ISSUE BRIEF

U.S. POLICY

PETs Use Case:

Differential Privacy for End-of-Life Data

Organizations: Oblivious; unnamed insurance company

FUTURE OF PRIVACY FORUM

Problem:
Meeting regulatory requirements to delete personal data at the end of the data's lifecycle presents a critical challenge for organizations. Under many privacy laws, companies are obligated to erase personal data once its retention period expires or when users exercise their "right to be forgotten." However, this can result in the loss of valuable macro-level insights and historical trends that could help inform strategic decisions. Differential Privacy can help organizations extract and retain aggregate-level information while ensuring that individual-level data is permanently erased and regulatory obligations are met. By anonymizing data before deletion, Differential Privacy allows businesses to generate summaries, trends, and patterns that do not compromise individual privacy. When data reaches its scheduled deletion, the organization processes it into Differentially Private flat files for future analysis. The flat file is saved, and the underlying microdata are then deleted. In this use case, an insurance company uses Differential Privacy to maintain more accurate actuarial tables.

PET Details

Technology Type:
The core Privacy Enhancing Technology (PET) in this use case is **Differential Privacy**.

What is Differential Privacy?

Differential Privacy is a controlled process of adding noise to data to protect privacy. Several factors inform how much noise is added to the data and how much analysis is allowed. These factors are combined into what is called a "privacy-loss budget," which is represented numerically. Differential Privacy may be applied at query time or to an entire dataset prior to sharing.

For query-based Differential Privacy, each interaction with a dataset that reveals an output consumes a portion of an overall "privacy budget," which represents the maximum allowable privacy loss. Imagine the privacy budget is set to 4. Each action that reveals information from the dataset might consume a fixed amount of this budget, for example, 0.5 per access. With a budget of 4, a user could interact with the dataset 8 times (0.5 × 8 = 4) before exhausting the budget. Once the budget is fully spent, any further access would be blocked until additional budget is assigned. One advantage of this budgeting approach is that it gives researchers flexibility to focus their budget on the most relevant parts of the dataset, allowing them to maximize the accuracy of the insights they care about most.

However, Differential Privacy may be applied to an entire dataset prior to release. In this case, the entire budget is spent to create a static dataset. Noise is added to individual records; the greater the noise, the lower the total privacy-loss budget. An advantage of this type of Differential Privacy is that it enables a one-time calculation of privacy loss and continual re-use of the same underlying data.

Functionality:
An insurance organization uses Differential Privacy to create anonymized “snapshots” of important statistics about a population at different points in time that it can keep for future use and analysis, even after the data the snapshots came from is deleted. The organization agrees upon the statistics they want from the data, which are influenced by the intended use and sector (health insurance, life insurance, etc.). A script automates statistical calculations without the need for ad-hoc analysis by any data analyst or data scientist. Differentially private snapshots are taken at regular time intervals. Determining the length of those intervals is critical as it influences the choice of epsilon. After the organization applies Differential Privacy to the data, the Differentially Private outputs are securely stored, and the original data is permanently deleted. The process requires integration with the organization’s enterprise cloud environment, including data connectors and a governance framework to manage privacy budgets.

Beneficiaries
Insurance Companies: Accurate actuarial data can lead to more accurate risk analysis for insurers. These companies assess risk by collecting historical data, and when they can gauge risk more accurately, they can set more accurate premiums. In insurance, differentially private aggregates could support the generation and continuous refinement of actuarial tables, including anonymized summaries of claim frequency by age group, policy duration, regional climate risk exposure, or chronic health indicators. These statistics are foundational to underwriting, risk pooling, and reserve planning.

Insurance Consumers: In a competitive market, insurance premiums could be strategically lowered based on the increased accuracy of premiums, giving the company a market advantage and potentially a better rate for the consumer.

Industry/Domain(s):
Insurance; Finance

Stakeholders:
- Can we name the insurance company involved in this?
- National Association of Insurance Commissioners (NAIC)?
- FTC or CFPB?
- Any state agencies?
- Insurers themselves?
- Re-insurers (who should get more comfortable with insurers who do this)?
- Insurance regulators?
- Privacy regulators?
- Insurance purchasers?

Current Regulatory Environment:
The companies operate in a multi-jurisdiction environment with multiple privacy frameworks that set the right to erasure/deletion and data retention laws, as well as local industry standards that impact data usage in particular:

European Union (EU):

- The **General Data Protection Regulation** (GDPR), the **Right to be Forgotten** (Article 17), the **Data Minimization Principle** (Article 5), and purpose limitation requirements apply to this use case.
- The **European Insurance and Occupational Pensions Authority** (EIOPA) sets guidelines for using data for actuarial modeling and risk assessment.

Canada

- The **Personal Information Protection and Electronic Documents Act** (PIPEDA) requires the deletion of personal data when it is no longer needed, but permits anonymization for aggregate analysis.

United Kingdom (UK):

- UK, GDPR, and the **Data Protection Act 2018** align with EU GDPR, mandating the right to erasure, data minimization, and processing limitations.
- The **UK Financial Conduct Authority**.

