SoK: Managing risks of linkage attacks on data privacy

Jovan Powar
Jovan.Powar@cst.cam.ac.uk
University of Cambridge
Cambridge, UK

Alastair R. Beresford
Alastair.Beresford@cst.cam.ac.uk
University of Cambridge
Cambridge, UK

ABSTRACT

Novel attacks on dataset privacy are usually met with the same range of responses: surprise that a route to information gain exists from information previously thought to be safe; disputes around the viability or validity of the attack in real-world contexts; and, in the case of the computer science community, a drive to produce techniques that provably protect against the new class of attack.

The result is a disjointed landscape with no shared approach to modelling threats to dataset privacy, and a toolbox of technically complex systems whose guarantees come with narrow assumptions and whose application in real-world contexts is hard to achieve.

In this paper we aim to understand these issues by charting the history of dataset privacy attacks and systematising breaches through the lens of data linkage. We show how identification or information gain on a dataset's subjects can be expressed as data linkage, and use this to present a taxonomy of threat models which we apply to ninety-four attacks from across the literature.

Our work demonstrates that dataset privacy must be approached first as a risk management problem, rather than one of strict guarantees, an approach which aligns well with law and practice. Our taxonomy of attacker intents provides a coherent language for expressing the wide variety of threat models in dataset privacy, and a framework for understanding how risks identified under one model can be understood within another. We also present insights around the factors that affect the feasibility and severity of attacks, and proposals for practical techniques that can be used for risk appraisal and management by practitioners, researchers, and regulators alike.

KEYWORDS

data sharing, identification, reidentification, anonymity, risk appraisal, data protection, record linkage, membership inference, attribute inference, threat modelling

1 INTRODUCTION

In 1997 Latanya Sweeney demonstrated the inadequacy of contemporary anonymisation techniques by retrieving the health records of the then-Governor of Massachussetts from a supposedly sanitised dataset. The breach motivated a rethinking of dataset privacy, including a new generation of privacy legislation.

In the decades since, we have seen a parade of dataset privacy breaches—from Narayanan and Shmatikov's attack on the anonymity of Netflix movie ratings to the outing of a Catholic priest via commercially available advertising data. Each new breach brings not only surprise but criticism over its validity or utility as an instructive example.

In 2012 the legal scholar Paul Ohm dubbed this issue 'the surprising failure of anonymisation' and argued for new approaches to anonymity. While the legal community accepted this challenge, the computer science community has continued to search for perfect, provable guarantees against unwanted disclosure.

The result is an approach to dataset privacy that is not fit for purpose: we lack a coherent threat modelling approach, which leaves us flat-footed in the face of novel attacks. Too often research focuses on narrowly applicable cryptographic protocols or statistical transformations which often severely limit data utility.

We propose a new, risk-based approach to dataset privacy based on two insights. First, the route to information gain in all dataset privacy attacks can be understood as data linkage—the combination of data with other sources of information. And secondly, we can rarely guarantee that there exists no potential linkage that breaks the data subject's privacy—instead we must use a risk-based approach to minimise the likelihood of those unwanted linkages.

In this paper we systematise the space of attacks on dataset privacy by describing the linkage process performed, and capturing the nature of an attack in terms of the descriptive record produced. The result is a threat modelling framework that yields a logically related taxonomy of attacker intents, which we apply to 94 dataset privacy attacks across the literature. From this systematisation, we extract a series of insights and recommendations for appraising risks of linkage attacks, and discuss how to improve upon the current practical state of the art in dataset privacy protection.

Our contributions are as follows:

- a critical analysis of the history of dataset privacy breaches;
- a definition of linkage attacks and a threat modelling framework that captures attacks in terms of the intended linkage;
- a categorisation of 94 dataset privacy attacks across technical and non-technical literature in this framework;
- an argument that protecting dataset privacy must be approached as a risk management exercise, with reference to legal framings;
- a discussion of the challenges to appraising linkage attack risk and the impropriety of a guarantees-focused approach;
- and recommendations for risk-based approaches to appraising threats to dataset privacy.

2 PRIOR PROBLEM DEFINITIONS

We begin by charting how the object of data protection law—personal data—as a concept has evolved over time away from the notion of identifiers toward a question of linkage. With reference to legislation and accompanying guidance, we discuss how the legal community approaches privacy threat modelling as a risk exercise, and to what extent the law formalises breaches of privacy.

Proceedings on Privacy Enhancing Technologies 2023(2), 97–116

https://doi.org/10.56553/popets-2023-0043

We then discuss the movement in technical privacy work that has mirrored this shift toward understanding identification and privacy attacks more generally as information gain through linkage, and present a review of attacks and their stated threat models.

2.1 Legal

The majority of data-related legal burdens arise when data allows the identification of an individual. Historically, identification was thought of in terms of 'identifiers'—data points that unambiguously refer to a natural person, subject to eligibility criteria taken from cultural norms or by decree. In the twentieth century, mass data collection and retention—particularly by governments in social welfare initiatives—became commonplace, and digital technologies allowed processing at scale and with granularity, greatly increasing the salience of privacy risks to individuals. This precipitated the task of data protection, and legislation was required to define which data should be subject to this protection.

2.1.1 PII: Personally Identifying Information. Early data protection legislation concerned the concept of Personally Identifiable Information (*PII*). Some narrowly scoped data protection laws, such as HIPAA in the United States and the California data breach law SB1386, prescribed datatypes that constitute PII, such as names alongside social security numbers or credit card numbers.

Such definitions described the presence of 'identifying information' which is *globally unique*, providing a perfect one-to-one mapping of data points to individuals. Datatypes declared in scope were limited to a few whose practical use in identification (linking against a database of social security numbers, for example) is already known. It is clear at first glance that a prescriptive, datatype-driven definition of personal information is insufficient for practical purposes, as it would be impossible to enumerate all data items that might be globally unique (or locally unique in some constrained context) to an individual. A more general definition of PII given by NIST [82] includes 'information that can be used to distinguish or trace an individual's identity' and 'any other information that is linked or linkable to an individual'.

This definition establishes a principle, rather than a prescription; instead of defining identification by reference to known datatypes and their usage, it considers the *potential use* of the data in distinguishing an individual. This definition's key contribution is the introduction of *data linkage* as the means of indirect identification. NIST defines *linkable information* as 'about or related to an individual for which there is a possibility of logical association with other information about the individual.'

This language—'can be used', 'linkable'—opens the door to new datatypes and new means of inferring logical relationships between data, which is essential in the modern landscape of novel datatypes and unexpected correlations. Despite the broader scope given by the NIST definition, the association of the term *PII* to specific datatypes has led to its use waning in recent years in favour of *personal data*, definitions of which are more reliably generalist and principled.

2.1.2 GDPR and personal data. The GDPR is the most internationally influential data protection regulation of recent years [61]—its economic centrality to much of the world's data processing has sparked a generation of data protection laws, with many jurisdictions following its lead in scope and definitions.

Finck and Pallas [50] provide a comprehensive overview of the definitional intricacies of the GDPR, including the extent to which it formalises the notions of singling out, linkability, and inference. We summarise the relevant material here.

The scope of the GDPR is all *personal data*, defined as 'any information relating to an identified or **identifiable** natural person ('data subject'); an identifiable natural person is one who **can be identified, directly or indirectly**...' (emphasis added)

As in the NIST definition of PII, note the centrality of the *potential use* of data in 'directly or indirectly' identifying an individual. This phrasing continues the definitional shift away from concrete datatypes—'identifiers' and pseudonyms, the tools of 'direct' identification—and towards the act of identification itself.

Recital 26 clarifies the definition of indirect identification: 'Personal data which have undergone pseudonymisation, which could be **attributed to a natural person by the use of additional information** should be considered to be information on an identifiable natural person.' This mirrors NIST's definition of linkage—the assertion of a relationship between a given set of data and a natural person with the aid of logically related information.

2.1.3 Data protection is a risk management exercise. While the task of seeking out direct identifiers is somewhat straightforward, by including indirect identification via linkage against arbitrary additional information, the GDPR imposes a difficult task on practitioners seeking to determine whether they hold 'personal data'.

On this question Recital 26 advises that '**account should be taken of all the means reasonably likely to be used**, such as singling out... to identify the natural person directly or indirectly... of all objective factors, such as the costs of and the amount of time required for identification, ... the available technology at the time of the processing and technological developments.'

In computer security parlance, this advice outlines the space of *threat models* that the GDPR requires the practitioner to consider. This space is undeniably vast, and a comprehensive appraisal of all potential third parties, the data they might use in linkage, and their technical and economic capabilities, is clearly an intractable problem. As such, the term 'reasonably likely' does significant heavy lifting here, and the consensus within the legal community is that the GDPR should be interpreted as imposing a risk management problem upon practitioners. This contrasts with the prior NIST approach to PII which required total risk mitigation.

Concrete risk appraisal tests are still a matter of debate amongst the legal community. For example, there is disagreement on the scope of the GDPR; while most national authorities have issued guidance instructing practitioners to reduce risks of reidentification to within acceptable levels of harm to individuals, the Article 29 Working Party takes a more hardline stance, proposing a skeleton test that requires data to be reduced to 'anonymous information'—i.e. completely eliminating the risk of reidentification.

Questions of scope notwithstanding, the practical burdens imposed by the GDPR where it does apply are explicitly risk management exercises, such as data minimisation or impact assessments.

2.1.4 Key takeaways. There are two key lessons to take from legal efforts to formalise the task of data protection. First, individual privacy risks associated with data arise from its potential to be linked against other data in a way that identifies a subject. Second, the task of data protection is fundamentally a risk management exercise—the practice of identifying potential linkages, appraising the risks they pose, and either mitigating or eliminating those risks.

2.2 Technical

Technical communities have understood for decades that protecting the privacy of dataset subjects is more than a matter of 'direct identifiers' or PII [90]. Dataset privacy work falls into the broad categories of: demonstrating that individual-level signals are retained even after measures that were thought to obscure them [45, 99, 100, 121]; showing statistical properties of certain categories of dataset or datatypes, particularly with regard to the unicity of subjects both within the dataset and globally [35, 37, 49, 57, 132]; or the construction of novel attacks which either identify subjects [85, 88, 133] or allow inferences to be made about them [105].

Where proposed attacks do not describe an attacker with access to a subset of the dataset under attack, or some limited profile of a target, they often explicitly cite linkage against side information (or auxiliary information) in their threat model as the route to privacy breach. For example, Shokri et al. [112] define an attacker in terms of two factors: the knowledge they have to support an attack and the nature of privacy breach they seek. A major theme in the past decade has been attacks that link seemingly innocuous information to enable more significant privacy breaches [55, 85, 87, 114]. The phrasing of linkage attacks is now common, as is 'linkability' to mean the propensity of data to retain signals that might be found in another dataset (the 'linkage set').

2.2.1 Technical guarantees and metrics. Technical research is prone to trends towards ever more concrete and provable guarantees, even for problems that are known to be intractable in practice. It has been known at least since 2002 [48] that the signals contained within rich datasets are largely unknown *a priori* (by definition, else there would be no need for data mining), and so the enumeration of all semantic connections to other sources of information (linkable data) is intractable. We discuss this problem in more depth in Section 6 but mention it here to clarify that while there is a wealth of good research into strong privacy guarantees—as Wagner and Eckhoff have thoroughly surveyed [126]— practical data protection problems are rarely solved by these approaches, either because the assumptions and threat models are too tightly specified or because the solutions impose unworkable constraints.

2.2.2 Threat models. There are some general terminologies used to describe attacks on dataset privacy, which we can categorise into the following threat model types.

Singling out. Following the definition of singling out given by the GDPR, these attacks pick an individual's data out of a dataset, perhaps by partially reversing aggregation or by demonstrating the unicity of individual-level data [35, 37, 49]. Notable work on unicity has studied location data—for example, Golle and Partridge [57] showed that the census block of a person's home and workplace is sufficient to uniquely identify the majority of US workers, while Farzanehfar et al. [49] showed that four points of location information are sufficient to uniquely specify 93 percent of subjects in a population of 60 million people.

Reidentification/deanonymisation. Reidentification (sometimes deanonymisation) attacks [33, 55, 88, 89, 104, 114] show that the identity of a data subject can be inferred from a dataset—this may be by demonstrating a means of linking the dataset to an 'identifier' for a person, or by showing that some set of attributes can be retrieved about the person that are sufficiently unique to be used as an identifier—a quasi-identifier (**QID**). Fingerprinting attacks demonstrate the construction of a profile about an entity that can be used to infer their presence in another dataset—in a sense this is the construction of a QID, but one that is often a derived, novel datatype [107]. Some attacks construct a fingerprint that is reliably reconstructable every time the target's data is present, and has a low chance of being computed for a different entity [133], while others may be more probabilistic, with some amount of uncertainty when the same fingerprint is computed from subsequent data [33].

Membership inference. These attacks determine whether or not data for a particular subject is in a given dataset. This has been shown to be possible after aggregation [100] and perturbation [99]. The problem has been formalised a number of times, for example as a risk parameter by Nergiz et al. [92], or with a game-based definition by Pyrgelis et al. [99]. The problem has been studied in many contexts, including DNA [67] and microRNA [12].

Attribute inference. Other attacks do not explicitly attempt to locate or extract the information of a subject from a dataset, but aim to gain information on a person. That person may be a subject of the dataset [45] or may not—for example, it has been shown that location visit information can be linked against aggregate location traces to infer the target's ethnicity [105].

2.2.3 Record linkage. Gkoulalas-Divanis et al. describe the evolution of record linkage [54], culminating in the current generation of privacy-preserving techniques. Record linkage [62], data matching [27], or entity reconciliation [28] is the task of asserting the equivalence of QIDs found in separate tables or databases that hold data in a record-based format. This assertion serves to combine information from multiple sources that describe the same entity.

This task has been well studied in in medical and social science since 1946 [42]. The use of record linkage as an automated or speculative process has gained traction in 'big data' [69, 91]. As an information retrieval technique, the quality of a linkage is described in terms of *precision* and *recall.*

Probabilistic record linkage [108] (or fuzzy matching) does not require correspondences to be proven, but incorporates degrees of belief and uncertainty into a proposed linkage. This enables data to be matched without a strict rule-based model, and instead matching weights can be trained from sample data. Probabilistic record linkage allows the reconciliation of data items where a linkage cannot be perfectly proven, at the cost of introducing uncertainty.

3 A BRIEF HISTORY OF REIDENTIFICATIONS IN THE WILD

In this section we discuss a few notable instances of reidentification or identity-related privacy breaches, and how each illuminates key tensions and disagreements surrounding the definition of identification. The purpose of this comparison is to show the variety of factors that one might consider when evaluating an 'identification'.

3.1 Governor William Weld

In one of the first high-profile health data privacy scares, in 1997 Latanya Sweeney [118] identified then-Governor of Massachussetts William Weld within a state health insurance dataset, which had been released publicly after being stripped of direct identifiers. The reidentification was a key motivating example for U.S. de-identification rules in the HIPAA legislation. The attack linked the dataset's gender, date of birth, and ZIP code fields against a voter registration list for the Cambridge, MA area. Following this linkage, Weld was identified by finding the record for his well-publicised hospitalisation following a collapse whilst giving a speech.

