical context (Dwork & Roth, 2013; McSherry, 2009; Nissim et al., 2016; Wasserman & Zhou, 2010).

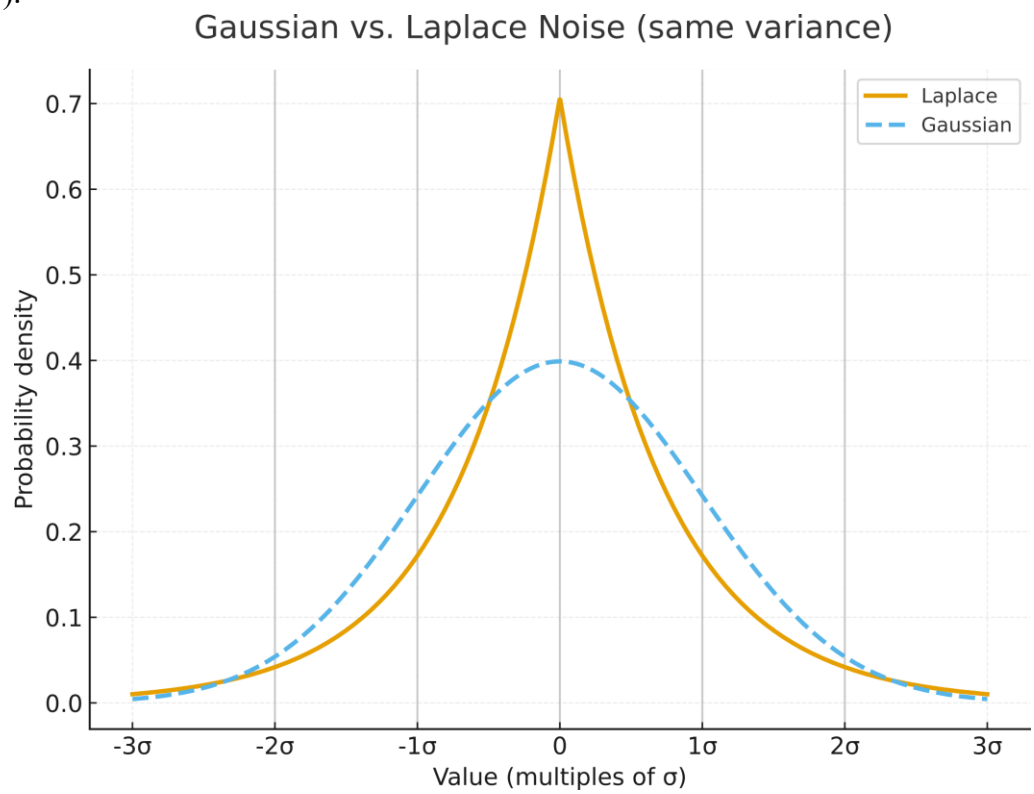

Figure 11: Gaussian vs. Laplace Noise at same variance

2.8.2.3 Implementation

The practical implementation of differential privacy as applied in digital platforms builds on the theoretical foundations outlined in Section 2.5.2, but requires additional considerations in the context of dynamic data streams. The key parameters for platform-specific implementation include:

1. **Sensitivity Assessment:**

   It is imperative for data analysts to evaluate the sensitivity of the data in order to ascertain the extent to which noise should be added. The parameter ε is pivotal in this context; lower values offer enhanced data protection, though at the expense of data accuracy (Dwork & Roth, 2013; Nissim et al., 2016).

2. **Mechanism Selection:**
   The selection of Laplace or Gaussian mechanisms is contingent upon the characteristics of the data and the analytical requirements. The utilisation of automated parameter optimisation systems in digital platforms is a common feature in the execution of such selection processes (Dwork & Roth, 2013; Nissim et al., 2016).

The implementation of such measures within existing platforms necessitates an iterative approach, incorporating continuous testing and calibration to optimise the balance between data privacy and data usability. The U.S. Census Bureau has demonstrated the successful implementation of differential privacy on a large scale, with their methods serving as a model for other organisations (Dwork & Roth, 2013; Nissim et al., 2016).

The primary challenge confronting these entities pertains to the inherent trade-off between the protection of personal privacy and the utilisation of data for commercial purposes. The employment of sophisticated algorithmic methodologies, including adaptive noise calibration and context-sensitive adjustments to the ε parameter, empowers platforms to optimise this trade-off while adhering to regulatory stipulations (Dwork & Roth, 2013; Nissim et al., 2016).

2.9 Existing Privacy Protection Tools for data subjects

The proliferation of location-based services and the increasing complexity of data collection mechanisms have driven the development of various privacy protection tools designed to protect users' sensitive location data. These tools encompass a range of approaches, from integrated features of mobile operating systems to specialized anonymisation systems based on artificial intelligence. Each of these tools utilises distinct methods to address specific aspects of location privacy risks. This chapter provides a comprehensive analysis of existing privacy protection solutions and examines their technical implementation, performance characteristics, and limitations in relation to the protection of text-based location data. The analysis is conducted with the objective of identifying current opportunities and gaps in the field of data protection. This process serves to establish a foundation for understanding the necessity for specialised solutions in the context of urban mobility.

2.9.1 Mobile and Browser-Based Privacy Tools

Mobile and browser-based privacy tools represent the most accessible category of privacy solutions for end users. These systems are typically integrated directly into operating systems or web browsers, thereby providing real-time protection during digital interactions. The primary function of these tools is to manage permissions, block trackers, and control consent mechanisms, with the objective of preventing unauthorised access to location data embedded in text communications. The following analysis examines four well-known tools in this category and evaluates their algorithmic approaches, development frameworks, and effectiveness in protecting location information in text form on various digital platforms and in different usage scenarios.

Safety Check

Apple's Safety Check, a feature integrated into iOS that is designed to manage textual location information sharing across applications and services, serves as a prime example of a sophisticated consent and access control system. The fundamental algorithm employed here is permission auditing, which proactively observes app permissions and user interactions with messaging and note-taking applications, with the aim of identifying any unauthorised sharing of location data embedded in text. The development is underpinned by Apple's Core Location API, a technology that facilitates secure permission management and data encryption at the OS level. From a scientific perspective, it aligns with privacy-enhancing technologies that focus on consent management and local permission tracking, with evaluations highlighting the effectiveness of the tool in emergency scenarios where immediate data withdrawal is required (Apple Inc, 2025).

MyPermissions Privacy Cleaner

The MyPermissions Privacy Cleaner application is a mobile software that has been developed for the specific purpose of monitoring and minimising location data exposure from textual content stored across a range of apps and cloud services. The app utilizes a permission-tracking algorithm that perpetually scans installed apps and registry files. These are then compared against a proprietary list of services known to access location data. The app's scientific underpinnings are rooted in permission minimization frameworks, employing a dual-layer detection model that integrates static analysis of app permissions with dynamic detection based on runtime activity. Utilising the Android Permission Manager API enhances transparency, and users have the ability to modify permissions in real-time through an interactive dashboard. Despite its strengths in terms of data transparency, however, the literature indicates limitations in terms of coverage when handling third-party software development kits (SDKs) embedded in applications (Torre et al., 2023).

DuckDuckGo Privacy Browser

The DuckDuckGo Privacy Browser is a web browser which has been engineered to limit an individual's geolocation data sharing during web browsing. The software utilises a dual-pronged approach, employing a tracker-blocking algorithm that operates on a filter basis. This algorithm draws from two distinct sources: the Disconnect blocklist and a bespoke heuristic analysis module. It has been designed to prevent the leakage of location-specific data when users interact with web forms and search queries by actively scanning HTTP requests and blocking those which contain geographic identifiers. The browser is built using Chromium and utilises WebAssembly and JavaScript to integrate privacy-focused browser extensions. From a scientific perspective, the system aligns with existing literature on browser fingerprinting and tracker suppression. However, studies have demonstrated limitations when handling embedded trackers within personalised search platforms (Englehardt & Narayanan, 2016).

Mozilla Firefox Enhanced Tracking Protection

The Firefox Enhanced Tracking Protection (ETP) module is designed to offer robust solutions for those concerned about the privacy implications of their online activity. One such concern pertains to the extraction of textual location data during web interactions. The ETP module employs a sophisticated blocking system that integrates Disconnect's tracking database and a real-time behavioural analysis algorithm, thus ensuring comprehensive protection. The mechanism's functionality is based on an analysis of user inputs in web forms; with a focus on location references. The objective is to prevent third parties from

extracting or storing such data. The development framework utilises the Mozilla Gecko engine, with the tracker database being drawn from open privacy datasets. Empirical research has validated the effectiveness of ETP in preventing data leakage during different browsing sessions (Roesner & Kohno, 2013).

2.9.2 AI-Based Location Anonymisation Systems

The utilisation of artificial intelligence in the domain of location anonymisation signifies a more sophisticated approach to safeguarding personal privacy. This approach employs natural language processing and machine learning methodologies to automatically identify, detect and anonymise location-specific information embedded within textual data. These systems utilise sophisticated algorithms, incorporating transformer-based language models and neural networks, to identify location information and substitute it with alternatives that are compliant with privacy regulations, while preserving semantic coherence. The development of such systems is driven by the complex challenge of balancing privacy protection with data usability, particularly in scenarios involving large text datasets that require automated processing capabilities.

Privacy BERT-LSTM

The Privacy BERT-LSTM model has been developed as a specialised solution for the anonymisation of textual location data contained within unstructured datasets by means of natural language processing (NLP). The tool integrates the BERT model, which is employed for the purpose of generating contextual embeddings, with LSTM networks to facilitate sequence detection and anonymisation. The transformer-based embeddings generated by the BERT model, in combination with the LSTM network's ability to detect and anonymise location-based named entities, constitutes the core functionality of the tool. These named entities are subsequently masked or replaced with placeholders, ensuring the utmost confidentiality of the underlying data. The tool has been developed using Python and TensorFlow, and has been trained on datasets such as CoNLL-2003 and OpenStreetMap. Recent empirical evaluations have demonstrated high accuracy for named entity anonymisation tasks, with precision and recall above 90% (Muralitharan & Arumugam, 2024).

Semantics-Preserved Distortion

Semantics-Preserved Distortion (SPD) is a methodology that employs a distortion algorithm for the anonymisation of textual location data. This is achieved by replacing terms with contextually similar but non-identifiable alternatives. The implementation of SPD involves the utilisation of SpaCy's Natural Language Processing (NLP) pipeline for the identification

of named entities and the generation of word embeddings. This is complemented by a cosine similarity scoring system, which is designed to ensure the preservation of semantic integrity. SPD systematically examines datasets for geographical locations and replaces the original terms with synthetic yet coherent alternatives. SPD has been developed using the Python programming language, with the objective of achieving a balance between privacy protection and data utility by preserving sentence coherence and structure. Nonetheless, it has been observed that the effectiveness of SPD is notably diminished in texts of a high level of complexity, particularly those wherein location indicators are extensively interwoven within technical discourses (Cui et al., 2018).

2.9.3 Comparative Analysis and Limitations

The evaluation of extant privacy tools necessitates a systematic comparative analysis in order to assess their relative effectiveness and to identify opportunities for the improvement of the protection of text location data. This section offers a comprehensive overview of the performance metrics, technical capabilities, and practical limitations of the tools analysed. By doing so, it aims to provide a thorough understanding of the current state of privacy technology. Through quantitative performance comparison and qualitative analysis of implementation approaches, this evaluation reveals significant gaps in existing solutions that limit their applicability to specific contexts, such as urban mobility applications. The identification of these research gaps is fundamental to comprehending the necessity for innovative approaches to location privacy protection, thus providing the rationale for the development of targeted solutions that address the particular requirements of location data sharing in contemporary digital environments.

Ethical Implications and Limitations

Whilst these tools offer efficacious mechanisms for the protection of privacy, there are a number of ethical and practical challenges associated with them. The utilisation of static blocklists, as observed in DuckDuckGo and Firefox ETP, gives rise to concerns regarding the completeness and currency of tracking databases. Furthermore, NLP-based tools, such as Privacy BERT-LSTM and SPD, are subject to limitations when dealing with multilingual contexts and non-standard datasets. Additionally, there is a risk of reduced data utility in scenarios where excessive masking occurs, thereby limiting the analytical value of datasets (Narayanan & Shmatikov, 2008).

Moreover, there is a lack of transparency in some applications, as users may not be provided with clear explanations of how privacy protection mechanisms work, particularly in permission management tools. Finally, the risk of re-identification through auxiliary datasets

remains a persistent challenge across most privacy-preserving software tools (Narayanan & Shmatikov, 2008).

Research Gaps

A comprehensive analysis of existing privacy tools reveals several critical research gaps that limit their effectiveness in protecting text location data, particularly in the context of urban mobility:

Limited contextual protection: Current tools predominantly use static filtering mechanisms and general anonymisation approaches that do not take into account the contextual nuances of location descriptions in urban environments. Although Privacy BERT-LSTM achieves 94% accuracy in general named entity recognition, it lacks specialized training for informal location references, cultural landmarks, and context-dependent location descriptions commonly used in urban mobility applications.

Insufficient real-time processing capabilities: Existing browser-based tools such as DuckDuckGo Privacy Browser and Firefox ETP focus primarily on preventing data transmission rather than real-time anonymisation of user-generated textual location content. This limitation is particularly evident in mobile messaging applications and social media platforms, where users actively share location information in textual form.

Limited multilingual and cultural adaptability: AI-based systems such as Semantics-Preserved Distortion show reduced effectiveness in multilingual contexts and struggle with culture-specific location references that are common in diverse urban environments. This limitation has a significant impact on their applicability in international urban mobility applications.

Lack of optimization of the trade-off between privacy and utility: Current solutions do not provide systematic mechanisms for users to balance privacy and information utility based on their specific use cases. The binary approach of either complete protection or no protection at all does not address scenarios where partial sharing of location information is necessary for the functionality of the service.

Gap in urban mobility: None of the tools analyzed specifically address the unique challenges of location privacy in urban mobility scenarios, where users need to share approximate location data for transportation services while protecting their precise location and movement patterns.

Summary of research gap: These identified limitations demonstrate a significant need for specialized privacy tools that enable contextual, customizable, and real-time anonymisation of textual location data and are specifically designed for urban mobility applications. This gap motivates the development of novel approaches that can bridge the divide between privacy and service functionality in location-based urban mobility systems.

The research gaps identified in existing privacy tools highlight the necessity for specialised solutions that address the unique challenges of protecting location data in the context of urban mobility. These limitations have led to the development of DeLocator, a novel mobile application that provides context-sensitive sharing of location data in urban environments. Contents provides a comprehensive overview of the methodological approach, technical implementation, and evaluation framework for this innovative privacy protection solution.

3 Methodology

This chapter presents the comprehensive research design and technical implementation strategy that was utilised in the development and evaluation of the DeLocator application. This application was developed as a novel solution for anonymising location information in texts. The methodology integrates rigorous software engineering practices with established human-computer interaction research methods to create a practical privacy tool that meets the real-world needs of users while maintaining high standards of usability and effectiveness.

The chapter opens with a comprehensive account of the software development environment and technical infrastructure that facilitated the creation of a functional Android application. The subsequent section provides a detailed exposition of the mixed-methods research approach, which integrates quantitative usability metrics with qualitative user feedback to facilitate a comprehensive evaluation of both technical performance and user acceptance. The technical implementation has been documented in great detail, including the novel substitutive anonymisation algorithm, geographical selection parameters, and integration strategies for external services such as geocoding and point-of-interest retrieval.

The evaluation methodology encompasses a systematic assessment of privacy features and an empirical evaluation of user experience through structured usability testing with expert and non-expert participants. The chapter concludes with a detailed presentation of the evaluation results, which provide empirical evidence of the effectiveness of the proposed approach and form the basis for the subsequent analysis and conclusions.

3.1 Software Development Environment

The Python programming language was utilised in the development of an Android application based on the Kivy framework. This application was designed to anonymise textual location information. The PyCharm Community Edition 2025.1 integrated development environment was utilised throughout the implementation process for the purposes of code editing, debugging and virtual environment management.

The application build process was performed in the Windows Subsystem for Linux (WSL) using Ubuntu 22.04.5 LTS. The utilisation of Buildozer in conjunction with this configuration facilitated the packaging of the application in an Android-compatible format. The Android Debug Bridge (ADB) was utilised to install and test the application on an Android device.

In accordance with the open access (OA) principle, the complete source code of the developed application is made publicly available. The corresponding GitHub repository was created to enable full reproducibility of the practical part of this study and to serve as a foundation for future research. The repository can be accessed here.

3.1.1 .Python

Python is an interpreted, general-purpose programming language that was developed by Guido van Rossum in 1991. The programming language under discussion is characterised by a clear syntax, dynamic typing and automatic memory management. Furthermore, it supports several programming paradigms, including object-oriented, functional and imperative approaches. Python's extensive standard library and active open-source community have contributed to its adoption in numerous application areas, including software development, data analysis, artificial intelligence and scientific computing (Python, 2025; van Rossum & Boer, 1991).

In the context of this thesis, Python was utilised as the primary programming language for the development of the mobile application. The implementation of logic pertaining to the anonymisation process, in conjunction with the graphical user interface (GUI) and interaction processing, was accomplished through the utilisation of the Python programming language. The language's compatibility with Kivy and its support for multiple platforms rendered it a suitable choice for Android development.

3.1.2 Kivy Framework

Kivy is a complimentary, open-source framework for the development of cross-platform applications with graphical user interfaces (GUIs) written in the Python programming language. It facilitates the development of interactive applications for a range of operating systems, including Windows, macOS, Linux, Android and iOS, utilising a standardised code base. Kivy is predicated on an event-based architecture model and provides support for multi-touch technology, rendering it particularly suitable for the development of mobile and touch-based applications (Kivy, 2025).

In this thesis, Kivy was utilised to facilitate the creation of the application's interface, to manage user input, and to define the application's structure and visual layout. The application's anonymisation logic was integrated directly into the Kivy event loop to enable real-time processing of user input. The cross-platform capabilities of the framework enabled local testing on desktop systems prior to deployment to Android.

3.1.3 PyCharm Community Edition 2025.1

PyCharm Community Edition is an integrated development environment (IDE) for the Python programming language that has been developed by the company JetBrains. The software in question provides support for essential functions, including syntax highlighting, code completion, refactoring, integrated version control and a debugging tool. The Community Edition is available at no cost and is open source, making it particularly well-suited for the development of pure Python applications. The software in question offers a user-friendly interface and tools for the efficient management of projects and virtual environments. The extensibility and integration of PyCharm with common development tools has led to its widespread use in software development, data analysis and scientific programming (JetBrains, 2025).

The primary development environment employed in this study was PyCharm. The process under discussion had a number of beneficial effects. Firstly, it facilitated the structuring of the code base. Secondly, it facilitated the testing of application modules. Thirdly, it facilitated the debugging of errors during runtime.

