Privacy Phenomenon Series

The Privacy Piñata Paradox™

When "*Better Service*" Incentives Drive PII Hoarding...

and *Risk* Compounds Faster than Value

Scott Foote

Last Updated: 12 February 2026

Phenomenati Consulting
www.phenomenati.com

6 Liberty Square, #2736
Boston, MA 02109
(508) 709-7990 (office)

CONFIDENTIALITY NOTICE: The contents of this document, including any attachments, are intended solely for stakeholders of Phenomenati Consulting, may contain confidential and/or privileged information, and are legally protected from disclosure.

<this page is intentionally blank>

Contents

Executive Summary

Modern organizations increasingly collect and retain personally identifiable information (PII) in the name of convenience, personalization, fraud prevention, and product improvement. This accumulation can produce real customer benefits, but it also creates a paradox: the more complete the "*customer understanding*" becomes, the more the organization resembles a high-value "***piñata***" for threat actors... an attractive trove that invites persistent attack, misuse, and legal exposure. This short paper defines the **Privacy Piñata Paradox™**, explains the economic and operational forces that cause PII collection to expand over time, and shows why risk grows *nonlinearly* as data becomes more concentrated, linkable, and durable. It concludes with practical design and governance strategies (e.g., data minimization, purpose limitation, privacy engineering, and modern security controls) that reduce "***piñata value***" while preserving legitimate customer outcomes ([1]; [2]; [3]).

Keywords: personally identifiable information (PII); privacy risk management; data minimization; retention; ransomware; zero trust

1 Introduction: The Paradox in Plain Terms

Companies often justify collecting more PII with a customer-centric narrative: "*We need your data to personalize experiences*." "*More signals help us prevent fraud and keep accounts safe*." "*Better analytics means better products*." Each rationale can be valid. Yet organizations rarely stop at "enough." Over time, PII collection tends to ratchet upward: new features ask for new fields; new partnerships require new identifiers; new tracking tools add new behavioral exhaust ([3]). What begins as "helpful" becomes structural dependence.

The Privacy Piñata Paradox™ captures this tension: The same data hoard that promises better service also becomes an increasingly irresistible target... amplifying security risk, misuse/abuse risk, and liability faster than customer value grows ([4]; [5]).

A piñata is colorful, desirable, and full of valuables... but it is also designed to be hit until it breaks. A rich PII repository can function similarly: it attracts attention, repeated attempts to compromise it, and escalating pressure until controls fail... or until the organization itself misuses the information under business incentives.

2 What Counts as PII... and Why the "Trove" Matters

PII is commonly thought of as direct identifiers (name, address, Social Security Number in the U.S.). In practice, the most powerful, and risky, datasets are *composite* identities, where multiple attributes become linkable into a unified profile ([1]).

These profiles typically include *direct* identifiers (e.g., name, email, phone, government IDs), *quasi*-identifiers (e.g., date of birth, ZIP code, device IDs, IP addresses), *behavioral* data (e.g., location history, browsing/app activity, purchase patterns), *sensitive* attributes (e.g., biometrics, health indicators, financial data, children's data), and *linkage keys* (e.g., customer IDs, hashed identifiers, advertising IDs, cross-device graphs).

Risk increases sharply when data is linkable... when separate pieces can be joined into a unified profile. A *single record* may be moderately sensitive; a *joined profile* can be profoundly revealing, durable, and monetizable ([6]).

3 Why Companies Keep Collecting: the PII Ratchet

Organizations rarely decide, in one moment, to build a giant PII hoard. Instead, accumulation happens via incremental choices that each look reasonable locally ([3]).

3.1 "Better Service" and Personalization

Personalization is a real differentiator. But the logic tends to expand: if a name improves service, then a birthday improves it too; if birthday improves it, then location history may help; if location helps, then device fingerprinting can reduce friction. This creates a feature-driven expansion cycle: new experiences depend on new data, which becomes embedded in product requirements ([3]).

3.2 Fraud Prevention and "Security Theater" Collection

Fraud prevention often motivates collecting extra identifiers, device telemetry, and behavioral signals. Some are necessary; some are "*just in case*." Over-collection can become a substitute for stronger design because "*more data*" feels like control... even when it increases breach impact ([7]; [8]).

3.3 Analytics, Machine Learning, and "Data as Optionality"

PII is frequently retained because it preserves future options: "*We don't know how we'll use it, but it might be useful later.*" "*More training data improves models*." "*We can monetize insights.*" This mindset treats PII as a strategic asset. The *paradox* is that optionality for the business becomes optionality for attackers ([3]; [5]).

3.4 Vendor Sprawl and Shadow Accumulation

Modern stacks include CRM platforms, marketing automation, data warehouses, Customer Data Platforms (CDP), analytics SDKs, A/B testing tools, customer support systems, and payment processors. Each adds copy paths, logs, and integrations. Even if a company *intends* restraint, the ecosystem can create unplanned replication of PII across systems with uneven security maturity ([3]; [8]).