Barth-Jones in 2012 published a critical analysis [16] of this breach and whether the 'astonishing ease' of reidentification holds in general and in practice, arguing that the ability to retrieve provably unique QID for Governor Weld was not a generalisable attack.

Barth-Jones describes 'the myth of the perfect population register', arguing that an attacker could not guarantee that the average resident of the Cambridge area was present within the dataset, which covered a subset of the local population. Similarly for the linkage set, not all residents of the area were registered to vote, introducing uncertainty into the correspondence between the datasets.

In the case of Governor Weld, his presence in both the target and linkage datasets were verifiable—his public hospitalisation on a date in the records' scope guaranteed his membership, and photo-ops showing him casting a ballot at past elections guaranteed that he was registered to vote, and so recorded in the linkage set. Each of these pieces of information reduces the reidentification problem: the question of identifying the Governor's records then only hinges upon whether he can uniquely be identified within the datasets by the quasi-identifier made up of gender, date of birth, and ZIP code.

3.1.1 Takeaways. ***Proving comembership.*** The preconditions that enabled Weld's reidentification and Barth-Jones' 'myth of the perfect population register' demonstrate how the feasibility of reidentification hinges upon proving that a given target is present within the target and linkage datasets. In cases where a globally unique identifier of a subject cannot be constructed, we must then calculate the probability of successful reidentification based on local uniqueness of records within both target and linkage datasets, degree of comembership across datasets, and the extent to which each dataset spans the potential space of subjects.

Targeted vs. opportunistic attacks. It is also important to note that this was a targeted attack—Sweeney had a particular subject in mind. The success of an attack under a different intent, such as to reidentify any person with no prior preference as to their identity, would be influenced by a different calculus: instead of matching a preselected quasi-identifer, the attacker's task is to piece together information to construct a globally-unique identifier. This is in one sense an easier task than a targeted attack and in another a more difficult one: Census or voter roll data can be trivially trawled for highly unique entries that are likely to be globally unique, even given only partial coverage of the dataset, and then this identifier is linked to the medical dataset; but the resulting identifier, while globally unique, may have low utility (a gender-ZIP-date of birth tuple is less valuable to the attacker than a name or email address).

Membership inference and local vs. global unicity. As Barth-Jones points out, the QID retrieved for William Weld may not have truly been globally unique, and an attack that requires a proof of global unicity is unlikely to be generally applicable.

In this case, retrieval of a globally unique QID is not necessary for the reidentification to hold; instead, the attack was successful by linking two *locally* unique (within each dataset) QIDs, paired with the side information that Weld was represented in each dataset.

Proving this fact is known as the **membership inference** problem [99, 111]. As the membership inference problem for Weld was simple for each dataset, finding locally unique QIDs that matched a known profile of his was sufficient to unambiguously tie the data to him. Relaxing either of these conditions—that membership was known in each dataset and that Weld's record happened to be locally unique—would weaken the claim that the data retrieved is unique to Weld. Each of these relaxations would reduce the attack to a probabilistic one. This intuition was the basis for Sweeney's later introduction of k-anonymity [119]. This example highlights that retrieving a globally unique data item is often not strictly necessary—a successful reidentification simply shows that the data in question was in fact collected on a particular individual.

Repeatability. The particulars of Weld's case do not hold for every, or even most subjects of the patient database. Even in another case where a locally unique QID could be matched across datasets, this would not constitute a reidentification unless the data assembled could be usefully interpreted—in this case, linked again to information held on a particular person.

This should not be taken as a sign that this attack is not salient, but that the success criteria for a reidentification can vary. The particular chain of inferences performed by Sweeney might not be repeatable for a majority of the data subjects, but it may be sufficient to pick out a few, high-value targets.

3.2 AOL search query logs

In 2006, AOL released a text file containing twenty million search strings for over 650,000 users over a 3-month period. The data was made available publicly for research purposes. Entries in the dataset were pseudonymised, such that all queries from one data subject carried a unique identifier for that individual. The privacy breach came via the presence of PII in search queries, most notably in the case of subject no. 4417749, Thelma Arnold [14]. Journalists were able to discover Ms Arnold's name, location, and marital status amongst other information. In a handful of other cases, users' sexual proclivities or personal interests could be discovered.

The main criticism of this breach as an instructive case study is the difficulty and maximum scope of the reidentification; of the 650,000 users, only very few have been reidentified, and even then by efforts that would not be considered scalable. Arguments that the scale of the breach was overblown point to the fact that it would take an especially motivated actor to reconstruct a profile for a subject, and even then the attacker would have no certainty that the subject could be reidentified or be of any particular value.

Nonetheless, it is clear that there is the capacity for harm to a user, even if they are a proverbial needle in a haystack. This distribution of victims illustrates a key quandary in data release: the reidentification of a subject is a 'black swan event'—highly difficult to predict, and with impact vastly disproportionate to its likelihood. This poses a difficult risk mitigation question for the data owner who must not only consider the average case of their subjects, but protect against unknown, uncommon worst-case scenarios.

3.2.1 Takeaways. ***Target scale and selection.*** This example motivates two distinctions around threat models: between attacks which reidentify subjects in the average case and/or at scale, versus only a small subset of subjects; and around the attacker's target, since a given attack may be likely to fail if targeted at a preselected subject, but could succeed if the attacker does not care who they reidentify.

3.3 The Netflix Prize

In 2007, Narayanan and Shmatikov [88] reidentified individual users within a dataset published by Netflix as part of a recommendation algorithm competition. The dataset contained film ratings (rating and date of rating) from 480, 189 users over 17, 770 movies. Each unique user and movie was given a unique integer pseudonym, which were consistent across entries.

Narayanan and Shmatikov reidentified subjects by linking the Netflix dataset against publicly available movie ratings from IMDB. Their algorithm was shown to be robust against incomplete or imprecise information in both the Netflix dataset and the linkage set; this result was one of the first notable examples of how supposedly sanitised or anonymised rich microdata retains reidentification risk through the presence of external sources of matching microdata.

The chief criticisms of this case as a cautionary tale of reidentification risk are twofold: first that it may be rare to find a linkage dataset with sufficient richness and similarity in datatype; and even if one exists, this may have been a case where the data subjects were uncommonly likely to be members of both—the type of Netflix user who rates many of the films they have watched is likely to also have an IMDB account where they do the same.

3.3.1 Takeaways. ***Linkage data availability.*** In data publishing, the presence of rich, correlated datasets which are not collected contemporaneously with the dataset in question is a rapidly growing contributor to reidentification risks.

Instead of data which was collected or derived alongside the published data, which could be used to directly reverse the anonymisation function, the trend towards high-dimensionality microdata provides globally-unique 'behavioural fingerprints' for individuals, which can be found and linked across datasets.

This is a difficult risk to appraise—data publishers must anticipate which implicit signals might be embedded in their dataset, and estimate the availability of other datasets now and in the future which might capture those same signals from the same subjects.

Unicity. The sparsity and high dimensionality of the Netflix ratings microdata yielded a high degree of unicity—as dimensionality and sparsity increases, local unicity of a record will approximate global unicity (i.e. unique amongst all possible subjects). This also ensures that even a partial overlap between datasets with sparse microdata has a high likelihood to truly capture comembership.

3.4 Data brokerage reidentifications

Our most recent examples are a set of publicised privacy breaches due to the data brokerage industry [32], where highly detailed microdata, including cellphone location data, is sold into complex data supply chains and aggregated by companies who provide 'identity resolution' services which can easily be used as tools of privacy invasion. One company was found to offer one-shot location services for as low as $4.95, and real-time tracking for $12.95 [30].

The trade in microdata includes the targeted advertising industry, which collects detailed user profiles, valuable to advertisers for targeting and attribution. This data collection has been shown to be almost ubiquitous in mobile apps and websites: third party libraries which collect targeting data are common in apps, to the point that their inclusion is a sufficient business model for 'free' apps [123].

This infrastructure is exploitable by malicious actors—Vines et al. [125] described an online attack in which buying targeted ads could be used as a means of extracting the profiling information held on data subjects, including but not limited to location information.

The New York Times demonstrated the viability of an offline attack after obtaining a database containing location traces of individuals involved in the storming of the U.S. Capitol on January 6, 2021, from which they were able to establish identities and potentially incriminating evidence.

In another case, a Catholic official was identified in data collected through the app Grindr [20, 98]; location traces were linked against his known residence and workplace (among other signals) to identify him, and showed his attendance at various gay bars. In a response to the reporting of this case, Grindr claimed the reidentification was 'infeasible from a technical standpoint and incredibly unlikely to occur'. Despite this claim, the official resigned his position—we argue this case demonstrates the viability 'in the wild' of the ad-based attack class described by Vines et al. [125].

3.4.1 Takeaways. ***Embedded signals.*** The Capitol riot dataset demonstrates that simple heuristics on the regularity of human behaviour can be used to invade privacy—location traces which did not contain a phone number or name, could be mined for home addresses, inferable by a lack of movement overnight, and workplaces. This echoes Golle and Partridge [57] who demonstrated that the anonymity set of U.S. residents given the census tracts of their home and workplace is 5 for 24.5% of the working population, and 2 or fewer for 7.4%. Even aggregate location data is not immune to regularity heuristics—Xu et al. demonstrated [131] reconstruction of location traces at scale from aggregate cell tower ping data.

Increasing availability of 'infeasible' side info. Criticisms of 'incredibly unlikely cases' dismiss the attacks demonstrated or theorised on the grounds that they cannot be perpetrated at scale, for a significant proportion of users, and require access to an unpredictable source of side information. However, this conflates an attack's *repeatability* and *transferability*—while the exact same attack using the same auxiliary dataset might not reidentify other subjects, it may be trivial to perform the same attack pattern using some other auxiliary information to identify another subject.

In some cases, as the availability of datasets that could be linked against increases, we expect that a data source is likely to be vulnerable to a range of partial attacks, which eventually form a patchwork that provides mass reidentification of the source's subjects.

3.5 Other takeaways

Validity of hypothesised attacks. There is often hand-wringing in the data science community regarding the validity of attacks shown 'in the lab' or at small scales. A key example is de Montjoye et al.'s work [35, 37] on the unicity of microdata, and an ensuing debate that occurred between the authors and Barth-Jones[15, 36] regarding the validity of the threat model. Barth-Jones argues that the attacks 'reflect unrealistic data intrusion threats', as they do not reidentify data subjects *en masse*, and claims that it would be a misplacement of effort to seek to mitigate these outlier cases.

This is something of a false equivalence between threat models—mass attacks and individual attacks, targeted and untargeted attacks alike can all result in harm to data subjects, and represent different but nonetheless valid threat models. Rather than arguing for the investigation only of a particular threat model, the response to this variety of threats ought to be to investigate each, and if possible construct an understanding of where risks are shared, so that mitigations designed for one model might be adapted for others.

Existence of side information. A related criticism of Barth-Jones' [15] levelled at de Montjoye et al. and similar works complains that the authors do not demonstrate that the hypothesised auxiliary data linked against in the attack actually exists. We would posit that this is, in fact, instructive in and of itself—there is evidence that data that resembles the required linkage dataset exists somewhere, and appreciating this fact is the best defence against what would otherwise be a 'black swan' reidentification attack.

Demonstrating that an attack is possible with a particular set of datasets, even 'in the lab', is useful because each of these datasets reflects a class of similar datasets—it is not unreasonable to expect that an attacker exists with access to such a similar dataset and thus is able to perpetrate a similar attack.

This criticism is also given in Barth-Jones' rebuttal of the Weld case [16], where the argument is made that the demonstrated attack only places an upper bound on per-subject reidentification risk, due to the absence of a 'perfect population register' to link against.

This may have been the case in 1997, but the modern availability of data (Section 3.4.1) suggests that a single 'perfect register' might be substituted for a large number of overlapping registers. Indeed, the Netflix prize case study demonstrated that sparse microdata yields high local unicity, and de Montjoye et al. showed that sparsity is often not needed for unicity—and as we saw from the Weld case, once membership of the dataset can be established local unicity is sufficient to reidentify the target.

Ease of linkability. The feasibility of a linkage also depends on the ease of extracting linkable attributes from the auxiliary data. A useful example might be to compare the 'sensitivity' of a photo album versus an address book (or contacts list). This is a common concern in mobile apps. Mobile operating systems prompt the user to grant or deny permissions to access the address book or photo library on a per-app basis—apps such as Clubhouse and Houseparty for the former, apps like Instagram for the latter.

In the past, the greater risk might have been thought to be the sharing of contacts—one has high confidence that it contains one-to-one 'direct identifiers', a goldmine of classic PII. However, once photos can be mined at scale with ease, extracting faces, time and location metadata, relationships between subjects, patterns of behaviour, all become extractable—as evidenced by mobile operating systems' 'memories' features. This is reflected by the current options given by the prompts for each of these datatypes—iOS allows all-or-nothing sharing of the contact book, whereas photos can be permissioned on a photo-by-photo basis.

This reflects a perennial trend of data technologies: where once only structured data could be handled mechanically and at scale, now useful information can be mined even from innocuous data sources; where the use-case of data had to be justified prior to extracting that utility, data is now amassed and linked speculatively.

The same trends apply to privacy risk: intuitively 'anonymous' data can be identifying, even if the information it confers about a person is implicit; and logical relationships between information can be found speculatively or after the fact.

4 SYSTEMATISATION

We have presented a range of notions of identification and related attacks on data subjects' privacy, and discussed ambiguities and contentions surrounding their use as definitions of privacy breach.

In this section we lay out our approach to resolving this ambiguity, by expressing all routes to privacy breach as variations on a single process—the **linkage attack**. Using the language of record linkage, we describe a linkage attack as the assembly of an **identifying record**—a collection of information that pertains to one or more natural persons—and discuss how to characterise the quality of the assertion that this record represents.

While this definition describes the *process* of the generic data linkage attack, a specific attack is characterised by the attacker's success criteria and the information in question that is being linked—the dataset under attack and the **linkage set**. We capture the success criteria as the attacker's **intent**—a description of a successful attack in terms of the final identifying record it produces. We will discuss the other aspect of the threat model, the linkage set, in Section 6.

We demonstrate the utility of our approach by defining natural (face-to-face) identification and the identification of persons by 'direct identifiers' as record linkage attacks, before presenting a family of intents that describes a range of threat models (with reference to known attacks). We round out the systematisation by presenting a discussion of distributional or semantic properties of the information context of an attack, and how the viability of certain threat models depends on these properties.

4.1 Defining linkage

4.1.1 The identifying record. Following the well-established concept of record linkage in information retrieval, and the QID-based conceptual model of dataset privacy used in k-anonymity [118], we define linkage attacks as the building of an identifying record.

The record consists of a set of facts which are either descriptive or contextual; descriptive facts assert information about a natural person, while contextual facts provide additional information that has some bearing on how other facts are to be interpreted or applied to individuals. Semantically, the identifying record is an assertion that all facts it contains are true of at least one natural person.