3.1.4 Ubuntu 22.04.5 LTS

Ubuntu 22.04.5 LTS (Long Term Support) is a stable, long-term supported version of the Debian-based Linux distribution Ubuntu, which is designed for use in production environments. Version 22.04.5 constitutes the fifth point release in the Jammy Jellyfish series, encompassing cumulative updates and enhancements to security protocols. It is noteworthy that this version is also available in the Microsoft Store for utilisation within the Windows Subsystem for Linux (WSL) (Ubuntu, 2025).

As a constituent element of the WSL, Ubuntu 22.04.5 LTS facilitates the execution of a complete Linux userland environment within the confines of the Windows operating system, thereby obviating the necessity for either a virtual machine or a dual-boot configuration. It facilitates the execution of native Linux commands, shell scripts and development tools (e.g. gcc, make, apt, python, git) directly on the Windows operating system. The version, which can be installed via the Microsoft Store, thus offers seamless integration into the Windows file system and enables cross-platform development workflows. In this thesis, Ubuntu via WSL was utilised as the build environment for the purpose of compiling the Kivy application into an Android package (.apk). The application was utilised for the purpose of installing and configuring Buildozer, in addition to executing terminal-based commands for the management of packages (for example, via APT) and the interaction with Android devices (Ubuntu, 2025).

3.1.5 Buildozer

Buildozer is an open-source tool that is primarily utilised for the automation of build processes in the domain of software development. The objective of its development was to facilitate the creation of software projects for different platforms by defining the build process in a standardised configuration. Buildozer facilitates the process of code compilation for multiple platforms, including Android, iOS, Windows, Linux and macOS, through the utilisation of a unified build script (Buildozer, 2025).

In this work the tool was particularly useful in the context of mobile application development, as it automates the integration of various dependencies and the creation of the associated packages for the respective platforms. Buildozer employs a configuration file, in which developers delineate the pertinent parameters and dependencies, thereby facilitating the execution of the build process with minimal effort and on multiple target systems. The Buildozer tool was utilised within the Ubuntu/WSL environment to construct the application and subsequently deploy it to a connected Android device.

3.1.6 Android Debug Bridge

Android Debug Bridge (adb) is a versatile command line tool that facilitates communication with a device. The ADB command can be utilised to execute a variety of device actions, including the installation and debugging of applications. ADB facilitates access to a Unix shell, which in turn enables the execution of a range of commands on a device (Android Developers, 2025). The program under consideration is of a client-server nature and is comprised of three distinct components:

- **A client** that is defined as the entity that initiates commands. The client is executed on the client's development computer. It is possible to initiate communication with a client by utilising the command line terminal and entering an ADB command.

- **A daemon (adbd)** which refers to a program that executes commands on a device. The daemon is executed on each device as a background process.

- **A server** that is responsible for the management of communication between the client and the daemon. The server is executed as a background process on the development computer.

In this thesis, ADB was used to install the generated .apk file on a physical Android device for testing. Additionally, ADB was employed to monitor runtime logs (adb logcat) and

interact with the app via the Android shell, facilitating error analysis and performance testing during development iterations.

3.1.7 GitHub

GitHub is a web-based platform for version management and collaborative software development based on the distributed version control system Git. The software in question facilitates the management of source code projects, the tracking of alterations, and collaborative work on these projects. GitHub provides a range of functions, including repositories, branches, pull requests, issue tracking and integrated CI/CD workflows. In addition to the management of source code, GitHub also fosters open exchange in the field of software development through the facilitation of open-source projects and the maintenance of an active developer community. The platform finds application in both professional software development and scientific projects, with the purpose of facilitating traceable and reproducible code management (GitHub, 2025).

In this thesis, the software development platform GitHub was utilised for the purpose of hosting the application's source code and to document the developmental process. A private repository was established for the purpose of storing various iterations of the project, facilitating the creation of backup copies, and ensuring the reproducibility of the work presented in this thesis. It is evident that GitHub has facilitated the seamless integration of version control functionality directly into PyCharm.

3.2 Research Design

This chapter outlines the methodological research design that underpins the development and evaluation of the DeLocator application. The framework under discussion combines technical development with a systematic scientific evaluation. Firstly, the methodological framework is presented, which forms the scientific-theoretical background of the work and explains its positioning in the context of design science research. The subsequent section will describe the mixed-methods approach that has been developed to address the complex research questions posed. This approach combines technical development methods with qualitative evaluation procedures. The subsequent section presents a detailed development and evaluation strategy, delineating a structured process that encompasses requirements analysis and the evaluation of the completed application. This three-stage approach facilitates a comprehensive understanding of the methodological procedure and ensures that all aspects of the research questions are adequately addressed.

3.2.1 Methodological Framework

The present work is informed by a design science research approach following Hevner et al. (2004), which integrates the development of a technological solution to address a tangible problem with a systematic scientific evaluation. Specifically, the development of a software application for anonymising location data is undertaken, and its effectiveness and user-friendliness is subsequently evaluated. This approach is especially well-suited to practice-oriented research problems in the field of information systems, where technological artefacts play a central role (Österle et al., 2011).

The methodological framework of this work combines the theoretical findings from the literature review with practical software engineering development methods. The underlying structure is based on an iterative process comprising the following phases: (1) identification of the problem and its relevance, (2) definition of solution goals, (3) design and development of the artifact, (4) demonstration of functionality, (5) evaluation, and (6) communication of results (Österle et al., 2011).

In this context, a construction-oriented research approach is pursued, whereby the research process is characterised by the construction of the artefact – in this case, the DeLocator app. The construction of scientific knowledge is not merely a means to an end, but rather an essential component of the scientific knowledge process (Österle et al., 2011). The development of the application provides insights into the practical feasibility of theoretical concepts of location data protection.

A fundamental component of the methodological framework under consideration is the methodological combination of:

1. technical methods of software development for implementing the functionality of the application.
2. qualitative methods for evaluating user-friendliness.
3. personal evaluation of the algorithm that has been implemented, with this evaluation informed by theoretical principles.

This combination enables the capture of both the technical functionality and the user-centred aspects of the application. The primary objective of this study is to establish a connection between theoretical data protection concepts and practical solutions that are applicable to end users.

The methodological orientation explicitly considers the ethical dimensions of the research project, particularly with regard to data protection and privacy. This is reflected in the special

attention paid to the balance between data protection and user-friendliness, a central challenge addressed in the first research question.

3.2.2 Mixed-Methods Approach

The present work employs a mixed-methods approach, integrating technical development methods and qualitative evaluation methods. The decision to adopt this integrated approach was made with the objective of addressing the multifaceted nature of the research questions, which encompass both technical and user-centred dimensions.

The technical development of the DeLocator app was approached via a prototype-based development approach. This process facilitates the iterative refinement of the system, informed by continuous feedback and technological insights (Sommerville, 2016). The development process was based on agile principles, in particular the incremental implementation of functionalities and regular reviews of technical feasibility. The selection of this methodology was made with the intention of ensuring the capacity to respond with agility to challenges encountered in the integration of various APIs (Nominatim, Overpass) and in the implementation of the anonymisation algorithm.

The qualitative evaluation was conducted using a structured questionnaire comprising both open and closed questions, which was completed by six participants. The decision to employ a qualitative approach with a limited number of participants was predicated on the realisation that in-depth insights into user behaviour are of greater importance for the evaluation of usability and user experience than statistical representativeness (Nielsen, 2000). The selection of participants was deliberate, with three individuals selected for their expertise in Geographic Information Systems (GIS) and three selected without expertise in this field. This enables a targeted comparison between experts and Non-Experts. This comparison is of particular pertinence when assessing the user-friendliness of a technically complex application intended for widespread use.

Quantitative elements were integrated into the qualitative survey in the form of Likert scale ratings for specific tasks. These tools facilitate a standardised evaluation of the user-friendliness of various functions, thereby providing an objective foundation for comparative analysis of different users and functionalities. The integration of qualitative and quantitative elements in the evaluation is underpinned by the principle of complementarity (Creswell & Creswell, 2017). This principle asserts that the combination of quantifiable usability aspects and in-depth insights into user perceptions and preferences can be achieved.

The methodological approaches were linked to the research questions in the following manner:

3. The first research question regarding the usability requirements for user-friendly location protection software is primarily addressed by the qualitative survey, which provides insights into user requirements and expectations.

4. The second research question regarding the prevention of re-identification is addressed by the technical implementation of the anonymisation algorithm, in conjunction with a personal evaluation of its conceptual strengths and limitations. This evaluation is founded upon the theoretical underpinnings delineated within the extant literature, eschewing the implementation of a formal technical analysis.

5. The third research question, which concerns the dependence of the level of protection on environmental factors, is investigated through a combination of personal assessment and qualitative user evaluation.

This integrated mixed-methods approach enables a comprehensive understanding of the research problem by taking into account technical, user-centred and contextual aspects of location privacy.

3.2.3 Development and Evaluation Strategy

The development and evaluation strategy of this thesis follows a structured process that ensures the systematic construction and evaluation of the DeLocator app. The process was meticulously engineered to comprehensively address the three research questions and to optimise both the technical functionality and the user-friendliness of the application.

Development Strategy

The DeLocator application was developed in five distinct phases:

1. **Requirement Definition:** The functional and non-functional requirements were defined on the basis of a review of the extant literature. It is evident that particular attention was paid to achieving a balance between data protection and user-friendliness. The requirements encompassed the conceptual anonymisation of location data, the development of an intuitive user interface, and the consideration of established data protection principles.

2. **Conception of the anonymisation method:** An anonymisation concept was developed based on the substitution of private locations with nearby public locations. The conceptual basis for this was formed by various anonymisation techniques described in the literature, whereby a non-deterministic approach was

chosen in order to reduce predictability. The emphasis in this instance was on the conceptualisation of the design rather than on a formal analysis of its effectiveness.

3. **Prototype Development:** The application was developed utilising the Python programming language and the Kivy framework, with the objective of ensuring platform independence and portability. The development process was characterised by an incremental approach, whereby a functional prototype encompassing the primary functionalities (address input, geocoding, anonymisation and map visualisation) was initially implemented.

4. **Integration of external services:** The Nominatim API was integrated for geocoding, while the Overpass API was used for querying nearby points of interest (POIs). The integration of these services necessitated particular consideration with regard to error handling and data processing.

5. **Iterative refinement**: It is evident that, in accordance with informal interim tests, enhancements were consistently implemented to the user interface and functionality. A particular emphasis was placed on the optimisation of map visualisation, the enhancement of response time for API requests, and the implementation of consistent error handling.

Evaluation Strategy

The DeLocator app was evaluated in a multi-stage process:

1. **Task-based usability evaluation:** The user-friendliness of the application was evaluated on the basis of four defined tasks, which encompass the core functionalities of the application.

    - Entering and anonymising an address
    - Copying the anonymised result
    - Saving an anonymised location
    - Managing saved locations

    The evaluation of these tasks was conducted through the utilisation of a 5-point Likert scale, ranging from 1 (designated as 'very difficult') to 5 (denoting 'very easy'), thus facilitating a quantitative assessment of usability. Furthermore, participants were given the opportunity to provide optional commentary on each task, thus facilitating the acquisition of qualitative insights into specific usability aspects.

2. **Comparative Analysis:** The reactions and evaluations of GIS-Experts and Non-Experts were compared in order to determine the extent to which prior knowledge in the field of geographical information systems influences the perception and use of the app. This comparison is of particular pertinence when assessing the accessibility of the app for diverse user groups.

3. **Personal evaluation of the algorithm:** In lieu of a formal technical analysis, a personal evaluation of the implemented anonymisation algorithm is conducted. The present study is founded on the theoretical foundations of location data anonymisation as outlined in the extant literature and reflects on the conceptual strengths and potential limitations of the chosen approach. This assessment is concerned with aspects of the second research question (2.2), which pertains to the prevention of re-identification.

4. **Analysis of user satisfaction and usability:** The questionnaire comprised a series of in-depth qualitative enquiries concerning diverse aspects of the application, thus yielding significant insights into its usability and user satisfaction:

    - The initial impression and comprehension of the application's purpose are paramount. The participants were invited to articulate the purpose of the application in a concise, one-to-two-sentence statement. This inquiry was designed to ascertain whether the notion of anonymising location data is conceptually graspable.

    - The following aspects of the visual design and user interface are to be considered in terms of evaluation: firstly, the application of colour, typography and icons, and secondly, the overall layout.

    - The identification of the tasks that are the simplest and most complex is imperative for the determination of the strengths and weaknesses of the user interface.

    - Understanding labels and icons to identify potential ambiguities in the user interface

    - Perception of error messages and feedback mechanisms: Evaluation of the comprehensibility and helpfulness of system feedback

    - Preferences for different types of feedback (e.g. acoustic signals, visual highlighting)

5. **Overall assessment and context of use analysis**: The final section of the questionnaire was designed to facilitate a comprehensive evaluation of the app, with a focus on its general utility and the various contexts in which it might be employed:

    - Positive aspects of the app from the user's perspective
    - Additional functions requested to improve the app
    - Willingness to recommend and the reasons behind it
    - Own intended use and usage scenarios
    - Open comments on further experiences with the app

These questions provide insights into various contextual factors that influence the user experience and thus address aspects of the third research question (2.3) on the dependence of the level of protection on various characteristics, such as the characteristics of the study area or user activities.

The data was collected using a qualitative questionnaire with seven main categories and a total of around 15 questions, which included both structured Likert scale ratings for the specific tasks and numerous open questions. The analysis combines the quantitative evaluation of the rating scales with a qualitative content analysis of the open-ended responses.

This comprehensive evaluation strategy facilitates a thorough investigation of the research questions concerning the usability and context of use of the developed solution. The employment of open-ended questions within a qualitative framework facilitates a more profound comprehension of the user perspective and the practical applicability of the DeLocator application in a range of scenarios.

3.3 DeLocator Application Architecture

The architecture of the DeLocator application constitutes the technical foundation for implementing the conceptual requirements for privacy-friendly location anonymisation. The development process was meticulous, ensuring that both the functional aspects of anonymisation and the intuitive user guidance were addressed. The subsequent discussion will firstly address the fundamental system requirements and design objectives. These will then be followed by a detailed description of the user interface components. The final component of the study presents the data flow architecture, which forms the technical core of the application and enables the efficient processing of location data while preserving privacy. The present chapter provides a technical illustration of the implementation of the

conceptual principles outlined in Literature Review, thereby addressing the practical aspects of the formulated research questions.

3.3.1 System Requirements and Design Goals

The DeLocator application was developed with specific system requirements and design goals in mind, with the aim of achieving an effective balance between privacy and usability. These requirements were derived from a review of the literature on location privacy, as well as from the research questions formulated.

System Requirements

Functional Requirements:

- Processing and anonymisation of textual location data.
- Address geocoding using the OpenStreetMap Nominatim API.
- Visualisation of the original and anonymised locations on an interactive map.
- Management and storage of anonymised locations in a JSON file.
- Copying anonymised location information to the clipboard.

Non-Functional Requirements:

- Platform independence (primarily Android, but also desktop environments).
- Minimal external dependencies through the use of established Python libraries.
- Reaction time of max. 3 seconds for anonymisation processes, with visual feedback mechanisms.
- Robust error handling in case of API failures or network problems.
- Data protection-compliant processing of personal data without permanent storage of original addresses.

Design Objectives

The primary design objectives of the DeLocator application encompass the following:

1. **User-friendliness:** The application's objective is to furnish an intuitive user interface that is comprehensible and straightforward to operate, even for those with no prior expertise in Geographic Information Systems (GIS). This encompasses the utilisation of clear labelling, the implementation of straightforward navigation structures, and the provision of consistent feedback. This objective directly addresses the primary research question (2.1) concerning the equilibrium between protection and service quality.

2. **Effective location data protection*:*** The implementation of an algorithm that ensures adequate concealment of individuals' locations without compromising the usability of data for its intended purpose is imperative. For this purpose, a substitution approach is employed, whereby private addresses are substituted with nearby public places (POIs).

3. **Flexibility:** The application should be designed to be usable in a range of contexts and for different user groups. The categorization of POIs (points of interest) into the following domains: restaurants, shopping, healthcare, services, transportation, and leisure, enables context-related anonymisation.

4. **Transparency:** Aiming to convey the concept of location anonymisation to users in a transparent manner, the application has been developed. This objective is realised through the presentation of both the original and anonymised locations on a map.

5. **Offline functionality:** While the application relies on online services for geocoding and POI queries, stored anonymised locations are also available offline and can be reused as predefined locations.

These requirements and design objectives served as the foundation for the architectural decisions and implementation of the DeLocator application. These decisions were continually evaluated based on their practical feasibility and theoretical requirements for location data privacy.

3.3.2 User Interface Components

The DeLocator application's user interface was constructed utilising the Kivy framework, a software framework that facilitates the uniform display of applications on Android devices and desktop platforms. The selection of Kivy was made on the basis of its platform independence and native support for interactive map visualisation through the MapView component.

1. *Start Screen:*
    - *Logo:* Central display of the app logo
    - *Generate New-Button:* Leads to the main function of address anonymisation
    - *Saved Locations-Button:* Opens the list of saved anonymised locations
    - *Info-Button:* Circular button in the upper right corner that provides information about the app

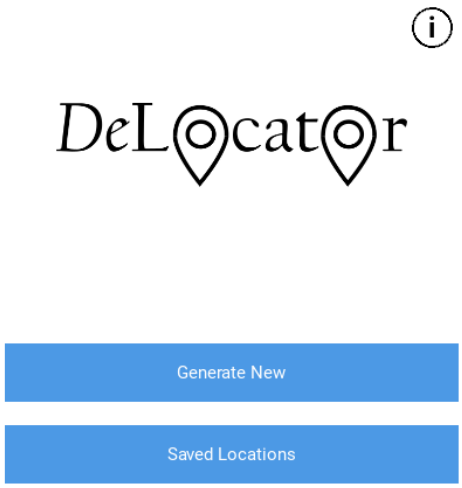

Figure 12: DeLocator Start Screen

2. *Map Screen:*
    - *Address Input Field:* Central text field for entering addresses or location descriptions.
    - *Submit Button:* The process of anonymisation is initiated, accompanied by the presentation of a loading status.
    - *Map View:* An interactive MapView component for visualizing the original and anonymised locations with different colored markers (red for anonymised, dark red for original).
    - *Map Legend:* Explains the meaning of the various markers on the map
    - *Taskbar:* Contains buttons for copying and saving anonymised location data
    - *Back Button:* Enables navigation to the start screen

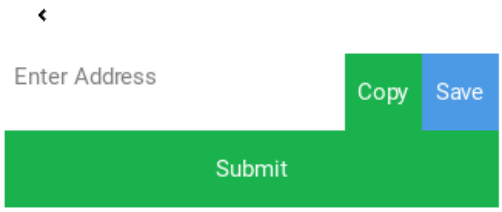

Figure 14: DeLocator Map Screen without Address

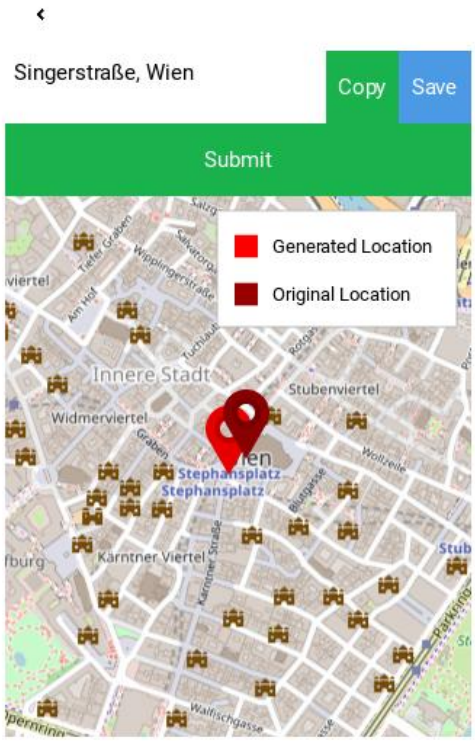

Figure 13: DeLocator Map Screen with Address

3. *Pop-ups:*
    - *Info Pop-up:* This text provides a comprehensive overview of the purpose of the application, its operational functionality, and the pertinent technical particulars. The content has been formatted using markup to enhance its readability.