United States (US):

- Federal Regulations: The **Health Information Portability and Accountability Act** (HIPAA) covers health insurance data.
- The **California Privacy Protection Act** (CCPA) and the **California Privacy Rights Act** (CPRA) grant consumers the right to delete their data while allowing aggregated, anonymized data to be retained.
- The **Virginia Consumer Data Protection Act** (VCDPA), the **Colorado Privacy Act** (CPA), the **Connecticut Data Privacy Act** (CTDPA), and the **Utah Consumer Privacy Act** (UCPA) all contain data deletion rights with provisions for anonymized data.

Australia:

- The **Privacy Act of 1988** and the **Australian Privacy Principles** (APPs) require organizations to delete or de-identify personal data at the end of its lifecycle.
- The **Insurance Contracts Act 1984** mandates insurers to maintain actuarial records while complying with privacy regulations.

Benefits of PET Implementation

Privacy Benefits:
The organization can permanently erase microdata while retaining analytical utility by processing the data into a differentially private flat file before deletion. This contrasts with traditional approaches

where the raw data is kept (risking privacy) or fully deleted (losing insights). Differential Privacy offers a privacy-preserving middle ground.

Operational Benefits:
Differential Privacy enables organizations to preserve key macro-level insights from historical data that would otherwise be deleted. This allows actuarial, risk, and forecasting teams to maintain modeling continuity. It prevents the operational disruption that typically results from hard data-deletion policies and supports more consistent decision-making and pricing strategies.

Data Security Benefits:
Organizations that use Differentially Private data rely more on descriptive statistics than record-level information, potentially reducing the risk of data leaks, breaches, and some forms of insider threat.

Potential for Broader Use:
If regulators recognize Differential Privacy-protected outputs more clearly as non-personal data, healthcare, education, retail, and finance organizations could confidently retain valuable statistical summaries after deleting raw data. Similarly, differentially private snapshots can have applications beyond the insurance sector. In the financial sector, retained statistics could be used to model credit risk trends, anonymized loan default rates across income brackets, savings behavior segmented by age or region, or time-series patterns of fraud detection, all while ensuring that no individual's financial data is retained. Differential Privacy thus provides a tool for maintaining institutional memory without retaining sensitive microdata.

Economic Impact:
In competitive markets with high variance, more accurate data could reduce people's premiums, increase insurance companies' profit, or both. This could be especially true for profit-regulated markets.

Research Impact:
Differential Privacy is instrumental in facilitating safe data sharing in research, particularly in fields handling sensitive information. Implementing mechanisms that inject noise into data or analysis processes allows for sharing insights and aggregate information without compromising individual privacy. This is crucial in research environments where data sharing and collaboration are essential, but privacy concerns are paramount. Differential Privacy enables researchers to access and analyze data safely, fostering collaboration and innovation while adhering to ethical data privacy standards. This approach is particularly beneficial in social and health sciences, where sensitive data is often critical for research but requires stringent privacy protection.

Risk and Ethical Analysis

Ethical Considerations:
Implementing Differential Privacy raises several ethical considerations. While it aims to protect individual privacy, determining the appropriate level of privacy protection involves ethical decision-making. For example, transparency in how Differential Privacy is implemented and communicated to data subjects is crucial for maintaining trust. Increasing insurance companies' ability to predict risk through differential privacy may also enable underserved markets to maintain coverage that private insurers would otherwise be unwilling to provide (e.g., homeowners' policies in areas at risk from increasingly severe weather events).

However, there is a central philosophical and ethical consideration in this use case: what does it mean to delete data? If deletion means scrubbing the ones and zeros that represent digital data, this process would not be appropriate. However, if deletion is to contain a broader meaning – to remove the ability for someone to act on an individual's data, to identify them with it, etc., then removing an individual from a differentially private data set actually introduces the very risks that deletion is supposed to protect against.

The fundamental ethical question in this use case is – as so many often are – about the preeminence of the rights of individuals or groups. What's different about utilizing differential privacy in this way is that it allows for careful titration of benefits between individuals and groups instead of a binary.

Ultimately, this question will require additional research, ranging from customers' expectations to how one might provide meaningful, informed consent for data use in this fashion.

Privacy Risks:
The primary privacy risk lies in setting privacy parameters (especially epsilon, ε) too high, which weakens the privacy guarantee. A poorly tuned configuration may leak more information than intended, especially if decision-makers prioritize utility over protection.

Trade-offs:
As noted above, the fundamental tradeoff here is best seen in shades of gray: with differentially private data retention, the ultimate "answer" need not be a full optimization of group or individual benefit, but instead a blend of the two.

Operational Risks:
Setting the privacy budget is particularly important because the system operates in a dynamic setting where personal data is permanently deleted after being transformed into differentially private aggregates. This process means decisions about the privacy budget (epsilon) and query frequency must be made upfront, with no opportunity to revisit the underlying microdata.

Known Regulatory Challenges and Barriers

Ambiguities:
Regulators could help clarify what constitutes an “acceptable” privacy budget in such a context — especially where statistical accuracy influences risk modeling and premium setting — and whether differentially private outputs are considered anonymous under legal definitions. Future guidance could also support best practices for managing privacy budgets when decisions must be locked in before data deletion, including how to document and justify parameter selection in audit-ready ways. Key questions include: How often should differentially private analyses be run? What epsilon level is appropriate to ensure individual privacy while preserving longitudinal utility for actuarial purposes?

1350 EYE STREET NW | SUITE 350 | WASHINGTON, DC 20005

info@fpf.org | FPF.ORG