4.1.2 Linking records. The process of record linkage is the incorporation of additional facts into the identifying record. For example, in the Weld case, Sweeney linked a record from a hospital dataset to information that uniquely described William Weld.

We might represent the candidate medical record as containing descriptive facts such as '*person* was hospitalised on May 18, 1996', and the information in Sweeney's 'profile' of Weld as a record containing descriptive facts such as '*person* is the natural person understood to be William Weld', '*person* was Governor of Massachusetts on May 18, 1996', and '*person* resides within ZIP 02138'.

4.1.3 Evaluating identifying records. The quality of an identifying record can be captured in two factors—the record's specificity and its coherence. During a linkage attack, we describe the attacker's evaluation of their identifying record—and by extension the success of the attack—as the factors: $\tilde{a}$ the **estimated size of its anonymity set** (all natural persons about whom the facts contained are true), and c the **linkage confidence**, a degree of belief that all facts describe the same person(s). Going forward we will use the notation $(\tilde{a}, c)$, and express c as a number between 0 and 1. Note that the anonymity set is not the set of all natural persons about whom the facts were collected—this is the *membership* which we will discuss later.

There are many quantitative metrics that might represent or model these qualities, and these have been a subject of much research [126]. For our purposes we are not concerned with strict quantitative modelling of these factors, but choose these two metrics as they are intuitive and align with existing work on reidentification. Similarly, we do not prescribe a means of estimating values for these factors, as this might be subjective or vary according to an attacker's methodology, and is not materially important to defining threat models, where we are primarily concerned with directionality—whether the anonymity set can be winnowed down to a useful number (1 in the case of many attacks on individuals), and how high a confidence can be achieved.

The linkage confidence is the attacker's confidence that all facts are jointly true about the described natural person(s). This is updated with each linkage—the strength of the linkage depends on the confidence that either the records were collected from the same subject, or that the information contained within one record is representative of the subject of the other. This measure is equivalent to the confidence in the estimate $\tilde{a}$, as all information used in that estimate is contained within the record—logical connections between facts, including probabilistic connections such as distributional information, are captured as contextual facts, and their addition can be used to strengthen the case for coherence.

The fact that the two factors cannot be captured jointly, or why the confidence is not simply a Bayesian probability of correctness of the estimate $\tilde{a}$, can be illustrated by the distinction between *probabilistic* and *fractional* attacks. Tradeoffs exist between specificity and linkage confidence. We might give up specificity of an identifying record in exchange for a higher linkage confidence. For example, we might assert fewer facts about the person, therefore describing a *fraction* of the previous anonymity set, but increasing our certainty that all the asserted facts are true. Similarly, the converse tradeoff might be made, where an absolutely certain assertion is weakened to a *probabilistic* claim about a more informative record.

4.2 Threat modelling

Note. We denote a person's natural identity ι, an atomic element that represents that person independently of any naming system or real-world descriptors.

4.2.1 Examples: intuitive or 'direct' identification. Before we systematise the whole space of privacy attacks discussed in Section 2.2, let us demonstrate how the record linkage-based definition of attacks given above can be used to understand common and intuitively understood notions of identity.

'Direct' identifiers. The term 'direct' identifier has historically been applied to special datatypes that were ordained by common wisdom or law to be a sufficient means of uniquely describing a natural person. The distinguishing characteristic of a direct identifier is that its use in linkage is so well known as to be implicit.

Consider a record containing an 11-character string and a contextual fact that states that this string is a driving license number. The assertion that this record uniquely identifies a person is the result of linking it to a linkage set that has information about the distribution of driving license numbers—specifically that the mapping of strings to persons is one-to-one. Because the contextual information in the record tells us that this collection of individual facts can only describe one person, we estimate $\tilde{a} = 1$, and if we believe that the linkage set truly describes the distribution of the driving license number in the record, we would also estimate $c = 1$. This is the linkage we implicitly perform when we take a driving license number to be a direct identifier.

Personal facial recognition.. Recognising a person on the street is a similar process to identification by 'direct' identifier, except that the data linked against is a part of one's personal memory; the visual stimulus of a face is our identifying record, linked against a private dataset which maps faces to natural identities.

Personal name recognition.. A slightly less trivial example is the identification that occurs when a letter sender is recognised from their full name, written on the sheet of paper. Let us informally describe the building and linking of an identifying record towards a satisfactory 'identification' end state. Consider an initial identifying record containing the facts 'the writer signed this letter P. Sherman', 'this letter is addressed to me', 'this letter was written by hand all in a single style of handwriting', and '*person* wrote this letter'.

A cursory evaluation would estimate the size of the anonymity set as 1, as it is unlikely that there exist multiple identities referenced by *person*, given the single style of handwriting, and $c = 1$, since there is no source of uncertainty in the joint truth of any two facts.

As with the previous example, the recipient implicitly links this record to a profile in their memory about a person ι. This profile record might include the facts '*person* goes by the name Priscilla Sherman', 'I know *person* personally', and '*person* is ι'. Because the record explicitly references one natural identity as the subject, this can be the only member of its anonymity set and so $(\tilde{a}, c) = (1, 1)$.

Finally, the recipient implicitly links in some contextual information about letters: 'handwritten letters are usually from people known personally by the recipient'. The recipient's internal evaluation of this record would yield $\tilde{a} = 1$ by the same reasoning as the initial record. Their linkage confidence would be less than certain due to potential sources of error: the name Priscilla Sherman is not equal to P. Sherman; neither name is likely to be globally unique;

and the writer may sign a name falsely. However, the coherence of the record is bolstered by the final contextual fact, which contributes confidence to the joint truth of the facts that state that the letter was hand-written by the described person, and that the recipient knows that person. The linkage confidence would thus be less than 1, but likely still high enough for the recipient to recognise the sender as their friend Priscilla.

4.2.2 Attacker intents. The criteria by which a dataset privacy attack might be considered successful can vary greatly. We will now present a language that captures these success criteria, which is the basis for threat modelling of linkage attacks. An evaluation $(\tilde{a}, c)$ describes the quality of the identifying record, and so factors into success criteria, but other—sometimes subjective—factors specific to an attacker's intended information gain must also be captured.

We describe a successful attack as the achievement of an attacker's **intent**. This is a description of the record an attacker starts with, and the final identifying record they produce—in mechanical terms such as the $(\tilde{a}, c)$ evaluation as well as subjective assessments such as the presence of a useful QID.

We present a taxonomy of intents that spans the most commonly discussed threat models in dataset privacy attacks. Figure 1 summarises this taxonomy, showing each intent and the relationships between them (i.e. where an intent is a strengthening or specific case of another, and how). Specifying the relationships between these intents helps to disambiguate between attacks, and should provide a basis for future threat modelling practices that consider attack escalation or adaptation.

4.2.3 Dimensions of intents. Before we define each intent in the taxonomy, we present some high-level properties of intents, which can be seen as orthogonal dimensions within the space of intents.

Perfect confidence vs. probabilistic. As we noted in the examples presented above, an attacker may be satisfied by some margin of error in their linkage confidence, as long as the confidence remains sufficiently high (close to 1). Therefore an intent may specify either that the final record evaluates to $c = 1$ or $c > c_t$ where c_t is their subjective threshold below 1. This applies to all intents that we might formulate, so we note it here rather than distinguishing for each intent its 'perfect confidence' and probabilistic variations.

Individual vs. class attacks. While an attacker will always seek to push c as close to 1 as possible, intents can vary in their acceptable values for $\tilde{a}$. In **individual attacks**, which comprise most of the attacks we discuss here, the attacker wishes to winnow down the anonymity set until they arrive at an evaluation of $\tilde{a} = 1$. All reidentifications are individual attacks.

In contrast to individual attacks, there are **class attacks**, where an attacker seeks to gain information on a group of persons. In class attacks, the attacker can strengthen their record by winnowing down their estimate of $\tilde{a}$ until it reaches the class size N: if the size is known a priori, we can capture this in the intent with the requirement $\tilde{a} = N$; if not the lowest $\tilde{a}$ the attacker is able to achieve is effectively their estimation of the class size.

Repeatability. The scalability of individual attacks is a matter of their repeatability. In some cases, such as the case of the AOL search logs, an attacker may only be able to successfully produce records for a very small number of a dataset's subjects—we call this an **outlier** attack. Where an attack can be repeated on multiple subjects, we call it a **mass attack**, and where it can be extended to all subjects of a dataset a **total attack**.

Starting records. Another significant differentiating factor between intents is the attacker's starting point—the record they begin with and attempt to incorporate new information into.

In **untargeted** attacks, the attacker is happy to produce any identifying record with high confidence, without seeking to establish any particular fact or identity about the person(s). In these cases, the starting record is empty.

In **profiling** attacks, the attacker begins with a collection of facts that describes their intended target—a profile record—and seeks to add information, so that the final identifying record is a proper superset of the profile. Intuitively, in a profiling attack the attacker is looking to learn information about *any* subject of the dataset that satisfies a particular description.

Targeted attacks take profiling attacks a step further, describing attacks where the attacker intends to produce an identifying record about a particular person. This can be described in terms of records as beginning with a target record that contains the natural identity ι of that person, for which $\tilde{a} = 1$ by definition.

4.2.4 Intent definitions. We will now define each intent in the taxonomy shown in Figure 1. We begin with two intents that are not necessarily attacks on datasets' subjects, but more general information gain efforts that serve as parent intents to many others. These are the two **attribute inference** attacks.

Class attribute inference. This intent describes the most basic profiling attack. The attacker begins with a profile record P and seeks to produce a final record R that is its proper superset. Intuitively, this means that the attacker begins with a description of a class of persons and seeks to infer with confidence any additional attribute about that class.

Targeted attribute inference. Similarly, this intent describes the most basic form of targeted information gain, where the attacker holds a profile that includes a person's natural identity (therefore trivially fixing $\tilde{a} = 1$ and making it an individual attack), and seeking to produce a proper superset of the target record, i.e. adding a fact (inferring an attribute) about that person.

Class membership estimation. This intent follows from class attribute inference by adding the requirement that the attribute inferred is membership of the dataset under attack. Intuitively, this intent describes an attacker asserting that some subjects of a dataset match the given profile. This is useful either as an estimation of the membership of that class amongst the dataset's subjects, given by the attacker's best estimate for $\tilde{a}$, or as the starting point for further attribute inference attacks on the class.

The remainder of the intents presented are individual attacks, which comprise the majority of attacks discussed in the literature.

Singling out. The simplest attack on a dataset subject, the extraction of a record that describes a single person. This is trivial in datasets that represent data at individual-level granularity, but non-trivial where data is stored at group- or population-level, such as in aggregate statistical datasets. The singling out attack is not truly a reidentification attack as it does not necessarily produce a means of identification that translates to other sources of information.

Untargeted reidentification. A strengthening of the singling out attack, adding the requirement that the final identifying record

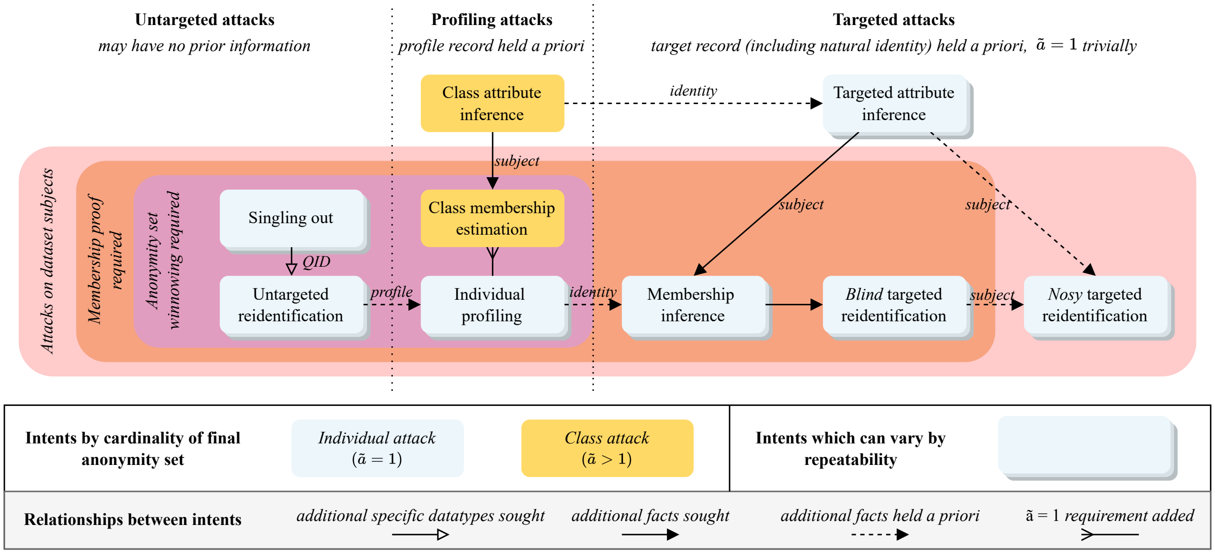
Untargeted attacks
may have no prior information
Profiling attacks
profile record held a priori
Targeted attacks
target record (including natural identity) held a priori, ã = 1 trivially
Class attribute inference
identity
Targeted attribute inference
subject
Class membership estimation
subject
subject
Attacks on dataset subjects
Membership proof required
Anonymity set winnowing required
Singling out
QID
Untargeted reidentification
profile
Individual profiling
identity
Membership inference
Blind targeted reidentification
subject
Nosy targeted reidentification
Intents by cardinality of final anonymity set
Individual attack (ã = 1)
Class attack (ã > 1)
Intents which can vary by repeatability
Relationships between intents
additional specific datatypes sought
additional facts sought
additional facts held a priori
ã = 1 requirement added

Figure 1: A taxonomy of notable linkage attack intents.

contains a useful QID—a subjective assessment particular to the attacker. The attacker might have a particular set of datatypes in mind as a useful QID prior to performing linkages, but this does not constitute a profile as a specification of datatypes is not itself a set of facts about a person.

Individual profiling. This intent is a child of both the untargeted reidentification and the class membership estimation. It is a reidentification attack in that it seeks to prove the existence of a single dataset subject that matches the profile—this is a strictly stronger prior requirement than simply seeking a QID. It follows from the class membership estimation essentially as its individual case—we simply add the requirement that $\tilde{a} = 1$, and so this intent is simply the process of finding a wider class and winnowing it down until the record describes only one of them.

The final three intents are all **targeted** attacks, and so the winnowing of $\tilde{a}$ is no longer necessary—as long as the facts in the target record (minus the stipulation of the natural identity ι) hold of any of the dataset's subject, they will hold of the target.

Membership inference. This is, in a sense, the minimal targeted attack on a dataset, as the simplest form of attribute inference, where the attribute is membership of the dataset. This intent is satisfied simply by proving that the target is a subject of the dataset.

Nosy targeted reidentification. This intent also follows from the targeted attribute inference attack, but assumes that the membership inference problem is pre-solved for the attacker (they know that their target is a subject of the dataset). The attack is thus simply attribute inference—the intent is satisfied by adding any information to the target record.