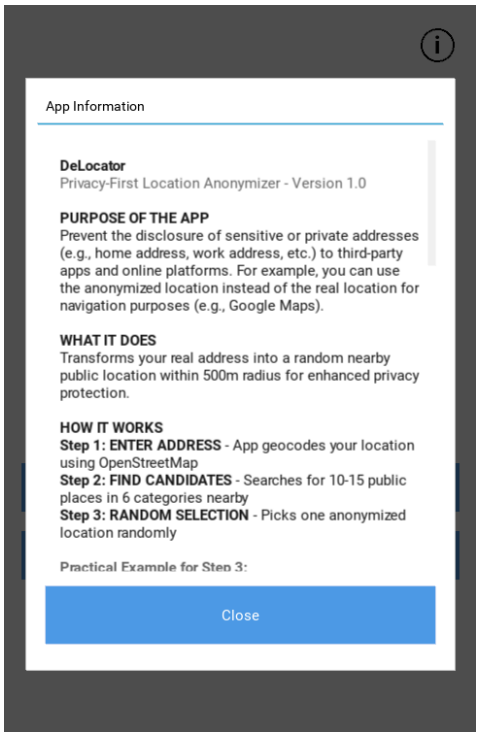

Figure 15: DeLocator Info Pop-up

- *Save Pop-up:* This function enables the user to store an anonymised location with icons representing different categories, such as *'Home', 'Work'* and *'Family'*, in addition to an optional description.

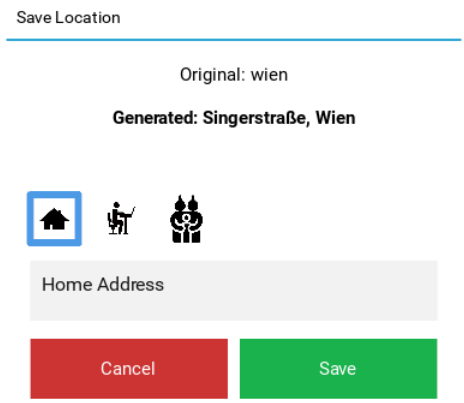

Figure 16: DeLocator Save Location Pop-up

- *Show saved locations Pop-up:* This functionality enables the user to access a list of previously saved locations, accompanied by the options to duplicate or erase these locations.

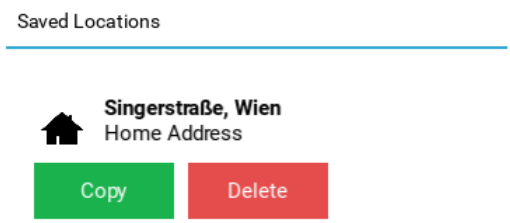

Figure 17: DeLocator Saved Locations Pop-up

- *Error Pop-ups:* A variety of context-sensitive pop-ups are utilised for the purpose of displaying error messages, including those pertaining to addresses that have not been found or API errors.

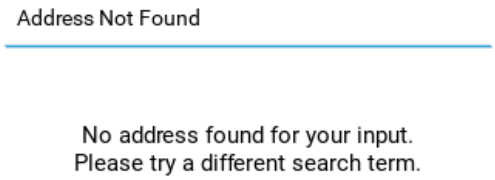

Figure 18: DeLocator Address not found Error

4. *Notification system:*
   - *Android notifications:* On devices utilising the Android operating system, system notifications are employed to facilitate expeditious access to saved locations.
   - *BroadcastReceiver:* Its implementation is for the purpose of managing notification actions on the Android operating system.
   - *Visual feedback:* Short-term colour alterations to buttons following the execution of confirmation actions.

UI-Design Principles

The principles outlined below were taken into account during the design process of the user interface.

- **Consistency:** An uniform colour scheme is employed, primarily white with blue and green accents, and there is consistent button design throughout the application.
- **Simplicity:** Focus on the core function of address anonymisation with minimal distraction.
- **Visual feedback:** Direct feedback on user actions is provided through colour changes and status indicators.
- **Error prevention:** Clear error messages and validation mechanisms for user input.
- **Self-explanatory components:** Clear labelling and informational texts elucidating functionality.

The user interface has been deliberately kept simple to facilitate usage by less technically proficient users and to prioritise the core function of location anonymisation.

3.3.3 Data Flow Architecture

DeLocator employs a pragmatic data flow architecture, underpinned by direct interaction between the user interface and external services. The data flow has been optimised with a view to efficiency and data protection.

Core components of the data flow architecture

1. *Input Processing:*
   - Direct capture of address input via the TextInput widget.
   - No automatic completion or caching of input data.
   - Error management through the utilisation of dedicated pop-up components.
2. *Geocoding:*
   - Integration of the OpenStreetMap Nominatim API via the geopy library.
   - Configuration of a secure SSL context with the certifi library.
   - Performing geocoding in a separate thread to ensure the user interface remains responsive.
   - Results obtained from this process contain geographic coordinates (latitude and longitude) and address information.

3. *POI-Query and Anonymisation:*
    - Implementation of direct HTTP requests to the Overpass API.
    - A categorisation of various POI categories is provided, including restaurants, shopping, healthcare, services, transportation and leisure.
    - Fallback mechanisms and test functions for the robustness of API queries.
    - Extraction of address information from POI tags is subject to stringent quality assurance rules.
    - A random selection of a POI as a substitute location.

4. *Data Visualization:*
    - The display of both original and anonymised location data is facilitated on an interactive map via the MapView interface.
    - The utilisation of distinct marker colours facilitates visual differentiation.
    - A map legend is provided to elucidate the significance of the markers.
    - Centring of the map automatically on the relevant area.

5. *Data Persistency:*
    - Storage of anonymised locations in a local JSON file.
    - Storage structure with original address, anonymised address, description, and icon path.
    - No permanent storage of original coordinates.
    - The ability to manage saved locations, including the functions of viewing, copying and deleting these locations.

Data flow

The typical sequence of operations in the DeLocator application is as follows:

1. **Input phase:** The user is required to enter an address in the designated text field and subsequently press the Submit button.

2. **Geocoding phase:**
    - The address is then transmitted to the Nominatim API.
    - In the event of a successful geocoding process, the geographic coordinates are returned.
    - In the event of failure, an error dialog is displayed.

3. **POI-Query phase:**
    - Subsequently, a request is transmitted to the Overpass API, utilising the coordinates that have been received.

- The query function searches for public places within a 500-metre radius, categorised into defined classifications.
- In the event of a successful query, a list of POIs accompanied by complete addresses is returned.
- In the event of failure, fallback methods or error dialogs are activated.

4. **Anonymisation phase:**
   - A location is selected at random from the list of POIs.
   - This random factor increases unpredictability and thus data protection.

5. **Visualization phase:**
   - The original and anonymised locations are displayed on the provided map.
   - The anonymised address is thereby displayed in the designated text field.
   - The map view is centred on both points.

6. **Interactive phase:**
   - The user is permitted to duplicate or store the anonymised address.
   - Upon executing the save function, a dialog box pertaining to categorisation is presented to the user. The available categories include 'Home', 'Work' and 'Family'.
   - Locations saved by the user may be accessed at a subsequent point via the start screen.

Android-specific components

A number of additional features have been incorporated for Android devices:

1. **Permission requests:** A request is hereby made for permission to enable POST_NOTIFICATIONS for system notifications.

2. **Notification system:**
   - Implementation of a NotificationChannel for Android 8.0+ (API 26+).
   - Creation of action buttons in notifications for saved locations.
   - BroadcastReceiver for handling notification actions.

3. **Life-Cycle management:**
   - Registration and deregistration of the BroadcastReceiver must be initiated at the beginning and end of the application's operation.
   - The display of notifications upon transitioning to the background (on_pause).

The data flow was designed so that sensitive information (e.g. the original address) is only temporarily stored in the working memory. Persistent storage is limited to anonymised data and takes place exclusively locally on the user's device, in accordance with the principles of data minimisation and privacy protection.

This pragmatic architecture facilitates the efficient implementation of core functionality with minimal resource consumption, a feature that is of particular importance for mobile devices. The direct coupling of UI components and logic facilitates a swift response to user input and API results, while the offloading of computationally intensive operations to separate threads ensures the responsiveness of the user interface.

3.4 Location Privacy Implementation

The implementation of location privacy measures in the DeLocator application provides a practical application of the theoretical concepts discussed in the literature review. The implemented approach prioritises both effective anonymisation and practical usability, ensuring that the resulting locations retain sufficient utility for common location-based services while providing meaningful privacy protection. This section delineates the algorithmic design decisions, geographic parameters, and randomisation techniques employed in the application.

3.4.1 Anonymisation Algorithm Design

The DeLocator application uses a substitution-based anonymisation approach that replaces private locations with nearby public locations. This approach was selected after careful consideration of various anonymisation techniques discussed in the literature, including k-anonymity, differential privacy, and geographic masking.

The substitution approach is predicated on the principle that public locations inherently offer greater data protection than private addresses while maintaining geographical relevance. The algorithm is comprised of the following sequential steps:

1. The original address was geocoded using the Nominatim API in order to obtain precise geographical coordinates.
2. The query of nearby public places within a defined radius is to be executed using the Overpass API, with filtering based on predefined categories of points of interest (POIs).

3. The retrieved locations must be validated to ensure that they contain complete address information, including the street and city.
4. A non-deterministic selection procedure is applied to select a public location from the filtered set.
5. The address of the selected public location is to be returned as an anonymised replacement for the original location.

This design offers several advantages over alternative approaches. In contradistinction to perturbation techniques, which involve the addition of noise to coordinates, the substitution approach guarantees that the anonymised location constitutes a valid, real address that can be utilised effectively in navigation and location-based services. Moreover, this approach circumvents the complexity of calculating spatial obfuscation regions or generating synthetic data, which frequently necessitates substantial computing resources or can yield unrealistic results.

The implementation deliberately prioritises practicality and user understanding over formal data protection guarantees. Whilst techniques such as differential privacy offer mathematical data protection assurances, they frequently result in locations that are challenging for users to conceptualise or apply in practical scenarios. The substitution approach engenders results that are immediately intelligible to users and directly applicable in everyday use cases.

The algorithm follows the principles of privacy by design, aiming to minimise data storage. During anonymisation, the original address data is processed only in memory. After the anonymisation is completed and the anonymised address is saved, the mapping between the original and anonymised address is stored, ensuring that the same input always yields the same anonymised output. This persistent mapping prevents attackers from narrowing down the original address by comparing repeated anonymised results.

3.4.2 Geographic Selection Parameters

The geographical parameters employed in the anonymisation process were defined with consideration for both data protection requirements and practical usability restrictions. These parameters delineate the spatial characteristics of anonymisation and exert a direct influence on the equilibrium between data protection and service quality.

The primary geographical parameter is the search radius, which has been set at 500 metres around the original location. The selection of this distance was arrived at as a compromise between competing objectives:

1. **Data protection:** It is imperative that the radius is sufficiently large in order to ensure a substantial separation between the original and anonymised locations, thereby rendering re-identification attempts more arduous. In the context of urban environments, a radius of less than 100 metres is likely to provide inadequate data protection.

2. **Service Usability:** Conversely, an excessively large radius would serve to reduce the usefulness of the anonymised location for services such as food delivery or transportation. It is reasonable to hypothesise that an anonymised location radius greater than one kilometre would be impractical for many common use cases.

3. **POI availability:** The radius must include a sufficient number of potential public locations to enable effective anonymisation. The 500-metre radius is generally considered to ensure that, even in less densely populated areas, there are still some public POIs to choose from.

4. **Contextual relevance:** The anonymised location should retain a certain contextual relevance to the original location and remain within the same general neighbourhood or district in order to maintain service quality.

The second important parameter is the categorization of the POIs used for anonymisation. The application employs the following categories, which were meticulously selected to ensure comprehensive coverage across diverse urban environments:

1. **Food service:** restaurants, cafés, bars, and fast-food establishments
2. **Shopping:** supermarkets, bakeries, and convenience stores
3. **Healthcare:** Pharmacies and similar healthcare-related facilities
4. **Services:** Banks, ATMs, post offices, and gas stations
5. **Transportation:** Bus stops and other public transportation hubs
6. **Recreation:** parks and personal service facilities

The selection of these categories was driven by the objective of providing a diverse array of potential anonymisation targets that are widely present in both urban and suburban environments. The distribution of these POI types varies in different urban contexts, but their ubiquity ensures that suitable anonymisation candidates can be identified in most populated areas.

The application implements rigorous validation rules for potential anonymisation targets, requiring each selected POI to possess complete address information, inclusive of street and city data. This validation process is integral in ensuring that the anonymised location

remains practical for utilisation in navigation and location services, which typically necessitate structured address information.

3.4.3 Randomization Techniques

The randomisation component of the DeLocator application's anonymisation algorithm is fundamental to its privacy strategy. The application employs non-deterministic selection to enhance privacy and reduce predictability, thereby augmenting resilience to re-identification attempts.

The core of the randomisation approach utilises the Python random module to execute a random selection from the set of validated POIs within the designated search radius. The following methodology is employed in this implementation:

$$selected\_place, (lon, lat) = random.choice(amenities)$$

This non-deterministic approach offers several important data protection advantages:

1. **Unpredictability:** It is important to note that each anonymisation process produces different results, even for the same input address. This unpredictability is a significant obstacle for attackers seeking to establish a consistent correlation between original and anonymised locations.

2. **Distribution across categories:** The random selection process is implemented across all valid POIs, irrespective of category, thereby ensuring a diverse distribution of possible anonymised locations. This diversity complicates the identification of patterns within the anonymisation process.

3. **Temporal variance:** The outcomes demonstrate variability not only between users, but also over time for a given user, thus hindering the establishment of a consistent anonymisation pattern that could be utilised for re-identification.

4. **Credible deniability:** The process of randomisation has been demonstrated to engender a form of credible deniability, insofar as the selection of a particular POI from the available set is rendered arbitrary and cannot be reliably linked to the original location beyond the general area defined by the search radius.

The randomisation process is implemented with equal probability distribution across all valid POIs, without weighting based on distance or POI type. The selection of this even distribution was made with the objective of maximising unpredictability and avoiding the introduction of biases that could be utilised to narrow down the original location.

In the context of stored locations, the application enables users to maintain consistent anonymised addresses for specific contexts (e.g. home, work, family) while leveraging randomisation for new addresses. This hybrid approach is predicated on striking a balance between the privacy benefits of randomisation and the practical need for consistency in certain scenarios.

The implementation does not utilise more complex randomisation techniques, such as differential privacy mechanisms or geometric perturbations. The decision to prioritise simplicity, computational efficiency, and real-world usability of the resulting locations was made with the intention of optimising the process. The random selection of points of interest (POI) from a validated real-world database offers a pragmatic balance between theoretical data protection and practical application.

3.5 Technical Services Integration

The DeLocator application integrates various technical services and components with a view to ensuring effective location anonymisation. The technical integration in question encompasses the implementation of geocoding services, the querying of points of interest (POIs), the visualization of map data, and local storage mechanisms. The subsequent subsections provide a comprehensive overview of the technical intricacies underlying these integrations. Collectively, these integrations facilitate the operational functionality of the application while adhering to the principles of data protection. The implementation process was meticulously executed with a pronounced emphasis on efficiency, robustness, and user-friendliness, while concurrently adhering to stringent data protection requirements.

3.5.1 Geocoding and Location Services

The DeLocator application utilises geocoding services to transform textual address entries into geographic coordinates. This functionality is intrinsic to the anonymisation process, as it forms the basis for the subsequent search for alternative public locations. The implementation utilises the Nominatim API, an open-source geocoding service that prioritises privacy.

The geocoding service has been integrated using the Python library geopy, which provides a structured interface to various geocoding services. In order to facilitate secure communication with the Nominatim API, it is necessary to configure an SSL context. The purpose of this is to ensure the integrity and confidentiality of the transmitted data:

```
ctx = ssl._create_unverified_context(cafile = certifi.where())
geopy.geocoders.options.default_ssl_context = ctx
loc = Nominatim(user_agent = "DeLocatorApp", timeout = 10)
```

The geocoding process itself is executed in a separate thread instance, with the purpose of ensuring the responsiveness of the user interface during the execution of the API request. It has been demonstrated that the implementation in question has the function of preventing the user interface from freezing during longer network requests:

```
def _perform_api_calls(self):
  """Performs the actual API calls"""
  self.original_address = self.address_input.text
  address = self.address_input.text
  # Geocoding
  location = loc.geocode(address)
  if not location:
    Clock.schedule_once(lambda dt: self.show_error_popup(
      "Address Not Found",
      "No address found for your input.\nPlease try a different search term."
    ))
    return
  print(f"Address found: {location.address}")
```

Error handling is an essential part of geocoding implementation. When geocoding requests fail (e.g., due to invalid addresses or network problems), user-friendly error messages are displayed that provide specific recommendations for action. This approach improves user-friendliness through transparent communication of error states.

A further pivotal element in the implementation of geocoding pertains to the verification of existing locations. Prior to the initiation of a new geocoding request, the application performs a check to determine whether the entered address already exists in the stored locations:

```
# Check saved locations
saved_locations = load_saved_locations()
for location in saved_locations:
  if location["original_address"] == address:
    print(f"Address already saved: {location['address']}")
    Clock.schedule_once(lambda dt: self._update_ui_with_saved_location(location, loc))
    return
```

This optimization reduces unnecessary API requests and improves response times for addresses that have already been anonymised. Concurrently, it guarantees uniform

anonymisation for recurring addresses, a feature of particular pertinence for frequently utilised locations such as "Home" or "Work".