3.5 Function Creep and Shifting Norms

Data collected for one purpose often migrates to another: support data becomes training data; usage telemetry becomes marketing segmentation; "safety" signals become product gating. This is **function creep**... a gradual broadening of use that increases privacy risk and regulatory exposure ([9]).

4 Why Risk Expands Faster than Value

A key insight of the Privacy Piñata Paradox is that ***risk is not linear***. As PII accumulates, multiple compounding effects occur ([4]).

4.1 Concentration Increases Attractiveness

A large dataset is a higher-value prize. Threat actors optimize effort-to-payoff. A company holding rich, linkable PII offers identity theft potential, account takeover leverage, social engineering fuel, resale value in criminal markets, and extortion and ransom opportunities. This means *attack probability rises* with the perceived value of the target ([5]).

4.2 Richness Increases Blast Radius

If a breach occurs, impact depends on what's inside. With richer PII: harm to individuals is more severe; remediation costs rise; reputational damage deepens; and legal/regulatory consequences intensify ([1]; [2]).

4.3 Linkability Turns "Data" into "Identity Infrastructure"

When PII is joinable across contexts, it can enable persistent tracking and de-anonymization of datasets that were assumed "*non-PII*," as well as *inference* of sensitive traits ([6]; [10]). This makes the dataset more valuable and more dangerous... often beyond what the original collectors anticipated.

4.4 Long Retention Multiplies Exposure Time

Retaining data "*forever*" ensures it will eventually face new vulnerabilities, new threat actors, new reuse scenarios, new regulatory standards, and new internal access patterns. Even if defenses improve, *time* is the *enemy*: the longer a trove exists, the more chances there are for compromise or misuse ([2]; [8]).

4.5 Liability Expands with Scope, Promises, and Expectations

As collection grows, so do *obligations*: compliance duties (privacy laws, sector regulations), contractual duties, representations in privacy policies and marketing, and litigation exposure after incidents ([2]; [11]). The paradox is that a "*customer-first*" data strategy can, if overextended, become a *customer-harm amplifier* in the incident scenario.

5 Threat Actors and the “Piñata Economy”

A PII trove attracts different categories of adversaries: **cybercriminals** seeking monetizable identity material; **ransomware groups** leveraging extortion; **fraud rings** using PII for synthetic identities and account takeovers; **insiders** abusing access for personal gain or grievance; **competitors** or **data brokers** seeking illicit advantage; and **nation-state** or **political actors** targeting high-profile datasets ([4]; [12]). The common denominator is *incentive*: more PII, more pathways to *profit* or *power*.

6 Measuring the Paradox: a Simple Risk Framing

Classic risk framing helps explain why things get worse quickly:

Risk ≈ **Probability** of compromise × **Impact** of compromise ([4]).

PII hoarding tends to increase *both*: ***probability*** increases because the trove becomes a more attractive and visible target, often distributed across more systems and vendors; ***impact*** increases because the trove becomes more sensitive, linkable, and consequential ([1]; [3]). That’s **the paradox** in operational terms: ***data accumulation can be a product strategy that doubles as a threat strategy***.

7 Escaping the Paradox Without Sacrificing Service

Any DPO would explain, the goal is not “*collect nothing*.” It is collect what you can *justify*, *protect* what you must keep, and *delete* what you don’t need... while designing products that deliver value *without* building a fragile ***identity stockpile*** ([2]; [8]). This is **Privacy Engineering 101**.

7.1 Data Minimization and Purpose Limitation as Product Requirements

Make *minimization* a first-class design constraint: ask what is the minimum data required for this feature today; refuse “*just in case*” fields by default; and treat new PII as a “*high-cost dependency*” requiring explicit approval ([2] ; [7]; [13]).

7.2 Progressive Profiling and Customer-Controlled Enrichment

Instead of collecting everything up front: request sensitive data only when a user triggers a feature that truly needs it; provide clear value exchange at the moment of collection; and offer ***dashboards*** for users to view, correct, and delete data ([2]; [3]).

7.3 Retention Limits and Deletion that Actually Works

Data *Retention* is one of the most underused risk levers: define retention schedules by data category and purpose; ensure backups, logs, warehouses, and vendors follow the same rules; and build deletion as an end-to-end system behavior, not a manual process ([2]; [8]).

7.4 Reduce Linkability: Tokenize, Segment, and Separate Duties

Architecturally reduce “*one dataset to rule them all*”: tokenize identifiers; avoid spreading raw PII across systems; *segment* sensitive attributes into higher-trust zones; enforce separation between marketing, support, and security datasets where feasible; and enforce least-privilege access and strong auditing ([1]; [3]).

7.5 Privacy Engineering Techniques for Analytics and ML

When personalization and analytics are needed: prefer *aggregation* and *sampling* over raw event hoarding; use *privacy-preserving* methods where appropriate; and create data use inventories that document what data fuels which outcomes ([3]).