Blind targeted reidentification. This final intent incorporates the tasks of both of the previous discussed intents—the attacker is seeking to show that their target is a subject of the dataset, and then seeks to gain some further information about them from that dataset. The attacker is thus performing membership inference plus an additional attribute inference.

4.2.5 Discussion of intents. ***Other intent formulations.*** No list of intents can be exhaustive, due to the subjectivity of attackers' goals. Some of the intents above might be strengthened, for example we might add additional attribute inference to the individual profiling attack: the resulting intent would require that the final record be a proper superset of the profile plus the statement of membership. Other, more complex intents could be formulated that do not fit neatly into the taxonomy given, such as for attacks that infer the relationships between dataset subjects. For example, an attack that reidentifies a pair of people who met at a particular location at a particular time could be represented as a profiling intent whose initial profile contains the facts 'was at location L at time T' and 'met a single other person at time T', and can be satisfied with two and only two distinct records each evaluating to $(\tilde{a}, c) = (1, 1)$

Fractional vs. probabilistic attacks. We discussed earlier that intents may have a margin of error in linkage confidence, such that $c < 1$. These intents represent *probabilistic attacks*. This uncertainty in the assertion may be acceptable due to its subjective utility—if the intent represents the inference of a highly sensitive attribute, a significant but not absolute probability of correctness may exceed a threshold of risk to the subject.

Consider an attack aiming to recover a characteristic of a class of N persons which yields a record evaluated as $(\tilde{a}, c) = (N, 0.7)$ might satisfy some attackers' intents—they would have a 70% confidence that the characteristic applies to *all* N subjects. This successful evaluation is distinct from *fractional attacks*, in which the goal of the attacker is to say that an identifying record describes a certain fraction of a class of N persons. A record that proves such an assertion would yield $(\tilde{a}, c) = (0.7N, 1)$.

Membership inference and global vs. local unicity. The difference between global and local unicity is important to understanding attack intents. Proving local unicity, i.e. if an attacker can show that a set of facts is only true of one subject of the dataset under attack—is sufficient for singling out attacks, and in nosy targeted reidentification attacks. However, in other reidentification attacks the attacker must find a set of facts within the dataset that they can show to be globally unique—i.e. construct an identifying record which can be shown to have a single-member anonymity set.

Corepresentative datasets. Rather than showing that the subject(s) of one identifying record are also subjects of a record to be linked against, a weaker logical relationship between datasets that can be used in a linkage is to say that two datasets are corepresentative in a certain profile. Assertion of corepresentation means that attributes that are true of all subjects that match the profile in one dataset will also be true of matching subjects in the other.

This captures why the targeted attribute inference intent is a weakening of blind targeted reidentification, as it replaces the embedded membership inference problem with one of proving corepresentativeness between target and identifying record.

Attacks without reidentification. Attack intents that do not include an assertion that the subject of the identifying record is a subject of the dataset under attack (the targeted and class attribute inference attacks) are not true reidentifications, as they can be successful without linking information within the dataset to their subject with unicity. These intents, as well as the nosy targeted reidentification, are only matters of record enrichment—using the dataset itself as a source of side information to link additional descriptive facts to the profile record.

Linkage sets. Untargeted attacks do not require an external source of data to link against, and can be totally implicit, assembling an identifying record from the dataset under attack and establishing its local unicity by statistical means.

5 APPLICATION TO LITERATURE SURVEY

We validate our model by applying it to an extensive survey of dataset privacy attacks in the wider literature. Our successful categorisation of a large space of attacks validates the expressiveness of our model, and in the remainder of the paper we will discuss the utility of this expressiveness using insights drawn from the experience of applying it.

5.0.1 Methodology. We performed an extensive survey of the academic literature, capturing a significant sample of papers published in top security and privacy venues, as well as any other significant papers which propose or demonstrate attacks on dataset privacy. This was achieved by searching all previous published proceedings of ACM CCS and ASIACCS, IEEE S&P and EuroS&P, NDSS, Usenix Security, and PETS, plus accompanying workshops, and the whole ACM Digital Library, for a range of keywords: deanonymisation, reidentification, linkability, linkage attacks, and variations on those terms. Snowball sampling was then employed to capture influential, notable, or highly cited examples from 'grey literature'—such as self-published research or media reports. From these searches we drew an initial sample of 418 papers which appeared to have some relevance to dataset privacy attacks. From this sample we identified 94 papers which described attacks on datasets which were in scope, and categorised each using our taxonomy of intents. The full categorisation, noting for each attack its intent, repeatability, and (where appropriate) the datatype of the linkage set used, is tabulated in Appendix A.

5.1 Notes on applying the model

Translating each attack surveyed into intents was straightforward in most cases, but we encountered some difficulties regarding the evaluation of deanonymisations—while some papers made a distinction between 'open-' and 'closed-world' evaluations, it was more often unclear whether an attack was evaluated with targets not present in the dataset. This made it difficult to determine whether a targeted reidentification was *blind* or *nosy*.

Furthermore, many of the attacks surveyed began with the retrieval of a highly unique signal within a dataset, which could then be used *en masse* as a quasi-identifier, either within a dataset or across similar data sources. How to express this as an intent is a matter of framing—if the thrust of the paper was to show a particular signal's global unicity, the attack is best described as an untargeted reidentification (i.e. a proof of a novel QID). However we more commonly found this approach used to show how a dataset can be deanonymised—in this sense the attack presented is better described as a mass (or total) nosy targeted reidentification.

6 RISK APPRAISAL AND MITIGATION

In Section 4 we stated that the threat model of a linkage attack is made up of two parts: the attacker's intent and the information they hold which can be linked against the dataset under attack—the linkage set. As we saw in Section 3, the availability of an appropriate linkage set is a crucial determinant of whether an attack is feasible.

Any risk-based assessment of privacy must consider attack intent as well as anticipate the linkage sets which might enable the most successful (and most invasive) attacks. Consideration must also be given as to whether those linkage sets are likely to exist, and whether or not they would be available to each attacker.

In this section we will describe the challenges inherent to this risk appraisal task, and how it must be approached as a best-effort exercise. We will discuss the various ways in which the uncertainty of linkage set availability has been addressed in prior work, and how our systematisation of linkage attacks can be leveraged to improve risk appraisal in practice.

6.1 Challenges of appraising linkage set availability

6.1.1 Perfect context anticipation is intractable. A core principle of 'big data' and 'data-driven' practices is that, by capturing data, we are able to identify useful and actionable signals that were previously unknown. This is reflected in the variety and novelty of privacy breaches seen over the last decade or so. Data collected about human beings is embedded with informative signals about those subjects, and those signals can be present in multiple instances of data capture, and later linked—in fact, this is the fundamental process of inference. (This has been noted as a key failure of synthetic data as a privacy protection [115]—signals remain in the data, but the nature of their retention is difficult to anticipate.)

We know that linkable data need not be contemporaneously captured [33], and that attributes of a subject can be reflected in data that has no direct semantic relationship to those attributes (such as location traces as an indicator for ethnicity [105]).

To anticipate all possible linkages is to have complete knowledge of what information is contained within a dataset, where that information was captured or may someday be captured, and whether or not the dataset in question will ever be inspected in a context where that other information is present. This is clearly an intractable problem—not least because to have complete knowledge of the information captured by a dataset would entail perfect data mining and analysis, which can also anticipate all future uses or interpretations of that information.

6.1.2 Speculative accrual of information and opaque dataflows. Data-driven economies motivate actors to accrue as much data as possible [135], with the expectation that signals will eventually be found within it, or that it will eventually prove a useful linkage set to interpret future data [91]. The result of this speculative accrual is not simply that large companies like Google and Facebook are likely to have the richest linkage sets around (perhaps even Barth-Jones' 'perfect population register') whose details are unknown. As we have seen (Section 3.4), data brokerage markets trade in this information, adding an extra layer of incentive to obscure the details of what information is moving where—as knowing where data resides, in what form, and at what scale is itself competitively sensitive information.

6.2 Appraisal state of the art

6.2.1 Expert wargaming. Today, in practice, risk appraisal for a novel dataset is done in one of two ways. The first is simply to call prior art or guidance in aid—a practitioner may recognise the datatypes or domains in question and look up sector-specific guidance, or often simply appeal to received wisdom or general consensus on what does and does not constitute personal data.

Where toolkits for risk appraisal do exist, they are often sector-specific (therefore constrained to narrow contexts) or prescribe outcomes, rather than providing practical advice for discovering routes to reidentification or information gain on the data subject.

The second, more intensive approach to identifying risks of identification we term *expert wargaming*. In this approach, the identification of data protection burden is performed by individuals who have specialised knowledge of the datatypes and domain. These experts are usually familiar with case studies and known failure modes, such as those discussed in Section 3, and have a (usually implicit) working knowledge of what datasets already exist that could be linked against, or the capabilities of interested adversaries.

Quantitative measures such as k-anonymity may aid in wargaming by establishing bounds on information gain, or orders of magnitude, but due to the *intractable context problem*, it is almost always essential to perform some amount of qualitative analysis. For example, quantitative metrics may not exist to describe the computational power available to an adversary, or the richness of their privately-collected datasets. In fact, the existence of those linkage datasets may be speculative, inferred from the visible data collection capabilities of the adversary—e.g. speculating that a social networking company could link against direct messaging metadata if they wished. This method of risk appraisal is clearly not optimal—it is labour intensive, and relies on domain-specific expertise that is often difficult to transpose to new settings.

This siloing effect also applies to threat models. For example, intensive familiarity with the William Weld reidentification case and the nuances of its critiques may be instructive to a practitioner applying privacy-enhancing transformations to a medical record dataset, in order to prevent individual reidentification. Comparison of the data's distribution to the case study may give the practitioner confidence that, while a one-in-a-million case of reidentification may be possible, it would be prohibitive to reidentify subjects en masse. However, that line of reasoning would give no sense of the risk of a *class attribute inference*, such as discovering the most common sexual health clinic attended by subjects of a particular ethnicity, despite such an attack sharing many common factors.

Furthermore, reliance on case studies leads to orthodoxy; assessments of an attack's infeasibility often fall, as advances in computational power make previously prohibitive costs manageable, or if a previously unknown and sufficiently overlapping linkage dataset arises. In these cases, it can be difficult to re-evaluate risks.

6.2.2 Inference control. An alternative way of approaching risk is to hedge against it—such approaches have a long history under the terminologies of data sanitisation, inference control, statistical disclosure control, and privacy-preserving data publishing. Generally, these approaches aim to prevent a given attacker from learning some specific information held in a dataset, by removing or attenuating information signals within it.

Early inference control [4, 38] focused on the protection of statistical databases which would be queried for aggregate statistics. Potential disclosures were identified by conceptual means—modelling the relationships between all information captured by the dataset and thus identifying routes to disclosure to disable—or by blanket perturbation of data such that the minimal specificity of any query result is bounded. The conceptual model was rendered obsolete by modern data science and the inability to completely characterise the information contained in data [48].

Modern privacy-preserving data publishing (PPDP) [53] is developing in the face of side information and linkage. Methods such as k-anonymity and randomised response [46], as well as sanitisation methods like generalisation and tokenisation, reduce risk by introducing uncertainty into the dataset.

While most defences still require a model of the potential linkage sets or attack intent, differential privacy [43, 44] promises strong protection in the face of arbitrary side information [73] by bounding the information contributed by an individual's inclusion in the database within a privacy budget ϵ, by perturbing every data item by some noise. In the polar opposite of the conceptual model, rather than capturing all possible logical relationships to information in all possible release contexts, differential privacy elides the consideration of contexts entirely. However, there are severe practical limitations of differential privacy, from setting privacy budgets, to its limitation to perturbable datatypes, to difficulty in applying it to time-varying or sequential releases [44]. As a result, its uptake in large-scale rich dataset release has been slow, with its largest (transparent) deployment to date being the 2020 U.S. Census [1].

It must be noted that even the strong promises of differential privacy are effectively a risk management solution to the problem of privacy—the privacy parameter ϵ represents the probabilistic nature of the protection, which is achieved by (albeit targeted) information attenuation. This reinforces the lesson of the intractable context problem—a perfect guarantee that no linkage can occur is impossible if data is to remain informative. (This insight should not be seen as controversial—almost all forms of encryption provide probabilistic protection, relying on the incredibly low likelihood of guessing the correct encryption key. An exception is the one-time pad where, without the key, all messages are equally likely.)

6.3 Better practices based on linkage risk

The way forward for the practice of dataset protection must be a toolkit of practices based on the appraisal and minimisation of linkage-based privacy breach. This toolkit must be clearly extensible, without relying on increasingly siloed expert experience, as is the current state of the art. We must also have clear ways of updating collective wisdom around what attacks are possible and where risks may arise, so that novel attacks and harms can be easily incorporated into the known attack surface that is tested against. We will now discuss how our record linkage framing, and intents in particular, provides a basis for this new risk-based toolkit.

6.3.1 A common language for novel attacks. The most straightforward application of our model is as a language for future literature, to avoid the historical confusion around threat models. This is useful not only to understand what a novel attack claims to achieve, but its limitations—a clear presentation of an attack would show not only how the intent is satisfied, but the extent to which escalations are possible. For example, if a novel dataset is published and a researcher identifies a possible individual profiling attack, by clearly stating the linkages performed that researcher should also be able to justify whether that profiling attack is repeatable as a *mass* or *total* attack, or whether the linkage data used is rare enough that the attack is only possible as an *outlier* attack. Our survey of papers validated this use case—once attacks are expressed in terms of intents their claims gain useful context, and the evolution of classes of attacks over time and by context can be charted. An example is the range seen in stylometry attacks [2, 5, 6, 23, 84, 128]. Variance in the attacked datasets (by language, context, or features captured) led to a range of attacks distinct in intent and repeatability (with some achieving total)—a prime candidate for comparative guidance for future publishers of text corpora.

6.3.2 Directed wargaming with risk pattern heuristics. The linkage attack language can also be used to replace the current status quo of expert wargaming with a more mechanical, directed methodology. Because potential sources of side information are so diverse and enumerating them is intractable, this will necessarily remain a best-effort exercise. We propose the building of a library of risk pattern heuristics, to augment qualitative and quantitative risk evaluations.

Risk pattern libraries could be collected around a number of commonalities, such as datatype, the availability of particular linkage datasets, or the specific failure modes prescribed by legislation. A comprehensive library of risk patterns dedicated to a certain datatype in a given domain would significantly lessen the need for imagination and extensive knowledge of information security on the part of the data protection practitioner.

These heuristics would be used as follows: if a data holder is preparing to publish or otherwise share a dataset, they would consult the library for risk patterns common to the datatypes involved, heuristics that might help identify potential sources of side information that their data could be linked against, and a checklist of data properties that are known to invite particular classes of risks.

Furthermore, once a potential attack under a particular intent is identified, patterns that leverage intent taxonomy could show when that risk might propagate to other intents—for example, the conditions that must be met for the attack to be repeatable.