The geocoding implementation also adheres to privacy by design principles, with a strong focus on data minimisation. During geocoding, the original address data is processed only in memory and is not stored permanently in the first place. Permanent storage only occurs if the user explicitly requests it by saving the anonymised address. This approach ensures that sensitive location data remains protected and reduces the risk of unauthorised disclosure, while still maintaining the flexibility needed for application functionality.

3.5.2 Point-of-Interest Data Retrieval

The DeLocator application's anonymisation algorithm is predicated on the querying of points of interest (POIs). Following the successful geocoding of an address, public locations in the vicinity are identified that have the potential to serve as anonymised alternatives. This functionality is implemented through the integration of OpenStreetMap's Overpass API, which enables detailed queries about geographic objects.

The implementation utilises direct HTTP requests to the Overpass API in order to ensure maximum control over query parameters and error handling. This approach was selected subsequent to the determination that the prevailing Python libraries for Overpass lacked the requisite flexibility and robustness:

```
def get_places_with_fallback(api, location, radius = 500):
    """
    DIRECT HTTP requests – bypass Overpass Python Library
    """
    import requests
    import json

    print(f"Direct HTTP request to Overpass API...")

    # Correct Overpass query for direct HTTP call
    overpass_query = f"""
[out:json][timeout:25];
(
  node(around:{radius},{location.latitude},{location.longitude})[amenity
= restaurant];
  node(around:{radius},{location.latitude},{location.longitude})[amenity = cafe];
  node(around:{radius},{location.latitude},{location.longitude})[amenity = bank];
  node(around:{radius},{location.latitude},{location.longitude})[shop = supermarket];
```

```
 node(around:{radius},{location.latitude},{location.longitude})[amenity
= pharmacy];
);
out body;
"""

  try:
    # Direct HTTP request
    url = "https://overpass-api.de/api/interpreter"
    response = requests.post(url, data = {'data': overpass_query.strip()}, timeout
= 30)

    if response.status_code != 200:
      print(f"HTTP Error: {response.status_code}")
      return []

    # Parse JSON
    data = response.json()
    elements = data.get('elements', [])
    print(f" Direct API: {len(elements)} elements found")
```

The Overpass query has been meticulously optimised to identify relevant POI categories. The categories were selected on the basis of their frequency in urban environments and their suitability as plausible alternative locations. The implemented query searches for restaurants, cafés, banks, supermarkets and pharmacies within a defined radius of 500 metres around the original location.

A fundamental component of POI queries pertains to the validation of the returned data. The application imposes stringent requirements on the quality of POI data, particularly with regard to the completeness of address information:

```
def extract_address_from_tags(tags):
  """
  REVISED VERSION: Extracts ONLY real addresses
  Discards all names, brands, operators and generic types
  """
  # Get relevant address parts
  street = tags.get("addr:street", "").strip()
  housenumber = tags.get("addr:housenumber", "").strip()
  postcode = tags.get("addr:postcode", "").strip()
  city = tags.get("addr:city", "").strip()

  # STRICT RULE: Only if street AND city are present
  if street and city:
```

```
    address_parts = []

    # Street + house number (if available)
    if housenumber:
        address_parts.append(f"{street} {housenumber}")
    else:
        address_parts.append(street)

    # Postal code + city
    if postcode:
        address_parts.append(f"{postcode} {city}")
    else:
        address_parts.append(city)

    result = ", ".join(address_parts)
    print(f"Real address accepted: {result}")
    return result

# DISCARD EVERYTHING ELSE – even if name, brand, operator are present
print(f"Discarded (no complete address): {tags.get('name', 'Unknown')}")
return None
```

This rigorous validation process guarantees that only POIs with complete address information (street and city) are considered as potential anonymisation targets. This decision was made to ensure the practical usability of the anonymised addresses, as many location-based services require structured address information.

The implementation of the system incorporates a number of fallback mechanisms to address potential API errors or network issues:

1. A simplified test function is utilised to verify the fundamental operational capacity of the Overpass API.
2. In the event of empty results or API errors, user-friendly error messages are displayed.
3. The timeout parameters have been optimised to ensure an equilibrium between reliability and responsiveness.

These mechanisms are of particular importance in the context of mobile applications, which are frequently utilised in environments characterised by unstable network connections.

3.5.3 Map Visualization Components

The DeLocator application incorporates a map visualisation component, which is instrumental in providing users with an intuitive representation of the anonymisation process. The visual representation of both the original and anonymised locations on an interactive map has been demonstrated to increase transparency and improve the user's understanding of the data protection achieved.

The map visualisation is implemented by integrating the MapView component from the Kivy Garden package. This component provides a cross-platform map view with support for markers and interactive features:

$$self.mapview = MapView(zoom = 15, lat = 0, lon = 0, size\_hint = [1, 0.8])$$
$$self.marker\_new\_address = MapMarker(lat = 0, lon = 0, color = (1, 0, 0, 1))$$
$$self.mapview.add\_widget(self.marker\_new\_address)$$
$$self.marker\_old\_address = MapMarker(lat = 0, lon = 0, color = (0.6, 0.0, 0.0, 1.0))$$
$$self.mapview.add\_widget(self.marker\_old\_address)$$
$$self.mapview.opacity = 0$$

A significant component of map visualisation pertains to the utilisation of colour differentiation between the original and anonymised locations. The original location is denoted by a dark red marker, while the anonymised location is indicated by a light red marker. This visual distinction facilitates immediate recognition of the anonymisation effect by users and comprehension of the spatial relationship between the two locations.

In order to enhance the user-friendliness of the system, a map legend has been incorporated to provide explanatory information regarding the different markers:

$$class\ MapLegend(BoxLayout):$$
$$\quad def\ \_init\_(self, ** kwargs):$$
$$\quad\quad super().\_init\_(** kwargs)$$
$$\quad\quad self.orientation = 'vertical'$$
$$\quad\quad self.size\_hint = (None, None)$$
$$\quad\quad self.size = (dp(165), dp(80))$$
$$\quad\quad self.spacing = dp(8)$$
$$\quad\quad self.padding = (dp(12), dp(10), dp(12), dp(10))$$

$$\quad\quad with\ self.canvas.before:$$
$$\quad\quad\quad Color(1, 1, 1, 1)$$
$$\quad\quad\quad self.bg\_rect = Rectangle(size = self.size, pos = self.pos)$$

$$\quad\quad self.bind(size = self.\_update\_bg, pos = self.\_update\_bg)$$

The MapLegend component is dynamically shown and hidden when the map is displayed or hidden. This is achieved through the implementation of a seamless animation sequence, which has been shown to enhance the user experience:

```
def show_legend(self):
    """Shows the legend with animation and adds it to MapView"""
    if self.map_legend not in self.mapview.children:
        self.mapview.add_widget(self.map_legend)

    def position_legend(*args):
        self.map_legend.pos = (
            self.mapview.right - self.map_legend.width - dp(10),
            self.mapview.top - self.map_legend.height - dp(10)
        )

    self.mapview.bind(pos=position_legend, size=position_legend)
    position_legend()

    def animate_legend(dt):
        if self.map_legend.opacity < 1:
            self.map_legend.opacity = min(1, self.map_legend.opacity + 0.1)
            Clock.schedule_once(animate_legend, 0.05)

    animate_legend(0)
    print(" Map Legend displayed")
```

Another significant component of map visualisation pertains to the automatic centring of maps. Following the process of anonymisation, the map view is automatically positioned so that both the original and anonymised locations are visible:

```
def _update_map_markers(self, new_lat, new_lon, old_lat, old_lon):
    """Update Map Marker positions"""
    self.marker_new_address.lat = new_lat
    self.marker_new_address.lon = new_lon
    self.marker_old_address.lat = old_lat
    self.marker_old_address.lon = old_lon

    # Center map
    center_lat = (new_lat + old_lat) / 2
    center_lon = (new_lon + old_lon) / 2
    self.mapview.center_on(center_lat, center_lon)

    # Show map
    self.mapview.opacity = 1
```

```
    # Show legend
    Clock.schedule_once(lambda dt: self.show_legend(), 0.5)
```

The cartographic representation was implemented with particular attention to its ease of use on mobile devices. The dimensions of the controls, the positioning of the legend, and the interaction mechanisms have been optimised for utilisation on touchscreens.

3.5.4 Local Data Storage Mechanisms

The local storage mechanisms of the DeLocator application were developed with particular attention to data protection principles. The application employs a minimalist and transparent approach to data storage, implementing the principle of data minimisation while ensuring practical usability for recurring use cases.

The primary storage format is JSON, which facilitates the structured, efficient, and accessible storage of location data. The implementation incorporates rudimentary functions for the loading and saving of anonymised locations:

```
def load_saved_locations():
  try:
    with open("saved_locations.json","r") as file:
      saved_locations = json.load(file)
  except FileNotFoundError:
    saved_locations = []
  return saved_locations

def save_saved_locations(saved_locations):
  with open("saved_locations.json","w") as file:
    json.dump(saved_locations, file)
```

The stored location data adheres to a defined structure that encompasses both the anonymised and original addresses, along with supplementary metadata such as a description and icon path. This configuration facilitates the subsequent reuse and management of the stored locations:

```
  def save_new_location(self, saved_locations):
    description = self.description_input.text
    saved_locations.append({
      "original_address": self.original_address,
      "address": self.address,
      "description": description,
      "icon": self.selected_icon
```

```
})
save_saved_locations(saved_locations)
```

A significant component of the storage mechanism pertains to the categorization of stored locations. The application enables users to store anonymised locations with predefined icons representing different contexts, such as "Home," "Work," or "Family." This categorization enhances usability by facilitating expeditious identification and selection of frequently used anonymised locations:

```
icons = {
  "Home": "home.png",
  "Work": "work.png",
  "Family": "family.png"
}

for name, icon_path in icons.items():
  btn = Button(
    size_hint = (None, None),
    size = (dp(40), dp(40)),
    background_normal = icon_path,
    background_down = icon_path,
    background_color = (1, 1, 1, 1),
    border = (5, 5, 5, 5)
  )
  btn.bind(on_release = lambda btn, path = icon_path: self.select_icon(btn, path))
  self.icon_buttons[icon_path] = btn
  icon_layout.add_widget(btn)
```

The application is equipped with safeguards that prevent the inadvertent overwriting of stored locations. In the event that a user endeavours to save a new location with an icon that has previously been used, a confirmation dialog box is displayed:

```
def ask_overwrite(self, existing_location, saved_locations):
  def overwrite(instance):
    saved_locations.remove(existing_location)
    self.save_new_location(saved_locations)
    confirm_popup.dismiss()

  def cancel(instance):
    confirm_popup.dismiss()
```

For Android devices, enhanced integration with the notification system has been implemented. Saved locations are made available in notifications, enabling swift access to frequently used anonymised addresses:

```
def send_notification(self):
  try:
    saved_locations = load_saved_locations()
    if not saved_locations:
      print("No saved locations found. No notification sent.")
      return

    for index, location in enumerate(saved_locations):
      address = location["address"]
      icon_path = location.get("icon","")
      icon_text = icon_mapping.get(icon_path,"Location")
```

The notifications are linked to a BroadcastReceiver, which facilitates the transfer of addresses to the clipboard without the necessity of opening the application. This integration enhances the user experience by facilitating expeditious access to stored locations from the Android notification area.

The local storage mechanism has been deliberately kept simple in order to maximize data security and minimize the attack surface. The application does not utilise online synchronisation or cloud storage, thereby ensuring that sensitive location data remains exclusively on the user's device. This decision aligns with the principle of data minimisation, thereby reducing the risk of unauthorised access to sensitive location information.

3.6 Evaluation Methodology

The development of a user-friendly application for the protection of location data requires a profound understanding of usability requirements, facilitating a balance between effective data protection and the quality of the requested services. The present chapter directly addresses Research Question 2.1: *"What are the usability requirements for a user-friendly location protection software application that balances both the protection and the quality of a requested service?"* A systematic evaluation will be conducted to ascertain the usability factors that are pivotal for the acceptance and effective utilisation of location protection applications.

3.6.1 Usability Assessment Framework

The development of privacy protection applications poses a particular challenge in terms of user-friendliness, as it necessitates the harmonisation of two potentially competing objectives: the maximisation of data protection, whilst ensuring that service quality is not unduly compromised. As posited by Iachello and Hong (2007), the development of data

protection applications must be informed by the necessity to achieve an equilibrium between the imperatives of protection and functionality, with the latter being characterised by transparency and user control.

To evaluate the developed location anonymisation application, a specialised framework was therefore developed. The framework is based on established usability principles and, at the same time, takes into account the special requirements of privacy-enhanced technologies (PETs). The theoretical basis underpinning this approach is a synthesis of:

1. **Nielsen's usability heuristics:** Nielsen (1994b) posited that usability is comprised of five core components: learnability, efficiency, memorability, error tolerance, and satisfaction. These general principles were employed as a foundational framework, albeit adapted to the context of data protection.

2. **Privacy by Design principles:** Cavoukian (2009) proposed seven fundamental principles for integrated data protection, including the concepts of user-centricity rather than technology-centricity and the importance of full functionality. These principles underscore the necessity of perceiving data protection and functionality as complementary aspects rather than as conflicting entities.

3. **Models of technology acceptance:** The Technology Acceptance Model (TAM), as proposed by Davis et al. (1989), was utilised to consider perceived usefulness and user-friendliness as pivotal factors in the acceptance of data protection technologies.

4. **Specific requirements for location data protection:** As posited by Krumm (2009) and Wernke et al. (2014), there exists a set of specific requirements for the usability of location protection technologies. These requirements include transparency of protection mechanisms, configurability of the degree of protection, and comprehensibility of the anonymisation strategy.

The integration of these theoretical approaches resulted in the following key dimensions for evaluating the location anonymisation application:

- **Learnability:** To what extent can the anonymisation features of the application be readily understood and utilised by users on their first encounter with the application?
- **Efficiency:** How expeditiously can users anonymise location data and make it available for desired services?
- **Memorability:** The extent to which users can readily utilise the application after an extended period of non-use.

- **Error tolerance:** To what extent does the mobile software application prevent data protection risks caused by user errors?
- **Satisfaction:** How satisfied are users with the data protection features?
- **Transparency:** How comprehensively do users comprehend the functionality of the anonymisation feature and its constraints?
- **Balance between protection and service quality:** In what manner do users rate the balance between data protection and the quality of services that can be used with anonymised data?

These dimensions constitute the conceptual framework within which the development of the evaluation methodology and the interpretation of the results are to be undertaken.

3.6.2 Task-Based Evaluation Design

It was decided that a task-based evaluation approach would constitute the most appropriate method by which to assess the defined dimensions of usability, and the balance between data protection and service quality. According to Hass (2019), this approach facilitates a realistic evaluation of usability by having participants perform typical tasks with the software and document their experiences.

The evaluation followed the scenario-based design as outlined by Rosson and Carroll (2002), wherein participants were tasked with performing four specific core functions of the application:

1. **Address entry and anonymisation:** The user is required to enter an address (e.g., "1 Sample Street, 1010 Vienna") in the text field and then anonymise it. The objective of this task is to evaluate the fundamental efficacy of anonymisation mechanisms and to ascertain the comprehensibility of the application's primary concept.

2. **Copying the anonymised result:** The anonymised location is transferred to the clipboard for subsequent utilisation. The present task is concerned with the interoperability of anonymised data with other applications and the practical usability of said data.

3. **Saving the anonymised location:** The selection of an appropriate icon is to be followed by the saving of its location in the application. The present task is concerned with the evaluation of the potential for the creation and reuse of consistent anonymised identities, a matter of particular significance in achieving a balance between data protection and service quality (Beresford & Stajano, 2003).

4. **Management of saved locations:** This function enables the user to access a list of previously saved locations and to delete an entry from the list. The present task is concerned with the evaluation of the application's control functions and identity management.

The tasks were meticulously selected to encompass the complete usage workflow of the application and to assess diverse facets of usability within the context of location anonymisation. As posited by Bødker et al. (2000), it is of particular significance for privacy tools that the entire usage cycle is evaluated, as opposed to merely individual functions.

For each task, participants were invited to rate its difficulty on a scale ranging from 1 (very difficult) to 5 (very easy). This quantitative assessment was supplemented by qualitative feedback questions to gain a deeper understanding of the user experience. This mixed-methods approach is consistent with the recommendations set out by (Albert & Tullis, 2010) for the evaluation of usability and user experience.

3.6.3 Participant Selection Strategy

The participants were selected through the author's personal contacts. Two distinct user groups were deliberately selected for the study to obtain varied perspectives on usability requirements:

1. **GIS-Experts:** The group comprised three of the author's longtime fellow students, who possess extensive expertise in the domains of geographic information systems, spatial analysis, and cartography. These individuals represent professional users who possess a profound understanding of spatial concepts and data protection aspects relating to location data.

2. **Non-Experts:** The group under discussion comprised three individuals with whom the author was personally acquainted, who had no specific professional background in geoinformation systems. These individuals embody the archetypal end users who would utilise the application in their quotidian lives.

This dichotomous division follows the concept of "*extreme groups*" proposed by Bødker et al. (2000), which uses contrasting perspectives to identify usability problems. The selection facilitates the identification of potential differences in the perception, utilisation, and acceptance of the application among specialist audiences and the general public, constituting a pivotal aspect for the determination of universal usability requirements.

The sample size of six participants (three per group) is commensurate with the 5-6 test subjects recommended by Nielsen and Landauer (1993) for usability tests. The authors posit that this sample size facilitates the identification of approximately 80% of usability problems while maintaining an optimal cost-benefit ratio.

The utilisation of personal contacts for the evaluation process resulted in the advantageous phenomenon of a high level of trust, which consequently engendered honest and detailed feedback. It is important to note, however, that the small sample size and the personal relationship with the author may impose potential limitations on the generalizability of the results. This is an aspect that is taken into account in the interpretation of the results.

3.6.4 Privacy Protection & Service Quality Metrics

A set of specific metrics, based on established data protection principles, was developed in order to evaluate the app's effectiveness in terms of data protection and balance with service quality. The metrics under discussion are based on Cavoukian's Privacy by Design Framework (2011) and the data protection principles of the GDPR (Voigt & Bussche, 2017), and have been specifically adapted for the context of location anonymisation:

1. **Perception of the purpose of data protection:** This metric is intended to evaluate the extent to which users comprehend the purpose of the application in terms of location anonymisation. As posited by Gerber et al. (2018), a comprehensive understanding of the data protection function is a pivotal factor in the acceptance of privacy-enhancing technologies.

2. **Willingness to use for data protection purposes:** By employing this metric, the degree of inclination to utilise the application for the purpose of anonymising personal location data can be ascertained. The model is grounded in the Technology Acceptance Model following Davis et al. (1989) and incorporates the perceived usefulness of the data protection function.

3. **Perceived effectiveness:** In essence, this metric is designed to evaluate the efficacy with which users perceive the application to safeguard their privacy. As emphasised by Acquisti et al. (2015), the perceived effectiveness of data protection measures is of crucial importance for their acceptance and utilisation.