7.6 Security Controls Aligned to “Piñata Value”

Match defenses to the value of what you hold: strong encryption at rest and in transit; mature key management; continuous monitoring and anomaly detection; zero-trust access patterns; phishing-resistant MFA for internal tools; secure SDLC practices; and vendor/third-party risk management (TPRM) with enforced security baselines ([1]; [3]; [14]).

7.7 Governance That Closes the Ratchet

Put ***friction*** in the right places: a PII review board for new collection (product + legal + security + privacy engineering); *data classification* schemes tied to controls; regular *data mapping* and system inventories; and metrics such as number of PII fields collected, retention duration, system replication count, and vendor copies ([3]). Governance should not merely document the trove... it should prevent uncontrolled growth.

8 Conclusion: Treat PII Like Explosive Material, not “Free Insight”

The Privacy Piñata Paradox is a predictable outcome of modern business incentives: PII collection promises *personalization*, safety, and innovation, but accumulation creates an irresistible *concentration* of value that attracts malicious attention and magnifies liability ([1]; [2]; [5]). The paradox is not solved by slogans or policies alone; it is solved by design.

Organizations that escape the paradox do three things consistently:

(1) *minimize* what they collect and how long they keep it;
(2) *decouple* systems so compromise doesn’t expose the whole identity graph; and
(3) *prove trustworthiness* through transparent controls, enforceable governance, and deletion that truly eliminates risk ([2]; [3]; [8]).

In practice, the safest “piñata” is the one that contains less candy... and is harder to hit.

9 References

[**1**] McCallister, E., Grance, T., & Scarfone, K. (2010). Guide to Protecting the Confidentiality of Personally Identifiable Information (PII) (NIST Special Publication 800-122). National Institute of Standards and Technology. https://csrc.nist.gov/pubs/sp/800/122/final

[**2**] European Union. (2016). Regulation (EU) 2016/679 (General Data Protection Regulation), Article 5 (Principles relating to processing of personal data). https://eur-lex.europa.eu/eli/reg/2016/679/oj/eng

[**3**] National Institute of Standards and Technology (NIST). (2020). NIST Privacy Framework: A Tool for Improving Privacy through Enterprise Risk Management (Version 1.0). https://www.nist.gov/privacy-framework/privacy-framework

[**4**] National Institute of Standards and Technology (NIST). (2012). Guide for Conducting Risk Assessments (Special Publication 800-30 Revision 1). https://csrc.nist.gov/pubs/sp/800/30/r1/final

[**5**] Cybersecurity and Infrastructure Security Agency (CISA). (n.d.). StopRansomware: Ransomware Guide. https://www.cisa.gov/stopransomware/ransomware-guide

[**6**] Sweeney, L. (2000). Simple Demographics Often Identify People Uniquely. Carnegie Mellon University, Data Privacy Laboratory. https://dataprivacylab.org/projects/identifiability/paper1.pdf

[**7**] National Institute of Standards and Technology (NIST). (2017). Digital Identity Guidelines: Enrollment and Identity Proofing (Special Publication 800-63A). https://pages.nist.gov/800-63-3/sp800-63a.html

[**8**] Federal Trade Commission (FTC). (n.d.). Protecting Personal Information: A Guide for Business. https://www.ftc.gov/business-guidance/resources/protecting-personal-information-guide-business-0

[**9**] Lyon, D. (2010). Surveillance, Power and Everyday Life. In K. D. Haggerty & R. V. Ericson (Eds.), The Routledge Handbook of Surveillance Studies (excerpt). https://www.dhi.ac.uk/san/waysofbeing/data/governance-crone-lyon-2010c.pdf

[**10**] Narayanan, A., & Shmatikov, V. (2008). Robust De-anonymization of Large Sparse Datasets. IEEE Symposium on Security and Privacy. https://www.cs.cornell.edu/~shmat/shmat_oak08netflix.pdf

[**11**] Information Commissioner's Office (ICO). (2023). A guide to the data protection principles (UK GDPR). https://ico.org.uk/for-organisations/uk-gdpr-guidance-and-resources/data-protection-principles/a-guide-to-the-data-protection-principles/

[**12**] Cybersecurity and Infrastructure Security Agency (CISA). (2025). #StopRansomware: Medusa Ransomware (AA25-071A). https://www.cisa.gov/news-events/cybersecurity-advisories/aa25-071a

[**13**] European Data Protection Supervisor (EDPS). (n.d.). Data minimisation (glossary). https://www.edps.europa.eu/data-protection/data-protection/glossary/d_en

[**14**] Rose, S., Borchert, O., Mitchell, S., & Connelly, S. (2020). Zero Trust Architecture (NIST Special Publication 800-207). National Institute of Standards and Technology. https://nvlpubs.nist.gov/nistpubs/specialpublications/NIST.SP.800-207.pdf