6.4 Example risk patterns

The remainder of this section gives an initial set of risk patterns drawn from our survey of the literature, which describe observed relationships between certain properties of the datasets or linkage sets and the risk due to the linkage.

6.4.1 Membership. ***Global-capture linkage sets.*** In some cases, the attacker may have access to a linkage set whose set of subjects contains all (or a vast majority) of the data subjects of the target set—what we might call an *almost-perfect population register*, per Barth-Jones [16]. A prime example of such a linkage set is the US Census, which is collected on the majority of US residents, though in many cases direct linkages are not possible due to pre-publishing transformations applied to combat singling-out, meaning its utility is greatest as a corepresentative linkage set.

Difficulty of membership inference. In many cases, especially where individuals' data is not sparse, there is insufficient information about the data collection to solve the membership inference problem—this prevents the success of blind targeted reidentification attacks. By contrast, there are cases where the contents of a dataset are high-risk, and solving the membership inference problem takes an unfeasible or probabilistic attack to a guaranteed breach.

6.4.2 Temporality. ***Fresh vs. stale data.*** The recency of data collection can impact the risk posed by attacks. Often, old data contributes minimal risk, as there are no longer clear routes to harm—for example in most cases where the subject is long deceased.

Proximity of linkage set to victim set. Even if there are routes to harm, linkage attacks may not be feasible if they rely on contemporaneous data, which may not have been collected at the time and might not be constructable after the fact.

Increasing availability of linkage data. Conversely, some data may exhibit an increase in risk due to an increased availability of linkage data, due to the creation of other semantically related datasets in the intervening time. This pattern applies when a dataset contains information that remains sensitive throughout the subject's (or their family's) lifetime, such as race or medical conditions, or where the subject has some temporally persistent property, (or 'fingerprint'), which enables linkages across long periods of time.

6.4.3 Linkage sets. ***Maximally informative linkage set.*** In some cases, the act of linking to a linkage set confers no information gain for the attacker in terms of features, over what already was available in the linkage set. This may happen in cases where the target dataset is a subset of the linkage set. The only information gained for the attacker is instead the solution to the membership inference problem—i.e. learning that a subject of the linkage dataset is also a subject of the target dataset.

Information gain without risk gain. In a similar but slightly subtler case, while a linkage attack may identify a subject within a dataset and confer information that was not present in the linkage set, this information may not contribute significant additional potential for harm. This often occurs with highly informative linkage sets—reducing to the previously mentioned case.

Public vs. private linkage sets. The availability of information that can be linked against is hard to appraise. If data is collected in a commercial or industrial setting, there may exist other datasets within the organisation that could be linked against—for example, employee payroll might form the perfect population register for a workplace. This is a complicating factor when assessing the capabilities of large data collectors such as Google or Facebook due to their dominance in data brokerage markets. In these cases, risk analyses may have to assume these actors have linkage sets with global coverage for certain datatypes, such as web search history.

6.4.4 Cardinality. ***High unicity datasets.*** The Netflix Prize [88] showed that sparse data or high-entropy datatypes (such as location traces) exhibit global unicity with even a handful of data points. Global unicity rapidly increases the chances of successful membership inference, as well as providing excellent candidate QIDs for linkage to other data sources.

Outlier attacks can't scale. In many cases, outlier reidentification attacks do not scale to mass attacks. This occurs when the former only retrieves with high precision a minority of subjects with extraordinarily high local unicity within the dataset.

Outlier risk. For some datasets the majority of information contained would not be particularly harmful in the hands of an adversary, but it is possible that highly sensitive information may be contained, particularly in free text fields. For example, in the Enron dataset a vast majority of emails discuss trivial matters such as scheduling meetings, but some contain details of doctor's appointments or the names, ages, and schools of a subject's children.

7 WAYS FORWARD FOR LINKAGE RISK MANAGEMENT

We expect that by aligning the language of threat modelling for dataset privacy with the legal burdens and risk-based nature of linkage attacks, our framework will provide a platform for bridging the gap between privacy attacks in the literature and practice.

The greatest practical utility of the linkage framework is to regulators and other bodies offering guidance on data protection practice. Beyond simply the disambiguation of threat models and a clear correspondence to the phrasing of legal burdens, risk patterns can alleviate the expertise burden on practitioners.

The approach of propagating risks across intents can be mirrored for risk mitigation—a complementary library of patterns could describe the risk-damping effect a certain PET has over a group of intents. For example, a k-anonymity pattern would reduce risks under intents which require $\tilde{a} < k$; similarly, l-diversity and t-closeness can be expressed as forcing lower bounds on $\tilde{a}$ as a function of the specificity of the facts in the record.

The attacks identified could also be used in privacy parameter tuning; since the differential privacy budget ϵ bounds the information gain for arbitrary membership inference attacks, the linkage confidence when linking data to a differentially private dataset is a function of ϵ. We could thus disrupt an identified linkage attack on this dataset by identifying a 'pressure point' fact—one which is crucial to proving comembership—and tuning ϵ so as to bound confidence in that assertion.

We also believe it is important to reassess the 'end-to-end principle' that many data protection frameworks have inherited from the original Privacy by Design framework [24], which charges practitioners with securing data from collection to destruction—we now know that if further information is derived from the use of data, information on the subjects will persist. In these cases, we must establish accountability for the 'downstream' effects of data collection, and design governance measures with them in mind.

We noted in Section 2.2 that our model of linkage builds on existing concepts of *probabilistic record linkage*. We illustrated an identifying record's specificity and coherence in this paper with the factors $\tilde{a}, c$; risk appraisal methodologies may pursue concrete information-theoretic metrics or statistical methods which build on probabilistic linkage literature or other inference techniques, to provide mathematically rigorous methods for well-specified subsets of intents and available side information. Recent work on privacy-preserving microdata sharing by Alvim et al. [7] is a good example of such a methodology; the work clearly defines its threat models (which map well into our intent formulation) and provides strong Quantitative Information Flow modelling tools for attacks that leverage well-defined types of side information.

The intents we present are by no means exhaustive of dataset privacy attacks and were chosen to cover prominent reidentification attacks. Future work will expand the set of intents to express other privacy failures, such as joint (two-subject) attacks—e.g. finding subject co-locations within a location trace dataset, or profile assertions that describe familial relationships. While we have defined the identifying record in terms of asserting information about natural persons, our conceptual framework can be extended to any information gain process where the goal is to produce a description of some entity.

8 CONCLUSIONS

In this paper we traced the development of the understanding of linkage attacks as the basis for dataset privacy breach in legal and technical work. Through a definition of linkage attacks, we were able to systematise the diverse and often ambiguous landscape of threat models in dataset privacy, and articulate the relationships between attacker intents.

We also discussed the shortcomings of and challenges facing existing risk appraisal and privacy protection methodologies as they relate to risks of side information. Through our risk management we approach the risk appraisal problem as a 'best effort' exercise, and propose a new methodology of heuristic risk patterns. We initiated the development of these heuristics with a selection of insights from our systematisation.

Finally, we have recommended a number of future directions for data protection practice—our risk-based problem framing harmonises the legal and technical approaches to data privacy breach, which we hope will provide a foundation for integrating existing and future work on privacy attacks with the legal and operational challenges of data protection.

ACKNOWLEDGMENTS

This work was supported by the Engineering and Physical Sciences Research Council [grant number 1940975].

REFERENCES

[1] John M Abowd. 2018. The US Census Bureau adopts differential privacy. In *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*. 2867–2867.

[2] Mohammed Abuhamad, Tamer AbuHmed, Aziz Mohaisen, and DaeHun Nyang. 2018. Large-Scale and Language-Oblivious Code Authorship Identification. In *Proceedings of the 2018 ACM SIGSAC Conference on Computer and Communications Security (CCS '18)*. Association for Computing Machinery, New York, NY, USA, 101–114. https://doi.org/10.1145/3243734.3243738 event-place: Toronto, Canada.

[3] Jagdish Prasad Achara, Gergely Acs, and Claude Castelluccia. 2015. On the Unicity of Smartphone Applications. In *Proceedings of the 14th ACM Workshop on Privacy in the Electronic Society (WPES '15)*. Association for Computing Machinery, New York, NY, USA, 27–36. https://doi.org/10.1145/2808138.2808146 event-place: Denver, Colorado, USA.

[4] Nabil R Adam and John C Worthmann. 1989. Security-control methods for statistical databases: a comparative study. *ACM Computing Surveys (CSUR)* 21, 4 (1989), 515–556.

[5] Sadia Afroz, Aylin Caliskan Islam, Ariel Stolerman, Rachel Greenstadt, and Damon McCoy. 2014. Doppelgänger Finder: Taking Stylometry to the Underground. In *2014 IEEE Symposium on Security and Privacy*. 212–226. https://doi.org/10.1109/SP.2014.21 ISSN: 2375-1207.

[6] Mishari Almishari and Gene Tsudik. 2012. Exploring Linkability of User Reviews. In *Computer Security – ESORICS 2012*, David Hutchison, Takeo Kanade, Josef Kittler, Jon M. Kleinberg, Friedemann Mattern, John C. Mitchell, Moni Naor, Oscar Nierstrasz, C. Pandu Rangan, Bernhard Steffen, Madhu Sudan, Demetri Terzopoulos, Doug Tygar, Moshe Y. Vardi, Gerhard Weikum, Sara Foresti, Moti Yung, and Fabio Martinelli (Eds.). Vol. 7459. Springer Berlin Heidelberg, Berlin, Heidelberg, 307–324. https://doi.org/10.1007/978-3-642-33167-1_18 Series Title: Lecture Notes in Computer Science.

[7] Mário S Alvim, Natasha Fernandes, Annabelle McIver, Carroll Morgan, and Gabriel H Nunes. 2022. Flexible and scalable privacy assessment for very large datasets, with an application to official governmental microdata. *arXiv preprint arXiv:2204.13734* (2022).

[8] Myrto Arapinis, Loretta Ilaria Mancini, Eike Ritter, and Mark Ryan. 2014. Privacy through Pseudonymity in Mobile Telephony Systems.. In *NDSS*. https://doi.org/10.14722/ndss.2014.23082

[9] Kerem Ayoz, Erman Ayday, and A. Ercument Cicek. 2021. Genome Reconstruction Attacks Against Genomic Data-Sharing Beacons. *Proceedings on Privacy Enhancing Technologies* 2021, 3 (July 2021), 28–48. https://doi.org/10.2478/popets-2021-0036

[10] Michael Backes, Pascal Berrang, Matthias Bieg, Roland Eils, Carl Herrmann, Mathias Humbert, and Irina Lehmann. 2017. Identifying Personal DNA Methylation Profiles by Genotype Inference. In *2017 IEEE Symposium on Security and Privacy (SP)*. 957–976. https://doi.org/10.1109/SP.2017.21 ISSN: 2375-1207.

[11] Michael Backes, Pascal Berrang, Anne Hecksteden, Mathias Humbert, Andreas Keller, and Tim Meyer. 2016. Privacy in epigenetics: temporal linkability of MicroRNA expression profiles. In *Proceedings of the 25th USENIX Conference on Security Symposium (SEC'16)*. USENIX Association, USA, 1223–1240.

[12] Michael Backes, Pascal Berrang, Mathias Humbert, and Praveen Manoharan. 2016. Membership privacy in MicroRNA-based studies. In *Proceedings of the 2016 ACM SIGSAC Conference on Computer and Communications Security*. 319–330.

[13] Michael Backes, Mathias Humbert, Jun Pang, and Yang Zhang. 2017. Walk2friends: Inferring Social Links from Mobility Profiles. In *Proceedings of the 2017 ACM SIGSAC Conference on Computer and Communications Security (CCS '17)*. Association for Computing Machinery, New York, NY, USA, 1943–1957. https://doi.org/10.1145/3133956.3133972 event-place: Dallas, Texas, USA.

[14] Michael Barbaro and Tom Zeller Jr. 2006. A Face Is Exposed for AOL Searcher No. 4417749. https://www.nytimes.com/2006/08/09/technology/09aol.html

[15] Daniel Barth-Jones, Khaled El Emam, Jane Bambauer, Ann Cavoukian, and Bradley Malin. 2015. Assessing data intrusion threats. *Science (New York, NY)* 348, 6231 (2015), 194.

[16] Daniel C. Barth-Jones. 2012. The 'Re-Identification' of Governor William Weld's Medical Information: A Critical Re-Examination of Health Data Identification Risks and Privacy Protections, Then and Now. *social science research network* (2012). https://doi.org/10.2139/ssrn.2076397

[17] Pascal Berrang, Mathias Humbert, Yang Zhang, Irina Lehmann, Roland Eils, and Michael Backes. 2018. Dissecting Privacy Risks in Biomedical Data. In *2018 IEEE European Symposium on Security and Privacy (EuroS&P)*. 62–76. https://doi.org/10.1109/EuroSP.2018.00013

[18] Sarah Bird, Ilana Segall, and Martin Lopatka. 2020. Replication: Why we Still Can't Browse in Peace: On the Uniqueness and Reidentifiability of Web Browsing Histories. In *Proceedings of the Sixteenth USENIX Conference on Usable Privacy and Security (SOUPS'20)*. USENIX Association, USA.

[19] Norbert Blenn, Christian Doerr, Nasireddin Shadravan, and Piet Van Mieghem. 2012. How Much Do Your Friends Know about You? Reconstructing Private Information from the Friendship Graph. In *Proceedings of the Fifth Workshop on Social Network Systems (SNS '12)*. Association for Computing Machinery, New York, NY, USA. https://doi.org/10.1145/2181176.2181178 event-place: Bern, Switzerland.

[20] Michelle Boorstein, Marisa Iati, and Annys Shin. 2021. Top U.S. Catholic Church official resigns after cellphone data used to track him on Grindr and to gay bars. https://www.washingtonpost.com/religion/2021/07/20/bishop-misconduct-resign-burrill/

[21] Erik Buchmann, Klemens Böhm, Thorben Burghardt, and Stephan Kessler. 2013. Re-identification of Smart Meter data. *Personal and Ubiquitous Computing* 17, 4 (April 2013), 653–662. https://doi.org/10.1007/s00779-012-0513-6

[22] Joseph A. Calandrino, Ann Kilzer, Arvind Narayanan, Edward W. Felten, and Vitaly Shmatikov. 2011. "You Might Also Like:" Privacy Risks of Collaborative Filtering. In *2011 IEEE Symposium on Security and Privacy*. 231–246. https://doi.org/10.1109/SP.2011.40 ISSN: 2375-1207.

[23] Aylin Caliskan-Islam, Richard Harang, Andrew Liu, Arvind Narayanan, Clare Voss, Fabian Yamaguchi, and Rachel Greenstadt. 2015. De-anonymizing programmers via code stylometry. In *24th USENIX security symposium (USENIX Security 15)*. 255–270.

[24] Ann Cavoukian et al. 2009. Privacy by design: The 7 foundational principles. *Information and privacy commissioner of Ontario, Canada* 5 (2009), 12.