4. **Willingness to recommend:** According to Reichheld (2004), this metric is indicative of the perceived value of a product, as it measures users' willingness to recommend the app for privacy purposes.

5. **Balancing the imperatives of data protection and service quality:** This specific metric examines how users rate the balance between the protection of their location data and the usability of anonymised data for services. As Krumm (2009) argue, this balance is a pivotal factor in the acceptance of location protection technologies.

The metrics in question were collected via the medium of qualitative questions in the questionnaire, with a particular focus on the sections pertaining to understanding of purpose, satisfaction, and willingness to recommend. The evaluation of these questions enables a differentiated view of data protection aspects and their integration into a user-friendly application.

3.6.5 Service Quality Analysis Methods

The analysis of the application's service quality employs a multidimensional approach, with a focus on achieving a balance between data protection and the quality of use. In accordance with the assertions put forth, digital services that are equipped with data protection functions are obligated to meet both the functional and emotional quality dimensions. The following methods were employed for the evaluation:

1. **Quality attributes according to Garvin (1987):**
    - **Performance:** How effective is the application in fulfilling its core function of location anonymisation?
    - **Features:** What is the rating of additional functions such as saving and copying?
    - **Reliability:** Has there been any error or delay in the anonymisation process?
    - **Compliance:** Does the application satisfy the criteria of a data protection application?
    - **Aesthetics:** What evaluation can be given to the visual design?

2. **SERVQUAL dimensions according to Parasuraman et al. (1988)**, adapted for mobile applications according to Ok (2025):
    - **Reliability:** Is the anonymisation function performed consistently?
    - **Response behaviour:** What is the response time of the application to user input?
    - **Security:** Does the application inspire confidence in its data protection features?
    - **Empathy:** Does the application satisfy the user's individual privacy requirements?

- **Material aspects:** What is the perception of the visual and interactive design of the application?

3. **Functional versus Emotional Quality Dimensions:** The questionnaire encompassed both functional aspects, pertaining to the completion of designated tasks, and emotional aspects, encompassing satisfaction and trust in the data protection function. Hassenzahl (2008) posits that the emotional dimension, encompassing sentiments of security and agency, plays a pivotal role in the satisfaction of users of data protection applications.

These methodologies enabled a comprehensive evaluation of service quality across multiple dimensions. A particular focus was placed on the perceived balance between data protection and service quality, an aspect that has been identified as being of crucial importance for the acceptance of privacy-enhancing technologies.

3.6.6 Data Collection and Analysis Procedures

The data was collected using a structured questionnaire developed according to the principles of user experience research by Albert and Tullis (2010) and comprising the following sections:

1. **Initial observations:** The utilisation of open-ended inquiries concerning the application's purpose and its visual design is intended to elicit impartial initial responses.

2. **Specific tasks:** The evaluation process involved the assessment of the complexity of four distinct tasks, utilising a 5-point Likert scale to quantify the perceived difficulty. This evaluation was complemented by the collection of qualitative feedback, which provided additional insights into the tasks' complexity and the perceptions of the participants.

3. **Post-task questions:** A reflective analysis was conducted on the tasks that were deemed to be both the simplest and most challenging, with the objective of eliciting feedback on the labels and icons in order to achieve a more profound comprehension of the usability of the system.

4. **Navigation:** Questions concerning the findability of functions and potential ambiguities, which represent critical aspects for mental modelling and user guidance.

5. **Feedback and error messages:** Experiences with error messages and preferences for feedback mechanisms are examined, drawing Nielsen's (1994a) heuristics for error prevention and error handling. This section also documented technical issues, including the occurrence of app crashes due to an absence of internet connection during the testing phase.

6. **Overall satisfaction and expressed desires:** Positive aspects and suggestions for improvement are identified in order to ascertain both strengths and potential for optimisation.

7. **Conclusion and recommendation:** The determination of the willingness to recommend and utilise the product is founded upon the Net Promoter Score (Reichheld, 2004) which is utilised as an indicator of overall satisfaction.

The participants were tasked with conducting the evaluation independently, and were required to complete the questionnaire immediately following the use of the app. The data was collected anonymously, with the only noted information being the subject's group affiliation (GIS-Expert or Non-Expert).

The following analysis methods were employed:

- **Quantitative analysis:** The calculation of mean values and the comparison of ratings between the two participant groups was undertaken, with particular emphasis placed on the equilibrium between data protection and service quality.

- **Qualitative content analysis:** Thematic coding and analysis of open-ended responses will be employed to identify recurring themes and patterns, utilising the method developed by (Braun & Clarke, 2006).

- **Comparative analysis:** A systematic comparison of responses from GIS-Experts and Non-Experts is required in order to identify different requirements and perceptions.

- **Usability problem identification:** The identification and classification of usability problems is to be conducted on the basis of user feedback, with the subsequent categorisation of said problems according to the heuristics proposed by Nielsen (1994a).

- **Triangulation:** The integration of quantitative and qualitative data is imperative in order to validate the results and identify both convergences and divergences (Creswell & Creswell, 2017).

This methodical approach enabled a comprehensive evaluation of the app in terms of user-friendliness, data protection functionality, and the balance between protection and service quality. The findings from this evaluation contribute directly to the answering of research question 2.1 by identifying the critical usability requirements for a user-friendly location protection application that enables both effective data protection and high service quality.

3.7 Evaluation Results

The evaluation of the location anonymisation application yielded substantial data on various aspects of user-friendliness, functionality, and acceptance. The ensuing sections are to be regarded as a presentation of the results of the survey, which have been organised in accordance with thematic focus areas. A particular focus of this study is the examination of the disparities and congruences between the assessments provided by GIS-Experts and Non-Experts. This investigation aims to elucidate the accessibility of the application for diverse user demographics. The results of the study are evaluated in two ways: quantitatively and qualitatively. Direct quotes from participants are included to supplement and contextualize the numerical evaluations.

3.7.1 Participant Demographics and Response Overview

The study comprised six participants, who were divided into two groups:

- Three GIS-Experts with specialist knowledge in the field of geographic information systems
- Three Non-Experts with no specific technical background in this field

All six participants completed the evaluation in full, providing answers to all questions. The depth and detail of the answers exhibited variation between participants, with an average of 1-3 sentences per open-ended question, thus providing a satisfactory foundation for qualitative analysis.

3.7.2 Conceptual Understanding and Initial Usability Assessment

The evaluation of initial user perception was conducted with the objective of assessing the intuitive comprehensibility of the anonymisation concept and the immediate impact of the interface design. The qualitative feedback on these aspects serves as an indicator of the effectiveness of the conceptual and visual communication of the app's functionality. Of particular interest in this study was whether the concept of location anonymisation could be

understood without prior knowledge and to what extent the minimalist user interface met user expectations.

App Comprehension

It was evident that all participants had a comprehensive understanding of the purpose of the application, irrespective of their professional background. The descriptions exhibited slight variations in their wording; however, the fundamental purpose was uniformly comprehensible: the application anonymises location data by generating alternative addresses in proximity to the actual location.

A GIS-Expert described the purpose in more technical detail:

"Generates an anoymized location that i can use instead of my actual adress. Therefore i can hide my personal information and offer made up coordinates to third-party apps."

An expert has outlined the key points in more colloquial language:

"Generating public addresses to avoid sharing private information."

This consistent purpose recognition is indicative of clear and effective communication of the app's concept, which is a fundamental usability goal.

Design and User Interface

The design and user interface received consistent positive feedback. Particular commendation was awarded to:

- **Simplicity and clarity of the design:**
  "It looks very clean (black and white). There are no major distractions from the essentials." (GIS-Expert)
  "Very good, minimalistic, easy to handle." (Non-Expert)

- **Intuitive operation:**
  "The App has an intuitive interface - straight to the point." (GIS-Expert)

- **Logo-Design:**
  "I really like the design of the logo with the geotags." (GIS-Expert)

The evaluation of design revealed no significant disparities between experts and Non-Expert s in GIS, suggesting that the visual design is universally appealing.

3.7.3 Task Performance Analysis

The analysis of task completion is a central component of the evaluation results. The app's practical user-friendliness is assessed, and the intuitiveness of its functions is evaluated, with a view to identifying areas for enhancement.

These preliminary texts have been designed to facilitate an optimal reading experience by ensuring that readers are provided with a comprehensive overview of the content and purpose of each section prior to the presentation of the detailed content.

Quantitative assessment of task difficulty

The assessment of task difficulty was conducted on a scale ranging from 1 (denoting extreme difficulty) to 5 (indicating extreme ease). The resulting average values are presented below:

Table 11: Quantitative assesment of task difficulty

Task	***GIS-Expert (Average)***	***Non-Expert (Average)***	*Overall Average*
2.1 Address entry and anonymisation	5,0	5,0	5,0
2.2 Copying the anonymised result	4,33	5,0	4,67
2.3 Saving the anonymised location	5,0	5,0	5,0
2.4 Management of saved locations	5,0	4,67	4,83
Overall Rating	***4,83***	***4,92***	***4,88***

The average ratings for all tasks were close to the maximum of 5 points, indicating that the app is very user-friendly. The average score of 4.88 out of 5 points on a scale from 1 to 5 indicates that the application is characterised by exceptionally good usability.

Task-specific feedback

Task 2.1: Address entry and anonymisation. The task was awarded the maximum possible rating (5/5) by all participants. The input mask was found to be intuitive and self-explanatory:

"I find it very easy, there is only one 'submit' button to push." (Non-Expert)

Task 2.2: Copying the anonymised result. This task was the subject of the lowest overall rating, particularly from those with expertise in GIS (4.33/5). One GIS-Expert commented:

"It seems like you first have to save it before I can copy the anonymised address -> I figured out that you can copy it. However, it was a bit confusing."

Task 2.3: Saving the anonymised location. This task received the highest possible rating (5/5) from all assessors. Although the majority of comments were positive, it was noted that the selection of an icon before saving was not immediately clear:

"The saving of the address was the 'most' difficult part because there was no hint that an icon has to be chosen first." (GIS-Expert)

Task 2.4: Management of saved locations. The task was assigned a rating of 5/5 by GIS-Experts and 4.67/5 by Non-Experts. One Non-Expert commented as follows:

"Maybe opening the list and deleting one entry could be one step easier."

Easiest and most challenging tasks

When asked to identify the most straightforward task, a number of aspects were highlighted:

- A GIS-Expert has indicated that the most straightforward approach is to **delete the saved locations**:
  "Deleting the saved location was the easiest because all there was to do was click on 'saved locations' and delete your selection."

- The simplicity of the **address entry** process was emphasised by a Non-Expert:
  "Type in the address into the field. Finding it."

In response to the question regarding the most challenging task, multiple participants asserted that none of the tasks were particularly difficult.

The few challenges mentioned pertained to:

- Saving the address without indicating that an icon must be selected
- Copying the result when it was not immediately clear that this was possible
- Lack of internet connection, which led to delays or crashes

3.7.4 User Interface Feedback

A detailed analysis of feedback on the user interface provides insights into the quality of the app's interactive elements and visual components. The following section will examine the clarity of the labels and icons used, the effectiveness of the navigation structure, and the

perception of the feedback mechanisms implemented. The data collected is primarily derived from responses to questions 3.3, 4.1, 4.2, and 4.3 to 4.5 of the Questionnaire, thus establishing a foundation for the identification of specific optimisation potential in the interface design.

Clarity of labels and icons

The majority of participants indicated that all labels and icons were clear and comprehensible.

Individual comments pertaining to:

- The meaning of the third icon when saving a location: *"With 'Save Location', the first two icons are clear, the third is unclear to me (at first I thought of kindergarten or an activity with the family)."* (GIS-Expert)
- The desire for different colours for the location icons: *"Everything was understandable, but the colors of the locations would be better in different colors."* (Non-Expert)

Navigation and user guidance

All participants found the navigation intuitive and easy to use. When asked whether there were any instances in which they lacked clear guidance, all respondents answered in the negative, suggesting a comprehensive understanding of the task at hand.

Feedback mechanisms and error messages

A limited number of participants encountered technical difficulties, which were found to be related to their internet connection.

The following requests were made regarding additional feedback:

- Confirmation feedback following the saving of a location: *"It would be nice to have an 'address saved' feedback, after saving the location"* (GIS-Expert)
- A additional loading indicator during address search: *"When entering locations, a loading symbol would be useful during the transition period, and the loading bar could be displayed in a lighter color"* (Non-Expert)

The notification feature was considered useful by all participants, especially with regard to facilitating the process of copying and pasting into other applications.

3.7.5 Privacy Perception Analysis

The present section is concerned with the participants' perception and assessment of the data protection aspects of the developed application. The analysis examined the extent to which the data protection function was recognised as relevant added value, which usage scenarios participants identified for the anonymisation function, and how they assessed its practical significance in the context of current digital data protection requirements. The findings are primarily derived from the responses to questions 6.1, 7.1, and 7.2 of the questionnaires, offering insight into the perceived relevance of the app's functionality.

Perception of data protection functionality

The data protection feature of the application was comprehended and endorsed by all participants. The significance of data protection was particularly evident in the responses to the question regarding the anonymisation feature:

"I find it practical, especially if privacy is very important to you and you don't want to pass your sensitive data on to companies or developers" (GIS-Expert)

Intent of use for data protection purposes

In response to the question regarding their own intended use, disparities became evident between the two groups:

- GIS-Experts tended to be more cautious in their comments:
  "Can't say. But I guess, I will not use it."
  "I would if I would use third-party apps that I haven't used before."

- Non-experts consistently showed a higher willingness to use the service:
  "Definitely yes. I have never seen an app like this and I would use it to anonymize addresses for example when ordering stuff online."
  "Yes. For keeping my information safe."

These discrepancies could signify that experts in GIS may have prior experience with alternative methodologies for safeguarding their location data, or alternatively, they could possess a more nuanced evaluation of the necessity for such anonymisation mechanisms.

Willingness to recommend in the context of data protection

It was reported by all participants that they would recommend the app to others, especially to people who value privacy.

However, it should be noted that the willingness to utilise the system may be influenced by the general awareness of data protection issues, as observed by one GIS-Expert:

"If someone wants more privacy, then yes. But it seems that most people no longer care who gets their data or what is done with it."

This statement suggests a potential discrepancy between the perceived value of the data protection function and the actual willingness to utilise it among the general population.

3.7.6 Comparative Analysis: Expert vs. Non-Expert Users

A fundamental question that underpins this evaluation pertains to the potential disparities in the perception and utilisation of the application between GIS-Experts and Non-Experts. This comparative analysis systematically examines the similarities and differences between the two user groups in terms of user-friendliness, understanding of functions, and perceived added value. The comparison enables conclusions to be drawn about the extent to which the developed application caters to both target groups equally, and whether specific adjustments are necessary for different user profiles.

Differences in the assessment of user-friendliness

The quantitative analysis demonstrated that the application received a slightly higher overall rating from Non-Experts (4.92/5) compared to GIS-Experts (4.83/5). This suggests that the app is also highly user-friendly for users without specific GIS expertise, which was an important goal during the development process.

Task-specific differences:

- In the estimation of GIS-Experts, the process of copying the result was considered to be marginally more challenging (4.33/5) in comparison to the perspective of Non-Experts (5.0/5).
- In the opinion of Non-Experts, the tasks of opening the list and deleting items were considered slightly more difficult (4.67/5) than those performed by GIS-Experts (5.0/5).

Nevertheless, these differences are minimal and not statistically significant given the small sample size.

Differences in the perception of the benefits of data protection

A clearer difference emerged in the perception of the benefits of the data protection function:

- GIS-Experts tended to take a more nuanced, sometimes more skeptical view:
  "If someone wants more privacy, then yes. But it seems that most people no longer care who gets their data or what is done with it."

- Non-experts showed a consistently more positive approach to the data protection function:
  "I find it fascinating that you can program something like this with Python, as described in the info box. Well done, keep up the good work!"

This could be indicative of a more critical perspective among GIS-Experts, who possess specialist knowledge of location data and data protection, regarding the actual effectiveness of such anonymisation measures.

Similarities between the groups

Despite the differences mentioned above, there were notable similarities between the two groups:

1. It was evident that both groups comprehended the objective of the application with relative ease.
2. The design and user interface were evaluated by both groups, with a high level of satisfaction expressed in both cases.
3. It was observed that both groups demonstrated an ability to complete all tasks with minimal difficulty.
4. The app is likely to be endorsed by both groups to individuals who prioritise privacy.

These similarities are indicative of the app's universal usability and practicality, irrespective of the users' technical backgrounds.

3.7.7 Feature Requests and Improvement Suggestions

The participants proposed a number of recommendations for enhancing the application, which can be categorised as follows:

Expansion of existing functionality

- **Expanded icon selection:**
  "Perhaps more icons for different occasions." (GIS-Expert)

"Maybe more icons added to label the different addresses." (Non-Expert)

- **Multiple locations per icon:**
  "Being able to generate multiple locations per icon." (GIS-Expert)

Improvement of visualization and navigation

- **Map overview of saved locations:**
  "It would be nice to have a 'map overview' of the saved locations because mostly, the new address names are unknown to the user but a visual representation on a map can help to connect the 'fake' address to the one that got switched." (GIS-Expert)

- **Colour differentiation:**
  "Everything was understandable, but the colors of the locations would be better in different colors" (Non-Expert)

Improvement of feedback and user guidance

- **Confirmation after saving:**
  "It would be nice to have an 'address saved' feedback, after saving the location." (GIS-Expert)

- **Visibility when entering a location description:**
  "While typing in which location you have generated (like 'Home'), to see what you have wrote." (Non-Expert)

- **Loading indicator:**
  "When entering locations, a loading symbol would be useful during the transition period, and the loading bar could be displayed in a lighter color." (Non-Expert)

Design improvements

- **Dark Mode:**
  "I would also really like a Dark mode." (Non-Expert)

- **Brief explanation below the logo:**
  "At first glance, it is not entirely clear what the app does (unless you click on the info button). Perhaps a short signature under the logo with e.g. 'Anonymizing Address' would be helpful." (GIS-Expert)

The suggestions for improvement focus primarily on enhancements and refinements to existing functionality, rather than fundamental changes to the app's structure or functionality. This underscores the generally high level of satisfaction with the app.

3.7.8 Conclusion and Implications

The evaluation results demonstrate a high level of satisfaction with the developed application for location anonymisation, both among GIS-Experts and among individuals with no expertise in this area. The average rating of 4.88 points on a scale of 5 points for user-friendliness demonstrates the exceptionally good usability of the application.

Key findings

1. **High user-friendliness:** The two user groups under consideration both perceive the app to be very user-friendly, with minimal differences in their ratings.
2. **Clear understanding of the concept:** It is evident that all users, irrespective of their level of expertise, comprehend the purpose and functionality of the application.
3. **Positive design feedback:** The minimalist design and clear structure have received a uniformly positive response.
4. **Different intended use:** It has been demonstrated that individuals lacking expertise in the field of Geographic Information Systems (GIS) demonstrate a greater propensity to utilise the application in question than those who are versed in the subject.
5. **Appreciation of the data protection function:** The data protection function is considered valuable by all participants, especially for users who prioritise data protection.

Implications for further development

1. **Retain the minimalist design:** The minimalist design should be retained as the core strength of the application.
2. **Focus on feedback optimisation:** Minor enhancements to the feedback system, such as the implementation of confirmation messages and loading indicators, have the potential to further enhance the user experience.

3. **Extended customisation options:** The incorporation of additional icons and the functionality to save multiple locations for each category would serve to enhance the app's utility.

4. **Visual representation of locations:** The provision of a map overview of saved locations has the potential to enhance the user's orientation.

5. **Advanced design options:** The implementation of a dark mode, coupled with the differentiation of icons in terms of colour, has the potential to enhance the visual appeal of the interface.

The evaluation confirms the successful approach of the app in offering location anonymisation in a user-friendly form that is accessible to both experts and the general public. The high level of satisfaction expressed, coupled with constructive suggestions for improvement, provides a solid foundation for the future development of the application.

4 Discussion

The present discussion is devoted to the critical analysis and classification of the developed DeLocator application and its evaluation results. As set out in the preceding chapters, the theoretical foundations, technical implementation and empirical evaluation have been presented. The present chapter thus provides a comprehensive interpretation and contextualisation of the findings. The chapter under scrutiny here demonstrates the implications of the results achieved for research and practice, whilst also providing a critical reflection on the strengths and limitations of the chosen approach.

The ensuing discourse is divided into six primary sections: Firstly, the DeLocator application is evaluated in the context of current privacy research, followed by a critical analysis of the implemented substitution-based anonymisation approach. This is followed by a methodological reflection on the evaluation carried out, before the practical applications of the developed solution are discussed. It is asserted that the results of the research are to be found embedded in a broader scientific and practical context through this structured analysis. Furthermore, it is claimed that the innovative contribution of the research to the field is highlighted.

4.1 Evaluation of the DeLocator application in the context of current privacy research

The evaluation results for the DeLocator application demonstrate remarkably high user acceptance, with an overall rating of 4.88/5.0 points. This score is considerably higher than the typical ratings for privacy applications documented in the literature. While comparable studies by Acquisti et al. (2015) report average usability ratings between 3.2 and 3.8 points for privacy tools, DeLocator demonstrates a significant improvement in the usability of privacy applications.

It is particularly noteworthy that the anonymisation concept is universally comprehensible. It was evident that all participants in the study, comprising both GIS-Experts and Non-Experts, comprehended the purpose and functionality of the application under investigation. This finding is in stark contrast to the research conducted by Bott and Young (2012), which demonstrated that complex privacy mechanisms are correctly understood by an average of only 34% of users. The DeLocator application provides a clear conceptual representation of the aforementioned principle, thereby validating the hypothesis concerning the utilisation of

conceptual models for the facilitation of intuitive technology acceptance. Furthermore, it is evident that privacy mechanisms do not necessitate an inherent level of complexity.

The minimal difference in the overall rating between GIS-Experts (4.83/5.0) and Non-Experts (4.92/5.0) is another significant finding. This minor discrepancy serves to illustrate the effective abstraction of technical intricacy, thereby addressing a pivotal critique of prevailing privacy instruments. According to Reay et al. (2013), existing tools are frequently "only usable by experts." The DeLocator application has been developed to overcome these limitations through the implementation of a minimalist design, an intuitive user interface, and a transparent visualisation of the anonymisation process.

An intriguing facet of the evaluation pertains to the discrepancy in usage intentions observed between the two participant groups. While experts demonstrated a more discerning and selective approach, as evidenced by their response, I would if I would use third-party apps that I haven't used before", Non-Experts exhibited a more enthusiastic inclination to adopt the application, as illustrated by their response, "Definitely yes. I have never seen an app like this". This divergence corroborates Nissenbaum's (2010) observations that technical expertise engenders a more nuanced yet also more sceptical evaluation of privacy technologies. Notwithstanding the divergent fundamental attitudes, both groups assigned a comparably favourable rating to the user-friendliness and concept of the application.

The DeLocator application is an exemplary implementation of Cavoukian (2009) privacy-by-design principles. The proactive approach, in which location data is anonymised prior to transmission, is consistent with the principle of "proactive rather than reactive". The principle of "privacy as standard" is realised by automatic anonymisation without the need for explicit user activation. The high usability ratings also confirm the principle of "full functionality," which states that data protection and functionality do not have to be mutually exclusive. The successful implementation of these principles in a practical application serves to extend the theoretical framework of privacy by design with empirical validation.

The evaluation results also extend the Technology Acceptance Model (TAM) of Davis et al. (1989) to include privacy-specific factors. The high correlation between perceived anonymisation effectiveness and intention to use indicates "perceived privacy effectiveness" as an important additional factor for the acceptance of privacy technologies. The favourable evaluation of the map display also emphasises the function of transparency as a moderator that impacts both perceived usefulness and perceived usability. These findings contribute to the further development of the Technology Acceptance Model (TAM) for the specific context of privacy technologies.

4.2 Critical analysis of the replacement-based anonymisation approach

The substitution-based anonymisation approach implemented in DeLocator represents a deliberate departure from traditional perturbation methods. While established academic approaches such as differential privacy (Dwork, 2008) and geo-indistinguishability (Chatzikokolakis et al., 2013) offer mathematically rigorous guarantees, they often lead to impractical results with limited usability. The high usability ratings of the DeLocator application provide validation for the fundamental design decision to utilise real points of interest (POIs) as substitutes for private addresses.

The substitution strategy that utilises real POIs from six distinct categories (restaurant, café, bank, supermarket, pharmacy, transportation) within a 500-meter radius offers numerous advantages over mathematical perturbation methods. Firstly, it guarantees the utilisation of authentic, navigable addresses that can be readily employed in subsequent applications. Secondly, the utilisation of random selection from validated POI categories facilitates non-deterministic randomisation, thereby impeding the emergence of predictable patterns and rendering systematic re-identification attempts more arduous. Thirdly, the random selection of data from a range of potential alternatives engenders the concept of "plausible deniability", an important concept in the realm of practical data protection.

The integration of contextual categorisation with an icon-based selection system (e.g., Home, Work, Family) represents a substantial enhancement of prevailing "one-size-fits-all" methodologies. This differentiation corresponds to Beresford and Stajano's (2003) concept of "mix zones" with contextual adaptation, and addresses the realization that privacy requirements vary greatly depending on context. The positive evaluation of this functionality reinforces the understanding that users recognise and appreciate the value of different anonymisation strategies for different areas of life.

The chosen approach positions DeLocator in a separate segment between mathematically rigorous but impractical academic solutions on the one hand and simplified commercial tools on the other. While differential privacy implementations, such as those by Chatzikokolakis et al. (2013), offer theoretically elegant guarantees, their usability studies show significantly lower user acceptance (2.1-3.4/5.0) compared to DeLocator's 4.88/5.0. Conversely, commercial privacy tools such as browser-based solutions (DuckDuckGo, Brave) offer reactive protection through tracker blocking; however, they do not offer the proactive anonymisation function implemented by DeLocator.

The implemented solution establishes a "pragmatic privacy" paradigm, which prioritises practical applicability over mathematical perfection. This approach acknowledges that achieving optimal privacy may often necessitate compromising usability, whereas a

pragmatic compromise can offer adequate protection for most routine scenarios while simultaneously enhancing the user experience. This finding aligns with the observation made by Acquisti et al. (2015) that privacy decisions are context-dependent and frequently made on the basis of practical considerations rather than abstract principles.

However, it should be noted that the substitution approach is also subject to deliberate limitations, which were carefully considered against the practical advantages offered. In contradistinction to differential privacy, it does not offer formal mathematical guarantees for the level of protection achieved. The effectiveness of this method is contingent upon the availability and quality of POI data, which can be problematic in less well-mapped or rural areas. Finally, it is important to note that the repeated use of the same original addresses poses a potential risk for pattern recognition attacks, as no systematic analysis of all possible deanonymisation strategies has been performed.

Notwithstanding the aforementioned limitations, the positive evaluation indicates that the selected approach is both effective and user-friendly for its intended use, namely the protection of private addresses in everyday text-based communication. The equilibrium between data protection and practicality is aligned with the real-world requirements of typical users, thereby providing a pragmatic yet effective option to the range of available privacy strategies.

4.3 Methodological reflection and validity of the evaluation

The methodological approach to evaluating the DeLocator application follows established principles of design science research (DSR) according to Hevner et al. (2004). The approach is characterised by a systematic combination of problem identification, solution design, implementation, and evaluation, ensuring both theoretical soundness and practical relevance. The methodology that has been selected is especially appropriate for the evaluation of privacy technologies, insofar as it considers both technical and human factors.

The evaluation design's robustness is attributable to the triangulation of disparate data sources and methodologies. The integration of quantitative usability metrics with qualitative user feedback facilitates a nuanced interpretation of the results, thereby enhancing their validity. This mixed-methods approach is consistent with the recommendations of Creswell and Creswell (2017), who argue that complex socio-technical systems, such as privacy tools, can be most effectively evaluated through methodological triangulation.

The "Extreme Groups Design" approach, which involves the participation of three GIS-Experts and three Non-Experts, facilitates a systematic investigation of the expertise-

dependent disparities in the perception and utilisation of the application. This approach is founded on the recommendations of Bødker et al. (2000), who advocate contrasting perspectives for the identification of usability problems. The data collected demonstrate significant variations in privacy perception and usage intention between the groups, while concurrently yielding congruently positive usability ratings.

The task-based evaluation, encompassing four distinct core functions (namely, address entry and anonymisation, copying the anonymised result, saving the anonymised location, and managing saved locations), adheres to the scenario-based design as outlined by Rosson and Carroll (2002). The tasks were meticulously selected to encompass the complete usage workflow of the application and to evaluate various aspects of usability in the context of location anonymisation. The high ratings for all tasks (between 4.67/5.0 and 5.0/5.0) confirm the consistently good usability of the application.

However, it is important to note that the selected evaluation methodology also has certain limitations that must be considered when interpreting the results. Despite the fact that the sample size of six participants adheres to the recommendations set out by Nielsen and Landauer (1993) for qualitative usability tests, this does limit the statistical generalizability of the results. The authors posit that this sample size is capable of identifying approximately 80% of usability problems while maintaining an optimal cost-benefit ratio. However, it is imperative to acknowledge the necessity for larger quantitative studies to be conducted in order to formulate definitive statistical statements.

Another methodological limitation is the short-term nature of the evaluation. The immediate post-task assessment does not capture long-term effects, habituation phenomena, or sustained usage patterns. However, these factors assume particular relevance in the context of privacy tools, as initial enthusiasm is often superseded by a state of "privacy fatigue" (Choi et al., 2018), which can result in diminished long-term utilisation. Longitudinal studies would provide a more complete picture of sustainable use and acceptance.

The utilisation of personal contacts for the evaluation process may have resulted in a positivity bias, despite the anonymous nature of the study design, which was intended to mitigate this potential effect. However, it is possible that social desirability and personal relationships may have influenced the ratings. It is imperative to consider this potential bias when interpreting the remarkably positive outcomes.

Notwithstanding the aforementioned limitations, the selected methodology provides a robust foundation for the evaluation of the DeLocator application. The integration of quantitative and qualitative data, the consideration of divergent user perspectives, and the systematic coverage of the entire usage workflow facilitate the formulation of valid statements

regarding the usability and user acceptance of the application. The methodological limitations identified also provide valuable insights for future, more comprehensive evaluations.

4.4 Practical implications and possible applications

The DeLocator application, which has been developed and evaluated in a positive manner, has far-reaching practical implications for various domains, from the development of location-based services to privacy practices and regulatory approaches. These findings have implications for the research beyond the academic context, and highlight concrete application possibilities for practice and industry.

For developers of location-based services, the DeLocator approach offers valuable design principles that can lead to improved user experiences while maintaining privacy. The transparency of the anonymisation process through map display is a particularly salient principle that has been demonstrated to engender increased user trust. In an era characterised by mounting concerns regarding privacy, the cultivation of trust can prove to be a pivotal factor in achieving a competitive advantage. The increasing personalisation of digital services is corresponded to by contextual differentiation with different anonymisation strategies for different life contexts. This enables a more nuanced balance between privacy and service quality.

DeLocator's implementation strategies, particularly the utilisation of real POIs as anonymisation targets, the employment of icon-based categorisation for intuitive contextualisation, and the implementation of local data processing, offer practical blueprints for the integration of privacy features into both existing and new applications. These strategies are of particular pertinence in the context of applications such as ridesharing services, food delivery, and social media, where location data is frequently disseminated and concerns regarding privacy are on the rise.

For privacy practitioners and data protection officers, the evaluation establishes significant criteria for the selection and evaluation of privacy tools. Ensuring that users have a comprehensive understanding of the privacy mechanisms should be prioritized as a core criterion. The practical applicability of anonymised data is also crucial; theoretically optimal privacy solutions are worthless if the resulting data cannot be used in real-world applications. In conclusion, it is imperative to consider contextual flexibility, given the significant variations in privacy requirements that are contingent on the context of utilisation.

In the domain of regulation and compliance, DeLocator exemplifies the efficacy of pragmatic anonymisation methodologies in ensuring GDPR compliance while maintaining cost-effectiveness. The local processing and data minimisation of the aforementioned data are in accordance with the fundamental principles of the EU General Data Protection Regulation. The approach can be applied to various regulated industries, such as healthcare (where the Health Insurance Portability and Accountability Act [HIPAA] sets the standard for the protection of patient data) and the financial sector, where both data protection and service quality are critical requirements.

In the context of urban mobility and smart city applications, DeLocator presents concrete solutions to a pivotal dilemma: the question of how personalised services can be delivered without compromising privacy. The application demonstrates the potential for location data to be utilised across various services while ensuring the confidentiality of the precise home address. This balance is of particular relevance to public transportation systems, sharing economy platforms, and urban navigation services.

The practical applicability of the DeLocator concepts is underscored by the minimal training required for effective utilisation. The intuitive interfaces and transparent visualisation significantly reduce adoption barriers. The evaluation demonstrated that all participants, irrespective of their technical expertise, were able to utilise the application effectively from the outset. This is a crucial aspect for real-world implementation in diverse user groups.

In summary, DeLocator demonstrates that pragmatic privacy solutions can offer significant practical value in various areas of application. The present study posits that the tension between data protection and service quality does not have to be conceptualised as a fundamental trade-off; rather, it can be optimised through user-centred design and contextual adaptation. This insight provides a novel framework for the practical implementation of privacy-by-design in commercial and public applications.

5 Conclusion

The objective of this master's thesis was to develop and evaluate a user-friendly solution for anonymising location data in textual form, with a particular focus on the identified privacy gaps in established mobility platforms such as Google Maps and WienMobil. The DeLocator application was designed and implemented as a pragmatic alternative, based on the theoretical framework of location privacy and the divergent data protection approaches of these platforms.

5.1 Summary of key findings

The theoretical analysis in Literature Review revealed clear differences in how location data is handled by Google Maps, a global provider, and WienMobil, a local mobility application. Google Maps operates an extensive ecosystem involving the collection and integration of data across various services (including GPS data, IP addresses and search history), whereas WienMobil takes a more focused approach based on data minimisation principles. However, it should be noted that both platforms utilise textual location data, which has the potential to divulge confidential information regarding users.

The evaluation of data protection strategies on both platforms, with particular reference to de-identification and differential privacy, revealed their effectiveness; however, it was also demonstrated that these strategies can present significant challenges for end users in terms of accessibility. Google's primary approach is predicated on generalisation and the introduction of noise, whereas WienMobil's emphasis lies in data minimisation. However, these mechanisms are implemented in a largely non-transparent manner for users, offering no direct control over the degree of anonymisation.

The DeLocator application was developed as a response to these shortcomings, and it successfully demonstrates that effective location anonymisation can be user-friendly. The exceptionally positive rating of 4.88/5 it received significantly exceeds the documented values for comparable privacy tools, validating the chosen substitution approach as a viable alternative to mathematical perturbation methods such as differential privacy.

5.2 Answering the Research Questions

SO1: Analysis of commonly used digital platforms

1.1. What are the different approaches in terms of collection, storage, protection, sharing, and terms of use of user specific textual location information in commonly used digital platforms?

An analysis of Google Maps and WienMobil was conducted, with the objective of ascertaining the fundamental differences in their approaches to handling textual location information.

- **Data Collection:** Google Maps has been developed to collect comprehensive location data, including GPS coordinates, IP addresses, search queries, and information about nearby Wi-Fi networks and Bluetooth devices. In contrast, WienMobil's primary focus is on specific mobility data, leveraging Google Firebase for app analytics and storing anonymised user interactions.

- **Data storage:** Google has established a categorisation system for data storage that encompasses four distinct deletion cycles, ranging from immediate deletion to long-term storage for compliance purposes. In accordance with the principle of data minimization, WienMobil retains data for no longer than is necessary for the stated purpose, with specific deletion periods in place (e.g., 14 months for Google Analytics data).

- **Data protection:** It is evident that both of these platforms rely on de-identification and generalization. Google employs differential privacy-like methods by adding noise, while WienMobil relies more heavily on anonymisation through Firebase IDs without personal references.

- **Data sharing:** Google shares data with third-party providers for integrated services and uses international frameworks, such as the EU-US Data Privacy Framework. WienMobil is a company that is known for its commitment to data protection. It limits the amount of data that it shares to only the essential functions, and it uses standard contractual clauses for transfers to third countries.

1.2. Do commonly used digital platforms employ location protection methods or user privacy enhancement tools for data subjects to lower the risk of re-identification? If so, what algorithms are being used and how can they be evaluated in terms of privacy protection offered?

It is evident that both platforms implement protective measures, albeit with different levels of user accessibility:

Google Maps employs a multi-layered algorithmic approach, encompassing generalisation through spatial and temporal abstraction, randomisation through controlled noise similar to differential privacy mechanisms, and automated K-anonymity for specific data sets. The evaluation indicates a high level of technical effectiveness, yet it also highlights a deficiency in user transparency and control.

WienMobil primarily relies on data minimisation and anonymisation through Firebase installation IDs without personal references. Protection is primarily facilitated through legal and technical measures (e.g. GDPR compliance, local data storage in Europe), with minimal provision for direct user control.

The evaluation of both systems indicates that the technical protection measures are effective in principle, yet a substantial gap exists between technical implementation and user understanding and control.

SO2: Development and evaluation of the DeLocator application

2.1. What are the usability requirements for a user-friendly location protection software application that balances both the protection and the quality of a requested service?

The evaluation identified five fundamental usability requirements for location protection software that is designed to be user-friendly:

1. **Conceptual clarity:** It is imperative that users possess a comprehensive understanding of the anonymisation process. DeLocator accomplished this by employing a transparent visualisation of location substitution via map display.

2. **Minimal interaction effort:** It is recommended that the workflow comprise a maximum of between 3 and 4 steps. DeLocator implemented an efficient 3-step process (input → anonymisation → use).