[25] Abdelberi Chaabane, Gergely Acs, Mohamed Ali Kaafar, and others. 2012. You are what you like! information leakage through users' interests. In *Proceedings of the 19th annual network & distributed system security symposium (NDSS)*. Citeseer.

[26] Terence Chen, Abdelberi Chaabane, Pierre Ugo Tournoux, Mohamed-Ali Kaafar, and Roksana Boreli. 2013. How much is too much? Leveraging ads audience estimation to evaluate public profile uniqueness. In *International Symposium on Privacy Enhancing Technologies Symposium*. Springer, 225–244. https://doi.org/10.1007/978-3-642-39077-7_12

[27] Peter Christen. 2012. The data matching process. In *Data matching*. Springer, 23–35.

[28] Vassilis Christophides, Vasilis Efthymiou, Themis Palpanas, George Papadakis, and Kostas Stefanidis. 2019. End-to-end entity resolution for big data: A survey. *arXiv preprint arXiv:1905.06397* (2019).

[29] Marco Cominelli, Francesco Gringoli, Paul Patras, Margus Lind, and Guevara Noubir. 2020. Even Black Cats Cannot Stay Hidden in the Dark: Full-band De-anonymization of Bluetooth Classic Devices. In *2020 IEEE Symposium on Security and Privacy (SP)*. 534–548. https://doi.org/10.1109/SP40000.2020.00091 ISSN: 2375-1207.

[30] Joseph Cox. 2019. T-Mobile, Sprint, and AT&T Are Selling Customers' Real-Time Location Data, And It's Falling Into the Wrong Hands. https://www.vice.com/en/article/nepxbz/i-gave-a-bounty-hunter-300-dollars-located-phone-microbilt-zumigo-tmobile

[31] Joseph Cox. 2021. The Inevitable Weaponization of App Data Is Here. https://www.vice.com/en/article/pkbxp8/grindr-location-data-priest-weaponization-app

[32] Joseph Cox. 2021. Inside the Industry That Unmasks People at Scale. https://www.vice.com/en/article/epnmvz/industry-unmasks-at-scale-maid-to-pii

[33] Ana-Maria Creţu, Federico Monti, Stefano Marrone, Xiaowen Dong, Michael Bronstein, and Yves-Alexandre de Montjoye. 2022. Interaction data are identifiable even across long periods of time. *Nature Communications* 13, 1 (2022), 1–11.

[34] Anupam Datta, Divya Sharma, and Arunesh Sinha. 2012. Provable De-anonymization of Large Datasets with Sparse Dimensions. In *Principles of Security and Trust (Lecture Notes in Computer Science)*, Pierpaolo Degano and Joshua D. Guttman (Eds.). Springer, Berlin, Heidelberg, 229–248. https://doi.org/10.1007/978-3-642-28641-4_13

[35] Yves-Alexandre De Montjoye, César A Hidalgo, Michel Verleysen, and Vincent D Blondel. 2013. Unique in the crowd: The privacy bounds of human mobility. *Scientific reports* 3, 1 (2013), 1–5.

[36] Yves-Alexandre de Montjoye and Alex Pentland. 2015. Assessing data intrusion threats—Response. *Science (New York, NY)* (2015).

[37] Yves-Alexandre De Montjoye, Laura Radaelli, Vivek Kumar Singh, et al. 2015. Unique in the shopping mall: On the reidentifiability of credit card metadata. *Science* 347, 6221 (2015), 536–539.

[38] Dorothy Elizabeth Robling Denning. 1982. *Cryptography and data security*. Vol. 112. Addison-Wesley Reading.

[39] Clemens Deußer, Steffen Passmann, and Thorsten Strufe. 2020. Browsing Unicity: On the Limits of Anonymizing Web Tracking Data. In *2020 IEEE Symposium on Security and Privacy (SP)*. 777–790. https://doi.org/10.1109/SP40000.2020.00018 ISSN: 2375-1207.

[40] Iman Deznabi, Mohammad Mobayen, Nazanin Jafari, Oznur Tastan, and Erman Ayday. 2018. An Inference Attack on Genomic Data Using Kinship, Complex Correlations, and Phenotype Information. *IEEE/ACM Trans. Comput. Biol. Bioinformatics* 15, 4 (July 2018), 1333–1343. https://doi.org/10.1109/TCBB.2017.2709740 Place: Washington, DC, USA Publisher: IEEE Computer Society Press.

[41] Xuan Ding, Lan Zhang, Zhiguo Wan, and Ming Gu. 2011. De-Anonymizing Dynamic Social Networks. In *2011 IEEE Global Telecommunications Conference - GLOBECOM 2011*. 1–6. https://doi.org/10.1109/GLOCOM.2011.6133607 ISSN: 1930-529X.

[42] Halbert L Dunn. 1946. Record linkage. *American Journal of Public Health and the Nations Health* 36, 12 (1946), 1412–1416.

[43] Cynthia Dwork. 2006. Differential privacy. In *International Colloquium on Automata, Languages, and Programming*. Springer, 1–12.

[44] Cynthia Dwork, Moni Naor, Toniann Pitassi, and Guy N Rothblum. 2010. Differential privacy under continual observation. In *Proceedings of the forty-second ACM symposium on Theory of computing*. 715–724.

[45] Cynthia Dwork, Adam Smith, Thomas Steinke, and Jonathan Ullman. 2017. Exposed! a survey of attacks on private data. *Annual Review of Statistics and Its Application* 4 (2017), 61–84.

[46] Úlfar Erlingsson, Vasyl Pihur, and Aleksandra Korolova. 2014. Rappor: Randomized aggregatable privacy-preserving ordinal response. In *Proceedings of the 2014 ACM SIGSAC conference on computer and communications security*. 1054–1067.

[47] Francesca Falzon, Evangelia Anna Markatou, Akshima, David Cash, Adam Rivkin, Jesse Stern, and Roberto Tamassia. 2020. Full Database Reconstruction in Two Dimensions. In *Proceedings of the 2020 ACM SIGSAC Conference on Computer and Communications Security (CCS '20)*. Association for Computing Machinery, New York, NY, USA, 443–460. https://doi.org/10.1145/3372297.3417275 event-place: Virtual Event, USA.

[48] Csilla Farkas and Sushil Jajodia. 2002. The inference problem: A survey. *ACM SIGKDD Explorations Newsletter* 4, 2 (2002), 6–11.

[49] Ali Farzanehfar, Florimond Houssiau, and Yves-Alexandre de Montjoye. 2021. The risk of re-identification remains high even in country-scale location datasets. *Patterns* 2, 3 (2021), 100204. https://doi.org/10.1016/j.patter.2021.100204

[50] Michèle Finck and Frank Pallas. 2019. They Who Must Not Be Identified - Distinguishing Personal from Non-Personal Data Under the GDPR. *social science research network* (2019). https://doi.org/10.2139/ssrn.3462948

[51] Hao Fu, Aston Zhang, and Xing Xie. 2015. Effective Social Graph Deanonymization Based on Graph Structure and Descriptive Information. *ACM Transactions on Intelligent Systems and Technology* 6, 4 (July 2015), 49:1–49:29. https://doi.org/10.1145/2700836 Place: New York, NY, USA Publisher: Association for Computing Machinery.

[52] Luoyi Fu, Jiapeng Zhang, Shuaiqi Wang, Xinyu Wu, Xinbing Wang, and Guihai Chen. 2020. De-Anonymizing Social Networks With Overlapping Community Structure. *IEEE/ACM Trans. Netw.* 28, 1 (Feb. 2020), 360–375. https://doi.org/10.1109/TNET.2019.2962731 Publisher: IEEE Press.

[53] Benjamin CM Fung, Ke Wang, Rui Chen, and Philip S Yu. 2010. Privacy-preserving data publishing: A survey of recent developments. *ACM Computing Surveys (Csur)* 42, 4 (2010), 1–53.

[54] Aris Gkoulalas-Divanis, Dinusha Vatsalan, Dimitrios Karapiperis, and Murat Kantarcioglu. 2021. Modern Privacy-Preserving Record Linkage Techniques: An Overview. *IEEE Transactions on Information Forensics and Security* (2021).

[55] Oana Goga, Howard Lei, Sree Hari Krishnan Parthasarathi, Gerald Friedland, Robin Sommer, and Renata Teixeira. 2013. Exploiting innocuous activity for correlating users across sites. In *Proceedings of the 22nd international conference on World Wide Web (WWW '13)*. Association for Computing Machinery, New York, NY, USA, 447–458. https://doi.org/10.1145/2488388.2488428

[56] Philippe Golle. 2006. Revisiting the uniqueness of simple demographics in the US population. In *Proceedings of the 5th ACM workshop on Privacy in electronic society (WPES '06)*. Association for Computing Machinery, New York, NY, USA, 77–80. https://doi.org/10.1145/1179601.1179615

[57] Philippe Golle and Kurt Partridge. 2009. On the anonymity of home/work location pairs. In *International Conference on Pervasive Computing*. Springer, 390–397.

[58] Neil Zhenqiang Gong and Bin Liu. 2016. You are who you know and how you behave: Attribute inference attacks via users' social friends and behaviors. In *25th USENIX Security Symposium (USENIX Security 16)*. 979–995.

[59] Neil Zhenqiang Gong and Bin Liu. 2018. Attribute Inference Attacks in Online Social Networks. *ACM Transactions on Privacy and Security* 21, 1 (Jan. 2018), 3:1–3:30. https://doi.org/10.1145/3154793

[60] José González-Cabañas, Ángel Cuevas, Rubén Cuevas, Juan López-Fernández, and David García. 2021. Unique on Facebook: Formulation and Evidence of (Nano)Targeting Individual Users with Non-PII Data. In *Proceedings of the 21st ACM Internet Measurement Conference (IMC '21)*. Association for Computing Machinery, New York, NY, USA, 464–479. https://doi.org/10.1145/3487552.3487861 event-place: Virtual Event.

[61] Graham Greenleaf. 2012. The influence of European data privacy standards outside Europe: implications for globalization of Convention 108. *International Data Privacy Law* 2, 2 (2012), 68–92.

[62] Lifang Gu, Rohan Baxter, Deanne Vickers, and Chris Rainsford. 2003. Record linkage: Current practice and future directions. *CSIRO Mathematical and Information Sciences Technical Report* 3 (2003), 83.

[63] Gábor György Gulyás, Benedek Simon, and Sándor Imre. 2016. An Efficient and Robust Social Network De-Anonymization Attack. In *Proceedings of the 2016 ACM on Workshop on Privacy in the Electronic Society (WPES '16)*. Association for Computing Machinery, New York, NY, USA, 1–11. https://doi.org/10.1145/2994620.2994632 event-place: Vienna, Austria.

[64] Xiaojie Guo, Ye Han, Zheli Liu, Ding Wang, Yan Jia, and Jin Li. 2022. Birds of a Feather Flock Together: How Set Bias Helps to Deanonymize You via Revealed Intersection Sizes. In *31st USENIX Security Symposium (USENIX Security 22)*. 1487–1504.

[65] Wajih Ul Hassan, Saad Hussain, and Adam Bates. 2018. Analysis of Privacy Protections in Fitness Tracking Social Networks -or- You can run, but can you hide?. In *27th USENIX Security Symposium (USENIX Security 18)*. 497–512.

[66] Nestor Hernandez, Mizanur Rahman, Ruben Recabarren, and Bogdan Carbunar. 2018. Fraud De-Anonymization for Fun and Profit. In *Proceedings of the 2018 ACM SIGSAC Conference on Computer and Communications Security (CCS '18)*. Association for Computing Machinery, New York, NY, USA, 115–130. https://doi.org/10.1145/3243734.3243770 event-place: Toronto, Canada.

[67] Nils Homer, Szabolcs Szelinger, Margot Redman, David Duggan, Waibhav Tembe, Jill Muehling, John V Pearson, Dietrich A Stephan, Stanley F Nelson, and David W Craig. 2008. Resolving individuals contributing trace amounts of DNA to highly complex mixtures using high-density SNP genotyping microarrays. *PLoS genetics* 4, 8 (2008), e1000167.

[68] Mathias Humbert, Kévin Huguenin, Joachim Hugonot, Erman Ayday, and J-P Hubaux. 2015. De-anonymizing genomic databases using phenotypic traits. *Proceedings on Privacy Enhancing Technologies* 2015, 2 (2015). https://doi.org/10.1515/popets-2015-0020

[69] Ekaterini Ioannou, Wolfgang Nejdl, Claudia Niederée, and Yannis Velegrakis. 2010. On-the-fly entity-aware query processing in the presence of linkage. *Proceedings of the VLDB Endowment* 3, 1-2 (2010), 429–438.

[70] Shouling Ji, Weiqing Li, Neil Zhenqiang Gong, Prateek Mittal, and Raheem Beyah. 2015. On Your Social Network De-anonymizablity: Quantification and Large Scale Evaluation with Seed Knowledge. In *Proceedings 2015 Network and Distributed System Security Symposium*. Internet Society, San Diego, CA. https://doi.org/10.14722/ndss.2015.23096

[71] Shouling Ji, Weiqing Li, Mudhakar Srivatsa, and Raheem Beyah. 2016. Structural Data De-Anonymization: Theory and Practice. *IEEE/ACM Trans. Netw.* 24, 6 (Dec. 2016), 3523–3536. https://doi.org/10.1109/TNET.2016.2536479 Publisher: IEEE Press.

[72] Shouling Ji, Weiqing Li, Mudhakar Srivatsa, Jing Selena He, and Raheem Beyah. 2016. General Graph Data De-Anonymization: From Mobility Traces to Social Networks. *ACM Trans. Inf. Syst. Secur.* 18, 4 (April 2016). https://doi.org/10.1145/2894760 Place: New York, NY, USA Publisher: Association for Computing Machinery.

[73] Shiva Prasad Kasiviswanathan and Adam Smith. 2008. A note on differential privacy: Defining resistance to arbitrary side information. *CoRR abs/0803.3946* (2008).

[74] M Koot, G Noordende, and Cees De Laat. 2010. A study on the re-identifiability of Dutch citizens. In *Workshop on Privacy Enhancing Technologies (PET 2010)*.

[75] Bin Lin and Alexander Serebrenik. 2016. Recognizing Gender of Stack Overflow Users. In *Proceedings of the 13th International Conference on Mining Software Repositories (MSR '16)*. Association for Computing Machinery, New York, NY, USA, 425–429. https://doi.org/10.1145/2901739.2901777 event-place: Austin, Texas.

[76] Yugeng Liu, Rui Wen, Xinlei He, Ahmed Salem, Zhikun Zhang, Michael Backes, Emiliano De Cristofaro, Mario Fritz, and Yang Zhang. 2022. ML-DOCTOR: Holistic Risk Assessment of Inference Attacks Against Machine Learning Models. In *31st USENIX Security Symposium (USENIX Security 22)*. 4525–4542.