3. **Contextual flexibility:** The level of protection required in any given area depends on the specific circumstances. The icon-based categorisation system (home, work, family) was universally positively rated.

4. **Practical applicability of the results:** Anonymised locations must retain their functionality in real-world applications. POI substitution ensures navigable, real-world addresses.

5. **Transparency and trust:** Users require insight into the protection process. The incorporation of a visual representation of the substitution in the map display resulted in a substantial enhancement in the level of trust.

2.2. To what degree can the developed software applications prevent re-identification?

DeLocator implements multi-layered re-identification protection measures:

- **Substitution-based protection:** The process of direct re-identification is rendered infeasible by the replacement of private addresses with a randomly selected set of points of interest (POIs) from six distinct categories, situated within a 500-metre radius. The utilisation of non-deterministic selection from a pool of available POIs serves to impede the efficacy of systematic pattern recognition attacks.

- **Consistency-based anonymisation:** A fundamental element of protection is the consistency of substitution, whereby once an original address has been anonymised and stored, the same anonymised version is utilised whenever the same address is re-entered. This is an essential step in preventing potential inconsistencies that could enable re-identification through correlation analyses. Attackers are unable to obtain different substitutes for the same address by entering it repeatedly, which would prevent effective tracing back to the original address.

- **Contextual anonymisation:** The utilisation of various categories (restaurant, café, bank, etc.) engenders a phenomenon known as "plausible deniability," thereby ensuring the true destination remains ambiguous.

- **Local processing**: All processes pertaining to anonymisation are executed exclusively on the device, with the stipulation that no raw data is transferred to external service providers.

- **Enhanced privacy through persistence:** While consistency-based anonymisation ensures that the same original address is always mapped to the same anonymised substitute (thus preventing attackers from exploiting inconsistencies), the persistence aspect goes a step further by maintaining this mapping over time and across sessions. This persistent assignment offers additional protection compared to approaches that generate a new anonymised substitute with every request, as it prevents systematic comparison attacks in which an attacker could recognise patterns by repeatedly submitting the same address. In summary, persistence not only guarantees consistency but also strengthens privacy by ensuring that the anonymisation cannot be undermined through repeated queries over time.

- **Limitations:** It is important to note that this approach does not provide formal mathematical privacy guarantees, such as those offered by differential privacy. A potential limitation arises if only a small number of original addresses are used infrequently: in this case, the pool of stored substitutes is also very small. As a result, it may still be possible, in theory, to draw conclusions about user behaviour or usage patterns based on these few anonymised addresses. However, because the mapping between the original and anonymised addresses always remains the same (consistency), it is much harder for an attacker to link the anonymised data back to the original addresses, which makes re-identification attempts significantly more difficult.

2.3. How is the protection level dependent on properties such as characteristics of the study area, user activity, or type of digital platform?

The analysis revealed clear dependencies in the level of protection:

- **Characteristics of the study area:** The analysis revealed that the protection level is not only influenced by the number of points of interest (POIs), but, more importantly, by the number of residential addresses or potential individuals in the vicinity of a given point. While urban areas such as Vienna generally feature a high density of POIs and a large number of available substitutes, thereby offering a higher level of protection, this is not always the case for all locations within a city. For example, in certain urban districts like Grunewald in Berlin, which are characterised by large plots, low building density, and extensive green spaces, the number of residential addresses around a given point can be very limited. In such scenarios, even though the area is part of a major city, the available alternatives for anonymisation are reduced, which can compromise the effectiveness of privacy protection. This challenge is even more pronounced in rural areas, where both the number of POIs and the density of residential addresses are generally low. Consequently, the risk of re-identification increases in areas with fewer plausible substitutes, regardless of whether the setting is urban or rural.

- **User activity:** The consistent mapping of original to anonymised addresses has been demonstrated to provide stable protection with regular use. This is due to the fact that no varying patterns exist that could enable re-identification. Users who possess a limited set of frequently used addresses benefit from consistency, as their anonymised substitutes remain predictable and trustworthy. The categorisation of usage contexts (e.g. home, work, family) is addressed through the implementation of contextual categorisation, with the allocation of separate substitution pools for each.

- **Type of digital platform:** The DeLocator solution is platform-independent and can be used for various text-based forms of communication. The persistent storage of mappings renders the solution particularly suitable for recurring communication scenarios, as users can employ consistent anonymised addresses.

- **Temporal factors:** The persistence of anonymisation provides long-term protection, as mappings remain constant over time. This eliminates the possibility of time-based correlation attacks, in which an attacker could detect patterns in the variation of anonymisation over different points in time.

5.3 Limitations of the Study

Notwithstanding the encouraging outcomes of this research, it is imperative to recognise the limitations inherent to both the technical aspects of the anonymisation approach and the methodological constraints of the evaluation process.

The following **technical limitations** must be taken into account:

A notable technical constraint associated with the selected substitution approach pertains to the absence of formal privacy guarantees. In contrast to established differential privacy methods, DeLocator does not offer mathematically provable levels of protection. This deliberate design decision prioritises practical applicability but limits theoretical robustness.

The application's urban orientation is a notable limitation. Its dependence on high POI density significantly limits its effectiveness in rural areas. This limitation is indicative of a fundamental problem with many privacy technologies, which are often optimised for environments that are densely populated and well-connected.

The utilisation of API dependencies on external services, such as OpenStreetMap (Nominatim, Overpass), has the potential to impact the reliability and scalability of the application. It is important to note that high payloads have the potential to exceed API limits, which can result in performance issues that compromise the user experience.

The simplified threat model of the application primarily considers passive attackers, while active correlation attacks or advanced de-anonymisation techniques have not been systematically analysed. This is of particular relevance in cases of repeated use of the same original addresses, where patterns could emerge that enable re-identification.

In contrast, the following **methodological limitations** are identified:

From a methodological perspective, the limited sample size of six participants is a major limitation. While this size is adequate for the collection of qualitative usability insights, it does not allow for the drawing of statistically significant generalisations.

The restriction of the study to German-speaking, urban users means that the results cannot be transferred to other cultural contexts. As demonstrated by Bellman et al. (2004), there is considerable variance in privacy perceptions and behaviours between different cultures.

The limited time frame of the evaluation precludes consideration of long-term effects and habituation phenomena, which are crucial for sustainable adoption. Privacy tools frequently encounter the phenomenon of "privacy fatigue" (Choi et al., 2018), which can result in diminished long-term utilisation.

5.4 Future Research Recommendations

These limitations, however, give rise to a number of promising areas for further research, which could lead to the further development of the concept of pragmatic location anonymisation.

Technical developments:

Future research could develop hybrid approaches that combine formally justifiable guarantees with practical usability, for example through the use of adaptive differential privacy with context-sensitive parameters. In order to address the challenges posed by areas with low density of residential addresses, alternative strategies must be developed. These may include intelligent radius adjustment, the inclusion of natural landmarks, or hybrid approaches utilising traditional perturbation methods.

Exploration of more robust architectural approaches is recommended, such as local POI databases with periodic updates or hybrid systems that can adaptively switch between local and remote data. It is imperative that comprehensive security analyses and advanced protection mechanisms against various attack scenarios are implemented.

Methodological extensions:

It is recommended that quantitative studies with larger samples ($n > 100$) be conducted, with a view to enabling generalizable statements about user acceptance and usability. International comparative studies could reveal cultural differences in privacy preferences and the acceptance of technological solutions, which would be helpful in developing culturally adapted privacy solutions.

Longitudinal studies over several months could analyse habituation effects, sustainable usage patterns, and evolving privacy preferences to provide a more complete picture of real-world application potential.

Innovative research directions:

- **Adaptive intelligence:** Machine learning-based systems have the capacity to learn user preferences and develop personalised anonymisation strategies that adapt to individual privacy utility preferences.
- **Cross-platform implementation:** The potential exists for expansion to iOS and web platforms, with the concomitant possibility of increased reach and practical applicability.
- **Advanced POI intelligence:** The selection of POIs based on semantic similarity, usage patterns, and contextual information has the potential to enhance the relevance and naturalness of anonymised locations.
- **Standardized privacy metrics:** The development of standardised metrics for substitution-based anonymisation would facilitate systematic comparisons and evaluations.

The advancement of these research domains has the potential to enhance the technical aspects of location anonymisation and to expand our understanding of user-centric privacy technologies. The interdisciplinary nature of these challenges necessitates collaboration between computer science, UX design, psychology, and privacy law—a collaboration that has the potential to result in innovative and practical privacy solutions for our increasingly connected digital society.

Bibliography

Abdelmoty, A. I., & Alrayes, F. (2017). Towards Understanding Location Privacy Awareness on Geo-Social Networks. *ISPRS International Journal of Geo-Information*, *6*(4), 109. https://doi.org/10.3390/ijgi6040109

Abowd, J. M. (2018). The U.S. Census Bureau Adopts Differential Privacy. In *KDD '18, Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining* (p. 2867). Association for Computing Machinery. https://doi.org/10.1145/3219819.3226070

Acquisti, A., Brandimarte, L., & Loewenstein, G. (2015). Privacy and human behavior in the age of information. *Science*, *347*(6221), 509–514. https://doi.org/10.1126/science.aaa1465

Aggarwal, C. C. (2005). On k-anonymity and the curse of dimensionality. In *VLDB '05, Proceedings of the 31st international conference on Very large data bases* (pp. 901–909). VLDB Endowment. https://dl.acm.org/doi/10.5555/1083592.1083696

Aggarwal, C. C., & Yu, P. S. (2008). A General Survey of Privacy-Preserving Data Mining Models and Algorithms. In C. C. Aggarwal & P. S. Yu (Eds.), *Privacy-Preserving Data Mining: Models and Algorithms* (pp. 11–52). Springer US. https://doi.org/10.1007/978-0-387-70992-5_2

Albert, W., & Tullis, T. (2010). *Measuring the user experience*. Elsevier. https://www.sciencedirect.com/book/9780128180808/measuring-the-user-experience

Android Developers. (2025). *Android Debug Bridge (adb) Android Studio*. https://developer.android.com/tools/adb

Apple Inc. (2025). *Differential Privacy Overview*. https://www.apple.com/privacy/docs/Differential_Privacy_Overview.pdf

Barreno, M., Nelson, B., Sears, R., Joseph, A. D., & Tygar, J. D. (2006). Can machine learning be secure? In *ASIACCS '06, Proceedings of the 2006 ACM Symposium on Information, computer and communications security* (pp. 16–25). Association for Computing Machinery. https://doi.org/10.1145/1128817.1128824

Beresford, A. R., & Stajano, F. (2003). Location privacy in pervasive computing. *IEEE Pervasive Computing*, *2*(1), 46–55. https://doi.org/10.1109/MPRV.2003.1186725

Bilogrevic, I. (2018). Privacy in Geospatial Applications and Location-Based Social Networks. In A. Gkoulalas-Divanis & C. Bettini (Eds.), *Handbook of Mobile Data*

Privacy (pp. 195–228). Springer International Publishing. https://doi.org/10.1007/978-3-319-98161-1_8

Bødker, S., Ehn, P., Sjögren, D., & Sundblad, Y. (2000). Cooperative design perspectives on 20 years with "the Scandinavian IT Design Model": Nordic conference on Human-computer interaction. NordiCHI 2000. *Proceedings of the First Nordic Conference on Human-Computer Interaction*. https://cid.nada.kth.se/pdf/cid_104.pdf

Bott, M., & Young, G. (2012). The Role of Crowdsourcing for Better Governance in International Development. *PRAXIS*, *The Fletcher Journal of Human Security*(VOLUME XXVII), 47–70. https://ecas.issuelab.org/resources/29626/29626.pdf

Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology*, *3*(2), 77–101. https://doi.org/10.1191/1478088706qp063oa

Buildozer. (2025). *Welcome to Buildozer's documentation! — Buildozer 0.11 documentation*. https://buildozer.readthedocs.io/en/latest/index.html

Bygrave, L. A. (2014). *Data Privacy Law: An International Perspective*. Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199675555.001.0001

Calderoni, L., Palmieri, P., & Maio, D. (2015). Location privacy without mutual trust: The spatial Bloom filter. *Computer Communications*, *68*, 4–16. https://doi.org/10.1016/j.comcom.2015.06.011

Camtepe, S. A., & Yener, B. (2005). Key distribution mechanisms for wireless sensor networks: a survey. *Rensselaer Polytechnic Institute, Troy, New York, Technical Report*, 5–7. http://www.cs.rpi.edu/research/pdf/05-07.pdf

Cao, Y., Yoshikawa, M., Xiao, Y., & Xiong, L. (2017). Quantifying Differential Privacy under Temporal Correlations. In *2017 IEEE 33rd International Conference on Data Engineering (ICDE)*. https://ieeexplore.ieee.org/document/7930028

Cavoukian, A. (2009). Privacy by design: The 7 foundational principles. *Information and Privacy Commissioner of Ontario, Canada*, *5*(2009), 12. https://student.cs.uwaterloo.ca/cs492/papers/7foundationalprinciples_longer.pdf

Chatzikokolakis, K., Andrés, M. E., Bordenabe, N. E., & Palamidessi, C. (2013). Broadening the Scope of Differential Privacy Using Metrics. In E. de Cristofaro & M. Wright (Eds.), *Privacy Enhancing Technologies* (pp. 82–102). Springer. https://doi.org/10.1007/978-3-642-39077-7_5

Choi, H., Park, J [Jonghwa], & Jung, Y. (2018). The role of privacy fatigue in online privacy behavior. *Computers in Human Behavior*, *81*, 42–51. https://doi.org/10.1016/j.chb.2017.12.001

Creswell, J. W., & Creswell, J. D. (2017). *Research design: Qualitative, quantitative, and mixed methods approaches*. Sage publications. https://books.google.at/books?hl=de&lr=&id=335ZDwAAQBAJ&oi=fnd&pg=PT16&dq=Research+Design:Qualitative,+Quantitative+and+Mixed+Methods+Approaches&ots=YExWJKytnO&sig=4F_Q2X26NmHd26JjmnEByCZkqzc

Cui, L., Xie, G., Qu, Y., Gao, L., & Yang, Y. (2018). Security and Privacy in Smart Cities: Challenges and Opportunities. *IEEE Access*, *6*, 46134–46145. https://doi.org/10.1109/ACCESS.2018.2853985

Davis, F. D., Bagozzi, R. P., & Warshaw, P. R. (1989). User Acceptance of Computer Technology: A Comparison of Two Theoretical Models. *Management Science*, *35*(8), 982–1003. https://doi.org/10.1287/mnsc.35.8.982

Du, W., & Atallah, M. (2001). Secure Multi-Party Computation Problems and Their Applications: A Review And Open Problems. *Electrical Engineering and Computer Science - All Scholarship*. https://surface.syr.edu/eecs/11

Dwork, C. (2008). Differential Privacy: A Survey of Results. In M. Agrawal, D. Du, Z. Duan, & A. Li (Eds.), *Theory and Applications of Models of Computation* (pp. 1–19). Springer. https://doi.org/10.1007/978-3-540-79228-4_1

Dwork, C., & Roth, A. (2013). The Algorithmic Foundations of Differential Privacy. *Foundations and Trends® in Theoretical Computer Science*, *9*(3-4), 211–407. https://doi.org/10.1561/0400000042

El Emam, K., & Arbuckle, L. (2013). *Anonymizing health data: case studies and methods to get you started*. O'Reilly Media, Inc. https://books.google.at/books?hl=de&lr=&id=3RtRAgAAQBAJ&oi=fnd&pg=PR2&dq=%E2%80%A2%09El+Emam,+K.,+%26+Arbuckle,+L.+Anonymizing+Health+Data&ots=qIcCpIIHwd&sig=huFo98UG_gt_R0WewnRmGSietsA

Englehardt, S., & Narayanan, A. (2016). Online Tracking: A 1-million-site Measurement and Analysis. In *CCS '16, Proceedings of the 2016 ACM SIGSAC Conference on Computer and Communications Security* (pp. 1388–1401). Association for Computing Machinery. https://doi.org/10.1145/2976749.2978313

Ester, M., Kriegel, H.-P., Sander, J., & Xu, X. (1996). A density-based algorithm for discovering clusters in large spatial databases with noise. In *KDD'96, Proceedings of the Second International Conference on Knowledge Discovery and Data Mining* (pp. 226–231). AAAI Press. https://dl.acm.org/doi/10.1145/2976749.2978313

European Commission. (2025). *Data protection - European Commission*. https://commission.europa.eu/law/law-topic/data-protection_en

Fung, B. C. M., Wang, K., Chen, R., & Yu, P. S. (2010). Privacy-preserving data publishing: A survey of recent developments. *ACM Comput. Surv.*, *42*(4), 14:1-14:53. https://doi.org/10.1145/1749603.1749605

Garvin, D. A. (1987). Competing on the Eight Dimensions of Quality. *Harvard Business Review*. https://hbr.org/1987/11/competing-on-the-eight-dimensions-of-quality

Gentry, C. (2009). Fully homomorphic encryption using ideal lattices. In *STOC '09, Proceedings of the forty-first annual ACM symposium on Theory of computing* (pp. 169–178). Association for Computing Machinery. https://doi.org/10.1145/1536414.1536440

Gerber, N., Gerber, P., & Volkamer, M. (2018). Explaining the privacy paradox: A systematic review of literature investigating privacy attitude and behavior. *Computers & Security*, *77*, 226–261. https://doi.org/10.1016/j.cose.2018.04.002

GitHub. (2025). *About GitHub and Git*. https://docs.github.com/de

Golle, P., & Partridge, K. (2009). On the Anonymity of Home/Work Location Pairs. In H. Tokuda, M. Beigl, A. Friday, A. J. B. Brush, & Y. Tobe (Eds.), *Pervasive Computing* (pp. 390–397). Springer. https://doi.org/10.1007/978-3-642-01516-8_26

Google Cloud. (2025). *Cybersecurity solutions: Secops, intelligence, and cloud security*. *Google Cloud*. https://cloud.google.com/security

Google Privacy & Terms. (2025). *Privacy Policy – Privacy & Terms – Google*. https://policies.google.com/privacy

Gruteser, M., & Grunwald, D. (2003). Anonymous Usage of Location-Based Services Through Spatial and Temporal Cloaking. In *MobiSys '03, Proceedings of the 1st international conference on Mobile systems, applications and services* (pp. 31–42). Association for Computing Machinery. https://doi.org/10.1145/1066116.1189037

Hartzog, W., & Selinger, E. (2013). Big data in small hands. *Stan. L. Rev. Online*, *66*, 81. https://heinonline.org/hol-cgi-bin/get_pdf.cgi?handle=hein.journals/slro66&section=12

Hass, C. (2019). A Practical Guide to Usability Testing. In M. Edmunds, C. Hass, & E. Holve (Eds.), *Consumer Informatics and Digital Health: Solutions for Health and Health Care* (pp. 107–124). Springer International Publishing. https://doi.org/10.1007/978-3-319-96906-0_6

Hassenzahl, M. (2008). User experience (UX): towards an experiential perspective on product quality. In *IHM '08, Proceedings of the 20th Conference on l'Interaction Homme-Machine* (pp. 11–15). Association for Computing Machinery. https://doi.org/10.1145/1512714.1512717

Hevner, A. R., March, S. T., Park, J [Jinsoo], & Ram, S. (2004). Design Science in Information Systems Research. *MIS Quarterly*, *28*(1), 75–105. https://doi.org/10.2307/25148625

Hong, J. I., & Landay, J. A. (2004). An architecture for privacy-sensitive ubiquitous computing. In *MobiSys '04, Proceedings of the 2nd international conference on Mobile systems, applications, and services* (pp. 177–189). Association for Computing Machinery. https://doi.org/10.1145/990064.990087

Iachello, G., & Hong, J. (2007). End-user privacy in human–computer interaction. *Foundations and Trends® in Human–Computer Interaction*, *1*(1), 1–137. https://www.nowpublishers.com/article/Details/HCI-004

Jagatic, T. N., Johnson, N. A., Jakobsson, M., & Menczer, F. (2007). Social phishing. *Commun. ACM*, *50*(10), 94–100. https://doi.org/10.1145/1290958.1290968

Jain, A. K., Murty, M. N., & Flynn, P. J. (1999). Data clustering: a review. *ACM Comput. Surv.*, *31*(3), 264–323. https://doi.org/10.1145/331499.331504

JetBrains. (2025). *Getting started PyCharm*. https://www.jetbrains.com/help/pycharm/getting-started.html

Kavitha S, Yamini S, & Raja Vadhana P (2015). An evaluation on big data generalization using k-Anonymity algorithm on cloud. In *2015 IEEE 9th International Conference on Intelligent Systems and Control (ISCO)*. https://ieeexplore.ieee.org/abstract/document/7282237

Kivy. (2025). *Introduction — Kivy 2.3.1 documentation*. https://kivy.org/doc/stable/gettingstarted/intro.html

Kounadi, O., & Resch, B. (2018). A Geoprivacy by Design Guideline for Research Campaigns That Use Participatory Sensing Data. *Journal of Empirical Research on Human Research Ethics*, *13*(3), 203–222. https://doi.org/10.1177/1556264618759877

Kounadi, O., Resch, B., & Petutschnig, A. (2018). Privacy Threats and Protection Recommendations for the Use of Geosocial Network Data in Research. *Social Sciences*, *7*(10), 191. https://doi.org/10.3390/socsci7100191

Krumm, J. (2009). A survey of computational location privacy. *Personal and Ubiquitous Computing*, *13*(6), 391–399. https://doi.org/10.1007/s00779-008-0212-5

Li, H., Zhu, H., Du, S., Liang, X., & Shen, X. (2018). Privacy Leakage of Location Sharing in Mobile Social Networks: Attacks and Defense. *IEEE Transactions on Dependable and Secure Computing*, *15*(4), 646–660. https://doi.org/10.1109/TDSC.2016.2604383

Li, N., Li, T., & Venkatasubramanian, S. (2007). t-Closeness: Privacy Beyond k-Anonymity and l-Diversity. In *2007 IEEE 23rd International Conference on Data Engineering*. https://ieeexplore.ieee.org/document/4221659

Li, Y., & Ren, J. (2009). Preserving Source-Location Privacy in Wireless Sensor Networks. In *2009 6th Annual IEEE Communications Society Conference on Sensor, Mesh and Ad Hoc Communications and Networks*. https://ieeexplore.ieee.org/document/5168962

Lloyd, S. (1982). Least squares quantization in PCM. *IEEE Transactions on Information Theory*, *28*(2), 129–137. https://doi.org/10.1109/TIT.1982.1056489

Machanavajjhala, A., Kifer, D., Gehrke, J., & Venkitasubramaniam, M. (2007). L-diversity: Privacy beyond k-anonymity. *ACM Trans. Knowl. Discov. Data*, *1*(1), 3-es. https://doi.org/10.1145/1217299.1217302

Mascetti, S., & Bettini, C. (2007). A Comparison of Spatial Generalization Algorithms for LBS Privacy Preservation. In *2007 International Conference on Mobile Data Management*. https://ieeexplore.ieee.org/document/4417161/

McKenzie, G., Romm, D., Zhang, H., & Brunila, M. (2022). Privyto: A privacy-preserving location-sharing platform. *Transactions in GIS*, *26*(4), 1703–1717. https://doi.org/10.1111/tgis.12924

McSherry, F. D. (2009). Privacy integrated queries: an extensible platform for privacy-preserving data analysis. In *SIGMOD '09, Proceedings of the 2009 ACM SIGMOD International Conference on Management of data* (pp. 19–30). Association for Computing Machinery. https://doi.org/10.1145/1559845.1559850

Montjoye, Y.-A. de, Hidalgo, C. A., Verleysen, M., & Blondel, V. D. (2013). Unique in the Crowd: The privacy bounds of human mobility. *Scientific Reports*, *3*(1), 1376. https://doi.org/10.1038/srep01376

Muralitharan, J., & Arumugam, C. (2024). Privacy BERT-LSTM: A novel NLP algorithm for sensitive information detection in textual documents. *Neural Computing and Applications*, *36*(25), 15439–15454. https://doi.org/10.1007/s00521-024-09707-w

Narayanan, A., & Shmatikov, V. (2008). Robust De-anonymization of Large Sparse Datasets. In *2008 IEEE Symposium on Security and Privacy (sp 2008)*. https://ieeexplore.ieee.org/document/4531148

Nielsen, J. (1994a). *Usability Engineering*. Morgan Kaufmann Publishers Inc. https://dl.acm.org/doi/10.5555/2821575

Nielsen, J. (1994b). Usability inspection methods. In *CHI '94, Conference Companion on Human Factors in Computing Systems* (pp. 413–414). Association for Computing Machinery. https://doi.org/10.1145/259963.260531

Nielsen, J. (2000). *Why You Only Need to Test with 5 Users*. *Nielsen Norman Group*. https://www.nngroup.com/articles/why-you-only-need-to-test-with-5-users/

Nielsen, J., & Landauer, T. K. (1993). A mathematical model of the finding of usability problems. In *CHI '93, Proceedings of the INTERACT '93 and CHI '93 Conference on Human Factors in Computing Systems* (pp. 206–213). Association for Computing Machinery. https://doi.org/10.1145/169059.169166

Nissenbaum, H. (2010). Privacy in Context: Technology, Policy, and the Integrity of Social Life. *Bibliovault OAI Repository, the University of Chicago Press*. https://crypto.stanford.edu/portia/papers/privacy_in_context.pdf

Nissim, K., Stemmer, U., & Vadhan, S. (2016). Locating a Small Cluster Privately. In *Proceedings of the 35th ACM SIGMOD-SIGACT-SIGAI Symposium on Principles of Database Systems* (pp. 413–427). ACM. https://doi.org/10.1145/2902251.2902296

Ohm, P. (2009). *Broken Promises of Privacy: Responding to the Surprising Failure of Anonymization*. Social Science Research Network. https://epic.org/wp-content/uploads/privacy/reidentification/ohm_article.pdf

OK, E. (2025). Privacy Concerns and Practices on Social Networking Sites: An Analysis of User Perspectives. https://www.researchgate.net/publication/389099385_Privacy_Concerns_and_Practices_on_Social_Networking_Sites_An_Analysis_of_User_Perspectives

onlinesicherheit.at. (2025). *ISO/IEC 27001 - Anforderungen an Informationssicherheits-Managementsysteme*. https://www.onlinesicherheit.gv.at/Themen/Experteninformation/Normen-und-Standards/ISO-IEC-27000-ISMS-Normenreihe/ISO-IEC-27001-Informationssicherheits-Managementsysteme-ISMS.html

Österle, H., Becker, J., Frank, U., Hess, T., Karagiannis, D., Krcmar, H., Loos, P., Mertens, P., Oberweis, A., & Sinz, E. J. (2011). Memorandum on design-oriented information systems research. *European Journal of Information Systems*, *20*(1), 7–10. https://doi.org/10.1057/ejis.2010.55

Parasuraman, A. P., Zeithaml, V., & Berry, L. (1988). SERVQUAL A Multiple-item Scale for Measuring Consumer Perceptions of Service Quality. *Journal of Retailing*, *64*, 12–40. https://www.researchgate.net/publication/200827786_SERVQUAL_A_Multiple-item_Scale_for_Measuring_Consumer_Perceptions_of_Service_Quality

Pingley, A., Zhang, N., Fu, X., Choi, H.-A., Subramaniam, S., & Zhao, W. (2011). Protection of query privacy for continuous location based services. In *2011 Proceedings IEEE INFOCOM*. https://ieeexplore.ieee.org/document/5934968

Polzin, F., & Kounadi, O. (2021). Adaptive Voronoi Masking: A Method to Protect Confidential Discrete Spatial Data. *LIPIcs, Volume 208, GIScience 2021*, *208*, 1:1-1:17. https://doi.org/10.4230/LIPICS.GISCIENCE.2021.II.1

Python. (2025). *The Python Tutorial*. https://docs.python.org/3/tutorial/index.html

Rao, J., Gao, S., Kang, Y., & Huang, Q. (2020). *LSTM-TrajGAN: A Deep Learning Approach to Trajectory Privacy Protection*. https://www.researchgate.net/publication/342302301_LSTM-TrajGAN_A_Deep_Learning_Approach_to_Trajectory_Privacy_Protection https://doi.org/10.48550/arXiv.2006.10521

Reay, I., Beatty, P., Dick, S., & Miller, J. (2013). Privacy policies and national culture on the internet. *Information Systems Frontiers*, *15*(2), 279–292. https://doi.org/10.1007/s10796-011-9336-7

Reichheld, F. (2004). The One Number you Need to Grow. *Harvard Business Review*, *81*, 46-54, 124. https://www.nashc.net/wp-content/uploads/2014/10/the-one-number-you-need-to-know.pdf

Richards, N., & Hartzog, W. (2020). A Relational Turn for Data Protection? *European Data Protection Law Review*, *6*(4), 492–497. https://doi.org/10.21552/edpl/2020/4/5

Roesner, F., & Kohno, T. (2013). Securing embedded user interfaces: Android and beyond. In *SEC'13, Proceedings of the 22nd USENIX conference on Security* (pp. 97–112). USENIX Association. https://dl.acm.org/doi/10.5555/2534766.2534776

Rosson, M. B., & Carroll, J. M. (2002). *Usability engineering: scenario-based development of human-computer interaction*. Morgan Kaufmann. https://books.google.at/books?hl=de&lr=&id=sRPg0IYhYFYC&oi=fnd&pg=PP2&dq=Usability+Engineering+Scenario-Based+Development+of+Human-Computer+Interaction&ots=mJJt2dRGCM&sig=SWm1UU0dXQvbt24IygsY_MLOQu8

Rothstein, M. A. (2010). Is Deidentification Sufficient to Protect Health Privacy in Research? *The American Journal of Bioethics*, *10*(9), 3–11. https://doi.org/10.1080/15265161.2010.494215

Samarati, P., & Sweeney, L. (1998). Protecting privacy when disclosing information: k-anonymity and its enforcement through generalization and suppression. https://dataprivacylab.org/dataprivacy/projects/kanonymity/paper3.pdf

Sandhu, R. S., Coyne, E. J., Feinstein, H. L., & Youman, C. E. (1996). Role-based access control models. *Computer*, *29*(2), 38–47. https://doi.org/10.1109/2.485845

Seidl, D. E., Jankowski, P., Clarke, K. C., & Nara, A. (2020). Please Enter Your Home Location: Geoprivacy Attitudes and Personal Location Masking Strategies of Internet Users. *Annals of the American Association of Geographers*, *110*(3), 586–605. https://doi.org/10.1080/24694452.2019.1654843

Shokri, R., Theodorakopoulos, G., Le Boudec, J.-Y., & Hubaux, J.-P. (2011). Quantifying Location Privacy. In *2011 IEEE Symposium on Security and Privacy*. https://ieeexplore.ieee.org/document/5958033

Solove, D. J. (2008). *Understanding Privacy*. Social Science Research Network. https://ssrn.com/abstract=1127888

Sommerville, I. (2016). *Software engineering* (Tenth edition). *Always learning*. Pearson. https://dn790001.ca.archive.org/0/items/bme-vik-konyvek/Software%20Engineering%20-%20Ian%20Sommerville.pdf

Sweeney, L. (2002). K-ANONYMITY: A MODEL FOR PROTECTING PRIVACY. *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, *10*(05), 557–570. https://doi.org/10.1142/S0218488502001648

Torre, D., Chennamaneni, A., & Rodriguez, A. (2023). Privacy-Preservation Techniques for IoT Devices: A Systematic Mapping Study. *IEEE Access*, *PP*, 1. https://doi.org/10.1109/ACCESS.2023.3245524

Ubuntu. (2025). *Ubuntu Desktop Leitfaden*. https://help.ubuntu.com/stable/ubuntu-help/index.html

van Rossum, G., & Boer, J. de (1991). Interactively testing remote servers using the Python programming language. *CWI Quarterly*, *4*(4), 283–304. https://ir.cwi.nl/pub/18204

Voigt, P., & Bussche, A. (2017). *The EU General Data Protection Regulation (GDPR): A Practical Guide*. https://link.springer.com/book/10.1007/978-3-319-57959-7 https://doi.org/10.1007/978-3-319-57959-7

Wasserman, L., & Zhou, S. (2010). A Statistical Framework for Differential Privacy. *Journal of the American Statistical Association*, *105*(489), 375–389. https://doi.org/10.1198/jasa.2009.tm08651

Wernke, M., Skvortsov, P., Dürr, F., & Rothermel, K. (2014). A classification of location privacy attacks and approaches. *Personal and Ubiquitous Computing*, *18*(1), 163–175. https://doi.org/10.1007/s00779-012-0633-z

WienMobil. (2025). *Privacy Policy*. https://www.wienmobil.at/en/services/content/privacy-policy

Wright, D., & Hert, P. de. (2012). Introduction to Privacy Impact Assessment. In P. Casanovas, G. Sartor, D. Wright, & P. de Hert (Eds.), *Privacy Impact Assessment* (Vol. 6, pp. 3–32). Springer Netherlands. https://doi.org/10.1007/978-94-007-2543-0_1

Xiao, Y., Xiong, L., Zhang, S [Si], & Cao, Y. (2017). LocLok: location cloaking with differential privacy via hidden markov model. *Proc. VLDB Endow.*, *10*(12), 1901–1904. https://doi.org/10.14778/3137765.3137804

Yuan, E., & Tong, J. (2005). Attributed based access control (ABAC) for Web services. In *IEEE International Conference on Web Services (ICWS'05)*. https://ieeexplore.ieee.org/document/1530847

Zang, H., & Bolot, J. (2011). Anonymization of location data does not work: a large-scale measurement study. In *MobiCom '11, Proceedings of the 17th annual international conference on Mobile computing and networking* (pp. 145–156). Association for Computing Machinery. https://doi.org/10.1145/2030613.2030630

Zhang, S [Shaobo], Li, X., Tan, Z., Peng, T., & Wang, G. (2019). A caching and spatial K-anonymity driven privacy enhancement scheme in continuous location-based services. *Future Generation Computer Systems*, *94*, 40–50. https://doi.org/10.1016/j.future.2018.10.053

Zhao, J., Qiu, J., Zhou, Y.-W., Hu, X.-J., & Yang, A.-F. (2020). Quality disclosure in the presence of strategic consumers. *Journal of Retailing and Consumer Services*, *55*, 102084. https://doi.org/10.1016/j.jretconser.2020.102084

Directory of resources

In accordance with the guidelines on independence and transparency in scientific work, the following overview lists all artificial intelligence-based aids and tools used during the work process. The table illustrates the utilisation of the respective aids and the manner in which their results were integrated into the work.

Table 12: Directory of resources

Resources	*Intended Purpose*	*Affected parts*	*Examples*
ChatGPT (OpenAI)	Phrasing assistance, Structuring, Technical explanation	Spelling and grammar checks, rephrasing,	Content has been revised, adapted, and critically reviewed.
DeepL & Deepl Write (DeepL SE)	Elaboration of language	Improvement of phrasing, syntax, and comprehensibility	Proposals were reviewed individually and partially adopted.
GitHub Copilot (GitHub, Inc.)	Technical explanation	Assistance with Python error messages	All content has been revised, adapted and critically reviewed.

All findings, analyses and evaluations are based on our own work and were developed independently. AI was not used to generate scientific content or to complete tasks independently. Instead, it served exclusively as a tool to support the work process.

Appendix - Questionnaire DeLocator

Instructions for Participants

By completing this Survey, you consent to participate in this research survey and agree to the anonymous publication of your survey responses for the master thesis of Aldina Kovacevic. All personal information of the participants will be held confidential and will not be made publicly available at any time.

Please complete the tasks below in the app and then answer the following questions as honestly and thoroughly as possible. Your feedback will help me improve the usability of the app.

Duration: ~10 minutes

---

1. First Impression

1. **What is the purpose of this app?**
   Describe in 1–2 sentences what the app does.

2. **How do you like the look and feel?**
   (colours, typography, icons, overall layout)

---

2. Specific Tasks

Please complete the following tasks in order. For each task, rate how easy or difficult it was.

Task	Ease Rating (1 = very difficult, 5 = very easy)	Comments (optional)
1. Enter an address (e.g., "Sample Street 1, 1010 Vienna") in the text field and anonymize it.	[1 2 3 4 5]	
2. Copy the anonymized result to the clipboard.	[1 2 3 4 5]	
3. Save the anonymized location.	[1 2 3 4 5]	
4. Open the list of saved locations and delete one entry.	[1 2 3 4 5]	

3. Post-Task Questions

1. **Which task was the easiest? Why?**

2. **Which task was the most difficult? Why?**

3. **Were any labels or icons unclear during the tasks? If yes, which ones?**

4. **Did you wish for additional feedback after any task? If so, for which task and what kind of feedback?**

---

4. Navigation

1. **How did you find the "Anonymize location" function?**

2. **Were there any points where you weren't sure what to do next?**
   If yes, please specify where.

---

5. Feedback & Error Messages

3. **Did you encounter any error messages or delays during the tasks?**
   – Were the messages understandable and helpful?

4. **Do you think the notification function is necessary?**
   – Why or why not?

5. **Would you prefer a different type of feedback (e.g., sound alert, visual highlight)?**

---

6. Overall Satisfaction & Wishes

6. **What did you like most about the app?**

7. **What feature would you like to see added to improve the app?**

---

7. Closing & Recommendation

8. **Would you recommend this app to your friends?**
   – Why or why not?

9. **Would you use this app yourself?**
   – Why or why not?

10. **Any additional comments or experiences you'd like to share?**

Thank you for contributing to this study!

I look forward to sharing my findings with you!