[77] Yunhui Long, Lei Wang, Diyue Bu, Vincent Bindschaedler, Xiaofeng Wang, Haixu Tang, Carl A. Gunter, and Kai Chen. 2020. A Pragmatic Approach to Membership Inferences on Machine Learning Models. In *2020 IEEE European Symposium on Security and Privacy (EuroS&P)*. 521–534. https://doi.org/10.1109/EuroSP48549.2020.00040

[78] Chris Y.T. Ma, David K.Y. Yau, Nung Kwan Yip, and Nageswara S.V. Rao. 2010. Privacy Vulnerability of Published Anonymous Mobility Traces. In *Proceedings of the Sixteenth Annual International Conference on Mobile Computing and Networking (MobiCom '10)*. Association for Computing Machinery, New York, NY, USA, 185–196. https://doi.org/10.1145/1859995.1860017 event-place: Chicago, Illinois, USA.

[79] Bradley Malin. 2005. Betrayed by my shadow: learning data identity via trail matching. *Journal of Privacy Technology* 2005 (2005), 20050609001.

[80] Dionysis Manousakas, Cecilia Mascolo, Alastair R Beresford, Dennis Chan, and Nikhil Sharma. 2018. Quantifying privacy loss of human mobility graph topology. *Proceedings on Privacy Enhancing Technologies* 2018, 3 (2018), 5–21. https://doi.org/10.1515/popets-2018-0018 Publisher: Walter de Gruyter GmbH.

[81] Mohamed Maouche, Sonia Ben Mokhtar, and Sara Bouchenak. 2017. AP-Attack: A Novel User Re-Identification Attack On Mobility Datasets. In *Proceedings of the 14th EAI International Conference on Mobile and Ubiquitous Systems: Computing, Networking and Services (MobiQuitous 2017)*. Association for Computing Machinery, New York, NY, USA, 48–57. https://doi.org/10.1145/3144457.3144494 event-place: Melbourne, VIC, Australia.

[82] Erika McCallister, Tim Grance, and Karen Scarfone. 2010. Identifiable Information (PII). *NIST Special Publication* 800 (2010), 122.

[83] Arsalan Mosenia, Xiaoliang Dai, Prateek Mittal, and Niraj K. Jha. 2018. PinMe: Tracking a Smartphone User around the World. *IEEE Transactions on Multi-Scale Computing Systems* 4, 3 (2018), 420–435. https://doi.org/10.1109/TMSCS.2017.2751462

[84] Mihir Nanavati, Nathan Taylor, William Aiello, and Andrew Warfield. 2011. Herbert West—Deanonymizer. In *6th USENIX Workshop on Hot Topics in Security (HotSec 11)*.

[85] Arvind Narayanan, Arvind Narayanan, and Vitaly Shmatikov. 2006. How To Break Anonymity of the Netflix Prize Dataset. *arXiv: Cryptography and Security* (2006). https://arxiv.org/abs/cs/0610105

[86] Arvind Narayanan, Hristo Paskov, Neil Zhenqiang Gong, John Bethencourt, Emil Stefanov, Eui Chul Richard Shin, and Dawn Song. 2012. On the Feasibility of Internet-Scale Author Identification. In *2012 IEEE Symposium on Security and Privacy*. IEEE, San Francisco, CA, USA, 300–314. https://doi.org/10.1109/SP.2012.46

[87] Arvind Narayanan, Elaine Shi, and Benjamin I. P. Rubinstein. 2011. Link prediction by de-anonymization: How We Won the Kaggle Social Network Challenge. In *The 2011 International Joint Conference on Neural Networks*. 1825–1834. https://doi.org/10.1109/IJCNN.2011.6033446 ISSN: 2161-4407.

[88] Arvind Narayanan and Vitaly Shmatikov. 2008. Robust De-anonymization of Large Sparse Datasets. In *2008 IEEE Symposium on Security and Privacy (sp 2008)*. 111–125. https://doi.org/10.1109/SP.2008.33 ISSN: 2375-1207.

[89] Arvind Narayanan and Vitaly Shmatikov. 2009. De-anonymizing Social Networks. In *2009 30th IEEE Symposium on Security and Privacy*. 173–187. https://doi.org/10.1109/SP.2009.22 ISSN: 2375-1207.

[90] Arvind Narayanan and Vitaly Shmatikov. 2010. Myths and fallacies of "Personally Identifiable Information". *Commun. ACM* 53, 6 (June 2010), 24–26. https://doi.org/10.1145/1743546.1743558

[91] Markus Nentwig, Michael Hartung, Axel-Cyrille Ngonga Ngomo, and Erhard Rahm. 2017. A survey of current link discovery frameworks. *Semantic Web* 8, 3 (2017), 419–436.

[92] Mehmet Ercan Nergiz, Maurizio Atzori, and Chris Clifton. 2007. Hiding the Presence of Individuals from Shared Databases. In *Proceedings of the 2007 ACM SIGMOD International Conference on Management of Data* (Beijing, China) *(SIGMOD '07)*. Association for Computing Machinery, New York, NY, USA, 665–676. https://doi.org/10.1145/1247480.1247554

[93] Peter Ney, Luis Ceze, and Tadayoshi Kohno. 2020. Genotype Extraction and False Relative Attacks: Security Risks to Third-Party Genetic Genealogy Services Beyond Identity Inference.. In *NDSS*. https://doi.org/10.14722/ndss.2020.23049

[94] Shirin Nilizadeh, Apu Kapadia, and Yong-Yeol Ahn. 2014. Community-Enhanced De-Anonymization of Online Social Networks. In *Proceedings of the 2014 ACM SIGSAC Conference on Computer and Communications Security (CCS '14)*. Association for Computing Machinery, New York, NY, USA, 537–548. https://doi.org/10.1145/2660267.2660324 event-place: Scottsdale, Arizona, USA.

[95] Salvador Ochoa, Jamie Rasmussen, Christine Robson, and Michael Salib. 2001. Reidentification of Individuals in Chicago's Homicide Database: A Technical and Legal Study.

[96] Lukasz Olejnik, Claude Castelluccia, and Artur Janc. 2012. Why johnny can't browse in peace: On the uniqueness of web browsing history patterns. In *5th Workshop on Hot Topics in Privacy Enhancing Technologies (HotPETs 2012)*.

[97] Wei Peng, Feng Li, Xukai Zou, and Jie Wu. 2012. Seed and Grow: An attack against anonymized social networks. In *2012 9th Annual IEEE Communications Society Conference on Sensor, Mesh and Ad Hoc Communications and Networks (SECON)*. 587–595. https://doi.org/10.1109/SECON.2012.6275831 ISSN: 2155-5494.

[98] The Pillar. 2021. Pillar Investigates: USCCB gen sec Burrill resigns after sexual misconduct allegations. https://www.pillarcatholic.com/p/pillar-investigates-usccb-gen-sec

[99] Apostolos Pyrgelis, Carmela Troncoso, and Emiliano De Cristofaro. 2018. Knock Knock, Who's There? Membership Inference on Aggregate Location Data. In *Proceedings 2018 Network and Distributed System Security Symposium*. Internet Society, San Diego, CA. https://doi.org/10.14722/ndss.2018.23183

[100] Apostolos Pyrgelis, Carmela Troncoso, and Emiliano De Cristofaro. 2020. Measuring Membership Privacy on Aggregate Location Time-Series. *Proc. ACM Meas. Anal. Comput. Syst.* 4, 2 (June 2020). https://doi.org/10.1145/3392154 Place: New York, NY, USA Publisher: Association for Computing Machinery.

[101] Shadi Rahimian, Tribhuvanesh Orekondy, and Mario Fritz. 2021. Differential Privacy Defenses and Sampling Attacks for Membership Inference. In *Proceedings of the 14th ACM Workshop on Artificial Intelligence and Security (AISec '21)*. Association for Computing Machinery, New York, NY, USA, 193–202. https://doi.org/10.1145/3474369.3486876 event-place: Virtual Event, Republic of Korea.

[102] Aditi Ramachandran, Lisa Singh, Edward Porter, and Frank Nagle. 2012. Exploring re-identification risks in public domains. In *2012 Tenth Annual International Conference on Privacy, Security and Trust*. 35–42. https://doi.org/10.1109/PST.2012.6297917

[103] Vikram Ravindra and Ananth Grama. 2021. De-Anonymization Attacks on Neuroimaging Datasets. In *Proceedings of the 2021 International Conference on Management of Data (SIGMOD '21)*. Association for Computing Machinery, New York, NY, USA, 2394–2398. https://doi.org/10.1145/3448016.3457234 event-place: Virtual Event, China.

[104] Christopher Riederer, Yunsung Kim, Augustin Chaintreau, Nitish Korula, and Silvio Lattanzi. 2016. Linking Users Across Domains with Location Data: Theory and Validation. In *Proceedings of the 25th International Conference on World Wide Web (WWW '16)*. International World Wide Web Conferences Steering Committee, Republic and Canton of Geneva, CHE, 707–719. https://doi.org/10.1145/2872427.2883002

[105] Christopher J. Riederer, Sebastian Zimmeck, Coralie Phanord, Augustin Chaintreau, and Steven M. Bellovin. 2015. "I Don't Have a Photograph, but You Can Have My Footprints.": Revealing the Demographics of Location Data. In *Proceedings of the 2015 ACM on Conference on Online Social Networks (COSN '15)*. Association for Computing Machinery, New York, NY, USA, 185–195. https://doi.org/10.1145/2817946.2817968 event-place: Palo Alto, California, USA.

[106] Nazir Saleheen, Md Azim Ullah, Supriyo Chakraborty, Deniz S. Ones, Mani Srivastava, and Santosh Kumar. 2021. WristPrint: Characterizing User Re-Identification Risks from Wrist-Worn Accelerometry Data. In *Proceedings of the 2021 ACM SIGSAC Conference on Computer and Communications Security (CCS '21)*. Association for Computing Machinery, New York, NY, USA, 2807–2823. https://doi.org/10.1145/3460120.3484799 event-place: Virtual Event, Republic of Korea.

[107] Pedro Miguel Sánchez Sánchez, Jose María Jorquera Valero, Alberto Huertas Celdrán, Gérôme Bovet, Manuel Gil Pérez, and Gregorio Martínez Pérez. 2021. A survey on device behavior fingerprinting: Data sources, techniques, application scenarios, and datasets. *IEEE Communications Surveys & Tutorials* (2021).

[108] Adrian Sayers, Yoav Ben-Shlomo, Ashley W Blom, and Fiona Steele. 2016. Probabilistic record linkage. *International journal of epidemiology* 45, 3 (2016), 954–964.

[109] Kumar Sharad. 2016. Change of Guard: The Next Generation of Social Graph De-Anonymization Attacks. In *Proceedings of the 2016 ACM Workshop on Artificial Intelligence and Security (AISec '16)*. Association for Computing Machinery, New York, NY, USA, 105–116. https://doi.org/10.1145/2996758.2996763 event-place: Vienna, Austria.

[110] Kumar Sharad and George Danezis. 2014. An Automated Social Graph De-Anonymization Technique. In *Proceedings of the 13th Workshop on Privacy in the Electronic Society (WPES '14)*. Association for Computing Machinery, New York, NY, USA, 47–58. https://doi.org/10.1145/2665943.2665960 event-place: Scottsdale, Arizona, USA.

[111] Reza Shokri, Marco Stronati, Congzheng Song, and Vitaly Shmatikov. 2017. Membership Inference Attacks Against Machine Learning Models. In *2017 IEEE Symposium on Security and Privacy (SP)*. 3–18. https://doi.org/10.1109/SP.2017.41 ISSN: 2375-1207.

[112] Reza Shokri, George Theodorakopoulos, Jean-Yves Le Boudec, and Jean-Pierre Hubaux. 2011. Quantifying Location Privacy. In *2011 IEEE Symposium on Security and Privacy*. 247–262. https://doi.org/10.1109/SP.2011.18 ISSN: 2375-1207.

[113] Anshu Singh, Shaojing Fan, and Mohan Kankanhalli. 2021. Human Attributes Prediction under Privacy-Preserving Conditions. In *Proceedings of the 29th ACM International Conference on Multimedia (MM '21)*. Association for Computing Machinery, New York, NY, USA, 4698–4706. https://doi.org/10.1145/3474085.3475687 event-place: Virtual Event, China.

[114] Mudhakar Srivatsa and Mike Hicks. 2012. Deanonymizing Mobility Traces: Using Social Network as a Side-Channel. In *Proceedings of the 2012 ACM Conference on Computer and Communications Security (CCS '12)*. Association for Computing Machinery, New York, NY, USA, 628–637. https://doi.org/10.1145/2382196.2382262 event-place: Raleigh, North Carolina, USA.

[115] Theresa Stadler, Bristena Oprisanu, and Carmela Troncoso. 2021. Synthetic Data–Anonymisation Groundhog Day. *arXiv preprint arXiv:2011.07018* (2021).

[116] Du Su, Hieu Tri Huynh, Ziao Chen, Yi Lu, and Wenmiao Lu. 2020. Re-Identification Attack to Privacy-Preserving Data Analysis with Noisy Sample-Mean. In *Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery &amp; Data Mining (KDD '20)*. Association for Computing Machinery, New York, NY, USA, 1045–1053. https://doi.org/10.1145/3394486.3403148 event-place: Virtual Event, CA, USA.

[117] Jessica Su, Ansh Shukla, Sharad Goel, and Arvind Narayanan. 2017. De-Anonymizing Web Browsing Data with Social Networks. In *Proceedings of the 26th International Conference on World Wide Web (WWW '17)*. International World Wide Web Conferences Steering Committee, Republic and Canton of Geneva, CHE, 1261–1269. https://doi.org/10.1145/3038912.3052714 event-place: Perth, Australia.

[118] Latanya Sweeney. 2000. Simple demographics often identify people uniquely. *Health (San Francisco)* 671, 2000 (2000), 1–34.

[119] Latanya Sweeney. 2002. K-anonymity: a model for protecting privacy. *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems* (2002). https://doi.org/10.1142/s0218488502001648

[120] Galini Tsoukaneri, George Theodorakopoulos, Hugh Leather, and Mahesh K. Marina. 2016. On the Inference of User Paths from Anonymized Mobility Data. In *2016 IEEE European Symposium on Security and Privacy (EuroS&P)*. 199–213. https://doi.org/10.1109/EuroSP.2016.25

[121] Zhen Tu, Fengli Xu, Yong Li, Pengyu Zhang, and Depeng Jin. 2018. A New Privacy Breach: User Trajectory Recovery From Aggregated Mobility Data. *IEEE/ACM Transactions on Networking* 26, 3 (2018), 1446–1459. https://doi.org/10.1109/TNET.2018.2829173

[122] Valentin Tudor, Magnus Almgren, and Marina Papatriantafilou. 2015. A Study on Data De-Pseudonymization in the Smart Grid. In *Proceedings of the Eighth European Workshop on System Security (EuroSec '15)*. Association for Computing Machinery, New York, NY, USA, 1–6. https://doi.org/10.1145/2751323.2751325 event-place: Bordeaux, France.

[123] Jennifer Valentino-DeVries, Natasha Singer, Michael Keller, and Aaron Krolik. 2018. Your Apps Know Where You Were Last Night, and They're Not Keeping It Secret. https://www.nytimes.com/interactive/2018/12/10/business/location-data-privacy-apps.html

[124] Giridhari Venkatadri, Athanasios Andreou, Yabing Liu, Alan Mislove, Krishna P. Gummadi, Patrick Loiseau, and Oana Goga. 2018. Privacy Risks with Facebook's PII-Based Targeting: Auditing a Data Broker's Advertising Interface. In *2018 IEEE Symposium on Security and Privacy (SP)*. 89–107. https://doi.org/10.1109/SP.2018.00014 ISSN: 2375-1207.

[125] Paul Vines, Franziska Roesner, and Tadayoshi Kohno. 2017. Exploring ADINT: Using Ad Targeting for Surveillance on a Budget - or - How Alice Can Buy Ads to Track Bob. In *Proceedings of the 2017 on Workshop on Privacy in the Electronic Society (WPES '17)*. Association for Computing Machinery, New York, NY, USA, 153–164. https://doi.org/10.1145/3139550.3139567 event-place: Dallas, Texas, USA.

[126] Isabel Wagner and David Eckhoff. 2018. Technical Privacy Metrics: A Systematic Survey. *ACM Comput. Surv.* 51, 3, Article 57 (jun 2018), 38 pages. https://doi.org/10.1145/3168389

[127] Huandong Wang, Chen Gao, Yong Li, Gang Wang, Depeng Jin, and Jingbo Sun. 2018. De-anonymization of mobility trajectories: Dissecting the gaps between theory and practice. In *The 25th Annual Network & Distributed System Security Symposium (NDSS'18)*. https://doi.org/10.14722/ndss.2018.23211

[128] Ningfei Wang, Shouling Ji, and Ting Wang. 2018. Integration of Static and Dynamic Code Stylometry Analysis for Programmer De-Anonymization. In *Proceedings of the 11th ACM Workshop on Artificial Intelligence and Security (AISec '18)*. Association for Computing Machinery, New York, NY, USA, 74–84. https://doi.org/10.1145/3270101.3270110 event-place: Toronto, Canada.

[129] Charlie Warzel and Stuart A. Thompson. 2021. They Stormed the Capitol. Their Apps Tracked Them. https://www.nytimes.com/2021/02/05/opinion/capitol-attack-cellphone-data.html

[130] Gilbert Wondracek, Thorsten Holz, Engin Kirda, and Christopher Kruegel. 2010. A Practical Attack to De-anonymize Social Network Users. In *2010 IEEE Symposium on Security and Privacy*. 223–238. https://doi.org/10.1109/SP.2010.21 ISSN: 2375-1207.

[131] Fengli Xu, Zhen Tu, Yong Li, Pengyu Zhang, Xiaoming Fu, and Depeng Jin. 2017. Trajectory recovery from ash: User privacy is not preserved in aggregated mobility data. In *Proceedings of the 26th international conference on world wide web*. 1241–1250.

[132] Hui Zang and Jean Bolot. 2011. Anonymization of location data does not work: A large-scale measurement study. In *Proceedings of the 17th annual international conference on Mobile computing and networking*. 145–156.

[133] Jiexin Zhang, Alastair R Beresford, and Ian Sheret. 2019. Sensorid: Sensor calibration fingerprinting for smartphones. In *2019 IEEE Symposium on Security and Privacy (SP)*. IEEE, 638–655.

[134] Qingchuan Zhao, Chaoshun Zuo, Giancarlo Pellegrino, and Li Zhiqiang. 2019. Geo-locating Drivers: A Study of Sensitive Data Leakage in Ride-Hailing Services.. In *Annual Network and Distributed System Security symposium, February 2019 (NDSS 2019)*. https://doi.org/10.14722/ndss.2019.23052

[135] Shoshana Zuboff. 2015. Big other: surveillance capitalism and the prospects of an information civilization. *Journal of information technology* 30, 1 (2015), 75–89.

A CATEGORISATION OF SURVEYED ATTACKS AS INTENTS

Our literature survey, described in Section 5, yielded 94 attacks from a sample of 418 papers. Table 1 shows these 94 attacks categorised by intent and repeatability (outlier, mass, or total), as well as the primary data type of the dataset in question. Repeatability is not applicable in cases of class attacks, as well as some papers which were ambiguously presented. Similarly, some attacks were presented in a datatype-agnostic way, sometimes as a break of a particular aggregation technology or PET, or were validated across a number of datatypes and contexts, and in those cases that column is left blank. There are entries in the table (membership inference on ML models and deanonymisation of online social networks) which represent multiple attacks each—this is in the case of attacks that all have the same intent and properties. These attacks often build on each other, representing their own mini-literatures.

Table 1: A systematisation of 94 identified attacks from across top academic literature and influential work in the field of anonymity and identifiability. Each paper's title is given except where series of work on the same problem have been condensed (in *italics*). Each attack is described in terms of its intent, its repeatability—○ for Outlier, ◐ for Mass, and ● for Total—and the data type under attack.

Title	Reference(s)	Intent	Repeatability	Data type
PinMe: Tracking a Smartphone User around the World	[83]	Singling out	○	Onboard phone sensors
Genotype Extraction and False Relative Attacks: Security Risks to Third-Party Genetic Genealogy Services Beyond Identity Inference.	[93]	Singling out	◐	Genomic
Trajectory recovery from ash: User privacy is not preserved in aggregated mobility data	[121]	Singling out	◐	Aggregate location data
Full Database Reconstruction in Two Dimensions	[47]	Singling out	●	-
On the Inference of User Paths from Anonymized Mobility Data	[120]	Singling out	●	Location instances
Unique in the crowd: The privacy bounds of human mobility	[35]	Singling out	●	Location traces
The risk of re-identification remains high even in country-scale location datasets	[49]	Singling out	●	Location traces
AOL search logs	[14]	Untargeted reidentification	○	Search logs
Herbert {West—Deanonymizer}	[84]	Untargeted reidentification	◐	Academic text
How much is too much? Leveraging ads audience estimation to evaluate public profile uniqueness	[26]	Untargeted reidentification	◐	Ad audiences
Browsing Unicity: On the Limits of Anonymizing Web Tracking Data	[39]	Untargeted reidentification	◐	Browsing history
Dissecting Privacy Risks in Biomedical Data	[17]	Untargeted reidentification	◐	DNA methylation
On the Unicity of Smartphone Applications	[3]	Untargeted reidentification	◐	App installs
Provable De-anonymization of Large Datasets with Sparse Dimensions	[34]	Untargeted reidentification	◐	Movie ratings
Re-identification of Smart Meter data	[21]	Untargeted reidentification	◐	Energy usage
Reidentification of Individuals in Chicago's Homicide Database: A Technical and Legal Study	[95]	Untargeted reidentification	◐	Municipal records
Revisiting the uniqueness of simple demographics in the US population	[56]	Untargeted reidentification	◐	Census data
Unique on Facebook: Formulation and Evidence of (Nano)Targeting Individual Users with Non-PII Data	[60]	Untargeted reidentification	◐	Facebook profiles
Why johnny can't browse in peace: On the uniqueness of web browsing history patterns	[96]	Untargeted reidentification	◐	Browsing history
A study on the re-identifiability of Dutch citizens	[74]	Untargeted reidentification	●	Census records
Even Black Cats Cannot Stay Hidden in the Dark: Full-band De-anonymization of Bluetooth Classic Devices	[29]	Untargeted reidentification	●	Bluetooth packets
Identifying Personal DNA Methylation Profiles by Genotype Inference	[10]	Untargeted reidentification	●	DNA methlyation
Privacy in epigenetics: temporal linkability of MicroRNA expression profiles	[11]	Untargeted reidentification	●	MicroRNA expression
Privacy through Pseudonymity in Mobile Telephony Systems.	[8]	Untargeted reidentification	●	Cell tower records
Quantifying privacy loss of human mobility graph topology	[80]	Untargeted reidentification	●	POI
Replication: Whywe Still Can't Browse in Peace: On the Uniqueness and Reidentifiability of Web Browsing Histories	[18]	Untargeted reidentification	●	Browsing history
WristPrint: Characterizing User Re-Identification Risks from Wrist-Worn Accelerometry Data	[106]	Untargeted reidentification	●	Wrist-worn accelerometer logs
SensorID: Sensor calibration fingerprinting for smartphones	[133]	Untargeted reidentification	●	Onboard phone sensors
On the anonymity of home/work location pairs	[57]	Untargeted reidentification	●	Coarse location pairs
You are what you like! information leakage through users' interests	[25]	Class attribute inference		Facebook profiles
Birds of a Feather Flock Together: How Set Bias Helps to Deanonymize You via Revealed Intersection Sizes	[64]	Class membership estimation		-
Exploring ADINT: Using Ad Targeting for Surveillance on a Budget - or - How Alice Can Buy Ads to Track Bob	[125]	Class membership estimation		Ad impressions
Privacy Risks with Facebook's PII-Based Targeting: Auditing a Data Broker's Advertising Interface	[124]	Class membership estimation		Facebook profiles
Exploring ADINT: Using Ad Targeting for Surveillance on a Budget - or - How Alice Can Buy Ads to Track Bob	[125]	Individual profiling		Ad impressions

Table 1: A systematisation of 94 identified attacks from across top academic literature and influential work in the field of anonymity and identifiability. Each paper's title is given except where series of work on the same problem have been condensed (in *italics*). Each attack is described in terms of its intent, its repeatability—○ for Outlier, ◐ for Mass, and ● for Total—and the data type under attack.

Title	Reference(s)	Intent	Repeatability	Data type
You Might Also Like: Privacy Risks of Collaborative Filtering	[22]	Individual profiling	◐	Transactions
Geo-locating Drivers: A Study of Sensitive Data Leakage in Ride-Hailing Services.	[134]	Individual profiling	◐	Ride share trips
Privacy Vulnerability of Published Anonymous Mobility Traces	[78]	Individual profiling	◐	Location traces
Capitol rioters	[129]	Individual profiling	◐	Location, advertising data
General Graph Data De-Anonymization: From Mobility Traces to Social Networks	[72]	Individual profiling	●	-
Re-Identification Attack to Privacy-Preserving Data Analysis with Noisy Sample-Mean	[116]	Individual profiling	●	MNIST
I Don't Have a Photograph, but You Can Have My Footprints.: Revealing the Demographics of Location Data	[105]	Targeted attribute inference		Location traces
Human Attributes Prediction under Privacy-Preserving Conditions	[113]	Targeted attribute inference		-
ML-DOCTOR: Holistic Risk Assessment of Inference Attacks Against Machine Learning Models	[76]	Targeted attribute inference	◐	-
How Much Do Your Friends Know about You? Reconstructing Private Information from the Friendship Graph	[19]	Targeted attribute inference	◐	Social network
An Inference Attack on Genomic Data Using Kinship, Complex Correlations, and Phenotype Information	[40]	Targeted attribute inference	◐	Genomic and phenotypic
Attribute Inference Attacks in Online Social Networks	[59]	Targeted attribute inference	◐	Social network
Recognizing Gender of Stack Overflow Users	[75]	Targeted attribute inference	◐	StackOverflow profiles
You are who you know and how you behave: Attribute inference attacks via users' social friends and behaviors	[58]	Targeted attribute inference	◐	Social network
Analysis of Privacy Protections in Fitness Tracking Social Networks-or-You can run, but can you hide?	[65]	Targeted attribute inference	●	Location traces
Membership inference on machine learning models	[76, 77, 99, 101, 111]	Membership inference	◐	-
The Inevitable Weaponization of App Data Is Here	[20, 31]	Blind targeted reidentification	○	Location, advertising data
Deanonymisation of online social networks	[41, 51, 52, 63, 89, 94, 109, 110, 130]	Blind targeted reidentification	◐	Social network
Integration of Static and Dynamic Code Stylometry Analysis for Programmer De-Anonymization	[128]	Blind targeted reidentification	◐	Code
De-Anonymizing Web Browsing Data with Social Networks	[117]	Blind targeted reidentification	◐	Browsing history
Exploiting innocuous activity for correlating users across sites	[55]	Blind targeted reidentification	◐	Social media posts
Fraud De-Anonymization for Fun and Profit	[66]	Blind targeted reidentification	◐	Fraud posts
Measuring Membership Privacy on Aggregate Location Time-Series	[100]	Blind targeted reidentification	◐	Location
Robust De-anonymization of Large Sparse Datasets	[88]	Blind targeted reidentification	◐	Movie ratings
The 'Re-Identification' of Governor William Weld's Medical Information: A Critical Re-Examination of Health Data Identification Risks and Privacy Protections, Then and Now	[16]	Nosy targeted reidentification	○	Medical records
Deanonymisation of online social networks	[19, 70, 71, 87, 97]	Nosy targeted reidentification	◐	Social networks
Doppelgänger Finder: Taking Stylometry to the Underground	[5]	Nosy targeted reidentification	◐	Forum posts
Exploring Linkability of User Reviews	[6]	Nosy targeted reidentification	◐	Yelp reviews
A Study on Data De-Pseudonymization in the Smart Grid	[122]	Nosy targeted reidentification	◐	Energy usage
AP-Attack: A Novel User Re-Identification Attack On Mobility Datasets	[81]	Nosy targeted reidentification	◐	Location traces
Betrayed by my shadow: learning data identity via trail matching	[79]	Nosy targeted reidentification	◐	Location traces
De-Anonymization Attacks on Neuroimaging Datasets	[103]	Nosy targeted reidentification	◐	MRI
De-anonymization of mobility trajectories: Dissecting the gaps between theory and practice	[127]	Nosy targeted reidentification	◐	Location traces
De-anonymizing genomic databases using phenotypic traits	[68]	Nosy targeted reidentification	◐	Genome, phenotype

Table 1: A systematisation of 94 identified attacks from across top academic literature and influential work in the field of anonymity and identifiability. Each paper's title is given except where series of work on the same problem have been condensed (in *italics*). Each attack is described in terms of its intent, its repeatability—○ for Outlier, ◐ for Mass, and ● for Total—and the data type under attack.

Title	Reference(s)	Intent	Repeatability	Data type
Deanonymizing Mobility Traces: Using Social Network as a Side-Channel	[114]	Nosy targeted reidentification	◐	Co-locations
Exploring re-identification risks in public domains	[102]	Nosy targeted reidentification	◐	Census, OSN
Genome Reconstruction Attacks Against Genomic Data-Sharing Beacons	[9]	Nosy targeted reidentification	◐	Genomic data
Linking Users Across Domains with Location Data: Theory and Validation	[104]	Nosy targeted reidentification	◐	Location
On the Feasibility of Internet-Scale Author Identification	[86]	Nosy targeted reidentification	◐	Blog posts
Walk2friends: Inferring Social Links from Mobility Profiles	[13]	Nosy targeted reidentification	◐	Location traces
De-anonymizing programmers via code stylometry	[23]	Nosy targeted reidentification	●	Code
Large-Scale and Language-Oblivious Code Authorship Identification	[2]	Nosy targeted reidentification	●	Code
Dissecting Privacy Risks in Biomedical Data	[17]	Nosy targeted reidentification	●	DNA methylation